USING MACHINE LEARNING TECHNIQUES TO ASSESS USE DECISIONS OF THAILAND'S FUTURE HIGH-SPEED RAIL

CHINNAKRIT BANYONG

มหาวิทยาลัยเทคโนโลยีสุรนารี

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Industrial and Logistics Management Engineering
Suranaree University of Technology
Academic Year 2025

การใช้เทคนิคการเรียนรู้ของเครื่องในการประเมินการตัดสินใจใช้รถไฟความเร็วสูงในอนาคตของประเทศไทย

นายชินกฤต บรรยงค์

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมและโลจิสติกส์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2568

USING MACHINE LEARNING TECHNIQUES TO ASSESS USE DECISIONS OF THAILAND'S FUTURE HIGH-SPEED RAIL

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

....

(Prof. Dr. Thaned Satiennam)
Chairperson

....

(Assoc. Prof. Dr. Sajjakaj Jomnonkwao)
Member (Thesis Advisor)

....

(Prof. Dr. Vatanavongs Ratanavaraha)
Member

....

(Asst. Prof. Dr. Somsak Siwadamrongpong)
Member

....

(Dr. I-soon Raungratanaamporn)
Member

Yupaporn
....

(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)
Vice Rector for Academic Affairs and
Quality Assurance

....

(Assoc. Prof. Dr. Pornsiri Jongkol)
Dean of Institute of Engineering

มหาวิทยาลัยเทคโนโลยีสุรนารี

ชินกฤต บรรยงค์: การใช้เทคนิคการเรียนรู้ของเครื่องในการประเมินการตัดสินใจใช้รถไฟความเร็วสูงในอนาคตของประเทศไทย (USING MACHINE LEARNING TECHNIQUES TO ASSESS USE DECISIONS OF THAILAND'S FUTURE HIGH-SPEED RAIL) อาจารย์ที่ปรึกษา: รองศาสตราจารย์ ดร. สัจจากาจ จอมโนนเขวา, 255 หน้า.

คำสำคัญ: รถไฟความเร็วสูง/พฤติกรรมการเดินทาง/การเรียนรู้ของเครื่อง/การเลือกโหมดการเดินทาง/การทำนาย

ประเทศไทยกำลังอยู่ในช่วงเปลี่ยนผ่านด้านโครงสร้างพื้นฐานคมนาคม โดยเฉพาะการพัฒนาระบบรถไฟความเร็วสูง (High-Speed Rail: HSR) ซึ่งถือเป็นโครงการสำคัญระดับชาติที่มุ่งเสริมสร้างการเชื่อมโยงระหว่างภูมิภาค เพิ่มประสิทธิภาพของระบบขนส่งมวลชน และยกระดับศักยภาพทางเศรษฐกิจของประเทศ ความร่วมมือระหว่างประเทศไทยกับประเทศจีนในการดำเนินโครงการ HSR ในหลายเส้นทาง เช่น สายกรุงเทพฯ–นครราชสีมา และสายเชื่อมต่อไปยังประเทศลาว ถือเป็นการเปลี่ยนแปลงเชิงโครงสร้างที่อาจส่งผลกระทบต่อรูปแบบการเดินทางของประชาชนอย่างมีนัยสำคัญ แม้ว่าแนวคิดการใช้ HSR จะประสบความสำเร็จในหลายประเทศทั่วโลก โดยแสดงให้เห็นถึงศักยภาพในการลดเวลาเดินทาง ลดมลพิษ เพิ่มความปลอดภัย และส่งเสริมการพัฒนาเมือง แต่สำหรับประเทศไทย ระบบ HSR ยังอยู่ในระยะเริ่มต้น และยังไม่มีข้อมูลเชิงประจักษ์เพียงพอในการทำความเข้าใจพฤติกรรมผู้โดยสาร การคาดการณ์ความต้องการใช้งาน รวมถึงการประเมินผลกระทบต่อระบบขนส่งเดิม จึงยังเป็นความท้าทายที่สำคัญสำหรับนักวางแผนและผู้กำหนดนโยบาย การศึกษาครั้งนี้จึงมีความสำคัญยิ่งในการเติมเต็มช่องว่างทางวิชาการและนโยบาย โดยมุ่งเน้นการวิเคราะห์พฤติกรรมการเลือกวิธีการเดินทางของประชาชนในบริบทของการพัฒนาระบบ HSR ในประเทศไทย โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และแบบสอบถามเชิงสมมุติ (Stated Preference: SP) เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการตัดสินใจของผู้โดยสาร และแนวโน้มการเปลี่ยนรูปแบบการเดินทาง (Modal Shift) จากระบบขนส่งเดิม โดยการวิจัยแบ่งออกเป็น 3 การศึกษาย่อย ดังนี้

การศึกษาที่ 1 มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการตัดสินใจเลือกใช้บริการรถไฟความเร็วสูงในประเทศไทย และเปรียบเทียบความแม่นยำของโมเดลการพยากรณ์ ได้แก่ Binary Logit Model (BL) และโมเดล Machine Learning ได้แก่ XGBoost, LightGBM และ CatBoost โดยใช้ข้อมูลจากแบบสอบถาม Stated Preference จำนวน 3,200 ตัวอย่างจาก 16 จังหวัด ผลการศึกษาแสดงว่า CatBoost เป็นโมเดลที่มีประสิทธิภาพสูงสุดในแง่ของความแม่นยำ (Accuracy =

มหาวิทยาลัยเทคโนโลยีสุรนารี

0.8853) และค่าพื้นที่ใต้โค้ง (AUC = 0.9584) ปัจจัยสำคัญที่ส่งผลต่อการเลือกเดินทางด้วยรถไฟความเร็วสูง ได้แก่ เวลาเดินทาง ค่าใช้จ่าย ระยะเวลาเข้าถึงสถานี ความถี่ของบริการ และรายได้ของผู้โดยสาร การใช้ Machine Learning ช่วยให้เข้าใจพฤติกรรมผู้โดยสารได้อย่างลึกซึ้ง และสามารถนำไปใช้ประกอบการกำหนดนโยบายด้านการคมนาคมที่เหมาะสม เพื่อสนับสนุนการใช้งานรถไฟความเร็วสูงในอนาคตของประเทศไทย

การศึกษาที่ 2 มีวัตถุประสงค์เพื่อศึกษาผลกระทบของรถไฟความเร็วสูง (High-Speed Rail: HSR) ต่อการเปลี่ยนแปลงส่วนแบ่งตลาดของรูปแบบการเดินทางระหว่างเมืองในประเทศไทย โดยใช้เทคนิค Machine Learning โดยเฉพาะ CatBoost เปรียบเทียบกับโมเดลเศรษฐมิติดั้งเดิมอย่าง Multinomial Logit (MNL) ผลการศึกษาใช้ข้อมูลจากผู้ตอบแบบสอบถามจำนวน 3,200 คน พบว่า HSR มีศักยภาพในการครองตลาดถึง 88.91% โดยลดส่วนแบ่งของรถโดยสาร รถไฟ และเครื่องบินลงอย่างมีนัยสำคัญ ปัจจัยที่มีอิทธิพลมากที่สุดต่อการเลือกเดินทางด้วย HSR ได้แก่ ค่าโดยสาร ความถี่บริการ เวลาเดินทาง ระยะเวลารอ และเวลาเข้าถึงสถานี ผลการวิเคราะห์ SHAP ช่วยให้เข้าใจลำดับความสำคัญของแต่ละปัจจัย ซึ่งสามารถนำไปประยุกต์ใช้ในการกำหนดนโยบายขนส่ง เช่น การกำหนดโครงสร้างราคาค่าโดยสารที่เหมาะสม พัฒนาคุณภาพบริการ และปรับปรุงการเข้าถึงสถานีเพื่อส่งเสริมให้เกิดการเปลี่ยนผ่านสู่ระบบขนส่งที่มีประสิทธิภาพและยั่งยืน

การศึกษาที่ 3 มีวัตถุประสงค์เพื่อการคาดการณ์พฤติกรรมการเลือกวิธีการเดินทางของประชาชนในบริบทของการพัฒนารถไฟความเร็วสูงในประเทศไทย โดยเปรียบเทียบประสิทธิภาพของโมเดล Machine Learning และ Deep Learning ได้แก่ Multinomial Logit (MNL), XGBoost, LightGBM, CatBoost, Deep Neural Network (DNN) และ Convolutional Neural Network (CNN) จากข้อมูลแบบสอบถาม Stated Preference จำนวน 3,200 รายการ ผลการวิเคราะห์พบว่า CatBoost มีประสิทธิภาพสูงสุด (AUC = 0.9113, Accuracy = 0.7557) ปัจจัยที่ส่งผลมากที่สุดได้แก่ ค่าโดยสาร ความถี่ของบริการ และเวลารอ ข้อมูลถูกวิเคราะห์เพิ่มเติมด้วย SHAP เพื่ออธิบายผลการพยากรณ์ของโมเดลอย่างโปร่งใสและเข้าใจง่าย ผลลัพธ์ชี้ให้เห็นว่าเทคนิค Machine Learning และ Deep Learning เหมาะสมในการวางแผนระบบขนส่งสมัยใหม่ และสามารถนำไปใช้เพื่อพัฒนานโยบายด้านโครงสร้างพื้นฐานขนส่งอย่างยั่งยืนในประเทศกำลังพัฒนา

มหาวิทยาลัยเทคโนโลยีสุรนารี

สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมและโลจิสติกส์ ลายมือชื่อนักศึกษา..

ปีการศึกษา 2568 ลายมือชื่ออาจารย์ที่ปรึกษา..

CHINNAKRIT BANYONG: USING MACHINE LEARNING TECHNIQUES TO ASSESS
USE DECISIONS OF THAILAND'S FUTURE HIGH-SPEED RAIL.
THESIS ADVISOR: SAJJAKAJ JOMNONKWAO, Ph.D. 255 PP.

Keyword: high-speed rail, travel behavior, machine learning, travel mode choice, prediction

Thailand is undergoing a pivotal transformation in its transportation infrastructure, with the development of high-speed rail (HSR) positioned as a key national initiative. This strategic effort is aimed at strengthening interregional connectivity, enhancing the efficiency of public transit systems, and driving long-term economic growth. In particular, the cooperation between Thailand and China on major HSR projects—such as the Bangkok–Nakhon Ratchasima line and its extension to Laos—marks a structural shift with significant implications for national travel behavior and modal choices. Although HSR has been successfully implemented in numerous countries, demonstrating benefits such as reduced travel times, lower emissions, improved safety, and urban development stimulation, Thailand remains in the nascent stages of adopting this technology. The country currently lacks sufficient empirical evidence regarding passenger preferences, travel demand forecasts, and the potential impact of HSR on existing modes of transport. These knowledge gaps present a pressing challenge for transportation planners and policymakers, who require robust data to support informed decision-making and policy development. In response, this research seeks to bridge these gaps by analyzing travel mode choice behavior within the context of Thailand’s planned HSR system. Leveraging machine learning (ML) techniques and stated preference (SP) survey data, the study investigates the factors influencing passengers’ travel decisions and assesses the likelihood of a modal shift from traditional transportation to HSR. The study is structured into three interrelated sub-studies, each contributing to the overarching goal of generating predictive behavioral insights through advanced data-driven methodologies.

Study 1 examines the key determinants influencing the choice to use HSR in Thailand and compares the predictive accuracy of the Binary Logit Model (BL) with

มหาวิทยาลัยเทคโนโลยีสุรนารี

three ML models: XGBoost, LightGBM, and CatBoost. Using SP data from 3,200 respondents across 16 provinces, the results show that CatBoost achieved the highest performance, with an accuracy of 0.8853 and an area under the curve (AUC) of 0.9584. Influential variables included travel time, cost, access time to stations, service frequency, and household income. These findings highlight the value of ML in enhancing the understanding of passenger decision-making processes and in supporting more effective transport policy design.

Study 2 evaluates the impact of HSR on the intercity travel market in Thailand by analyzing modal share shifts using the CatBoost model in comparison with the conventional Multinomial Logit (MNL) model. Based on the same dataset, findings indicate that HSR could capture up to 88.91% of the market share, significantly displacing bus, traditional rail, and domestic air services. Key factors influencing this shift include fare levels, service frequency, travel time, waiting time, and station accessibility. SHAP (SHapley Additive exPlanations) analysis was employed to interpret the relative importance of these variables, offering actionable insights for pricing strategies, service enhancements, and station accessibility improvements.

Study 3 focuses on predicting travel behavior using both traditional and deep learning models, including MNL, XGBoost, LightGBM, CatBoost, Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN). Utilizing the same SP dataset, CatBoost again demonstrated superior predictive performance, with an AUC of 0.9113 and an accuracy of 0.7557. Fare, service frequency, and waiting time emerged as the most significant predictors. SHAP analysis provided additional transparency and interpretability of model outputs, underscoring the suitability of ML and DL techniques for modern transportation modeling and infrastructure planning in developing economies.

มหาวิทยาลัยเทคโนโลยีสุรนารี

School of Industrial and Logistics Management Engineering Student's Signature........................
Academic Year 2025 Advisor's Signature........................

ACKNOWLEDGEMENT

This thesis, consisting of three integrated studies on travel behavior and mode choice in the context of High-Speed Rail (HSR) development in Thailand, would not have been possible without the support and guidance of many individuals.

I am profoundly grateful to Assoc. Prof. Dr. Sajjajak Jomnonkwao, my thesis advisor, for his exceptional mentorship, insightful guidance, and unwavering encouragement throughout this academic journey. His expertise in transportation planning and behavioral modeling served as the foundation for this research.

I would also like to express sincere appreciation to the Thesis Examining Committee for their valuable time and constructive feedback. Special thanks go to Prof. Dr. Thaned Satiennam (Chairperson), Prof. Dr. Vatanavongs Ratanavaraha, Asst. Prof. Dr. Somsak Sivalerttanapong, and Dr. I-soon Ruangratanamon. Their expertise significantly enhanced the academic quality of this thesis.

My heartfelt thanks go to Suranaree University of Technology for granting the doctoral scholarship that supported the completion of this research.

I also gratefully acknowledge all respondents from 16 provinces who contributed to the stated preference surveys. Their participation was essential to the data collection process.

Lastly, I would like to thank my family, friends, and colleagues for their constant support and belief in me. This thesis is dedicated to all who contribute to sustainable transport development in Thailand and beyond.

CHINNAKRIT BANYONG

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS

Page

CHAPTER

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

Page

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

มหาวิทยาลัยเทคโนโลยีสุรนารี

TABLE OF CONTENTS (Continued)

Page

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF TABLES

มหาวิทยาลัยเทคโนโลยีสุรนาร

LIST OF TABLES (Continued)

Table **Page**

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF FIGURES

Figure Page

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF FIGURES (Continued)

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF ABBREVIATIONS

β	Estimated coefficient
SD	Standard deviation
SK	Skewness
KU	Kurtosis
AUC	Area under the curve
ROC	Receiver Operating Characteristic
S.E.	Standard Error
Sig.	Significance (p-value)
Exp(β)	Exponentiated Coefficient

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER I

INTRODUCTION

1.1 Rationale for research

1.1.1 Excursion

Over the past few decades, Thailand has continuously encountered structural challenges in its intercity transportation systems, largely stemming from the pressures of rapid urbanization, sustained economic growth, and dynamic changes in socio-spatial configurations. These transformations have contributed to a significant surge in travel demand, especially among major urban corridors such as Bangkok, Chiang Mai, and the northeastern provinces. Despite this increasing demand, existing public transport modes—namely, intercity buses and conventional rail—remain plagued by chronic inefficiencies, service delays, inadequate comfort, and safety limitations. Compounding these issues is the nation’s growing reliance on private automobiles, which not only exacerbates traffic congestion but also contributes to rising carbon emissions and deteriorating urban environmental conditions. Meanwhile, domestic air travel is increasingly constrained by saturated airport capacities and bottlenecks in high-demand routes.

To mitigate these systemic problems, the Thai government has launched a strategic initiative to develop a high-speed rail (HSR) network, with a flagship project focusing on the Bangkok–Nakhon Ratchasima–Nong Khai corridor. This corridor is a critical component of the broader regional connectivity vision that aims to link Thailand with the Lao People's Democratic Republic and the southern region of China. Far beyond its implications as a transportation megaproject, the HSR system is envisioned as a transformative instrument for fostering inclusive economic growth, promoting regional tourism, and addressing the geographic disparities in infrastructure access across Thailand’s secondary cities and rural provinces.

มหาวิทยาลัยเทคโนโลยีสุรนารี

The introduction of HSR is therefore expected to play a pivotal role in reshaping the national transportation landscape. It offers not only a faster and more reliable travel alternative but also the potential to enhance regional economic competitiveness by improving accessibility, reducing travel time, and stimulating investment in surrounding areas. Furthermore, by encouraging a modal shift away from fossil fuel-dependent transport modes, the HSR system could contribute to Thailand's commitments under global climate agreements and long-term sustainability goals. In this context, high-speed rail is not merely a technological advancement but a socio-economic catalyst with far-reaching implications.

However, while the anticipated introduction of HSR is expected to dramatically reshape Thailand's intercity travel landscape—particularly in redistributing modal market shares away from buses, conventional trains, and short-haul flights—it also raises complex challenges in behavioral forecasting. Predicting how travelers will respond to this new transport option involves navigating a highly uncertain landscape influenced by social, economic, technological, and psychological variables. Understanding this behavioral shift is essential for effective planning, pricing strategies, infrastructure investment, and policy development.

Traditional econometric models, though widely adopted in the transportation literature, are often limited by assumptions of linearity, independence of irrelevant alternatives, and their inability to process high-dimensional, non-linear, or imbalanced datasets. These limitations become especially pronounced in the context of developing countries like Thailand, where transportation behavior is deeply influenced by regional disparities in income, education, urbanization, and transport service availability. The heterogeneity of traveler profiles—ranging from daily commuters and intercity workers to tourists and rural migrants—adds another layer of complexity that traditional models struggle to address.

Consequently, there is a growing need to adopt more advanced, data-driven analytical approaches—such as machine learning (ML) and deep learning (DL) techniques—that can more accurately capture the multifaceted nature of travel mode decision-making in emerging contexts. These methods are capable of identifying

มหาวิทยาลัยเทคโนโลยีสุรนารี

complex, nonlinear relationships among variables and can process large volumes of data from various sources, including survey data, smart card usage, GPS trajectories, and mobile applications. Moreover, ensemble learning models and neural networks have shown particular promise in improving predictive accuracy for discrete choice problems, outperforming classical methods such as multinomial logit or probit models.

By integrating ML and DL techniques into travel demand forecasting, researchers and policymakers can gain deeper insights into the factors that influence traveler preferences, willingness to switch modes, and sensitivity to price and travel time. These insights are invaluable for scenario analysis, demand management, and evaluating the broader socio-economic impacts of high-speed rail investments. Ultimately, a better understanding of future travel behavior through advanced analytics not only enhances academic knowledge but also strengthens the evidence base for infrastructure development and sustainable transport policy in Thailand.

1.1.2 The Importance of Applying Machine Learning to Understand Passenger Behavior in High-Speed Rail Contexts in Thailand

Thailand has embarked on the development of a high-speed rail (HSR) system in collaboration with China as part of its national transportation infrastructure strategy. This initiative aims to transform the country into a regional hub for mobility and logistics within the ASEAN region (P. Tissayakorn, 2021). The HSR project marks a major step toward upgrading the national transportation network, especially in a context where the public transport system is still undergoing transition and development. However, infrastructure investment alone is not sufficient without a clear understanding of user behavior—particularly in terms of travel mode choice, which reflects the population’s readiness and willingness to adopt HSR. International studies have highlighted the multifaceted benefits of HSR, such as travel time reduction, economic stimulation, improved regional connectivity, and lower environmental impacts (S. A. Kashifi, S. Jamal, M. Kashefi, M. Almoshaogeh, & M. A. Rahman, 2022; Y. Sun, X. Wang, L. Liu, & Y. Wang, 2021). Yet, empirical insights into the travel intentions and attitudes of future users in Thailand remain limited. Moreover, traditional methods for analyzing travel behavior, such as Binary Logit and Multinomial

Logit models, often struggle to capture complex, nonlinear interactions between variables. This limitation creates the need for advanced analytical approaches. Recent research has demonstrated the superior predictive power of machine learning (ML) techniques, especially boosting models such as XGBoost, in modeling travel mode choice (A. Abulibdeh, 2023; F. Wang & C. Ross, 2018). These models, when combined with interpretability techniques like SHAP, can also provide transparency by identifying key decision-making factors (Y. Cheng, X. Chen, J. De Vos, X. Lai, & F. Witlox, 2019; J. Díaz-Ramírez, N. Estrada-García, & M. Figueroa-Sayago, 2023).

Accordingly, the rationale for this study lies in the urgent need to assess the potential of HSR to reshape travel behavior in Thailand, using robust and interpretable machine learning models. The insights gained are expected to support the design of evidence-based transportation policies and strategic planning for sustainable HSR adoption in the near future.

1.1.3 The Importance of Assessing the Impact of High-Speed Rail on Passenger Travel Behavior in Thailand

Thailand is currently facing challenges in its intercity transportation system, which lacks efficiency in several aspects, including long travel times, limited service frequency, high travel costs, and unequal accessibility. The country’s primary transport modes—intercity buses, conventional rail, and domestic airlines—are widely used, yet they still fall short of fully meeting the needs of the traveling public (Office of, Traffic, & Planning, 2024).

To support population and economic growth and to upgrade national infrastructure, the Thai government has launched the High-Speed Rail (HSR) project, particularly the Bangkok–Nakhon Ratchasima–Nong Khai route. This initiative aims to enhance regional logistics connectivity and reduce dependence on air travel and private vehicles. The project is expected to significantly shorten travel times, increase convenience, and stimulate economic decentralization to regional areas (Banyong et al., 2024). While international studies have widely examined the impacts of HSR on market share and travel behavior—such as those conducted in Europe, China, and the Middle East (Ammar Abulibdeh, 2023; N. Avogadro, E. Pels, & R. Redondi, 2023; F. Xu, Liu, Zheng, Cao, & Yang, 2022)—empirical research analyzing the effects of HSR in Thailand remains limited. In particular, there is a lack of evidence-based insights into how travelers may shift their behavior once HSR services are operational.

Most previous studies in Thailand have relied on traditional econometric models, such as the multinomial logit (MNL) model, which have limitations in capturing the complex, nonlinear relationships in human travel behavior. In contrast, emerging techniques such as machine learning—especially the CatBoost algorithm—have shown great potential in international studies due to their ability to efficiently handle categorical and imbalanced data (Banyong et al., 2024; X. Zhao, Yan, Yu, & Van Hentenryck, 2020).

Given this context, the present study aims to analyze the impact of HSR on travel mode choice in Thailand by applying CatBoost in combination with Shapley Additive Explanations (SHAP). This approach identifies key factors influencing travel decisions and forecasts the potential shifts in market share among buses, trains, airplanes, and HSR. The findings are expected to provide valuable policy insights for enhancing the efficiency and sustainability of Thailand's transportation system.

1.1.4 Evaluating Intercity Travel Behavior in Response to HSR in Thailand: Insights from Machine Learning and Econometric Models

Rapid urbanization and the increasing demand for intercity travel in Thailand have led to congestion and inefficiencies in the existing public transportation system. One proposed solution to address these challenges is the development of a high-speed rail (HSR) network, which is expected to reduce travel time, enhance regional connectivity, and stimulate economic growth (C. Blanquart & M. Koning, 2017; Liang et al., 2020). Thailand's HSR project, developed in collaboration with China, primarily focuses on the Bangkok–Nong Khai corridor and aligns with the Sustainable Development Goals (SDGs), particularly in areas related to infrastructure development, sustainable cities, and carbon emission reduction (Office of the National & Social Development, 2025).

However, analyzing travel mode choice behavior in response to HSR remains a complex task due to the interplay of multiple factors such as travel cost, duration, and station accessibility (Y. Deng, Y. Bai, L. Cui, & R. He, 2023; Y. Zhang, K. Liang, E. Yao, & M. Gu, 2024). While the Multinomial Logit (MNL) model has been widely adopted in travel behavior studies, its limitations in capturing nonlinear relationships and flexibility are well recognized (Akter & Alam, 2024; M. E. Ben-Akiva & S. R. Lerman, 1985).

To address these challenges, this study employs machine learning (ML) and deep learning (DL) techniques alongside traditional econometric models to analyze travel mode choice in the context of HSR development. The analysis is based on stated preference (SP) survey data, and SHapley Additive exPlanations (SHAP) are utilized to improve model interpretability and identify key influencing factors (V. Dahmen, S. Weikl, & K. Bogenberger, 2024).

1.2 Purpose of Research

The research aims to achieve the following objectives:

1) To examine the effectiveness of machine learning methods and analyze the factors associated with future high-speed rail adoption in Thailand.

2) To propose policy recommendations that support the adoption of high-speed rail in the future.

3) To analyze the projected modal shift in transportation choices within Thailand.

4) To compare the predictive performance of machine learning techniques between ensemble learning and deep learning approaches in analyzing travel mode choice factors in Thailand.

1.3 Scope of Research

The research scope encompasses the following aspects:

1) Population: The target population includes Thai passengers aged 18 years and older who travel between provinces.

มหาวิทยาลัยเทคโนโลยีสุรนารี

2) Study Areas: The study covers 16 provinces across four main regions of Thailand—North, Northeast, Central, and South—focusing on areas with proposed or planned HSR stations.

3) Research Focus: The research focuses on analyzing intercity travel mode choice behavior using stated preference (SP) survey data.

1.4 Research Questions

The research seeks to address the following questions:

1) To what extent can machine learning methods effectively predict the factors influencing the future adoption of high-speed rail in Thailand?

2) What are the key factors that influence the decision to adopt high-speed rail in Thailand in the future?

3) How is the proportion of travel mode usage in Thailand expected to change following the introduction of high-speed rail?

4) How do ensemble learning and deep learning techniques differ in their ability to analyze factors influencing travel mode choice in Thailand?

1.5 Contribution of Research

The research is expected to make the following contributions:

1) The development of effective machine learning models for predicting factors influencing the future adoption of high-speed rail in Thailand, which can inform transport policy-making and strategic planning.

2) Forecasting potential changes in travel mode shares following the introduction of high-speed rail, which can assist in transportation infrastructure planning and investment.

3) A comparative evaluation of the performance of ensemble learning and deep learning methods in analyzing travel mode choice factors, providing valuable insights for practitioners to select appropriate analytical techniques in transportation studies.

มหาวิทยาลัยเทคโนโลยีสุรนารี

1.6 Organization of Research

The research comprises five chapters, structured as follows:

1) Chapter 1: Research Principles and Justification, including Background Information, Research Objectives, Scope, Research Questions, and Contributions.

2) Chapter 2: A Machine Learning Comparison of Transportation Mode Changes from High-Speed Railway Promotion in Thailand

3) Chapter 3: Analyzing High-Speed Rail's Transformative Impact on Public Transport in Thailand Using Machine Learning

4) Chapter 4: Machine Learning-Based Analysis of Travel Mode Preferences: Neural and Boosting Model Comparison Using Stated Preference Data from Thailand's Emerging High-Speed Rail Network

5) Chapter 5: Summary and Analysis of the Three Studies (Chapters 2–4).

1.7 Reference

Abulibdeh, A. (2023a). Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis. *Transportation Research Interdisciplinary Perspectives, 20*, 100852. doi:10.1016/j.trip.2023.100852

Abulibdeh, A. (2023b). Machine learning models for travel behavior analysis in Qatar: Comparing XGBoost, MNL, and BL models. *Transportation Research Record, 2677*(3), 245-260.

Akter, T., & Alam, B. M. (2024). Travel mode choice behavior analysis using multinomial logit models towards creating sustainable college campus: A case study of the University of Toledo, Ohio. *Frontiers in Future Transportation, 5*, 1389614.

Avogadro, N., Pels, E., & Redondi, R. (2023). Policy impacts on the propensity to travel by HSR in the Amsterdam–London market. *Socio-Economic Planning Sciences, 87B*, 101585. doi:10.1016/j.seps.2023.101585

มหาวิทยาลัยเทคโนโลยีสุรนารี

Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Seefong, M., . . . Jomnonkwao, S. (2024). A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand. *Results in Engineering, 24*, 103110. doi:https://doi.org/10.1016/j.rineng.2024.103110

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Blanquart, C., & Koning, M. (2017). The local economic impacts of high-speed railways: Theories and facts. *European Transport Research Review, 9*, 12.

Cheng, Y., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying random forest to analyze household car ownership decision: A comparison of performance and interpretation with multinomial logit model. *Transportation Research Part A: Policy and Practice, 125*, 89-101.

Dahmen, V., Weikl, S., & Bogenberger, K. (2024). Interpretable machine learning for mode choice modeling on tracking-based revealed preference data. *Transportation Research Record, 2678*, 2075-2091.

Deng, Y., Bai, Y., Cui, L., & He, R. (2023). Travel Mode Choice Behavior for High-Speed Railway Stations Based on Multi-Source Data. *Transportation Research Record, 2677*, 525-540.

Díaz-Ramírez, J., Estrada-García, N., & Figueroa-Sayago, M. (2023). Machine learning for mode choice prediction: A comparison of models in Mexico. *Transportation Research Part C, 148*, 103970-103970.

Kashifi, S. A., Jamal, S., Kashefi, M., Almoshaogeh, M., & Rahman, M. A. (2022). Evaluating high-speed rail impact on modal shift and environmental benefits. *Environmental Science and Pollution Research, 29*, 40926-40940.

Liang, Y., Zhou, K., Li, X., Zhou, Z., Sun, W., & Zeng, J. (2020). Effectiveness of high-speed railway on regional economic growth for less developed areas. *Journal of Transport Geography, 82*, 102621.

Office of, T., Traffic, P., & Planning. (2024). *Proportion of Public Transportation Usage for Intercity Travel (2017-2022)*. Retrieved from

มหาวิทยาลัยเทคโนโลยีสุรนารี

https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45

Office of the National, E., & Social Development, C. (2025). Sustainable Development Goals: SDGs. Retrieved from https://sdgs.nesdc.go.th/goals-and-indicators/

Sun, Y., Wang, X., Liu, L., & Wang, Y. (2021). High-speed rail and urban development in China: Evidence from multiple indicators. *Journal of Urban Planning and Development, 147*(4), 05021047-05021047.

Tissayakorn, P. (2021). Thailand's high-speed rail progress and regional integration. *ASEAN Transportation Review, 14*(2), 55-68.

Wang, F., & Ross, C. (2018). Machine learning travel mode choice modeling: A comparison with traditional logit models. *Transportation Research Record, 2672*(47), 1-11.

Xu, F., Liu, Q., Zheng, X., Cao, L., & Yang, M. (2022). Research on the impact of China's high-speed rail opening on enterprise market power: Based on the perspective of market segmentation. *Transport Policy, 128*, 121-137. doi:https://doi.org/10.1016/j.tranpol.2022.09.019

Zhang, Y., Liang, K., Yao, E., & Gu, M. (2024). Measuring Reliable Accessibility to High-Speed Railway Stations by Integrating the Utility-Based Model and Multimodal Space–Time Prism under Travel Time Uncertainty. *ISPRS International Journal of Geo-Information, 13*, 263.

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society, 20*, 22-35. doi:10.1016/j.tranpol.2022.09.019

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER II

A MACHINE LEARNING COMPARISON OF TRANSPORTATION MODE CHANGES FROM HIGH-SPEED RAILWAY PROMOTION IN THAILAND

2.1 Abstract

Thailand's collaboration with China to develop High-Speed Rail (HSR) represents a crucial step in enhancing transportation infrastructure and promoting regional economic growth. While most research has focused on existing travel modes such as cars, buses, and planes, there is a notable lack of analysis on future transport choices, particularly in developing countries like Thailand. This study addresses this gap by analyzing the factors influencing High-Speed Railways adoption to inform future traveler decisions. The comparison of models for predicting High-Speed Railways usage revealed that CatBoost consistently outperformed the other models, with cross-validation confirming its superior performance across all key metrics. The Binary Logit Model (BL) demonstrated moderate effectiveness, achieving an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a relatively lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were moderate, but it struggled with handling imbalanced data. In contrast, XGBoost delivered significantly stronger results, with an accuracy of 0.8846 ± 0.0068, sensitivity of 0.9210 ± 0.0094, and an AUC of 0.9583 ± 0.0036. XGBoost also achieved high specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and an F1 score (0.8910 ± 0.0067). LightGBM also performed well, achieving an accuracy of 0.8763 ± 0.0077 and a relatively high sensitivity of 0.9358 ± 0.0071. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. CatBoost demonstrated the highest overall performance, achieving an accuracy of 0.8853 ± 0.0061, sensitivity of 0.9245 ± 0.0072, specificity of 0.8441 ± 0.0068, precision of 0.8616 ± 0.0058, an F1 score of 0.8920 ± 0.0058, and an AUC of 0.9584 ± 0.0034. The decision to use

มหาวิทยาลัยเทคโนโลยีสุรนารี

High-Speed Railways in Thailand is influenced by several key factors that reflect the country's unique transportation context. These factors include travel time, access time, service frequency, cost, waiting time, household income, automobile ownership, gender, and the purpose of travel. This information is valuable for shaping policies to support High-Speed Railways adoption in Thailand.

2.2 Introduction

Thailand and China initially agreed to collaborate on developing railway infrastructure in Thailand as part of the country's transportation infrastructure strategy for 2015–2022. In 2022, the agreement was renewed to continue the development of high-speed rail projects. Thailand, as an ASEAN transport hub, is developing the Thailand–China high-speed railway to enhance its railway network and boost regional economic growth, though the project is still incomplete (K. Tissayakorn, 2021). High-Speed Railways are not only critical for improving transportation infrastructure but also directly influence people's travel choices, as they provide an alternative mode of transport that competes with traditional travel methods, such as buses, airplanes, and private cars. This competition has the potential to reshape traveler behavior, especially in developing nations like Thailand, where transportation options are limited.

While most studies have focused on current travel modes such as private cars, trains, buses, and planes, there is a notable lack of analysis on future transportation mode choices, especially in developing nations such as Thailand, which plans to introduce a High-Speed Railways —an unprecedented mode of travel in the country.

High-Speed Railways are the primary infrastructure for transporting passengers and are used in many countries (Jin, Lin, Shi, Lee, & Li, 2020). Research indicates that High-Speed Railways considerably benefit a country's economic development, particularly in underdeveloped regions (Aalami, Anari, Shafighfard, & Talatahari, 2013). High-Speed Railways reduce travel time, improve labor market efficiency, and foster economic connections, while also lowering road accidents and pollution (M. T. Kashifi, A. Jamal, M. S. Kashefi, M. Almoshaogeh, & S. M. Rahman, 2022; Liang et al., 2020; W. Sun, C. Wang, C. Liu, & L. Wang, 2021). These benefits contribute to shifts in traveler preferences and mode choices, as individuals increasingly favor quicker, safer, and more environmentally friendly travel alternatives. Effective transportation planning is necessary to achieve these benefits, including evaluating travel options (Chang et al., 2019). In Thailand, where High-Speed Railways is yet to be implemented, understanding traveler attitudes is critical for assessing its future impact (Heinen, Harshfield, Panter, Mackett, & Ogilvie, 2017; Shasha Liu, Yao, Cheng, & Zhang, 2017). By influencing factors such as travel time, cost, and convenience, High-Speed Railways could potentially shift a significant portion of the population from existing travel modes to High-Speed Railways, necessitating government policies to encourage such a transition once the High-Speed Railways is operational. High-Speed Railways are expected to compete with or replace airlines, buses, and private cars, requiring government policies that promote their use after construction.

Several factors influence travel mode choices, such as demographics, travel attributes (distance, cost, convenience), safety, and land use (Belgiawan, Ilahi, & Axhausen, 2019; Cui, Li, Zhu, & Ma, 2023; Gross & Grimm, 2018; M. T. Kashifi et al., 2022; Xiaowei Li et al., 2024; Xiaowei Li, Tang, Hu, & Wang, 2020; Mirzaei, Kheyroddin, & Mignot, 2021; Nurhidayat, Widyastuti, Sutikno, & Upahita, 2023; Yang, Chen, & Yang, 2022). Traditional models like binary logit (BL) and multinomial logit (MNL) have been widely used to study travel behavior (Gunay & Gokasar, 2021; Y. Jiang, Yu, Guan, Gao, & Feng, 2021; Ning, Lyu, & Wang, 2021; Pasha, Hickman, & Prato, 2020; Sharma et al., 2021), but recent research shows that machine learning (ML) techniques can handle complex relationships more effectively (Sarker, 2022). M. T. Kashifi et al. (2022) found that light gradient boosting decision trees outperformed traditional models, while Ammar Abulibdeh (2023) demonstrated that XGBoost was more efficient than Binary Logit (BL) and MNL in predicting metro mode choice. F. Wang and C. L. Ross (2018) confirmed that XGBoost outperformed MNL, and Xiaowei Li et al. (2024) used XGBoost to analyze High-Speed Railways choices in China. These studies support the growing use of machine learning for travel mode analysis, especially with ensemble learning techniques (Ammar Abulibdeh, 2023; L. Cheng, X. Chen, J. De Vos, X. Lai, & F. Witlox, 2019; Darwish et al., 2024; J. Díaz-Ramírez, J. A. Estrada-García, & J. Figueroa-Sayago, 2023; Elharoun, El-Badawy, & Shahdah, 2023; M. T. Kashifi et al., 2022; Kim, 2021; Salas, De la Fuente, Astroza, & Carrasco, 2022; F. Wang & C. L. Ross, 2018; X. Zhao et al., 2020). Travel mode choice is influenced by subjective factors (e.g., age, income, car ownership) (Armah, Yawson, & Pappoe, 2010; Cascetta, Cartenì, Henke, & Pagliara, 2020; Givoni, 2006; D. Wu & Martín, 2022; X. Yu et al., 2018), area determinants (e.g., population density), and passenger perceptions (e.g., policies, and fees) (Calvo, Eboli, Forciniti, & Mazzulla, 2019; Juan de Dios Ortúzar & Willumsen, 2011; J. Li, Lo, & Guo, 2018; Pitombo, Salgueiro, da Costa, & Isler, 2015; Puan et al., 2019; Wójcik, 2019). Spatial factors like land use and travel costs also affect choices, with studies showing different preferences in the USA and Germany due to varying transit policies (Rattanakijsuntorn, Suwannarat, Samittivate, & Nithikittiwat, 2024; Sresakoolchai & Kaewunruen, 2020; Weerawat, Samitiwantikul, & Torpanya, 2020). In Thailand, public-private partnerships (PPP) play a critical role in High-Speed Railways development. Sresakoolchai and Kaewunruen (2020) highlighted how PPPs improve accessibility and cost control, promoting High-Speed Railways adoption. However, operational issues in Bangkok’s Airport Rail Link have hindered public transport use, though improvements could address these problems (Weerawat et al., 2020). Rattanakijsuntorn et al. (2024) emphasized air-rail connectivity's importance for High-Speed Railways promotion. Machine learning is increasingly used in travel mode modeling for its accuracy in handling large datasets (Abduljabbar, Dia, Liyanage, & Bagloee, 2019; Bhavsar, Safro, Bouaynaya, Polikar, & Dera, 2017; Khan, Adnan, & Iqbal, 2022). Studies by F. Wang and C. L. Ross (2018), L. Cheng et al. (2019), and X. Zhao et al. (2020) found that ML models like XGBoost, random forest, and neural networks outperform traditional models. Ammar Abulibdeh (2023) and Kim (2021) also validated XGBoost and neural networks for predicting travel modes, which are detailed in Table 2.1 and Figure 2.1.

While traditional machine learning models are robust and widely used, they often struggle with complex, nonlinear relationships and large datasets. In contrast, advanced ensemble methods like XGBoost, LightGBM, and CatBoost offer superior performance by handling high-dimensional data, improving interpretability, and reducing overfitting. This research compares the efficiency of these boosting methods in analyzing future travel mode choices in developing countries, offering insights for policy recommendations in Thailand to support High-Speed Rail adoption and assess unintroduced travel modes.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.1 Summary of selected literature on the use of machine learning for predicting travel mode choices

Authors	Regions	Statistical approaches	Input Data	Key findings	Accuracy
F. Wang and C. L. Ross (2018)	United States of America	XGBoost, MNL	Travel time, socio-demographic.	XGBoost outperformed MNL, with travel time being the most significant factor.	93.12%
Ammar Abulibdeh (2023)	Qatar	XGBoost, MNL, Binomial Logit	Travel behavior, metro introduction.	XGBoost was most effective, with trip parameters and socioeconomic factors crucial.	91.2%
Jenny Díaz-Ramírez et al. (2023)	Mexico	RF, XGBoost, GB, NB, DT, LR, SVM (polynomial), SVM (RBF)	Transport preferences, student travel data.	XGBoost showed excellent performance for mode choice estimation.	92.5%
L. Cheng et al. (2019)	China	RF, AdaBoost, and SVM	Travel instances (age, land use, traffic data)	Random Forest outperformed others with higher accuracy and lower computational cost.	91%
Elharoun et al. (2023)	Egypt	DNN, KNN, AdaBoost, SVM, NB, RF, GB, XGB, LDA, DT, QDA	Household size, income, travel time, distance	DNN achieved 97.81% accuracy, outperforming RF and XGBoost with larger sample sizes.	97.81%
Darwish et al. (2024)	Egypt	MNL, SVM, MLP, KNN, AdaBoost, DT, RF, GBDT, and XGB	Mode choice responses, Alexandria city.	GBDT achieved the highest prediction accuracy, exceeding MNL.	93.6%

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.1 Summary of selected literature on the use of machine learning for predicting travel mode choices (Continued)

Authors	Regions	Statistical approaches	Input Data	Key findings	Accuracy
X. Zhao et al. (2020)	United States of America	CART, NB, Bag, Boost, RF, SVM, and NN	Socio-economic data, travel characteristics.	Logit and ML models differed in outcomes, requiring balance between accuracy and validity.	89.5%
Kim (2021)	South Korea	ANN, XGBoost, RF	Age, activity duration, land use	XGBoost and ANN performed best, with age, trip time, and land use being key factors.	94.3%
Salas et al. (2022)	-	MNL, MMNL, ANN, SVM, RF, and XGBoost	Travel responses, socio-economic data	XGBoost, RF, NN outperformed logit models.	95.2%
M. T. Kashifi et al. (2022)	Netherland	LR, RF, DT, MLP, and LightGBDT	Trip density, income, car/bike ownership, age	LightGBDT outperform others; key factors were trip density, income, ownership, and age.	94.5%

มหาวิทยาลัยเทคโนโลยีสุรนารี

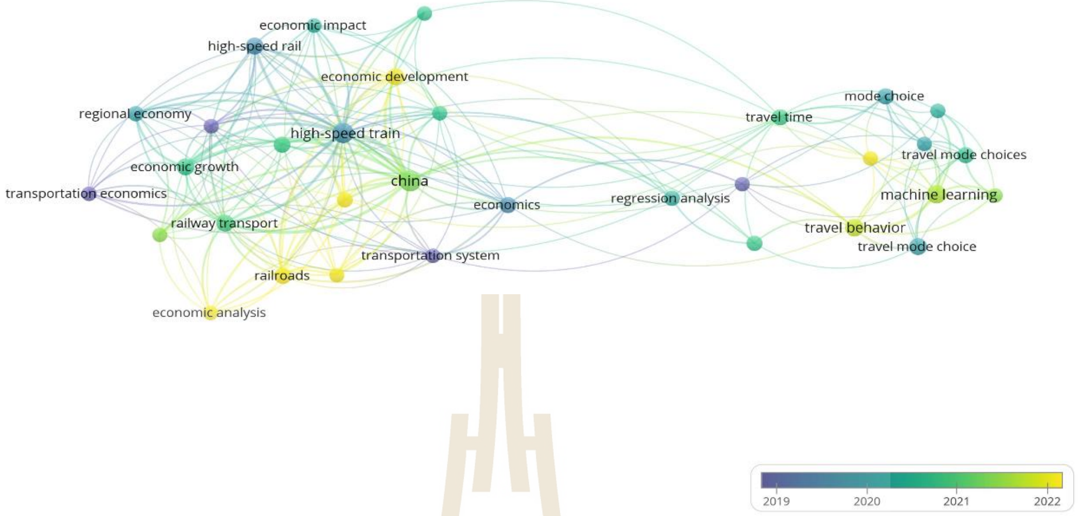


Figure 2.1 VOS viewer Network Visualization of Keywords Related

2.3 Materials and Methods

2.3.1 Stated Preference Surveys

The survey process examines the factors influencing the decision to use High-Speed Railways across four regions of Thailand: Northern, Northeastern, Southern, and Central. Each region will select an area from the province where an High-Speed Railways station is planned for construction (Department of International Trade Promotion, 2023) based on the Gross Regional and Provincial Product Chain Volume Measures for 2022 (Office of the national economic and social development council, 2023), focusing on the highest value in each of the region's top four priorities, as shown in Figure 2.2.The survey design aims to understand the factors that influenced to decision on high-speed rail (HSR) service in each region of Thailand by choosing the important economic provinces and collecting data from many critical provinces in Thailand because the economic characteristics and infrastructure of each region are different. Especially in Thailand, each province has clearly different infrastructure even the province that is located in the same region, it is distinctly difference (Srithongrung & Kriz, 2019) which directly affects the travel behavior. The survey in many provinces helps the researcher access to extensive data and reflect the various of travel behaviors at the country level (Habib, Poole, & McCarthy, 2022).

มหาวิทยาลัยเทคโนโลยีสุรนารี

Face-to-face interviews will be performed with individuals aged 18 and older at transportation stations, shopping malls, and public parks across 16 provinces in Thailand. consists of northern zone: Nakhon Sawan, Phitsanulok, Lamphun, and Chiang Mai, Northeastern zone: Nakhon Ratchasima, Khon Kaen, Udon Thani, and Nong Khai, Central zone: Bangkok, Chonburi, Rayong, and Ayutthaya, and Southern zone: Prachuap Khiri Khan, Chumphon, Surat Thani, and Songkhla. The weather of these 4 zones is tropical making Thailand is inappropriate to travel as an active journey (Al-Atroush, Mustaffa, & Sebeay, 2022) and travelers always use transportation systems with air conditioners which affects travel mode choice of intercity travel result to main vehicles like private cars, buses, planes, air conditioner trains, or high-speed rails (Xiaowei Li, Ma, Wang, & Wang, 2021; Otim, Dörfer, Ahmed, & Munoz Diaz, 2022). A total of 200 samples per province will be collected, using Convenience Sampling to ensure a diverse and efficient selection of participants. This approach guarantees balanced representation across all regions, mitigating bias and enhancing the representativeness of the findings at the national level, totaling 3,200 samples, is sufficient for machine learning analysis, as the sample size should be at least 50 to 1000 times the number of prediction classes. This guideline helps ensure that the model has enough data to perform effectively. Since there are 2 prediction classes in this research (Pavlou et al., 2024).

มหาวิทยาลัยเทคโนโลยีสุรนารี

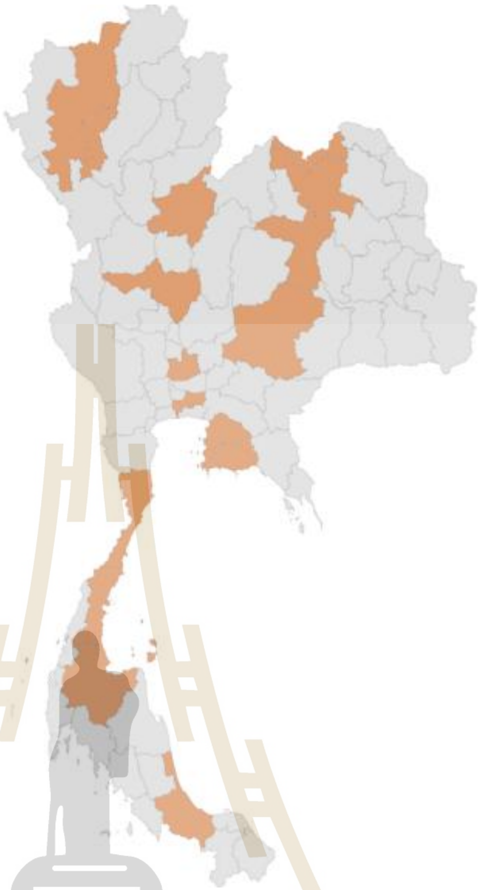


Figure 2.2 Provinces Surveyed for Data Collection in Thailand

The survey questionnaire is divided into two sections. The first section focuses on social and economic data, including gender, marital status, household size, presence of children in the family, household income, vehicle ownership, purpose of intercity travel, and travel frequency between cities. The second section comprises a stated preference (SP) questionnaire, where the stated choice (SC) experiment allows for the description of alternatives through various attribute combinations at different levels. This method helps estimate the marginal utility of each attribute for the High-Speed Railways option, enabling the assessment of in-mode choices, access times, waiting times, travel times, travel costs, and travel frequency related to the proposed mode of transportation (Arencibia, Feo-Valero, García-Menéndez, & Román, 2015; Kujala, Weckström, Mladenović, & Saramäki, 2018). The stated choice (SC) experiment required travelers to choose between alternative High-Speed Railways options and their current public transportation mode. Collecting data on travel mode

มหาวิทยาลัยเทคโนโลยีสุรนารี

preferences—taking into account factors such as access time, waiting time, travel time, travel cost, and frequency—is crucial for understanding passenger behavior and preferences. This study will reflect real-world scenarios, demonstrating that passengers who wish to reduce their travel time in the public transportation system may need to incur higher costs (Arencibia et al., 2015). These data can drive the development and enhancement of transport services by improving service frequency, optimizing routes to reduce travel time, and setting appropriate fare structures. Additionally, they assist facilitators in assessing competitiveness with other travel modes and making informed business decisions. While the stated preference (SP) questionnaire may have lower reliability owing to potential discrepancies between respondents' stated preferences and their actual behavior in real situations, it remains a widely used method for collecting relevant data on travel modes (Becker & Carmi, 2019; Birolini, Malighetti, Redondi, & Deforza, 2019; Y. Guo, Li, Anastasopoulos, Peeta, & Lu, 2021; Nam & Cho, 2020; Tarriño-Ortiz, Gómez, Soria-Lara, & Vassallo, 2022).

2.3.1.1 Statistical Design of the Choice Experiment

The attributes and their levels used in the stated choice (SC) experiment are detailed in Table 2.2. These attributes encompass various factors that influence travel mode decisions, such as access time, waiting time, travel time, travel cost, and frequency. Different levels are assigned to each attribute to represent a range of scenarios and respondent preferences. This structured approach enables a thorough analysis of how different factors affect travel decisions and the trade-offs individuals are willing to make.

In this experiment, hypothetical scenarios were based on a 700-kilometer journey from Bangkok to Chiang Mai, a route earmarked for future high-speed rail development, with attribute levels applied to this travel context in Thailand (State Railway of Thailand, 2019). As the High-Speed Railways infrastructure had not yet been constructed at the time of the study, there was no empirical data available regarding actual travel times, pricing, waiting times, service schedules (frequency of service), or operational details. Consequently, data for pricing, travel times, waiting times, and service frequency were referenced from governmental sources that provided estimates to the public through news outlets and websites in Thailand, which may still be subject to change (Economic Base, 2017; POST TO DAY, 2017). However, within the hypothetical scenarios, these estimates were adjusted to reflect a range of possible outcomes that are reasonably consistent and close to the projected figures. Waiting time was considered based on recommendations from ticket booking sources that advise passengers to arrive in advance. For other travel modes, the hypothetical scenario referenced pricing, travel times, waiting times, and service frequency from real-world sources: for buses, data was taken from national ticket booking centers(BusOnlineTicket.co.th, 2024); for trains, the estimates were primarily based on data from the State Railway of Thailand's booking system(State Railway of Thailand 2024); and for airplanes, pricing, schedules, and frequencies were referenced from airline ticketing websites(AirAsia move, 2024). The different pricing, time ranges, waiting times, and frequencies in these hypothetical scenarios were designed to reflect realistic variations across service providers. Access times were set within reasonable limits to ensure fairness between buses, trains, and high-speed rail, while access to airports was higher due to the more challenging reach compared to other transport stations in Thailand. The figures used in the simulation were within the realistic range that passengers would typically encounter. This route, with a distance suitable for competition between High-Speed Railways nd other modes of transport, provided an ideal context for analyzing the potential impacts of introducing High-Speed Railways.

In our study, we evaluated five attributes across two groups: traditional modes (bus, train, and airplane) with one level each and High-Speed Railways with two levels. A full factorial design would yield 96 choice sets $3 \times 1^5 \times 2^5$, which would be overwhelming for respondents. Instead, we used an orthogonal fractional factorial design, creating 96 profiles divided into eight sets of 12 choice sets each. As a result, each respondent only evaluated 12 of the 96 profiles during the survey (Hensher, Rose, & Greene, 2005). To ensure an efficient choice design, our approach incorporated key properties such as orthogonality, level balance, minimal overlap, and utility balance. Orthogonality was achieved by arranging attribute levels in the design correlation matrix so that any two columns were uncorrelated, thereby minimizing collinearity. Level balance was maintained by ensuring that each attribute level appeared an equal number of times across the profile sets. Utility balance was addressed by minimizing

มหาวิทยาลัยเทคโนโลยีสุรนารี

the differences in utility between alternatives, achieved by rotating the dominant alternatives in each choice set.

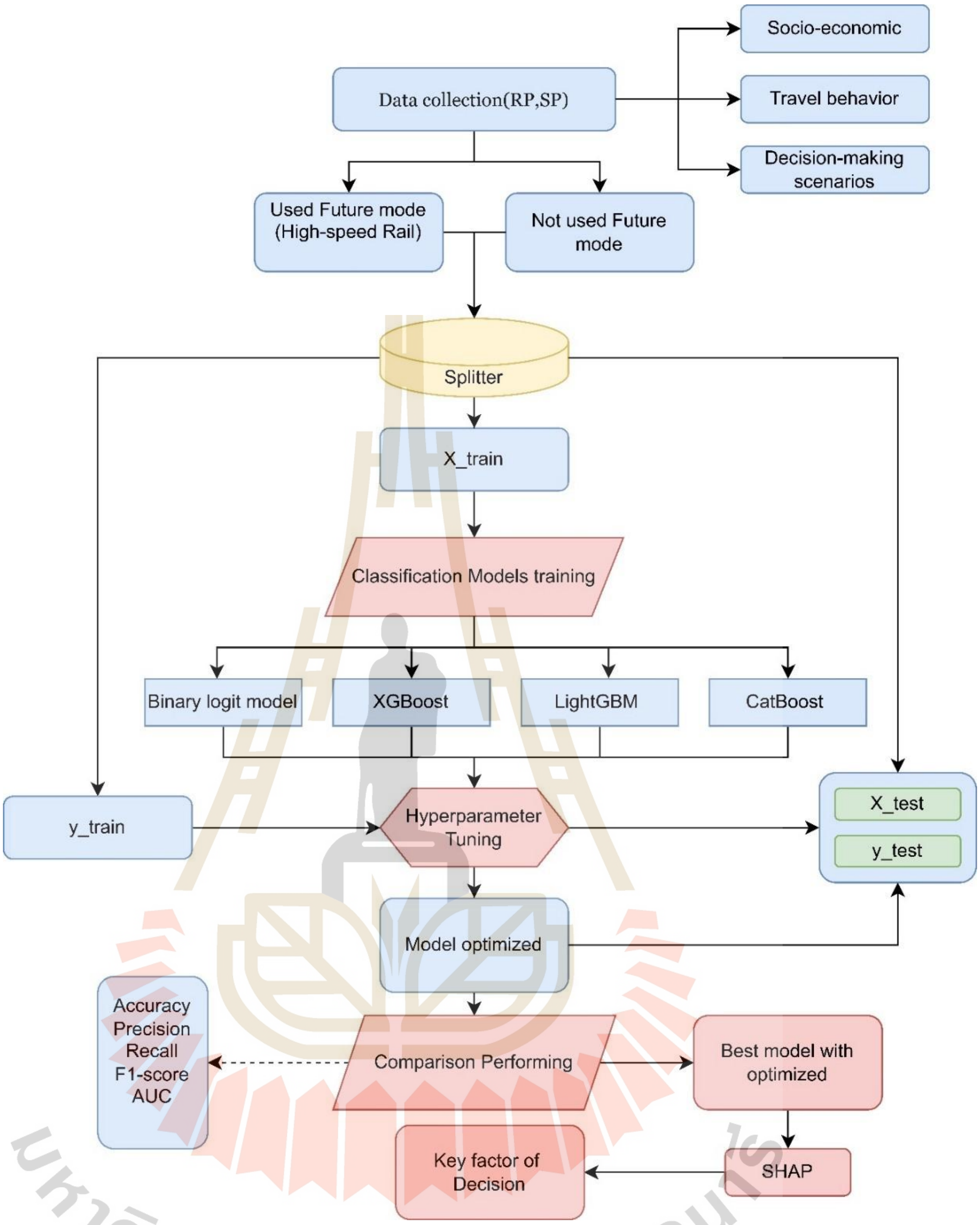
Table 2.2 Typical cards from stated choice (SC) sets

Attribute/Levels	Mode Choice 1			Mode Choice 2
	Bus	Train	Airplane	High-speed rail
access time (min)	10	10	30	10, 15
waiting time (min)	15	10	120	15, 10
travel time (min)	720	720	135	190, 220
travel cost (baht)	750	300	3000	1050, 1400
Frequency times (min)	30	150	120	190, 220

2.3.2 Methodology

The research methodology, outlined in the flowchart in Figure 2.3, begins with the data collection phase, where both Revealed Preference (RP) and Stated Preference (SP) data are gathered. This information captures current transportation behaviors and potential future choices in the context of adopting high-speed rail. Additionally, the survey process examines the factors influencing the decision to use High-Speed Railways across four regions of Thailand, selecting provinces where High-Speed Railways stations are planned based on their economic significance, to ensure a diverse and representative sample. Three key factors are considered in the data collection process: socio-economic aspects, travel behavior, and various decision-making scenarios that could influence an individual's choice of transportation mode. Once the data is collected, it is passed through a splitter, which divides it into two main categories: those who would adopt the future transportation mode (high-speed rail) and those who would continue using current transportation modes. This separation of data helps to analyze and train models effectively based on the distinct behavior patterns of these two groups. The next step involves getting the data ready for model training. The dependent variable, which represents the selected mode of transportation, is allocated to y_train, while the independent variables, which comprise socioeconomic and travel behavior data, are assigned to X_train. Python was used for the analysis through Anaconda. As for the analysis time, after data correction and final adjustments, the analysis took approximately 4-5 hours per model. The computer specifications used for the analysis are as follows: Processor: 13th Gen Intel(R) Core (TM) i9-13900H 2.60 GHz, RAM 32.0 GB, where a variety of categorization models, including more sophisticated machine learning methods like XGBoost, LightGBM, and CatBoost, as well as more conventional models like the Binary Logit Model, are taught to anticipate transit choices. Because these models can manage intricate relationships in the data, they are used to improve prediction accuracy. The models are hyperparameter tuned to maximize performance after training. This process involves adjusting the internal parameters of the models to ensure the best possible outcomes. The optimized models are then tested using a separate dataset, X_test and y_test, to evaluate their accuracy and generalization capabilities. Finally, the performance of each model is compared using various evaluation metrics are Accuracy, Precision, Recall, F1-score, and AUC. These metrics help determine which model performs the best. The final step includes identifying the key factors influencing transportation decisions using SHAP analysis, which allows for an interpretability of model predictions, highlighting the most important variables that drive decisions regarding future transportation modes like high-speed rail.

มหาวิทยาลัยเทคโนโลยีสุรนารี



มหาวิทยาลัยเทคโนโลยีสุรนารี

Figure 2.3 Research process flowchart

2.3.3 Binary Logit Model in Mode Choice (BL)

The Binary Logit Model (BL) is applied in mode choice studies as a discrete choice model because it can predict the probability of specific events occurring based on independent variables. The probability of choosing mode (i) for travel is determined by the likelihood that the utility of mode (i) is equal to or greater than the utility of an alternative mode (j). As a result, travelers select the mode of transportation that offers the highest utility.

The probability calculation for developing a binary logit model (BL) is as follows:

$$P_{n1} = \frac{\exp(\beta X_{1n})}{\exp\left(\beta X_{1n}\right) + \exp\left(\beta X_{2n}\right)} = \frac{1}{1 + \exp(\beta X_{2n} - \beta X_{1n})} = \frac{1}{1 + \exp(\Delta U)} \qquad (2\text{-}1)$$

Where,

P_{n1} indicates the likelihood that traveler n will select the first option.

βX_{1n} denotes the utility function associated with traveler n selecting the first mode.

βX_{2n} denotes the utility function associated with traveler n selecting the second mode.

$$\Delta U = \beta X_{2n} - \beta X_{1n} = \sum (a_i - b_i) Z_i ,$$

where Z_i is the i^{th} variable, a_i is the coefficient of the i^{th} variable in βX_{1n}, and b_i is the coefficient of the i^{th} variable in βX_{2n}

2.3.4. Machine Learning

Machine learning (ML) has significantly advanced transportation mode choice modeling by enhancing predictive accuracy and providing deeper insights into traveler behavior, which are crucial for urban planning and policy-making. Models such as random forests and XGBoost have consistently outperformed traditional approaches like the multinomial logit (MNL) model, as demonstrated in studies from Bengaluru and Chengdu, where ML models achieved superior accuracy (Ghosh & Nagaraj, 2024; Tang, Tang, Fu, & Ma, 2024). Similarly, in other fields such as material production, research has confirmed that ML surpasses traditional analytical methods (Bagherzadeh, Shafighfard, Khan, Szczuko, & Mieloszyk, 2023; Kazemi, Asgarkhani, Shafighfard, Jankowski, & Yoo, 2024; Shafighfard, Kazemi, Asgarkhani, & Yoo, 2024). Furthermore, techniques like SHAP have improved the interpretability of ML models, allowing for the identification of key factors such as travel costs and demographic characteristics that influence decisions (Dahmen, Weikl, & Bogenberger, 2023; Tang et al., 2024). Feature importance analysis facilitates more informed policy decisions, and ML models

มหาวิทยาลัยเทคโนโลยีสุรนารี

have demonstrated greater sensitivity to policy changes, as evidenced by studies in Alexandria City (Darwish et al., 2024).

