

TABLE OF CONTENTS

	Page
ABSTRACT (ENGLISH)	I
ABSTRACT (THAI)	III
ACKNOWLEDGEMENT	V
TABLE OF CONTENTS	VI
LIST OF TABLES	VIII
LIST OF FIGURES.....	IX
LIST OF ABBREVIATIONS	X
CHAPTER	
1 INTRODUCTION	1
1.1 The significance of the research question	1
1.2 The research objective:.....	1
1.3 Contribute of the research.....	2
1.4 Justifications for conducting research in this population	2
2 PREDICTIVE ANALYSIS OF A HIGHWAY ROAD ACCIDENT IN THAILAND: USING MACHINE LEARNING APPROACH	4
2.1 Abstract	4
2.2 Introduction.....	5
2.3 Data Description and Methodology.....	15
2.4 Descriptive Statistics and Result	21
2.5 Conclusion and Discussion.....	32
2.6 Study limitation and future study	34
2.7 Reference.....	34
3 USING A DECISION TREE TO COMPARE RURAL VERSUS HIGHWAY MOTORCYCLE FATALITIES.....	44
3.1 Abstract	44

TABLE OF CONTENTS (Continued)

	Page
3.2 Introduction.....	45
3.3 Literature Review	47
3.4 Methodology.....	50
3.5 Results	57
3.6 Conclusion and Discussion.....	65
3.7 Limitations and Future Studies	67
3.8 References.....	68
4 COMPARISON OF MACHINE LEARNING PREDICTABILITY PERFORMANCE: THE CASE OF MOTORCYCLE ACCIDENT IN THAILAND.....	74
4.1 Abstract	74
4.2 Introduction.....	75
4.3 Literature Review	78
4.4 Methodology and Data	81
4.5 Results	89
4.6 Conclusion & Discussion.....	95
4.7 Limitations and Future Studies	97
4.8 Reference.....	98
5 CONCLUSION AND RECOMMENDATION	105
APPENDIX A LIST OF PUBLICATIONS	106
BIOGRAPHY	120

LIST OF TABLES

Table	Page
2.1 Road accident using data mining and Machine learning	7
2.2 Previous research has identified the factors that determine the severity of driving injuries.....	11
2.3 the driver who was the caused in those accident divided by highway vs Non highway.....	16
2.4 Total 34 Attribute with setting description.....	16
2.5 Focusing Rule with high lift and widely gap between support and confidence.....	28
3.1 The Machine Learning Models Used in Extant Traffic Accident Studies.....	50
3.2 The Categorical Variables and Their Descriptive Statistics	52
3.3 The Measurement Categories for the 27 Identified Motorcycle Accident Variables.....	54
3.4 The Final HW and RR Sets by Rider Speed.....	63
3.5 The Model Evaluation Results	64
4.1 Comparison of the advantages and disadvantages of ML models.....	76
4.2 Machine learning models in traffic accident study.....	78
4.3 Categorical Attribute and descriptive statistics.....	82
4.4 Total 28 Attributes with setting description.....	86
4.5 Info. Gain Ranking by model.....	89
4.6 evaluation result from models	91
4.7 Confusion Metrix for each model.....	94

LIST OF FIGURES

Figures	Page
2.1 Highway accident stacked column chart by year	6
2.2 Data analysis process step	15
2.3 Associate Rules Mining Diagram	19
2.4 Highway accident distribution plot by 24-hour time series w/ Kernel density as line chart	22
2.5 Frequency itemset extraction	23
2.6 Associate Rules Mining total 1558 rules.....	24
2.7 and 2.8 Support and Confidence distribution from 1,558 rules discovered	25
2.8 1,558 discovered rules with scatter plot Support VS Confidence	26
2.9 Dendrogram for 1,558 rules discovered on Antecedent	27
2.10 Confidence and support chart gap trend chart by interesting rules.....	31
3.1 Total number of vehicles and motorcycles registered in Thailand from 2015 to 2020.....	45
3.2 The steps in the process for the study.	51
3.3 Diagram of the confusion matrix.....	57
3.4 HW and RR fatality probabilities at different times of the day.	58
3.5 The HW tree model.....	59
3.6 The RR tree model.	60
3.7 The confusion matrix actual and predicted results for HWs.	65
3.8 The confusion matrix actual and predicted results for RRs.....	65
3.9 Key accident factors: HWs versus RRs.....	66
4.1 Machine learning Process flow.....	81
4.2 Confusion matrix diagram.....	88
4.3 Performance Measurement models	92
4.4 Model-specific ROC plot for predicting non-fatality	93

LIST OF ABBREVIATIONS

LHS	=	Left hand Side
RHS	=	Right Hand Side
HW	=	Highway
RR	=	Rural Road
TP	=	True Positive
FP	=	False Positive
TN	=	True Negative
FN	=	False Negative
DT	=	Decision Tree
SVM	=	Support Vector Machine
RF	=	Random Forest
kNN	=	K-Nearest-Neighbors
NN	=	Neural Network
LR	=	Logistic Regression
GB	=	Gradient Boosting
AUC	=	Area Under Curve
CA	=	Classification Accuracy