2.3.4.1 Ensemble Models

Ensemble modeling has garnered a lot of interest in machine learning applications due to its broad range of dependent variable prediction capabilities. Typically, tree-based techniques like random forests, additional trees, gradient boosting, or adaptive boosting are used to build ensemble models. On the other hand, new research has demonstrated the potency of innovative ensembles that blend different machine learning techniques to enhance prediction accuracy. These ensembles are sometimes referred to as multi-step-ahead modeling or inclusive multiple models (Başakın, Ekmekcioğlu, Çıtakoğlu, & Özger, 2022; Zouzou & Citakoglu, 2023). Notably, boosting is used in prominent ensemble techniques like XGBoost, LightGBM, and CatBoost to improve prediction accuracy and classification problems. XGBoost is more effective at preventing overfitting than LightGBM, which combines gradient-based one-side sampling and leaf-wise tree growth that is targeted for speed and efficiency on big datasets. This is because XGBoost uses regularization and second-order derivatives. Random Permutation is an effective technique that CatBoost excels at using to handle categorical data with little bias and no overfitting. These models are routinely used to analyze difficult data with extraordinary accuracy, highlighting the increasing significance of sophisticated ensemble techniques in machine learning.

Three popular gradient boosting libraries, XGBoost, LightGBM, and CatBoost (So, 2024), are highly effective for predicting transportation mode choices by analyzing factors like travel cost, time, and socio-demographic attributes. XGBoost handles complex relationships in large datasets, LightGBM excels in speed and scalability, and CatBoost is ideal for categorical data and preventing overfitting, making all three well-suited for transportation mode analysis with minimal preprocessing and high accuracy. These models XGBoost, LightGBM, and CatBoost are well-suited for analyzing transportation mode choices due to their ability to capture complex, nonlinear interactions between variables such as travel time, cost, convenience, and demographic factors like income and car ownership (Xiaowei Li et al., 2024; J. Yu, Chang, Hu, Yin, & Wu, 2024; L. Zhu, Shu, & Zou, 2022; X. Zhu, Shen, Chen, & Zhang,

มหาวิทยาลัยเทคโนโลยีสุรนารี

2024). XGBoost and LightGBM efficiently manage large datasets, while CatBoost makes it ideal for analyzing traveler preferences. Their ability to manage complex data relationships helps identify key factors influencing travel decisions, offering valuable insights for transportation planning.

2.3.4.2 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a machine learning technique that enhances the efficiency of constructing scalable decision trees and is widely used across various fields (T. Chen & Guestrin, 2016). XGBoost was developed from the Gradient Boosting Decision Tree (GBDT) algorithm, which consists of multiple decision trees and consistently uses classification and regression trees (Y. Xu, Zhao, Chen, & Yang, 2019). However, XGBoost differs from GBDT in two key ways. First, XGBoost uses the second-order Taylor expansion for the loss function, whereas GBDT only uses the first-order Taylor expansion. Second, XGBoost incorporates regularization in the objective function to prevent overfitting and reduce model complexity (Krizhevsky, Sutskever, & Hinton, 2012). XGBoost has been widely used for many years owing to its efficiency and high accuracy, which stem from its interpretability, flexibility, and scalability. For example, it can efficiently handle large datasets in parallel and iteratively improve the model, often outperforming other algorithms. Additionally, XGBoost assesses feature importance to aid in understanding the prediction process, supports custom loss functions and evaluation metrics, and allows users to adjust model parameters to meet various requirements, offering greater flexibility (Xiaowei Li et al., 2024).In the context of modeling by XGBoost, we start with an explanation of the learning objective that was adjusted to be regularity. In terms of mathematics, the specified D data set with k sample and m characteristic is shown as $D = \{(X_i, Y_i)\}$ $([D] = k, X_i \in R^m, Y_i \in R)$. A GB model, such as XGBoost, uses an additional regularization function N to predict results (T. Chen & Guestrin, 2016).

$$\hat{Y}_i = \phi(X_i) = \sum_{n=1}^{N} f_n(X_i),\ f_n \in F, \qquad (2\text{-}2)$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

The value of $F=\left\{f(x)=\omega_{q(x)}\right\}(q:R^m\rightarrow T,\omega\in R^T)$ refers to the regression tree area (specifically, CART). N is the number of trees, F encompasses all tree regions, q is the tree structure, T is the number of leaves, f_n is a tree with structure q, and the leaf weights are independent. The function $q(x)$ aligns with the input data to learn function groups used in the model, enabling the objectives to be regularized as follows:

$$\wedge(\theta)=\sum_i l\left(\hat{Y}_i, Y_i\right)+\sum_n \Omega\left(f_n\right)$$

$$\text{where }\Omega\left(f_n\right)=\gamma T+\frac{1}{2}\lambda\|\omega\|^2 \tag{2-3}$$

The term l represents different convex loss functions used to measure the prediction of $\hat{Y}_i$ against the target of $\hat{Y}_i$. The second term, Ω, penalizes model complexity and includes the regression tree functions γ and λ, which are regularization parameters. This regularization term smooths the final learned weights, helping to prevent overfitting. In summary, the objective of regularization is to select a simpler model that still provides efficient predictions. However, adding a group of efficient tree models cannot enhance area efficiency using the traditional Euclidean method because the model will be trained by adding this loss function.

$$\wedge^{(t)}=\sum_{i=1}^{k} l(Y_i,\hat{Y}_i^{t-1}+f_t\left(X_i\right))+\delta\left(f_t\right) \tag{2-4}$$

This equation will gradually add a function f_t, which helps to improve the model as much as the equation.

2.3.4.3 Light Gradient Boosting Machine (LightGBM)

LightGBM is a machine learning algorithm developed by Ke et al. (2017) to improve the efficiency and accuracy of models based on Gradient Boosting Decision Trees (GBDT). LightGBM introduces several key innovations, including gradient-based one-side sampling (GOSS), leaf-wise tree growth, histogram-based approaches, and exclusive feature bundling (EFB). These developments allow LightGBM to work with GBDT models up to twenty times faster than traditional frameworks, without sacrificing prediction accuracy (Ke et al., 2017). LightGBM significantly enhances GBDT,

มหาวิทยาลัยเทคโนโลยีสุรนารี

making it an excellent choice for solving complex classification problems. One of its primary advantages is that GBDT models are known for their superior predictive power compared to general models (Ding, Cao, & Næss, 2018). Additionally, GBDT-based algorithms, like LightGBM, are not constrained by the distribution of independent and dependent variables and can effectively manage multicollinearity, outliers, and missing values. LightGBM also does not require specifying the relationship between independent and dependent variables in advance, allowing it to better capture nonlinear relationships.

Moreover, LightGBM automatically determines the contribution of independent variables to the prediction by considering the interactions among them (Elith, Leathwick, & Hastie, 2008). This is a distinct advantage over traditional statistical models, which rely on standard coefficients and marginal effects. By accounting for auxiliary effects between independent variables, LightGBM can more precisely depict the effect size of those variables, making it particularly effective for handling complex, high-dimensional datasets.

The main procedure of LightGBM begins with calculating the gradient for data example i and tree t. The loss value L is then calculated using the gradient value g_i (Ke et al., 2017) given by

$$g_i = \frac{\partial L(y_i, y_i)}{\partial y_i} \tag{2-5}$$

y_i is the actual value, and $\hat{y}_i$ is the predicted value from the tree. LightGBM uses gradient-based one-side sampling (GOSS) to select only the examples with the highest gradients while subsampling those with lower gradients to reduce the data used for model training. Next, exclusive feature bundling (EFB) is applied to group mutually exclusive features together, reducing the number of features and speeding up tree construction. During tree building, LightGBM constructs the tree by distributing the best data based on the calculated gradients, adjusting parameters to enhance efficiency. Finally, the predicted values are obtained by aggregating the results from all trees as follows:

$$y_i = \sum_{t=1}^{T} f_t(x_i) \tag{2-6}$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

T represents all the trees, and $f_n\left(X_i\right)$ is the function from tree t. The LightGBM framework is designed to efficiently evaluate these functions and handle large datasets effectively. Additionally, parameter tuning and technique adjustments are used to prevent overfitting and enhance prediction accuracy.

2.3.4.4 Categorical Boosting (CatBoost)

CatBoost is a machine learning method based on the Gradient Boosting Decision Tree (GBDT) framework, developed by Yandex engineers in 2017 (Dorogush, Ershov, & Gulin, 2018; Prokhorenkova, Gusev, Vorobev, Dorogush, & Gulin, 2018). It refines the GBDT approach to efficiently use all available data during training by enhancing the level method. Initially, the training process assigns equal weights to all samples. After this initial phase, the weights are adjusted to give more emphasis to samples that deviate from the predictions. The next basic training sample will be repeated in each round, with weights adjusted based on the errors of previous predictions until all training samples are processed. Finally, the overall prediction is derived from aggregating the predictions of all training samples, with weights adjusted accordingly to achieve the optimal result (Zhai et al., 2023). This approach helps mitigate overfitting issues during training (Pham et al., 2020).

CatBoost enhances the efficiency of gradient evaluation through a priority-based procedure. It begins by sampling permutations σ of [1, n] to sort the training samples and then initiates n different models, M1, M2,..., Mn. Each model Mi is trained using the i-th sample from the permutation. In each iteration, the gradient estimation bias of the traditional GBDT is addressed, with the sample j being processed by the Mj−1 model (G. Huang et al., 2019). This method can decrease traditional gradient-boosting decision tree gradient estimation bias and help the model gain higher generalization. However, despite these strengths, CatBoost can still face challenges when dealing with highly imbalanced data. While it adjusts class weights to manage imbalance, extreme cases where one class significantly outweighs the other may result in difficulty accurately predicting the minority class, increasing error rates for those samples. Moreover, incorrect tuning of parameters such as learning rate, depth, and iterations can lead to overfitting or underfitting, particularly impacting the minority

มหาวิทยาลัยเทคโนโลยีสุรนารี

class. Additionally, if not properly adjusted for imbalanced data, CatBoost may introduce biases into the predictions, resulting in poor accuracy for underrepresented classes (Prokhorenkova et al., 2018).

The CatBoost algorithm comprises several key steps. It starts with classification adjustment using the target-based statistics technique, which calculates statistics based on the prediction target, such as the average target for each classification during the building process. CatBoost uses boosting learning to construct multiple trees, with each tree aimed at correcting the errors of the previous one. The gradient value is calculated from the loss function L to refine the predicted value y_i and enhance the accuracy of subsequent predictions (Prokhorenkova et al., 2018). The following formula is used:

$$y_i^{(t)} = y_i^{(t-1)} + \eta f_t(x_i) \qquad (2\text{-}7)$$

η is the learning rank.

2.3.5 Shapley Additive Explanations (SHAP)

SHAP is a tool designed to explain the predictions of machine learning models, particularly complex ones such as gradient boosting machines (GBM). While these models may perform well during training, their internal functioning can be difficult to interpret, and their effectiveness in real-world situations may not always be clear. SHAP helps to elucidate how these models make predictions, providing insights into their decision-making processes (Hong, Naghibi, Pourghasemi, & Pradhan, 2016). This problem considerably affects the risk and reliability of the model (Lapuschkin et al., 2019). Enhancing the ability to explain the "black box" problem in models is a crucial approach to improving the performance and reliability of machine learning algorithms (Mangalathu, Hwang, & Jeon, 2020; Mao et al., 2021). SHAP provides a clear analysis of each feature’s impact on the predicted result by estimating feature importance and integrating it with the SHAP values of other features used in the model. This makes SHAP a key tool for interpreting GBM models. The results of the SHAP value analysis are typically presented in a ranked diagram, where the horizontal position of each variable indicates its importance and effect on the prediction. The interpret

มหาวิทยาลัยเทคโนโลยีสุรนารี

ranking will use the color to show the relation size for the observation: + (red) or – (blue) (Adland, Jia, Lode, & Skontorp, 2021; Vega & Aznarte José, 2020).

SHAP value can be calculated using the following formula:

$$\phi_k(val) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(n-|S|-1!}{n!}(val(S\cup\{i\})-val(S)) \tag{2-8}$$

where val is the feature relevance to the algorithm target and $\phi_k(val)$ is the weighted summing of the feature contributions to the model target result overall feature combinations. $\frac{|S|!(n-|S|-1!}{n!}$ is the weight of $|S|$, val(s) is the expected value of $|S|$, The number of features on the data framework is n, and the subset of the model features is s. The letter i stands for the vector of feature values for the sections that require interpretation (Adland et al., 2021; Cascetta et al., 2020; Lundberg et al., 2020; Lundberg & Lee, 2017; Shafighfard, Kazemi, Bagherzadeh, Mieloszyk, & Yoo, 2024)

A linear summation of the standard features and the SHAP of all features were the goals of the method. It was shown as:

$$g(x)=l(x_8)=\phi_0+\sum_{t=1}^{n}\phi_t x_8^{i} \tag{2-9}$$

where ϕ_0 is the standard value for the undefined features and n is the number of features in the data framework, for feature t, the value ϕ_t represents the SHAP value, whereas x_8 is the vector of reduced input variables (Shafighfard, Kazemi, Bagherzadeh, et al., 2024).

2.4 Results

2.4.1 Descriptive Analysis

Traveler data collection in Thailand involved a self-administered questionnaire with 3,200 samples. After data collection, a comprehensive data integrity assessment was conducted to ensure the completeness and reliability of the dataset. This assessment included verifying that respondents answered all required questions and calculating basic statistical measures such as the mean, skewness, and kurtosis to identify any potential anomalies in the data distribution. Additionally, under sampling techniques were applied to address data imbalance by reducing the size of the majority class, ensuring a more balanced dataset that would not disproportionately

มหาวิทยาลัยเทคโนโลยีสุรนารี

affect subsequent analysis and model performance. The results are presented in Table 2.3.

Table 2.3 General social, economic, and travel data

Variable	Description	Categorical variable (%)	Mean	*SD*	*KU*	*SK*
GENDER	Male = 1	52.5	0.525	0.499	-1.989	-0.100
	Female = 0	47.5				
	Total	100				
HOUSEHOLD_SIZE	Household members 1 person = 1	8.6	3.210	1.12	-0.675	-0.301
	2 people = 2	18.0				
	3 people = 3	28.8				
	4 people = 4	33.4				
	More than four people = 5	11.4				
	Total	100				
CHILDREN	Have children in household (younger than 18 years old) = 1	62.0	0.619	0.485	-1.756	-0.493
	Do not have children in household (younger than 18 years old) = 0	38.0				
	Total	100				
HOUSEHOLD_INCOME	Less than 15,000 = 1	2.7	2.944	0.896	-1.306	-0.113
	15,000–30,000 = 2	35.1				
	30,001–45,000 = 3	27.3				

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 General social, economic, and travel data (Continued)

Variable	Description	Categorical variable (%)	Mean	*SD*	*KU*	*SK*
	More than 45,000 = 4	34.9				
	Total	100				
CARS	Have = 1	75.2	0.752	0.431	-0.634	-1.168
	Do not have = 0	24.8				
	Total	100				
OBJ1	Travel purpose for study/work Yes = 1	38.6	0.386	0.486	-1.782	0.466
	No = 0	61.4				
	Total	100				
OBJ2	Travel purpose for leisure/vacation Yes = 1	48.8	0.488	0.499	-1.99	0.046
	No = 0	51.2				
	Total	100				
OBJ3	Travel purpose for shopping Yes = 1	10.3	0.102	0.303	4.864	2.619
	No = 0	89.7				
	Total	100				
TRAVEL_FREQUENCY	Travel frequency between provinces per year 1–3 times = 1	32.0	2.170	1.092	-1.056	0.502
	3–6 times = 2	30.9				
	6–9 times = 3	14.4				

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 2.3 General social, economic, and travel data (Continued)

Variable	Description	Categorical variable (%)	Mean	*SD*	*KU*	*SK*
	More than nine times = 4	22.7				
	Total	100				
MODE_CHOICE	Choose new mode (High-Speed Railways) = 1	51.2	0.512	0.499	-1.997	-0.048
	Do not choose new mode (bus, train, and airplane) = 0	48.8				
	Total	100				

<u>Note</u>: *SD* = standard deviation, *KU*= kurtosis, and *SK*= skewness.

The table shows that the research population comprised 52.5% males and 47.5% females. Most households had four members (33.4%) or three members (28.8%), and 62% of respondents had children. The majority reported incomes between 15,000 and 30,000 baht (35.1%) or above 45,000 baht (34.9%). Approximately 75.2% owned a private car. The primary travel purposes were for study or work (38.6%) and leisure or vacation (48.8%), with only 10.3% traveling for shopping. In terms of travel frequency between provinces per year, most respondents traveled 1–3 times (32%) or 3–6 times (30.9%). Most preferred traveling by High-Speed Railways (51.2%) over the current transportation system (48.8%). Additionally, the descriptive statistical data indicated a normal distribution, as the skewness (SK) values were within the acceptable range of –3.0 to +3.0, and kurtosis (KU) values were lower than 10.0 (Champahom et al., 2023).

2.4.2 Model Fitting and Performance

This study used four models to verify and predict factors influencing future travel mode choices for High-Speed Railways, analyzed through Python

มหาวิทยาลัยเทคโนโลยีสุรนารี

programming. To enhance model evaluation, a 10-fold Cross-Validation technique was applied to assess model performance consistently.

Focusing on improving model efficiency for classifying unseen data and addressing overfitting issues. The verification process used data from the stated preference (SP) questionnaire to compare the performance of the Binary Logit Model (BL), XGBoost, LightGBM, and CatBoost. Grid search optimization was applied to critical hyperparameters for XGBoost, LightGBM, and CatBoost to decrease overfitting and ensure robust model predictions, as shown in Table 2.4. Model performance was assessed by comparing the log loss values between the training and test datasets, as detailed in Table 2.5 and Table 2.6.

This study, XGBoost, LightGBM, and CatBoost algorithms were utilized to model transportation mode choice, with each model's performance optimized through a grid search algorithm to determine the optimal set of hyperparameters. For XGBoost, key hyperparameters calibrated included the learning rate (eta), maximum tree depth (max_depth), and regularization terms (alpha and lambda), with logloss as the evaluation metric. Similarly, for LightGBM, parameters such as the learning rate, maximum tree depth, and number of leaves were fine-tuned, using binary_logloss as the evaluation metric. CatBoost, chosen for its effective handling of imbalanced datasets, was optimized by adjusting the learning rate, tree depth, and number of iterations. CatBoost addresses data imbalance effectively through its inbuilt mechanism, specifically utilizing Gradient Boosting with Ordered Boosting. This technique helps reduce overfitting by constructing decision trees that are more flexible and robust, even when one class significantly outweighs others (Fernández et al., 2018; Y. Zhang & Haghani, 2015). These adjustments aimed to enhance both the predictive accuracy and robustness of the models, with final hyperparameter values summarized in Table 2.4.

The application of 10-fold Cross-Validation ensured that each model’s performance was rigorously tested across different data partitions. This methodology not only improved the reliability of the results but also provided insight into the variance of the model performance metrics across different folds. As shown in Table

2.7, the cross-validated performance metrics, including Accuracy, Sensitivity, Specificity, Precision, F1 score, and AUC, highlight how each model performed under consistent conditions.

Table 2.4 Hyperparameter values determined using grid search algorithms for XGBoost, LightGBM, and CatBoost

Model	Parameter	Description	Value
XGBoost	eta	Learning rate: [0,1]	0.2
	max_depth	Maximum depth of a tree: [0,20]	7
	gamma	Min_split_loss: [0,10]	0
	alpha	L1 regularization term on weights: [0,10]	0.1
	lambda	L2 regularization term on weights: [0,10]	1
	objective	Objective function	binary: logistic
	eval_metric	Evaluation metric	"logloss"
LightGBM	learning_rate	Learning rate: [0,1]	0.2
	max_depth	Maximum depth of a tree: [0,20]	6
	num_leaves	Maximum tree leaves	41
	reg_alpha	L1 regularization term	0.1
	reg_lambda	L2 regularization term	0.1
	objective	Objective function	binary
	metric	Evaluation metric	binary_ logloss
CatBoost	depth	Maximum depth of a tree	9
	iterations	Number of iterations	500
	l2_leaf_reg	L2 regularization term	1
	learning_rate	Learning rate: [0,1]	0.05

มหาวิทยาลัยเทคโนโลยีสุรนารี

The evaluation of model efficiency considered metrics such as accuracy, precision, recall, and F1 score, all of which offer different insights into model performance across various dimensions. The use of macro average (Macro avg) was chosen for this study as it treats each class equally, regardless of the number of samples in each class. This approach is crucial in multiclass classification problems with imbalanced data, where some classes have significantly fewer samples than others. The macro average involves two main steps: 1) calculating the evaluation metrics (precision, recall, and F1 score) for each individual class, and 2) averaging these values to provide a more reliable and acceptable comparison of model performance (Mokhtarimousavi, Anderson, Azizinamini, & Hadi, 2020; Sokolova & Lapalme, 2009). The values were considered as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-9}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{2-10}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2-11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2-12}$$

$$\text{F-Score} = \frac{2TP}{2TP + FN + FP} \tag{2-13}$$

$$\text{AUC} = \frac{1}{2}(\text{Recall} + \text{Specificity}) \tag{2-14}$$

These metrics were calculated from a confusion matrix, which includes:

TP is a true positive rate.

TN is a true negative rate.

FP is a false positive rate.

FN is a false negative rate.

Accuracy represents the proportion of correct predictions over the total instances, while recall (sensitivity) measures the model's ability to correctly identify relevant instances, and specificity focuses on identifying negative instances accurately. Precision assesses the proportion of correct positive predictions, and the F1 score balances precision and recall by calculating their harmonic mean. Given the imbalanced nature of the dataset, AUC-ROC (Area Under the Receiver Operating Characteristic Curve) was also utilized as a crucial metric. Unlike accuracy, which can be misleading with imbalanced data, AUC-ROC provides a robust evaluation by

มหาวิทยาลัยเทคโนโลยีสุรนารี

assessing the trade-off between true positive and false positive rates across different thresholds. This metric is particularly valuable as it remains unaffected by class imbalance and offers a more comprehensive view of the model's performance in distinguishing between minority and majority classes. By combining traditional metrics like accuracy, precision, and recall with AUC-ROC, the study ensures a thorough and reliable assessment of the model's ability to handle imbalanced data.

This comprehensive metric facilitates thorough evaluation across various dimensions, including accuracy, the rates of true positive and true negative predictions, prediction values, imbalance management, and discrimination.

Table 2.5 Performance Comparison of Train Models for High-Speed Railways Usage Prediction

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7388	0.7646	0.7118	0.7355	0.7497	0.8149
XGBoost	0.9312	0.9542	0.9070	0.9151	0.9342	0.9306
LightGBM	0.8966	0.9362	0.8549	0.8714	0.8962	0.9681
CatBoost	0.9461	0.9643	0.9270	0.9327	0.9460	0.9910

Table 2.6 Performance Comparison of Test Models for High-Speed Railways Usage Prediction

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7445	0.7720	0.7154	0.7410	0.7562	0.8187
XGBoost	0.8887	0.9245	0.8512	0.8672	0.8949	0.8878
LightGBM	0.8762	0.9299	0.8197	0.8443	0.8755	0.9526
CatBoost	0.8970	0.9292	0.8631	0.8771	0.8966	0.9669

มหาวิทยาลัยเทคโนโลยีสุรนารี

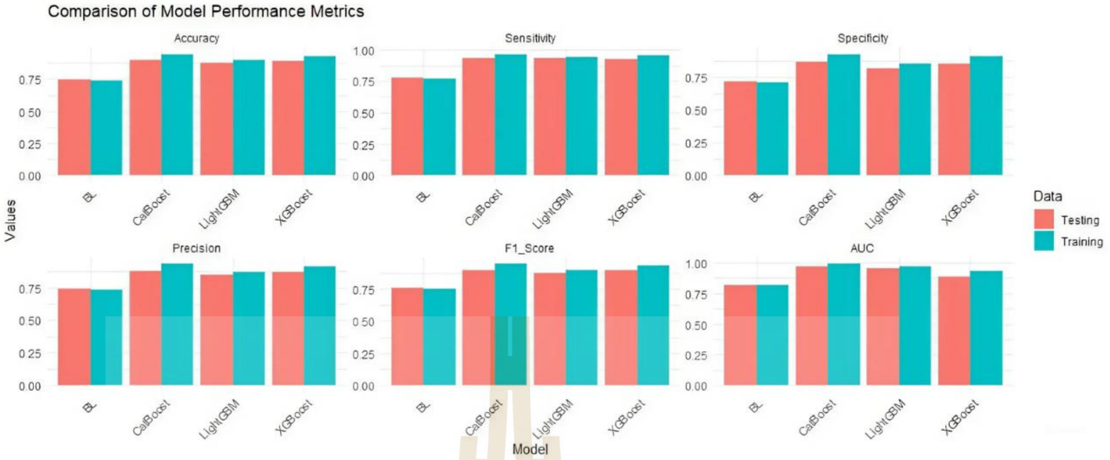


Figure 2.4 Comparison Model Performance Metrics

As shown in Table 2.5, Table 2.6 and Figure 2.4, the efficiency of High-Speed Railways travel behavior prediction varied significantly across models. The Binary Logit Model (BL) achieved training and testing accuracies of 0.7388 and 0.7445, with sensitivities of 0.7646 and 0.7720, and specificities of 0.7118 and 0.7154, respectively. Its precision and F1 score were moderate. In contrast, XGBoost performed notably better, with accuracies of 0.9312 for training and 0.8887 for testing, high sensitivities of 0.9542 and 0.9245, and strong specificities of 0.9070 and 0.8512. Precision and F1 scores were also high, at 0.9151 and 0.9342 for training, and 0.8672 and 0.8949 for testing. LightGBM achieved accuracies of 0.8966 for training and 0.8762 for testing, with high sensitivity of 0.9362 and 0.9299, and specificity of 0.8549 and 0.8197. Precision and F1 scores remained high, with training values of 0.8714 and 0.8962, and testing values of 0.8443 and 0.8755. CatBoost demonstrated the highest efficiency, with accuracies of 0.9461 for training and 0.8970 for testing. Its sensitivity of 0.9643 and 0.9292, and specificity of 0.9270 and 0.8631 were notably high, as were its precision and F1 scores, at 0.9327 and 0.9460 for training, and 0.8771 and 0.8966 for testing. CatBoost outperformed XGBoost and LightGBM, delivering the highest overall predictive efficiency for High-Speed Railways usage.

The close performance between the training and test sets, with the test set surpassing the training set in some metrics, can be explained by several factors. Regularization techniques, commonly used in models like Gradient Boosting, intentionally reduce the training performance to prevent overfitting and improve

มหาวิทยาลัยเทคโนโลยีสุรนารี

generalization (T. Chen & Guestrin, 2016). Additionally, if k-fold cross-validation is used, the reported training performance is averaged across folds, which may differ from the single test set performance (Krstajic, Buturovic, Leahy, & Thomas, 2014). Finally, overfitting prevention methods, such as early stopping, allow the model to halt training before fully fitting the training data, resulting in better performance on the test set (Natekin & Knoll, 2013).

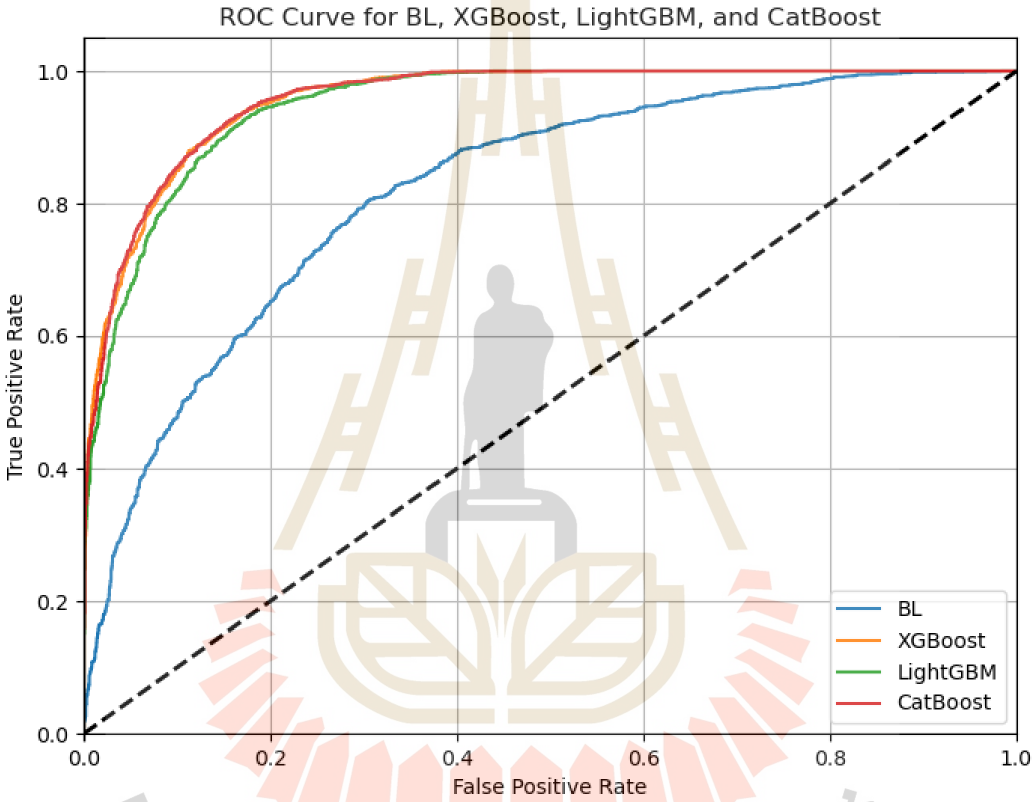


Figure 2.5 ROC Curve model

มหาวิทยาลัยเทคโนโลยีสุรนารี

As shown in Figure 2.5, the ROC curve comparison indicates that machine learning models significantly outperform the baseline logistic regression (BL). The Binary Logit Model (BL) has the lowest AUC, indicating moderate performance. XGBoost and LightGBM show strong classification abilities, while CatBoost performs best with the highest AUC, making it the most effective in predicting High-Speed Railways travel mode choices.

Table 2.7 Cross-Validation Analysis of Model Performance in Transforming Travel Patterns to High-Speed Railways

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7404 ± 0.006	0.7655 ± 0.006	0.7140 ± 0.0129	0.7376 ± 0.0085	0.7513 ± 0.0055	0.8161 ± 0.0067
XGBoost	0.8846 ± 0.0068	0.9210 ± 0.0094	0.8464 ± 0.0062	0.8630 ± 0.0055	0.8910 ± 0.0067	0.9583 ± 0.0036
LightGBM	0.8763 ± 0.0077	0.9358 ± 0.0071	0.8139 ± 0.0107	0.8408 ± 0.0083	0.8857 ± 0.0069	0.9506 ± 0.0034
CatBoost	0.8853 ± 0.0061	0.9245 ± 0.0072	0.8441 ± 0.0068	0.8616 ± 0.0058	0.8920 ± 0.0058	0.9584 ± 0.0034

As shown in Table 2.7, the 10-fold Cross-Validation results reveal noticeable differences in model performance when predicting High-Speed Railways travel behavior. The CatBoost model achieved the highest overall efficiency, with an accuracy of 0.8853 ± 0.0061, sensitivity of 0.9245 ± 0.0072, and an AUC score of 0.9584 ± 0.0034. Its specificity of 0.8441 ± 0.0068 and precision of 0.8616 ± 0.0058, combined with a strong F1 score of 0.8920 ± 0.0058, further solidify its role as the most effective model for predicting changes in travel patterns. The XGBoost model followed closely, showing an accuracy of 0.8846 ± 0.0068, which was slightly lower than CatBoost. However, its sensitivity of 0.9210 ± 0.0094 and AUC of 0.9583 ± 0.0036 indicate strong predictive capability. XGBoost performed similarly in other metrics, including specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and F1 score (0.8910 ± 0.0067). The LightGBM model demonstrated comparable performance, achieving an accuracy of 0.8763 ± 0.0077 and a relatively higher sensitivity of 0.9358 ± 0.0071, surpassing both CatBoost and XGBoost. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. In contrast, the Binary Logit (BL) model displayed significantly lower performance, with an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were also substantially lower compared to the machine learning models, reflecting its struggles with imbalanced data. Overall, the results highlight the significant performance advantages of the machine learning models over BL, with CatBoost leading in most metrics and providing a reliable, consistent approach for predicting High-Speed Railway usage

2.4.3 Analysis result of the Binary Logit Model (BL)

The analysis of transportation mode using the Binary Logit Model (BL), as shown in Table 2.8, identified several important and statistically significant variables. For example, the gender variable (GEN) had a β value of −0.0734 and a significance level of 0.013, indicating that males were significantly more likely to choose the new public transportation system, such as High-Speed Railways, compared to females. This finding aligns with a previous study by Ta, Wang, Hu, and Liu (2022). They studied travel behavior and the factors influencing transportation mode choices between males and females in China. They determined that females were more inclined to use new public transportation systems, such as High-Speed Railways, than males. This preference was attributed to females' more positive perceptions and attitudes toward High-Speed Railways, including their safety, convenience, and environmental benefits (F. Chen, Hao, & Chen, 2021; Xiaowei Li et al., 2021; Xie et al., 2022). Additionally, the house size variable had a β value of −0.3779 with a significance level of less than 0.001, indicating that individuals with larger house sizes are less likely to travel by High-Speed Railways. This tendency is linked to factors such as higher travel expenses for larger families, the convenience of traveling with a large group, and the management of time and activities for family members. Those living in larger homes often prefer using private cars or other transportation modes over High-Speed Railways because these alternatives can be more cost-effective (Rubin, Mulder, & Bertolini, 2014). The variable for the number of children in a family (CHIL_SIZE) had a β value of 0.9693 and a significance level of less than 0.001, indicating that families with children under 18 years old were more likely to use High-Speed Railways compared to families without children. This preference is likely attributed to the additional considerations for vehicle space and the higher comfort costs associated with traveling with children, which make High-Speed Railways a more appealing option (Dowling, 2015). High-Speed Railways offer considerable advantages in terms of convenience over buses and trains and cost benefits compared to airplanes. The variable for household income (MON_HOUSE) had a β value of 0.0925 and a significance level of less than 0.001, indicating that higher-income households were more likely to use High-Speed Railways. This is attributed to their financial security, which allows them to afford the superior convenience, speed, and quality of service offered by High-Speed Railways compared to other transportation modes. High-income households are more likely to overlook travel costs, valuing the time saved and the ability to engage in other activities, which is a key factor for them in choosing High-Speed Railways (Ren et al., 2020; P. Zhao & Yuan, 2023). The number of cars in the household had a β value of –0.0476 with a significance level of 0.198, indicating it was not statistically significant. Among the variables related to travel purposes, traveling for study or work (OBJ1) did not show statistical significance. However, traveling for leisure or vacation (OBJ2) had a β value of –0.2670 and a significance level of 0.008, indicating a statistically significant influence on travel decisions. Travelers who prefer leisure or vacation travel are less likely to choose High-Speed Railways, which is consistent with a previous study by Xiaowei Li et al. (2021) This indicates that travelers who seek relaxation are more likely to use trains or buses rather than High-Speed Railways. In contrast, travel for work or business purposes, which is often mandatory or required, typically involves using High-Speed Railways (Xiaowei Li et al., 2020). Travel for shopping (OBJ3) had a significance level of 0.061, indicating it was not statistically significant. In contrast, the variable for traveling between cities (FR_TRA) had a β value of 0.3582 and a significance level of less than 0.001, showing that frequent travelers were more likely to use High-Speed Railways. This is attributed to the convenience and speed of High-Speed Railways, which meet their needs for reliability and time savings. High-Speed Railways reduce travel fatigue and enhance efficiency, leading to greater satisfaction for customers who frequently travel between cities. Additionally, the variable for access time (ACCESS) was considered in this analysis. The variables of waiting time (WAITING), travel time (TRAVEL), and service frequency (Headway) (FR_TRA) all had negative β values with a significance level of less than 0.001. This indicates that if access time, waiting time, travel time, and service frequency for High-Speed Railways were to increase, customers would be more likely to choose High-Speed Railways for their travel. The accessibility of transport stations plays a crucial role in this decision. Regarding access time, Martín, Román, García-Palomares, and Gutiérrez (2014) have explored strategies for enhancing service by improving station access, either via private car or direct public transport connections. These improvements increase the competitiveness of the public transport system relative to other modes of transportation. The emphasis on saving time while waiting for transportation between cities is valued more than the time spent accessing the station (Román et al., 2014). Waiting time plays a crucial role in customer decision-making and satisfaction (Abenoza, Cats, & Susilo, 2019). If the waiting time for High-Speed Railways is excessively long, it will likely reduce demand and decrease its competitive edge compared to other transportation modes (L. Wu, Zhang, & Wang, 2015). Travel time is a crucial factor in travel planning, with travelers often preferring transportation systems that offer shorter travel times (Danapour, Nickkar, Jeihani, & Khaksar, 2018; Lhéritier, Bocamazo, Delahaye, & Acuna-Agost, 2019). Service frequency considerably impacts the competitiveness of the public transport system. A lower headway improves the service level and provides a competitive advantage in the market (Jiaoe Wang, Huang, & Jing, 2020). The study by Yuanqing Wang et al. (2014) has shown that the frequency of High-Speed Railways service effectively segments the market. The cost (COST) has a β value of 2.1997 and a significance level of <0.001, indicating that higher costs are associated with increased travel by High-Speed Railways, consistent with the study by Karmarkar, Jana, and Velaga (2023). A study on the willingness to pay for High-Speed Railways services in India found that travelers were willing to pay an average of 300 rupees, which is higher than their current travel expenses, in exchange for saving time on their journeys. This finding aligns with the study by Shelat, Cats, and van Lint (2021), which revealed that travelers were willing to pay more to reduce waiting time and improve travel accuracy. The Binary Logit Model (BL) analysis yielded a Cox and Snell R^2 value of 0.287 and a Nagelkerke R^2 value of 0.383. According to relevant literature, R^2 values between 0.2 and 0.4 are generally considered indicative of a good model fit (Abulibdeh, 2018).

Table 2.8 Model parameter estimates for mode choice

Variable	β	Standard error	Sig.
CONST	13.6393	0.462	<0.001*
GEN	−0.0734	0.029	0.013*
HOUSE_SIZE	−0.3779	0.016	<0.001*
CHIL_SIZE	0.9693	0.035	<0.001*
MON_HOUSE	0.0925	0.019	<0.001*
N_CARS	−0.0476	0.037	0.198
OBJ1	0.0112	0.101	0.912
OBJ2	−0.2670	0.101	0.008*
OBJ3	−0.2042	0.109	0.061
FR_TRA	0.3582	0.014	<0.001*
ACCESS	−3.3402	0.148	<0.001*
WAITING	−9.5750	0.471	<0.001*
TRAVEL	−14.6052	0.447	<0.001*
COST	2.1997	0.195	<0.001*
FREQ	−1.0164	0.071	<0.001*

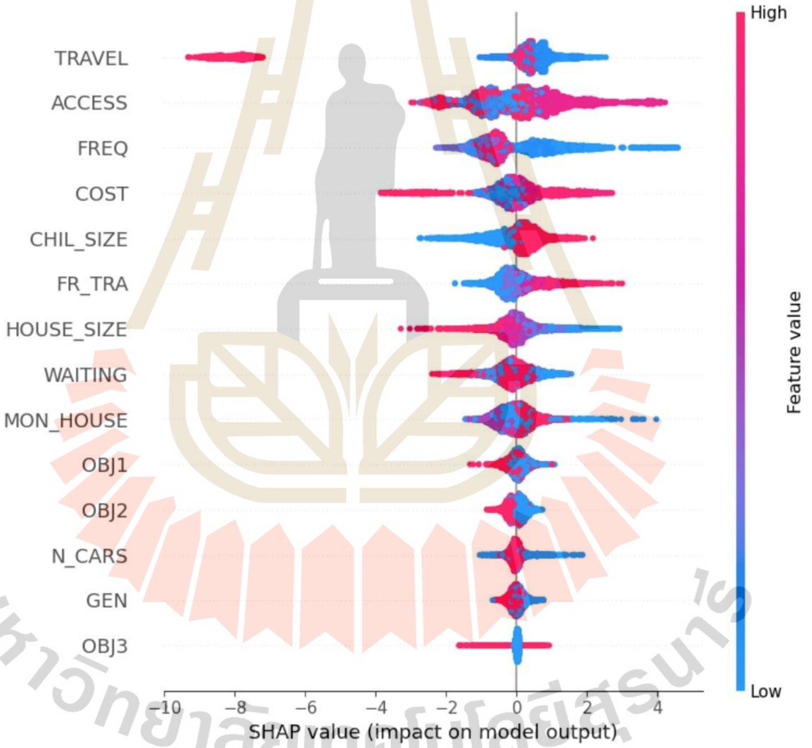
* Sig. = Significance < 0.05

2.4.4 Assessing Future Travel Mode Choice

The evaluation of mode choice is crucial for understanding travelers' economic attributes and the conditions influencing their travel mode decisions based on survey response rankings. The SHAP method offers a deeper analysis of how economic and social attributes, along with travel conditions, affect travel mode choice. This methodology primarily aims to relate each data instance response with travel behavior patterns for more accurate insights (Zaidan, Abulibdeh, Alban, & Jabbar, 2022).

Figures 2.6 to 2.8 show the mapping of responses to understand the influence of various variables on mode choice. The importance of the input variables, as displayed in Figures 2.6 to 2.8, was determined by calculating the average of the absolute Shapley values. The input features were then ranked based on their significance, with a higher mean SHAP value indicating greater importance. Figure 2.6

presents the XGBoost analysis results, Figure 2.7 shows the LightGBM analysis results, and Figure 2.8 depicts the CatBoost analysis results, revealing similar variable prioritization across the models. The most important factors for choosing High-Speed Railways in the future were travel time, service frequency, access time, travel cost, waiting time, number of children in the family, frequency of inter-provincial travel per year, number of household members, household income, private car ownership, travel purpose for study/work, gender, travel purpose for leisure/vacation, and travel purpose for shopping, respectively. The machine learning analysis of important factors for High-Speed Railways choice in the future aligned with the results obtained using the Binary Logit Model (BL) method.



มหาวิทยาลัยเทคโนโลยีสุรนารี

Figure 2.6 Impact of input features on High-Speed Railways usage prediction using SHAP with XGBoost

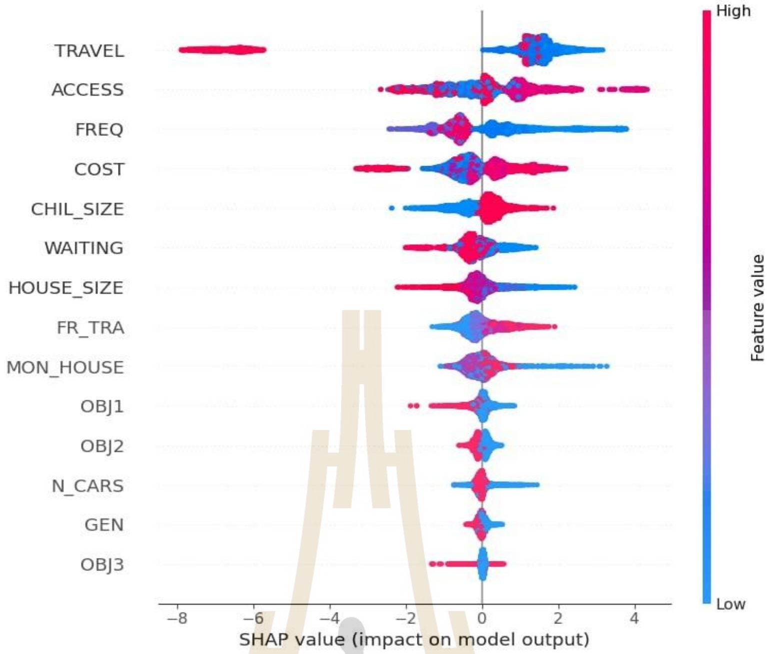


Figure 2.7 Impact of input features on High-Speed Railways usage prediction using SHAP with LightGBM

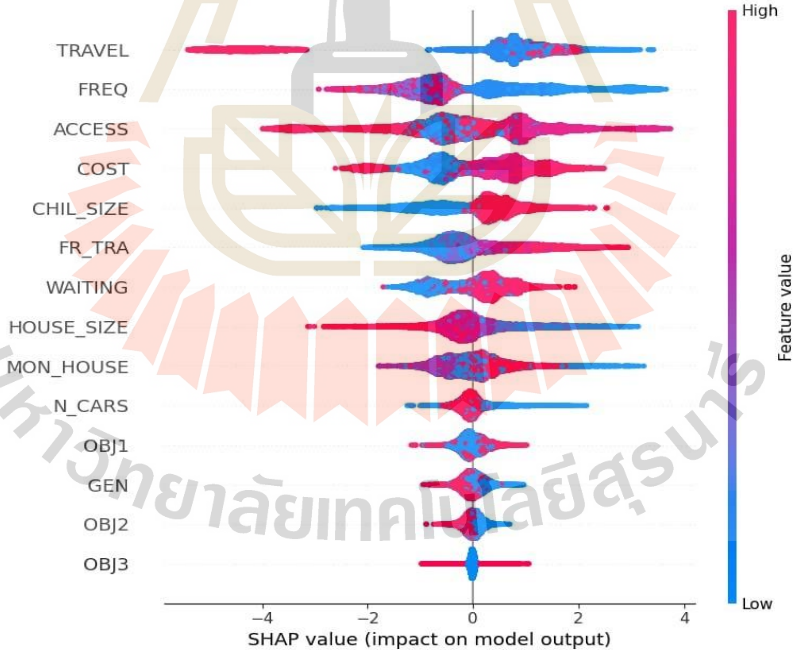


Figure 2.8 Impact input features on High-Speed Railways usage prediction using SHAP with CatBoost

The decision to use High-Speed Rail in Thailand is influenced by several key factors, as highlighted by the SHAP analysis in Figures 2.6 to 2.8. Travel time is crucial, as High-Speed Rail offers a substantial time-saving advantage over conventional buses and trains, which are often slower and more prone to delays. Passengers who value time efficiency, especially frequent inter-provincial travelers, find High-Speed Rail particularly attractive in Thailand, where long bus rides and traffic congestion are common. However, if the time savings offered by High-Speed Rail are not as great as air travel or if passengers prefer the flexibility of buses or personal vehicles, some may opt for these alternatives. Another significant factor is access time to High-Speed Rail stations. Longer access times can reduce the appeal of High-Speed Rail, especially for passengers in rural areas with less developed public transportation infrastructure. Conversely, passengers in urban areas or those prioritizing speed, comfort, and reliability may still choose High-Speed Rail, even if access times are longer compared to other modes of transport like buses or conventional trains, which typically offer more flexible access points. Service frequency also plays a critical role in passengers' decisions. When High-Speed Rail services are infrequent or the intervals between departures are too long, passengers may turn to buses, which operate more frequently and offer greater flexibility. This is particularly relevant in Thailand, where buses are a common mode of transportation between provinces and have flexible schedules that accommodate passengers' travel needs, especially for last-minute trips. Cost is another key factor influencing passenger behavior. While some passengers, particularly those with higher incomes, are willing to pay more for the speed, comfort, and reliability of High-Speed Rail, others are more price-sensitive and opt for cheaper alternatives like buses or traditional trains. When High-Speed Rail fares are perceived as too high, passengers often choose buses, which are much less expensive, or trains if speed is not a priority. Waiting time between High-Speed Rail services is a major deciding factor. Long waiting times make High-Speed Rail less attractive, particularly for passengers who value flexibility and frequent departures. In Thailand, where buses and airplanes offer more frequent services, passengers may find these alternatives more convenient. Moreover, longer waiting times can disrupt travel plans and extend the overall journey duration, making High-Speed Rail less appealing to those who prioritize quick and efficient travel.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Household income higher earners may opt for it because it saves them time, but passengers with lesser incomes may not believe that purchasing more expensive tickets is worthwhile. Additionally, having an automobile has a variety of effects. While some passengers who own cars prefer to be able to drive, others could choose to take the High-Speed Rail in order to escape traffic and parking issues. Gender data indicates that women are more likely than males to utilize high-speed trains, most likely due to their awareness of the benefits' security and comfort. Passengers traveling for business or study show a range of behaviors when it comes to their travel habits. Some may prioritize money and select less expensive options, while others may value the time efficiency of High-Speed Rail. Tourists and leisure travelers prefer to steer clear of high-speed rail, probably due to the variable nature of vacation travel, where flexibility in scheduling and cost-effectiveness trump speed. Similarly, shopping trips result in mixed behavior, with some passengers choosing High-Speed Rail for convenience and others prioritizing flexibility or cost savings.

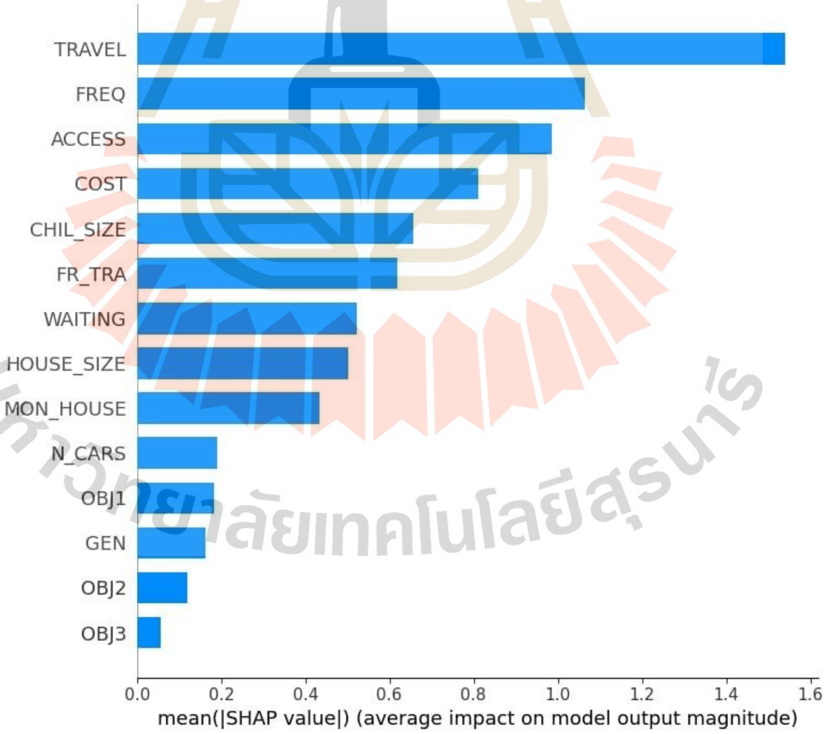


Figure 2.9 Feature importance for High-Speed Railways usage using SHAP

In Figure 2.9, the importance of each feature was determined by calculating the average absolute values of Shapley across the dataset. The input features were

มหาวิทยาลัยเทคโนโลยีสุรนารี

ranked by their significance based on the mean SHAP values, with higher values indicating greater importance. Additionally, the figure emphasizes the significance of each input variable for the High-Speed Railways mode.

2.5 Discussion

The comparison between the four models shows that the Binary Logit Model had the lowest efficiency in analyzing travel mode choices, primarily due to its limitations in handling complex, imbalanced datasets. Binary Logit Mode, being a traditional model, often struggles with non-linear relationships and data variability (Tang et al., 2024). In contrast, the three machine learning models—XGBoost, LightGBM, and CatBoost—demonstrated superior efficiency, with CatBoost outperforming the others. The strength of CatBoost lies in its ability to handle imbalanced datasets effectively, a common challenge in travel mode prediction studies. Its resilience to overfitting further enhances its utility, making it particularly suitable for complex datasets like those used in this analysis. Moreover, CatBoost's robust handling of missing data and integration with under sampling techniques provides a more balanced and accurate prediction of less common travel modes, while maintaining interpretability and accuracy across the model (Huanfa Chen & Cheng, 2023).

The results from the CatBoost model underline that travel time plays a decisive role in the choice of High-Speed Railways as a new mode of transportation. The significant reduction in travel time is a key factor driving passengers to opt for High-Speed Rail, especially in Thailand, where conventional modes like buses or trains are often hindered by traffic congestion and delays. This advantage makes High-Speed Rail a highly efficient and attractive alternative, particularly during peak hours when time efficiency is crucial. As previous studies have shown, passengers tend to prioritize faster travel options that minimize travel time (Yajuan Deng, Yu Bai, Liangbin Cui, & Renjie He, 2023; L. Liu & Zhang, 2018; Jianqiang Wang, Zhao, Liu, & Huang, 2023).In addition to travel time, access time is another critical factor. The convenience of accessing High-Speed Rail stations, especially in urban areas, significantly improves passenger preference. As Yongsheng Zhang, Kangyu Liang, Enjian Yao, and Mingyi Gu (2024) pointed out, reliable and predictable access to stations enhances overall convenience, making High-Speed Rail more appealing. However, in more rural or less developed regions, longer access times can reduce the attractiveness of High-Speed Rail as an option, which is why improving station accessibility remains a key policy recommendation (Zhou et al., 2020). Service frequency further impacts passengers' choices. Short intervals between train arrivals improve service efficiency and reduce waiting times, which are critical in maintaining the system’s competitive edge over other modes like buses or airplanes. If services are too infrequent, passengers may opt for more flexible options like buses, which run more frequently, especially between provinces (L. Liu & Zhang, 2018). The cost of High-Speed Railways services is also a significant factor. Passengers with higher incomes are more willing to pay for faster, more reliable transportation, as studies in other countries like India and Italy have shown (Cartenì, Pariota, & Henke, 2017; Karmarkar et al., 2023). However, for price-sensitive passengers, especially in developing regions, the higher cost of High-Speed Rail could deter them from using the service. Balancing affordability with service quality is thus essential to increase ridership across diverse income groups. Waiting time between services is crucial in shaping the passenger experience. Long waiting times negatively affect satisfaction and reduce the likelihood of passengers opting for High-Speed Rail, particularly when more frequent and flexible alternatives like buses and airplanes are available (Pei, Zhou, & Peng, 2013). Managing wait times efficiently not only enhances the travel experience but also boosts the competitiveness of High-Speed Railways by ensuring smoother operations and improved passenger flow.

Family structure is one of the most important social and economic elements influencing travel behavior in Thailand. Compared to bigger families, smaller families travel more frequently, especially for leisure purposes. In contrast, larger households with higher incomes are more likely to own private cars, influencing their mode of travel. The size of the home, the family's income, whether or not a car is owned, and whether or not there are small children or elderly family members all have an impact on travel behavior (P. Zhao & Yuan, 2023). For instance, because of their convenience and safety characteristics, families with children are thought to find particular value in High-Speed Railways (HSR). In Thailand, where traffic congestion frequently makes long-distance travel difficult, High-Speed Rail (HSR) offers a quick and effective substitute.

มหาวิทยาลัยเทคโนโลยีสุรนารี

This makes it appealing to families looking for dependable modes of transportation. High-Speed Railways is a better option for families who value safety and punctuality because it is more socially and economically beneficial than traditional trains and has higher traffic efficiency (Hongsheng Chen, Sun, Zhu, & Zeng, 2016). Moreover, choosing to use high-speed rail is heavily influenced by household income. Families with lower incomes also express a desire to utilize High-Speed Railways because it provides greater reliability, safety, and comfort, even though passengers with higher incomes are more likely to choose it because of its time-saving advantages. These advantages are especially crucial in places like Thailand, where traffic and accidents frequently make road travel uncertain. For special trips where safety and dependability are more important than cost, like family holidays or necessary travel, lower-income families may consider High-Speed Railways to be worth the additional expense. This indicates that although cost is still an obstacle, for some population segments the benefit offered by High-Speed Railways can outweigh the increased price (Y. Huang, Song, & Jia, 2014).

Travel preferences are influenced by owning a car. While some people choose to use High-Speed Railways to avoid traffic and parking issues, others, especially in large cities like Bangkok, like having the freedom to drive. Behavior is influenced by gender as well; women are more likely to use High-Speed Railways because it feels secure and comfortable. In Thailand, where concerns about safety are common, this makes High-Speed Railways more alluring to female travelers. Business and study travelers have different preferences when it comes to transportation. While some choose more affordable options, others enjoy High-Speed Railways 's ability to save time, especially when traveling from an urban location to a rural one. Meanwhile, High-Speed Railways is usually avoided by tourists and leisure visitors in favor of more flexible and cost-effective options like buses or private cars. High-Speed Railways networks offer quick and effective transportation, making it possible for frequent travelers to go where they're going without difficulty, as Y. Huang et al. (2014) further highlights. When it comes to passengers with tight schedules, including those who are going for business or school, High-Speed Railways 's timeliness guarantees dependable travel. Because of its dependability, High-Speed Railways is a competitive choice for long-distance or intercity travel, particularly in Thailand where road traffic can be unpredictable. In conclusion, considering its benefits in terms of speed, safety, and dependability, High-Speed Railways is a desirable alternative for families, women, and those with better incomes in Thailand. Lower-class visitors may find High-Speed Railways less appealing because of its expense, but many of them still think it's a good choice for special occasions because of its total value. While travelers and those looking for flexibility could favor less expensive options, measures to alleviate these socioeconomic gaps and enhance accessibility to High-Speed Railways might encourage use among a wider range of income levels.

2.6 Conclusion and Policy Implications

Thailand's collaboration with China to develop High-Speed Rail (HSR) is a crucial step in enhancing transportation infrastructure and boosting regional economic growth. While most research focuses on existing travel modes like cars, buses, and planes, there's a lack of analysis on future transport choices, particularly in developing nations like Thailand. This study addresses this gap by analyzing factors influencing High-Speed Railways adoption, such as access time, service frequency, and socioeconomic characteristics, to shape future traveler decisions.

The comparison of models for predicting High-Speed Rail usage revealed that CatBoost consistently outperformed the other models, with cross-validation confirming its superior performance across all key metrics. The Binary Logit Model (BL) demonstrated moderate effectiveness, achieving an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a relatively lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were moderate, but it struggled with handling imbalanced data. In contrast, XGBoost showed significantly stronger results, with an accuracy of 0.8846 ± 0.0068, a sensitivity of 0.9210 ± 0.0094, and an AUC of 0.9583 ± 0.0036. XGBoost also delivered high specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and F1 score (0.8910 ± 0.0067). LightGBM also performed well, with an accuracy of 0.8763 ± 0.0077 and a relatively high sensitivity of 0.9358 ± 0.0071. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. CatBoost demonstrated the highest overall performance, with its ability to handle imbalanced datasets, avoid overfitting, and provide clear interpretability through SHAP values, making it the most effective model for predicting High-Speed Rail usage, outperforming XGBoost, LightGBM, and BL. The relatively higher performance of CatBoost can be generalizable to similar studies in different regions or countries, depending on the structure and characteristics of the dataset in each region, such as social and demographic diversity, transportation options, and regional transportation policies.

The decision to use High-Speed Rail (High-Speed Railways) in Thailand is influenced by several key factors, reflecting the unique transportation context of the country. Travel time is a major advantage, with High-Speed Railways offering faster journeys compared to buses and conventional trains, which are prone to delays, making it appealing to frequent inter-provincial travelers. However, if time savings are not significant compared to air travel or if passengers prefer the flexibility of buses or private cars, High-Speed Railways may not be their preferred choice. Access time to stations also plays a critical role, especially for rural passengers where public transport infrastructure is less developed, though urban travelers may still prioritize High-Speed Railways for its speed and reliability. Mixed responses regarding access time are observed; some passengers still choose High-Speed Railways despite longer station access due to its overall comfort and speed, while others prefer buses or personal vehicles for greater flexibility. Service frequency is another crucial factor; infrequent departures could drive passengers to opt for buses, which provide more flexible schedules. Cost, whether high or low, leads to mixed behavior: some passengers are willing to pay more for the convenience and time savings High-Speed Railways offers, while others may opt for cheaper alternatives regardless of High-Speed Railways's affordability. Waiting time between services can further dissuade passengers, especially in Thailand, where buses and airplanes offer more frequent options. Household income and automobile ownership also influence preferences, with higher earners and those without cars more inclined to choose High-Speed Railways. Women are more likely than males to select High-Speed Railways in Thailand, probably because of its comfort and safety characteristics. This suggests that gender plays a part. The reason for the trip also counts. While consumers and leisure travelers have mixed tastes and frequently value flexibility and cost savings over speed, business and study commuters prefer High-Speed Railways for its time efficiency. The use of High-Speed Railways in Thailand is influenced by a number of factors that together show the population's varied requirements and habits.

By taking into account elements related to travel behavior, policymakers can gain a better understanding of how people make transportation decisions and develop solutions that can be customized to support high-speed rail and remove adoption hurdles. Policymakers can more successfully focus policies to promote sustainable transportation choices and remove barriers to HSR adoption by understanding reactions to various circumstances.To effectively promote High-Speed Rail (HSR) usage in Thailand, several key policy actions can be implemented:

1) Enhancing Confidence in High-Speed Railways: Publicizing data on reduced travel times compared to cars and other modes will build public trust in the speed and punctuality of High-Speed Railways.
2) Increasing Service Frequency: Adjusting High-Speed Railways schedules to meet peak demand during mornings, evenings, and weekends, and adding extra services during festivals, will reduce waiting times and enhance service availability.
3) Competitive Pricing: Offering discount packages for groups like students, families, and seniors, and seasonal promotions, will make High-Speed Railways more attractive, especially compared to other transport options.
4) Reducing Station Waiting Time: Streamlining boarding processes through more efficient ticketing systems, online reservations, and real-time updates on train schedules will improve passenger experience and reduce delays.
5) Developing Infrastructure Connections: Expanding public transport links (buses, metro) to High-Speed Railways stations will improve accessibility, particularly from community areas, encouraging more people to use High-Speed Railways.

6) Family-Friendly Services: Providing family-oriented services like child-friendly areas, diaper-changing stations, and family travel discounts will make High-Speed Railways more appealing to families.
7) Targeting Car Users: Highlighting benefits like reduced travel fatigue, lower accident risks, and long-distance cost savings will encourage car users to switch to High-Speed Railways. Offering secure parking options at stations will facilitate this shift.
8) Ongoing Public Relations Campaigns: A continuous public relations effort is essential to raise awareness of High-Speed Railways 's benefits and address public concerns, ensuring widespread adoption of this mode of travel.

While these policies can promote High-Speed Railways, there are practical limitations to consider. Financial constraints, such as funding for discounts and infrastructure improvements, must be addressed. Increasing train frequency is challenging due to physical infrastructure limits. Human resource management, legal approvals, and compliance with environmental regulations must also be considered. Regular data collection and evaluation will ensure that policies are adapted to passenger needs, maximizing the long-term success of High-Speed Railways in Thailand.

2.7 Limitations and Further Research

This study has several limitations. First, the use of hypothetical scenarios and reliance on estimated data, as the High-Speed Railways infrastructure was not yet constructed, may introduce uncertainty into the findings. Additionally, the absence of real-world behavioral data and potential variations in travel preferences over time could affect the generalizability of the results. There are also potential biases inherent in the self-administered questionnaires used in the study, and the short data collection period may not fully capture long-term travel behavior changes. Furthermore, the reliance on a single method of data collection may limit the comprehensiveness of the findings, particularly as the study focused on travel mode choice during a specific timeframe. These factors may influence the accuracy and generalizability of the models used, especially as transportation preferences evolve due to economic or infrastructural changes.

มหาวิทยาลัยเทคโนโลยีสุรนารี

For future research, it is recommended to integrate a variety of data sources, such as real-time tracking technologies (e.g., UAV-based object tracking) and longer-term data collection. This approach would provide more dynamic insights and enhance the predictive power of machine learning models. Additionally, future studies could explore the broader economic and environmental impacts of High-Speed Railways development in Thailand, incorporating post-launch customer feedback surveys to assess satisfaction and service effectiveness. Such data would be invaluable for refining transportation policies and improving the High-Speed Railways services to better meet the evolving needs of the population.

2.8 References

Aalami, M. R., Anari, A., Shafighfard, T., & Talatahari, S. (2013). A robust finite element analysis of the rail-wheel rolling contact. *Advances in Mechanical Engineering,* 5, 272350.

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability,* 11(1), 189. doi:10.3390/su11010189

Abenoza, R. F., Cats, O., & Susilo, Y. O. (2019). How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. *Transportation,* 46(5), 1615-1642.

Abulibdeh, A. (2018). Implementing congestion pricing policies in a MENA Region City: Analysis of the impact on travel behaviour and equity. *Cities,* 74, 196-207. doi:https://doi.org/10.1016/j.cities.2017.12.003

Abulibdeh, A. (2023). Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis. *Transportation Research Interdisciplinary Perspectives,* 20, 100852. doi:10.1016/j.trip.2023.100852

Adland, R., Jia, H., Lode, T., & Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering & System Safety,* 209, 107480. doi:10.1016/j.ress.2021.107480

AirAsia move. (2024). Flight. Retrieved from https://www.airasia.com/th/th

Al-Atroush, M., Mustaffa, Z., & Sebeay, T. (2022). *Emerging trends in overcoming the weather barrier to sustainable mobility in gulf and tropical cities.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

Arencibia, A. I., Feo-Valero, M., García-Menéndez, L., & Román, C. (2015). Modelling mode choice for freight transport using advanced choice experiments. *Transportation Research Part A: Policy and Practice,* 75, 252-267. doi:10.1016/j.tra.2015.03.027

Armah, F. A., Yawson, D. O., & Pappoe, A. A. N. M. (2010). A Systems Dynamics Approach to Explore Traffic Congestion and Air Pollution Link in the City of Accra, Ghana. *Sustainability,* 2(1), 252-265. doi:10.3390/su2010252

Bagherzadeh, F., Shafighfard, T., Khan, R. M. A., Szczuko, P., & Mieloszyk, M. (2023). Prediction of maximum tensile stress in plain-weave composite laminates with interacting holes via stacked machine learning algorithms: A comparative study. *Mechanical Systems and Signal Processing,* 195, 110315. doi:https://doi.org/10.1016/j.ymssp.2023.110315

Başakın, E. E., Ekmekcioğlu, Ö., Çıtakoğlu, H., & Özger, M. (2022). A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Computing and Applications,* 34(1), 783-812.

Becker, N., & Carmi, N. (2019). Changing trip behavior in a higher education institution: The role of parking fees. *International Journal of Sustainable Transportation,* 13(4), 268-277. doi:10.1080/15568318.2018.1463418

Belgiawan, P. F., Ilahi, A., & Axhausen, K. W. (2019). Influence of pricing on mode choice decision in Jakarta: A random regret minimization model. *Case Studies on Transport Policy,* 7(1), 87-95. doi:10.1016/j.cstp.2018.12.002

Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems* (pp. 283-307): Elsevier.

Birolini, S., Malighetti, P., Redondi, R., & Deforza, P. (2019). Access mode choice to low-cost airports: Evaluation of new direct rail services at Milan-Bergamo airport. *Transport policy,* 73, 113-124. doi:10.1016/j.tranpol.2018.10.008

BusOnlineTicket.co.th. (2024). Book Thailand Bus Tickets Online. Retrieved from https://www.busonlineticket.co.th/

Calvo, F., Eboli, L., Forciniti, C., & Mazzulla, G. (2019). Factors influencing trip generation on metro system in Madrid (Spain). *Transportation Research Part D: Transport and Environment,* 67, 156-172.

Cartenì, A., Pariota, L., & Henke, I. (2017). Hedonic value of high-speed rail services: Quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities. *Transportation Research Part A: Policy and Practice,* 100, 348-365. doi:10.1016/j.tra.2017.04.018

Cascetta, E., Cartenì, A., Henke, I., & Pagliara, F. (2020). Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post evaluation and future perspectives. *Transportation Research Part A: Policy and Practice,* 139, 412-428. doi:https://doi.org/10.1016/j.tra.2020.07.008

Champahom, T., Banyong, C., Hantanong, N., Se, C., Jomnonkwao, S., & Ratanavaraha, V. (2023). Factors influencing the willingness to pay for motorcycle safety improvement: A structural equation modeling approach. *Transportation Research Interdisciplinary Perspectives,* 22, 100950. doi:https://doi.org/10.1016/j.trip.2023.100950

Chang, X., Wu, J., Liu, H., Yan, X., Sun, H., & Qu, Y. (2019). Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data. *Transportmetrica A: Transport Science,* 15(2), 1587-1612. doi:10.1080/23249935.2019.1620380

Chen, F., Hao, X., & Chen, Z. (2021). Can high-speed rail improve health and alleviate health inequality? Evidence from China. *Transport policy,* 114, 266-279. doi:https://doi.org/10.1016/j.tranpol.2021.10.007

Chen, H., & Cheng, Y. (2023). Travel mode choice prediction using imbalanced machine learning. *IEEE Transactions on Intelligent Transportation Systems,* 24(4), 3795-3808. doi:10.1109/TITS.2023.3237681

มหาวิทยาลัยเทคโนโลยีสุรนารี

Chen, H., Sun, D., Zhu, Z., & Zeng, J. (2016). The impact of high-speed rail on residents' travel behavior and household mobility: A case study of the Beijing-Shanghai Line, China. *Sustainability,* 8(11), 1187. doi:10.3390/su8111187

Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society,* 14, 1-10. doi:10.1016/j.tbs.2018.09.002

Cui, H., Li, M., Zhu, M., & Ma, X. (2023). Investigating the Impacts of Urban–Rural Bus Service Quality on Rural Residents' Travel Choices Using an SEM–MNL Integration Model. *Sustainability,* 15(15), 11950.

Dahmen, V., Weikl, S., & Bogenberger, K. (2023). Interpretable Machine Learning for Mode Choice Modeling on Tracking-Based Revealed Preference Data. *Transportation Research Record*, 03611981241246973.

Danapour, M., Nickkar, A., Jeihani, M., & Khaksar, H. (2018). Competition between high-speed rail and air transport in Iran: The case of Tehran–Isfahan. *Case Studies on Transport Policy,* 6(4), 456-461. doi:https://doi.org/10.1016/j.cstp.2018.05.006

Darwish, A. M., Almansour, M., Salah, A., Zagow, M., Saeed, K., & Elkafoury, A. (2024). Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: a case study of Alexandria City, Egypt. *Transportation Research Interdisciplinary Perspectives,* 24, 101052.

Deng, Y., Bai, Y., Cui, L., & He, R. (2023). Travel Mode Choice Behavior for High-Speed Railway Stations Based on Multi-Source Data. *Transportation Research Record,* 2677(8), 525-540.

Department of International Trade Promotion. (2023). High-speed rail 4 routes. Retrieved from https://www.rentconnected.com/blogs/high-speed-rail-4-routes-for-thailand/

Díaz-Ramírez, J., Estrada-García, J. A., & Figueroa-Sayago, J. (2023). Predicting transport mode choice preferences in a university district with decision tree-based models. *City and Environment Interactions,* 20, 100118. doi:https://doi.org/10.1016/j.cacint.2023.100118

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice,* 110, 107-117.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:*1810.11363.

Dowling, R. (2015). Parents, children and automobility: Trends, challenges and opportunities. *Handbook on transport and development*, 526-538.

Economic Base. (2017). Knock on the Bangkok-Chiang Mai high-speed train fare 1,089 baht. Retrieved from https://www.thansettakij.com/business/242231

Elharoun, M., El-Badawy, S. M., & Shahdah, U. E. (2023). Artificial intelligence techniques for predicting individuals' mode choice behavior in Mansoura city, Egypt. *Transportation Research Record,* 2677(10), 605-623.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology,* 77(4), 802-813.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10): Springer.

Ghosh, T., & Nagaraj, N. (2024). Evaluating the Determinants of Mode Choice Using Statistical and Machine Learning Techniques in the Indian Megacity of Bengaluru. *arXiv preprint arXiv:*2401.13977.

Givoni, M. (2006). Development and Impact of the Modern High-speed Train: A Review. *Transport Reviews,* 26(5), 593-611. doi:10.1080/01441640600589319

Gross, S., & Grimm, B. (2018). Sustainable mode of transport choices at the destination–public transport at German destinations. *Tourism Review,* 73(3), 401-420.

Gunay, G., & Gokasar, I. (2021). Market segmentation analysis for airport access mode choice modeling with mixed logit. *Journal of Air Transport Management,* 91, 102001.

Guo, Y., Li, Y., Anastasopoulos, P. C., Peeta, S., & Lu, J. (2021). China’s millennial car travelers’ mode shift responses under congestion pricing and reward policies: A case study in Beijing. *Travel Behaviour and Society,* 23, 86-99.

Habib, M., Poole, K., & McCarthy, S. (2022). *NovaTRAC: A multi-year travel survey data collection program.* Paper presented at the the 12th International Conference on Transport Survey Methods.

Heinen, E., Harshfield, A., Panter, J., Mackett, R., & Ogilvie, D. (2017). Does exposure to new transport infrastructure result in modal shifts? Patterns of change in commute mode choices in a four-year quasi-experimental cohort study. *Journal of Transport & Health,* 6, 396-410. doi:https://doi.org/10.1016/j.jth.2017.07.009

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*: Cambridge university press.

Hong, H., Naghibi, S. A., Pourghasemi, H. R., & Pradhan, B. (2016). GIS-based landslide spatial modeling in Ganzhou City, China. *Arabian Journal of Geosciences,* 9, 1-26.

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., . . . Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology,* 574, 1029-1041.

Huang, Y., Song, H., & Jia, C. Y. (2014). Analysis on the main factors influencing the high-speed railway passenger flow. *Advanced Materials Research,* 919, 1085-1090.

Jiang, Y., Yu, S., Guan, W., Gao, S., & Feng, T. (2021). Ground access behaviour of air-rail passengers: A case study of Dalian ARIS. *Travel Behaviour and Society,* 24, 152-163.

Jin, M., Lin, K.-C., Shi, W., Lee, P. T., & Li, K. X. (2020). Impacts of high-speed railways on economic growth and disparity in China. *Transportation Research Part A: Policy and Practice,* 138, 158-171.

Juan de Dios Ortúzar, & Willumsen, L. G. (2011). Trip Generation Modelling. In *Modelling Transport* (pp. 139-173).

Karmarkar, O. D., Jana, A., & Velaga, N. R. (2023). Willingness to pay for reduced travel time: Case study of Mumbai - Ahmedabad High Speed Railway. *Case Studies on Transport Policy,* 12, 100974. doi:https://doi.org/10.1016/j.cstp.2023.100974

Kashifi, M. T., Jamal, A., Kashefi, M. S., Almoshaogeh, M., & Rahman, S. M. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behaviour and Society,* 29, 279-296.

Kazemi, F., Asgarkhani, N., Shafighfard, T., Jankowski, R., & Yoo, D.-Y. (2024). Machine-Learning Methods for Estimating Performance of Structural Concrete Members Reinforced with Fiber-Reinforced Polymers. *Archives of Computational Methods in Engineering*, 1-33.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems,* 30.

Khan, S., Adnan, A., & Iqbal, N. (2022). *Applications of artificial intelligence in transportation.* Paper presented at the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET).

Kim, E.-J. (2021). Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation,* 2021(1), 6685004.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems,* 25.

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics,* 6, 1-15.

Kujala, R., Weckström, C., Mladenović, M. N., & Saramäki, J. (2018). Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. *Computers, Environment and Urban Systems,* 67, 41-54. doi:https://doi.org/10.1016/j.compenvurbsys.2017.08.012

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications,* 10(1), 1096.

Lhéritier, A., Bocamazo, M., Delahaye, T., & Acuna-Agost, R. (2019). Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling,* 31, 198-209.

Li, J., Lo, K., & Guo, M. (2018). What influences the choice between private car and public transport for ShoppingTrips? Impact of socio-economic and built environment factors. *Journal of Asian Energy Studies,* 2(1), 28-42.

Li, X., Ma, Q., Wang, W., & Wang, B. (2021). Influence of weather conditions on the intercity travel mode choice: A case of Xi'an. *Computational Intelligence and Neuroscience,* 2021(1), 9969322.

Li, X., Shi, L., Shi, Y., Tang, J., Zhao, P., Wang, Y., & Chen, J. (2024). Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Applied Geography,* 166, 103264.

Li, X., Tang, J., Hu, X., & Wang, W. (2020). Assessing intercity multimodal choice behavior in a touristy city: a factor analysis. *Journal of Transport Geography,* 86, 102776.

Liang, Y., Zhou, K., Li, X., Zhou, Z., Sun, W., & Zeng, J. (2020). Effectiveness of high-speed railway on regional economic growth for less developed areas. *Journal of Transport Geography,* 82, 102621.

Liu, L., & Zhang, M. (2018). High-speed rail impacts on travel times, accessibility, and economic productivity: A benchmarking analysis in city-cluster regions of China. *Journal of Transport Geography,* 73, 25-40.

Liu, S., Yao, E., Cheng, X., & Zhang, Y. (2017). Evaluating the impact of new lines on entrance/exit passenger flow of adjacent existing stations in urban rail transit system. *Transportation Research Procedia,* 25, 2625-2638. doi:https://doi.org/10.1016/j.trpro.2017.05.316

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., . . . Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence,* 2(1), 56-67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems,* 30.

Mangalathu, S., Hwang, S.-H., & Jeon, J.-S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures,* 219, 110927.

Mao, H., Deng, X., Jiang, H., Shi, L., Li, H., Tuo, L., . . . Guo, F. (2021). Driving safety assessment for ride-hailing drivers. *Accident Analysis & Prevention,* 149, 105574.

Martín, J. C., Román, C., García-Palomares, J. C., & Gutiérrez, J. (2014). Spatial analysis of the competitiveness of the high-speed train and air transport: The role of access to terminals in the Madrid–Barcelona corridor. *Transportation Research Part A: Policy and Practice,* 69, 392-408.

Mirzaei, E., Kheyroddin, R., & Mignot, D. (2021). Exploring the effect of the built environment, weather condition and departure time of travel on mode choice decision for different travel purposes: evidence from Isfahan, Iran. *Case Studies on Transport Policy,* 9(4), 1419-1430.

Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks. *International journal of transportation science and technology,* 9(2), 100-115.

Nam, D., & Cho, J. (2020). Deep neural network design for modeling individual-level travel mode choice behavior. *Sustainability,* 12(18), 7481.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics,* 7, 21.

Ning, J., Lyu, T., & Wang, Y. (2021). Exploring the built environment factors in the metro that influence the ridership and the market share of the elderly and students. *Journal of Advanced Transportation,* 2021(1), 9966794.

Nurhidayat, A. Y., Widyastuti, H., Sutikno, & Upahita, D. P. (2023). Research on Passengers’ Preferences and Impact of High-Speed Rail on Air Transport Demand. *Sustainability,* 15(4), 3060.

Office of the national economic and social development council. (2023). Gross Regional and Provincial Product Chain Volume Measure 2022 Edition. Retrieved from https://www.nesdc.go.th/ewt_dl_link.php?nid=15104&filename=gross_regional

Otim, T., Dörfer, L., Ahmed, D. B., & Munoz Diaz, E. (2022). Modeling the impact of weather and context data on transport mode choices: A case study of gps trajectories from beijing. *Sustainability,* 14(10), 6042.

Pasha, M. M., Hickman, M. D., & Prato, C. G. (2020). Modeling Mode Choice of Air Passengers' Ground Access to Brisbane Airport. *Transportation Research Record,* 2674(11), 756-767.

Pavlou, M., Ambler, G., Qu, C., Seaman, S. R., White, I. R., & Omar, R. Z. (2024). An evaluation of sample size requirements for developing risk prediction models with binary outcomes. *BMC Medical Research Methodology,* 24(1), 146.

Pei, Y. L., Zhou, K., & Peng, T. (2013). Prediction model of passenger waiting time in high-speed rail hub based on BP neural network. *Applied Mechanics and Materials,* 321, 1903-1906.

Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., . . . Takeuchi, W. (2020). Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam. *Remote Sensing,* 12(8), 1334.

Pitombo, C. S., Salgueiro, A. R., da Costa, A. S. G., & Isler, C. A. (2015). A two-step method for mode choice estimation with socioeconomic and spatial information. *Spatial Statistics,* 11, 45-64.

POST TO DAY. (2017). Transportation" expects the ticket price of the Bangkok-Chiang Mai high-speed train to run for 3 hours and a half 1,200 baht,. Retrieved from https://www.posttoday.com/business/531218

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems,* 31.

Puan, O., Hassan, Y., Mashros, N., Idham, M., Hassan, N., Warid, M., & Hainin, M. (2019). *Transportation mode choice binary logit model: A case study for Johor Bahru* *city.* Paper presented at the IOP conference series: Materials science and engineering.

Rattanakijsuntorn, W., Suwannarat, B., Samittivate, N., & Nithikittiwat, C. (2024). Success Factors and Partnership Evaluation of Air–Rail Integration Development: A Case of a High-Speed Rail Project Linking Three Airports in Thailand. *Infrastructures,* 9(7), 115.

Ren, X., Chen, Z., Wang, F., Dan, T., Wang, W., Guo, X., & Liu, C. (2020). Impact of high-speed rail on social equity in China: Evidence from a mode choice survey. *Transportation Research Part A: Policy and Practice,* 138, 422-441. doi:https://doi.org/10.1016/j.tra.2020.05.018

Román, C., Martín, J. C., Espino, R., Cherchi, E., de Dios Ortúzar, J., Rizzi, L. I., . . . Amador, F. J. (2014). Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor. *Transport policy,* 36, 105-117.

Rubin, O., Mulder, C. H., & Bertolini, L. (2014). The determinants of mode choice for family visits – evidence from Dutch panel data. *Journal of Transport Geography,* 38, 137-147. doi:https://doi.org/10.1016/j.jtrangeo.2014.06.004

Salas, P., De la Fuente, R., Astroza, S., & Carrasco, J. A. (2022). A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Systems with Applications,* 193, 116253.

Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science,* 3(2), 158.

Shafighfard, T., Kazemi, F., Asgarkhani, N., & Yoo, D.-Y. (2024). Machine-learning methods for estimating compressive strength of high-performance alkali-activated concrete. *Engineering Applications of Artificial Intelligence,* 136, 109053. doi:https://doi.org/10.1016/j.engappai.2024.109053

Shafighfard, T., Kazemi, F., Bagherzadeh, F., Mieloszyk, M., & Yoo, D. Y. (2024). Chained machine learning model for predicting load capacity and ductility of steel fiber–reinforced concrete beams. *Computer‐Aided Civil and Infrastructure Engineering*.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Sharma, A., Gani, A., Asirvatham, D., Ahmed, R., Hamzah, M., & Asli, M. F. (2021). Travel Behavior Modeling: Taxonomy, Challenges, and Opportunities. *International Journal of Advanced Computer Science and Applications (IJACSA),* 12(5), 774-786.

Shelat, S., Cats, O., & van Lint, J. W. C. (2021). Quantifying travellers' evaluation of waiting time uncertainty in public transport networks. *Travel Behaviour and Society,* 25, 209-222. doi:https://doi.org/10.1016/j.tbs.2021.07.009

So, B. (2024). Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM. *Scandinavian Actuarial Journal*, 1-23.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management,* 45(4), 427-437.

Sresakoolchai, J., & Kaewunruen, S. (2020). Comparative studies into public private partnership and traditional investment approaches on the high-speed rail project linking 3 airports in Thailand. *Transportation Research Interdisciplinary Perspectives,* 5, 100116.

Srithongrung, A., & Kriz, K. A. (2019). Thai Public Capital Budget and Management Process. In *Capital Management and Budgeting in the Public Sector* (pp. 206-235): IGI Global.

State Railway of Thailand. (2019). *COMPLETE GUIDE TO SRT.* 4.0. Retrieved from https://www.railway.co.th/RailwayMiddleFile/NewsAndActImg/1307/131951934207171332_2_2562.pdf

State Railway of Thailand (2024). The State Railway of Thailand Easy book, Convenient check. Retrieved from https://dticket.railway.co.th/DTicketPublicWeb/home/Home

Sun, W., Wang, C., Liu, C., & Wang, L. (2021). High-speed rail network expansion and its impact on regional economic sustainability in the Yangtze River Delta, China, 2009–2018. *Sustainability,* 14(1), 155.

Ta, N., Wang, X., Hu, L., & Liu, Z. (2022). Gender difference in commuting travel: A comparative study of suburban residents in Beijing and Shanghai. *Travel Behaviour and Society,* 28, 196-203.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Tang, L., Tang, C., Fu, Q., & Ma, C. (2024). Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis. *IET Intelligent Transport Systems*.

Tarriño-Ortiz, J., Gómez, J., Soria-Lara, J. A., & Vassallo, J. M. (2022). Analyzing the impact of Low Emission Zones on modal shift. *Sustainable Cities and Society,* 77, 103562.

TISSAYAKORN, K. (2021). A study on transit-oriented development strategies in the high-speed rail project in Thailand.

Vega, G. M., & Aznarte José, L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics,* 56, 101039.

Wang, F., & Ross, C. L. (2018). Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record,* 2672(47), 35-45.

Wang, J., Huang, J., & Jing, Y. (2020). Competition between high-speed trains and air travel in China: From a spatial to spatiotemporal perspective. *Transportation Research Part A: Policy and Practice,* 133, 62-78.

Wang, J., Zhao, W., Liu, C., & Huang, Z. (2023). A System Optimization Approach for Trains' Operation Plan with a Time Flexible Pricing Strategy for High-Speed Rail Corridors. *Sustainability,* 15(12), 9556.

Wang, Y., Li, L., Wang, L., Moore, A., Staley, S., & Li, Z. (2014). Modeling traveler mode choice behavior of a new high-speed rail corridor in China. *Transportation Planning and Technology,* 37(5), 466-483.

Weerawat, W., Samitiwantikul, L., & Torpanya, R. (2020). Operational challenges of the Bangkok airport rail link. *Urban rail transit,* 6, 42-55.

Wójcik, S. (2019). The determinants of travel mode choice: the case of ŁódŹ, Poland. *Bulletin of Geography. Socio-economic Series*(44), 93-101.

Wu, D., & Martín, J. C. (2022). Research on Passengers' Preference for High-Speed Railways (HSRs) and High-Speed Trains (HSTs). *Sustainability,* 14(3). doi:10.3390/su14031473

มหาวิทยาลัยเทคโนโลยีสุรนารี

Wu, L., Zhang, M., & Wang, Y. (2015). Analysis of Ninghang intercity travel mode choice behavior based on panel data. *Journal of Transportation Systems Engineering and Information Technology,* 15(1), 226-231.

Xie, J., Zhan, S., Wong, S., Wen, K., Qiang, L., & Lo, S. (2022). High-speed rail services for elderly passengers: Ticket-booking patterns and policy implications. *Transport policy,* 125, 96-106.

Xu, Y., Zhao, X., Chen, Y., & Yang, Z. (2019). Research on a mixed gas classification algorithm based on extreme random tree. *Applied Sciences,* 9(9), 1728.

Yang, W., Chen, Q., & Yang, J. (2022). Factors Affecting Travel Mode Choice between High-Speed Railway and Road Passenger Transport—Evidence from China. *Sustainability,* 14(23), 15745.

Yu, J., Chang, X., Hu, S., Yin, H., & Wu, J. (2024). Combining travel behavior in metro passenger flow prediction: A smart explainable Stacking-Catboost algorithm. *Information Processing & Management,* 61(4), 103733. doi:https://doi.org/10.1016/j.ipm.2024.103733

Yu, X., Lang, M., Gao, Y., Wang, K., Su, C.-H., Tsai, S.-B., . . . Li, S. (2018). An empirical study on the design of China high-speed rail express train operation plan—From a sustainable transport perspective. *Sustainability,* 10(7), 2478.

Zaidan, E., Abulibdeh, A., Alban, A., & Jabbar, R. (2022). Motivation, preference, socioeconomic, and building features: New paradigm of analyzing electricity consumption in residential buildings. *Building and Environment,* 219, 109177.

Zhai, W., Li, C., Fei, S., Liu, Y., Ding, F., Cheng, Q., & Chen, Z. (2023). CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values. *Computers and Electronics in Agriculture,* 214, 108306. doi:https://doi.org/10.1016/j.compag.2023.108306

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies,* 58, 308-324. doi:https://doi.org/10.1016/j.trc.2015.02.019

Zhang, Y., Liang, K., Yao, E., & Gu, M. (2024). Measuring Reliable Accessibility to High-Speed Railway Stations by Integrating the Utility-Based Model and Multimodal

มหาวิทยาลัยเทคโนโลยีสุรนารี

Space–Time Prism under Travel Time Uncertainty. *ISPRS International Journal of Geo-Information,* 13(8), 263.

Zhao, P., & Yuan, D. (2023). Effects of Family Structure on Travel Behaviour. In *Population Growth and Sustainable Transport in China* (pp. 233-266): Springer.

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society,* 20, 22-35.

Zhou, Y., Zhao, M., Tang, S., Lam, W. H., Chen, A., Sze, N., & Chen, Y. (2020). Assessing the relationship between access travel time estimation and the accessibility to high speed railway station by different travel modes. *Sustainability,* 12(18), 7827.

Zhu, L., Shu, S., & Zou, L. (2022). XGBoost-Based Travel Time Prediction between Bus Stations and Analysis of Influencing Factors. *Wireless Communications and Mobile Computing,* 2022(1), 3504704.

Zhu, X., Shen, X., Chen, K., & Zhang, Z. (2024). Research on the prediction and influencing factors of heavy duty truck fuel consumption based on LightGBM. *Energy,* 296, 131221. doi:https://doi.org/10.1016/j.energy.2024.131221

Zouzou, Y., & Citakoglu, H. (2023). General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. *Acta Geophysica,* 71(2), 927-947.

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER III

ANALYZING HIGH-SPEED RAIL’S TRANSFORMATIVE IMPACT ON PUBLIC TRANSPORT IN THAILAND USING MACHINE LEARNING

3.1 Abstract

This study investigates the impact of high-speed rail (HSR) on Thailand’s public transportation market and evaluates the effectiveness of machine learning techniques in predicting travel mode choices. A stated preference survey was conducted with 3200 respondents across 16 provinces, simulating travel scenarios involving buses, trains, airplanes, and HSR. The dataset, consisting of 38,400 observations, was analyzed using the CatBoost model and the multinomial logit (MNL) model. CatBoost demonstrated superior predictive performance, achieving an accuracy of 0.853 and an AUC of 0.948, compared to MNL’s accuracy of 0.749 and AUC of 0.879. Shapley additive explanations (SHAP) analysis identified key factors influencing travel behavior, including cost, service frequency, waiting time, travel time, and station access time. The results predict that HSR will capture 88.91% of the intercity travel market, significantly reducing market shares for buses (4.76%), trains (5.11%), and airplanes (1.22%). The findings highlight the transformative role of HSR in reshaping travel patterns and offer policy insights for optimizing pricing, service frequency, and accessibility. Machine learning enhances predictive accuracy and enables a deeper understanding of mode choice behavior, providing a robust analytical framework for transportation planning.

3.2 Introduction

Thailand is modernizing its public transportation system to support economic growth and population mobility, aiming to enhance travel efficiency for the public. Central to this is the high-speed rail (HSR) project, anticipated to launch soon, which will drastically reduce travel times between major cities such as Bangkok and Chiang Mai. This project is expected to draw passengers from buses, trains, and domestic airlines because of its speed and convenience (Banyong et al., 2024; Sresakoolchai & Kaewunruen, 2020). HSR, with its modern features and efficiency, is considered an appealing option for passengers seeking faster travel solutions (Shuyuan Liu & Putro, 2024; Yang et al., 2022).

However, HSR's introduction will inevitably impact the competitive dynamics of Thailand's public transportation market, especially affecting the bargaining power and market share of existing systems (Bergantino & Madio, 2020; F. Xu et al., 2022; Q. Zhang, Yang, Wang, Zhang, & Zhang, 2020). Currently, intercity public transportation in Thailand is dominated by buses, trains, and airplanes. As of 2022, buses are the primary choice for 78.25% of passengers because of their extensive network and affordability. Trains serve 7.47% of passengers, mainly on routes between major provinces, although they lack speed and convenience. Airplanes hold a 14.28% market share, favored for routes like Bangkok-Chiang Mai and Bangkok-Phuket, where speed and comfort are priorities (Office of Transport and Traffic Policy and Planning, 2024).

Figure 3.1 illustrates the proportion of passengers using buses, trains, and airplanes between cities in Thailand from 2017 to 2022. The advent of HSR, with its advantages, raises critical questions regarding how existing transportation systems—airlines, trains, and buses—will adjust their strategies to retain market share. The long-term impact of HSR on the market structure is also a concern. Studies on market share changes following the introduction of new travel systems in various countries reveal significant effects because of increased competition. F. Xu et al. (2022) found that HSR reduced leading transportation providers' market dominance. Similarly, Nicolò Avogadro, Eric Pels, and Renato Redondi (2023) emphasized that increasing service frequency and reducing travel time boosted the HSR market share in the London–Amsterdam corridor. Ammar Abulibdeh (2023) found that the introduction of a subway in Doha reduced reliance on cars and taxis, with XGBoost emerging as the most accurate market share prediction model. Ristić, Stojadinović, and Trifunović (2022) examined the competition between old and new HSR services, finding that competition expanded market share, evidenced by Trenitalia's success on the Paris-Lyon route. S. Wu and Han (2022) demonstrated that poor station accessibility in suburban China increased demand for airplanes, especially for distances under 1500 km. Meanwhile, Yuniaristanto, Sutopo, Hisjam, and Wicaksono (2024) predicted a 5.7% increase in Indonesia's electric motorcycle market share by 2030 through a System Dynamics model. For rail transportation, Shi, Hu, Zhou, and Liu (2014) showed that train speeds exceeding 50 km/h significantly increased market share, influenced by container capacity and transportation distance. In Sweden, Vigren (2017) reported lower ticket prices and higher ridership after the rail market was opened to competition. However, Tomeš and Jandová (2018) noted that while competition in Europe increased passenger numbers, it also reduced midrange income because of discount pricing.

Despite these international insights, existing research fails to provide a comprehensive understanding of how HSR introduction may reshape Thailand's transport sector. While studies in other countries have analyzed the effects of HSR on market competition, pricing strategies, and passenger modal shifts, similar investigations in Thailand remain scarce. Although the Thai government has projected significant passenger demand for HSR, there is limited empirical evidence to support these forecasts, particularly concerning how different traveler segments will respond to HSR services.

Prior studies in Thailand have primarily relied on traditional econometric models like the multinomial logit (MNL) model, which, despite their widespread use, may not fully capture the complexity of real-world travel behavior. These models assume fixed relationships between variables, limiting their ability to reflect nonlinear interactions and evolving market dynamics. In contrast, machine learning (ML) techniques offer a more advanced approach by uncovering hidden patterns in large datasets, yet their application in HSR adoption analysis remains underexplored.

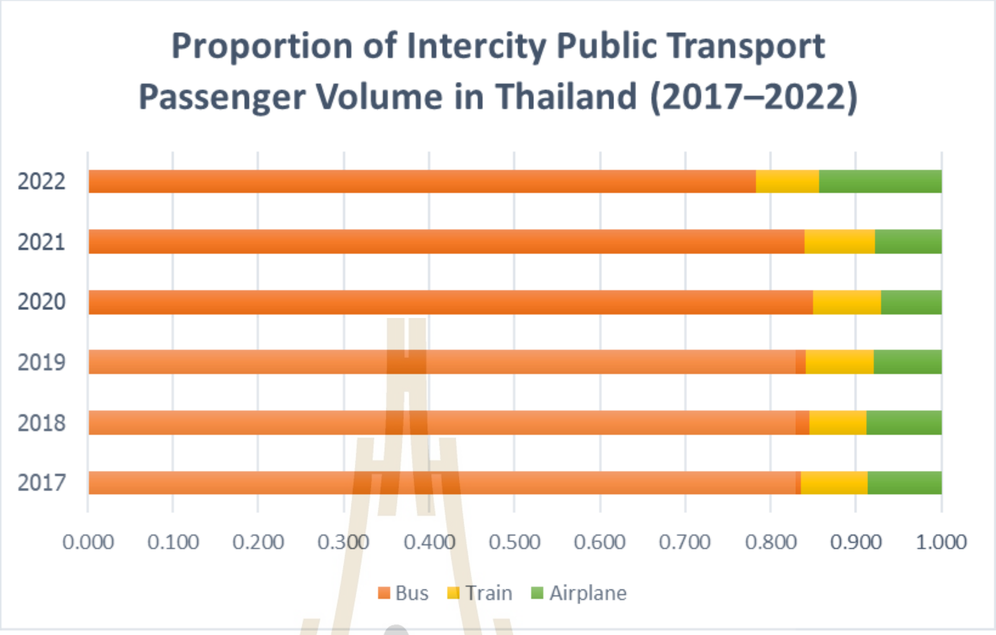


Figure 3.1 The proportion of intercity public transport passenger volume in Thailand (2017–2022)

Transportation mode choices impact system efficiency and sustainability, influenced by personal, area-specific, and perception-based factors. Personal factors like age, gender, income, and vehicle ownership affect preferences, with higher-income individuals favoring private vehicles (Armah et al., 2010; Cascetta et al., 2020; Givoni, 2006; Puan et al., 2019; Wójcik, 2019; D. Wu & Martín, 2022; X. Yu et al., 2018). Area-specific factors, including population density, land use, and accessibility, also shape mode choice, as seen in U.S. and German studies linking high-density areas to greater public transport use (Buehler, 2011; Shen, Sakata, & Hashimoto, 2006). Perception-based factors such as government policies, pricing, service quality, and environmental concerns further influence decisions. Research suggests passengers prefer public transport with waiting times under 10 min, while compact urban development enhances trust and convenience (Al Doori, 2017; J. Chen & Li, 2017; D. Wang & Zhou, 2017). In Thailand, key determinants of HSR adoption include travel time, station accessibility, service frequency, cost, waiting time, demographics, and private vehicle ownership. Travel purpose, such as commuting for work, study, or leisure, also plays a role, underscoring the need for improved regional transportation policies (Banyong et al., 2024).

Market segmentation significantly affects competition and economic efficiency. Monopolistic conditions can hinder innovation, increase costs, and limit accessibility, restricting mobility and economic growth (Osawa, Akamatsu, & Takayama, 2017; F. Xu et al., 2022). HSR infrastructure can mitigate these barriers by lowering transaction costs, improving connectivity, and promoting resource flow. However, Thailand lacks detailed studies on how airlines, buses, and trains will respond to HSR competition. This study addresses that gap by using machine learning to analyze key competitive factors and predict market share shifts. The findings will provide insights for policymakers to develop strategies that foster fair competition and sustainable transportation planning.

Accurate predictions of the travel demand for new transportation systems are essential for estimating future modal share, understanding travelers' attitudes, and predicting potential behavioral changes. Machine learning (ML) has emerged as a powerful analytical tool in the transportation sector, offering data-driven insights that enhance decision-making and predictive modeling. In recent years, ML techniques have been widely applied to assess travel behavior, optimize transportation networks, and forecast market trends. By leveraging large datasets, ML algorithms can identify patterns and relationships that traditional statistical models may overlook, making them particularly suitable for analyzing complex systems such as public transportation. Currently, machine learning methods are widely used for this analysis (Ammar Abulibdeh, 2023; L. Cheng et al., 2019; Darwish et al., 2024; Jenny Díaz-Ramírez et al., 2023; Elharoun et al., 2023; M. T. Kashifi et al., 2022; Kim, 2021; Xiaowei Li et al., 2024; Salas et al., 2022; F. Wang & C. L. Ross, 2018; X. Zhao et al., 2020). Among these methods, XGBoost is a popular ensemble learning algorithm known for its efficiency. However, recent research by Banyong et al. (2024) found that CatBoost outperforms other methods in analyzing traveler behavior. In this study, ML is employed to examine the transformative impact of high-speed rail (HSR) on Thailand's public transport landscape. Specifically, the CatBoost algorithm—a gradient-boosting method designed to handle categorical variables efficiently—is utilized to predict travel mode choices. CatBoost has demonstrated superior performance over traditional models, such as the multinomial logit (MNL) model, due to its ability to mitigate overfitting and handle imbalanced data effectively. Additionally, CatBoost offers robust performance in scenarios with high-dimensional data, making it a compelling choice for transportation demand modeling. To enhance interpretability, this study also employs Shapley Additive Explanations (SHAP), a widely used method for explaining ML model predictions. SHAP values provide insights into the relative importance of various factors influencing travel mode decisions, such as cost, service frequency, waiting time, and station accessibility. By incorporating SHAP analysis, the study improves the transparency of ML-driven findings and facilitates policy recommendations based on data-driven evidence.

The structure of this paper is as follows: Section 2 details the methodology, encompassing data collection procedures, model selection criteria, and analytical techniques employed in the study. Section 3 presents the research findings, including market share projections and an evaluation of model performance. Section 4 provides an in-depth discussion of the key findings, policy implications, and potential strategic responses from existing transport modes. Finally, Section 5 concludes the study with a summary of the main insights and recommendations for future research directions.

3.3 Materials and Methods

3.3.1 Stated Choice (SC) Experiment Design

The market share analysis of transportation modes, including buses, trains, airplanes, and HSR, began with data collection through questionnaires simulating travel mode choices. Data validation ensured completeness, correctness, and balance, minimizing deviations among users of different travel modes. The data were processed and analyzed using Python 3.12.7 via Anaconda, requiring approximately two hours of final preparation. The analysis used an Intel Core i9-13900H processor (2.60 GHz, 13th Generation) and 32 GB RAM. Market share calculations evaluated the usage proportions of each mode, and the multinomial logit model predicted travel mode choices based on influencing factors. Machine learning techniques, especially CatBoost, were employed to enhance prediction accuracy and address overfitting. Hyperparameter tuning optimized the model, and cross-validation ensured reliability by testing on multiple data segments. SHAP (Shapley Additive Explanations) was used to analyze factor importance, identifying the positive and negative influences on passengers' decisions. Finally, the elasticity analysis examined how changes in factors such as price or travel time affected market share, providing insights into traveler behavior. To ensure that the research results meet the required accuracy and performance criteria, a decision point was incorporated into the framework, allowing for model refinement through hyperparameter tuning and retraining if necessary. These results support the development of strategies and policies to improve transportation systems. Figure 3.2 illustrates the detailed conceptual framework.

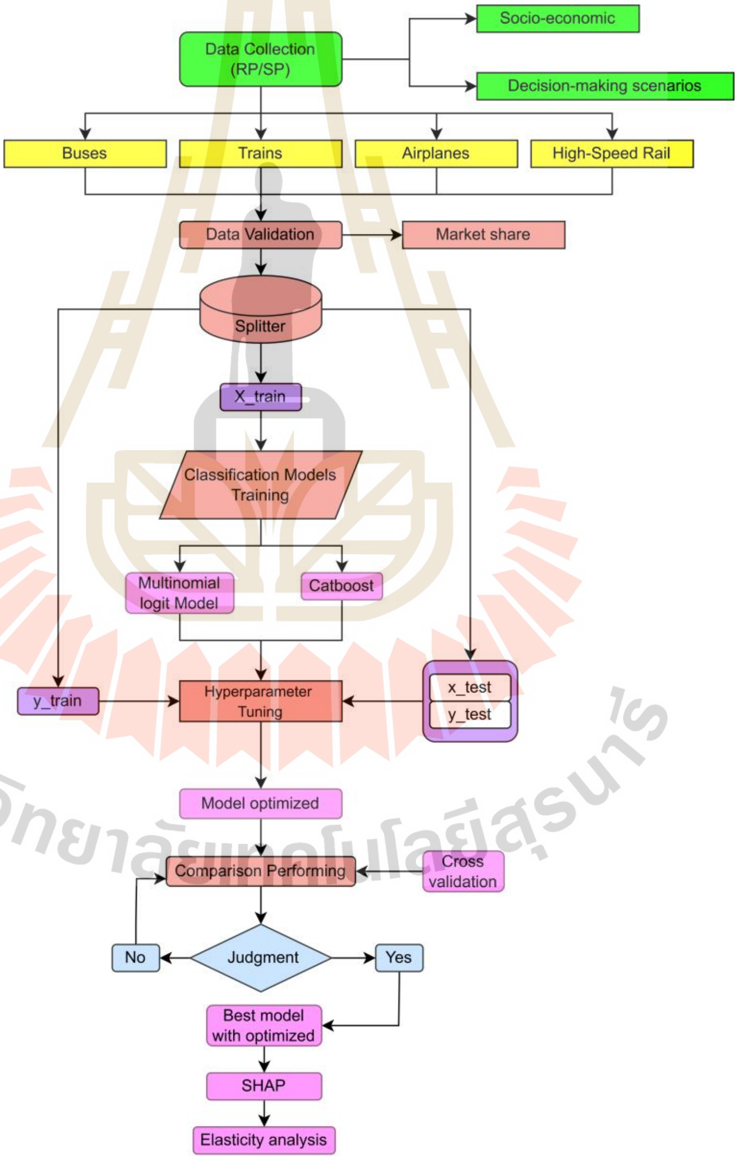


Figure 3.2 Research process flowchart

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.3.2 Stated Preference (SP) Survey

This survey aimed to predict shifts in travel mode proportions following the introduction of HSR and to examine the factors influencing decisions regarding future public transportation options, including buses, trains, airplanes, and HSR. Machine learning techniques were applied to cover Thailand's four main regions: north, northeast, central, and south. The study focuses on identifying key factors that influence travel mode choices and understanding their impact on passenger behavior (Banyong et al., 2024). The survey targeted significant provinces in each region with planned HSR stations, selecting the highest-potential provinces based on 2022 economic variables (NESDC, 2022), as illustrated in Figure 3.3. Regional differences in economic conditions and infrastructure were also considered because these factors play a crucial role in shaping travel behavior. By conducting the survey across multiple provinces, the analysis provides a comprehensive reflection of passenger travel behavior on a national scale (Srithongrung & Kriz, 2019).

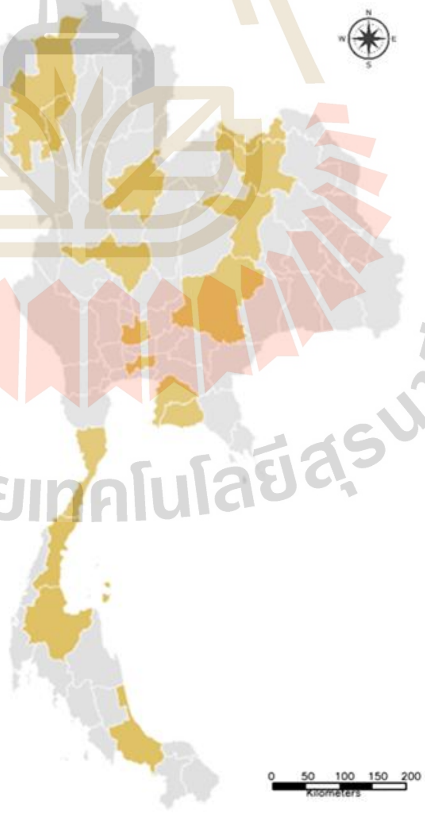


Figure 3.3 Provinces selected for data collection in Thailand

มหาวิทยาลัยเทคโนโลยีสุรนารี

Data collection involved one-on-one interviews with an 18-year-old sample group across 16 provinces. The total sample size was 3200 individuals (200 per province). Convenience sampling was used to ensure regional coverage, reduce bias, and increase the study's reliability. The collected data are robust enough for machine learning analysis, meeting the requirement of 50–1000 times the number of predicted classes. This study involves four classes (Pavlou et al., 2024). The survey had two parts: the first collected economic and social data, while the second gathered stated preferences (SPs). The SP section used a choice experiment (SC) where participants selected future travel modes, taking into account factors such as station approach duration, delay intervals, journey duration, financial expenditure, and operational regularity. This design accurately assesses travel behavior in the future public transportation context (Arencibia et al., 2015; Kujala et al., 2018). The results from this analysis are crucial for improving public transportation services. Key applications include increasing service frequency, optimizing routes, and setting appropriate fare structures. Additionally, the data offers valuable insights for evaluating the competitive potential of different transportation modes and supporting strategic business decisions.

3.3.2.1 Stated Choice (SC) Experiment Design

Table 3.1 presents a detailed summary of the attributes and levels employed in the stated choice (SC) experiment, emphasizing the essential factors influencing travel mode selection: station approach duration, delay periods, journey length, financial expenditure, and operational frequency. Each attribute has multiple levels to simulate various scenarios, ensuring accuracy in responses. The SC experiment focused on the Bangkok–Chiang Mai route (700 km), a crucial case study for analyzing travel behavior changes because of its competitive distance among HSR and other modes.

Because HSR infrastructure was not yet constructed during the study, data on travel time, costs, waiting time, and service frequency were derived from public sector estimates disseminated through the media (Economic Base, 2017).To ensure realistic scenario representation, other key parameters were referenced from existing research studies, incorporating empirical findings on passenger behavior, operational performance, and market conditions. While these estimates provided a structured foundation for modeling potential travel behavior shifts, their reliance on assumptions rather than real-world operations introduces uncertainty and potential biases, affecting the reliability and accuracy of the model's predictions.

A primary concern is that forecasted values for travel time, cost, and service frequency may differ from actual HSR operations once implemented. Projected travel times may not fully capture operational delays, infrastructure constraints, or network efficiency, while cost estimates might not accurately reflect future pricing strategies, government subsidies, or evolving market competition. Additionally, service frequency projections are often based on expected demand, which may vary due to economic conditions, policy changes, or unforeseen passenger adoption rates. Such discrepancies could lead to inaccuracies in predicting modal shifts, impacting the credibility of demand forecasts and policy recommendations.

Data were then adjusted to reflect realistic scenarios. To ensure that the hypothetical scenarios in the stated preference (SP) survey accurately represented real-world travel conditions, the collected data underwent an adjustment process based on official sources and observed travel patterns. Waiting times were derived from ticket reservation recommendations, ensuring alignment with actual passenger experiences, with data sourced from national ticket reservation centers for buses and trains (BusOnlineTicket.co.th, 2024), as well as airline ticketing websites (AirAsia move, 2024) for flights. For current travel modes (bus, train, and airplane), waiting times, travel times, costs, and service frequencies were based on official schedules and ticketing systems (State Railway of Thailand 2024), ensuring that the comparison between HSR and existing transport options was grounded in realistic and currently available travel conditions rather than theoretical estimates. Access times were adjusted based on mode-specific travel challenges, with longer access times for airports due to additional processes such as security checks and check-in procedures. The final dataset was cross-validated using historical travel data and user surveys to ensure plausibility and consistency. By integrating actual scheduling, pricing, and travel time data for existing transport modes, this approach ensured that the hypothetical scenarios closely mirrored real-world conditions, strengthening the reliability of the stated preference experiment for forecasting potential travel mode shifts after HSR implementation.

This study analyzed five variables by comparing three modes of current public transportation, with each variable defined at the same level and using two levels of HSR. The full design comprises 96 choice sets, calculated from ($3 \times 1^5 \times 2^5$). To reduce complexity for respondents, the researcher used a fractional factorial design, dividing the choice sets into eight subsets, with 12 choices per set. Each respondent evaluated only 12 choice sets (Hensher, Rose, & Greene, 2015). This design maintained efficient qualities such as level balance and minimized complexity in choices.

Table 3.1 Typical card from the SC sets.

Variable	Attribute	Bus	Train	Airplane	HSR	
		Levels			Levels 1	Levels 2
ACCESS	Access time (minute)	10	10	30	10	15
WAITING	Waiting time (minute)	15	10	120	15	10
TRAVEL	Travel time (minute)	720	70	135	190	220
COST	Travel cost (bath)	750	300	3000	1050	1400
FREQ	Frequency times (min)	30	150	120	190	220

3.3.2.2 Data Processing and Transformation

After collecting responses from the stated preference (SP) survey, the data underwent a structured transformation process to create a dataset suitable for machine learning analysis. This involved data cleaning, encoding categorical variables, structuring choice sets, and formatting variables for model input. First, responses were digitized and reviewed to remove incomplete or inconsistent entries, ensuring data reliability. Key preprocessing steps included removing missing or

มหาวิทยาลัยเทคโนโลยีสุรนารี

contradictory responses, standardizing numerical values, and encoding categorical variables to ensure compatibility with machine learning models.

Each respondent was presented with 12 choice sets, each containing four travel modes: bus, train, airplane, and high-speed rail (HSR). Since 3200 respondents participated, the final dataset consisted of 38,400 observations (3200 × 12). Each choice set included attributes such as access time, waiting time, travel time, and travel cost, which defined different travel conditions across modes. To prepare the dataset for predictive modeling, categorical and numerical data were structured accordingly. The dependent variable (Y) represented the chosen travel mode, converted into numerical categories (1 = bus, 2 = train, 3 = airplane, 4 = HSR), while the independent variables (X) included numerical values such as cost, waiting time, and travel time, which remained unchanged, and categorical variables like gender, income level, and car ownership, which were encoded using one-hot encoding or ordinal encoding as required.

The dataset was structured such that each row corresponded to a specific choice set presented to a respondent, with one mode being chosen (Y = 1) while others were not chosen (Y = 0). For example, if a respondent selected train, the dataset recorded train as Y = 1, while the other three modes were set to Y = 0. After transformation, the final dataset contained 38,400 records, incorporating both travel mode attributes and socio-demographic factors. To ensure robust model evaluation, the dataset was split into 80% training and 20% testing sets, optimizing performance assessment for machine learning algorithms.

3.4 Methodology

3.4.1 Multinomial Logit Model

The analysis is based on each option in the choice set offering passengers identifiable benefits. The likelihood of a traveler selecting an option increases with the utility gap between options. The multinomial logit (MNL) model is expressed as

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$P_{(i)} = \frac{e^{U_i}}{\sum_{j \in J} e^{U_i}} \tag{3-1}$$

where

$P_{(i)}$ is the likelihood of selecting option i;

U_i and U_j are the utilities of options i and j;

J is the number of alternatives.

3.4.2 Machine Learning Model

Machine learning (ML) has revolutionized transportation mode choice modeling, enhancing predictive accuracy and providing deeper insights into traveler behavior, crucial for urban planning and policy-making. Studies in Bengaluru and Chengdu confirm that ML models, such as random forests and XGBoost, outperform traditional multinomial logit (MNL) models in accuracy (Ghosh & Nagaraj, 2024; Tang et al., 2024). Beyond transportation, ML has also surpassed traditional methods in material production analysis (Bagherzadeh et al., 2023; Kazemi et al., 2024; Shafighfard, Kazemi, Asgarkhani, et al., 2024). Additionally, SHAP analysis improves model interpretability by identifying key decision factors, such as travel costs and demographics (Dahmen et al., 2023; Tang et al., 2024). Feature importance analysis further enhances policy-making, and ML models exhibit greater sensitivity to policy changes, as demonstrated in Alexandria City (Darwish et al., 2024).

Ensemble modeling has gained significant attention in machine learning applications due to its ability to enhance prediction accuracy across various domains. Tree-based techniques such as random forests, gradient boosting, and adaptive boosting are widely used to build ensemble models, with recent research demonstrating the effectiveness of hybrid ensembles that combine different ML techniques to improve predictive performance (Başakın et al., 2022; Zouzou & Citakoglu, 2023).

มหาวิทยาลัยเทคโนโลยีสุรนารี

Boosting algorithms, including XGBoost, LightGBM, and CatBoost, are commonly applied in classification problems and transportation mode choice modeling. XGBoost mitigates overfitting through regularization and second-order derivatives, while LightGBM enhances speed and scalability using gradient-based one-side sampling and leaf-wise tree growth. CatBoost, designed for handling categorical data, employs random permutation techniques to reduce bias and overfitting. These models have demonstrated high accuracy in complex datasets, reinforcing the importance of advanced ensemble methods in ML.

Among these, CatBoost proves particularly effective for predicting transportation mode choices, capturing nonlinear interactions between variables such as travel cost, time, convenience, and socio-demographic factors (BusOnlineTicket.co.th, 2024). Its ability to process categorical data efficiently with minimal preprocessing makes it ideal for transportation analysis (Xiaowei Li et al., 2024; J. Yu et al., 2024; L. Zhu et al., 2022; X. Zhu et al., 2024). By accurately modeling complex travel behavior, these ensemble methods provide valuable insights for transportation planning and policy development.

3.4.3 Categorical Boosting (CatBoost)

CatBoost is based on the Gradient Boosting Decision Tree (GBDT) framework, introducing the level method to optimize data utilization and reduce overfitting. Initially, all samples have equal weights, which adjust iteratively to emphasize misclassified samples (Dorogush et al., 2018; Prokhorenkova et al., 2018). This process continues until comprehensive training is achieved, enhancing model performance (Pham et al., 2020; Zhai et al., 2023).

The algorithm uses a priority-based process to address gradient bias in the traditional GBDT. It rearranges data samples through random permutations and constructs multiple sub-models, processing data in varied sequences to mitigate gradient estimation bias and improve generalization. Challenges include handling imbalanced datasets, where underrepresented class prediction accuracy may decline without proper parameter optimization—such as optimizing the step size, model depth, and number of iterations. (G. Huang et al., 2019).

The key CatBoost components include target-based statistics metrics like average target value per class. The algorithm constructs decision trees using boosting, where each tree corrects its predecessor's errors. Gradient calculations from the loss function further refine predictions, improving overall model accuracy (Prokhorenkova et al., 2018).The following formula is applied:

$$y_i^{(t)} = y_i^{(t-1)} + \eta f_t(x_i), \tag{3-2}$$

η is the learning rank.

3.4.4. Shapley Additive Explanations (SHAP)

SHAP explains the predictions of complex models, like gradient boosting machines, by showing how features affect outcomes. It addresses the "black box" issue, improving model interpretability, performance, and reliability (Hong et al., 2016; Lapuschkin et al., 2019; Mangalathu et al., 2020; Mao et al., 2021). SHAP values are displayed in ranked diagrams, indicating each feature's impact and direction on predictions, with red and blue colors showing positive and negative effects, respectively (Adland et al., 2021; Vega & Aznarte José, 2020).

The SHAP value uses this formula for calculation:

$$\varphi_k(val) = \sum s \subseteq N \setminus \{i\} \frac{|s|!(n-|s|-1!}{n!}(val(s \cup \{i\}) - val(s)), \tag{3-3}$$

In this context, "val" represents the feature's relevance to the algorithm's target and $\varphi_k(val)$ enotes the weighted sum of feature contributions to the target result across all feature combinations. The term $\frac{|s|!(n-|s|-1!}{n!}$ explains the weight of |s|, where val(s) is the expected value of |s|. Here, "n" is the total number of features, "s" is a subset of the model features, and "i" indicates the vector of feature values requiring interpretation (Adland et al., 2021; Cascetta et al., 2020; Lundberg et al., 2020; Lundberg & Lee, 2017; Shafighfard, Kazemi, Bagherzadeh, et al., 2024).

มหาวิทยาลัยเทคโนโลยีสุรนารี

This method determines the SHAP of all features through a linear aggregation of standardized attributes. It appears as follows:

$$g(x)=l(x_8)=\varphi_0+\sum_{t=1}^{n}\varphi_t x_8^{i} \tag{3-4}$$

where n is the number of features and φ_0 is the standard value for the undefined features. For feature t, the SHAP value is represented by φ_t, whereas x_8 signifies the vector of the condensed input variables (Shafighfard, Kazemi, Bagherzadeh, et al., 2024).

3.4.5. Market Share

Market share in public transportation refers to the proportion of users of buses, trains, BTS, or ships compared to other modes such as private cars, motorcycles, bicycles, or walking. It is crucial for evaluating the efficiency and attractiveness of public transportation within an area or country. Factors affecting market share include infrastructure accessibility, high service frequency, cost-effectiveness compared to private cars, especially in congested cities, and travel time benefits during rush hours. Increased environmental awareness and government policies promoting public transportation also play significant roles (Ammar Abulibdeh, 2023).

Regarding market share measurement, this research references Lackner and Zulehner (2020) as follows:

$$MS_{i,T}=\frac{Rev_{i,T}}{\sum_{i\in I} Rev_{i,T}}, \tag{3-5}$$

มหาวิทยาลัยเทคโนโลยีสุรนารี

Remark

$MS_{i,T}$ is the market share of the sample traveling mode i in the year T;

$Rev_{i,T}$ is the quantity of customers who use sample traveling mode i in the year T;

$\sum_{i\in I} Rev_{i,T}$ is the quantity of customers who use all traveling modes in year T.

A high market share provides companies with greater market power (M.-W. Wu & Shen, 2011), especially in incomplete competition markets (Y Wang, Sun, Mark-up, & D, 2018).

3.4.6. Elasticity Analysis

Elasticity analysis characterizes the sensitivity of socioeconomic characteristics and trip condition variables to shifts in mode choice through cross-elasticity computations. The goal is to examine the relationship between the probability of choosing a specific transportation mode for current and future selections and the percentage of significant independent variables.

The socioeconomic traits of the passengers and circumstances of their journeys were two important independent elements considered in this research, which was connected to direct elasticity computations (Larranaga, Arellana, & Senna, 2017).

$$E_{Piq,Xikq} = \theta_{ik}.X_{ikp}\left(1 - P_{iq}\right), \tag{3-6}$$

$$E_{Piq,Xjkq} = -\theta_{jk}.X_{jkp}P_{jq}, \tag{3-7}$$

Thus, given a marginal change (1% increase) in a particular variable X_{ikp} relative to the base scenario, $E_{Piq,Xjkq}$ represents the elasticity of the choice probability of the alternative i for the individual q (Piq) of selecting the Ai mode.

θ is a constant, and $E_{Piq,Xjkq}$ is the elasticity of the probability of selecting the Ai mode, taking into account slight changes in the value of the th variable of the alternative Aj for the individual q.

3.5 Results

3.5.1 Descriptive Analysis

Traveler data in Thailand was collected through self-administered questionnaires from a total sample of 3200 respondents. After data collection, the completeness and reliability of the data were extensively checked to ensure all necessary questions were answered. Initial statistical calculations, such as mean,

มหาวิทยาลัยเทคโนโลยีสุรนารี

skewness, and kurtosis, were conducted to identify any abnormalities in the data distribution (Table 3.2).

Table 3.2 Data results of the sample

Variable	Description	Categorical Variable (%)	Mean	*SD*
Switching the current mode to the HSR	Bus ▶ high-speed rail	0.754	-	-
	Train ▶ high-speed rail	0.744		
	Airplane ▶ high-speed rail	0.844		
GENDER	Male = 1	53.002	0.530	0.499
	Female = 0	46.998		
	Total	100		
HOUSEHOLD_SIZE	Family members		3.154	1.165
	There is one person = 1	10.006		
	There are two people = 2	19.387		
	There are three people = 3	27.829		
	There are four people = 4	30.801		
	There are more than four people = 5	11.976		
	Total	100		
CHILDREN	There are children in the family (age less than 18 years old) = 1	40.869	0.591	0.492
	There are no children in the family (age less than 18 years old) = 0	59.131		
	Total	100		
HOUSEHOLD_INCOME	Less than 15,000 = 1	11.069	3.009	0.915
	15,000–30,000 = 2	48.906		
	30,001–45,000 = 3	28.268		
	More than 45,000 = 4	11.757		
	Total	100		
N_CARS	Have own car = 1	78.111	1.049	0.800
	Do not have their own car = 0	21.889		
	Total	100		
OBJ1	Travel plan for study/work, Yes = 1	33.333	0.333	0.471
	Other plans besides study/work = 0	66.667		
	Total	100		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 3.2 Data results of the sample (Continued)

Variable	Description	Categorical Variable (%)	Mean	*SD*
OBJ2	Travel plan for leisure/vacation		0.510	0.500
	Yes = 1	51.032		
	Other plans besides leisure/vacation = 0	48.968		
	Total	100		
OBJ3	Travel plan for shopping		0.126	0.332
	Yes = 1	12.633		
	Other plans besides shopping = 0	87.367		
	Total	100		
TRAVEL_ FREQUENCY	Frequency of travel across the provinces per year		2.279	1.138
	Frequency for travel across the provinces per year: 1–3 times = 1	31.957		
	Frequency for travel across the provinces per year: 3–6 times = 2	30.894		
	Frequency for travel across the provinces per year: 6–9 times = 3	14.415		
	Frequency for travel across the provinces per year: more than nine times = 4	22.733		
	Total	100		

Note: *SD* = standard deviation.

The survey indicates that HSR has strong potential to attract passengers from existing travel modes. Notably, those who currently travel by buses (75.4%), trains (74.4%), and airplanes (84.4%) are likely to switch to HSR. There is a demographic variation in HSR preference, with males slightly more inclined to choose HSR than females (53.002% and 46.998%, respectively). Medium and large households, particularly those with four members (30.801%), show the highest proportion of HSR preference. Medium-income households (15,000–30,000 baht) also display the highest proportion (48.906%). Thus, access to private cars in the sample is high at 78.111%, affecting travel via HSR. The primary purpose of travel being leisure or tourism (51.032%) underscores HSR's role in facilitating interprovincial travel for various

มหาวิทยาลัยเทคโนโลยีสุรนารี

activities. Furthermore, the descriptive statistical analysis confirmed a normal distribution, as the skewness (SK) values remained within the acceptable range of −3.0 to +3.0, while the kurtosis (KU) values did not exceed 10.0 (Champahom et al., 2023).

3.5.2 Statistical Analysis Verses of Machine Learning Modeling

This study utilizes three models to predict travel mode choice in Thailand, employing 10-fold cross-validation to improve efficiency and address overfitting and underfitting issues. The analysis contrasts the multinomial logit (MNL) and CatBoost models, with CatBoost fine-tuned through hyperparameter optimization to enhance performance. Specifically, Table 3.3 presents the optimized hyperparameter values obtained using the Grid Search algorithm, where the tree depth was set to 6, the total number of iterations to 700, the L2 penalty coefficient (l2_leaf_reg) to 1, and the learning rate to 0.1. These values were selected to balance model complexity and generalization ability, ensuring the best trade-off between predictive accuracy and overfitting prevention. Additionally, the dataset was split into an 80:20 ratio for training and testing, ensuring that the model was trained on a sufficiently large dataset while preserving a validation set for robust performance evaluation.

A 10-fold cross-validation approach was employed, dividing the dataset into 10 equal parts, with the model trained on nine parts and validated on the remaining one, repeating the process across all sections. This method improves model reliability by ensuring that performance metrics are averaged over multiple iterations, reducing the risk of biased evaluation. Model performance was assessed using accuracy, recall, precision, F1-score, and AUC, with averages computed for comparison, as shown in Table 3.4.

To address class imbalance, under sampling techniques were applied, particularly for overrepresented travel modes. This approach ensures that the model does not disproportionately favor dominant classes, such as bus travel, while improving its ability to distinguish minority classes, such as airplane and HSR users. Additionally, SHAP analysis was performed to identify key factors influencing public transportation mode choices, as illustrated in Figure 3.4. The results provide insights into the most influential variables affecting travelers' decisions, supporting policy recommendations for optimizing Thailand's transport system.

Accuracy measures the proportion of correctly classified instances among the total instances. It provides a general measure of model performance.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3-8}$$

Recall quantifies the proportion of actual positive cases correctly identified by the model. It is important when the cost of false negatives is high.

$$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3-9}$$

Precision measures the proportion of correctly predicted positive cases among all cases predicted as positive. It is particularly useful when the cost of false positives is high.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3-10}$$

The F1-score provides a balance between precision and recall, particularly useful in cases where the dataset is imbalanced. It is calculated using the harmonic mean of precision and recall.

$$\text{F1} - \text{Score} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{3-11}$$

The AUC measures the model’s ability to distinguish between classes. It is derived from the Receiver Operating Characteristic (ROC) curve, where a higher AUC value indicates a better performing model.

$$\text{AUC} = \frac{1}{2}(\text{Recall} + \text{Specificity}) \tag{3-12}$$

These meter values were calculated from a confusion matrix that featured the following: TP is a true-positive rate; TN is a true-negative rate; FP is a false-positive rate; and FN is a false-negative rate.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 3.3 Hyperparameter values were extracted using the Grid Search algorithms for CatBoost

Model	Parameter	Description	Value
CatBoost	Depth	Tree depth limit	6
	Iterations	Total iteration count	700
	l2_leaf_reg	L2 penalty coefficient	1
	learning_rate	Range for step size: [0, 1]	0.1

Table 3.4 Comparison of average evaluation metrics between the MNL and XGBoost models

	MNL Model					CatBoost				
	Accuracy	AUC	Precision	Recall	F1-Score	Accuracy	AUC	Precision	Recall	F1-Score
Total	0.749	0.879	0.700	0.749	0.689	0.853	0.948	0.848	0.853	0.832
Bus			0.499	0.129	0.204			0.790	0.490	0.600
Train			0.776	0.948	0.854			0.800	0.530	0.640
Airplane			0.781	0.944	0.855			0.840	0.200	0.320
HSR			0.746	0.978	0.846			0.860	0.970	0.910

The comparison between the multinomial logit (MNL) model and the CatBoost model in Table 3.4 highlights CatBoost's overall superiority in predicting travel mode choices. The most critical performance metrics in assessing model effectiveness are accuracy and AUC (area under the curve), as they indicate general predictive reliability and the model's ability to distinguish between travel modes. CatBoost significantly outperforms MNL in both measures, achieving an accuracy of 0.853 and an AUC of 0.948, compared to MNL's 0.749 and 0.879, respectively. Additionally, precision, recall, and F1-score are key indicators for class-specific performance, with F1-score being particularly crucial as it balances precision and recall, ensuring a model's ability to make consistent predictions across different travel modes. For high-speed rail (HSR) travel, CatBoost achieves the best performance, with an F1-score of 0.910, precision of 0.860, and recall of 0.970, outperforming the MNL model, which records an F1-score of 0.846, precision of 0.746, and recall of 0.978. This underscores CatBoost's strength in predicting HSR adoption, making it the most effective model for this mode. For bus travel, CatBoost substantially outperforms MNL, achieving an F1-score of 0.600 (0.204

for MNL), with higher precision (0.790, 0.499) and recall (0.490, 0.129). This suggests that CatBoost captures bus travelers more accurately, significantly reducing misclassification errors. For train travel, the MNL model exhibits better recall (0.948) and F1-score (0.854) than CatBoost (recall of 0.530, F1-score of 0.640), despite CatBoost having slightly higher precision (0.800, 0.776). This indicates that CatBoost misses a considerable number of train users, making MNL the preferred model for this mode. For airplane travel, MNL achieves a higher F1-score (0.855) and recall (0.944) than CatBoost, which, despite having a slightly better precision (0.840, 0.781), suffers from a substantial drop in recall (0.200), leading to a much lower F1-score (0.320). This suggests that CatBoost struggles to capture airplane users effectively, likely misclassifying them into other modes. Overall, CatBoost emerges as the superior model for predicting travel mode choices, particularly for HSR and bus travel, where it significantly enhances classification accuracy.

Figure 3.4 illustrates the significance of various features, determined by their average absolute Shapley values. Features with higher SHAP values have a greater impact on public transportation mode choices in Thailand. The results highlight that COST, FREQ, and WAITING are the most influential factors. While TRAVEL and ACCESS also contribute to the predictions, their impact is relatively lower. Other variables, such as HOUSE_SIZE, FR_TRA, and MON_HOUSE, have progressively smaller effects.

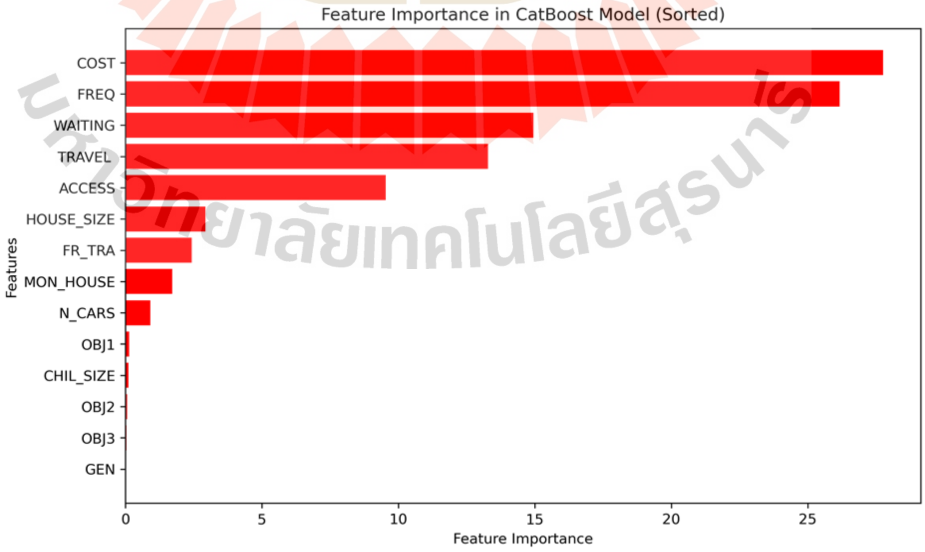


Figure 3.4 Feature importance of using public transport by SHAP

มหาวิทยาลัยเทคโนโลยีสุรนารี

As presented in Table 3.5, the analysis of future travel mode preferences in comparison to HSR identifies key factors influencing decisions among buses, trains, and airplanes. Household size (HOUSE_SIZE) positively correlates with travel mode selection, indicating that higher-income households may prefer comfortable travel choices. Conversely, household income (MON_HOUSE) negatively impacts the likelihood of selecting other modes, suggesting that households with greater financial resources tend to be more likely to opt for HSR. Car ownership (N_CARS) decreases the likelihood of selecting buses and trains but increases the probability of selecting airplanes. Simulation results indicate that factors such as station approach duration, delay intervals, monetary costs, journey length, and operational frequency significantly influence the choice between buses, trains, airplanes, and HSR. Longer HSR access times increase the probability of choosing other modes, with train usage increasing 3.3-fold, followed by airplanes and buses. Lengthy HSR waiting times lead to a 66.7-fold increase in train choice, whereas bus and airplane options decline, emphasizing the importance of minimal waiting times for HSR adoption. When HSR travel times are extended, passengers are much more likely to opt for airplanes (81.9-fold increase), with train choice preference increasing 12.2-fold and bus preference decreasing. Rising HSR costs lead to a significant increase in the train option (235.6-fold) as passengers prioritize financial value, whereas the airplane option rises modestly (7.3-fold). Conversely, higher HSR service frequency reduces the likelihood of choosing buses (374.3-fold reduction), trains, and airplanes, indicating that frequent services attract more passengers to HSR. The model's statistical measures—Cox and Snell R Square (0.221), Nagelkerke R Square (0.415), and McFadden R Square (0.328)—demonstrate its efficiency in explaining results, highlighting the importance of cost, convenience, and service features in determining travel mode preferences.

Comparative analysis shows that CatBoost outperforms the MNL model in evaluating travel mode choices. Its ability to manage unbalanced data and mitigate overfitting is crucial for complex predictions. Additionally, CatBoost effectively manages missing data and enhances the accuracy of travel mode predictions, even when combined with sampling techniques. The SHAP analysis identifies key factors influencing public transportation options in Thailand, with price being a significant factor in competition. Pricing strategies, such as accessible and flexible structures including flat-rate tickets, can attract cost-sensitive passengers. Research reveals that appropriate pricing not only increases public transportation usage but also decreases private car dependency, alleviating congestion and pollution. A well-designed pricing structure enhances the long-term sustainability and competitiveness of public transportation against other travel modes (Bursa, Mailer, & Axhausen, 2022; Da Silva et al., 2022; Daganzo & Ouyang, 2019; Jara-Díaz, Gschwender, & Hörcher, 2023; Y. Liu, Wang, & Xie, 2019).

Table 3.5 Model parameter estimates for using the HSR, future mode choice (HSR is the reference)

	Bus				Train				Airplane			
	β	S.E.	Sig.	Exp (β)	β	S.E.	Sig.	Exp (β)	β	S.E.	Sig.	Exp (β)
CONST	1.827 *	0.730	0.012	6.215	–9.101 **	0.794	0.001	0.001	–7.003 **	1.029	0.001	0.001
GEN	–0.063	0.043	0.138	0.938	–0.041	0.042	0.327	0.938	0.038	0.050	0.477	1.038
HOUSE_SIZE	0.399 **	0.022	0.001	1.490	0.332 **	0.022	0.001	1.490	0.331 **	0.026	0.001	1.392
CHIL_SIZE	–0.050	0.050	0.314	0.951	0.051	0.050	0.302	0.951	0.093	0.060	0.119	1.097
MON_HOUSE	–0.172 **	0.026	0.001	0.841	–0.097 **	0.026	0.001	0.841	–0.139 **	0.031	0.001	0.870
N_CARS	–0.247 **	0.031	0.001	0.781	–0.339 **	0.031	0.001	0.781	0.073 *	0.034	0.030	1.075
OBJ1	0.420 **	0.148	0.004	1.521	0.314 *	0.140	0.025	1.521	–0.429 **	0.155	0.006	0.651
OBJ2	0.389 **	0.146	0.008	1.475	0.280 *	0.140	0.044	1.475	0.175	0.151	0.246	1.191
OBJ3	0.115	0.158	0.467	1.121	–0.182	0.153	0.232	0.833	–0.078	0.165	0.634	0.924
FR_TRA	–0.259 **	0.021	0.001	0.771	–0.279 **	0.021	0.001	0.756	–0.136 **	0.024	0.001	0.872
ACCESS	0.776 *	0.311	0.012	2.172	1.194 **	0.312	0.001	3.300	0.797 *	0.346	0.021	2.218
WAITING	–6.546 **	0.866	0.001	0.001	4.200 **	0.895	0.001	66.686	–6.520 **	0.967	0.001	0.001
TRAVEL	–13.470 **	0.660	0.001	0.001	2.500 **	0.669	0.001	12.182	4.405 **	0.975	0.001	81.859
COST	–1.333 **	0.278	0.001	0.263	5.462 **	0.271	0.001	235.568	1.990 **	0.426	0.001	7.315
FREQ	5.925 **	0.154	0.001	374.278	–5.490 **	0.173	0.001	0.004	0.802	0.459	0.081	2.229

Note: Reference is the HSR, * = p-value < 0.05, ** = p-value < 0.01. Cox and Snell R Square: 0.221, Nagelkerke R Square: 0.415, McFadden: 0.328.

A well-organized service timetable is crucial for enhancing the competitiveness of public transportation. Passengers value convenience, and systems with regular

มหาวิทยาลัยเทคโนโลยีสุรนารี

schedules and high service frequency better meet their needs, particularly during peak hours or intercity travel. Research indicates that frequent services provide comfort and encourage increased usage. Transportation systems that adjust schedules to match demand, such as increasing service frequency during rush hours, become more competitive by reducing waiting times and instilling confidence in passengers to plan their trips efficiently. Effective timetable management gives transportation systems a unique advantage over those with inconsistent or irregular schedules (Lachapelle & Boisjoly, 2023; Orlando, Baquela, Bhouri, & Lotito, 2023; Tirachini, Godachevich, Cats, Muñoz, & Soza-Parra, 2022).

Waiting time plays a crucial role in passenger satisfaction and retention. Shorter waiting times enhance reliability and comfort, encouraging repeat usage. Passengers who experience punctual schedules and minimal delays tend to develop positive attitudes toward the transportation system, leading to continued patronage. Efficiently managing waiting times is thus a critical competitive advantage (Drabicki, Cats, Kucharski, Fonzone, & Szarata, 2023; Lachapelle & Boisjoly, 2023; Shelat, Cats, & van Cranenburgh, 2022).

Travel time is crucial for public transportation’s competitiveness. Passengers prioritize fast, efficient travel that minimizes delays and ensures punctuality. Systems providing shorter and more reliable travel times are highly appealing to frequent travelers and those aiming to reduce their daily commute. Studies show that decreasing travel time and enhancing service reliability significantly enhance passenger satisfaction and encourage repeat usage. Reliable services reduce stress and create a positive travel experience, reinforcing the system’s reputation as a dependable option. Transportation systems that effectively manage both travel time and punctuality gain a competitive advantage and attract more passengers (Esfeh, Kattan, Lam, Esfe, & Salari, 2020; Lee, Kawakita, Huai, Lo, & Zhang, 2024; Zang, Xu, Qu, Chen, & Chen, 2022).

Access time to stations is a critical factor in public transportation decisions. Passengers are more inclined to use public transportation if station access is easily accessible and time-efficient. Systems with well-distributed stations covering a wide range of areas and providing easy access have a competitive advantage. These systems cater to passengers seeking fast, comfortable travel and help reduce private car usage

(He, Zhao, & Tsui, 2018; Xinghua Li et al., 2022; Soczówka, Kłos, Żochowska, & Sobota, 2021; Tanwar & Kumar Agarwal, 2024). Key determinants of public transportation choices include household size, intercity travel frequency, household income, and vehicle ownership. Lu, Prato, Sipe, and Corcoran (2023) found that larger households use public transportation more often due to the high cost and inconvenience of traveling together by private vehicle. Providing group fares and designing systems with convenient access can make public transportation more appealing to larger households. Frequent city travelers prioritize affordable costs, efficient travel times, and frequent schedules (Banyong et al., 2024; Y. Huang et al., 2014). Public transportation is often relied upon by low-income households, whereas high-income households prioritize time savings and are willing to pay for comfort and speed. Binsuwadan, Wardman, de Jong, Batley, and Wheat (2023) found a direct relationship between income levels and the perceived value of time savings, leading high-income passengers to opt for faster transportation modes. Moreover, households with more vehicles tend to prefer private transport, which necessitates strategies such as value-based pricing, increased service frequency, and reduced waiting times to encourage public transport usage (Ibraeva & Sousa, 2014).

The CatBoost model highlights several key factors influencing the adoption of high-speed rail (HSR), particularly price, service rounds, waiting time, travel time, and access time. Price is a major factor, as travelers are willing to pay more for time savings—especially for longer trips—with studies showing that in Mumbai, travelers are prepared to pay over 300 INR for each hour saved, and in Italy, they are willing to pay 40% more for HSR tickets (Carteni et al., 2017; Karmarkar et al., 2023). In addition, short service rounds contribute to more frequent departures, thereby reducing waiting times and enhancing overall operational efficiency, which in turn boosts passenger convenience and satisfaction. Furthermore, reduced waiting times are linked to improved perceptions of service quality; for example, in China, HSR services have been shown to reduce travel times by an average of 10 h, with even more significant benefits in less developed regions (L. Liu & Zhang, 2018). Travel time itself is a critical factor, as the substantial time savings offered by HSR compared to conventional modes of transport can heavily influence passengers' choices, particularly during peak hours

[103–105]. Similarly, access time to HSR stations is vital; improvements in station accessibility—especially when integrated with efficient public transit—enhance the overall travel experience and make HSR a more attractive option in larger cities (Yongsheng Zhang et al., 2024; Zhou et al., 2020). Ultimately, these factors collectively enhance the appeal of HSR services, making them a competitive and cost-effective option for travelers.

3.5.3 Market Share Shifts in Public Transportation

Table 3.6 illustrates the current and projected market share for public transportation modes in Thailand, including the adoption of HSR. When considering the current market share, the bus has the highest share at 78.25%, airplanes at 14.28%, and trains at 7.47%. However, with the introduction of HSR, significant shifts are anticipated: bus usage is expected to drop to 4.76%, train usage to 5.11%, and airplane usage to 1.22%. HSR is predicted to have a high market share of 88.91%. This trend indicates a future shift toward HSR among public transportation users in Thailand.

Table 3.6 Market share of public transportation mode choice in Thailand (2029)

Mode	Current Market Share (%)	Transition to the HSR	Predicted Market Share (%)
Bus	78.25	59.82	4.76
Train	7.47	57.98	5.11
Airplane	14.28	84.55	1.22
HSR	0	0	88.91

Market share predictions indicate that HSR will dominate intercity travel once launched, capturing 88.91% of the market, whereas buses, trains, and airplanes will decrease to 4.76%, 5.11%, and 1.22%, respectively. This shift highlights passengers' high expectations for HSR's speed, comfort, and safety compared to traditional modes. HSR's ability to significantly reduce travel time, particularly for long distances, aligns with travelers' demands for rapid and reliable services (H. Wang & Huang, 2024). Passengers expect HSR to save time and enable confident trip planning. Features such as spacious seating, onboard amenities, and consistent service further enhance its appeal. HSR also offers direct travel to destinations, avoiding the connectivity issues often faced with buses and airplanes. Additionally, passengers perceive HSR to provide superior safety standards compared to other modes, reinforcing its attractiveness (Anagnostopoulos, 2024). Lastly, fare expectations are crucial; passengers expect HSR fees to reflect service quality, balancing speed, convenience, and cost, making HSR an appealing option for future travel (Hu, Huang, Gao, & Lin, 2023; Meesit, 2024).

Table 3.7 Market Share of Public Transportation Mode Choice in Thailand (2029) Estimated Using the Multinomial Logit Model

Mode	Current Market Share (%)	Transition to the HSR	Predicted Market Share (%)
Bus	78.25	55.72	13.89
Train	7.47	77.31	5.33
Airplane	14.28	84.50	4.34
HSR	0	0	76.44

Table 3.7 reports the projected market shares of public transportation modes in Thailand for 2029 under the HSR introduction scenario, as estimated by the Multinomial Logit (MNL) model. The MNL results indicate that, following the deployment of HSR, the predicted market shares would be 13.89% for buses, 5.33% for conventional trains, 4.34% for air transport, and 76.44% for HSR. The transition matrix shows that the highest shift is expected from air passengers (84.50%) and train passengers (77.31%) to HSR, while more than half of bus passengers (55.72%) would also convert to the new mode. In contrast, the CatBoost-based estimation yields a markedly higher predicted HSR market share of 88.91%, with substantially lower post-HSR shares for buses (4.76%), trains (5.11%), and airplanes (1.22%). This divergence highlights that the CatBoost model anticipates a more pronounced substitution effect toward HSR, particularly from bus and air segments. Two methodological considerations may explain these differences. First, the MNL model is constrained by the Independence of Irrelevant Alternatives (IIA) property, which imposes proportional substitution patterns and may underrepresent asymmetric mode shifts when introducing a new alternative. Second, the CatBoost algorithm, as a gradient boosting decision tree method, flexibly captures complex, non-linear relationships, higher-order

interactions between explanatory variables, and heterogeneous traveler preferences across market segments. These capabilities allow CatBoost to identify stronger and more targeted substitution effects that MNL may not detect, particularly when mode attractiveness changes are non-proportional.

While both approaches predict that HSR will dominate the intercity travel market after its introduction, the magnitude of the shift differs considerably. The findings underscore the value of complementing traditional discrete choice modeling with advanced machine learning methods to capture nuanced behavioral responses and improve the robustness of market share forecasts in large-scale transportation policy evaluations.

3.5.4 Elasticity Analysis and Strategic Adaptations in Response to High-Speed Rail

The direct and cross-elasticity analysis evaluates how changes in key variables influence the probability of travel mode choices, specifically in relation to HSR (HSR). According to Table 3.8, the cost of HSR is a significant factor. A 1% increase in HSR costs reduces HSR usage by 1%, reflecting passengers' sensitivity to price. Conversely, if the costs of buses, trains, or airplanes increase by 1%, HSR usage rises, with airplane costs showing the greatest impact, increasing HSR usage by 16.204%, followed by buses (9.889%) and trains (9.478%). Service frequency has a clear positive relationship with HSR choice. A 1% increase in HSR waiting time reduces its usage by 1%. Meanwhile, increases in the service frequency of buses, trains, or airplanes lead to greater HSR adoption. Specifically, airplane service frequency caused a 16.230% rise in HSR usage, highlighting passengers' preference for HSR under competitive conditions. Waiting time is another crucial factor. A 1% increase in HSR waiting time reduces HSR usage by 1%. Conversely, longer waiting times for other modes, especially airplanes (16.154%), significantly increased HSR usage, followed by buses (9.858%) and trains (9.447%). Similarly, travel time impacts mode choice; a 1% increase in HSR travel time reduces HSR usage by 1%, but longer travel times for airplanes (15.796%), buses (9.632%), or trains (9.230%) encourage passengers to switch to HSR. Station access time also affects HSR usage. A 1% increase in HSR access time reduces its usage by 1%, while more difficult access to airplanes (15.082%), buses (9.180%), or trains (8.795%) drives passengers toward HSR due to its comparative convenience.

Social and economic factors, including household size, number of companions, income, and car ownership, significantly influence HSR usage. Larger households increase HSR adoption by 1%, particularly as airplane households grow (13.400%). Similarly, more companions boost HSR usage by 1%, especially for group travelers who find airplanes less suitable (14.637%). Higher household income increases HSR usage by 1%, with a notable 14.239% rise when air travel costs escalate. Additionally, a higher number of cars leads to a 1% increase in HSR usage, with airplane households showing a 14.934% rise in HSR adoption when travel costs increase.

The elasticity results provide crucial insights into demand responses and competitive dynamics with the introduction of a new transportation mode. These findings help transport planners assess demand shifts and develop strategies to promote HSR adoption while maintaining a balance with existing modes.

Table 3.8 Elasticity analysis

Variable	Mode	Bus	Train	Airplane	HSR
COST	Bus	1.00	0.239	−0.978	9.889
	Train	−0.943	1.00	−0.978	9.478
	Airplane	−0.906	0.958	1.00	16.204
	HSR	−0.993	−0.870	−0.997	1.00
FREQ	Bus	1.00	−0.988	−0.878	9.906
	Train	0.126	1.00	−0.883	9.494
	Airplane	0.849	−0.982	1.00	16.230
	HSR	−0.877	−0.998	−0.987	1.00
WAITING	Bus	1.00	−0.038	−0.910	9.858
	Train	−0.710	1.00	−0.914	9.447
	Airplane	−0.524	0.518	1.00	16.154
	HSR	−0.968	−0.899	−0.990	1.00
TRAVEL	Bus	1.00	−0.406	−0.348	9.632
	Train	−0.679	1.00	−0.373	9.230
	Airplane	−0.473	−0.062	1.00	15.796

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 3.8 Elasticity analysis (Continued)

Variable	Mode	Bus	Train	Airplane	HSR
	HSR	−0.965	−0.937	−0.931	1.00
ACCESS	Bus	1.00	−0.016	−0.947	9.180
	Train	−0.042	1.00	−0.949	8.795
	Airplane	0.571	0.553	1.00	15.082
	HSR	−0.895	−0.896	−0.994	1.00
HOUSE_SIZE	Bus	1.00	0.134	−0.218	8.115
	Train	0.134	1.00	−0.248	7.770
	Airplane	0.862	0.792	1.00	13.400
	HSR	−0.876	−0.881	−0.918	1.00
FR_TRA	Bus	1.00	−0.118	−0.425	8.898
	Train	−0.176	1.00	−0.447	8.523
	Airplane	0.351	0.393	1.00	14.6374
	HSR	−0.910	−0.907	−0.939	1.00
MON_HOUSE	Bus	1.00	0.020	−0.403	8.646
	Train	−0.088	1.00	−0.426	8.281
	Airplane	0.496	0.611	1.00	14.239
	HSR	−0.900	−0.893	−0.937	1.00
N_CARS	Bus	1.00	−0.199	−0.435	9.086
	Train	−0.269	1.00	−0.456	8.704
	Airplane	0.199	0.264	1.00	14.934
	HSR	−0.920	−0.916	−0.940	1.00

มหาวิทยาลัยเทคโนโลยีสุรนารี

The elasticity analysis highlights HSR’s ability to draw passengers and maintain a competitive advantage over other travel modes by leveraging key factors that influence passenger options. These factors include pricing, service frequency, waiting times, travel duration, station accessibility, and socioeconomic. HSR demonstrates a high sensitivity to price fluctuations, with passengers more inclined to switch to HSR when it offers competitive fares. Interestingly, when air travel costs increase, passengers often prefer HSR, finding it to be a more economical alternative (C. Jiang &

Wang, 2021; Ma, Wang, Yang, Zhang, & Zhang, 2019; Raturi & Verma, 2020; Xia, Jiang, Wang, & Zhang, 2019). HSR shines in its service frequency, providing consistent and frequent services that align with passenger expectations for convenience and reliability. This makes HSR a preferred choice for structured travel plans. Additionally, reduced waiting times give HSR a notable advantage over other modes, including airplanes and buses. Passengers value punctuality and swift service, making shorter waiting times a critical factor in their travel decisions (Nurhidayat et al., 2023; Yuan et al., 2021). HSR's ability to minimize travel time and ensure efficiency, particularly for long-distance trips, further enhances its attractiveness. It offers a considerable reduction in travel time compared to buses and trains, highlighting its value and comfort. Additionally, the ease of station access and the presence of conveniently located stations improve passenger satisfaction, making HSR a preferred option (P. Chen, Zhang, & Gao, 2024; Pan, 2024; Wei, Wang, Li, & Li, 2024).

The elasticity analysis Table 3.7 highlights key factors influencing HSR adoption, particularly the sensitivity of passengers to fare changes, waiting time, travel time, and access time. These findings provide critical insights for policymakers in designing effective pricing strategies, prioritizing infrastructure investments, and optimizing service operations to maximize ridership while maintaining financial sustainability.

The results indicate that HSR demand is highly elastic with respect to fare changes, meaning that even small price reductions can significantly increase ridership. However, indiscriminate fare reductions may negatively impact long-term revenue sustainability, particularly given the high capital and operational costs of HSR infrastructure. Policymakers must balance affordability with financial viability through dynamic pricing strategies, such as time-based pricing models that reduce off-peak fares to encourage travel during low-demand periods, improving seat utilization. Additionally, subscription-based and loyalty programs can offer discounts for frequent travelers while ensuring stable revenue streams, and tiered service pricing can provide different levels of comfort (e.g., economy vs. premium) to cater to diverse passenger needs without compromising overall profitability. To maintain financial sustainability, these pricing strategies must be data-driven and responsive to passenger behavior, ensuring that HSR remains attractive while covering operational costs. A comprehensive

มหาวิทยาลัยเทคโนโลยีสุรนารี

pricing model could integrate machine learning techniques to dynamically adjust fares based on demand fluctuations, enhancing efficiency and revenue management.

The elasticity results also show that station access time is a significant determinant of HSR adoption, suggesting that improvements in accessibility could lead to substantial increases in ridership. Long access times deter potential users, making it essential to integrate HSR with other transport networks seamlessly. Policymakers should consider enhancing first-mile and last-mile connectivity through dedicated feeder buses, metro linkages, and pedestrian-friendly infrastructure such as skywalks and bicycle lanes. Strategic station placement is also crucial, ensuring that HSR stations are conveniently located in key urban centers or near existing transport hubs to minimize transfer times. Furthermore, smart infrastructure investments, such as real-time travel information systems and multimodal journey planning apps, can improve the overall passenger experience and reduce perceived inconvenience. Reducing station access time does not only increase ridership but also fosters multi-modal integration, enhancing the competitiveness of HSR as part of a broader, well-connected transport ecosystem.

In addition to pricing and station access, the analysis highlights that waiting time significantly influences passenger decisions, emphasizing the need to minimize delays and improve service efficiency. Longer waiting times discourage usage, particularly for time-sensitive travelers, necessitating strategies such as increasing service frequency during peak hours to ensure minimal waiting time and reducing congestion at stations. Synchronizing schedules with other transport modes, such as buses, local trains, and metro systems, can enable smooth transfers and reduce total travel time. Additionally, implementing flexible ticketing systems that allow passengers to modify bookings with minimal penalties can increase confidence in HSR reliability. By optimizing service design, HSR can compete more effectively with alternative transport modes, particularly short-haul flights and long-distance buses, where schedule reliability is a key factor in passenger choice.

The elasticity results underscore the importance of balancing affordability, convenience, and financial sustainability in HSR planning. While lowering fares or improving station access can significantly increase ridership, these decisions must be weighed against the need for cost recovery and long-term viability. Infrastructure investments should be strategically aligned with demand patterns, ensuring that improvements generate tangible benefits in terms of increased passenger uptake. A holistic transport policy integrating HSR with existing transport modes through multimodal ticketing systems, optimized scheduling, and infrastructure enhancements will be essential in ensuring that HSR contributes to a sustainable, efficient, and inclusive public transport system in Thailand. Future research should explore the real-world impact of these policy measures, using empirical data to refine strategies that enhance both ridership and operational efficiency in a competitive transport landscape.

While the projected market share shifts suggest a significant decline in the usage of buses, trains, and airplanes following the introduction of HSR, it is crucial to explore how existing transport providers might respond strategically to mitigate ridership losses and retain competitiveness. Rather than viewing HSR as a direct competitor, incumbent transport modes can implement targeted counterstrategies to maintain market relevance and adapt to changing travel dynamics. These adaptations can be broadly categorized into fare adjustments, service frequency enhancements, route optimization, and multimodal integration, all of which align with sustainable transport planning principles.

One of the primary responses available to existing transport operators is implementing competitive and flexible pricing strategies. Bus and train operators may introduce dynamic fare adjustments, offering discounted off-peak pricing, seasonal promotions, or bulk ticketing options (e.g., monthly or annual passes) to attract cost-sensitive travelers. Airlines, facing substantial competition from HSR on short-haul routes, might leverage frequent flyer programs, promotional fares, and bundled services (such as flight-and-hotel packages) to retain business and leisure passengers who prioritize speed and convenience. Additionally, transport operators could introduce fare integration mechanisms, where passengers receive discounts when using multiple modes (e.g., bus + HSR), making multimodal travel more appealing.

Increasing service frequency and optimizing operational efficiency can serve as key countermeasures to offset passenger migration to HSR. Bus operators may increase

มหาวิทยาลัยเทคโนโลยีสุรนารี

express services on high-demand corridors or introduce premium services with added comfort and direct connectivity to urban centers. Similarly, regional and intercity rail services can adjust their schedules to provide more frequent departures, particularly in areas where HSR is not expected to have full coverage. Airlines may increase flight frequency on medium-to-long-haul routes, where HSR is less competitive, to focus on segments where speed and convenience remain a priority. Additionally, transport providers could explore technological advancements such as real-time scheduling, AI-driven demand forecasting, and digital ticketing systems to enhance passenger experience and reduce inefficiencies.

Rather than competing directly, existing transport providers can reposition their services to complement HSR and leverage network integration to sustain ridership levels. Bus operators can shift their focus toward feeder routes that provide first-mile and last-mile connectivity to major HSR stations, ensuring passengers from smaller towns and rural areas can seamlessly access high-speed rail services. Likewise, conventional rail networks could realign their services to operate as secondary intercity connections, synchronizing schedules with HSR departures and arrivals to facilitate seamless multimodal transfers. Airlines, facing declining demand on short-haul domestic routes, may pivot towards expanding regional and international connectivity, focusing on flights where HSR cannot serve as a viable substitute.

To ensure a balanced and sustainable transition within the transportation sector, policymakers should consider strategies that promote cooperation rather than competition between HSR and existing transport modes. Integrated ticketing systems, where passengers can book a single ticket covering multiple transport modes (e.g., HSR + bus or train), would improve accessibility and encourage multimodal travel. Additionally, investment in multimodal transport hubs—where bus, train, and HSR services are co-located—can enhance transfer efficiency and reduce travel friction. Targeted government subsidies or incentives for bus and train operators that provide crucial feeder services to HSR stations could help maintain equitable transport access, particularly in rural and underserved regions.

Beyond infrastructure investments, regulatory frameworks should also be adapted to support transport providers in adapting their business models to align with evolving passenger needs. Encouraging innovation through public–private partnerships (PPPs), digital transformation initiatives, and mobility-as-a-service (MaaS) platforms can further integrate various transport modes into a cohesive and user-friendly network. By fostering an ecosystem where HSR enhances rather than disrupts existing transport services, Thailand can achieve a more efficient, resilient, and sustainable transportation landscape.

In summary, while HSR is poised to become the dominant intercity travel mode, buses, trains, and airlines still have viable pathways to retain market share through targeted strategic adaptations. Fare adjustments, service frequency enhancements, and route optimization—when combined with integrated transport planning and supportive government policies—can ensure that all transport modes coexist and contribute to a well-connected mobility network. Future research should assess the effectiveness of these adaptation strategies through empirical case studies, evaluating how various transport operators adjust their services in response to HSR's real-world impact.

3.6 Conclusions and Policy Implications

This study investigates the impact of HSR on Thailand's public transportation market and evaluates the effectiveness of the CatBoost and MNL models in predicting factors influencing travel mode choices. The results indicate that HSR has the potential to significantly change passenger travel behavior in Thailand. CatBoost outperformed MNL in terms of accuracy (0.853 vs. 0.749) and AUC (0.948 vs. 0.879). Additionally, precision, recall, and F1-scores for CatBoost were 0.848, 0.853, and 0.832, respectively, demonstrating its ability to analyze complex travel behaviors. In contrast, MNL recorded lower values for precision (0.700), recall (0.749), and F1-score (0.689), indicating its comparatively lower predictive performance in capturing travel mode choice patterns. SHAP analysis identified key factors influencing mode choice, including cost, service frequency, waiting time, travel time, access time, household size, intercity travel frequency, income, and car ownership. Market share predictions indicated that HSR will dominate with 88.91%, drawing passengers from buses (78.25–4.76%), trains (7.47–5.11%), and airplanes (14.28–1.22%), showing a strong preference for HSR in Thailand.

Several key factors drive the shift from buses, trains, and airplanes to HSR, including cost, service frequency, waiting time, travel time, and station access time. A 1% increase in HSR costs leads to a 1% decrease in usage, while fare increases for buses, trains, or airplanes drive passengers toward HSR, with airplane fare hikes having the greatest effect (16.204%). Similarly, a 1% increase in HSR service frequency boosts usage by 1%, while higher airplane service frequency further enhances HSR's appeal (16.230%). Passengers are highly sensitive to waiting times; a 1% increase in HSR waiting time results in a 1% decline in usage, while longer waiting times for airplanes, buses, or trains significantly boost HSR adoption, particularly for airplanes (16.154%). Travel time follows the same pattern, where longer HSR travel times reduce usage by 1%, whereas increased travel times for competing modes increase HSR adoption, with airplanes exerting the strongest influence (15.796%). Station access time also plays a crucial role; a 1% increase in HSR access time reduces usage by 1%, while more difficult access to airplanes, buses, or trains pushes passengers toward HSR, with airplanes having the greatest impact (15.082%). These findings highlight the importance of optimizing HSR pricing, frequency, and accessibility to maximize passenger adoption, ensuring it remains a competitive and attractive travel alternative.

These findings highlight the transformative potential of HSR in Thailand and provide key insights for planning and expansion before its launch. To ensure successful adoption, policymakers should focus on strategic pricing, high service frequency, waiting time management, and improved station access. Competitive fares, discounts, loyalty programs, and integrated travel solutions can encourage early adoption. Enhancing station accessibility, efficient booking systems, and multimodal connectivity will further strengthen passenger confidence. Public awareness campaigns and trial services can facilitate smoother transitions. To retain market share, buses, trains, and airplanes should emphasize their unique advantages through competitive pricing, enhanced service quality, and improved accessibility. Strategies such as discounted fares, expanded service frequency, reduced waiting times, and superior onboard amenities can attract passengers. Integrated ticketing and multimodal partnerships will also be crucial in ensuring long-term competitiveness. These strategies will help optimize Thailand’s transportation landscape, balancing HSR adoption with the sustainability of existing travel modes.

3.7 Limitations and Further Research

This study relies on simulated data and public sector estimates for key high-speed rail (HSR) parameters, including travel time, cost, and waiting time, as actual infrastructure is not yet operational. While this approach is necessary given the current lack of real-world HSR data, it introduces potential biases and uncertainties that may impact the accuracy of market share predictions. The assumed values for travel parameters may not fully capture the actual passenger experience once the HSR system is implemented, potentially leading to deviations in projected ridership patterns.

One major limitation is the reliance on publicly available estimates for travel attributes, which may not precisely reflect future operational conditions. Real-world travel behavior is influenced by additional factors such as dynamic pricing strategies, service disruptions, infrastructure delays, and passenger perception of comfort and convenience, all of which are difficult to account for in a simulated environment. Moreover, the study assumes a static competitive landscape, whereas existing transport modes (buses, trains, and airlines) may adjust their pricing, service frequency, or promotional strategies in response to the introduction of HSR, affecting market dynamics in ways not captured by the current model.

To enhance the reliability of these findings, future research should incorporate post-implementation validation using actual passenger data once the HSR service is operational. Collecting empirical data on passenger preferences, travel times, real ticket prices, and service frequency will allow for a more precise assessment of HSR’s impact on travel behavior and market share. Comparative analyses between predicted and observed market trends will help refine forecasting models and improve their applicability for future infrastructure projects.

Additionally, conducting a sensitivity analysis would strengthen the robustness of these projections. By systematically varying key input parameters—such as increasing or decreasing ticket prices by 10%, adjusting waiting times by ±5 min, or modifying travel times based on potential congestion or delays—it would be possible to assess how fluctuations in travel conditions influence market share predictions. This approach would provide a confidence interval around the estimated adoption rates and offer insights into the stability of the findings under different operational scenarios. Further integration of probabilistic modeling techniques or Monte Carlo simulations could help account for uncertainties in future travel conditions and improve the predictive power of the study. Beyond these methodological improvements, future research should also consider the long-term behavioral adaptation of travelers. Initial ridership patterns may differ significantly from long-term adoption trends, as passengers gradually adjust their habits in response to HSR availability. Investigating how habit formation, word-of-mouth recommendations, and social influence contribute to sustained mode shifts could provide a more comprehensive understanding of HSR's long-term market position. Despite these limitations, the study presents valuable preliminary insights into the transformative potential of HSR in Thailand. By understanding potential adoption trends and market share projections, policymakers and transportation planners can design more informed strategies for infrastructure development. Future studies should integrate real-world data, more sophisticated modeling techniques, and behavioral research methodologies to refine these findings further. By doing so, decision-makers can develop more effective policies and investments that align with passenger needs and optimize the role of HSR within the broader transportation network.

3.8 References

Abulibdeh, A. (2023). Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis. *Transportation Research Interdisciplinary Perspectives 20*, 100852. doi:https://doi.org/10.1016/j.trip.2023.100852

Adland, R., Jia, H., Lode, T., & Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering & System Safety 209*, 107480. doi:10.1016/j.ress.2021.107480

AirAsia move. (2024). Flight. Retrieved from https://www.airasia.com/th/th

Al Doori, A. (2017). Waiting time factor in public transport by binary logistic regression. *Australian Journal of Basic and Applied Sciences, 11*(4), 72-76.

Anagnostopoulos, A. (2024). High-speed railway and safety: Insights from a bibliometric approach. *High-speed Railway, 2*(3), 187-196. doi:https://doi.org/10.1016/j.hspr.2024.08.004

Arencibia, A. I., Feo-Valero, M., García-Menéndez, L., & Román, C. (2015). Modelling mode choice for freight transport using advanced choice experiments. *Transportation Research Part A: Policy and Practice 75*, 252-267.

Armah, F. A., Yawson, D. O., & Pappoe, A. A. N. M. (2010). A Systems Dynamics Approach to Explore Traffic Congestion and Air Pollution Link in the City of Accra, Ghana. *Sustainability 2*(1), 252-265. doi:10.3390/su2010252

Avogadro, N., Pels, E., & Redondi, R. (2023). Policy impacts on the propensity to travel by HSR in the Amsterdam – London market. *Socio-Economic Planning Sciences, 87*, 101585. doi:https://doi.org/10.1016/j.seps.2023.101585

Bagherzadeh, F., Shafighfard, T., Khan, R. M. A., Szczuko, P., & Mieloszyk, M. (2023). Prediction of maximum tensile stress in plain-weave composite laminates with interacting holes via stacked machine learning algorithms: A comparative study. *Mechanical Systems and Signal Processing 195*, 110315. doi:https://doi.org/10.1016/j.ymssp.2023.110315

Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Seefong, M., . . . Jomnonkwao, S. (2024). A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand. *Results in Engineering, 24*, 103110. doi:https://doi.org/10.1016/j.rineng.2024.103110

Başakın, E. E., Ekmekcioğlu, Ö., Çıtakoğlu, H., & Özger, M. (2022). A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Computing and Applications 34*(1), 783-812.

Bergantino, A. S., & Madio, L. (2020). Intermodal competition and substitution. HSR versus air transport: Understanding the socio-economic determinants of modal choice. *Research in Transportation Economics, 79*, 100823. doi:https://doi.org/10.1016/j.retrec.2020.100823

Binsuwadan, J., Wardman, M., de Jong, G., Batley, R., & Wheat, P. (2023). The income elasticity of the value of travel time savings: A meta-analysis. *Transport Policy 136*, 126-136. doi:https://doi.org/10.1016/j.tranpol.2023.03.013

Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography 19*(4), 644-657.

Bursa, B., Mailer, M., & Axhausen, K. W. (2022). Intra-destination travel behavior of alpine tourists: a literature review on choice determinants and the survey work. *Transportation 49*(5), 1465-1516.

BusOnlineTicket.co.th. (2024). Book Thailand Bus Tickets Online. Retrieved from https://www.busonlineticket.co.th/

Cartenì, A., Pariota, L., & Henke, I. (2017). Hedonic value of high-speed rail services: Quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities. *Transportation Research Part A: Policy and Practice 100*, 348-365. doi:10.1016/j.tra.2017.04.018

Cascetta, E., Cartenì, A., Henke, I., & Pagliara, F. (2020). Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post evaluation and future perspectives. *Transportation Research Part A: Policy and Practice 139*, 412-428. doi:https://doi.org/10.1016/j.tra.2020.07.008

Champahom, T., Banyong, C., Hantanong, N., Se, C., Jomnonkwao, S., & Ratanavaraha, V. (2023). Factors influencing the willingness to pay for motorcycle safety improvement: A structural equation modeling approach. *Transportation Research Interdisciplinary Perspectives 22*, 100950. doi:https://doi.org/10.1016/j.trip.2023.100950

Chen, J., & Li, S. (2017). Mode choice model for public transport with categorized latent variables. *Mathematical Problems in Engineering, 2017*.

Chen, P., Zhang, X., & Gao, D. (2024). Preference heterogeneity analysis on train choice behaviour of high-speed railway passengers: A case study in China. *Transportation Research Part A: Policy and Practice 188*, 104198. doi:https://doi.org/10.1016/j.tra.2024.104198

Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society 14*, 1-10.

Da Silva, D., Klumpenhouwer, W., Karner, A., Robinson, M., Liu, R., & Shalaby, A. (2022). Living on a fare: Modeling and quantifying the effects of fare budgets on transit access and equity. *Journal of Transport Geography 101*, 103348. doi:https://doi.org/10.1016/j.jtrangeo.2022.103348

Daganzo, C. F., & Ouyang, Y. (2019). *Public transportation systems: Principles of system design, operations planning and real-time control*: World Scientific.

Dahmen, V., Weikl, S., & Bogenberger, K. (2023). Interpretable Machine Learning for Mode Choice Modeling on Tracking-Based Revealed Preference Data. *Transportation Research Record* 03611981241246973.

Darwish, A. M., Almansour, M., Salah, A., Zagow, M., Saeed, K., & Elkafoury, A. (2024). Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: a case study of Alexandria City, Egypt. *Transportation Research Interdisciplinary Perspectives 24*, 101052.

Díaz-Ramírez, J., Estrada-García, J. A., & Figueroa-Sayago, J. (2023). Predicting transport mode choice preferences in a university district with decision tree-based models. *City and Environment Interactions 20*, 100118. doi:https://doi.org/10.1016/j.cacint.2023.100118

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.

Drabicki, A., Cats, O., Kucharski, R., Fonzone, A., & Szarata, A. (2023). Should I stay or should I board? Willingness to wait with real-time crowding information in urban public transport. *Research in Transportation Business & Management, 47*, 100963. doi:https://doi.org/10.1016/j.rtbm.2023.100963

Economic Base. (2017). Knock on the Bangkok-Chiang Mai high-speed train fare 1,089 baht. Retrieved from https://www.thansettakij.com/business/242231

Elharoun, M., El-Badawy, S. M., & Shahdah, U. E. (2023). Artificial intelligence techniques for predicting individuals' mode choice behavior in Mansoura city, Egypt. *Transportation Research Record 2677*(10), 605-623.

Esfeh, M. A., Kattan, L., Lam, W. H., Esfe, R. A., & Salari, M. (2020). Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for vulnerability analysis of road network. *Transportation Research Part C: Emerging Technologies 120*, 102808.

Ghosh, T., & Nagaraj, N. (2024). Evaluating the Determinants of Mode Choice Using Statistical and Machine Learning Techniques in the Indian Megacity of Bengaluru. *arXiv preprint arXiv:2401.13977*.

Givoni, M. (2006). Development and Impact of the Modern High‐speed Train: A Review. *Transport Reviews 26*(5), 593-611. doi:10.1080/01441640600589319

He, Y., Zhao, Y., & Tsui, K. L. (2018). *An analysis of factors influencing metro station ridership: Insights from Taipei metro.* Paper presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC).

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis*: Cambridge university press.

Hong, H., Naghibi, S. A., Pourghasemi, H. R., & Pradhan, B. (2016). GIS-based landslide spatial modeling in Ganzhou City, China. *Arabian Journal of Geosciences 9*, 1-26.

Hu, C., Huang, J., Gao, Y., & Lin, R. (2023). Can high-speed railway promote regional market integration? Evidence from China. *Research in Transportation Business & Management, 51*, 101057. doi:https://doi.org/10.1016/j.rtbm.2023.101057

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., . . . Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology 574*, 1029-1041.

Huang, Y., Song, H., & Jia, C. Y. (2014). Analysis on the main factors influencing the high-speed railway passenger flow. *Advanced Materials Research 919*, 1085-1090.

Ibraeva, A., & Sousa, J. F. d. (2014). Marketing of Public Transport and Public Transport Information Provision. *Procedia - Social and Behavioral Sciences, 162*, 121-128. doi:https://doi.org/10.1016/j.sbspro.2014.12.192

มหาวิทยาลัยเทคโนโลยีสุรนารี

Jara-Díaz, S., Gschwender, A., & Hörcher, D. (2023). Public transport: design, scale, and pricing. In *Handbook on Transport Pricing and Financing* (pp. 171-189): Edward Elgar Publishing.

Jiang, C., & Wang, C. (2021). High-speed rail pricing: Implications for social welfare. *Transportation Research Part E: Logistics and Transportation Review, 155*, 102484.

Karmarkar, O. D., Jana, A., & Velaga, N. R. (2023). Willingness to pay for reduced travel time: Case study of Mumbai - Ahmedabad High Speed Railway. *Case Studies on Transport Policy 12*, 100974. doi:https://doi.org/10.1016/j.cstp.2023.100974

Kashifi, M. T., Jamal, A., Kashefi, M. S., Almoshaogeh, M., & Rahman, S. M. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behaviour and Society 29*, 279-296.

Kazemi, F., Asgarkhani, N., Shafighfard, T., Jankowski, R., & Yoo, D.-Y. (2024). Machine-Learning Methods for Estimating Performance of Structural Concrete Members Reinforced with Fiber-Reinforced Polymers. *Archives of Computational Methods in Engineering* 1-33.

Kim, E.-J. (2021). Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation 2021*(1), 6685004.

Kujala, R., Weckström, C., Mladenović, M. N., & Saramäki, J. (2018). Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. *Computers, Environment and Urban Systems, 67*, 41-54.

Lachapelle, U., & Boisjoly, G. (2023). Breaking down public transit travel time for more accurate transport equity policies: A trip component approach. *Transportation Research Part A: Policy and Practice 175*, 103756. doi:https://doi.org/10.1016/j.tra.2023.103756

Lackner, M., & Zulehner, C. (2020). Rent sharing and gender discrimination: causal evidence from collegiate athletics. *Social Science Quarterly, 101*(2), 678-696.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications 10*(1), 1096.

Larranaga, A. M., Arellana, J., & Senna, L. A. (2017). Encouraging intermodality: A stated preference analysis of freight mode choice in Rio Grande do Sul. *Transportation Research Part A: Policy and Practice 102*, 202-211.

Lee, E., Kawakita, T., Huai, Y., Lo, H. K., & Zhang, A. (2024). Airline and high-speed rail collaboration and competition under travel time variability. *Transportation Research Part A: Policy and Practice 185*, 104104. doi:https://doi.org/10.1016/j.tra.2024.104104

Li, X., Shi, L., Shi, Y., Tang, J., Zhao, P., Wang, Y., & Chen, J. (2024). Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Applied Geography 166*, 103264.

Li, X., Xing, G., Qian, X., Guo, Y., Wang, W., & Cheng, C. (2022). Subway station accessibility and its impacts on the spatial and temporal variations of its outbound ridership. *Journal of Transportation Engineering, Part A: Systems, 148*(12), 04022106.

Liu, L., & Zhang, M. (2018). High-speed rail impacts on travel times, accessibility, and economic productivity: A benchmarking analysis in city-cluster regions of China. *Journal of Transport Geography 73*, 25-40.

Liu, S., & Putro, U. S. (2024). Passenger Service Satisfaction Evaluation of Jakarta-Bandung High-Speed Railway. *European Journal of Business and Management Research, 9*(4), 115-126.

Liu, Y., Wang, S., & Xie, B. (2019). Evaluating the effects of public transport fare policy change together with built and non-built environment features on ridership: The case in South East Queensland, Australia. *Transport Policy 76*, 78-89. doi:https://doi.org/10.1016/j.tranpol.2019.02.004

Lu, Y., Prato, C. G., Sipe, N., & Corcoran, J. (2023). Understanding the role of household modality style on first and last mile travel mode choice and public transit station choice. *Travel Behaviour and Society 32*, 100580. doi:https://doi.org/10.1016/j.tbs.2023.100580

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., . . . Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence 2*(1), 56-67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems 30*.

Ma, W., Wang, Q., Yang, H., Zhang, A., & Zhang, Y. (2019). Effects of Beijing-Shanghai high-speed rail on air travel: Passenger types, airline groups and tacit collusion. *Research in Transportation Economics, 74*, 64-76.

Mangalathu, S., Hwang, S.-H., & Jeon, J.-S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures 219*, 110927.

Mao, H., Deng, X., Jiang, H., Shi, L., Li, H., Tuo, L., . . . Guo, F. (2021). Driving safety assessment for ride-hailing drivers. *Accident Analysis & Prevention 149*, 105574.

Meesit, R. (2024). Factors affecting the future utilisation of high-speed train services in Eastern Economic Corridor (EEC). *Transportation Research Interdisciplinary Perspectives 26*, 101162. doi:https://doi.org/10.1016/j.trip.2024.101162

NESDC. (2022). Gross regional and provincial product chain volume measure 2020 edition. In: Office of the National Economic and Social Development Council (NESDC) Bangkok.

Nurhidayat, A. Y., Widyastuti, H., Sutikno, & Upahita, D. P. (2023). Research on passengers' preferences and impact of high-speed rail on air transport demand. *Sustainability 15*(4), 3060.

Office of Transport and Traffic Policy and Planning. (2024). Proportion of Public Transportation Usage for Intercity Travel (2017-2022). Retrieved from https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45. Retrieved 11 August 2024 https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45

Orlando, V. M., Baquela, E. G., Bhouri, N., & Lotito, P. A. (2023). Public transport demand estimation by frequency adjustments. *Transportation Research*

Interdisciplinary Perspectives 19, 100832. doi:https://doi.org/10.1016/j.trip.2023.100832

Osawa, M., Akamatsu, T., & Takayama, Y. (2017). Harris And Wilson (1978) Model Revisited: The Spatial Period-Doubling Cascade In An Urban Retail Model. *Journal of Regional Science, 57*(3), 442-466.

Pan, J. Y. (2024). High-Speed Rail in the US—Mode Choice Decision and Impact of COVID-19. *Sustainability 16*(10), 4041.

Pavlou, M., Ambler, G., Qu, C., Seaman, S. R., White, I. R., & Omar, R. Z. (2024). An evaluation of sample size requirements for developing risk prediction models with binary outcomes. *BMC Medical Research Methodology 24*(1), 146.

Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., . . . Takeuchi, W. (2020). Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam. *Remote Sensing 12*(8), 1334.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems 31*.

Puan, O., Hassan, Y., Mashros, N., Idham, M., Hassan, N., Warid, M., & Hainin, M. (2019). *Transportation mode choice binary logit model: A case study for Johor Bahru city.* Paper presented at the IOP conference series: Materials science and engineering.

Raturi, V., & Verma, A. (2020). A game-theoretic approach to analyse inter-modal competition between high-speed rail and airlines in the Indian context. *Transportation Planning and Technology 43*(1), 20-47.

Ristić, B., Stojadinović, N., & Trifunović, D. (2022). Conditions for effective on-track competition in the European passenger railway market: A yardstick for regulations. *Transport Policy 119*, 1-15. doi:https://doi.org/10.1016/j.tranpol.2022.02.006

Salas, P., De la Fuente, R., Astroza, S., & Carrasco, J. A. (2022). A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Systems with Applications 193*, 116253.

Shafighfard, T., Kazemi, F., Asgarkhani, N., & Yoo, D.-Y. (2024). Machine-learning methods for estimating compressive strength of high-performance alkali-activated concrete. *Engineering Applications of Artificial Intelligence 136*, 109053. doi:https://doi.org/10.1016/j.engappai.2024.109053

Shafighfard, T., Kazemi, F., Bagherzadeh, F., Mieloszyk, M., & Yoo, D. Y. (2024). Chained machine learning model for predicting load capacity and ductility of steel fiber–reinforced concrete beams. *Computer‐Aided Civil and Infrastructure Engineering*.

Shelat, S., Cats, O., & van Cranenburgh, S. (2022). Traveller behaviour in public transport in the early stages of the COVID-19 pandemic in the Netherlands. *Transportation Research Part A: Policy and Practice 159*, 357-371.

Shen, J., Sakata, Y., & Hashimoto, Y. (2006). A comparison between latent class model and mixed logit model for transport mode choice: evidences from two datasets of Japan. *Discussion papers in economics and business*, 06-05.

Shi, R., Hu, Z., Zhou, Y., & Liu, P. (2014). Research on Railway Freight Market Share Based on the Whole Process of Transport. *Procedia - Social and Behavioral Sciences, 138*, 298-304. doi:https://doi.org/10.1016/j.sbspro.2014.07.207

Soczówka, P., Kłos, M. J., Żochowska, R., & Sobota, A. (2021). An analysis of the influence of travel time on access time in public transport. *Zeszyty Naukowe. Transport/Politechnika Śląska*(111), 137-149.

Sresakoolchai, J., & Kaewunruen, S. (2020). Comparative studies into public private partnership and traditional investment approaches on the high-speed rail project linking 3 airports in Thailand. *Transportation Research Interdisciplinary Perspectives 5*, 100116.

Srithongrung, A., & Kriz, K. A. (2019). Thai Public Capital Budget and Management Process. In *Capital Management and Budgeting in the Public Sector* (pp. 206-235): IGI Global.

State Railway of Thailand (2024). The State Railway of Thailand Easy book, Convenient check. Retrieved from https://dticket.railway.co.th/DTicketPublicWeb/home/Home

Tang, L., Tang, C., Fu, Q., & Ma, C. (2024). Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis. *IET Intelligent Transport Systems*

Tanwar, R., & Kumar Agarwal, P. (2024). Investigating the factors determining service quality in Bhopal's multimodal public transport system: A factor analysis. *Case Studies on Transport Policy* 101317. doi:https://doi.org/10.1016/j.cstp.2024.101317

Tirachini, A., Godachevich, J., Cats, O., Muñoz, J. C., & Soza-Parra, J. (2022). Headway variability in public transport: A review of metrics, determinants, effects for quality of service and control strategies. *Transport Reviews 42*(3), 337-361.

Tomeš, Z., & Jandová, M. (2018). Open access passenger rail services in Central Europe. *Research in Transportation Economics, 72*, 74-81. doi:https://doi.org/10.1016/j.retrec.2018.10.002

Vega, G. M., & Aznarte José, L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics 56*, 101039.

Vigren, A. (2017). Competition in Swedish passenger railway: Entry in an open access market and its effect on prices. *Economics of Transportation, 11-12*, 49-59. doi:https://doi.org/10.1016/j.ecotra.2017.10.005

Wang, D., & Zhou, M. (2017). The built environment and travel behavior in urban China: A literature review. *Transportation Research Part D: Transport and Environment 52*, 574-585.

Wang, F., & Ross, C. L. (2018). Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record 2672*(47), 35-45.

Wang, H., & Huang, H.-J. (2024). Effects of high-speed rail on intercity travels, utility and social welfare in urban agglomerations: A game-theoretic perspective. *Transportation Research Part E: Logistics and Transportation Review, 192*, 103800. doi:https://doi.org/10.1016/j.tre.2024.103800

มหาวิทยาลัยเทคโนโลยีสุรนารี

Wang, Y., Sun, X., Mark-up, I. S., & D, I. (2018). A Reexamination of the Relationship between Market Power and Technological Innovation. *RD Res. Manag, 39*, 144-152.

Wei, W., Wang, F., Li, T., & Li, F. (2024). Comprehensive impacts of high-speed rail and air transport on tourism development in China. *Transportation Research Part A: Policy and Practice 190*, 104263. doi:https://doi.org/10.1016/j.tra.2024.104263

Wójcik, S. (2019). The determinants of travel mode choice: the case of ŁódŹ, Poland. *Bulletin of Geography. Socio-economic Series* (44), 93-101.

Wu, D., & Martín, J. C. (2022). Research on Passengers' Preference for High-Speed Railways (HSRs) and High-Speed Trains (HSTs). *Sustainability 14*(3). doi:10.3390/su14031473

Wu, M.-W., & Shen, C.-H. (2011). The elusive effect of bank size on profits. *The Service Industries Journal, 31*(11), 1703-1724.

Wu, S., & Han, D. (2022). Accessibility of high-speed rail (HSR) stations and HSR–air competition: Evidence from China. *Transportation Research Part A: Policy and Practice 166*, 262-284. doi:https://doi.org/10.1016/j.tra.2022.10.015

Xia, W., Jiang, C., Wang, K., & Zhang, A. (2019). Air-rail revenue sharing in a multi-airport system: Effects on traffic and social welfare. *Transportation research part B: methodological, 121*, 304-319.

Xu, F., Liu, Q., Zheng, X., Cao, L., & Yang, M. (2022). Research on the impact of China's high-speed rail opening on enterprise market power: Based on the perspective of market segmentation. *Transport Policy 128*, 121-137. doi:https://doi.org/10.1016/j.tranpol.2022.09.019

Yang, W., Chen, Q., & Yang, J. (2022). Factors Affecting Travel Mode Choice between High-Speed Railway and Road Passenger Transport—Evidence from China. *Sustainability 14*(23), 15745. doi: https://doi.org/10.3390/su142315745

Yu, J., Chang, X., Hu, S., Yin, H., & Wu, J. (2024). Combining travel behavior in metro passenger flow prediction: A smart explainable Stacking-Catboost algorithm. *Information Processing & Management 61*(4), 103733. doi:https://doi.org/10.1016/j.ipm.2024.103733

มหาวิทยาลัยเทคโนโลยีสุรนารี

Yu, X., Lang, M., Gao, Y., Wang, K., Su, C.-H., Tsai, S.-B., . . . Li, S. (2018). An empirical study on the design of China high-speed rail express train operation plan—From a sustainable transport perspective. *Sustainability 10*(7), 2478.

Yuan, Y., Yang, M., Feng, T., Rasouli, S., Ruan, X., Wang, X., & Li, Y. (2021). Analyzing heterogeneity in passenger satisfaction, loyalty, and complaints with air-rail integrated services. *Transportation Research Part D: Transport and Environment 97*, 102950.

Yuniaristanto, Sutopo, W., Hisjam, M., & Wicaksono, H. (2024). Estimating the market share of electric motorcycles: A system dynamics approach with the policy mix and sustainable life cycle costs. *Energy Policy, 195*, 114345. doi:https://doi.org/10.1016/j.enpol.2024.114345

Zang, Z., Xu, X., Qu, K., Chen, R., & Chen, A. (2022). Travel time reliability in transportation networks: A review of methodological developments. *Transportation Research Part C: Emerging Technologies 143*, 103866. doi:https://doi.org/10.1016/j.trc.2022.103866

Zhai, W., Li, C., Fei, S., Liu, Y., Ding, F., Cheng, Q., & Chen, Z. (2023). CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values. *Computers and Electronics in Agriculture 214*, 108306. doi:https://doi.org/10.1016/j.compag.2023.108306

Zhang, Q., Yang, H., Wang, Q., Zhang, A., & Zhang, Y. (2020). Impact of high-speed rail on market concentration and Lerner index in China's airline market. *Journal of Air Transport Management 83*, 101755. doi:https://doi.org/10.1016/j.jairtraman.2019.101755

Zhang, Y., Liang, K., Yao, E., & Gu, M. (2024). Measuring Reliable Accessibility to High-Speed Railway Stations by Integrating the Utility-Based Model and Multimodal Space–Time Prism under Travel Time Uncertainty. *ISPRS International Journal of Geo-Information 13*(8), 263.

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society 20*, 22-35.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Zhou, Y., Zhao, M., Tang, S., Lam, W. H., Chen, A., Sze, N., & Chen, Y. (2020). Assessing the relationship between access travel time estimation and the accessibility to high speed railway station by different travel modes. *Sustainability 12*(18), 7827.

Zhu, L., Shu, S., & Zou, L. (2022). XGBoost‐Based Travel Time Prediction between Bus Stations and Analysis of Influencing Factors. *Wireless Communications and Mobile Computing 2022*(1), 3504704.

Zhu, X., Shen, X., Chen, K., & Zhang, Z. (2024). Research on the prediction and influencing factors of heavy duty truck fuel consumption based on LightGBM. *Energy 296*, 131221. doi:https://doi.org/10.1016/j.energy.2024.131221

Zouzou, Y., & Citakoglu, H. (2023). General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. *Acta Geophysica 71*(2), 927-947.

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER IV

MACHINE LEARNING-BASED ANALYSIS OF TRAVEL MODE PREFERENCES: NEURAL AND BOOSTING MODEL COMPARISON USING STATED PREFERENCE DATA FROM THAILAND'S EMERGING HIGH-SPEED RAIL NETWORK

4.1 Abstract

This study examines travel mode choice behavior within the context of Thailand's emerging high-speed rail (HSR) development. It conducts a comparative assessment of predictive capabilities between the conventional Multinomial Logit (MNL) framework and advanced data-driven methodologies, including gradient boosting algorithms (Extreme Gradient Boosting, Light Gradient Boosting Machine, Categorical Boosting) and neural network architectures (Deep Neural Network, Convolutional Neural Network). The analysis leverages stated preference (SP) data and employs Bayesian optimization in conjunction with a stratified 10-fold cross-validation scheme to ensure model robustness. CatBoost emerges as the top-performing model (area under the curve = 0.9113; accuracy = 0.7557), highlighting travel cost, service frequency, and waiting time as the most influential determinants. These findings underscore the effectiveness of machine learning approaches in capturing complex behavioral patterns, providing empirical evidence to guide high-speed rail policy development in low- and middle-income countries. Practical implications include optimizing fare structures, enhancing service quality, and improving station accessibility to support sustainable adoption.

4.2 Introduction

In recent decades, Thailand has experienced rapid urbanization, resulting in a substantial increase in travel demand. This surge has led to congestion in public transportation systems and a growing reliance on private vehicles. High-speed rail (HSR) has emerged as a key solution to address these issues, offering the potential to significantly reduce travel times, enhance intercity connectivity, and promote economic development across various regions (Corinne Blanquart & Martin Koning, 2017; Liang et al., 2020; W. Wu, Liang, & Wu, 2016). In Thailand, collaboration with China on the development of HSR is expected to substantially reshape the nation's transportation landscape by providing an alternative to private cars, intercity buses, and air travel (K. Tissayakorn, 2021).

The development of HSR also aligns with Thailand's commitments under the Sustainable Development Goals (SDGs), particularly SDG 9, which focuses on building resilient infrastructure; SDG 11, aimed at making cities inclusive, safe, resilient, and sustainable; and SDG 13, which calls for urgent action to combat climate change and its impacts. HSR development is not only anticipated to reduce reliance on fossil fuels but also to foster sustainable urban connectivity and mitigate carbon emissions from long-distance transportation (Office of the National Economic and Social Development Council, 2025). Furthermore, the Thai government has incorporated HSR into its National Action Plan for advancing the SDGs, positioning the rail network as a crucial strategy for reducing infrastructure disparities and promoting inclusive growth nationwide.

From an infrastructure perspective, Thailand is currently undertaking the construction of four major HSR lines—the northern, eastern, northeastern, and southern corridors—designed to enhance economic development, trade, tourism, and regional connectivity. A particularly significant project is the Thai–Lao–Chinese HSR, a collaboration between the Thai and Chinese governments. This line will connect Bangkok's Bang Sue Grand Station with Laos and onward to China, spanning approximately 377 miles across eight provinces through eleven stations. The project is divided into two phases: the Bangkok–Nakhon Ratchasima section (155 miles), currently under construction, and the Nakhon Ratchasima–Nong Khai section (222 miles), which will link to China's HSR network via Laos. Full completion is expected by 2030 (JTTRI-AIRO, 2023; State Railway of Thailand, 2020). This project is anticipated to significantly influence intercity travel behavior and attract users away from private vehicles and low-cost airlines.

Nevertheless, predicting travel mode choice in response to HSR introduction remains challenging due to the complex interplay of factors such as travel time, cost, station accessibility, waiting time, and service frequency (Yajuan Deng et al., 2023; L. Liu & Zhang, 2018; Jianqiang Wang et al., 2023; Yang et al., 2022; Yongsheng Zhang et al., 2024; Zhou et al., 2020). Traditionally, travel mode choice modeling has been grounded in the Random Utility Maximization (RUM) framework, with the Multinomial Logit (MNL) model serving as the dominant tool due to its theoretical rigor, ease of interpretation, and suitability for analyzing both individual-specific and alternative-specific variables (Akter & Alam, 2024; Moshe E Ben-Akiva & Steven R Lerman, 1985; Ning et al., 2021; Török, Szalay, Uti, & Verebélyi, 2020). However, MNL models suffer from several limitations, including the Independence of Irrelevant Alternatives (IIA) assumption, limited capacity to capture nonlinear relationships, and reduced predictive performance when dealing with high-dimensional or correlated datasets.

To overcome these limitations, extensions such as the Nested Logit (NL), Cross-Nested Logit (CNL), and Mixed Logit (ML) models have been developed, offering improved behavioral flexibility (Akter & Alam, 2024; Moshe E Ben-Akiva & Steven R Lerman, 1985; Ning et al., 2021; Török et al., 2020). Nevertheless, these models still require predefined functional forms and often involve considerable computational complexity. Recent advancements in machine learning (ML) and deep learning (DL) offer promising alternatives, capable of capturing complex, nonlinear relationships without assuming specific functional forms. Ensemble tree methods, such as Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting(CatBoost), have proven to deliver superior predictive performance in travel mode choice applications (Ammar Abulibdeh, 2023; T. Chen & Guestrin, 2016; Jenny Díaz-Ramírez et al., 2023; Ke et al., 2017; Prokhorenkova et al., 2018), while Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) have shown remarkable capabilities in processing structured and high-dimensional data (Banyong et al., 2024; L. Guo et al., 2022; Hillel, Bierlaire, Elshafie, & Jin, 2021; Wen & Chen, 2025).

Although machine learning (ML) and deep learning (DL) models are widely acknowledged for their superior predictive capabilities, concerns have been persistently raised regarding their limited interpretability and opacity, commonly referred to as the “black-box” phenomenon (Grinsztajn, Oyallon, & Varoquaux, 2022; Hillel et al., 2021). To address this challenge and enhance model explain ability, the present study employs Shapley Additive Explanations (SHAP), a solution grounded in cooperative game theory, to elucidate the internal logic of the models and identify the most influential predictors associated with travel mode decisions (Victoria Dahmen, Simone Weikl, & Klaus Bogenberger, 2024). While the use of ML techniques in travel behavior analysis has expanded, rigorous comparative studies between ML/DL frameworks and conventional econometric models, particularly those utilizing stated preference (SP) data in the context of high-speed rail (HSR) development in Thailand, remain limited. This research endeavors to bridge this empirical gap by conducting a systematic evaluation of the predictive performance of MNL, XGBoost, LightGBM, CatBoost, DNN, and CNN models using large-scale SP data collected from 3200 respondents across 16 provinces. A standardized data preprocessing framework, Bayesian hyperparameter optimization, and 10-fold cross-validation are employed to ensure rigorous model evaluation. Furthermore, SHAP analysis is utilized to enhance transparency and provide insights into the key factors driving travel mode choices. This research offers one of the first comprehensive evaluations integrating traditional econometric models and advanced ML/DL techniques in the emerging context of HSR in a developing country, contributing valuable insights for future transportation planning and sustainable mobility policymaking.

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.3 Methodology and Data Analysis

The research methodology, as depicted in Figure 4.1, follows a systematic and structured approach, commencing with stated preference surveys and data validation to ensure data accuracy and reliability. Subsequently, the dataset is partitioned into training (X_train, y_train) and testing (X_test, y_test) sets to facilitate model development. The study employs multiple analytical frameworks, including deep learning comprising Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) as well as Multinomial Logit (MNL) and Gradient Boosting methods, such as

XGBoost, LightGBM, and CatBoost. To enhance model performance, hyperparameter tuning is conducted, followed by a comparative evaluation utilizing cross-validation techniques and performance metrics. In the final stage, a judgment assessment is performed to determine the acceptability of the model results. Upon validation, the most optimized model undergoes further interpretability analysis using Shapley Additive Explanations (SHAP) to identify the key determinants influencing travel mode choice. The dataset was handled and examined utilizing Python 3.11 through Anaconda, with the final preparation phase taking approximately two hours. A 13th-generation Intel Core i9-13900H processor running at 2.60 GHz, combined with 32 GB of RAM, powered the system used for computational analysis, supporting high efficiency and processing speed.

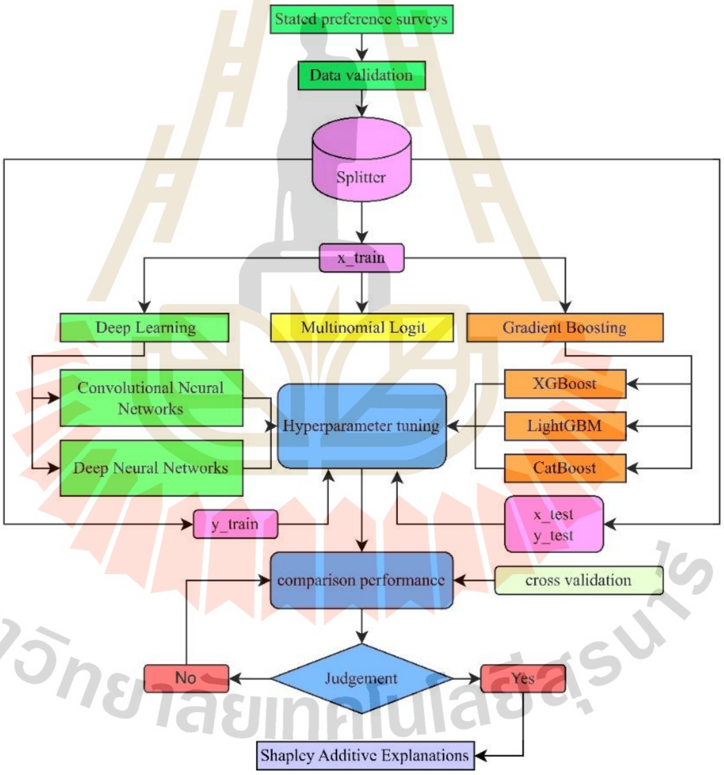


Figure 4.1 Research process flowchart.

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.3.1 Survey Design

This study examines future travel mode preferences, encompassing buses, traditional rail, air transport, and high-speed rail (HSR), through advanced machine learning techniques. Data were obtained from a nationwide survey encompassing four major geographic zones of Thailand: The north, northeast, central, and south. The analysis aims to uncover the key determinants influencing travel behavior and choice decisions (Banyong et al., 2024).

Target provinces were selected based on their strategic potential for future HSR development, with selection criteria including economic vitality and infrastructural significance as of 2022 (NESDC, 2022). Incorporating provinces with varying socioeconomic and infrastructural profiles across the regions supports the national representativeness of the findings (Srithongrung & Kriz, 2019).

The survey involved in-person interviews with adults aged 18 and above, covering 16 provinces and yielding 3200 valid responses, equally distributed across provinces (n = 200 each). A non-probability sampling strategy was employed to ensure regional coverage, mitigate selection bias, and enhance result robustness. The sample size adheres to machine learning best practices, falling within the suggested range of 50 to 1000 observations per target class, four in this case (Pavlou et al., 2024).

The questionnaire comprised two components: respondents' socioeconomic information and a stated preference (SP) experiment using a choice-based design. Participants indicated their preferred travel mode under hypothetical future HSR scenarios, considering factors such as station accessibility, waiting duration, total journey time, cost, and service frequency, providing realistic behavioral insights (Arencibia et al., 2015; Kujala et al., 2018).

4.3.2 Questionnaire Design

The set of explanatory variables incorporated into the choice-based survey design reflects critical service attributes known to influence individual travel mode decisions. These include the duration required to reach the departure station, the delay experienced prior to boarding, and the in-vehicle journey duration between the origin and destination. Additionally, the survey accounts for the direct financial expenditure borne by the traveler in the form of out-of-pocket fare, as well as the interval between consecutive scheduled services. These attributes were carefully selected to capture the key trade-offs faced by travelers in real-world decision-making scenarios. Collectively, they provide a robust foundation for analyzing modal preferences and quantifying the influence of service characteristics on travel behavior. A summary of these variables and their operational definitions is presented in Table

4.1. Each feature was defined across multiple value tiers to mirror plausible travel conditions, thereby enhancing the reliability of respondents' stated selections. This stated preference task concentrated along the Bangkok–Chiang Mai axis, spanning approximately 435 miles, was recognized as a strategic route due to the competitive dynamic between high-speed rail (HSR) and traditional transport alternatives.

As HSR had not yet been implemented during the survey period, assumptions related to travel time, fare levels, service intervals, and waiting durations were based on government forecasts (Economic Base, 2017). Additional scenario parameters were informed by previous empirical research to incorporate insights into traveler preferences, operational performance, and market dynamics. Given the reliance on forecasted data, uncertainties may arise due to unexpected disruptions, infrastructure constraints, fare policy changes, or broader system dynamics. These uncertainties may affect projected demand and policy decisions. To improve scenario realism, adjustments were applied to align assumptions with actual system parameters. Reference values for waiting times were obtained from national bus and rail booking platforms (BusOnlineTicket.co.th, 2024) and airline reservation systems (AirAsia move, 2024). For competing modes, information on travel time, pricing, and scheduling was derived from official fare tables and transit timetables (State Railway of Thailand 2024), allowing direct comparison with HSR. Adjustments were also made to access durations to account for specific constraints—such as extended airport check-in and security procedures. The finalized dataset was validated using empirical travel history and survey-based feedback to strengthen its reliability. This study assessed five core attributes across three existing travel alternatives, along with two hypothetical HSR options using a standardized factorial design.

The experimental design produced 96 total choice sets, derived from three service types, 15 attribute combinations, and 25 replications. To reduce respondent burden, a fractional factorial design was employed, dividing the sets into eight blocks containing twelve alternatives each, with each participant assigned to a single block. This structure balanced statistical precision with manageable cognitive demand (Hensher et al., 2015).

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.1 Attribute Levels for Travel Modes

Attribute	Bus	Train	Airplane	HSR (Levels 1)	HSR (Levels 2)
Access time (Station approach duration: minute)	10	10	30	10	15
Waiting time (Pre-departure delay: minute)	15	10	120	15	10
Travel (Time In-vehicle journey duration: minute)	720	720	135	190	220
Travel cost (Out-of-pocket fare: bath)	750	300	3000	1050	1400
Frequency times (Scheduled service interval: minute)	30	150	120	190	220

4.3.3 Data and Variables

To prepare the stated preference (SP) survey data for machine learning applications, a preprocessing phase was carried out involving data cleansing, the handling of missing or inconsistent entries, and the normalization of continuous variables. Categorical variables were transformed using one-hot or ordinal encoding techniques. Each of the 3200 respondents participated in twelve hypothetical choice tasks, with each task presenting four alternatives: bus, train, airplane, and high-speed rail (HSR), yielding a total of 38,400 data points. The dependent variable (Y) denoted the travel mode selected by the respondent (coded as 1 = bus, 2 = train, 3 = airplane, 4 = HSR), while the predictor variables (X) comprised both quantitative factors (e.g., fare, duration) and sociodemographic indicators (e.g., gender, income level, vehicle ownership).

The data structure was organized such that each respondent's choice set was represented by four separate records, one for each mode, where the selected alternative was labeled as Y = 1 and non-chosen alternatives as Y = 0. For instance, if train was selected, then that record was marked Y = 1 while the others were assigned Y = 0. This encoding process produced a final dataset of 38,400 rows encompassing both trip characteristics and individual-level features. To facilitate model development and validation, the dataset was randomly partitioned into training (80%) and testing (20%) subsets.

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.3.3.1 Dataset Structuring and Preprocessing

Following the collection of stated preference (SP) survey responses, the dataset underwent a comprehensive preprocessing workflow tailored to support predictive modeling using machine learning techniques. This phase involved cleaning raw entries, restructuring categorical data, and formatting the choice set architecture. Any incomplete or inconsistent records were removed, continuous variables were normalized, and categorical features including gender, household income, and private vehicle ownership were transformed using suitable encoding strategies, such as binary (one-hot) and ordinal schemes.

Each of the 3200 participants evaluated 12 hypothetical travel scenarios, yielding a dataset comprising 38,400 observations. In each scenario, respondents were asked to choose among four transportation modes: bus, conventional rail, air travel, and high-speed rail (HSR). These options were differentiated based on key service characteristics, namely access time, wait duration, total travel time, and fare cost. The response variable (Y) indicated the selected alternative, coded as 1 for the chosen mode and 0 for the others. Each option was assigned a numeric identifier from 1 to 4. The predictor variables (X) encompassed both operational attributes and sociodemographic details.

The dataset was organized such that each row represented one travel alternative within a respondent’s choice set. After preprocessing, the dataset was divided into training (80%) and testing (20%) subsets to facilitate model validation and performance evaluation.

4.3.4 Multinomial Logit Model for Mode Choice Estimation

The modeling framework is grounded in the assumption that each transport alternative embedded in the experimental design delivers distinct utility to decision-makers, thereby influencing their travel behavior. The probability that an individual selects a given option increases as its associated utility becomes relatively more favorable compared to competing alternatives. The Multinomial Logit (MNL) model expresses this probability as follows:

$$P_{(i)} = \frac{e^{U_i}}{\sum_{j \in J} e^{U_i}} \tag{4-1}$$

where:

$P_{(i)}$ represents the probability that alternative i is selected;

U_i and U_j denote the systematic utilities corresponding to alternatives i and j, respectively;

J is the total number of competing alternatives considered in the model.

4.3.5 Deep Neural Network (DNN)

Deep Neural Networks (DNNs) are particularly effective in handling problems characterized by high nonlinearity, especially in cases involving unstructured data such as images, videos, or continuous data streams. In such complex scenarios, incorporating multiple hidden layers into a traditional Multilayer Perceptron (MLP) significantly enhances its ability to capture intricate patterns and improve predictive performance. These architectures, commonly referred to as Deep Multilayer Perceptron (DMLPs) or Deep Neural Networks (DNNs), form the foundation of modern deep learning methodologies (García-García, García-Ródenas, López-Gómez, & Martín-Baos, 2022; García-Ródenas, Linares, & López-Gómez, 2017).

4.3.6 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) represent one of the most widely used and effective deep learning architectures, particularly in tasks requiring the extraction of hierarchical features to classify input data. These networks are particularly well suited for applications involving image and video processing, where features need to be captured at multiple levels of abstraction. The core operation of a CNN is convolution, which is performed using filters (also referred to as kernels). In the case of image data, these filters are typically two-dimensional and slide across the entire input sample to extract spatial features. However, when dealing with structured matrix data, such as in this study, a one-dimensional convolution can be applied instead. The depth of the network determines the number of filters, where a deeper architecture enables the extraction of more complex and interrelated patterns across different

มหาวิทยาลัยเทคโนโลยีสุรนารี

abstraction levels. In this research, a CNN model has been designed with two one-dimensional convolutional layers, each containing 64 kernels of 2 units. To further optimize model performance, various hyperparameters related to the training process, such as learning rate, batch size, and regularization techniques, have been fine-tuned. This ensures that the CNN effectively captures relevant features from the dataset while maintaining high generalization capability (García-García et al., 2022; García-Ródenas et al., 2017).

4.3.7 Extreme Gradient Boosting (XGBoost)

Widely adopted across a range of domains, Extreme Gradient Boosting (XGBoost) is a sophisticated machine learning method designed to enhance the performance and efficiency of decision tree ensembles (T. Chen & Guestrin, 2016). XGBoost, which builds upon the foundational architecture of Gradient Boosting Decision Trees (GBDTs), introduces two key innovations: the integration of the second-order Taylor expansion to optimize the loss function, surpassing GBDT's reliance on first-order gradients and the implementation of regularization mechanisms to mitigate overfitting and promote robust model generalization (Krizhevsky et al., 2012; Y. Xu et al., 2019). These enhancements make XGBoost highly efficient, scalable, and accurate, supporting parallel processing, feature importance analysis, and customizable loss functions, making it a flexible and widely applicable technique (Friedman, 2001). At its core, XGBoost follows an iterative additive learning process, where low-depth decision trees are constructed sequentially to minimize a predefined loss function. Unlike conventional decision trees, XGBoost assigns greater weights to misclassified instances in each iteration, gradually refining predictions while balancing bias and variance (Friedman, 2001). The ultimate prediction is obtained by consolidating the contributions of individual weak learners, resulting in a resilient ensemble architecture. In this study, the XGB model was implemented using the "XGBoost" package in Python 3.11, with core hyperparameters such as the step size for weight updates, maximum depth of individual trees, and the total number of boosting iterations alongside regularization components, meticulously tuned to optimize predictive performance. In machine learning, hyperparameters are model settings that must be specified before training, as opposed to parameters learned from data. Their proper tuning is essential for enhancing model generalization and preventing overfitting. To achieve this, cross-validation techniques were applied, allowing for the systematic evaluation of different hyperparameter configurations to determine the optimal combination for improved predictive accuracy.

4.3.8 Light Gradient Boosting (LightGBM)

LightGBM, proposed by Ke et al. (2017), is a high-efficiency gradient boosting algorithm built to outperform traditional GBDT models in both speed and predictive strength. It incorporates several algorithmic innovations such as one-sided sampling based on gradient magnitudes (GOSS), a leaf-wise growth policy, histogram-based data partitioning, and feature bundling techniques (EFB), which together significantly reduce training time while maintaining high model accuracy. These enhancements enable LightGBM to train significantly faster than conventional GBDT frameworks, achieving up to twenty-fold improvements in computational speed while maintaining high classification accuracy (Ding et al., 2018). A notable advantage of LightGBM lies in its capacity to model complex, nonlinear relationships without assuming a predefined functional form between independent and dependent variables—an inherent limitation of traditional statistical approaches (Ding et al., 2018). Furthermore, the algorithm is robust to issues commonly encountered in real-world datasets, including multicollinearity, outliers, and missing values. Instead of focusing on marginal effects or regression coefficients, LightGBM quantifies feature importance based on variable interactions and their contribution to prediction outcomes (Elith et al., 2008).

This enables the model to provide more precise insights into variable influence, an advantage over traditional statistical models. The training process in LightGBM follows a gradient-based optimization approach, where data points with the highest gradient values are prioritized for training, reducing computational overhead without sacrificing accuracy. Furthermore, exclusive feature bundling (EFB) groups non-overlapping features together, improving memory efficiency and accelerating model performance. By leveraging these optimizations, LightGBM is capable of efficiently handling large datasets while preventing overfitting through hyperparameter tuning and regularization techniques. Overall, LightGBM stands out as a highly efficient and scalable gradient boosting algorithm, making it a preferred choice for tackling complex machine learning problems where both speed and accuracy are crucial.

4.3.9 Categorical Boosting (CatBoost)

Developed by Yandex researchers in 2017, CatBoost is a high-performance machine learning algorithm that builds upon the Gradient Boosting Decision Tree (GBDT) framework (Dorogush et al., 2018; Prokhorenkova et al., 2018). It enhances traditional GBDT methods by implementing a more efficient training scheme that fully leverages the input data and streamlines the boosting process to achieve improved predictive accuracy and computational efficiency. In contrast to traditional gradient boosting methods—which iteratively update weak learners based solely on previous errors—CatBoost initially applies equal weighting to all samples and subsequently increases the focus on instances with higher prediction errors. This iterative refinement continues until all training samples are incorporated, culminating in a final prediction through the aggregation of weighted outputs (G. Huang et al., 2019).

This process reduces overfitting, making the model more robust in practical applications (G. Huang et al., 2019). A key feature of CatBoost is its priority-based gradient evaluation, which reduces estimation bias, a common issue in traditional GBDT models. The algorithm first samples permutations of the training data, generating multiple models trained on different data orderings. By adjusting gradients dynamically, CatBoost enhances generalization and prevents over-reliance on specific training sequences (G. Huang et al., 2019). Despite its advantages, CatBoost can struggle with highly imbalanced datasets. Although it adjusts for class weights, extreme class imbalances may still result in poor predictions for minority classes, leading to higher misclassification rates. The incorrect tuning of hyperparameters like learning rate, tree depth, and iteration count can cause overfitting or underfitting, which adversely affects model performance, particularly for under-represented samples (Prokhorenkova et al., 2018). CatBoost's training process consists of several stages. The training process initiates with a classification adjustment phase that leverages target-based statistics. In this step, the algorithm calculates statistical indicators—such as the mean target value—within each categorical grouping, thereby incorporating class-specific distributional information into the model learning process. This adjustment allows the model to handle categorical variables effectively without extensive preprocessing. Next, boosting learning is applied, iteratively constructing trees that correct errors from previous predictions. By refining gradient values based on the loss function, CatBoost continuously enhances prediction accuracy (Prokhorenkova et al., 2018). Overall, CatBoost stands out as a highly efficient and scalable gradient boosting algorithm, particularly suited for datasets with categorical features and complex relationships. Its ability to minimize overfitting, optimize training efficiency, and improve generalization makes it a competitive choice for both classification and regression tasks.

4.3.10 Hyperparameter Tuning

Hyperparameter tuning is an essential process in optimizing machine learning models, ensuring better predictive accuracy, improved generalization, and a reduced risk of overfitting. Unlike model parameters, which are learned during training, hyperparameters are predefined settings that control the learning process, such as learning rate, tree depth, number of estimators, and regularization terms (Banyong et al., 2024; García-García et al., 2022). In this study, Bayesian optimization (BO) was employed for hyperparameter tuning, offering an efficient and systematic approach to finding the optimal configuration. BO operates by constructing a probabilistic surrogate model, typically a Gaussian Process (GP) or a Tree-structured Parzen Estimator (TPE), to approximate the objective function, which in this case is the model's performance metric (e.g., accuracy, AUC, or cross-entropy loss). Unlike exhaustive methods such as Grid Search, which tests all possible combinations, or Random Search, which selects configurations arbitrarily, Bayesian optimization strategically balances exploitation (focusing on promising hyperparameter regions) and exploration (searching for potentially better configurations). This is achieved through an acquisition function, such as Expected Improvement (EI) or Upper Confidence Bound (UCB), which determines the next set of hyperparameters to evaluate. After each iteration, the surrogate model is updated with new observations, refining predictions for subsequent searches. This iterative process continues until an optimal set of hyperparameters is found, minimizing computational costs while maximizing model performance. By leveraging Bayesian optimization, this study efficiently tuned hyperparameters in complex machine learning models, ensuring an optimal balance between training efficiency and predictive capability (Chongzhi, Lin, & Zhang, 2021; Shakya, Biswas, & Pal, 2022).

4.3.11 Model Comparison

To comprehensively conduct a comparative analysis of model effectiveness in identifying the key factors influencing the dependent variable (Y), this study explores the Multinomial Logit Model (MNL), Deep Neural Network (DNN), Convolutional Neural Network (CNN), Extreme Gradient Boosting (XGBoost), LightGBM, and CatBoost (Categorical Boosting). Given the diversity in modeling approaches, multiple evaluation metrics are employed to ensure a fair and robust comparison. To enhance the reliability and consistency of the assessment, a 10-fold cross-validation technique was applied to all models. This approach serves to attenuate overfitting and facilitates a more robust and generalizable assessment of model performance by systematically partitioning the dataset into distinct training and validation subsets. The comparative evaluation framework is anchored in three principal dimensions: predictive accuracy, model interpretability, and computational efficiency. With regard to predictive performance, the analysis underscores each model's capacity to effectively capture the structural relationships between explanatory variables and the target outcome, rather than relying exclusively on conventional classification accuracy. To rigorously assess this dimension, a suite of strategically selected performance indicators is employed. In pursuit of a comprehensive appraisal of classification efficacy, this study adopts a multidimensional evaluation scheme that encapsulates a broader spectrum of diagnostic insights across model outputs. Rather than relying solely on classification accuracy, which may be insufficient in the presence of class imbalance, additional metrics are employed to capture different aspects of predictive performance. Accuracy serves as a general measure of correctly classified instances, while recall (sensitivity) and specificity provide insights into the model's ability to correctly identify positive and negative cases, respectively. Furthermore, precision evaluates correctly predicted positives, while by integrating both precision and recall, the F1 score offers a single metric that reflects the model's ability to maintain consistency between correct positive predictions and sensitivity, which is especially beneficial in skewed class distributions, especially for imbalanced datasets. The area under the ROC curve (AUC-ROC) is also used to evaluate the models' ability to distinguish between different outcome categories. This metric is particularly valuable for multiclass classification, offering a threshold-independent assessment of performance. By applying these evaluation metrics uniformly across the Multinomial Logit Model (MNL), Deep Neural Network (DNN), Convolutional Neural Network (CNN), XGBoost, LightGBM, and CatBoost, this study ensures a robust and equitable comparison. This approach supports both the interpretability and predictive strength of each model in identifying the determinants of the dependent variable (Y).

Additionally, the area under the ROC curve (AUC-ROC) is employed to evaluate each model's ability to distinguish between outcome classes across multiple prediction scenarios (Banyong et al., 2024). In parallel, accuracy is used as a baseline indicator for classification performance, quantifying the ratio of correct predictions to the overall number of cases. While accuracy is widely adopted, it is often augmented with supplementary metrics to address potential biases arising from class imbalance (Banyong et al., 2024; García-García et al., 2022; X. Zhao, Yan, Yu, & Van Hentenryck, 2018). Considering the distinct methodological foundations of econometric versus machine learning techniques, the use of multiple evaluation criteria offers a more comprehensive and impartial basis for model comparison, capturing both interpretability and predictive reliability in estimating the determinants of the response variable.

Model performance was assessed using a set of evaluation metrics, including log loss, accuracy, precision, recall, and F1 score. To account for class imbalance and ensure uniform consideration across all outcome categories, macro-averaging was employed, thereby assigning equal weight to each class irrespective of its frequency (Mokhtarimousavi et al., 2020; Sokolova & Lapalme, 2009).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4-2}$$

$$\text{Recall (Sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4-3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{4-4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4-5}$$

$$\text{F1} - \text{Score} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{4-6}$$

$$\text{AUC} = \frac{1}{2}(\text{Recall} + \text{Specificity}) \tag{4-7}$$

These performance metrics were derived from the confusion matrix, which summarizes the classification outcomes as follows:

TP (True Positive): instances where actual positives were accurately identified as positive;

TN (True Negative): instances where actual negatives were correctly recognized as negative;

FP (False Positive): cases in which negative instances were mistakenly classified as positive;

FN (False Negative): cases in which positive instances were wrongly labeled as negative.

4.3.12 Shapley Additive Explanations (SHAP)

SHAP interprets complex machine learning models by showing how each feature influences predictions. It improves transparency and reliability through ranked plots, with position and color (red for positive, blue for negative) indicating each feature's impact (Mokhtarimousavi et al., 2020; Sokolova & Lapalme, 2009).

4.4 Results and Discussion

4.4.1 Descriptive Analysis

Below are the descriptive statistics for the key variables used in the analysis, including their distributions and summary measures relevant to the modeling process.

Table 4.2 summarizes the respondents' sociodemographic and travel behavior characteristics. The sample consisted of 52.43% male and 47.57% female respondents. The average household size was 3.36 members, with single-person households being the most common (33.88%), followed by two-person households

(28.92%). Regarding children under 18 in the household, 63.3% of respondents reported having children in this age group. In terms of household income, the majority of respondents had a monthly income in the range of 15,000 to 45,000 baht, with 33.55% earning more than 45,000 baht. The average income level (coded as categorical values) was 2.96. The most frequently reported travel purpose was for leisure or vacation (53.57%), followed by work or study (33.39%), and shopping (10.29%). The frequency of interprovincial travel averaged 2.11 times per month. Most respondents traveled one to three times per month (35.72%), while 33.93% reported traveling three to six times. In terms of mode choice, high-speed rail was the most preferred option, accounting for 29.42% of the responses, followed by bus (27.45%), conventional train (26.41%), and airplane (16.72%).

An examination of the skewness values across the variables showed a range between −0.55 and +0.70, indicating that most variables were approximately symmetrically distributed and did not display severe skewness. The kurtosis values were generally negative, suggesting that the distributions were flatter than the normal distribution (Champahom et al., 2023).

To investigate the interrelationships among the explanatory variables and to assess potential multicollinearity, a Pearson correlation analysis was conducted. As shown in Figure 4.2, all pairwise correlation coefficients remain below the standard threshold of 0.5 (Vatcheva, Lee, McCormick, & Rahbar, 2016), indicating that no substantial multicollinearity is present. The relatively low levels of intercorrelation confirm that the inclusion of these variables in a multivariate framework does not pose multicollinearity concerns, thereby validating the statistical independence required for

Table 4.2 Statistical Overview of Personal Characteristics and Mobility Behavior

Variable	Description	Categorical Variable (%)	Mean	*SD*	*SK*	*KU*
Gender	Male = 1	52.43	0.5243	0.4994	−0.0975	−1.9908
	Female = 0	47.57				
	Total	100				

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.2 Statistical Overview of Personal Characteristics and Mobility Behavior (Continued)

Variable	Description	Categorical Variable (%)	Mean	*SD*	*SK*	*KU*
Household members	Household members	33.88	3.3561	1.1090	−0.3527	−0.5709
	1 person = 1					
	2 people = 2	28.92				
	3 people = 3	15.67				
	4 people = 4	15.11				
	More than four people = 5	6.41				
	Total	100				
Children	Have children under 18 in the household = 1	63.3	0.6330	0.4820	−0.5520	−1.6956
	No children under 18 in the household = 0	36.7				
	Total	100				
Income	Less than 15,000 = 1	2.22	2.9557	0.8705	−0.1163	−1.2520
	15,000–30,000 = 2	30.7				
	30,001–45,000 = 3	33.54				
	More than 45,000 = 4	33.55				
	Total	100				
Work	Travel for study or work.	33.39	0.3339	0.4716	0.7043	−1.5043
	Yes = 1					
	No = 0	66.61				

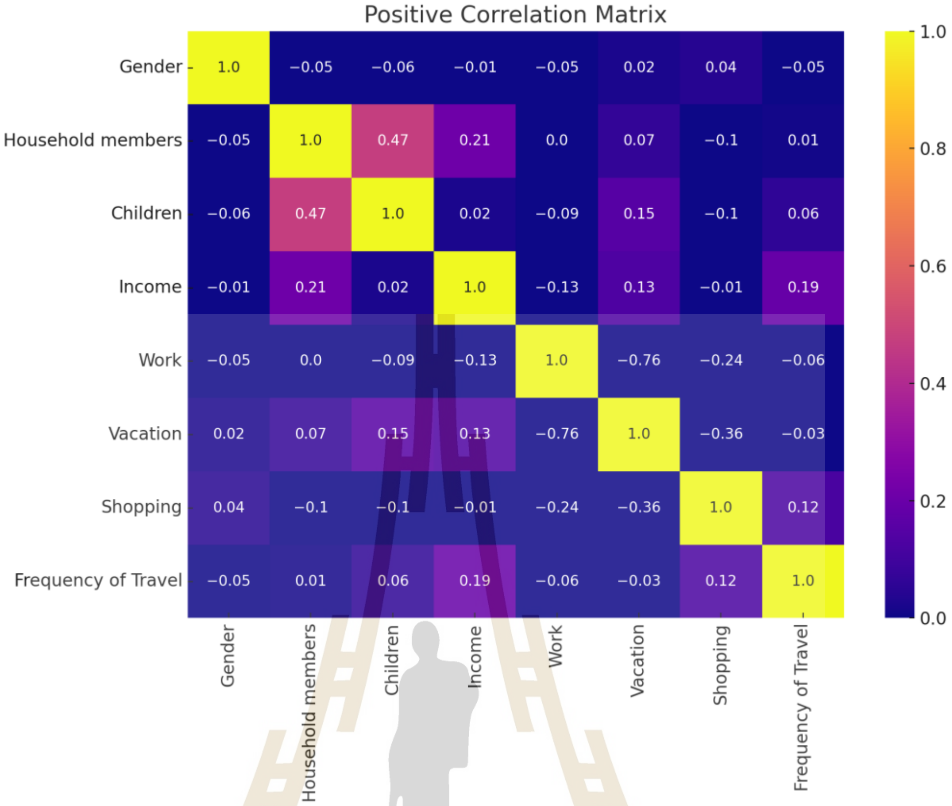
มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.2 Statistical Overview of Personal Characteristics and Mobility Behavior (Continued)

Variable	Description	Categorical Variable (%)	Mean	*SD*	*SK*	*KU*
	Total	100				
Vacation	Travel for leisure or vacation. Yes = 1	53.57	0.5357	0.4987	−0.1431	−1.9799
	No = 0	46.43				
	Total	100				
Shopping	Travel for shopping. Yes = 1	10.29	0.1029	0.3038	2.6146	4.837
	No = 0	89.71				
	Total	100				
Frequency of Travel	1–3 times = 1	35.72	2.1097	1.0673	0.5877	−0.9058
	3–6 times = 2	33.93				
	6–9 times = 3	16.35				
	More than nine times = 4	14				
	Total	100				
Mode	High speed railways = 1	29.42				
	Bus = 2	27.45				
	Train = 3	26.41				
	Airplane = 4	16.72				
	Total	100				

Note: *SD* = Standard Deviation, *SK* = Skewness, and *KU* = Kurtosis

มหาวิทยาลัยเทคโนโลยีสุรนารี



	Gender	Household members	Children	Income	Work	Vacation	Shopping	Frequency of Travel
Gender	1.0	−0.05	−0.06	−0.01	−0.05	0.02	0.04	−0.05
Household members	−0.05	1.0	0.47	0.21	0.0	0.07	−0.1	0.01
Children	−0.06	0.47	1.0	0.02	−0.09	0.15	−0.1	0.06
Income	−0.01	0.21	0.02	1.0	−0.13	0.13	−0.01	0.19
Work	−0.05	0.0	−0.09	−0.13	1.0	−0.76	−0.24	−0.06
Vacation	0.02	0.07	0.15	0.13	−0.76	1.0	−0.36	−0.03
Shopping	0.04	−0.1	−0.1	−0.01	−0.24	−0.36	1.0	0.12
Frequency of Travel	−0.05	0.01	0.06	0.19	−0.06	−0.03	0.12	1.0

Figure 4.2 Heatmap of positive correlation among socioeconomic factors.

4.4.2. Hyperparameter Optimization Using Bayesian Optimization

In the development of high-performance predictive systems, hyperparameter optimization plays a pivotal role in enhancing model precision and generalizability. This consideration is particularly salient in the present study, which incorporates a diverse array of model architectures, including XGBoost, LightGBM, CatBoost, Deep Neural Network (DNN), and Convolutional Neural Network (CNN). Selecting optimal hyperparameter configurations is essential for maximizing model efficacy. Accordingly, this research utilizes Bayesian hyperparameter optimization, a probabilistic and data-efficient method for identifying optimal parameter settings under uncertainty. A summary of the optimized hyperparameter values for each algorithm is provided in Table 4.3.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.3 Optimized Hyperparameters of the Compared Models

Model	Description	Value
XGBoost	n_estimators	210
	max_depth	6
	learning_rate	0.21977629940065888
	subsample	0.8166577433928425
	colsample_bytree	0.6713509802187799
	gamma	0.016849547970738232
	reg_alpha	4.443653782167797
	reg_lambda	2.162026822349147
LightGBM	n_estimators	493
	max_depth	5
	learning_rate	0.0441
	subsample	0.8957
	colsample_bytree	0.9840
	reg_alpha	0.0754
	reg_lambda	0.0788
	random_state	42
CatBoost	iterations	157
	depth	7
	learning_rate	0.16214535070336702
	l2_leaf_reg	4.815982341211366
	random_strength	5.493473561258114
	bagging_temperature	0.5270768048053522
	border_count	113
	loss_function	‘MultiClass’
	random_state	42
Deep Neural Network	first_dense_units	191
	second_dense_units	87
	dropout_rate	0.2294
	optimizer	Adam

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 4.3 Optimized Hyperparameters of the Compared Models (Continued)

Model	Description	Value
	learning_rate	0.00033001201097314586
	epochs	32
	batch_size	16
Convolutional Neural Network	filters	64
	kernel_size	3
	activation	‘relu’
	pool_size	2
	dense_units	64
	dropout_rate	0.3
	output_activation	‘softmax’
	optimizer	Adam
	learning_rate	0.001
	loss	‘categorical_crossentropy’
	batch_size	16
	epochs	30

4.4.3. Model Performance

Model performance evaluation is critical for assessing the predictive accuracy of travel mode choice models, particularly in the context of promoting high-speed rail (HSR) as a competitive alternative to conventional transport modes such as buses, trains, and airplanes. To ensure methodological rigor and reduce sampling bias, a stratified 10-fold cross-validation procedure was employed, whereby the dataset was partitioned into ten mutually exclusive subsets. Each subset was sequentially designated as a validation set, while the remaining nine were used for model training, thereby enhancing the generalizability of the results. Key evaluation metrics—accuracy, precision, recall, F1 score (macro-averaged), and AUC for multiclass classification—were utilized to benchmark model performance. The outcomes are systematically presented in Figures 4.3 and 4.4 to facilitate comparative analysis.

มหาวิทยาลัยเทคโนโลยีสุรนารี

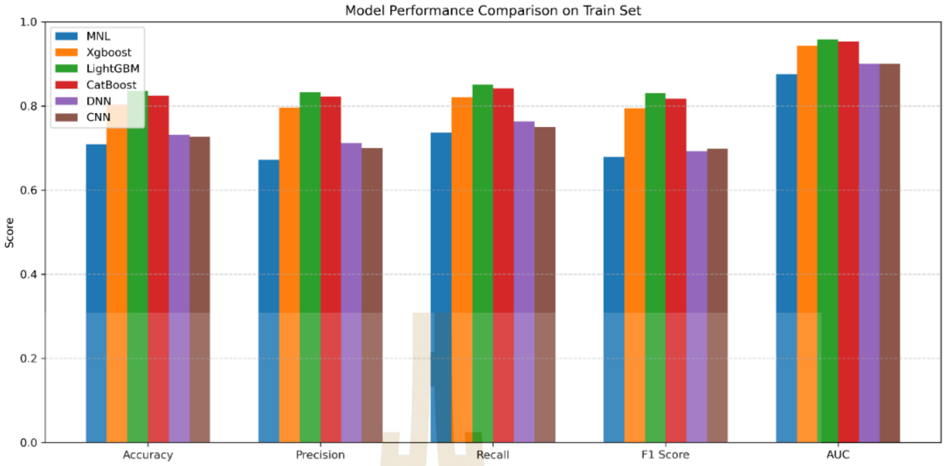


Figure 4.3 Model performance comparison on the training set.

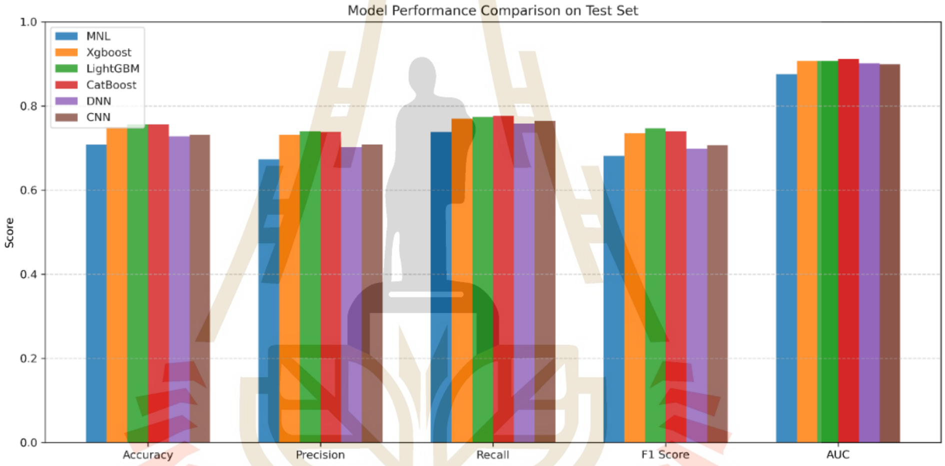


Figure 4.4 Model performance comparison on the test set.

Figures 4.3 and 4.4 present the cross-validated results for five predictive models—XGBoost, LightGBM, CatBoost, Deep Neural Network (DNN), and Convolutional Neural Network (CNN)—in forecasting travel mode transitions toward high-speed rail (HSR). Performance was assessed using standard classification metrics, including accuracy, precision, recall, F1 score, and area under the ROC curve (AUC), reported for both training and test sets to evaluate generalization capability.

Among the gradient boosting techniques, LightGBM achieved the highest scores on the training set, yielding peak values for accuracy (0.8355), F1 score (0.8299), and AUC (0.9584). Although CatBoost exhibited slightly lower training performance, it outperformed all models on the test set with the highest AUC (0.9113) and demonstrated consistent performance across all metrics—accuracy (0.7557), precision

(0.7384), recall (0.7760), and F1 score (0.7404)—indicating robust generalization to unseen data.

The deep learning models (DNN and CNN) delivered stable yet slightly lower results on the test set, with accuracy and F1 scores ranging between 0.727 and 0.734, and comparatively reduced AUC values. Notably, the DNN model achieved the highest recall (0.7579), reflecting its sensitivity to correctly identifying positive cases.

In contrast, the Multinomial Logit (MNL) model underperformed across all evaluation criteria—accuracy (0.7081), precision (0.6731), recall (0.7373), F1 score (0.6812), and AUC (0.8753)—reflecting its limitations in handling nonlinearities and complex variable interactions. Despite its popularity in transport research for interpretability, MNL’s structural constraints limit its predictive accuracy. Overall, CatBoost emerged as the most effective model for HSR adoption prediction, striking the optimal balance between accuracy and generalization across all performance metrics.

Figure 4.5 illustrates the comparison of ROC curves across five classification models for predicting travel mode transitions toward high-speed rail (HSR) adoption. The CatBoost model achieved the highest macro-average AUC (0.911), followed closely by XGBoost and LightGBM (0.907), while DNN and CNN showed slightly lower but comparable performance (0.901 and 0.899, respectively). The results indicate that CatBoost demonstrates superior discriminatory power in differentiating among multiple travel mode classes. Only machine learning models are visualized in this figure, as the traditional Multinomial Logit (MNL) model showed substantially lower AUC and overall predictive performance, making it less effective for visual comparison.

มหาวิทยาลัยเทคโนโลยีสุรนารี

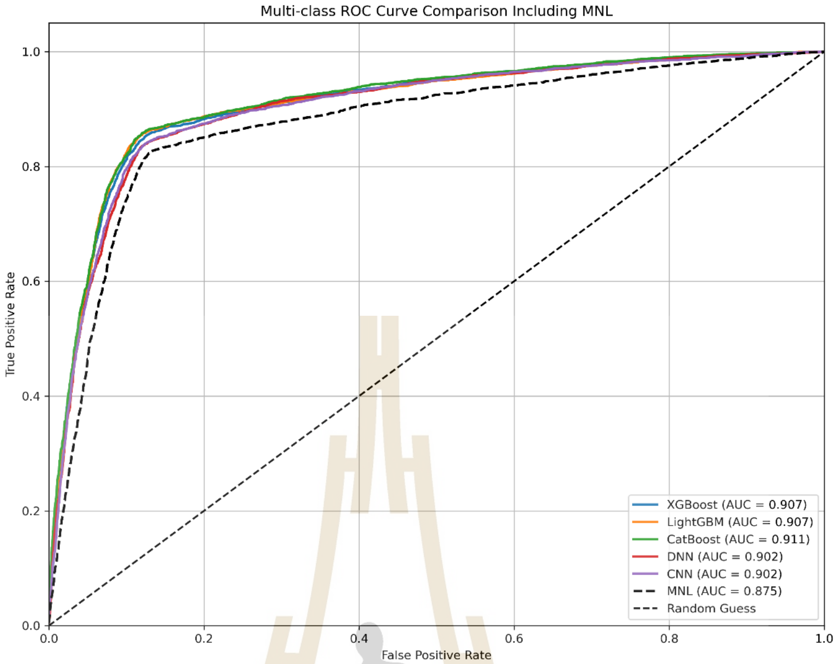


Figure 4.5 Multiclass ROC curve comparison across models including Multinomial Logit.

The comparison of models in this study highlights significant performance differences across various approaches with distinct structures and assumptions. The Multinomial Logit (MNL) model, valued for its simplicity and behavioral interpretability, faces limitations due to its assumptions of linear relationships and the Independence of Irrelevant Alternatives (IIA), which hinder its ability to capture complex, nonlinear interactions (Kalantari, Sabouri, Brewer, Ewing, & Tian, 2025). As a result, MNL's predictive performance, particularly in accuracy and AUC, was outperformed by machine learning (ML) models (Shahdah, Elharoun, Ali, Elbany, & Elagamy, 2025; Yin, Wu, Sun, Meng, & Lee, 2024).

In contrast, machine learning models, particularly CatBoost, LightGBM, and XGBoost, exhibited significantly superior performance. These models excel at capturing complex variable interactions and can process categorical data directly, minimizing the need for extensive preprocessing (Champahom et al., 2023). Notably, CatBoost demonstrated particular strengths in addressing overfitting through integrated regularization mechanisms, which contributed to its robust accuracy and stability in the present dataset (Champahom et al., 2023). CatBoost utilizes ordered boosting (also known as ordered target statistics), which minimizes prediction shift during training and employs a symmetric tree-growing algorithm that promotes balanced and

มหาวิทยาลัยเทคโนโลยีสุรนารี

generalizable trees. Furthermore, its native handling of categorical variables through advanced encoding techniques allows for the preservation of essential information without manual transformation. The model also incorporates gradient-based regularization to enhance robustness, enabling it to perform well on structured datasets with minimal hyperparameter tuning (Himeur et al., 2022).

Although Deep Neural Networks (DNNs) are capable of learning intricate, high-level data representations (Himeur et al., 2022), their use on small datasets often leads to overfitting, as the number of parameters greatly exceeds the available data (Himeur et al., 2022). This leads the model to memorize training data rather than generalize underlying patterns, thereby reducing its effectiveness on unseen test data. Moreover, DNNs are susceptible to learning noise instead of meaningful structure (Himeur et al., 2022). Mitigating overfitting in such cases requires advanced techniques, including data augmentation, transfer learning, or other regularization strategies (Himeur et al., 2022). These complexities make the practical deployment of DNNs in small-data contexts challenging, where simpler models such as gradient boosting often outperform them. While Convolutional Neural Networks (CNNs) can technically be adapted for tabular data by arranging features into a matrix format resembling an image, thereby enabling the model to exploit spatial structure through convolutional filters, this approach may not be well suited in contexts where the features are inherently unordered and lack spatial correlation (Shwartz-Ziv & Armon, 2022). In the present study, the data were structured in a tabular format with semantically distinct and logically independent variables (Chehreh Chelgani, Homafar, Nasiri, & Rezaei laksar, 2024). Therefore, convolutional learning was unable to extract salient features effectively (Shwartz-Ziv & Armon, 2022). While the deep learning models in this study, particularly DNN and CNN, were constrained by the limited sample size, several strategies were employed to mitigate overfitting and enhance generalization. Although the original sample consisted of 3200 respondents, the stated preference (SP) design included 12 hypothetical choice tasks per person, resulting in a dataset of 38,400 observations. This enriched dataset allowed for more effective model training despite the moderate respondent base. To further address overfitting, we implemented stratified 10-fold cross-validation, Bayesian hyperparameter tuning, and incorporated dropout layers and early stopping during the training of DNN and CNN models. These regularization techniques aimed to reduce model complexity and prevent memorization of the training data. In contrast, models tailored for tabular data—such as CatBoost and XGBoost—demonstrated greater efficiency in handling categorical variables without preprocessing and capturing nonlinear patterns, resulting in higher accuracy and more stable performance on this dataset (Borisov et al., 2022; Grinsztajn et al., 2022; Shwartz-Ziv & Armon, 2022).

Figure 4.6 presents the SHAP-based global feature importance analysis, offering key insights into the relative influence of individual factors driving the shift in travel mode choices, particularly toward the adoption of high-speed rail (HSR). The results emphasize that economic and temporal service characteristics play a far more dominant role than sociodemographic attributes in shaping travelers' decision-making processes. Travel cost emerges as the most influential factor across all travel modes, particularly in the context of airplane and bus choices. This underscores the high price sensitivity among travelers, especially in low-to-middle-income contexts like Thailand, where affordability remains a crucial determinant in mode selection (Salas et al., 2022; M. Xu, Shuai, Wang, Liu, & Zhou, 2023). This finding aligns with established transport economics theory, which identifies cost as a central constraint in mode choice decisions (Salas et al., 2022; M. Xu et al., 2023).

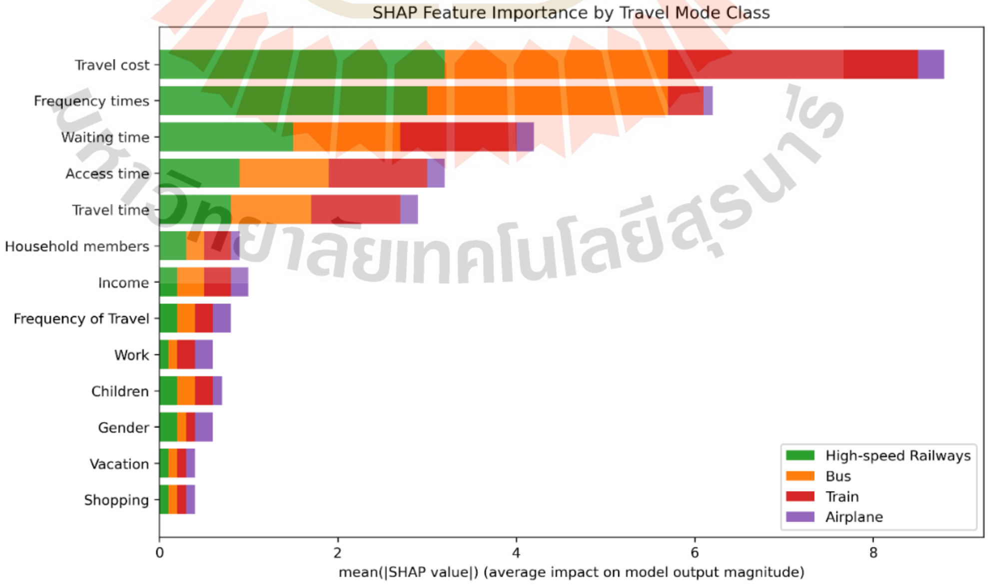


Figure 4.6 SHAP feature importance disaggregated by travel mode class

มหาวิทยาลัยเทคโนโลยีสุรนารี

Beyond cost, service frequency and waiting time emerge as key determinants, especially among train and bus users. These results suggest that operational predictability and schedule reliability are key considerations for travelers (Salas et al., 2022; M. Xu et al., 2023). Low service frequency and prolonged waiting times pose substantial barriers to HSR adoption, reinforcing the need for strategic interventions to enhance service regularity (Tiong, Ma, & Palmqvist, 2023). This reflects the concept of perceived temporal utility, where delays and inconsistencies in departure or arrival times reduce the perceived value of the service. Travelers tend to associate high-frequency services with lower opportunity costs and greater flexibility, which enhances the attractiveness of the mode. Access time and travel time, representing components of total journey effort, show moderate SHAP values across modes. Their consistent influence across all alternatives highlights their importance in overall convenience, particularly for users with limited access to HSR stations (Salas et al., 2022; M. Xu et al., 2023). This is consistent with the theory of generalized travel cost, which incorporates not only monetary cost but also time-related burdens such as first-mile/last-mile travel and in-vehicle duration. Inadequate access and long travel times increase the perceived disutility of HSR, especially for travelers who value efficiency or have time constraints.

In contrast, sociodemographic variables—household size, income, gender, and trip purpose—show comparatively low SHAP importance. This indicates that mode choice is driven more by service characteristics than by traveler attributes, a trend that diverges from traditional behavioral models that emphasize demographics (Salas et al., 2022; M. Xu et al., 2023). Although variables like income are commonly used in travel behavior studies, the current results reveal that actual trip cost has a more direct and substantial impact on decision-making than income levels. Nonetheless, variables such as travel frequency and gender, though ranked lower, retain non-zero SHAP values, suggesting latent behavioral patterns. For instance, regular travelers may prioritize punctuality and convenience more heavily, increasing their likelihood of shifting to HSR when reliability is perceived to be superior (Salas et al., 2022; M. Xu et al., 2023). This finding implies that frequent intercity travelers such as commuters, business travelers, or students who travel regularly between provinces represent a key target segment for early HSR adoption. These groups are often time-sensitive and more responsive to improvements in service frequency and travel time, making them ideal candidates for tailored marketing strategies and service design.

In summary, the SHAP analysis confirms the primacy of economic and temporal attributes notably cost, service frequency, and waiting time in determining mode shifts toward HSR. These findings have direct implications for transportation policy and service design, underscoring the need to enhance affordability and reliability to foster greater adoption of HSR systems. In addition, the results highlight specific passenger groups—such as frequent travelers, larger households, and higher-income individuals—who exhibit greater sensitivity to improvements in HSR-related service attributes. These segments represent strategic targets for early adoption and should be prioritized in marketing, fare policy, and service design to maximize uptake and long-term viability.

The disaggregated SHAP analysis by travel mode reveals a distinctive behavioral pattern among high-speed rail (HSR) users (Figure 4.7), in which household size emerges as a prominent determinant. This effect is not observed in the other transport modes. Moreover, income demonstrates contrasting influences: individuals with higher income levels show a greater propensity to choose HSR, while their likelihood of using bus services declines. These findings suggest a clear socioeconomic differentiation in travel mode preferences.

Across all four transport modes (HSR, bus, train, and airplane), travel cost and time-related service attributes, including service frequency, waiting time, and access time, consistently emerge as the most influential factors in shaping mode choice. These variables exert particularly strong effects among bus and train users (Figures 4.8 and 4.9), highlighting the critical importance of operational efficiency and schedule reliability. In the case of air travel (Figure 4.10), access time and cost are especially salient, indicating heightened sensitivity to first- and last-mile connectivity as well as fare levels in mode selection.

These findings underscore the policy importance of enhancing affordability and temporal reliability in HSR service design. Tailored strategies such as competitive fare structures, increased service frequency, and improved station accessibility are vital to promoting HSR adoption. Moreover, targeted incentives for specific user segments, such as larger households and higher-income travelers, may further enhance the attractiveness of HSR in emerging intercity transport systems.

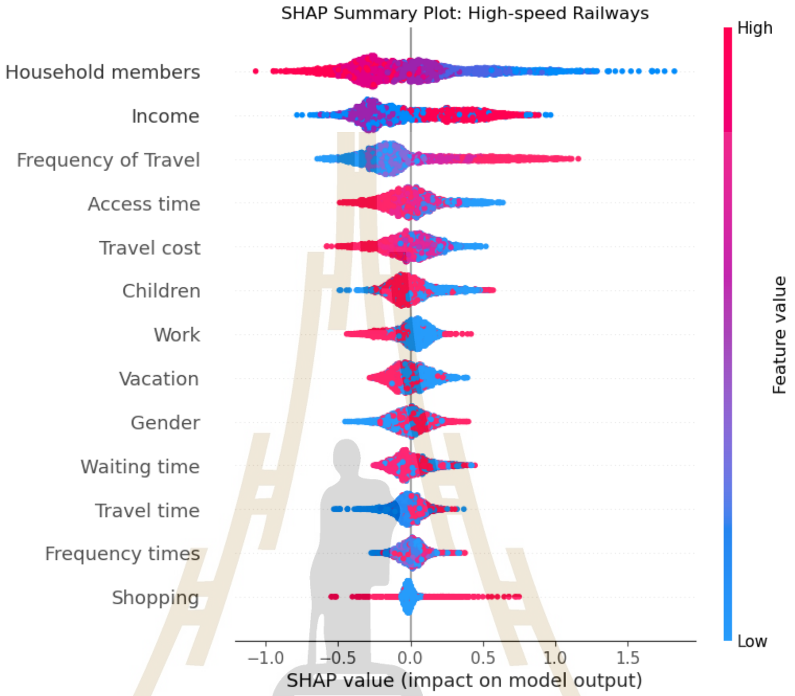


Figure 4.7 SHAP feature importance and impact on high-speed rail mode choice.

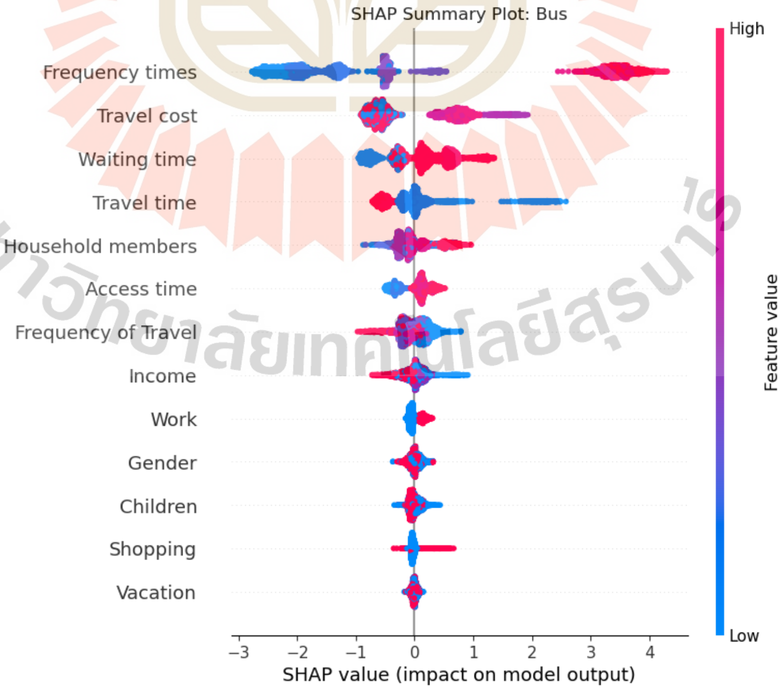


Figure 4.8 SHAP feature importance and impact on bus mode choice.

มหาวิทยาลัยเทคโนโลยีสุรนารี

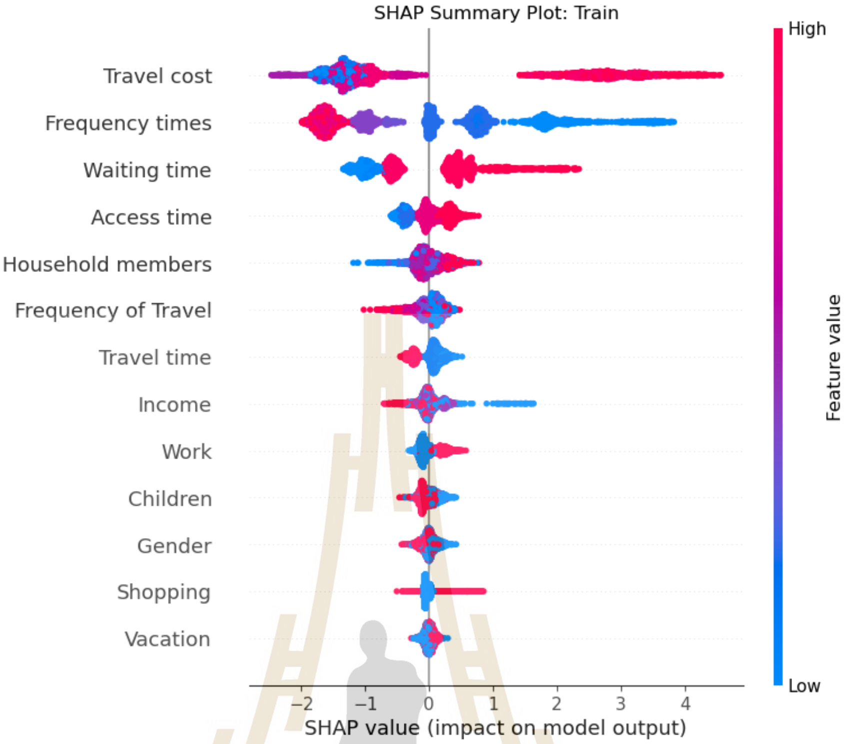


Figure 4.9 SHAP feature importance and impact on train mode choice.

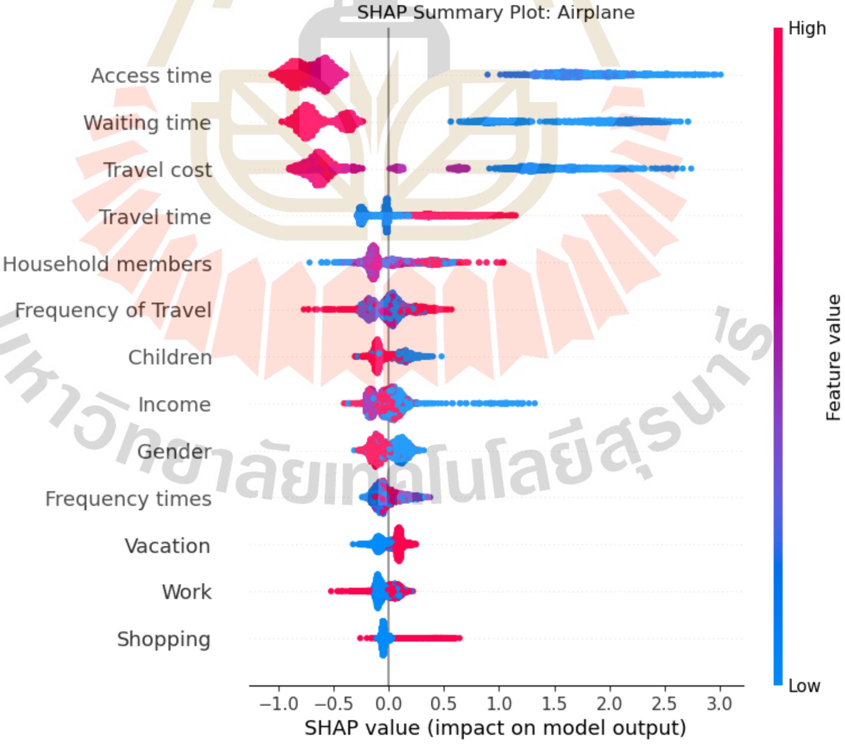


Figure 4.10 SHAP feature importance and impact on airplane mode choice.

มหาวิทยาลัยเทคโนโลยีสุรนารี

4.5 Conclusions

This study compared the predictive performance of traditional and advanced modeling techniques in forecasting travel mode choice amid the introduction of high-speed rail (HSR) in Thailand. Using stated preference data from diverse geographic and socioeconomic contexts, the analysis evaluated the Multinomial Logit (MNL) model alongside advanced models including Extreme Gradient Boosting, LightGBM, CatBoost, Deep Neural Networks (DNNs), and Convolutional Neural Networks (CNNs), with model training enhanced through cross-validation and Bayesian optimization. The proposed modeling framework not only demonstrates strong predictive performance in the context of Thailand’s HSR, but also holds potential for application in similar large-scale transportation infrastructure studies across different countries or systems. CatBoost achieved the highest performance on the test set, with an AUC of 0.9113 and an accuracy of 0.7557, demonstrating superior generalization compared to all other models. While LightGBM obtained the highest training AUC of 0.9584, CatBoost maintained more stable performance on unseen data, attributed to its algorithmic features such as native handling of categorical variables and ordered boosting. The MNL model, although interpretable and easy to implement, was constrained by structural assumptions like linearity and the Independence of Irrelevant Alternatives (IIA), resulting in a lower AUC of 0.8753. Deep learning models (DNN and CNN) showed limited performance due to data volume constraints and the mismatch between CNN architecture and tabular data structure.

Feature importance analysis using SHAP revealed that travel cost, service frequency, and waiting time were the most influential factors in mode choice, particularly for bus and air travel. In contrast, sociodemographic variables had limited impact, indicating that service attributes play a more dominant role in shaping HSR-related travel decisions. Although this study did not perform an explicit segmentation analysis, the results suggest certain passenger groups are more likely to adopt high-speed rail services. In particular, travelers with higher income levels, larger households, and those who travel frequently for business or educational purposes demonstrate greater sensitivity to HSR’s economic and temporal attributes. These findings point to the practical value of targeting such groups in future marketing and service planning efforts, especially during the early stages of HSR implementation.

4.5.1 Policy Recommendations

Informed by the empirical findings, this study proposes the following policy recommendations to promote the adoption of high-speed rail, as depicted in Figure 4.11:

1) Design competitive fare structures by offering promotional pricing, monthly passes, or integrated fare bundles with other public transportation services;
2) Increase service frequency and minimize waiting times to enhance the reliability and attractiveness of high-speed rail operations;
3) Improve accessibility to rail stations through feeder systems such as shuttle buses or local transit networks that facilitate first-mile and last-mile connectivity;
4) Encourage regular intercity travelers to adopt high-speed rail through loyalty programs or targeted fare incentives;
5) Simplify the ticketing process by developing a seamless and intuitive platform for booking and payment, supporting mobile access, QR code usage, and electronic wallets.

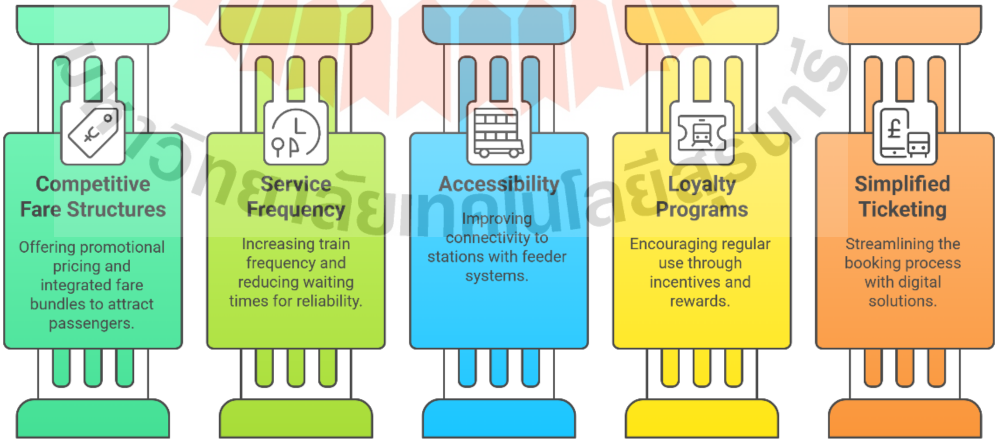


Figure 4.11 Enhancing high-speed rail adoption through strategic policy recommendations.

In conclusion, this study demonstrates not only the superior predictive performance of machine learning models, especially Categorical Boosting, in modeling

มหาวิทยาลัยเทคโนโลยีสุรนารี

travel mode choice but also underscores the critical role of service-related attributes in influencing behavioral shifts. These insights provide a practical foundation for designing efficient, inclusive, and user-centered transportation policies, particularly in the context of high-speed rail development in Thailand and other emerging economies.

4.5.2 Limitations and Further Research

This study provides valuable insights into travel mode choice behavior in the context of high-speed rail (HSR) development in Thailand. However, several limitations should be acknowledged. First, the use of stated preference (SP) data introduces potential hypothetical bias, as respondents' stated intentions under hypothetical scenarios may not align with their actual behavior. This is especially relevant for emerging systems like HSR, for which respondents have no prior experience. Future research should incorporate revealed preference (RP) data following real-world implementation to validate model predictions. In addition, future studies may consider developing multiple hypothetical travel scenarios and evaluating different predictive targets beyond mode choice, such as user satisfaction, likelihood of repeated use, or pricing sensitivity.

Second, although advanced machine learning and deep learning models were employed, the modest dataset size may have limited the performance of complex architectures, particularly Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs), which are typically optimized for large-scale data. Despite the use of cross-validation and hyperparameter tuning, these models may still be prone to overfitting. Expanding the dataset and employing techniques such as transfer learning or ensemble methods may enhance performance in future studies.

Third, the study utilized a non-probability sampling approach, which, despite ensuring regional coverage, may limit the generalizability of the results due to potential selection bias. Future research should adopt stratified or probability-based sampling techniques to ensure more representative and externally valid conclusions.

Fourth, the models do not explicitly account for latent psychological constructs such as attitudes, perceptions, or satisfaction, which are increasingly recognized as key determinants of travel behavior. This unexpectedly weak influence of demographic

มหาวิทยาลัยเทคโนโลยีสุรนารี

variables may be attributed to specific cultural characteristics in Thailand, where travel decisions are more influenced by practical constraints (e.g., cost and access) than personal traits. Additionally, limitations in the questionnaire design may have restricted the ability to capture attitudinal or psychological dimensions that mediate the effect of sociodemographic factors. Incorporating latent variable modeling techniques such as hybrid choice models or Structural Equation Modeling (SEM) could improve explanatory power.

Fifth, while Shapley Additive Explanations (SHAP) was used to interpret global feature importance, this method does not account for temporal dynamics or context-specific variability. Future work should consider longitudinal data and time series modeling to capture behavioral adaptation following HSR implementation. In addition, future research could extend SHAP analysis by incorporating interaction effects between key service attributes such as travel cost, waiting time, and service frequency to uncover how combinations of factors jointly influence mode choice decisions. This would provide deeper behavioral insights beyond individual variable importance. Cross-country comparative studies, particularly in other developing nations, may also offer valuable insights into the transferability and generalizability of machine learning approaches in travel behavior research. Moreover, future studies could extend SHAP analysis by incorporating interaction effects among key variables, such as cost, waiting time, and service frequency. Assessing these combined influences may reveal compound sensitivities that are not apparent when considering each factor in isolation. Additionally, clustering algorithms or latent class analysis could be applied to segment travelers into subgroups—such as budget-sensitive, time-sensitive, or frequent users—and assess model performance and behavioral responses within each cluster. These segmentation techniques would allow for more targeted policy recommendations and service planning based on distinct traveler profiles.

Finally, future research should explore traveler segmentation techniques such as clustering, latent class analysis, or subgroup-specific modeling to uncover distinct user profiles and heterogeneity in preferences. This would allow for more tailored policy interventions and service strategies that address the unique needs of different passenger groups, such as frequent travelers, students, or price-sensitive individuals.

4.6 References

Aalami, M. R., Anari, A., Shafighfard, T., & Talatahari, S. (2013). A robust finite element analysis of the rail-wheel rolling contact. *Advances in Mechanical Engineering, 5*, 272350.

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability, 11*(1), 189. doi:10.3390/su11010189

Abenoza, R. F., Cats, O., & Susilo, Y. O. (2019). How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips. *Transportation, 46*(5), 1615-1642.

Abulibdeh, A. (2018). Implementing congestion pricing policies in a MENA Region City: Analysis of the impact on travel behaviour and equity. *Cities, 74*, 196-207. doi:https://doi.org/10.1016/j.cities.2017.12.003

Abulibdeh, A. (2023). Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis. *Transportation Research Interdisciplinary Perspectives 20*, 100852. doi:https://doi.org/10.1016/j.trip.2023.100852

Abulibdeh, A. (2023). Machine learning models for travel behavior analysis in Qatar: Comparing XGBoost, MNL, and BL models. *Transportation Research Record, 2677*(3), 245-260.

Adland, R., Jia, H., Lode, T., & Skontorp, J. (2021). The value of meteorological data in marine risk assessment. *Reliability Engineering & System Safety 209*, 107480. doi:10.1016/j.ress.2021.107480

AirAsia move. (2024). Flight. Retrieved from https://www.airasia.com/th/th

Akter, T., & Alam, B. M. (2024). Travel mode choice behavior analysis using multinomial logit models towards creating sustainable college campus: A case study of the University of Toledo, Ohio. *Frontiers in Future Transportation, 5*, 1389614.

Al-Atroush, M., Mustaffa, Z., & Sebeay, T. (2022). *Emerging trends in overcoming the weather barrier to sustainable mobility in gulf and tropical cities.* Paper presented at the IOP Conference Series: Earth and Environmental Science.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Al Doori, A. (2017). Waiting time factor in public transport by binary logistic regression. *Australian Journal of Basic and Applied Sciences, 11*(4), 72-76.

Anagnostopoulos, A. (2024). High-speed railway and safety: Insights from a bibliometric approach. *High-speed Railway, 2*(3), 187-196. doi:https://doi.org/10.1016/j.hspr.2024.08.004

Arencibia, A. I., Feo-Valero, M., García-Menéndez, L., & Román, C. (2015). Modelling mode choice for freight transport using advanced choice experiments. *Transportation Research Part A: Policy and Practice 75*, 252-267.

Armah, F. A., Yawson, D. O., & Pappoe, A. A. N. M. (2010). A Systems Dynamics Approach to Explore Traffic Congestion and Air Pollution Link in the City of Accra, Ghana. *Sustainability, 2*(1), 252-265. doi:10.3390/su2010252

Avogadro, N., Pels, E., & Redondi, R. (2023). Policy impacts on the propensity to travel by HSR in the Amsterdam–London market. *Socio-Economic Planning Sciences, 87B*, 101585. doi:10.1016/j.seps.2023.101585

Avogadro, N., Pels, E., & Redondi, R. (2023). Policy impacts on the propensity to travel by HSR in the Amsterdam – London market. *Socio-Economic Planning Sciences, 87*, 101585. doi:https://doi.org/10.1016/j.seps.2023.101585

Bagherzadeh, F., Shafighfard, T., Khan, R. M. A., Szczuko, P., & Mieloszyk, M. (2023). Prediction of maximum tensile stress in plain-weave composite laminates with interacting holes via stacked machine learning algorithms: A comparative study. *Mechanical Systems and Signal Processing, 195*, 110315. doi:https://doi.org/10.1016/j.ymssp.2023.110315

Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Seefong, M., . . . Jomnonkwao, S. (2024). A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand. *Results in Engineering, 24*, 103110. doi:https://doi.org/10.1016/j.rineng.2024.103110

Başakın, E. E., Ekmekcioğlu, Ö., Çıtakoğlu, H., & Özger, M. (2022). A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Computing and Applications, 34*(1), 783-812.

Becker, N., & Carmi, N. (2019). Changing trip behavior in a higher education institution: The role of parking fees. *International Journal of Sustainable Transportation, 13*(4), 268-277. doi:10.1080/15568318.2018.1463418

Belgiawan, P. F., Ilahi, A., & Axhausen, K. W. (2019). Influence of pricing on mode choice decision in Jakarta: A random regret minimization model. *Case Studies on Transport Policy, 7*(1), 87-95. doi:10.1016/j.cstp.2018.12.002

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9): MIT press.

Bergantino, A. S., & Madio, L. (2020). Intermodal competition and substitution. HSR versus air transport: Understanding the socio-economic determinants of modal choice. *Research in Transportation Economics, 79*, 100823. doi:https://doi.org/10.1016/j.retrec.2020.100823

Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems* (pp. 283-307): Elsevier.

Binsuwadan, J., Wardman, M., de Jong, G., Batley, R., & Wheat, P. (2023). The income elasticity of the value of travel time savings: A meta-analysis. *Transport Policy 136*, 126-136. doi:https://doi.org/10.1016/j.tranpol.2023.03.013

Birolini, S., Malighetti, P., Redondi, R., & Deforza, P. (2019). Access mode choice to low-cost airports: Evaluation of new direct rail services at Milan-Bergamo airport. *Transport policy, 73*, 113-124. doi:10.1016/j.tranpol.2018.10.008

Blanquart, C., & Koning, M. (2017). The local economic impacts of high-speed railways: Theories and facts. *European Transport Research Review, 9*, 12.

Blanquart, C., & Koning, M. (2017). The local economic impacts of high-speed railways: Theories and facts. *European Transport Research Review 9*, 1-14.

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE transactions on neural networks and learning systems*

Buehler, R. (2011). Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography 19*(4), 644-657.

Bursa, B., Mailer, M., & Axhausen, K. W. (2022). Intra-destination travel behavior of alpine tourists: a literature review on choice determinants and the survey work. *Transportation 49*(5), 1465-1516.

BusOnlineTicket.co.th. (2024). Book Thailand Bus Tickets Online. Retrieved from https://www.busonlineticket.co.th/

Calvo, F., Eboli, L., Forciniti, C., & Mazzulla, G. (2019). Factors influencing trip generation on metro system in Madrid (Spain). *Transportation Research Part D: Transport and Environment, 67*, 156-172.

Cartenì, A., Pariota, L., & Henke, I. (2017). Hedonic value of high-speed rail services: Quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities. *Transportation Research Part A: Policy and Practice 100*, 348-365. doi:10.1016/j.tra.2017.04.018

Cascetta, E., Cartenì, A., Henke, I., & Pagliara, F. (2020). Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post evaluation and future perspectives. *Transportation Research Part A: Policy and Practice 139*, 412-428. doi:https://doi.org/10.1016/j.tra.2020.07.008

Champahom, T., Banyong, C., Hantanong, N., Se, C., Jomnonkwao, S., & Ratanavaraha, V. (2023). Factors influencing the willingness to pay for motorcycle safety improvement: A structural equation modeling approach. *Transportation Research Interdisciplinary Perspectives, 22*, 100950. doi:https://doi.org/10.1016/j.trip.2023.100950

Chang, X., Wu, J., Liu, H., Yan, X., Sun, H., & Qu, Y. (2019). Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data. *Transportmetrica A: Transport Science, 15*(2), 1587-1612. doi:10.1080/23249935.2019.1620380

Chehreh Chelgani, S., Homafar, A., Nasiri, H., & Rezaei laksar, M. (2024). CatBoost-SHAP for modeling industrial operational flotation variables – A "conscious lab" approach. *Minerals Engineering 213*, 108754. doi:https://doi.org/10.1016/j.mineng.2024.108754

Chen, F., Hao, X., & Chen, Z. (2021). Can high-speed rail improve health and alleviate health inequality? Evidence from China. *Transport policy, 114*, 266-279. doi:https://doi.org/10.1016/j.tranpol.2021.10.007

Chen, H., & Cheng, Y. (2023). Travel mode choice prediction using imbalanced machine learning. *IEEE Transactions on Intelligent Transportation Systems, 24*(4), 3795-3808. doi:10.1109/TITS.2023.3237681

Chen, H., Sun, D., Zhu, Z., & Zeng, J. (2016). The impact of high-speed rail on residents' travel behavior and household mobility: A case study of the Beijing-Shanghai Line, China. *Sustainability, 8*(11), 1187. doi:10.3390/su8111187

Chen, J., & Li, S. (2017). Mode choice model for public transport with categorized latent variables. *Mathematical Problems in Engineering, 2017*.

Chen, P., Zhang, X., & Gao, D. (2024). Preference heterogeneity analysis on train choice behaviour of high-speed railway passengers: A case study in China. *Transportation Research Part A: Policy and Practice 188*, 104198. doi:https://doi.org/10.1016/j.tra.2024.104198

Chen, T., & Guestrin, C. (2016). *Xgboost: A scalable tree boosting system.* Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.

Cheng, L., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society 14*, 1-10.

Cheng, Y., Chen, X., De Vos, J., Lai, X., & Witlox, F. (2019). Applying random forest to analyze household car ownership decision: A comparison of performance and interpretation with multinomial logit model. *Transportation Research Part A: Policy and Practice, 125*, 89-101.

Chongzhi, W., Lin, W., & Zhang, W. (2021). Chapter 14 - Assessment of undrained shear strength using ensemble learning based on Bayesian hyperparameter optimization. In P. Samui, S. Kumari, V. Makarov, & P. Kurup (Eds.), *Modeling in Geotechnical Engineering* (pp. 309-326): Academic Press.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Cui, H., Li, M., Zhu, M., & Ma, X. (2023). Investigating the Impacts of Urban–Rural Bus Service Quality on Rural Residents' Travel Choices Using an SEM–MNL Integration Model. *Sustainability, 15*(15), 11950.

Da Silva, D., Klumpenhouwer, W., Karner, A., Robinson, M., Liu, R., & Shalaby, A. (2022). Living on a fare: Modeling and quantifying the effects of fare budgets on transit access and equity. *Journal of Transport Geography 101*, 103348. doi:https://doi.org/10.1016/j.jtrangeo.2022.103348

Daganzo, C. F., & Ouyang, Y. (2019). *Public transportation systems: Principles of system design, operations planning and real-time control*: World Scientific.

Dahmen, V., Weikl, S., & Bogenberger, K. (2023). Interpretable Machine Learning for Mode Choice Modeling on Tracking-Based Revealed Preference Data. *Transportation Research Record* 03611981241246973.

Dahmen, V., Weikl, S., & Bogenberger, K. (2024). Interpretable machine learning for mode choice modeling on tracking-based revealed preference data. *Transportation Research Record, 2678*, 2075-2091.

Dahmen, V., Weikl, S., & Bogenberger, K. (2024). Interpretable machine learning for mode choice modeling on tracking-based revealed preference data. *Transportation Research Record 2678*(11), 2075-2091.

Danapour, M., Nickkar, A., Jeihani, M., & Khaksar, H. (2018). Competition between high-speed rail and air transport in Iran: The case of Tehran–Isfahan. *Case Studies on Transport Policy, 6*(4), 456-461. doi:https://doi.org/10.1016/j.cstp.2018.05.006

Darwish, A. M., Almansour, M., Salah, A., Zagow, M., Saeed, K., & Elkafoury, A. (2024). Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: a case study of Alexandria City, Egypt. *Transportation Research Interdisciplinary Perspectives, 24*, 101052.

Deng, Y., Bai, Y., Cui, L., & He, R. (2023). Travel Mode Choice Behavior for High-Speed Railway Stations Based on Multi-Source Data. *Transportation Research Record, 2677*, 525-540.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Deng, Y., Bai, Y., Cui, L., & He, R. (2023). Travel Mode Choice Behavior for High-Speed Railway Stations Based on Multi-Source Data. *Transportation Research Record, 2677*(8), 525-540.

Department of International Trade Promotion. (2023). High-speed rail 4 routes. Retrieved from https://www.rentconnected.com/blogs/high-speed-rail-4-routes-for-thailand/

Díaz-Ramírez, J., Estrada-García, J. A., & Figueroa-Sayago, J. (2023). Predicting transport mode choice preferences in a university district with decision tree-based models. *City and Environment Interactions 20*, 100118. doi:https://doi.org/10.1016/j.cacint.2023.100118

Díaz-Ramírez, J., Estrada-García, N., & Figueroa-Sayago, M. (2023). Machine learning for mode choice prediction: A comparison of models in Mexico. *Transportation Research Part C, 148*, 103970-103970.

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice 110*, 107-117.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.

Dowling, R. (2015). Parents, children and automobility: Trends, challenges and opportunities. *Handbook on transport and development*, 526-538.

Drabicki, A., Cats, O., Kucharski, R., Fonzone, A., & Szarata, A. (2023). Should I stay or should I board? Willingness to wait with real-time crowding information in urban public transport. *Research in Transportation Business & Management, 47*, 100963. doi:https://doi.org/10.1016/j.rtbm.2023.100963

Economic Base. (2017). Knock on the Bangkok-Chiang Mai high-speed train fare 1,089 baht. Retrieved from https://www.thansettakij.com/business/242231

Elharoun, M., El-Badawy, S. M., & Shahdah, U. E. (2023). Artificial intelligence techniques for predicting individuals' mode choice behavior in Mansoura city, Egypt. *Transportation Research Record, 2677*(10), 605-623.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology 77*(4), 802-813.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Esfeh, M. A., Kattan, L., Lam, W. H., Esfe, R. A., & Salari, M. (2020). Compound generalized extreme value distribution for modeling the effects of monthly and seasonal variation on the extreme travel delays for vulnerability analysis of road network. *Transportation Research Part C: Emerging Technologies 120*, 102808.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10): Springer.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189-1232.

García-García, J. C., García-Ródenas, R., López-Gómez, J. A., & Martín-Baos, J. Á. (2022). A comparative study of machine learning, deep neural networks and random utility maximization models for travel mode choice modelling. *Transportation Research Procedia 62*, 374-382.

García-Ródenas, R., Linares, L. J., & López-Gómez, J. A. (2017). *On the performance of classic and deep neural models in image recognition.* Paper presented at the Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26.

Ghosh, T., & Nagaraj, N. (2024). Evaluating the Determinants of Mode Choice Using Statistical and Machine Learning Techniques in the Indian Megacity of Bengaluru. *arXiv preprint arXiv:2401.13977*.

Givoni, M. (2006). Development and Impact of the Modern High-speed Train: A Review. *Transport Reviews 26*(5), 593-611. doi:10.1080/01441640600589319

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in neural information processing systems 35*, 507-520.

Gross, S., & Grimm, B. (2018). Sustainable mode of transport choices at the destination–public transport at German destinations. *Tourism Review, 73*(3), 401-420.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Gunay, G., & Gokasar, I. (2021). Market segmentation analysis for airport access mode choice modeling with mixed logit. *Journal of Air Transport Management, 91*, 102001.

Guo, L., Huang, J., Ma, W., Sun, L., Zhou, L., Pan, J., & Yang, W. (2022). Convolutional Neural Network-Based Travel Mode Recognition Based on Multiple Smartphone Sensors. *Applied Sciences 12*(13), 6511. Retrieved from https://www.mdpi.com/2076-3417/12/13/6511

Guo, Y., Li, Y., Anastasopoulos, P. C., Peeta, S., & Lu, J. (2021). China's millennial car travelers' mode shift responses under congestion pricing and reward policies: A case study in Beijing. *Travel Behaviour and Society, 23*, 86-99.

Habib, M., Poole, K., & McCarthy, S. (2022). *NovaTRAC: A multi-year travel survey data collection program.* Paper presented at the the 12th International Conference on Transport Survey Methods.

He, Y., Zhao, Y., & Tsui, K. L. (2018). *An analysis of factors influencing metro station ridership: Insights from Taipei metro.* Paper presented at the 2018 21st International Conference on Intelligent Transportation Systems (ITSC).

Heinen, E., Harshfield, A., Panter, J., Mackett, R., & Ogilvie, D. (2017). Does exposure to new transport infrastructure result in modal shifts? Patterns of change in commute mode choices in a four-year quasi-experimental cohort study. *Journal of Transport & Health, 6*, 396-410. doi:https://doi.org/10.1016/j.jth.2017.07.009

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied choice analysis: a primer*: Cambridge university press.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis*: Cambridge university press.

Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y. (2021). A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of choice modelling 38*, 100221. doi:https://doi.org/10.1016/j.jocm.2020.100221

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., . . . Amira, A. (2022). Next-generation energy systems for sustainable smart cities: Roles of transfer learning. *Sustainable Cities and Society 85*, 104059. doi:https://doi.org/10.1016/j.scs.2022.104059

Hong, H., Naghibi, S. A., Pourghasemi, H. R., & Pradhan, B. (2016). GIS-based landslide spatial modeling in Ganzhou City, China. *Arabian Journal of Geosciences 9*, 1-26.

Hu, C., Huang, J., Gao, Y., & Lin, R. (2023). Can high-speed railway promote regional market integration? Evidence from China. *Research in Transportation Business & Management, 51*, 101057. doi:https://doi.org/10.1016/j.rtbm.2023.101057

Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., . . . Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology 574*, 1029-1041.

Huang, Y., Song, H., & Jia, C. Y. (2014). Analysis on the main factors influencing the high-speed railway passenger flow. *Advanced Materials Research 919*, 1085-1090.

Ibraeva, A., & Sousa, J. F. d. (2014). Marketing of Public Transport and Public Transport Information Provision. *Procedia - Social and Behavioral Sciences, 162*, 121-128. doi:https://doi.org/10.1016/j.sbspro.2014.12.192

Jara-Díaz, S., Gschwender, A., & Hörcher, D. (2023). Public transport: design, scale, and pricing. In *Handbook on Transport Pricing and Financing* (pp. 171-189): Edward Elgar Publishing.

Jiang, C., & Wang, C. (2021). High-speed rail pricing: Implications for social welfare. *Transportation Research Part E: Logistics and Transportation Review, 155*, 102484.

Jiang, Y., Yu, S., Guan, W., Gao, S., & Feng, T. (2021). Ground access behaviour of air-rail passengers: A case study of Dalian ARIS. *Travel Behaviour and Society, 24*, 152-163.

Jin, M., Lin, K.-C., Shi, W., Lee, P. T., & Li, K. X. (2020). Impacts of high-speed railways on economic growth and disparity in China. *Transportation Research Part A: Policy and Practice, 138*, 158-171.

JTTRI-AIRO. (2023). Progress of the Thailand-China High-Speed Railway Retrieved from https://www.jttri-airo.org/en/dll.php?id=20&s=pdf1&t=repo

มหาวิทยาลัยเทคโนโลยีสุรนารี

Juan de Dios Ortúzar, & Willumsen, L. G. (2011). Trip Generation Modelling. In *Modelling Transport* (pp. 139-173).

Kalantari, H. A., Sabouri, S., Brewer, S., Ewing, R., & Tian, G. (2025). Machine Learning in Mode Choice Prediction as Part of MPOs' Regional Travel Demand Models: Is It Time for Change? *Sustainability 17*(8), 3580.

Karmarkar, O. D., Jana, A., & Velaga, N. R. (2023). Willingness to pay for reduced travel time: Case study of Mumbai - Ahmedabad High Speed Railway. *Case Studies on Transport Policy, 12*, 100974. doi:https://doi.org/10.1016/j.cstp.2023.100974

Kashifi, M. T., Jamal, A., Kashefi, M. S., Almoshaogeh, M., & Rahman, S. M. (2022). Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behaviour and Society, 29*, 279-296.

Kashifi, S. A., Jamal, S., Kashefi, M., Almoshaogeh, M., & Rahman, M. A. (2022). Evaluating high-speed rail impact on modal shift and environmental benefits. *Environmental Science and Pollution Research, 29*, 40926-40940.

Kazemi, F., Asgarkhani, N., Shafighfard, T., Jankowski, R., & Yoo, D.-Y. (2024). Machine-Learning Methods for Estimating Performance of Structural Concrete Members Reinforced with Fiber-Reinforced Polymers. *Archives of Computational Methods in Engineering* 1-33.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems, 30*.

Khan, S., Adnan, A., & Iqbal, N. (2022). *Applications of artificial intelligence in transportation.* Paper presented at the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET).

Kim, E.-J. (2021). Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation 2021*(1), 6685004.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics, 6*, 1-15.

Kujala, R., Weckström, C., Mladenović, M. N., & Saramäki, J. (2018). Travel times and transfers in public transport: Comprehensive accessibility analysis based on Pareto-optimal journeys. *Computers, Environment and Urban Systems, 67*, 41-54. doi:https://doi.org/10.1016/j.compenvurbsys.2017.08.012

Lachapelle, U., & Boisjoly, G. (2023). Breaking down public transit travel time for more accurate transport equity policies: A trip component approach. *Transportation Research Part A: Policy and Practice 175*, 103756. doi:https://doi.org/10.1016/j.tra.2023.103756

Lackner, M., & Zulehner, C. (2020). Rent sharing and gender discrimination: causal evidence from collegiate athletics. *Social Science Quarterly, 101*(2), 678-696.

Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications 10*(1), 1096.

Larranaga, A. M., Arellana, J., & Senna, L. A. (2017). Encouraging intermodality: A stated preference analysis of freight mode choice in Rio Grande do Sul. *Transportation Research Part A: Policy and Practice 102*, 202-211.

Lee, E., Kawakita, T., Huai, Y., Lo, H. K., & Zhang, A. (2024). Airline and high-speed rail collaboration and competition under travel time variability. *Transportation Research Part A: Policy and Practice 185*, 104104. doi:https://doi.org/10.1016/j.tra.2024.104104

Lhéritier, A., Bocamazo, M., Delahaye, T., & Acuna-Agost, R. (2019). Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling, 31*, 198-209.

Li, J., Lo, K., & Guo, M. (2018). What influences the choice between private car and public transport for ShoppingTrips? Impact of socio-economic and built environment factors. *Journal of Asian Energy Studies, 2*(1), 28-42.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Li, X., Ma, Q., Wang, W., & Wang, B. (2021). Influence of weather conditions on the intercity travel mode choice: A case of Xi'an. *Computational Intelligence and Neuroscience, 2021*(1), 9969322.

Li, X., Shi, L., Shi, Y., Tang, J., Zhao, P., Wang, Y., & Chen, J. (2024). Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Applied Geography 166*, 103264.

Li, X., Tang, J., Hu, X., & Wang, W. (2020). Assessing intercity multimodal choice behavior in a touristy city: a factor analysis. *Journal of Transport Geography, 86*, 102776.

Li, X., Xing, G., Qian, X., Guo, Y., Wang, W., & Cheng, C. (2022). Subway station accessibility and its impacts on the spatial and temporal variations of its outbound ridership. *Journal of Transportation Engineering, Part A: Systems, 148*(12), 04022106.

Liang, Y., Zhou, K., Li, X., Zhou, Z., Sun, W., & Zeng, J. (2020). Effectiveness of high-speed railway on regional economic growth for less developed areas. *Journal of Transport Geography 82*, 102621.

Liu, L., & Zhang, M. (2018). High-speed rail impacts on travel times, accessibility, and economic productivity: A benchmarking analysis in city-cluster regions of China. *Journal of Transport Geography, 73*, 25-40.

Liu, S., & Putro, U. S. (2024). Passenger Service Satisfaction Evaluation of Jakarta-Bandung High-Speed Railway. *European Journal of Business and Management Research, 9*(4), 115-126.

Liu, S., Yao, E., Cheng, X., & Zhang, Y. (2017). Evaluating the impact of new lines on entrance/exit passenger flow of adjacent existing stations in urban rail transit system. *Transportation Research Procedia, 25*, 2625-2638. doi:https://doi.org/10.1016/j.trpro.2017.05.316

Liu, Y., Wang, S., & Xie, B. (2019). Evaluating the effects of public transport fare policy change together with built and non-built environment features on ridership: The case in South East Queensland, Australia. *Transport Policy 76*, 78-89. doi:https://doi.org/10.1016/j.tranpol.2019.02.004

Lu, Y., Prato, C. G., Sipe, N., & Corcoran, J. (2023). Understanding the role of household modality style on first and last mile travel mode choice and public transit station choice. *Travel Behaviour and Society 32*, 100580. doi:https://doi.org/10.1016/j.tbs.2023.100580

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., . . . Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence, 2*(1), 56-67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*.

Ma, W., Wang, Q., Yang, H., Zhang, A., & Zhang, Y. (2019). Effects of Beijing-Shanghai high-speed rail on air travel: Passenger types, airline groups and tacit collusion. *Research in Transportation Economics, 74*, 64-76.

Mangalathu, S., Hwang, S.-H., & Jeon, J.-S. (2020). Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Engineering Structures, 219*, 110927.

Mao, H., Deng, X., Jiang, H., Shi, L., Li, H., Tuo, L., . . . Guo, F. (2021). Driving safety assessment for ride-hailing drivers. *Accident Analysis & Prevention, 149*, 105574.

Martín, J. C., Román, C., García-Palomares, J. C., & Gutiérrez, J. (2014). Spatial analysis of the competitiveness of the high-speed train and air transport: The role of access to terminals in the Madrid–Barcelona corridor. *Transportation Research Part A: Policy and Practice, 69*, 392-408.

Meesit, R. (2024). Factors affecting the future utilisation of high-speed train services in Eastern Economic Corridor (EEC). *Transportation Research Interdisciplinary Perspectives 26*, 101162. doi:https://doi.org/10.1016/j.trip.2024.101162

Mirzaei, E., Kheyroddin, R., & Mignot, D. (2021). Exploring the effect of the built environment, weather condition and departure time of travel on mode choice decision for different travel purposes: evidence from Isfahan, Iran. *Case Studies on Transport Policy, 9*(4), 1419-1430.

Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks. *International journal of transportation science and technology, 9*(2), 100-115.

Nam, D., & Cho, J. (2020). Deep neural network design for modeling individual-level travel mode choice behavior. *Sustainability, 12*(18), 7481.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics, 7*, 21.

NESDC. (2022). Gross regional and provincial product chain volume measure 2020 edition. In: Office of the National Economic and Social Development Council (NESDC) Bangkok.

Ning, J., Lyu, T., & Wang, Y. (2021). Exploring the built environment factors in the metro that influence the ridership and the market share of the elderly and students. *Journal of Advanced Transportation 2021*(1), 9966794.

Nurhidayat, A. Y., Widyastuti, H., Sutikno, & Upahita, D. P. (2023). Research on passengers' preferences and impact of high-speed rail on air transport demand. *Sustainability 15*(4), 3060.

Office of, T., Traffic, P., & Planning. (2024). *Proportion of Public Transportation Usage for Intercity Travel (2017-2022)*. Retrieved from https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45

Office of the National, E., & Social Development, C. (2025). Sustainable Development Goals: SDGs. Retrieved from https://sdgs.nesdc.go.th/goals-and-indicators/

Office of the national economic and social development council. (2023). Gross Regional and Provincial Product Chain Volume Measure 2022 Edition. Retrieved from https://www.nesdc.go.th/ewt_dl_link.php?nid=15104&filename=gross_regional

Office of the National Economic and Social Development Council. (2025). Sustainable Development Goals: SDGs. Retrieved from https://sdgs.nesdc.go.th/goals-and-indicators/

Office of Transport and Traffic Policy and Planning. (2024). Proportion of Public Transportation Usage for Intercity Travel (2017-2022). Retrieved from

มหาวิทยาลัยเทคโนโลยีสุรนารี

https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45. Retrieved 11 August 2024 https://otp.gdcatalog.go.th/dataset/otp_65_03/resource/64e110f9-9cc7-44b0-810d-38b03544eb45

Orlando, V. M., Baquela, E. G., Bhouri, N., & Lotito, P. A. (2023). Public transport demand estimation by frequency adjustments. *Transportation Research Interdisciplinary Perspectives 19*, 100832. doi:https://doi.org/10.1016/j.trip.2023.100832

Osawa, M., Akamatsu, T., & Takayama, Y. (2017). Harris And Wilson (1978) Model Revisited: The Spatial Period-Doubling Cascade In An Urban Retail Model. *Journal of Regional Science, 57*(3), 442-466.

Otim, T., Dörfer, L., Ahmed, D. B., & Munoz Diaz, E. (2022). Modeling the impact of weather and context data on transport mode choices: A case study of gps trajectories from beijing. *Sustainability, 14*(10), 6042.

Pan, J. Y. (2024). High-Speed Rail in the US—Mode Choice Decision and Impact of COVID-19. *Sustainability 16*(10), 4041.

Pasha, M. M., Hickman, M. D., & Prato, C. G. (2020). Modeling Mode Choice of Air Passengers' Ground Access to Brisbane Airport. *Transportation Research Record, 2674*(11), 756-767.

Pavlou, M., Ambler, G., Qu, C., Seaman, S. R., White, I. R., & Omar, R. Z. (2024). An evaluation of sample size requirements for developing risk prediction models with binary outcomes. *BMC Medical Research Methodology 24*(1), 146.

Pei, Y. L., Zhou, K., & Peng, T. (2013). Prediction model of passenger waiting time in high-speed rail hub based on BP neural network. *Applied Mechanics and Materials, 321*, 1903-1906.

Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., . . . Takeuchi, W. (2020). Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam. *Remote Sensing, 12*(8), 1334.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Pitombo, C. S., Salgueiro, A. R., da Costa, A. S. G., & Isler, C. A. (2015). A two-step method for mode choice estimation with socioeconomic and spatial information. *Spatial Statistics, 11*, 45-64.

POST TO DAY. (2017). Transportation" expects the ticket price of the Bangkok-Chiang Mai high-speed train to run for 3 hours and a half 1,200 baht,. Retrieved from https://www.posttoday.com/business/531218

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems 31*.

Puan, O., Hassan, Y., Mashros, N., Idham, M., Hassan, N., Warid, M., & Hainin, M. (2019). *Transportation mode choice binary logit model: A case study for Johor Bahru city.* Paper presented at the IOP conference series: Materials science and engineering.

Rattanakijsuntorn, W., Suwannarat, B., Samittivate, N., & Nithikittiwat, C. (2024). Success Factors and Partnership Evaluation of Air–Rail Integration Development: A Case of a High-Speed Rail Project Linking Three Airports in Thailand. *Infrastructures, 9*(7), 115.

Raturi, V., & Verma, A. (2020). A game-theoretic approach to analyse inter-modal competition between high-speed rail and airlines in the Indian context. *Transportation Planning and Technology 43*(1), 20-47.

Ren, X., Chen, Z., Wang, F., Dan, T., Wang, W., Guo, X., & Liu, C. (2020). Impact of high-speed rail on social equity in China: Evidence from a mode choice survey. *Transportation Research Part A: Policy and Practice, 138*, 422-441. doi:https://doi.org/10.1016/j.tra.2020.05.018

Ristić, B., Stojadinović, N., & Trifunović, D. (2022). Conditions for effective on-track competition in the European passenger railway market: A yardstick for regulations. *Transport Policy 119*, 1-15. doi:https://doi.org/10.1016/j.tranpol.2022.02.006

Román, C., Martín, J. C., Espino, R., Cherchi, E., de Dios Ortúzar, J., Rizzi, L. I., . . . Amador, F. J. (2014). Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor. *Transport policy, 36*, 105-117.

Rubin, O., Mulder, C. H., & Bertolini, L. (2014). The determinants of mode choice for family visits – evidence from Dutch panel data. *Journal of Transport Geography, 38*, 137-147. doi:https://doi.org/10.1016/j.jtrangeo.2014.06.004

Salas, P., De la Fuente, R., Astroza, S., & Carrasco, J. A. (2022). A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Systems with Applications 193*, 116253. doi:https://doi.org/10.1016/j.eswa.2021.116253

Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science, 3*(2), 158.

Shafighfard, T., Kazemi, F., Asgarkhani, N., & Yoo, D.-Y. (2024). Machine-learning methods for estimating compressive strength of high-performance alkali-activated concrete. *Engineering Applications of Artificial Intelligence, 136*, 109053. doi:https://doi.org/10.1016/j.engappai.2024.109053

Shafighfard, T., Kazemi, F., Bagherzadeh, F., Mieloszyk, M., & Yoo, D. Y. (2024). Chained machine learning model for predicting load capacity and ductility of steel fiber–reinforced concrete beams. *Computer‐Aided Civil and Infrastructure Engineering*.

Shahdah, U. E., Elharoun, M., Ali, E. K., Elbany, M., & Elagamy, S. R. (2025). Stated preference survey for predicting eco-friendly transportation choices among Mansoura University students. *Innovative Infrastructure Solutions 10*(5), 1-23.

Shakya, A., Biswas, M., & Pal, M. (2022). Chapter 9 - Classification of Radar data using Bayesian optimized two-dimensional Convolutional Neural Network. In P. K. Srivastava, D. K. Gupta, T. Islam, D. Han, & R. Prasad (Eds.), *Radar Remote Sensing* (pp. 175-186): Elsevier.

Sharma, A., Gani, A., Asirvatham, D., Ahmed, R., Hamzah, M., & Asli, M. F. (2021). Travel Behavior Modeling: Taxonomy, Challenges, and Opportunities. *International Journal of Advanced Computer Science and Applications (IJACSA), 12*(5), 774-786.

Shelat, S., Cats, O., & van Cranenburgh, S. (2022). Traveller behaviour in public transport in the early stages of the COVID-19 pandemic in the Netherlands. *Transportation Research Part A: Policy and Practice 159*, 357-371.

Shelat, S., Cats, O., & van Lint, J. W. C. (2021). Quantifying travellers' evaluation of waiting time uncertainty in public transport networks. *Travel Behaviour and Society, 25*, 209-222. doi:https://doi.org/10.1016/j.tbs.2021.07.009

Shen, J., Sakata, Y., & Hashimoto, Y. (2006). A comparison between latent class model and mixed logit model for transport mode choice: evidences from two datasets of Japan. *Discussion papers in economics and business*, 06-05.

Shi, R., Hu, Z., Zhou, Y., & Liu, P. (2014). Research on Railway Freight Market Share Based on the Whole Process of Transport. *Procedia - Social and Behavioral Sciences, 138*, 298-304. doi:https://doi.org/10.1016/j.sbspro.2014.07.207

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion 81*, 84-90. doi:https://doi.org/10.1016/j.inffus.2021.11.011

So, B. (2024). Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM. *Scandinavian Actuarial Journal*, 1-23.

Soczówka, P., Kłos, M. J., Żochowska, R., & Sobota, A. (2021). An analysis of the influence of travel time on access time in public transport. *Zeszyty Naukowe. Transport/Politechnika Śląska*(111), 137-149.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management 45*(4), 427-437.

Sresakoolchai, J., & Kaewunruen, S. (2020). Comparative studies into public private partnership and traditional investment approaches on the high-speed rail project linking 3 airports in Thailand. *Transportation Research Interdisciplinary Perspectives, 5*, 100116.

Srithongrung, A., & Kriz, K. A. (2019). Thai Public Capital Budget and Management Process. In *Capital Management and Budgeting in the Public Sector* (pp. 206-235): IGI Global.

มหาวิทยาลัยเทคโนโลยีสุรนารี

State Railway of Thailand. (2019). *COMPLETE GUIDE TO SRT. 4.0*. Retrieved from https://www.railway.co.th/RailwayMiddleFile/NewsAndActImg/1307/131951934207171332_2_2562.pdf

State Railway of Thailand. (2020). รถไฟความเร็วสูง ช่วง กทม. - หนองคาย (ระยะที่ 2 นครราชสีมา - หนองคาย),. Retrieved from https://www.hsrkorat-nongkhai.com/

State Railway of Thailand (2024). The State Railway of Thailand Easy book, Convenient check. Retrieved from https://dticket.railway.co.th/DTicketPublicWeb/home/Home

Sun, W., Wang, C., Liu, C., & Wang, L. (2021). High-speed rail network expansion and its impact on regional economic sustainability in the Yangtze River Delta, China, 2009–2018. *Sustainability, 14*(1), 155.

Sun, Y., Wang, X., Liu, L., & Wang, Y. (2021). High-speed rail and urban development in China: Evidence from multiple indicators. *Journal of Urban Planning and Development, 147*(4), 05021047-05021047.

Ta, N., Wang, X., Hu, L., & Liu, Z. (2022). Gender difference in commuting travel: A comparative study of suburban residents in Beijing and Shanghai. *Travel Behaviour and Society, 28*, 196-203.

Tang, L., Tang, C., Fu, Q., & Ma, C. (2024). Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis. *IET Intelligent Transport Systems*

Tanwar, R., & Kumar Agarwal, P. (2024). Investigating the factors determining service quality in Bhopal's multimodal public transport system: A factor analysis. *Case Studies on Transport Policy* 101317. doi:https://doi.org/10.1016/j.cstp.2024.101317

Tarriño-Ortiz, J., Gómez, J., Soria-Lara, J. A., & Vassallo, J. M. (2022). Analyzing the impact of Low Emission Zones on modal shift. *Sustainable Cities and Society, 77*, 103562.

Tiong, K. Y., Ma, Z., & Palmqvist, C.-W. (2023). Analyzing factors contributing to real-time train arrival delays using seemingly unrelated regression models. *Transportation Research Part A: Policy and Practice 174*, 103751. doi:https://doi.org/10.1016/j.tra.2023.103751

มหาวิทยาลัยเทคโนโลยีสุรนารี

Tirachini, A., Godachevich, J., Cats, O., Muñoz, J. C., & Soza-Parra, J. (2022). Headway variability in public transport: A review of metrics, determinants, effects for quality of service and control strategies. *Transport Reviews 42*(3), 337-361.

Tissayakorn, K. (2021). A study on transit-oriented development strategies in the high-speed rail project in Thailand.

Tissayakorn, P. (2021). Thailand's high-speed rail progress and regional integration. *ASEAN Transportation Review, 14*(2), 55-68.

Tomeš, Z., & Jandová, M. (2018). Open access passenger rail services in Central Europe. *Research in Transportation Economics, 72*, 74-81. doi:https://doi.org/10.1016/j.retrec.2018.10.002

Török, Á., Szalay, Z., Uti, G., & Verebélyi, B. (2020). Rerepresenting autonomated vehicles in a macroscopic transportation model. *Periodica Polytechnica Transportation Engineering 48*(3), 269-275.

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.), 6*(2), 227.

Vega, G. M., & Aznarte José, L. (2020). Shapley additive explanations for NO2 forecasting. *Ecological Informatics 56*, 101039.

Vigren, A. (2017). Competition in Swedish passenger railway: Entry in an open access market and its effect on prices. *Economics of Transportation, 11-12*, 49-59. doi:https://doi.org/10.1016/j.ecotra.2017.10.005

Wang, D., & Zhou, M. (2017). The built environment and travel behavior in urban China: A literature review. *Transportation Research Part D: Transport and Environment 52*, 574-585.

Wang, F., & Ross, C. (2018). Machine learning travel mode choice modeling: A comparison with traditional logit models. *Transportation Research Record, 2672*(47), 1-11.

Wang, F., & Ross, C. L. (2018). Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record 2672*(47), 35-45.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Wang, H., & Huang, H.-J. (2024). Effects of high-speed rail on intercity travels, utility and social welfare in urban agglomerations: A game-theoretic perspective. *Transportation Research Part E: Logistics and Transportation Review, 192*, 103800. doi:https://doi.org/10.1016/j.tre.2024.103800

Wang, J., Huang, J., & Jing, Y. (2020). Competition between high-speed trains and air travel in China: From a spatial to spatiotemporal perspective. *Transportation Research Part A: Policy and Practice, 133*, 62-78.

Wang, J., Zhao, W., Liu, C., & Huang, Z. (2023). A System Optimization Approach for Trains' Operation Plan with a Time Flexible Pricing Strategy for High-Speed Rail Corridors. *Sustainability 15*(12), 9556.

Wang, Y., Li, L., Wang, L., Moore, A., Staley, S., & Li, Z. (2014). Modeling traveler mode choice behavior of a new high-speed rail corridor in China. *Transportation Planning and Technology, 37*(5), 466-483.

Wang, Y., Sun, X., Mark-up, I. S., & D, I. (2018). A Reexamination of the Relationship between Market Power and Technological Innovation. *RD Res. Manag, 39*, 144-152.

Weerawat, W., Samitiwantikul, L., & Torpanya, R. (2020). Operational challenges of the Bangkok airport rail link. *Urban rail transit, 6*, 42-55.

Wei, W., Wang, F., Li, T., & Li, F. (2024). Comprehensive impacts of high-speed rail and air transport on tourism development in China. *Transportation Research Part A: Policy and Practice 190*, 104263. doi:https://doi.org/10.1016/j.tra.2024.104263

Wen, X., & Chen, X. (2025). A New Breakthrough in Travel Behavior Modeling Using Deep Learning: A High-Accuracy Prediction Method Based on a CNN. *Sustainability 17*(2), 738. Retrieved from https://www.mdpi.com/2071-1050/17/2/738

Wójcik, S. (2019). The determinants of travel mode choice: the case of ŁódŹ, Poland. *Bulletin of Geography. Socio-economic Series*(44), 93-101.

Wu, D., & Martín, J. C. (2022). Research on Passengers' Preference for High-Speed Railways (HSRs) and High-Speed Trains (HSTs). *Sustainability, 14*(3). doi:10.3390/su14031473

มหาวิทยาลัยเทคโนโลยีสุรนารี

Wu, L., Zhang, M., & Wang, Y. (2015). Analysis of Ninghang intercity travel mode choice behavior based on panel data. *Journal of Transportation Systems Engineering and Information Technology, 15*(1), 226-231.

Wu, M.-W., & Shen, C.-H. (2011). The elusive effect of bank size on profits. *The Service Industries Journal, 31*(11), 1703-1724.

Wu, S., & Han, D. (2022). Accessibility of high-speed rail (HSR) stations and HSR–air competition: Evidence from China. *Transportation Research Part A: Policy and Practice 166*, 262-284. doi:https://doi.org/10.1016/j.tra.2022.10.015

Wu, W., Liang, Y., & Wu, D. (2016). Evaluating the impact of China's rail network expansions on local accessibility: A market potential approach. *Sustainability 8*(6), 512.

Xia, W., Jiang, C., Wang, K., & Zhang, A. (2019). Air-rail revenue sharing in a multi-airport system: Effects on traffic and social welfare. *Transportation research part B: methodological, 121*, 304-319.

Xie, J., Zhan, S., Wong, S., Wen, K., Qiang, L., & Lo, S. (2022). High-speed rail services for elderly passengers: Ticket-booking patterns and policy implications. *Transport policy, 125*, 96-106.

Xu, F., Liu, Q., Zheng, X., Cao, L., & Yang, M. (2022). Research on the impact of China's high-speed rail opening on enterprise market power: Based on the perspective of market segmentation. *Transport Policy, 128*, 121-137. doi:https://doi.org/10.1016/j.tranpol.2022.09.019

Xu, M., Shuai, B., Wang, X., Liu, H., & Zhou, H. (2023). Analysis of the accessibility of connecting transport at High-speed rail stations from the perspective of departing passengers. *Transportation Research Part A: Policy and Practice 173*, 103714. doi:https://doi.org/10.1016/j.tra.2023.103714

Xu, Y., Zhao, X., Chen, Y., & Yang, Z. (2019). Research on a mixed gas classification algorithm based on extreme random tree. *Applied Sciences, 9*(9), 1728.

Yang, W., Chen, Q., & Yang, J. (2022). Factors Affecting Travel Mode Choice between High-Speed Railway and Road Passenger Transport—Evidence from China. *Sustainability 14*(23), 15745. doi: https://doi.org/10.3390/su142315745

Yin, C., Wu, J., Sun, X., Meng, Z., & Lee, C. (2024). Road transportation emission prediction and policy formulation: Machine learning model analysis. *Transportation Research Part D: Transport and Environment 135*, 104390. doi:https://doi.org/10.1016/j.trd.2024.104390

Yu, J., Chang, X., Hu, S., Yin, H., & Wu, J. (2024). Combining travel behavior in metro passenger flow prediction: A smart explainable Stacking-Catboost algorithm. *Information Processing & Management 61*(4), 103733. doi:https://doi.org/10.1016/j.ipm.2024.103733

Yu, X., Lang, M., Gao, Y., Wang, K., Su, C.-H., Tsai, S.-B., . . . Li, S. (2018). An empirical study on the design of China high-speed rail express train operation plan—From a sustainable transport perspective. *Sustainability 10*(7), 2478.

Yuan, Y., Yang, M., Feng, T., Rasouli, S., Ruan, X., Wang, X., & Li, Y. (2021). Analyzing heterogeneity in passenger satisfaction, loyalty, and complaints with air-rail integrated services. *Transportation Research Part D: Transport and Environment 97*, 102950.

Yuniaristanto, Sutopo, W., Hisjam, M., & Wicaksono, H. (2024). Estimating the market share of electric motorcycles: A system dynamics approach with the policy mix and sustainable life cycle costs. *Energy Policy, 195*, 114345. doi:https://doi.org/10.1016/j.enpol.2024.114345

Zaidan, E., Abulibdeh, A., Alban, A., & Jabbar, R. (2022). Motivation, preference, socioeconomic, and building features: New paradigm of analyzing electricity consumption in residential buildings. *Building and Environment, 219*, 109177.

Zang, Z., Xu, X., Qu, K., Chen, R., & Chen, A. (2022). Travel time reliability in transportation networks: A review of methodological developments. *Transportation Research Part C: Emerging Technologies 143*, 103866. doi:https://doi.org/10.1016/j.trc.2022.103866

Zhai, W., Li, C., Fei, S., Liu, Y., Ding, F., Cheng, Q., & Chen, Z. (2023). CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values. *Computers and Electronics in Agriculture 214*, 108306. doi:https://doi.org/10.1016/j.compag.2023.108306

มหาวิทยาลัยเทคโนโลยีสุรนารี

Zhang, Q., Yang, H., Wang, Q., Zhang, A., & Zhang, Y. (2020). Impact of high-speed rail on market concentration and Lerner index in China's airline market. *Journal of Air Transport Management 83*, 101755. doi:https://doi.org/10.1016/j.jairtraman.2019.101755

Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies, 58*, 308-324. doi:https://doi.org/10.1016/j.trc.2015.02.019

Zhang, Y., Liang, K., Yao, E., & Gu, M. (2024). Measuring Reliable Accessibility to High-Speed Railway Stations by Integrating the Utility-Based Model and Multimodal Space–Time Prism under Travel Time Uncertainty. *ISPRS International Journal of Geo-Information, 13*, 263.

Zhang, Y., Liang, K., Yao, E., & Gu, M. (2024). Measuring Reliable Accessibility to High-Speed Railway Stations by Integrating the Utility-Based Model and Multimodal Space–Time Prism under Travel Time Uncertainty. *ISPRS International Journal of Geo-Information 13*(8), 263.

Zhao, P., & Yuan, D. (2023). Effects of Family Structure on Travel Behaviour. In *Population Growth and Sustainable Transport in China* (pp. 233-266): Springer.

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2018). Modeling Stated preference for mobility-on-demand transit: a comparison of Machine Learning and logit models. *arXiv preprint arXiv:1811.01315*

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel behaviour and society 20*, 22-35.

Zhou, Y., Zhao, M., Tang, S., Lam, W. H., Chen, A., Sze, N., & Chen, Y. (2020). Assessing the relationship between access travel time estimation and the accessibility to high speed railway station by different travel modes. *Sustainability 12*(18), 7827.

Zhu, L., Shu, S., & Zou, L. (2022). XGBoost-Based Travel Time Prediction between Bus Stations and Analysis of Influencing Factors. *Wireless Communications and Mobile Computing, 2022*(1), 3504704.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Zhu, X., Shen, X., Chen, K., & Zhang, Z. (2024). Research on the prediction and influencing factors of heavy duty truck fuel consumption based on LightGBM. *Energy 296*, 131221. doi:https://doi.org/10.1016/j.energy.2024.131221

Zouzou, Y., & Citakoglu, H. (2023). General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. *Acta Geophysica 71*(2), 927-947.

มหาวิทยาลัยเทคโนโลยีสุรนารี

CHAPTER V

CONCLUSION AND RECOMMENDATION

5.1 Conclusions

This dissertation set out to investigate the transformative impact of high-speed rail (HSR) development on intercity travel behavior in Thailand through a series of three integrated research articles. Drawing on stated preference data and advanced modeling techniques, the studies aimed to enhance understanding of modal shifts in response to HSR implementation. Each paper offered distinct yet complementary insights into traveler preferences, market share dynamics, and predictive modeling accuracy. Collectively, the findings underscore the limitations of traditional econometric models and highlight the utility of machine learning in transportation planning. The concluding chapter synthesizes the key contributions of all three studies.

5.1.1 A Machine Learning Comparison of Transportation Mode Changes from High-Speed Railway Promotion in Thailand

Study 1: This study aimed to investigate travel mode choice behavior in the context of Thailand's high-speed rail (HSR) development using advanced predictive analytics. By applying machine learning techniques—particularly the XGBoost algorithm—and comparing them with traditional models such as the Multinomial Logit (MNL), the research demonstrated that machine learning offers superior predictive performance in terms of both accuracy and flexibility.

Moreover, the integration of SHAP (SHapley Additive Explanations) enabled interpretable insights into the underlying decision factors, helping to identify which variables most strongly influence individuals' willingness to adopt HSR. Key determinants included travel cost, travel time, income, and frequency of current intercity travel.

The study's findings contribute to the growing body of literature on transport behavior modeling, particularly in the context of emerging economies where

มหาวิทยาลัยเทคโนโลยีสุรนารี

HSR adoption is still in the planning phase. Practically, the results offer evidence-based recommendations for policymakers to develop strategies that support effective and sustainable HSR implementation in Thailand. These may include fare structuring, targeted marketing campaigns, and investment in multimodal connectivity.

In conclusion, this research reinforces the value of machine learning as a robust and interpretable tool for transportation behavior analysis and highlights its practical relevance for infrastructure planning in developing countries.

5.1.2 Analyzing High-Speed Rail's Transformative Impact on Public Transport in Thailand Using Machine Learning

Study 2: This study examined the potential impact of high-speed rail (HSR) on intercity travel behavior in Thailand by applying advanced machine learning techniques, particularly the CatBoost algorithm, in conjunction with Shapley Additive Explanations (SHAP) for model interpretability. Based on data collected from 3,200 passengers across 16 provinces, the results reveal that HSR is likely to significantly alter the landscape of intercity travel, with notable modal shifts anticipated from traditional transport modes—especially buses and conventional trains—to HSR.

The machine learning model outperformed the traditional multinomial logit (MNL) model in terms of predictive accuracy and its ability to capture non-linear, high-dimensional relationships in travel behavior data. Key influencing factors identified include travel cost, total travel time, waiting time, and access time, all of which substantially affect the likelihood of choosing HSR.

Moreover, the findings suggest that HSR adoption will be most prominent among middle-income travelers and those who currently rely on buses or short-haul flights. These insights offer valuable implications for transport policy design, particularly in areas of pricing strategy, service frequency, and last-mile accessibility. By using empirical, data-driven modeling approaches, this research contributes to filling a critical knowledge gap in the Thai context and supports the formulation of sustainable, efficient, and user-centered public transport development strategies in the era of HSR.

มหาวิทยาลัยเทคโนโลยีสุรนารี

5.1.3 Machine Learning-Based Analysis of Travel Mode Preferences: Neural and Boosting Model Comparison Using Stated Preference Data from Thailand's Emerging High-Speed Rail Network

Study 3: This study aimed to compare the predictive performance of traditional econometric models—specifically the Multinomial Logit (MNL) model—with advanced machine learning techniques, including CatBoost, XGBoost, LightGBM, Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN), in forecasting intercity travel mode choice behavior in the context of Thailand's high-speed rail (HSR) development. The analysis was based on stated preference (SP) data collected from 3,200 respondents across the country.

The results revealed that CatBoost outperformed all other models, achieving the highest predictive performance with an AUC of 0.9113 and an accuracy of 0.7557. In contrast, the MNL model—though interpretable—exhibited significantly lower predictive accuracy and struggled to capture nonlinear relationships.

SHAP analysis indicated that travel cost, service frequency, and waiting time were the most influential factors in mode choice, surpassing the influence of sociodemographic variables such as income and gender. Notably, individuals with higher incomes, frequent travel patterns, or larger household sizes were more likely to choose HSR.

These findings highlight the potential of machine learning techniques in uncovering complex behavioral patterns and demonstrate their practical value in supporting data-driven, sustainable transport planning—particularly in the areas of fare policy, service frequency design, and station accessibility for Thailand's future HSR system.

มหาวิทยาลัยเทคโนโลยีสุรนารี

5.2 Recommendations

Based on the findings from the three integrated research articles presented in this dissertation, it is evident that artificial intelligence techniques—particularly machine learning and deep learning algorithms—hold substantial promise for modeling and forecasting travel mode choices in the Thai context. These insights provide a foundation for formulating policy recommendations aimed at infrastructure planning and the design of responsive public transport services. Moreover, the research highlights practical considerations in data collection, survey design, and model application that can enhance future studies. This section offers both policy-oriented and academic recommendations aligned with the empirical results and suggests directions for further research to support sustainable and data-driven mobility development in Thailand.

5.2.1 Recommendations for A Machine Learning Comparison of Transportation Mode Changes from High-Speed Railway Promotion in Thailand

Based on the research findings, which demonstrate the effectiveness of machine learning techniques—particularly XGBoost—in predicting travel mode choice and identifying influential decision factors, the following recommendations are proposed:

1) Policy Recommendations

(1) Develop a competitive pricing structure for high-speed rail (HSR) fares to ensure affordability and encourage mode shift from private vehicles and air transport.

(2) Promote multimodal integration by improving connections between HSR and local transit systems, such as metros, buses, and vans, to support seamless “door-to-door” travel.

(3) Launch public awareness campaigns that emphasize the benefits of HSR in terms of speed, safety, environmental sustainability, and long-term travel cost savings to increase early public acceptance.

2) Academic and Research Recommendations

(1) Conduct replication studies in other HSR corridors or regions to validate the generalizability of the findings and compare

มหาวิทยาลัยเทคโนโลยีสุรนารี

behavioral patterns across different geographic and demographic groups.

(2) Explore hybrid modeling approaches that integrate machine learning with behavioral theories, such as the Theory of Planned Behavior (TPB) or Discrete Choice Models, to enhance both predictive accuracy and psychological interpretability.

(3) Develop interactive dashboards or decision-support tools based on real-time or survey-based travel behavior data to assist policymakers in dynamic and data-driven transport planning.

5.2.2 Recommendations for Analyzing High-Speed Rail's Transformative Impact on Public Transport in Thailand Using Machine Learning

Based on the analysis, high-speed rail (HSR) is expected to significantly influence intercity travel behavior among passengers. To ensure the effective implementation of HSR, this study proposes policy recommendations and strategic guidelines for planning a sustainable public transportation system in the future, as follows.

1) Develop targeted pricing strategies Policymakers should consider implementing dynamic or tiered pricing structures that make HSR attractive to middle-income and price-sensitive travelers. Fare levels should be competitive with intercity buses and low-cost airlines to promote modal shift.

2) Enhance service frequency and travel time efficiency To fully capitalize on HSR's advantages, planners should ensure frequent departures and minimal total travel time, including reduced waiting and transfer durations. These operational improvements will increase perceived convenience and encourage adoption.

มหาวิทยาลัยเทคโนโลยีสุรนารี

3) Improve first-mile and last-mile connectivity Investments in feeder systems—such as shuttle buses, park-and-ride facilities, or integrated ticketing with local transit—are critical to enhance access to HSR stations and ensure seamless multimodal journeys.
4) Conduct public awareness and behavioral campaigns A nationwide communication strategy should be developed to educate the public about the benefits, reliability, and convenience of HSR, particularly in regions where travel habits are firmly rooted in bus or car usage.
5) Utilize machine learning for continuous transport planning Agencies should incorporate machine learning methods, such as CatBoost and SHAP, into routine travel behavior analysis and forecasting. These tools offer more accurate and interpretable insights, especially in complex, data-rich environments.
6) Reassess infrastructure investments in conventional modes As HSR adoption grows, policymakers must reassess the role and investment priorities of traditional modes (e.g., long-distance buses and conventional rail), including strategies for complementary or restructured services.
7) Expand research using revealed preference data Future studies should build on this research by incorporating revealed preference (RP) data after the actual launch of HSR services in Thailand, to validate and refine predictive models and policy recommendations.

มหาวิทยาลัยเทคโนโลยีสุรนารี

5.2.3 Recommendations for Machine Learning-Based Analysis of Travel Mode Preferences: Neural and Boosting Model Comparison Using Stated Preference Data from Thailand’s Emerging High-Speed Rail Network

Based on the findings of this study, the following recommendations are proposed to enhance the adoption and efficiency of high-speed rail (HSR) in Thailand:

1) Implement Competitive and Flexible Fare Structures Offer tiered pricing, promotional discounts, and integrated fare options (e.g., monthly passes or multimodal bundles) to improve affordability and attract cost-sensitive travelers.

2) Increase Service Frequency and Reduce Waiting Times Enhancing schedule regularity and minimizing waiting time can significantly improve service reliability, a key factor in influencing mode choice.

3) Improve Accessibility to HSR Stations Develop efficient first-mile and last-mile connectivity—such as shuttle buses, park-and-ride facilities, and integration with local transit systems—to reduce access time and inconvenience.

4) Target Frequent and High-Value Travelers Focus marketing and service design on traveler segments with high potential adoption rates, such as business travelers, regular intercity commuters, and households with higher income.

5) Leverage Data-Driven Planning Tools Integrate machine learning techniques and interpretable models (e.g., SHAP analysis) into transport planning processes to better understand user preferences and adapt policies accordingly.

6) Enhance Public Awareness and User Confidence Promote HSR benefits—such as time savings, comfort, and sustainability—through public communication campaigns to foster early adoption and build trust among new users.

5.3 Practical Applications of the Developed Models

The operational architecture of the proposed models comprises three core components: inputs, processes, and outputs. The input stage involves travel attribute data such as fares, in-vehicle travel time, station access time, service frequency/waiting time, as well as socioeconomic and demographic information, station location and accessibility, historical ticket sales, and mode choice data (when available). The process begins with data cleaning and preprocessing, followed by feature engineering tailored to the research objectives. The models are then tuned or calibrated to align with real-world conditions before conducting scenario simulations and calculating elasticities, which are subsequently transformed into decision-oriented performance indicators. The output stage delivers key metrics including mode share by passenger segment, forecasted ridership by route–station–time period, revenue projections, load factors, train/capacity requirements, greenhouse gas (GHG) reduction estimates, and socio-economic and financial feasibility indicators.

In terms of policy formulation and strategic planning, the models can directly address key policy questions such as "How would changes in fares, service frequency, or station accessibility affect HSR acceptance and market share?". Policy levers are operationalized into model variables, with fares represented as travel cost, service frequency as waiting time, in-vehicle travel time, and station access time. Baseline models are calibrated to reflect observed data, and policy scenarios such as 10–30% fare reductions, 20–50% frequency increases, or improved first/last-mile connectivity are tested. The Multinomial Logit Model (MNL) provides clear interpretability for policy rationale, while CatBoost, LightGBM, and XGBoost enhance predictive accuracy for identifying cost-effective policy packages.

For market segmentation and targeted promotion, the models can identify passenger segments with a high likelihood of adopting HSR. CatBoost effectively handles categorical variables such as income, occupation, and trip purpose without complex data transformations, enabling precise segmentation. Feature importance methods such as SHAP reveal the key drivers of choice within each segment, allowing the design of tailored marketing measures including monthly passes for frequent travelers, off-peak discounts, or integrated first/last-mile bundles. These measures can then be evaluated through scenario simulations to maximize return on marketing investment.

In operational optimization, the models can forecast daily and time-specific demand by incorporating seasonal effects, day-of-week patterns, special events, and recent ticket sales data. This supports operational decisions such as setting optimal headways, allocating trainsets, and managing seating to maintain target load factors.

มหาวิทยาลัยเทคโนโลยีสุรนารี

By updating the models with near-real-time data and recalibrating at regular intervals, operators can dynamically adjust operations while continuously monitoring key performance indicators (KPIs) such as load factor, on-time performance, seat spoilage, and EBITDA per train-kilometer to improve efficiency and service quality.

Regarding scenario analysis and risk management, the models can perform "what-if" analyses to assess the impacts of uncertain external factors such as fuel price volatility, competition from alternative transport modes, or macroeconomic downturns. Monte Carlo simulations combined with CatBoost or XGBoost can quantify these risks, while sensitivity analysis can identify the most influential variables. This enables the development of trigger-based contingency plans that can be activated when specific risk conditions occur.

In the context of sustainability and long-term planning, the models can quantify environmental benefits such as GHG emissions reductions from mode shifts from road or air to rail. They can estimate the social value of time savings and conduct benefit–cost analyses (BCA) to inform infrastructure investment decisions. Additionally, the models can support the preparation of sustainability reports aligned with Sustainable Development Goals (SDGs) and Net-Zero targets, positioning HSR as both an economic growth driver and a long-term environmental asset. Model selection should be aligned with application objectives, with MNL suited for policy interpretation, CatBoost/LightGBM/XGBoost for fast and accurate simulation and forecasting, and DNN/CNN for multi-dimensional or spatial–urban planning integration in long-range analyses.

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX A

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 1
Code...........................
Province.......................
District........................
Sub-district................

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province........................... District...........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people 4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle
Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel
3) shopping 4) Others.....................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 5	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 6	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 7	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 8	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 10	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 11	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 12	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Questionary Set 2
Code...........................
Province.......................
District........................
Sub-district................

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

__Explanation:__ Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province.......................... District...........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people 4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle
Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel

3) shopping 4) Others.....................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?	☐	☐

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?	☐	☐

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?	☐	☐

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?	☐	☐

Point 5	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 6	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 7	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 8	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 10	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 11	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 12	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 3
Code……………………….
Province…………………..
District……………………
Sub-district…………….

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province……………………… District………………………Sub-district……………………….

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people
4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ……………………………baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle Motorcycle……………….vehicles
Car………………..vehicles
Others ………………..vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel
3) shopping 4) Others…………………

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 5	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?	☐	☐

Point 6	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?	☐	☐

Point 7	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?	☐	☐

Point 8	Travel Modes	
	Train	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?	☐	☐

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 10	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 11	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 12	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 4
Code…………………………
Province……………………..
District……………………..
Sub-district……………..

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

<u>Explanation:</u> Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province.......................... District..........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people
4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)
1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle
Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?
1) study/ work 2) rest/ travel
3) shopping 4) Others.....................

1.8) How much frequency for travel between the provinces per year?
1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?
1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 5	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 6	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 7	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 8	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 10	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 11	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 12	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 5
Code…………………………
Province…………………..
District……………………
Sub-district……………

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province.......................... District..........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people 4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle

Motorcycle...................vehicles

Car...................vehicles

Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel

3) shopping 4) Others....................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?	☐	☐

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?	☐	☐

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?	☐	☐

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?	☐	☐

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 5	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 6	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 7	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 8	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?	☐	☐

Point 10	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?	☐	☐

Point 11	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?	☐	☐

Point 12	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?	☐	☐

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 6
Code...........................
Province.......................
District........................
Sub-district................

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province.......................... District..........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people 4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle

Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel
3) shopping 4) Others.....................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 5	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 6	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 7	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 8	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 10	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 11	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 12	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	10
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Questionary Set 7
Code...........................
Province......................
District.......................
Sub-district...............

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province........................... District...........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people
4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel
3) shopping 4) Others......................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 2	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 3	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 4	**Travel Modes**	
	Bus	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	10
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 5	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?	☐	☐

Point 6	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?	☐	☐

Point 7	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?	☐	☐

Point 8	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	10
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?	☐	☐

มหาวิทยาลัยเทคโนโลยีสุรนาร

Point 9	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 10	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 11	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 12	**Travel Modes**	
	Airplane	**HSR**
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	10
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Questionary Set 8
Code...........................
Province.......................
District........................
Sub-district................

The Questionary of Artificial Intelligence Methods Applying for the Prediction of High-Speed Rail Passengers

Purpose: to develop the model for the amount prediction of high-speed rail passenger

The questionnaire was divided into 2 parts which are

Section I: General Information of the Answerers

Section II: Travel Information by High-Speed Rail in the Future of Thailand

I agree to answer the questionnaire ○ agree ○ disagree

Section I: General Information of the Respondents

Explanation: Please mark in front of the answer that corrects to the truth

1.1) Gender 1) Male 2) Female 3) N/A

1.2) Residence Province.......................... District...........................Sub-district............................

1.3) The amount of family members 1) 1 person 2) 2 people 3) 3 people 4) 4 people 5) more than 4 people

1.4) The amount of family children (lower than 18 years old)

1) none 2) 1 person 3) 2 people 4) more than 3 people

1.5) Average household income ....baht/month *(The information is important to the analysis, please give this information and it will be secret)*

1.6) The amount of possessive vehicle

Motorcycle...................vehicles
Car....................vehicles
Others ....vehicles

1.7) What are the most travel objectives when you travel between the provinces?

1) study/ work 2) rest/ travel
3) shopping 4) Others.....................

1.8) How much frequency for travel between the provinces per year?

1) 1-3 times 2) 3-6 times 3) 6-9 times 4) more than 9 times

1.9) If the high-speed rail service will open in the future, would you like to choose its service?

1) Yes 2) No

มหาวิทยาลัยเทคโนโลยีสุรนารี

Section II: Travel Information by High-Speed Rail in the Future of Thailand

Explanation: Please consider the following situations carefully. If you would like to travel in these mentioned situations, which travel mode will you choose for travel as the conditions in the charts?

Set 1: Simulation, if you would like to travel from Bangkok to Chiangmai for an errand of 700 kilometers

Point 1	Travel Modes	
	Bus	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 2	Travel Modes	
	Bus	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	190
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 3	Travel Modes	
	Bus	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1050
Waiting time for the next round of the ride (minutes)	30	220
Which travel mode will you choose?		

Point 4	Travel Modes	
	Bus	HSR
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	15	15
Travel time (minutes)	720	220
Travel cost (baht)	750	1400
Waiting time for the next round of the ride (minutes)	30	190
Which travel mode will you choose?		

Point 5	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 6	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	190
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

Point 7	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1050
Waiting time for the next round of the ride (minutes)	150	220
Which travel mode will you choose?		

Point 8	**Travel Modes**	
	Train	**HSR**
Time for traveling from house to the station (minutes)	10	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	10	15
Travel time (minutes)	720	220
Travel cost (baht)	300	1400
Waiting time for the next round of the ride (minutes)	150	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

Point 9	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 10	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	190
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

Point 11	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1050
Waiting time for the next round of the ride (minutes)	120	220
Which travel mode will you choose?		

Point 12	Travel Modes	
	Airplane	HSR
Time for traveling from house to the station (minutes)	30	15
Time for waiting to use the service (minutes) (time for waiting to get into the ride)	120	15
Travel time (minutes)	135	220
Travel cost (baht)	3000	1400
Waiting time for the next round of the ride (minutes)	120	190
Which travel mode will you choose?		

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX B

Descriptive Statistics of Survey Variables

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table B.1 Descriptive Statistics of Survey Variables

Variable	Value	Count	Percent	Mean	Standard Deviation	Variance
GENDER	0=Female	1507	47.09	0.53	0.50	0.25
	1=Male	1693	52.91			
Household members	1=1 person	321	10.03	3.16	1.17	1.36
	2=2 person	615	19.22			
	3=3 person	890	27.81			
	4=4 person	991	30.97			
	5=More than four people	383	11.97			
Have children in household (younger than 18 years old)	0=Do not have children	1307	40.84	0.59	0.49	0.24
	1=Have children	1893	59.16			
Household income	1=Less than 15,000 Baht	90	2.81	2.9997	0.917851114	0.84245
	2=15,000 – 30,000 Baht	1,078	33.69			
	3=30,001 – 45,000 Baht	775	24.22			
	4=More than 45,000 Baht	1,257	39.28			
Number of Motorcycles	0=None	366	11.44	1.33	0.79	0.62
	1=One vehicle	1648	51.50			
	2=Two vehicles	980	30.63			

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table B.1 Descriptive Statistics of Survey Variables (Continued)

Variable	Value	Count	Percent	Mean	Standard Deviation	Variance
	3=Three vehicles	178	5.56			
	4=Four vehicles	23	0.72			
	5=Five or more vehicles	5	0.16			
Number of Cars	0=None	705	22.03	1.05	0.80	0.64
	1=One vehicle	1826	57.06			
	2=Two vehicles	533	16.66			
	3=Three vehicles	101	3.16			
	4=Four vehicles	23	0.72			
	5=Five vehicles	11	0.34			
	6=Six or more vehicles	1	0.03			
Number of Other Vehicles	0=None	3193	99.78	0.00	0.08	0.01
	1=	3	0.09			
	2=	4	0.12			
Travel purpose for study/work	0=No	2130	66.56	0.33	0.47	0.22
	1=Yes	1070	33.44			

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table B.1 Descriptive Statistics of Survey Variables (Continued)

Variable	Value	Count	Percent	Mean	Standard Deviation	Variance
Travel purpose for leisure/vacation	0=No	1572	49.12	0.51	0.50	0.25
	1=Yes	1628	50.88			
Travel purpose for shopping	0=No	2795	87.34	0.13	0.33	0.11
	1=Yes	405	12.66			
Travel frequency between provinces per year	1=1–3 times	1018	31.81	2.29	1.14	1.30
	2=3–6 times	986	30.81			
	3=6–9 times	461	14.41			
	4=More than nine times	735	22.97			
Choose High-Speed Railways	0=Not choose	393	12.28	0.88	0.33	0.11
	1=Choose	2807	87.72			

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table B.2 Mode Choice between Current Transport and High-Speed Rail by BOX Group

BOX	Mode	Current	HSR	Current %	HSR %
1	BUS	445	1189	27.23	72.77
	TRAIN	435	1179	26.95	73.05
	AIR	227	1397	13.98	86.02
2	BUS	406	1170	25.76	74.24
	TRAIN	397	1179	25.19	74.81
	AIR	281	1295	17.83	82.17
3	BUS	434	1110	28.11	71.89
	TRAIN	437	1107	28.3	71.7
	AIR	314	1230	20.34	79.66
4	BUS	336	1160	22.46	77.54
	TRAIN	345	1151	23.06	76.94
	AIR	239	1257	15.98	84.02
5	BUS	350	1382	20.21	79.79
	TRAIN	421	1311	24.31	75.69
	AIR	210	1522	12.12	87.88
6	BUS	359	1253	22.27	77.73
	TRAIN	380	1232	23.57	76.43
	AIR	180	1432	11.17	88.83
7	BUS	455	1229	27.02	72.98
	TRAIN	447	1237	26.54	73.46
	AIR	305	1379	18.11	81.89
8	BUS	388	1144	25.33	74.67
	TRAIN	405	1127	26.44	73.56
	AIR	246	1286	16.06	83.94

มหาวิทยาลัยเทคโนโลยีสุรนารี

APPENDIX C

Published in Q1 Journal “Results in Engineering”: A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand

มหาวิทยาลัยเทคโนโลยีสุรนารี

Results in Engineering 24 (2024) 103110

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand

Chinnakrit Banyong [a], Natthaporn Hantanong [b], Panuwat Wisutwattanasak [c], Thanapong Champahom [d], Kestsirin Theerathitichaipa [b], Manlika Seefong [b], Vatanavongs Ratanavaraha [b], Sajjakaj Jomnonkwao [b,*]

[a] Industrial and Logistics Management Engineering, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand
[b] School of Transportation Engineering, Institute of Engineering, Suranaree University of Technology, 111 University Avenue, Suranaree Sub-district, Muang District, Nakhon Ratchasima, 30000, Thailand
[c] Institute of Research and Development, Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand
[d] Department of Management, Faculty of Business Administration, Rajamangala University of Technology Isan, Nakhon Ratchasima, 30000, Thailand

ARTICLE INFO

Keywords:
Travel mode choice
Stated preference survey
Travel behavior
Gradient boosting
SHAP analysis

ABSTRACT

Thailand's collaboration with China to develop High-Speed Rail (HSR) represents a crucial step in enhancing transportation infrastructure and promoting regional economic growth. While most research has focused on existing travel modes such as cars, buses, and planes, there is a notable lack of analysis on future transport choices, particularly in developing countries like Thailand. This study addresses this gap by analyzing the factors influencing High-Speed Railways adoption to inform future traveler decisions. The comparison of models for predicting High-Speed Railways usage revealed that CatBoost consistently outperformed the other models, with cross-validation confirming its superior performance across all key metrics. The Binary Logit Model (BL) demonstrated moderate effectiveness, achieving an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a relatively lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were moderate, but it struggled with handling imbalanced data. In contrast, XGBoost delivered significantly stronger results, with an accuracy of 0.8846 ± 0.0068, sensitivity of 0.9210 ± 0.0094, and an AUC of 0.9583 ± 0.0036. XGBoost also achieved high specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and an F1 score (0.8910 ± 0.0067). LightGBM also performed well, achieving an accuracy of 0.8763 ± 0.0077 and a relatively high sensitivity of 0.9358 ± 0.0071. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. CatBoost demonstrated the highest overall performance, achieving an accuracy of 0.8853 ± 0.0061, sensitivity of 0.9245 ± 0.0072, specificity of 0.8441 ± 0.0068, precision of 0.8616 ± 0.0058, an F1 score of 0.8920 ± 0.0058, and an AUC of 0.9584 ± 0.0034. The decision to use High-Speed Railways in Thailand is influenced by several key factors that reflect the country's unique transportation context. These factors include travel time, access time, service frequency, cost, waiting time, household income, automobile ownership, gender, and the purpose of travel. This information is valuable for shaping policies to support High-Speed Railways adoption in Thailand.

1. Introduction

Thailand and China initially agreed to collaborate on developing railway infrastructure in Thailand as part of the country's transportation infrastructure strategy for 2015–2022. In 2022, the agreement was renewed to continue the development of high-speed rail projects. Thailand, as an ASEAN transport hub, is developing the Thailand–China high-speed railway to enhance its railway network and boost regional economic growth, though the project is still incomplete [1]. High-Speed Railways are not only critical for improving transportation

---

* Corresponding author.

E-mail addresses: d6500733@g.sut.ac.th (C. Banyong), d6040086@g.sut.ac.th (N. Hantanong), panuwat.w@g.sut.ac.th (P. Wisutwattanasak), thanapong.ch@rmuti.ac.th (T. Champahom), d6500573@g.sut.ac.th (K. Theerathitichaipa), d6500580@g.sut.ac.th (M. Seefong), vatanavongs@g.sut.ac.th (V. Ratanavaraha), sajjakaj@g.sut.ac.th (S. Jomnonkwao).

https://doi.org/10.1016/j.rineng.2024.103110
Received 14 August 2024; Received in revised form 26 September 2024; Accepted 10 October 2024
Available online 11 October 2024
2590-1230/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

มหาวิทยาลัยเทคโนโลยีสุรนาร

infrastructure but also directly influence people`s travel choices, as they provide an alternative mode of transport that competes with traditional travel methods, such as buses, airplanes, and private cars. This competition has the potential to reshape traveler behavior, especially in developing nations like Thailand, where transportation options are limited.

While most studies have focused on current travel modes such as private cars, trains, buses, and planes, there is a notable lack of analysis on future transportation mode choices, especially in developing nations such as Thailand, which plans to introduce a High-Speed Railways —an unprecedented mode of travel in the country.

High-Speed Railways are the primary infrastructure for transporting passengers and are used in many countries [2]. Research indicates that High-Speed Railways considerably benefit a country`s economic development, particularly in underdeveloped regions [3]. High-Speed Railways reduce travel time, improve labor market efficiency, and foster economic connections, while also lowering road accidents and pollution [4–6]. These benefits contribute to shifts in traveler preferences and mode choices, as individuals increasingly favor quicker, safer, and more environmentally friendly travel alternatives. Effective transportation planning is necessary to achieve these benefits, including evaluating travel options [7]. In Thailand, where High-Speed Railways is yet to be implemented, understanding traveler attitudes is critical for assessing its future impact [8,9]. By influencing factors such as travel time, cost, and convenience, High-Speed Railways could potentially shift a significant portion of the population from existing travel modes to High-Speed Railways, necessitating government policies to encourage such a transition once the High-Speed Railways is operational. High-Speed Railways are expected to compete with or replace airlines, buses, and private cars, requiring government policies that promote their use after construction.

Several factors influence travel mode choices, such as demographics, travel attributes (distance, cost, convenience), safety, and land use [4, 10–17]. Traditional models like binary logit (BL) and multinomial logit (MNL) have been widely used to study travel behavior [18–22], but recent research shows that machine learning (ML) techniques can handle complex relationships more effectively [23]. Kashifi, Jamal [4] found that light gradient boosting decision trees outperformed traditional models, while Abulibdeh [24] demonstrated that XGBoost was more efficient than Binary Logit (BL) and MNL in predicting metro mode choice. Wang and Ross [25] confirmed that XGBoost outperformed MNL, and Li, Shi [12] used XGBoost to analyze High-Speed Railways choices in China. These studies support the growing use of machine learning for travel mode analysis, especially with ensemble learning techniques [4,24–32]. Travel mode choice is influenced by subjective factors (e.g., age, income, car ownership) [33–37], area determinants (e.g., population density), and passenger perceptions (e.g., policies, and fees) [38–43]. Spatial factors like land use and travel costs also affect choices, with studies showing different preferences in the USA and Germany due to varying transit policies [44–46]. In Thailand, public-private partnerships (PPP) play a critical role in High-Speed Railways development. Sresakoolchai and Kaewunruen [44] highlighted how PPPs improve accessibility and cost control, promoting High-Speed Railways adoption. However, operational issues in Bangkok`s Airport Rail Link have hindered public transport use, though improvements could address these problems [45]. Rattanakijsuntorn, Suwannarat [46] emphasized air-rail connectivity`s importance for High-Speed Railways promotion. Machine learning is increasingly used in travel mode modeling for its accuracy in handling large datasets [47–49]. Studies by Wang and Ross [25], Cheng, Chen [32], and Zhao, Yan [31] found that ML models like XGBoost, random forest, and neural networks outperform traditional models. Abulibdeh [24] and Kim [30] also validated XGBoost and neural networks for predicting travel modes, which are detailed in Table 1 and Fig. 1.

While traditional machine learning models are robust and widely used, they often struggle with complex, nonlinear relationships and large datasets. In contrast, advanced ensemble methods like XGBoost, LightGBM, and CatBoost offer superior performance by handling high-dimensional data, improving interpretability, and reducing overfitting. This research compares the efficiency of these boosting methods in analyzing future travel mode choices in developing countries, offering insights for policy recommendations in Thailand to support High-Speed Rail adoption and assess unintroduced travel modes.

2. Material and methods

2.1. *Stated preference surveys*

The survey process examines the factors influencing the decision to use High-Speed Railways across four regions of Thailand: Northern, Northeastern, Southern, and Central. Each region will select an area from the province where an High-Speed Railways station is planned for construction [50] based on the Gross Regional and Provincial Product Chain Volume Measures for 2022 [51], focusing on the highest value in each of the region`s top four priorities, as shown in Fig. 2.The survey design aims to understand the factors that influenced to decision on high-speed rail (HSR) service in each region of Thailand by choosing the important economic provinces and collecting data from many critical

Table 1
Summary of selected literature on the use of machine learning for predicting travel mode choices.

Authors	Regions	Statistical approaches	Input Data	Key findings	Accuracy
[25]	United States of America	XGBoost, MNL	Travel time, socio-demographic.	XGBoost outperformed MNL, with travel time being the most significant factor.	93.12 %
[24]	Qatar	XGBoost, MNL, Binomial Logit	Travel behavior, metro introduction.	XGBoost was most effective, with trip parameters and socioeconomic factors crucial.	91.2 %
[28]	Mexico	RF, XGBoost, GB, NB, DT, LR, SVM (polynomial), SVM (RBF)	Transport preferences, student travel data.	XGBoost showed excellent performance for mode choice estimation.	92.5 %
[32]	China	RF, AdaBoost, and SVM	Travel instances (age, land use, traffic data)	Random Forest outperformed others with higher accuracy and lower computational cost.	91 %
[27]	Egypt	DNN, KNN, AdaBoost, SVM, NB, RF, GB, XGB, LDA, DT, QDA	Household size, income, travel time, distance	DNN achieved 97.81 % accuracy, outperforming RF and XGBoost with larger sample sizes.	97.81 %
[26]	Egypt	MNL, SVM, MLP, KNN, AdaBoost, DT, RF, GBDT, and XGB	Mode choice responses, Alexandria city.	GBDT achieved the highest prediction accuracy, exceeding MNL.	93.6 %
[31]	United States of America	CART, NB, Bag, Boost, RF, SVM, and NN	Socio-economic data, travel characteristics.	Logit and ML models differed in outcomes, requiring balance between accuracy and validity.	89.5 %
[30]	South Korea	ANN, XGBoost, RF	Age, activity duration, land use	XGBoost and ANN performed best, with age, trip time, and land use being key factors.	94.3 %
[29]	–	MNL, MMNL, ANN, SVM, RF, and XGBoost	Travel responses, socio-economic data	XGBoost, RF, NN outperformed logit models.	95.2 %
[4]	Netherland	LR, RF, DT, MLP, and LightGBDT	Trip density, income, car/bike ownership, age	LightGBDT outperform others; key factors were trip density, income, ownership, and age.	94.5 %

มหาวิทยาลัยเทคโนโลยีสุรนารี

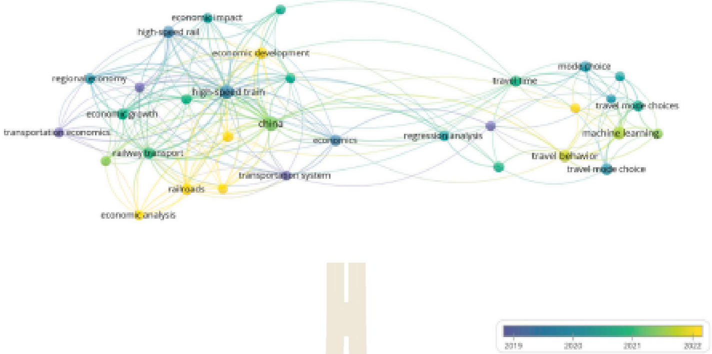


Fig. 1. VOS viewer Network Visualization of Keywords Related.

provinces in Thailand because the economic characteristics and infrastructure of each region are different. Especially in Thailand, each province has clearly different infrastructure even the province that is located in the same region, it is distinctly difference [52]which directly affects the travel behavior. The survey in many provinces helps the researcher access to extensive data and reflect the various of travel behaviors at the country level [53].

Face-to-face interviews will be performed with individuals aged 18 and older at transportation stations, shopping malls, and public parks across 16 provinces in Thailand. consists of northern zone: Nakhon Sawan, Phitsanulok, Lamphun, and Chiang Mai, Northeastern zone: Nakhon Ratchasima, Khon Kaen, Udon Thani, and Nong Khai, Central zone: Bangkok, Chonburi, Rayong, and Ayutthaya, and Southern zone: Prachuap Khiri Khan, Chumphon, Surat Thani, and Songkhla. The weather of these 4 zones is tropical making Thailand is inappropriate to travel as an active journey [54]and travelers always use transportation systems with air conditioners which affects travel mode choice of intercity travel result to main vehicles like private cars, buses, planes, air conditioner trains, or high-speed rails [55,56]. A total of 200 samples per province will be collected, using Convenience Sampling to ensure a diverse and efficient selection of participants. This approach guarantees balanced representation across all regions, mitigating bias and enhancing the representativeness of the findings at the national level, totaling 3200 samples, is sufficient for machine learning analysis, as the sample size should be at least 50 to 1000 times the number of prediction classes. This guideline helps ensure that the model has enough data to perform effectively. Since there are 2 prediction classes in this research [57].

The survey questionnaire is divided into two sections. The first section focuses on social and economic data, including gender, marital status, household size, presence of children in the family, household income, vehicle ownership, purpose of intercity travel, and travel frequency between cities. The second section comprises a stated preference (SP) questionnaire, where the stated choice (SC) experiment allows for the description of alternatives through various attribute combinations at different levels. This method helps estimate the marginal utility of each attribute for the High-Speed Railways option, enabling the assessment of in-mode choices, access times, waiting times, travel times, travel costs, and travel frequency related to the proposed mode of transportation [58, 59]. The stated choice (SC) experiment required travelers to choose between alternative High-Speed Railways options and their current public transportation mode. Collecting data on travel mode preferences—taking into account factors such as access time, waiting time, travel time, travel cost, and frequency—is crucial for understanding passenger behavior and preferences. This study will reflect real-world scenarios, demonstrating that passengers who wish to reduce their travel time in the public transportation system may need to incur higher costs [59]. These data can drive the development and enhancement of transport services by improving service frequency, optimizing routes to reduce travel time, and setting appropriate fare structures. Additionally, they assist facilitators in assessing competitiveness with other travel modes and making informed business decisions. While the stated preference (SP) questionnaire may have lower reliability owing to potential discrepancies between respondents' stated preferences and their actual behavior in real situations, it remains a widely used method for collecting relevant data on travel modes [60–64].

2.1.1. *Statistical design of the choice experiment*

The attributes and their levels used in the stated choice (SC) experiment are detailed in Table 2. These attributes encompass various factors that influence travel mode decisions, such as access time, waiting time, travel time, travel cost, and frequency. Different levels are assigned to each attribute to represent a range of scenarios and respondent preferences. This structured approach enables a thorough analysis of how different factors affect travel decisions and the trade-offs individuals are willing to make.

In this experiment, hypothetical scenarios were based on a 700-km journey from Bangkok to Chiang Mai, a route earmarked for future high-speed rail development, with attribute levels applied to this travel context in Thailand [65]. As the High-Speed Railways infrastructure had not yet been constructed at the time of the study, there was no empirical data available regarding actual travel times, pricing, waiting times, service schedules (frequency of service), or operational details. Consequently, data for pricing, travel times, waiting times, and service frequency were referenced from governmental sources that provided estimates to the public through news outlets and websites in Thailand, which may still be subject to change [66,67]. However, within the hypothetical scenarios, these estimates were adjusted to reflect a range of possible outcomes that are reasonably consistent and close to the projected figures. Waiting time was considered based on recommendations from ticket booking sources that advise passengers to arrive in advance. For other travel modes, the hypothetical scenario referenced pricing, travel times, waiting times, and service frequency from real-world sources: for buses, data was taken from national ticket booking centers [68]; for trains, the estimates were primarily based on data from the State Railway of Thailand's booking system [69]; and for airplanes,

วิทยาลัยเทคโนโลยีสุร

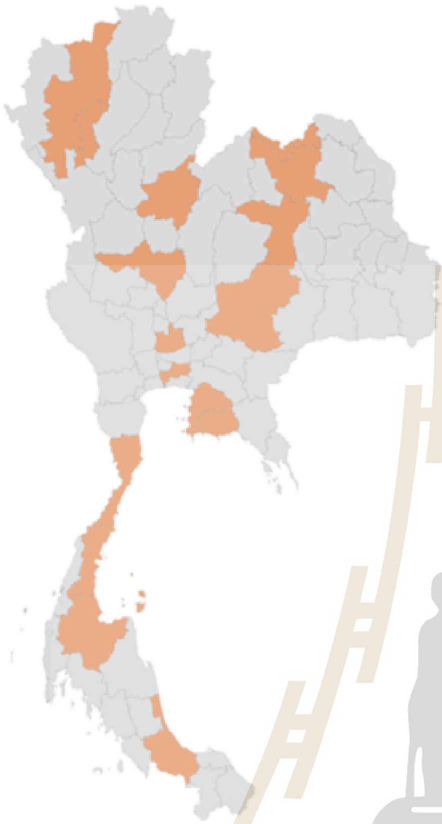


Fig. 2. Provinces Surveyed for data collection in Thailand.

Table 2
Typical cards from stated choice (SC) sets.

Attribute/Levels	Mode Choice 1			Mode Choice 2
	bus	train	airplane	high-speed rail
access time (min)	10	10	30	10, 15
waiting time (min)	15	10	120	15, 10
travel time (min)	720	720	135	190, 220
travel cost (baht)	750	300	3000	1050, 1400
Frequency times (min)	30	150	120	190, 220

pricing, schedules, and frequencies were referenced from airline ticketing websites [70]. The different pricing, time ranges, waiting times, and frequencies in these hypothetical scenarios were designed to reflect realistic variations across service providers. Access times were set within reasonable limits to ensure fairness between buses, trains, and high-speed rail, while access to airports was higher due to the more challenging reach compared to other transport stations in Thailand. The figures used in the simulation were within the realistic range that passengers would typically encounter. This route, with a distance suitable for competition between High-Speed Railways nd other modes of transport, provided an ideal context for analyzing the potential impacts of introducing High-Speed Railways.

In our study, we evaluated five attributes across two groups: traditional modes (bus, train, and airplane) with one level each and High-Speed Railways with two levels. A full factorial design would yield 96 choice sets ($3 \times 1^5 \times 2^5$), which would be overwhelming for respondents. Instead, we used an orthogonal fractional factorial design, creating 96 profiles divided into eight sets of 12 choice sets each. As a result, each respondent only evaluated 12 of the 96 profiles during the survey [71]. To ensure an efficient choice design, our approach incorporated key properties such as orthogonality, level balance, minimal overlap, and utility balance. Orthogonality was achieved by arranging attribute levels in the design correlation matrix so that any two columns were uncorrelated, thereby minimizing collinearity. Level balance was maintained by ensuring that each attribute level appeared an equal number of times across the profile sets. Utility balance was addressed by minimizing the differences in utility between alternatives, achieved by rotating the dominant alternatives in each choice set.

2.2. *Methodology*

The research methodology, outlined in the flowchart in Fig. 3, begins with the data collection phase, where both Revealed Preference (RP) and Stated Preference (SP) data are gathered. This information captures current transportation behaviors and potential future choices in the context of adopting high-speed rail. Additionally, the survey process examines the factors influencing the decision to use High-Speed Railways across four regions of Thailand, selecting provinces where High-Speed Railways stations are planned based on their economic significance, to ensure a diverse and representative sample. Three key factors are considered in the data collection process: socio-economic aspects, travel behavior, and various decision-making scenarios that could influence an individual's choice of transportation mode. Once the data is collected, it is passed through a splitter, which divides it into two main categories: those who would adopt the future transportation mode (high-speed rail) and those who would continue using current transportation modes. This separation of data helps to analyze and train models effectively based on the distinct behavior patterns of these two groups. The next step involves getting the data ready for model training. The dependent variable, which represents the selected mode of transportation, is allocated to y_train, while the independent variables, which comprise socioeconomic and travel behavior data, are assigned to X_train. Python was used for the analysis through Anaconda. As for the analysis time, after data correction and final adjustments, the analysis took approximately 4–5 h per model. The computer specifications used for the analysis are as follows: Processor: 13th Gen Intel(R) Core (TM) i9-13900H 2.60 GHz, RAM 32.0 GB, where a variety of categorization models, including more sophisticated machine learning methods like XGBoost, LightGBM, and CatBoost, as well as more conventional models like the Binary Logit Model, are taught to anticipate transit choices. Because these models can manage intricate relationships in the data, they are used to improve prediction accuracy. The models are hyperparameter tuned to maximize performance after training. This process involves adjusting the internal parameters of the models to ensure the best possible outcomes. The optimized models are then tested using a separate dataset, X_test and y_test, to evaluate their accuracy and generalization capabilities. Finally, the performance of each model is compared using various evaluation metrics are Accuracy, Precision, Recall, F1-score, and AUC. These metrics help determine which model performs the best. The final step includes identifying the key factors influencing transportation decisions using SHAP analysis, which allows for an interpretability of model predictions, highlighting the most important variables that drive decisions regarding future transportation modes like high-speed rail.

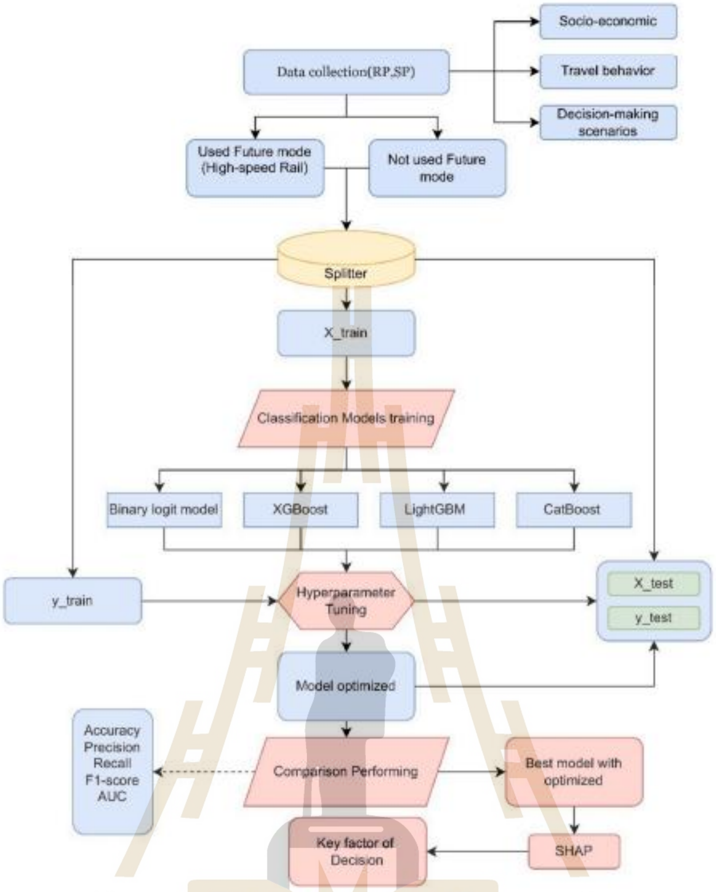


Fig. 3. Research process flowchart.

2.3. Binary logit model in mode choice (BL)

The Binary Logit Model (BL) is applied in mode choice studies as a discrete choice model because it can predict the probability of specific events occurring based on independent variables. The probability of choosing mode (i) for travel is determined by the likelihood that the utility of mode (i) is equal to or greater than the utility of an alternative mode (j). As a result, travelers select the mode of transportation that offers the highest utility.

The probability calculation for developing a binary logit model (BL) is as follows:

$$P_{n1} = \frac{\exp\left(\beta X_{1n}\right)}{\exp(\beta X_{1n}) + \exp(\beta X_{2n})} = \frac{1}{1 + \exp\left(\beta X_{2n} - \beta X_{1n}\right)} = \frac{1}{1 + \exp\left(\Delta U\right)} \quad (1)$$

Where,

P_{n1} indicates the likelihood that traveler n will select the first option.

βX_{1n} denotes the utility function associated with traveler n selecting the first mode.

βX_{2n} denotes the utility function associated with traveler n selecting the second mode.

$\Delta U = \beta X_{2n} - \beta X_{1n} = \sum (a_i - b_i) Z_i$, where Z_i is the i^{th} variable, a_i is the coefficient of the i^{th} variable in βX_{n1}, and b_i is the coefficient of the i^{th} variable in βX_{n2}.

2.4. Machine learning

Machine learning (ML) has significantly advanced transportation mode choice modeling by enhancing predictive accuracy and providing

มหาวิทยาลัยเทคโนโลยีสุรนารี

deeper insights into traveler behavior, which are crucial for urban planning and policy-making. Models such as random forests and XGBoost have consistently outperformed traditional approaches like the multinomial logit (MNL) model, as demonstrated in studies from Bengaluru and Chengdu, where ML models achieved superior accuracy [72,73]. Similarly, in other fields such as material production, research has confirmed that ML surpasses traditional analytical methods [74–76]. Furthermore, techniques like SHAP have improved the interpretability of ML models, allowing for the identification of key factors such as travel costs and demographic characteristics that influence decisions [73,77]. Feature importance analysis facilitates more informed policy decisions, and ML models have demonstrated greater sensitivity to policy changes, as evidenced by studies in Alexandria City [26].

2.4.1. *Ensemble models*

Ensemble modeling has garnered a lot of interest in machine learning applications due to its broad range of dependent variable prediction capabilities. Typically, tree-based techniques like random forests, additional trees, gradient boosting, or adaptive boosting are used to build ensemble models. On the other hand, new research has demonstrated the potency of innovative ensembles that blend different machine learning techniques to enhance prediction accuracy. These ensembles are sometimes referred to as multi-step-ahead modeling or inclusive multiple models [78,79]. Notably, boosting is used in prominent ensemble techniques like XGBoost, LightGBM, and CatBoost to improve prediction accuracy and classification problems. XGBoost is more effective at preventing overfitting than LightGBM, which combines gradient-based one-side sampling and leaf-wise tree growth that is targeted for speed and efficiency on big datasets. This is because XGBoost uses regularization and second-order derivatives. Random Permutation is an effective technique that CatBoost excels at using to handle categorical data with little bias and no overfitting. These models are routinely used to analyze difficult data with extraordinary accuracy, highlighting the increasing significance of sophisticated ensemble techniques in machine learning.

Three popular gradient boosting libraries, XGBoost, LightGBM, and CatBoost [80], are highly effective for predicting transportation mode choices by analyzing factors like travel cost, time, and socio-demographic attributes. XGBoost handles complex relationships in large datasets, LightGBM excels in speed and scalability, and CatBoost is ideal for categorical data and preventing overfitting, making all three well-suited for transportation mode analysis with minimal preprocessing and high accuracy. These models XGBoost, LightGBM, and CatBoost are well-suited for analyzing transportation mode choices due to their ability to capture complex, nonlinear interactions between variables such as travel time, cost, convenience, and demographic factors like income and car ownership [12,81–83]. XGBoost and LightGBM efficiently manage large datasets, while CatBoost makes it ideal for analyzing traveler preferences. Their ability to manage complex data relationships helps identify key factors influencing travel decisions, offering valuable insights for transportation planning.

2.4.2. *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) is a machine learning technique that enhances the efficiency of constructing scalable decision trees and is widely used across various fields [84]. XGBoost was developed from the Gradient Boosting Decision Tree (GBDT) algorithm, which consists of multiple decision trees and consistently uses classification and regression trees [85]. However, XGBoost differs from GBDT in two key ways. First, XGBoost uses the second-order Taylor expansion for the loss function, whereas GBDT only uses the first-order Taylor expansion. Second, XGBoost incorporates regularization in the objective function to prevent overfitting and reduce model complexity [86]. XGBoost has been widely used for many years owing to its efficiency and high accuracy, which stem from its interpretability, flexibility, and scalability. For example, it can efficiently handle large datasets in parallel and iteratively improve the model, often outperforming other algorithms. Additionally, XGBoost assesses feature importance to aid in understanding the prediction process, supports custom loss functions and evaluation metrics, and allows users to adjust model parameters to meet various requirements, offering greater flexibility [12].In the context of modeling by XGBoost, we start with an explanation of the learning objective that was adjusted to be regularity. In terms of mathematics, the specified D data set with k sample and m characteristic is shown as $D = \{(X_i, Y_i)\}\ (\lceil D \rceil = k, X_i \in \mathbb{R}^m, Y_i \in \mathbb{R})$. A GB model, such as XGBoost, uses an additional regularization function N to predict results [84].

$$\widehat{Y}_i = \phi(X_i) = \sum_{n=1}^{N} f_n(X_i), f_n \in F, \tag{2}$$

The value of $F = \{f(x) = \omega_{q(x)}\}\ (q : \mathbb{R}^m \rightarrow T, \omega \in \mathbb{R}^T)$ refers to the regression tree area (specifically, CART). N is the number of trees, F encompasses all tree regions, q is the tree structure, T is the number of leaves, f_n is a tree with structure q, and the leaf weights are independent. The function $q(x)$ aligns with the input data to learn function groups used in the model, enabling the objectives to be regularized as follows:

$$\begin{aligned} \Lambda(\theta) &= \sum\nolimits_i l(\widehat{Y}_i, Y_i) + \sum\nolimits_n \Omega(f_n) \text{ where} \\ \Omega(f_n) &= \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \end{aligned} \tag{3}$$

The term l represents different convex loss functions used to measure the prediction of $\widehat{Y}_i$ against the target of $\widehat{Y}_i$. The second term, Ω, penalizes model complexity and includes the regression tree functions γ and λ, which are regularization parameters. This regularization term smooths the final learned weights, helping to prevent overfitting. In summary, the objective of regularization is to select a simpler model that still provides efficient predictions. However, adding a group of efficient tree models cannot enhance area efficiency using the traditional Euclidean method because the model will be trained by adding this loss function.

$$\Lambda^{(t)} = \sum_{i=1}^{k} l\left(Y_i, \widehat{Y}_i^{\,t-1} + f_t(X_i)\right) + \Omega(f_t) \tag{4}$$

This equation will gradually add a function f_t, which helps to improve the model as much as the equation.

2.4.3. *Light gradient boosting machine (LightGBM)*

LightGBM is a machine learning algorithm developed by Ke, Meng [87] to improve the efficiency and accuracy of models based on Gradient Boosting Decision Trees (GBDT). LightGBM introduces several key innovations, including gradient-based one-side sampling (GOSS), leaf-wise tree growth, histogram-based approaches, and exclusive feature bundling (EFB). These developments allow LightGBM to work with GBDT models up to twenty times faster than traditional frameworks, without sacrificing prediction accuracy [87]. LightGBM significantly enhances GBDT, making it an excellent choice for solving complex classification problems. One of its primary advantages is that GBDT models are known for their superior predictive power compared to general models [88]. Additionally, GBDT-based algorithms, like LightGBM, are not constrained by the distribution of independent and dependent variables and can effectively manage multicollinearity, outliers, and missing values. LightGBM also does not require specifying the relationship between independent and dependent variables in advance, allowing it to better capture nonlinear relationships.

Moreover, LightGBM automatically determines the contribution of independent variables to the prediction by considering the interactions among them [89]. This is a distinct advantage over traditional statistical models, which rely on standard coefficients and marginal effects. By accounting for auxiliary effects between independent variables, LightGBM can more precisely depict the effect size of those variables, making it particularly effective for handling complex, high-dimensional datasets.

The main procedure of LightGBM begins with calculating the gradient for data example i and tree t. The loss value L is then calculated using the gradient value g_i [87] given by

$$g_i = \frac{\partial L(y_i, \widehat{y_i})}{\partial \widehat{y_i}} \tag{5}$$

y_i is the actual value, and $\widehat{y_i}$ is the predicted value from the tree. LightGBM uses gradient-based one-side sampling (GOSS) to select only the examples with the highest gradients while subsampling those with lower gradients to reduce the data used for model training. Next, exclusive feature bundling (EFB) is applied to group mutually exclusive features together, reducing the number of features and speeding up tree construction. During tree building, LightGBM constructs the tree by distributing the best data based on the calculated gradients, adjusting parameters to enhance efficiency. Finally, the predicted values are obtained by aggregating the results from all trees as follows:

$$\widehat{y}_i = \sum_{t=1}^{T} f_t(x_i) \tag{6}$$

T represents all the trees, and $f_t(x_i))$ is the function from tree t. The LightGBM framework is designed to efficiently evaluate these functions and handle large datasets effectively. Additionally, parameter tuning and technique adjustments are used to prevent overfitting and enhance prediction accuracy.

2.4.4. *Categorical boosting (CatBoost)*

CatBoost is a machine learning method based on the Gradient Boosting Decision Tree (GBDT) framework, developed by Yandex engineers in 2017 [90,91]. It refines the GBDT approach to efficiently use all available data during training by enhancing the level method. Initially, the training process assigns equal weights to all samples. After this initial phase, the weights are adjusted to give more emphasis to samples that deviate from the predictions. The next basic training sample will be repeated in each round, with weights adjusted based on the errors of previous predictions until all training samples are processed. Finally, the overall prediction is derived from aggregating the predictions of all training samples, with weights adjusted accordingly to achieve the optimal result [92]. This approach helps mitigate overfitting issues during training [93].

CatBoost enhances the efficiency of gradient evaluation through a priority-based procedure. It begins by sampling permutations σ of [1, n] to sort the training samples and then initiates n different models, M1, M2, …, Mn. Each model Mi is trained using the i-th sample from the permutation. In each iteration, the gradient estimation bias of the traditional GBDT is addressed, with the sample j being processed by the Mj−1 model [94]. This method can decrease traditional gradient-boosting decision tree gradient estimation bias and help the model gain higher generalization. However, despite these strengths, CatBoost can still face challenges when dealing with highly imbalanced data. While it adjusts class weights to manage imbalance, extreme cases where one class significantly outweighs the other may result in difficulty accurately predicting the minority class, increasing error rates for those samples. Moreover, incorrect tuning of parameters such as learning rate, depth, and iterations can lead to overfitting or underfitting, particularly impacting the minority class. Additionally, if not properly adjusted for imbalanced data, CatBoost may introduce biases into the predictions, resulting in poor accuracy for underrepresented classes [90].

The CatBoost algorithm comprises several key steps. It starts with classification adjustment using the target-based statistics technique, which calculates statistics based on the prediction target, such as the average target for each classification during the building process. CatBoost uses boosting learning to construct multiple trees, with each tree aimed at correcting the errors of the previous one. The gradient value is calculated from the loss function L to refine the predicted value y_i and enhance the accuracy of subsequent predictions [90]. The following formula is used:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \tag{7}$$

η is the learning rank.

2.5. *Shapley additive explanations (SHAP)*

SHAP is a tool designed to explain the predictions of machine learning models, particularly complex ones such as gradient boosting machines (GBM). While these models may perform well during training, their internal functioning can be difficult to interpret, and their effectiveness in real-world situations may not always be clear. SHAP helps to elucidate how these models make predictions, providing insights into their decision-making processes [95]. This problem considerably affects the risk and reliability of the model [96]. Enhancing the ability to explain the "black box" problem in models is a crucial approach to improving the performance and reliability of machine learning algorithms [97,98]. SHAP provides a clear analysis of each feature's impact on the predicted result by estimating feature importance and integrating it with the SHAP values of other features used in the model. This makes SHAP a key tool for interpreting GBM models. The results of the SHAP value analysis are typically presented in a ranked diagram, where the horizontal position of each variable indicates its importance and effect on the prediction. The interpret ranking will use the color to show the relation size for the observation: + (red) or − (blue) [99,100].

SHAP value can be calculated using the following formula:

$$\varphi_k(val) = \sum_{s \subseteq N \setminus \{i\}} \frac{|s|!(n - |s| - 1!}{n!} (\mathrm{val}(s \cup \{i\}) - val(s)) \tag{8}$$

where val is the feature relevance to the algorithm target and $\varphi_k(val)$ is the weighted summing of the feature contributions to the model target result overall feature combinations. $\frac{|s|!(n-|s|-1!}{n!}$ is the weight of $|s|$, $val(s)$ is the expected value of $|s|$, The number of features on the data framework is n, and the subset of the model features is s. The letter i stands for the vector of feature values for the sections that require interpretation [35, 99,101–103]

A linear summation of the standard features and the SHAP of all features were the goals of the method. It was shown as:

$$g(x) = l(x_s) = \varphi_0 + \sum_{t=1}^{n} \varphi_t x'_s \tag{9}$$

where φ_0 is the standard value for the undefined features and n is the number of features in the data framework, for feature t, the value φ_t represents the SHAP value, whereas x_s is the vector of reduced input variables [103].

3. Results

3.1. *Descriptive analysis*

Traveler data collection in Thailand involved a self-administered questionnaire with 3200 samples. After data collection, a comprehensive data integrity assessment was conducted to ensure the completeness and reliability of the dataset. This assessment included verifying that respondents answered all required questions and calculating basic statistical measures such as the mean, skewness, and kurtosis to identify any potential anomalies in the data distribution. Additionally, under sampling techniques were applied to address data imbalance by reducing the size of the majority class, ensuring a more balanced dataset that would not disproportionately affect subsequent analysis and model performance. The results are presented in Table 3.

The table shows that the research population comprised 52.5 % males and 47.5 % females. Most households had four members (33.4 %)

Table 3
General social, economic, and travel data.

Variable	Description	Categorical variable (%)	Mean	SD	KU	SK
GENDER	Male = 1	52.5	0.525	0.499	−1.989	−0.100
	Female = 0	47.5				
	Total	100				
HOUSEHOLD_SIZE	Household members	8.6	3.210	1.12	−0.675	−0.301
	1 person = 1					
	2 people = 2	18.0				
	3 people = 3	28.8				
	4 people = 4	33.4				
	More than four people = 5	11.4				
	Total	100				
CHILDREN	Have children in household (younger than 18 years old) = 1	62.0	0.619	0.485	−1.756	−0.493
	Do not have children in household (younger than 18 years old) = 0	38.0				
	Total	100				
HOUSEHOLD_INCOME	Less than 15,000 = 1	2.7	2.944	0.896	−1.306	−0.113
	15,000–30,000 = 2	35.1				
	30,001–45,000 = 3	27.3				
	More than 45,000 = 4	34.9				
	Total	100				
CARS	Have = 1	75.2	0.752	0.431	−0.634	−1.168
	Do not have = 0	24.8				
	Total	100				
OBJ1	Travel purpose for study/work Yes = 1	38.6	0.386	0.486	−1.782	0.466
	No = 0	61.4				
	Total	100				
OBJ2	Travel purpose for leisure/vacation	48.8	0.488	0.499	−1.99	0.046
	Yes = 1					
	No = 0	51.2				
	Total	100				
OBJ3	Travel purpose for shopping	10.3	0.102	0.303	4.864	2.619
	Yes = 1					
	No = 0	89.7				
	Total	100				
TRAVEL_FREQUENCY	Travel frequency between provinces per year	32.0	2.170	1.092	−1.056	0.502
	1–3 times = 1					
	3–6 times = 2	30.9				
	6–9 times = 3	14.4				
	More than nine times = 4	22.7				
	Total	100				
MODE_CHOICE	Choose new mode (High-Speed Railways) = 1	51.2	0.512	0.499	−1.997	−0.048
	Do not choose new mode (bus, train, and airplane) = 0	48.8				
	Total	100				

Note: SD = standard deviation.
KU = kurtosis.
SK = skewness.

or three members (28.8 %), and 62 % of respondents had children. The majority reported incomes between 15,000 and 30,000 baht (35.1 %) or above 45,000 baht (34.9 %). Approximately 75.2 % owned a private car. The primary travel purposes were for study or work (38.6 %) and leisure or vacation (48.8 %), with only 10.3 % traveling for shopping. In terms of travel frequency between provinces per year, most respondents traveled 1–3 times (32 %) or 3–6 times (30.9 %). Most preferred traveling by High-Speed Railways (51.2 %) over the current transportation system (48.8 %). Additionally, the descriptive statistical data indicated a normal distribution, as the skewness (SK) values were within the acceptable range of −3.0 to +3.0, and kurtosis (KU) values were lower than 10.0 [104].

3.2. Model fitting and performance

This study used four models to verify and predict factors influencing future travel mode choices for High-Speed Railways, analyzed through Python programming. To enhance model evaluation, a 10-fold Cross-Validation technique was applied to assess model performance consistently.

Focusing on improving model efficiency for classifying unseen data and addressing overfitting issues. The verification process used data from the stated preference (SP) questionnaire to compare the performance of the Binary Logit Model (BL), XGBoost, LightGBM, and CatBoost. Grid search optimization was applied to critical hyperparameters for XGBoost, LightGBM, and CatBoost to decrease overfitting and ensure robust model predictions, as shown in Table 4. Model performance was assessed by comparing the log loss values between the training and test datasets, as detailed in Tables 5 and 6.

This study, XGBoost, LightGBM, and CatBoost algorithms were utilized to model transportation mode choice, with each model's performance optimized through a grid search algorithm to determine the optimal set of hyperparameters. For XGBoost, key hyperparameters calibrated included the learning rate (eta), maximum tree depth (max_depth), and regularization terms (alpha and lambda), with logloss as the evaluation metric. Similarly, for LightGBM, parameters such as the learning rate, maximum tree depth, and number of leaves were fine-tuned, using binary_logloss as the evaluation metric. CatBoost, chosen for its effective handling of imbalanced datasets, was optimized by adjusting the learning rate, tree depth, and number of iterations. CatBoost addresses data imbalance effectively through its inbuilt mechanism, specifically utilizing Gradient Boosting with Ordered Boosting. This technique helps reduce overfitting by constructing decision trees that are more flexible and robust, even when one class significantly outweighs others [105,106]. These adjustments aimed to enhance both the predictive accuracy and robustness of the models, with final hyperparameter values summarized in Table 4.

The application of 10-fold Cross-Validation ensured that each

วิทยาลัยเทคโนโลยีสุร

Table 4
Hyperparameter values determined using grid search algorithms for XGBoost, LightGBM, and CatBoost.

Model	Parameter	Description	Value
XGBoost	eta	Learning rate: [0,1]	0.2
	max_depth	Maximum depth of a tree: [0,20]	7
	gamma	Min_split_loss: [0,10]	0
	alpha	L1 regularization term on weights: [0,10]	0.1
	lambda	L2 regularization term on weights : [0,10]	1
	objective	Objective function	binary: logistic
	eval_metric	Evaluation metric	"logloss"
LightGBM	learning_rate	Learning rate: [0,1]	0.2
	max_depth	Maximum depth of a tree: [0,20]	6
	num_leaves	Maximum tree leaves	41
	reg_alpha	L1 regularization term	0.1
	reg_lambda	L2 regularization term	0.1
	objective	Objective function	binary
	metric	Evaluation metric	binary_ logloss
CatBoost	depth	Maximum depth of a tree	9
	iterations	Number of iterations	500
	l2_leaf_reg	L2 regularization term	1
	learning_rate	Learning rate: [0,1]	0.05

model's performance was rigorously tested across different data partitions. This methodology not only improved the reliability of the results but also provided insight into the variance of the model performance metrics across different folds. As shown in Table 7, the cross-validated performance metrics, including Accuracy, Sensitivity, Specificity, Precision, F1 score, and AUC, highlight how each model performed under consistent conditions.

The evaluation of model efficiency considered metrics such as accuracy, precision, recall, and F1 score, all of which offer different insights into model performance across various dimensions. The use of macro average (Macro avg) was chosen for this study as it treats each class equally, regardless of the number of samples in each class. This approach is crucial in multiclass classification problems with imbalanced data, where some classes have significantly fewer samples than others. The macro average involves two main steps: 1) calculating the evaluation metrics (precision, recall, and F1 score) for each individual class, and 2) averaging these values to provide a more reliable and acceptable comparison of model performance [107,108]. The values were considered as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \quad (10)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

$$F - \text{Score} = \frac{2TP}{2TP + FN + FP} \quad (13)$$

$$AUC = \frac{1}{2}(\text{Recall} + \text{Specificity}) \quad (14)$$

These metrics were calculated from a confusion matrix, which includes:

TP is a true positive rate.
TN is a true negative rate.
FP is a false positive rate.
FN is a false negative rate.

Accuracy represents the proportion of correct predictions over the total instances, while recall (sensitivity) measures the model's ability to correctly identify relevant instances, and specificity focuses on identifying negative instances accurately. Precision assesses the proportion of correct positive predictions, and the F1 score balances precision and recall by calculating their harmonic mean. Given the imbalanced nature

Table 5
Performance comparison of train models for high-speed railways usage prediction.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7388	0.7646	0.7118	0.7355	0.7497	0.8149
XGBoost	0.9312	0.9542	0.9070	0.9151	0.9342	0.9306
LightGBM	0.8966	0.9362	0.8549	0.8714	0.8962	0.9681
CatBoost	0.9461	0.9643	0.9270	0.9327	0.9460	0.9910

Table 6
Performance comparison of test models for high-speed railways usage prediction.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7445	0.7720	0.7154	0.7410	0.7562	0.8187
XGBoost	0.8887	0.9245	0.8512	0.8672	0.8949	0.8878
LightGBM	0.8762	0.9299	0.8197	0.8443	0.8755	0.9526
CatBoost	0.8970	0.9292	0.8631	0.8771	0.8966	0.9669

Table 7
Cross-validation analysis of model performance in Transforming Travel patterns to high-speed railways.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 score	AUC
BL	0.7404 ±0.006	0.7655 ±0.006	0.7140 ±0.0129	0.7376 ±0.0085	0.7513 ±0.0055	0.8161 ±0.0067
XGBoost	0.8846 ±0.0068	0.9210 ±0.0094	0.8464 ±0.0062	0.8630 ±0.0055	0.8910 ±0.0067	0.9583 ±0.0036
LightGBM	0.8763 ±0.0077	0.9358 ±0.0071	0.8139 ±0.0107	0.8408 ±0.0083	0.8857 ±0.0069	0.9506 ±0.0034
CatBoost	0.8853 ±0.0061	0.9245 ±0.0072	0.8441 ±0.0068	0.8616 ±0.0058	0.8920 ±0.0058	0.9584 ±0.0034

มหาวิทยาลัยเทคโนโลยีสุรนารี

of the dataset, AUC-ROC (Area Under the Receiver Operating Characteristic Curve) was also utilized as a crucial metric. Unlike accuracy, which can be misleading with imbalanced data, AUC-ROC provides a robust evaluation by assessing the trade-off between true positive and false positive rates across different thresholds. This metric is particularly valuable as it remains unaffected by class imbalance and offers a more comprehensive view of the model's performance in distinguishing between minority and majority classes. By combining traditional metrics like accuracy, precision, and recall with AUC-ROC, the study ensures a thorough and reliable assessment of the model's ability to handle imbalanced data.

This comprehensive metric facilitates thorough evaluation across various dimensions, including accuracy, the rates of true positive and true negative predictions, prediction values, imbalance management, and discrimination.

As shown in Tables 5 and 6 and Fig. 4, the efficiency of High-Speed Railways travel behavior prediction varied significantly across models. The Binary Logit Model (BL) achieved training and testing accuracies of 0.7388 and 0.7445, with sensitivities of 0.7646 and 0.7720, and specificities of 0.7118 and 0.7154, respectively. Its precision and F1 score were moderate. In contrast, XGBoost performed notably better, with accuracies of 0.9312 for training and 0.8887 for testing, high sensitivities of 0.9542 and 0.9245, and strong specificities of 0.9070 and 0.8512. Precision and F1 scores were also high, at 0.9151 and 0.9342 for training, and 0.8672 and 0.8949 for testing. LightGBM achieved accuracies of 0.8966 for training and 0.8762 for testing, with high sensitivity of 0.9362 and 0.9299, and specificity of 0.8549 and 0.8197. Precision and F1 scores remained high, with training values of 0.8714 and 0.8962, and testing values of 0.8443 and 0.8755. CatBoost demonstrated the highest efficiency, with accuracies of 0.9461 for training and 0.8970 for testing. Its sensitivity of 0.9643 and 0.9292, and specificity of 0.9270 and 0.8631 were notably high, as were its precision and F1 scores, at 0.9327 and 0.9460 for training, and 0.8771 and 0.8966 for testing. CatBoost outperformed XGBoost and LightGBM, delivering the highest overall predictive efficiency for High-Speed Railways usage.

The close performance between the training and test sets, with the test set surpassing the training set in some metrics, can be explained by several factors. Regularization techniques, commonly used in models like Gradient Boosting, intentionally reduce the training performance to prevent overfitting and improve generalization [84]. Additionally, if k-fold cross-validation is used, the reported training performance is averaged across folds, which may differ from the single test set performance [109]. Finally, overfitting prevention methods, such as early stopping, allow the model to halt training before fully fitting the training data, resulting in better performance on the test set [110].

As shown in Fig. 5, the ROC curve comparison indicates that machine learning models significantly outperform the baseline logistic regression (BL). The Binary Logit Model (BL) has the lowest AUC, indicating moderate performance. XGBoost and LightGBM show strong classification abilities, while CatBoost performs best with the highest AUC, making it the most effective in predicting High-Speed Railways travel mode choices.

As shown in Table 7, the 10-fold Cross-Validation results reveal noticeable differences in model performance when predicting High-Speed Railways travel behavior. The CatBoost model achieved the highest overall efficiency, with an accuracy of 0.8853 ± 0.0061, sensitivity of 0.9245 ± 0.0072, and an AUC score of 0.9584 ± 0.0034. Its specificity of 0.8441 ± 0.0068 and precision of 0.8616 ± 0.0058, combined with a strong F1 score of 0.8920 ± 0.0058, further solidify its role as the most effective model for predicting changes in travel patterns. The XGBoost model followed closely, showing an accuracy of 0.8846 ± 0.0068, which was slightly lower than CatBoost. However, its sensitivity

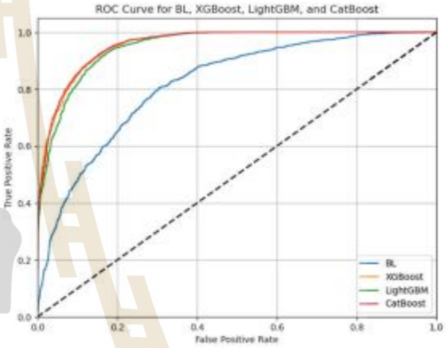


Fig. 5. ROC Curve model.

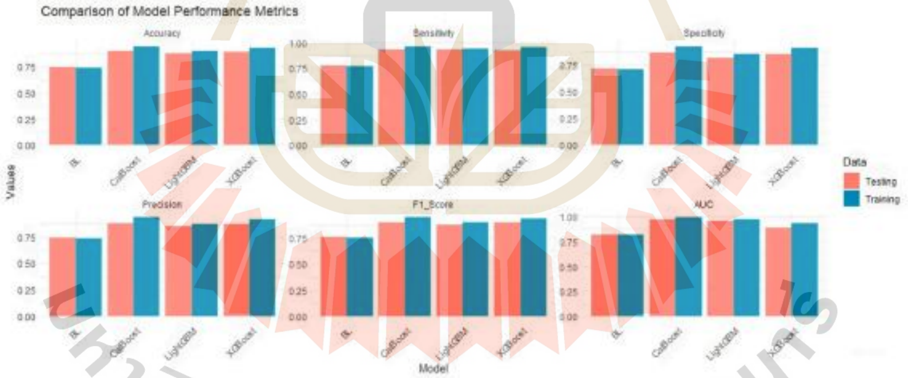


Fig. 4. Comparison model performance metrics.

มหาวิทยาลัยเทคโนโลยีสุรนารี

of 0.9210 ± 0.0094 and AUC of 0.9583 ± 0.0036 indicate strong predictive capability. XGBoost performed similarly in other metrics, including specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and F1 score (0.8910 ± 0.0067). The LightGBM model demonstrated comparable performance, achieving an accuracy of 0.8763 ± 0.0077 and a relatively higher sensitivity of 0.9358 ± 0.0071, surpassing both CatBoost and XGBoost. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. In contrast, the Binary Logit (BL) model displayed significantly lower performance, with an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were also substantially lower compared to the machine learning models, reflecting its struggles with imbalanced data. Overall, the results highlight the significant performance advantages of the machine learning models over BL, with CatBoost leading in most metrics and providing a reliable, consistent approach for predicting High-Speed Railway usage.

3.3. *Analysis result of the binary logit model (BL)*

The analysis of transportation mode using the Binary Logit Model (BL), as shown in Table 8, identified several important and statistically significant variables. For example, the gender variable (GEN) had a β value of −0.0734 and a significance level of 0.013, indicating that males were significantly more likely to choose the new public transportation system, such as High-Speed Railways, compared to females. This finding aligns with a previous study by Ta, Wang [111]. They studied travel behavior and the factors influencing transportation mode choices between males and females in China. They determined that females were more inclined to use new public transportation systems, such as High-Speed Railways, than males. This preference was attributed to females' more positive perceptions and attitudes toward High-Speed Railways, including their safety, convenience, and environmental benefits [55,112,113]. Additionally, the house size variable had a β value of −0.3779 with a significance level of less than 0.001, indicating that individuals with larger house sizes are less likely to travel by High-Speed Railways. This tendency is linked to factors such as higher travel expenses for larger families, the convenience of traveling with a large group, and the management of time and activities for family members. Those living in larger homes often prefer using private cars or other transportation modes over High-Speed Railways because these alternatives can be more cost-effective [114]. The variable for the number of children in a family (CHIL_SIZE) had a β value of 0.9693 and a significance level of less than 0.001, indicating that families with children under 18 years old were more likely to use High-Speed Railways compared to families without children. This preference is likely attributed to the additional considerations for vehicle space and the higher comfort costs associated with traveling with children, which make High-Speed Railways a more appealing option [115]. High-Speed Railways offer considerable advantages in terms of convenience over buses and trains and cost benefits compared to airplanes. The variable for household income (MON_HOUSE) had a β value of 0.0925 and a significance level of less than 0.001, indicating that higher-income households were more likely to use High-Speed Railways. This is attributed to their financial security, which allows them to afford the superior convenience, speed, and quality of service offered by High-Speed Railways compared to other transportation modes. High-income households are more likely to overlook travel costs, valuing the time saved and the ability to engage in other activities, which is a key factor for them in choosing High-Speed Railways [116, 117]. The number of cars in the household had a β value of −0.0476 with a significance level of 0.198, indicating it was not statistically significant. Among the variables related to travel purposes, traveling for study or work (OBJ1) did not show statistical significance. However, traveling for leisure or vacation (OBJ2) had a β value of −0.2670 and a significance level of 0.008, indicating a statistically significant influence on travel decisions. Travelers who prefer leisure or vacation travel are less likely to choose High-Speed Railways, which is consistent with a previous study by Li, Ma [55] This indicates that travelers who seek relaxation are more likely to use trains or buses rather than High-Speed Railways. In contrast, travel for work or business purposes, which is often mandatory or required, typically involves using High-Speed Railways [16]. Travel for shopping (OBJ3) had a significance level of 0.061, indicating it was not statistically significant. In contrast, the variable for traveling between cities (FR_TRA) had a β value of 0.3582 and a significance level of less than 0.001, showing that frequent travelers were more likely to use High-Speed Railways. This is attributed to the convenience and speed of High-Speed Railways, which meet their needs for reliability and time savings. High-Speed Railways reduce travel fatigue and enhance efficiency, leading to greater satisfaction for customers who frequently travel between cities. Additionally, the variable for access time (ACCESS) was considered in this analysis. The variables of waiting time (WAITING), travel time (TRAVEL), and service frequency (Headway) (FR_TRA) all had negative β values with a significance level of less than 0.001. This indicates that if access time, waiting time, travel time, and service frequency for High-Speed Railways were to increase, customers would be more likely to choose High-Speed Railways for their travel. The accessibility of transport stations plays a crucial role in this decision. Regarding access time, Martín, Román [118] have explored strategies for enhancing service by improving station access, either via private car or direct public transport connections. These improvements increase the competitiveness of the public transport system relative to other modes of transportation. The emphasis on saving time while waiting for transportation between cities is valued more than the time spent accessing the station [119]. Waiting time plays a crucial role in customer decision-making and satisfaction [120]. If the waiting time for High-Speed Railways is excessively long, it will likely reduce demand and decrease its competitive edge compared to other transportation modes [121]. Travel time is a crucial factor in travel planning, with travelers often preferring transportation systems that offer shorter travel times [122,123]. Service frequency considerably impacts the competitiveness of the public transport system. A lower headway improves the service level and provides a competitive advantage in the market [124]. The study by Wang, Li [125] has shown that the frequency of High-Speed Railways service effectively segments the market. The cost (COST) has a β value of 2.1997 and a significance level of <0.001, indicating that higher costs are associated with increased travel by High-Speed Railways, consistent with the study by Karmarkar, Jana [126]. A study on the willingness to pay for High-Speed Railways services in India found that travelers were willing to pay an average of 300 rupees, which is higher than their current travel expenses, in exchange for saving time on their journeys. This finding aligns with the study by

Table 8
Model parameter estimates for mode choice.

Variable	β	Standard error	Sig.
CONST	13.6393	0.462	<0.001[a]
GEN	−0.0734	0.029	0.013[a]
HOUSE_SIZE	−0.3779	0.016	<0.001[a]
CHIL_SIZE	0.9693	0.035	<0.001[a]
MON_HOUSE	0.0925	0.019	<0.001[a]
N_CARS	−0.0476	0.037	0.198
OBJ1	0.0112	0.101	0.912
OBJ2	−0.2670	0.101	0.008[a]
OBJ3	−0.2042	0.109	0.061
FR_TRA	0.3582	0.014	<0.001[a]
ACCESS	−3.3402	0.148	<0.001[a]
WAITING	−9.5750	0.471	<0.001[a]
TRAVEL	−14.6052	0.447	<0.001[a]
COST	2.1997	0.195	<0.001[a]
FREQ	−1.0164	0.071	<0.001[a]

[a] Sig. – Significance <0.05.

มหาวิทยาลัยเทคโนโลยีสุ

Shelat, Cats [127], which revealed that travelers were willing to pay more to reduce waiting time and improve travel accuracy. The Binary Logit Model (BL) analysis yielded a Cox and Snell R^2 value of 0.287 and a Nagelkerke R^2 value of 0.383. According to relevant literature, R^2 values between 0.2 and 0.4 are generally considered indicative of a good model fit [128].

3.4. Assessing future Travel mode choice

The evaluation of mode choice is crucial for understanding travelers' economic attributes and the conditions influencing their travel mode decisions based on survey response rankings. The SHAP method offers a deeper analysis of how economic and social attributes, along with travel conditions, affect travel mode choice. This methodology primarily aims to relate each data instance response with travel behavior patterns for more accurate insights [129].

Figs. 6–8 show the mapping of responses to understand the influence of various variables on mode choice. The importance of the input variables, as displayed in Figs. 6–8, was determined by calculating the average of the absolute Shapley values. The input features were then ranked based on their significance, with a higher mean SHAP value indicating greater importance. Fig. 6 presents the XGBoost analysis results, Fig. 7 shows the LightGBM analysis results, and Fig. 8 depicts the CatBoost analysis results, revealing similar variable prioritization across the models. The most important factors for choosing High-Speed Railways in the future were travel time, service frequency, access time, travel cost, waiting time, number of children in the family, frequency of inter-provincial travel per year, number of household members, household income, private car ownership, travel purpose for study/work, gender, travel purpose for leisure/vacation, and travel purpose for shopping, respectively. The machine learning analysis of important factors for High-Speed Railways choice in the future aligned with the results obtained using the Binary Logit Model (BL) method.

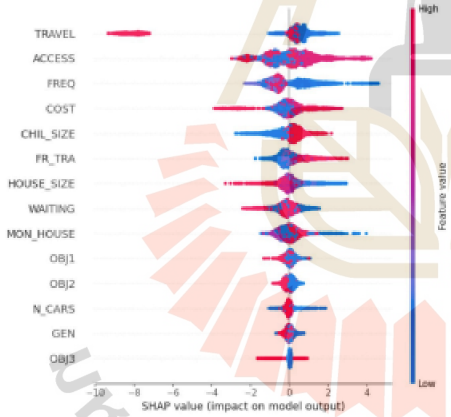


Fig. 6. Impact of input features on High-Speed Railways usage prediction using SHAP with XGBoost.

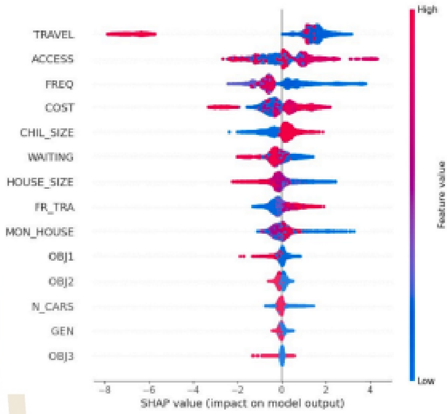


Fig. 7. Impact of input features on High-Speed Railways usage prediction using SHAP with LightGBM.

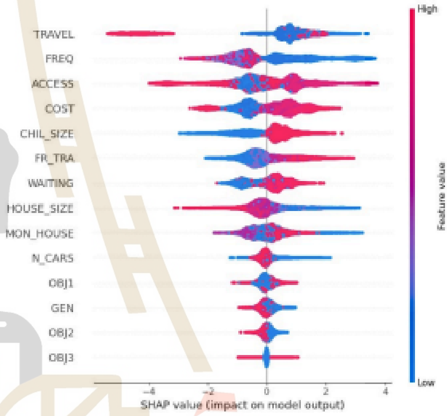


Fig. 8. Impact input features on High-Speed Railways usage prediction using SHAP with CatBoost.

The decision to use High-Speed Rail in Thailand is influenced by several key factors, as highlighted by the SHAP analysis in Figs. 6–8. Travel time is crucial, as High-Speed Rail offers a substantial time-saving advantage over conventional buses and trains, which are often slower and more prone to delays. Passengers who value time efficiency, especially frequent inter-provincial travelers, find High-Speed Rail particularly attractive in Thailand, where long bus rides and traffic congestion are common. However, if the time savings offered by High-Speed Rail are not as great as air travel or if passengers prefer the flexibility of buses or personal vehicles, some may opt for these alternatives. Another significant factor is access time to High-Speed Rail stations. Longer access times can reduce the appeal of High-Speed Rail, especially for passengers in rural areas with less developed public transportation infrastructure. Conversely, passengers in urban areas or those prioritizing speed, comfort, and reliability may still choose High-Speed Rail, even if access times are longer compared to other modes of transport like buses or conventional trains, which typically offer more flexible access points. Service frequency also plays a critical role in passengers' decisions. When High-Speed Rail services are infrequent or the intervals between departures are too long, passengers may turn to buses, which operate

มหาวิทยาลัยเทคโนโลยีสุรนารี

more frequently and offer greater flexibility. This is particularly relevant in Thailand, where buses are a common mode of transportation between provinces and have flexible schedules that accommodate passengers' travel needs, especially for last-minute trips. Cost is another key factor influencing passenger behavior. While some passengers, particularly those with higher incomes, are willing to pay more for the speed, comfort, and reliability of High-Speed Rail, others are more price-sensitive and opt for cheaper alternatives like buses or traditional trains. When High-Speed Rail fares are perceived as too high, passengers often choose buses, which are much less expensive, or trains if speed is not a priority. Waiting time between High-Speed Rail services is a major deciding factor. Long waiting times make High-Speed Rail less attractive, particularly for passengers who value flexibility and frequent departures. In Thailand, where buses and airplanes offer more frequent services, passengers may find these alternatives more convenient. Moreover, longer waiting times can disrupt travel plans and extend the overall journey duration, making High-Speed Rail less appealing to those who prioritize quick and efficient travel.

Household income higher earners may opt for it because it saves them time, but passengers with lesser incomes may not believe that purchasing more expensive tickets is worthwhile. Additionally, having an automobile has a variety of effects. While some passengers who own cars prefer to be able to drive, others could choose to take the High-Speed Rail in order to escape traffic and parking issues. Gender data indicates that women are more likely than males to utilize high-speed trains, most likely due to their awareness of the benefits' security and comfort. Passengers traveling for business or study show a range of behaviors when it comes to their travel habits. Some may prioritize money and select less expensive options, while others may value the time efficiency of High-Speed Rail. Tourists and leisure travelers prefer to steer clear of high-speed rail, probably due to the variable nature of vacation travel, where flexibility in scheduling and cost-effectiveness trump speed. Similarly, shopping trips result in mixed behavior, with some passengers choosing High-Speed Rail for convenience and others prioritizing flexibility or cost savings.

In Fig. 9, the importance of each feature was determined by calculating the average absolute values of Shapley across the dataset. The input features were ranked by their significance based on the mean SHAP values, with higher values indicating greater importance. Additionally, the figure emphasizes the significance of each input variable for the High-Speed Railways mode.

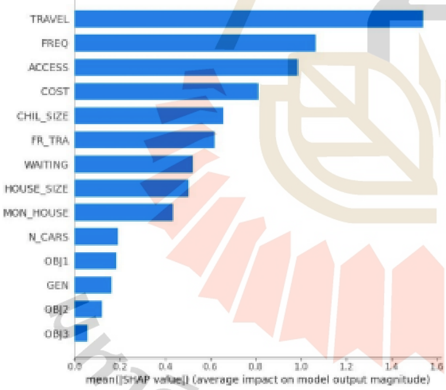


Fig. 9. Feature importance for High-Speed Railways usage using SHAP.

4. Discussion

The comparison between the four models shows that the Binary Logit Model had the lowest efficiency in analyzing travel mode choices, primarily due to its limitations in handling complex, imbalanced datasets. Binary Logit Mode, being a traditional model, often struggles with non-linear relationships and data variability [73]. In contrast, the three machine learning models—XGBoost, LightGBM, and CatBoost—demonstrated superior efficiency, with CatBoost outperforming the others. The strength of CatBoost lies in its ability to handle imbalanced datasets effectively, a common challenge in travel mode prediction studies. Its resilience to overfitting further enhances its utility, making it particularly suitable for complex datasets like those used in this analysis. Moreover, CatBoost's robust handling of missing data and integration with under sampling techniques provides a more balanced and accurate prediction of less common travel modes, while maintaining interpretability and accuracy across the model [130].

The results from the CatBoost model underline that travel time plays a decisive role in the choice of High-Speed Railways as a new mode of transportation. The significant reduction in travel time is a key factor driving passengers to opt for High-Speed Rail, especially in Thailand, where conventional modes like buses or trains are often hindered by traffic congestion and delays. This advantage makes High-Speed Rail a highly efficient and attractive alternative, particularly during peak hours when time efficiency is crucial. As previous studies have shown, passengers tend to prioritize faster travel options that minimize travel time [131–133].In addition to travel time, access time is another critical factor. The convenience of accessing High-Speed Rail stations, especially in urban areas, significantly improves passenger preference. As Zhang, Liang [134] pointed out, reliable and predictable access to stations enhances overall convenience, making High-Speed Rail more appealing. However, in more rural or less developed regions, longer access times can reduce the attractiveness of High-Speed Rail as an option, which is why improving station accessibility remains a key policy recommendation [135]. Service frequency further impacts passengers' choices. Short intervals between train arrivals improve service efficiency and reduce waiting times, which are critical in maintaining the system's competitive edge over other modes like buses or airplanes. If services are too infrequent, passengers may opt for more flexible options like buses, which run more frequently, especially between provinces [133]. The cost of High-Speed Railways services is also a significant factor. Passengers with higher incomes are more willing to pay for faster, more reliable transportation, as studies in other countries like India and Italy have shown [126,136]. However, for price-sensitive passengers, especially in developing regions, the higher cost of High-Speed Rail could deter them from using the service. Balancing affordability with service quality is thus essential to increase ridership across diverse income groups. Waiting time between services is crucial in shaping the passenger experience. Long waiting times negatively affect satisfaction and reduce the likelihood of passengers opting for High-Speed Rail, particularly when more frequent and flexible alternatives like buses and airplanes are available [137]. Managing wait times efficiently not only enhances the travel experience but also boosts the competitiveness of High-Speed Railways by ensuring smoother operations and improved passenger flow.

Family structure is one of the most important social and economic elements influencing travel behavior in Thailand. Compared to bigger families, smaller families travel more frequently, especially for leisure purposes. In contrast, larger households with higher incomes are more likely to own private cars, influencing their mode of travel. The size of the home, the family's income, whether or not a car is owned, and whether or not there are small children or elderly family members all have an impact on travel behavior [116]. For instance, because of their convenience and safety characteristics, families with children are thought to find particular value in High-Speed Railways (HSR). In Thailand, where traffic congestion frequently makes long-distance

ทยาลัยเทคโนโลยีสุ

travel difficult, High-Speed Rail (HSR) offers a quick and effective substitute. This makes it appealing to families looking for dependable modes of transportation. High-Speed Railways is a better option for families who value safety and punctuality because it is more socially and economically beneficial than traditional trains and has higher traffic efficiency [138]. Moreover, choosing to use high-speed rail is heavily influenced by household income. Families with lower incomes also express a desire to utilize High-Speed Railways because it provides greater reliability, safety, and comfort, even though passengers with higher incomes are more likely to choose it because of its time-saving advantages. These advantages are especially crucial in places like Thailand, where traffic and accidents frequently make road travel uncertain. For special trips where safety and dependability are more important than cost, like family holidays or necessary travel, lower-income families may consider High-Speed Railways to be worth the additional expense. This indicates that although cost is still an obstacle, for some population segments the benefit offered by High-Speed Railways can outweigh the increased price [139].

Travel preferences are influenced by owning a car. While some people choose to use High-Speed Railways to avoid traffic and parking issues, others, especially in large cities like Bangkok, like having the freedom to drive. Behavior is influenced by gender as well; women are more likely to use High-Speed Railways because it feels secure and comfortable. In Thailand, where concerns about safety are common, this makes High-Speed Railways more alluring to female travelers. Business and study travelers have different preferences when it comes to transportation. While some choose more affordable options, others enjoy High-Speed Railways ˋs ability to save time, especially when traveling from an urban location to a rural one. Meanwhile, High-Speed Railways is usually avoided by tourists and leisure visitors in favor of more flexible and cost-effective options like buses or private cars. High-Speed Railways networks offer quick and effective transportation, making it possible for frequent travelers to go where theyˋre going without difficulty, as Huang, Song [139] further highlights. When it comes to passengers with tight schedules, including those who are going for business or school, High-Speed Railways ˋs timeliness guarantees dependable travel. Because of its dependability, High-Speed Railways is a competitive choice for long-distance or intercity travel, particularly in Thailand where road traffic can be unpredictable. In conclusion, considering its benefits in terms of speed, safety, and dependability, High-Speed Railways is a desirable alternative for families, women, and those with better incomes in Thailand. Lower-class visitors may find High-Speed Railways less appealing because of its expense, but many of them still think itˋs a good choice for special occasions because of its total value. While travelers and those looking for flexibility could favor less expensive options, measures to alleviate these socioeconomic gaps and enhance accessibility to High-Speed Railways might encourage use among a wider range of income levels.

5. Conclusion and policy implications

Thailandˋs collaboration with China to develop High-Speed Rail (HSR) is a crucial step in enhancing transportation infrastructure and boosting regional economic growth. While most research focuses on existing travel modes like cars, buses, and planes, thereˋs a lack of analysis on future transport choices, particularly in developing nations like Thailand. This study addresses this gap by analyzing factors influencing High-Speed Railways adoption, such as access time, service frequency, and socioeconomic characteristics, to shape future traveler decisions.

The comparison of models for predicting High-Speed Rail usage revealed that CatBoost consistently outperformed the other models, with cross-validation confirming its superior performance across all key metrics. The Binary Logit Model (BL) demonstrated moderate effectiveness, achieving an accuracy of 0.7404 ± 0.006, sensitivity of 0.7655 ± 0.006, and a relatively lower AUC of 0.8161 ± 0.0067. Its specificity (0.7140 ± 0.0129), precision (0.7376 ± 0.0085), and F1 score (0.7513 ± 0.0055) were moderate, but it struggled with handling imbalanced data. In contrast, XGBoost showed significantly stronger results, with an accuracy of 0.8846 ± 0.0068, a sensitivity of 0.9210 ± 0.0094, and an AUC of 0.9583 ± 0.0036. XGBoost also delivered high specificity (0.8464 ± 0.0062), precision (0.8630 ± 0.0055), and F1 score (0.8910 ± 0.0067). LightGBM also performed well, with an accuracy of 0.8763 ± 0.0077 and a relatively high sensitivity of 0.9358 ± 0.0071. However, it exhibited lower specificity (0.8139 ± 0.0107) and precision (0.8408 ± 0.0083), resulting in a slightly reduced F1 score of 0.8857 ± 0.0069 and an AUC of 0.9506 ± 0.0034. CatBoost demonstrated the highest overall performance, with its ability to handle imbalanced datasets, avoid overfitting, and provide clear interpretability through SHAP values, making it the most effective model for predicting High-Speed Rail usage, outperforming XGBoost, LightGBM, and BL. The relatively higher performance of CatBoost can be generalizable to similar studies in different regions or countries, depending on the structure and characteristics of the dataset in each region, such as social and demographic diversity, transportation options, and regional transportation policies.

The decision to use High-Speed Rail (High-Speed Railways) in Thailand is influenced by several key factors, reflecting the unique transportation context of the country. Travel time is a major advantage, with High-Speed Railways offering faster journeys compared to buses and conventional trains, which are prone to delays, making it appealing to frequent inter-provincial travelers. However, if time savings are not significant compared to air travel or if passengers prefer the flexibility of buses or private cars, High-Speed Railways may not be their preferred choice. Access time to stations also plays a critical role, especially for rural passengers where public transport infrastructure is less developed, though urban travelers may still prioritize High-Speed Railways for its speed and reliability. Mixed responses regarding access time are observed; some passengers still choose High-Speed Railways despite longer station access due to its overall comfort and speed, while others prefer buses or personal vehicles for greater flexibility. Service frequency is another crucial factor; infrequent departures could drive passengers to opt for buses, which provide more flexible schedules. Cost, whether high or low, leads to mixed behavior: some passengers are willing to pay more for the convenience and time savings High-Speed Railways offers, while others may opt for cheaper alternatives regardless of High-Speed Railways´s affordability. Waiting time between services can further dissuade passengers, especially in Thailand, where buses and airplanes offer more frequent options. Household income and automobile ownership also influence preferences, with higher earners and those without cars more inclined to choose High-Speed Railways. Women are more likely than males to select High-Speed Railways in Thailand, probably because of its comfort and safety characteristics. This suggests that gender plays a part. The reason for the trip also counts. While consumers and leisure travelers have mixed tastes and frequently value flexibility and cost savings over speed, business and study commuters prefer High-Speed Railways for its time efficiency. The use of High-Speed Railways in Thailand is influenced by a number of factors that together show the population´s varied requirements and habits.

By taking into account elements related to travel behavior, policymakers can gain a better understanding of how people make transportation decisions and develop solutions that can be customized to support high-speed rail and remove adoption hurdles. Policymakers can more successfully focus policies to promote sustainable transportation choices and remove barriers to HSR adoption by understanding reactions to various circumstances.To effectively promote High-Speed Rail (HSR) usage in Thailand, several key policy actions can be implemented.

- Enhancing Confidence in High-Speed Railways: Publicizing data on reduced travel times compared to cars and other modes will build public trust in the speed and punctuality of High-Speed Railways.

ทยาลัยเทคโนโลยีสุร

- Increasing Service Frequency: Adjusting High-Speed Railways schedules to meet peak demand during mornings, evenings, and weekends, and adding extra services during festivals, will reduce waiting times and enhance service availability.
- Competitive Pricing: Offering discount packages for groups like students, families, and seniors, and seasonal promotions, will make High-Speed Railways more attractive, especially compared to other transport options.
- Reducing Station Waiting Time: Streamlining boarding processes through more efficient ticketing systems, online reservations, and real-time updates on train schedules will improve passenger experience and reduce delays.
- Developing Infrastructure Connections: Expanding public transport links (buses, metro) to High-Speed Railways stations will improve accessibility, particularly from community areas, encouraging more people to use High-Speed Railways.
- Family-Friendly Services: Providing family-oriented services like child-friendly areas, diaper-changing stations, and family travel discounts will make High-Speed Railways more appealing to families.
- Targeting Car Users: Highlighting benefits like reduced travel fatigue, lower accident risks, and long-distance cost savings will encourage car users to switch to High-Speed Railways. Offering secure parking options at stations will facilitate this shift.
- Ongoing Public Relations Campaigns: A continuous public relations effort is essential to raise awareness of High-Speed Railways 's benefits and address public concerns, ensuring widespread adoption of this mode of travel.

While these policies can promote High-Speed Railways, there are practical limitations to consider. Financial constraints, such as funding for discounts and infrastructure improvements, must be addressed. Increasing train frequency is challenging due to physical infrastructure limits. Human resource management, legal approvals, and compliance with environmental regulations must also be considered. Regular data collection and evaluation will ensure that policies are adapted to passenger needs, maximizing the long-term success of High-Speed Railways in Thailand.

5.1. Limitations and further research

This study has several limitations. First, the use of hypothetical scenarios and reliance on estimated data, as the High-Speed Railways infrastructure was not yet constructed, may introduce uncertainty into the findings. Additionally, the absence of real-world behavioral data and potential variations in travel preferences over time could affect the generalizability of the results. There are also potential biases inherent in the self-administered questionnaires used in the study, and the short data collection period may not fully capture long-term travel behavior changes. Furthermore, the reliance on a single method of data collection may limit the comprehensiveness of the findings, particularly as the study focused on travel mode choice during a specific timeframe. These factors may influence the accuracy and generalizability of the models used, especially as transportation preferences evolve due to economic or infrastructural changes.

For future research, it is recommended to integrate a variety of data sources, such as real-time tracking technologies (e.g., UAV-based object tracking) and longer-term data collection. This approach would provide more dynamic insights and enhance the predictive power of machine learning models. Additionally, future studies could explore the broader economic and environmental impacts of High-Speed Railways development in Thailand, incorporating post-launch customer feedback surveys to assess satisfaction and service effectiveness. Such data would be invaluable for refining transportation policies and improving the High-Speed Railways services to better meet the evolving needs of the population.

CRediT authorship contribution statement

Chinnakrit Banyong: Writing – original draft, Methodology, Formal analysis, Data curation. **Natthaporn Hantanong:** Data curation, Conceptualization. **Panuwat Wisutwattanasak:** Writing – review & editing, Validation, Formal analysis. **Thanapong Champahom:** Visualization, Validation. **Kestsirin Theerathitichaipa:** Formal analysis, Data curation. **Manlika Seefong:** Formal analysis, Data curation. **Vatanavongs Ratanavaraha:** Supervision, Software. **Sajjakaj Jomnonkwao:** Writing – review & editing, Visualization, Supervision, Conceptualization.

Institutional review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Suranaree University of Technology (COE No 9/2567, February 5, 2024).

Funding

This work was supported by Suranaree University of Technology (SUT), Thailand Science Research and Innovation, and National Science, Research and Innovation Fund (NSRF) (Project code: 195602).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their gratitude to the Suranaree University of Technology (SUT), Thailand Science Research and Innovation, and National Science, Research and Innovation Fund for their support in undertaking this research.

Data availability

The authors do not have permission to share data.

References

[1] K. Tissayakorn, A study on transit-oriented development strategies in the high-speed rail project in Thailand (2021).
[2] M. Jin, et al., Impacts of high-speed railways on economic growth and disparity in China, Transport. Res. Pol. Pract. 138 (2020) 158–171.
[3] M.R. Aalami, et al., A robust finite element analysis of the rail-wheel rolling contact, Adv. Mech. Eng. 5 (2013) 272350.
[4] M.T. Kashifi, et al., Predicting the travel mode choice with interpretable machine learning techniques: a comparative study, Travel Behaviour and Society 29 (2022) 279–296.
[5] Y. Liang, et al., Effectiveness of high-speed railway on regional economic growth for less developed areas, J. Transport Geogr. 82 (2020) 102621.
[6] W. Sun, et al., High-speed rail network expansion and its impact on regional economic sustainability in the Yangtze River Delta, China, 2009–2018, Sustainability 14 (1) (2021) 155.
[7] X. Chang, et al., Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data, Transportmetrica: Transport Science 15 (2) (2019) 1587–1612.
[8] E. Heinen, et al., Does exposure to new transport infrastructure result in modal shifts? Patterns of change in commute mode choices in a four-year quasi-experimental cohort study, J. Transport Health 6 (2017) 396–410.
[9] S. Liu, et al., Evaluating the impact of new lines on entrance/exit passenger flow of adjacent existing stations in urban rail transit system, Transport. Res. Procedia 25 (2017) 2625–2638.
[10] W. Yang, Q. Chen, J. Yang, Factors affecting travel mode choice between high-speed railway and road passenger transport—evidence from China, Sustainability 14 (23) (2022) 15745.
[11] A.Y. Nurhidayat, et al., Research on passengers' preferences and impact of high-speed rail on air transport demand, Sustainability 15 (4) (2023) 3060.
[12] X. Li, et al., Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost, Appl. Geogr. 166 (2024) 103264.

มหาวิทยาลัยเทคโนโลยีสุรนารี

[13] P.F. Belgiawan, A. Ilahi, K.W. Axhausen, Influence of pricing on mode choice decision in Jakarta: a random regret minimization model, Case Studies on Transport Policy 7 (1) (2019) 87–95.

[14] E. Mirzaei, R. Kheyroddin, D. Mignot, Exploring the effect of the built environment, weather condition and departure time of travel on mode choice decision for different travel purposes: evidence from Isfahan, Iran, Case studies on transport policy 9 (4) (2021) 1419–1430.

[15] S. Gross, B. Grimm, Sustainable mode of transport choices at the destination–public transport at German destinations, Tourism Review 73 (3) (2010) 401–420.

[16] X. Li, et al., Assessing intercity multimodal choice behavior in a touristy city: a factor analysis, J. Transport Geogr. 86 (2020) 102776.

[17] H. Cui, et al., Investigating the impacts of urban–rural bus service quality on rural residents' travel choices using an SEM-MNL integration model, Sustainability 15 (15) (2023) 11950.

[18] J. Ning, T. Lyu, Y. Wang, Exploring the built environment factors in the metro that influence the ridership and the market share of the elderly and students, J. Adv. Transport. 2021 (1) (2021) 9966794.

[19] A. Sharma, et al., Travel behavior modeling: taxonomy, challenges, and opportunities, Int. J. Adv. Comput. Sci. Appl. 12 (5) (2021) 774–786.

[20] Y. Jiang, et al., Ground access behaviour of air-rail passengers: a case study of Dalian ARIS, Travel Behaviour and Society 24 (2021) 152–163.

[21] M.M. Pasha, M.D. Hickman, C.G. Prato, Modeling mode choice of air passengers' ground access to brisbane airport, Transport. Res. Rec. 2674 (11) (2020) 756–767.

[22] G. Gunay, I. Gokasar, Market segmentation analysis for airport access mode choice modeling with mixed logit, J. Air Transport. Manag. 91 (2021) 102001.

[23] I.H. Sarker, AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems, SN Computer Science 3 (2) (2022) 158.

[24] A. Abulibdeh, Analysis of mode choice affects from the introduction of Doha Metro using machine learning and statistical analysis, Transp. Res. Interdiscip. Perspect. 20 (2023) 100852.

[25] F. Wang, C.L. Ross, Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model, Transport. Res. Rec. 2672 (47) (2018) 35–45.

[26] A.M. Darwish, et al., Sensitivity evaluation of machine learning-based calibrated transportation mode choice models: a case study of Alexandria City, Egypt, Transp. Res. Interdiscip. Perspect. 24 (2024) 101052.

[27] M. Elharoun, S.M. El-Badawy, U.E. Shahdah, Artificial intelligence techniques for predicting individuals' mode choice behavior in Mansoura city, Egypt, Transport. Res. Rec. 2677 (10) (2023) 605–623.

[28] J. Díaz-Ramírez, J.A. Estrada-García, J. Figueroa-Sayago, Predicting transport mode choice preferences in a university district with decision tree-based models, City and Environment Interactions 20 (2023) 100118.

[29] P. Salas, et al., A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity, Expert Syst. Appl. 193 (2022) 116253.

[30] E.-J. Kim, Analysis of travel mode choice in Seoul using an interpretable machine learning approach, J. Adv. Transport. 2021 (1) (2021) 6685004.

[31] X. Zhao, et al., Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and logit models, Travel behaviour and society 20 (2020) 22–35.

[32] L. Cheng, et al., Applying a random forest method approach to model travel mode choice behavior, Travel behaviour and society 14 (2019) 1–10.

[33] D. Wu, J.C. Martín, Research on passengers' preference for high-speed railways (HSRs) and high-speed trains (HSTs), Sustainability 14 (2022), https://doi.org/10.3390/su14031473.

[34] M. Givoni, Development and impact of the modern high-speed train: a review, Transport Rev. 26 (5) (2006) 593–611.

[35] E. Cascetta, et al., Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: ten years ex post evaluation and future perspectives, Transport. Res. Pol. Pract. 139 (2020) 412–428.

[36] F.A. Armah, D.O. Yawson, A.A.N.M. Pappoe, A systems dynamics approach to explore traffic congestion and air pollution link in the city of accra, Ghana, Sustainability 2 (2010) 252–265, https://doi.org/10.3390/su2010252.

[37] X. Yu, et al., An empirical study on the design of China high-speed rail express train operation plan—from a sustainable transport perspective, Sustainability 10 (7) (2018) 2478.

[38] O. Puan, et al., Transportation mode choice binary logit model: a case study for Johor Bahru city, in: IOP Conference Series: Materials Science and Engineering, IOP Publishing, 2019.

[39] S. Wójcik, The determinants of travel mode choice: the case of Łódź, Poland, Bull. Geogr. (44) (2019) 93–101.

[40] J. Li, K. Lo, M. Guo, What influences the choice between private car and public transport for ShoppingTrips? Impact of socio-economic and built environment factors, Journal of Asian Energy Studies 2 (1) (2018) 28–42.

[41] C.S. Pitombo, et al., A two-step method for mode choice estimation with socioeconomic and spatial information, Spatial Statistics 11 (2015) 45–64.

[42] Juan de Dios Ortúzar, L.G. Willumsen, Trip generation modelling, Modelling Transport (2011) 139–173.

[43] F. Calvo, et al., Factors influencing trip generation on metro system in Madrid (Spain), Transport. Res. Transport Environ. 67 (2019) 156–172.

[44] J. Sresakoolchai, S. Kaewunruen, Comparative studies into public private partnership and traditional investment approaches on the high-speed rail project linking 3 airports in Thailand, Transp. Res. Interdiscip. Perspect. 5 (2020) 100116.

[45] W. Weerawat, L. Samitthiwantikul, R. Torpanya, Operational challenges of the Bangkok airport rail link, Urban rail transit 6 (2020) 42–55.

[46] W. Rattanakijsuntorn, et al., Success factors and partnership evaluation of air–rail integration development: a case of a high-speed rail project linking three airports in Thailand, Infrastructures 9 (7) (2024) 115.

[47] P. Bhavsar, et al., Machine learning in transportation data analytics, in: Data Analytics for Intelligent Transportation Systems, Elsevier, 2017, pp. 283–307.

[48] S. Khan, A. Adnan, N. Iqbal, Applications of artificial intelligence in transportation, in: 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), IEEE, 2022.

[49] R. Abduljabbar, et al., Applications of artificial intelligence in transport: an overview, Sustainability 11 (1) (2019) 189.

[50] Department of International Trade Promotion, High-speed rail 4 routes, Available from: https://www.rentconnected.com/blogs/high-speed-rail-4-routes-for-thailand/, 2023.

[51] Office of the national economic and social development council, Gross regional and provincial Product Chain Volume measure 2022 edition, Available from: https://www.nesdc.go.th/ewt_dl_link.php?nid=15104&filename=gross_regional, 2023.

[52] A. Srithongrung, K.A. Kriz, Thai public capital budget and management process, in: Capital Management and Budgeting in the Public Sector, IGI Global, 2019, pp. 206–235.

[53] M. Habib, K. Poole, S. McCarthy NovaTRAC, A multi-year travel survey data collection program, in: The 12th International Conference on Transport Survey Methods, 2022. Lisbon, Portugal.

[54] M. Al-Atroush, Z. Mustaffa, T. Sebeay, Emerging trends in overcoming the weather barrier to sustainable mobility in gulf and tropical cities, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2022.

[55] X. Li, et al., Influence of weather conditions on the intercity travel mode choice: a case of Xi'an, Comput. Intell. Neurosci. 2021 (1) (2021) 9969322.

[56] T. Otim, et al., Modeling the impact of weather and context data on transport mode choices: a case study of gps trajectories from beijing, Sustainability 14 (10) (2022) 6042.

[57] M. Pavlou, et al., An evaluation of sample size requirements for developing risk prediction models with binary outcomes, BMC Med. Res. Methodol. 24 (1) (2024) 146.

[58] R. Kujala, et al., Travel times and transfers in public transport: comprehensive accessibility analysis based on Pareto-optimal journeys, Comput. Environ. Urban Syst. 67 (2018) 41–54.

[59] A.I. Arencibia, et al., Modelling mode choice for freight transport using advanced choice experiments, Transport. Res. Pol. Pract. 75 (2015) 252–267.

[60] S. Birolini, et al., Access mode choice to low-cost airports: evaluation of new direct rail services at Milan-Bergamo airport, Transport Pol. 73 (2019) 113–124.

[61] N. Becker, N. Carmi, Changing trip behavior in a higher education institution: the role of parking fees, International Journal of Sustainable Transportation 13 (4) (2019) 268–277.

[62] D. Nam, J. Cho, Deep neural network design for modeling individual-level travel mode choice behavior, Sustainability 12 (18) (2020) 7481.

[63] Y. Guo, et al., China's millennial car travelers' mode shift responses under congestion pricing and reward policies: a case study in Beijing, Travel Behaviour and Society 23 (2021) 86–99.

[64] J. Tarriño-Ortiz, et al., Analyzing the impact of low emission zones on modal shift, Sustain. Cities Soc. 77 (2022) 103562.

[65] S.R.o. Thailand (Ed.), State Railway of Thailand, *COMPLETE GUIDE TO SRT*, 2019, 4.0.

[66] POST TO DAY, Transportation" expects the ticket price of the Bangkok-Chiang Mai high-speed train to run for 3 hours and a half 1,200 baht [cited 2024 10 January]; Available from: https://www.posttoday.com/business/531218, 2017.

[67] Economic Base, Knock on the Bangkok-Chiang Mai high-speed train fare 1,089 baht [cited 2024 10 January]; Available from: https://www.thansettakij.com/business/242231, 2017.

[68] BusOnlineTicket.co, Book Thailand Bus Tickets Online (2024) [cited 2024 10 January]; Available from: https://www.busonlineticket.co.th/.

[69] State railway of Thailand *the state Railway of Thailand Easy book, Convenient check* [cited 2024 10 January]; Available from: https://dticket.railway.co.th/DTicketPublicWeb/home/Home, 2024.

[70] AirAsia move, Flight (2024) [cited 2024 10 January]; Available from: https://www.airasia.com/th/th.

[71] D.A. Hensher, J.M. Rose, W.H. Greene, Applied Choice Analysis: a Primer, Cambridge university press, 2005.

[72] T. Ghosh, N. Nagaraj, Evaluating the determinants of mode choice using statistical and machine learning techniques in the Indian megacity of Bengaluru, arXiv preprint arXiv:2401.13977 (2024).

[73] L. Tang, et al., Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis, IEEE Intell. Transp. Syst. (2024).

[74] T. Shafighfard, et al., Machine-learning methods for estimating compressive strength of high-performance alkali-activated concrete, Eng. Appl. Artif. Intell. 136 (2024) 109053.

[75] F. Kazemi, et al., Machine-learning methods for estimating performance of structural concrete members reinforced with fiber-reinforced polymers, Arch. Comput. Methods Eng. (2024) 1–33.

[76] F. Bagherzadeh, et al., Prediction of maximum tensile stress in plain-weave composite laminates with interacting holes via stacked machine learning algorithms: a comparative study, Mech. Syst. Signal Process. 195 (2023) 110315.

[77] V. Dahmen, S. Weikl, K. Bogenberger, Interpretable machine learning for mode choice modeling on tracking-based revealed preference data, Transport. Res. Rec. (2023) 03611981241246973.

[78] E.E. Başakın, et al., A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment, Neural Comput. Appl. 34 (1) (2022) 783–812.

[79] Y. Zouzou, H. Citakoglu, General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration, Acta Geophysica 71 (2) (2023) 927–947.

[80] B. So, Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM, Scand. Actuar. J. (2024) 1–23.

[81] L. Zhu, S. Shu, L. Zou, XGBoost-based travel time prediction between bus stations and analysis of influencing factors, Wireless Commun. Mobile Comput. 2022 (1) (2022) 3504704.

[82] J. Yu, et al., Combining travel behavior in metro passenger flow prediction: a smart explainable Stacking-Catboost algorithm, Inf. Process. Manag. 61 (4) (2024) 103733.

[83] X. Zhu, et al., Research on the prediction and influencing factors of heavy duty truck fuel consumption based on LightGBM, Energy 296 (2024) 131221.

[84] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.

[85] Y. Xu, et al., Research on a mixed gas classification algorithm based on extreme random tree, Appl. Sci. 9 (9) (2019) 1728.

[86] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).

[87] G. Ke, et al., Lightgbm: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017).

[88] C. Ding, X.J. Cao, P. Næss, Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo, Transport. Res. Pol. Pract. 110 (2018) 107–117.

[89] J. Elith, J.R. Leathwick, T. Hastie, A working guide to boosted regression trees, J. Anim. Ecol. 77 (4) (2008) 802–813.

[90] L. Prokhorenkova, et al., CatBoost: unbiased boosting with categorical features, Adv. Neural Inf. Process. Syst. 31 (2018).

[91] A.V. Dorogush, V. Ershov, A. Gulin, CatBoost: gradient boosting with categorical features support (2018) arXiv preprint arXiv:1810.11363.

[92] W. Zhai, et al., CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values, Comput. Electron. Agric. 214 (2023) 108306.

[93] T.D. Pham, et al., Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the red river delta biosphere reserve, Vietnam, Rem. Sens. 12 (8) (2020) 1334.

[94] G. Huang, et al., Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions, J. Hydrol. 574 (2019) 1029–1041.

[95] H. Hong, et al., GIS-based landslide spatial modeling in Ganzhou City, China, Arabian J. Geosci. 9 (2016) 1–26.

[96] S. Lapuschkin, et al., Unmasking Clever Hans predictors and assessing what machines really learn, Nat. Commun. 10 (1) (2019) 1096.

[97] S. Mangalathu, S.-H. Hwang, J.-S. Jeon, Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach, Eng. Struct. 219 (2020) 110927.

[98] H. Mao, et al., Driving safety assessment for ride-hailing drivers, Accid. Anal. Prev. 149 (2021) 105574.

[99] R. Adland, et al., The value of meteorological data in marine risk assessment, Reliab. Eng. Syst. Saf. 209 (2021) 107480.

[100] G.M. Vega, L. Aznarte José, Shapley additive explanations for NO2 forecasting, Ecol. Inf. 56 (2020) 101039.

[101] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[102] S.M. Lundberg, et al., From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 56–67.

[103] T. Shafighfard, et al., Chained machine learning model for predicting load capacity and ductility of steel fiber–reinforced concrete beams, Comput. Aided Civ. Infrastruct. Eng. (2024).

[104] T. Champahom, et al., Factors influencing the willingness to pay for motorcycle safety improvement: a structural equation modeling approach, Transp. Res. Interdiscip. Perspect. 22 (2023) 100950.

[105] Y. Zhang, A. Haghani, A gradient boosting method to improve travel time prediction, Transport. Res. C Emerg. Technol. 58 (2015) 308–324.

[106] A. Fernández, et al., Learning from Imbalanced Data Sets, vol. 10, Springer, 2018.

[107] S. Mokhtarimousavi, et al., Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks, International journal of transportation science and technology 9 (2) (2020) 100–115.

[108] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009) 427–437.

[109] D. Krstajic, et al., Cross-validation pitfalls when selecting and assessing regression and classification models, J. Cheminf. 6 (2014) 1–15.

[110] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorob. 7 (2013) 21.

[111] N. Ta, et al., Gender difference in commuting travel: a comparative study of suburban residents in Beijing and Shanghai, Travel behaviour and society 28 (2022) 196–203.

[112] F. Chen, X. Hao, Z. Chen, Can high-speed rail improve health and alleviate health inequality? Evidence from China, Transport Pol. 114 (2021) 266–279.

[113] J. Xie, et al., High-speed rail services for elderly passengers: ticket-booking patterns and policy implications, Transport Pol. 125 (2022) 96–106.

[114] O. Rubin, C.H. Mulder, L. Bertolini, The determinants of mode choice for family visits – evidence from Dutch panel data, J. Transport Geogr. 38 (2014) 137–147.

[115] R. Dowling, Parents, children and automobility: trends, challenges and opportunities, Handbook on transport and development (2015) 526–538.

[116] P. Zhao, D. Yuan, Effects of family structure on travel behaviour, in: Population Growth and Sustainable Transport in China, Springer, 2023, pp. 233–266.

[117] X. Ren, et al., Impact of high-speed rail on social equity in China: evidence from a mode choice survey, Transport. Res. Pol. Pract. 138 (2020) 422–441.

[118] J.C. Martín, et al., Spatial analysis of the competitiveness of the high-speed train and air transport: the role of access to terminals in the Madrid–Barcelona corridor, Transport. Res. Pol. Pract. 69 (2014) 392–408.

[119] C. Román, et al., Valuation of travel time savings for intercity travel: the Madrid-Barcelona corridor, Transport Pol. 36 (2014) 105–117.

[120] R.F. Abenoza, O. Cats, Y.O. Susilo, How does travel satisfaction sum up? An exploratory analysis in decomposing the door-to-door experience for multimodal trips, Transportation 46 (5) (2019) 1615–1642.

[121] L. Wu, M. Zhang, Y. Wang, Analysis of Ninghang intercity travel mode choice behavior based on panel data, Journal of Transportation Systems Engineering and Information Technology 15 (1) (2015) 226–231.

[122] M. Danapour, et al., Competition between high-speed rail and air transport in Iran: the case of Tehran–Isfahan, Case Studies on Transport Policy 6 (4) (2018) 456–461.

[123] A. Lhéritier, et al., Airline itinerary choice modeling using machine learning, Journal of choice modelling 31 (2019) 198–209.

[124] J. Wang, J. Huang, Y. Jing, Competition between high-speed trains and air travel in China: from a spatial to spatiotemporal perspective, Transport. Res. Pol. Pract. 133 (2020) 62–78.

[125] Y. Wang, et al., Modeling traveler mode choice behavior of a new high-speed rail corridor in China, Transport. Plann. Technol. 37 (5) (2014) 466–483.

[126] O.D. Karmarkar, A. Jana, N.R. Velaga, Willingness to pay for reduced travel time: case study of Mumbai - ahmedabad high speed railway, Case Studies on Transport Policy 12 (2023) 100974.

[127] S. Shelat, O. Cats, J.W.C. van Lint, Quantifying travellers' evaluation of waiting time uncertainty in public transport networks, Travel Behaviour and Society 25 (2021) 209–222.

[128] A. Abulibdeh, Implementing congestion pricing policies in a MENA Region City: analysis of the impact on travel behaviour and equity, Cities 74 (2018) 196–207.

[129] E. Zaidan, et al., Motivation, preference, socioeconomic, and building features: new paradigm of analyzing electricity consumption in residential buildings, Build. Environ. 219 (2022) 109177.

[130] H. Chen, Y. Cheng, Travel mode choice prediction using imbalanced machine learning, IEEE Trans. Intell. Transport. Syst. 24 (4) (2023) 3795–3808.

[131] Y. Deng, et al., Travel mode choice behavior for high-speed railway stations based on multi-source data, Transport. Res. Rec. 2677 (8) (2023) 525–540.

[132] J. Wang, et al., A system optimization approach for trains' operation plan with a time flexible pricing strategy for high-speed rail corridors, Sustainability 15 (12) (2023) 9556.

[133] L. Liu, M. Zhang, High-speed rail impacts on travel times, accessibility, and economic productivity: a benchmarking analysis in city-cluster regions of China, J. Transport Geogr. 73 (2018) 25–40.

[134] Y. Zhang, et al., Measuring reliable accessibility to high-speed railway stations by integrating the utility-based model and multimodal space–time prism under travel time uncertainty, ISPRS Int. J. Geo-Inf. 13 (8) (2024) 263.

[135] Y. Zhou, et al., Assessing the relationship between access travel time estimation and the accessibility to high speed railway station by different travel modes, Sustainability 12 (18) (2020) 7827.

[136] A. Carteni, L. Pariota, I. Henke, Hedonic value of high-speed rail services: quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities, Transport. Res. Pol. Pract. 100 (2017) 348–365.

[137] Y.L. Pei, K. Zhou, T. Peng, Prediction model of passenger waiting time in high-speed rail hub based on BP neural network, Appl. Mech. Mater. 321 (2013) 1903–1906.

[138] H. Chen, et al., The impact of high-speed rail on residents' travel behavior and household mobility: a case study of the Beijing-Shanghai Line, China, Sustainability 8 (11) (2016) 1187.

[139] Y. Huang, H. Song, C.Y. Jia, Analysis on the main factors influencing the high-speed railway passenger flow, Adv. Mater. Res. 919 (2014) 1085–1090.

มหาวิทยาลัยเทคโนโลยีสุรนารี

LIST OF PUBLICATIONS

Banyong, C., Hantanong, N., Nanthawong, S., Se, C., Wisutwattanasak, P., Champahom, T., Jomnonkwao, S. (2025). Machine Learning-Based Analysis of Travel Mode Preferences: Neural and Boosting Model Comparison Using Stated Preference Data from Thailand's Emerging High-Speed Rail Network. *Big Data and Cognitive Computing, 9*(6). doi:10.3390/bdcc9060155

Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Kasemsri, R., Jomnonkwao, S. (2025). Analyzing High-Speed Rail's Transformative Impact on Public Transport in Thailand Using Machine Learning. *Infrastructures, 10*(3). doi:10.3390/infrastructures10030057

Banyong, C., Hantanong, N., Wisutwattanasak, P., Champahom, T., Theerathitichaipa, K., Seefong, M., Jomnonkwao, S. (2024). A machine learning comparison of transportation mode changes from high-speed railway promotion in Thailand. *Results in Engineering, 24*, 103110. doi:https://doi.org/10.1016/j.rineng.2024.103110

มหาวิทยาลัยเทคโนโลยีสุรนารี

BIOGRAPHY

Mr. Chinnakrit Banyong was born on August 12, 1994, in Bangkok, Thailand. He completed his Bachelor's degree with First Class Honors in Transportation and Logistics Engineering from the School of Transportation Engineering, Suranaree University of Technology in 2016. He then earned a Master's degree in Transportation Engineering from the same university in 2019.

In 2022, he was awarded a doctoral scholarship from Suranaree University of Technology to pursue a Ph.D. in Industrial and Logistics Management Engineering. His academic interests lie in travel behavior analysis, transport system planning, and the application of machine learning techniques to support data-driven decision-making in public transportation.

His Ph.D. dissertation, "Using Machine Learning Techniques to Assess Use Decisions of Thailand's Future High-Speed Rail," applies machine learning to predict travel mode choices and support evidence-based planning for Thailand's high-speed rail development.

Throughout his doctoral studies, Mr. Banyong has published several peer-reviewed journal articles in internationally recognized academic outlets, including

1) Machine Learning-Based Analysis of Travel Mode Preferences: Neural and Boosting Model Comparison Using Stated Preference Data from Thailand's Emerging High-Speed Rail Network, published in Big Data and Cognitive Computing (Vol. 9, Issue 6).
2) Analyzing High-Speed Rail's Transformative Impact on Public Transport in Thailand Using Machine Learning, published in Infrastructures (Vol. 10, Issue 3).
3) A Machine Learning Comparison of Transportation Mode Changes from High-Speed Railway Promotion in Thailand, published in Results in Engineering.

His research reflects a strong commitment to integrating engineering knowledge with intelligent systems to address emerging transportation challenges and promote sustainable mobility in Thailand and the ASEAN region.

มหาวิทยาลัยเทคโนโลยีสุรนารี