CHAPTER 2

PREDICTIVE ANALYSIS OF A HIGHWAY ROAD ACCIDENT IN THAILAND: USING MACHINE LEARNING APPROACH

2.1 Abstract

Accidents are a major obstacle to economic development and quality of life in developing countries. The same challenges are perceived today as major issues in Thailand. This research aims to assess the frequency and most common causes of road accidents that are most likely to result in fatalities. Machine learning technique is employed to examine the relation of factors in accidents, which are then applied to policymaking to lower the rate of road accidents, economic and human resource losses, as well as improve the overall efficiency of a country's healthcare system. The researcher has included information of road accidents in Thailand during the years 2015–2020; a total of 167,820 events, with total damages costing some 1.13 billion Thai baht (34 million USD). Although the overall data comprises the elements influencing the accidents, this article only considers the drivers who were the causes of fatal highway accidents. As a result, the factors that enhance the likelihood of fatality in highway road accidents are as follows: driver info, male; driver behavior, over speed limit; vehicle type, motorbike; roadway, straight, dry surface; and weather, clear. All these variables are related, as the association rule shows an increased risk of injury or death in traffic accidents.

2.1.1 Highlights:

1) Driver risk perception was discovered to have the strongest influence on road accidents.

2) The factors that enhance the likelihood of fatality in highway road accidents are as follows: driver info, male; driver behavior, over speed limit; vehicle type, motorbike; roadway, straight, dry surface; and weather, clear.

3) Most accidents occur during daytime (08.00–18.00), while peaks occur at 19.00–20.00 and 22.00–23.00 and high fatality rate at night (19.00–07.00).

4) The higher the number of elements involved, the greater the possibility of an accident.

2.2 Introduction

Road traffic accidents are a worldwide issue that have been troubling civilization for a long time. Specifically, road accidents in Southeast Asia and Africa, the two previously mentioned regions, have been continuously increasing for at least the last 10 years (2008–2018) WHO (2018). According to WHO data in 2018, Thailand was ranked No. 1 for road accidents in Asia and No. 9 in the world. An average of 32.7 Thais per 100,000 population die in road accidents every year (WHO, 2018). Not only has it caused an economic upheaval, but it has impacted the country's public health system. Road accidents have also caused the country's limited resources to be used in ways harmful to its progress. It also negatively impacts the country's human resources, resulting in the death or disability of its residents.

In Thailand, examples of road safety policies include law enforcement (e.g., for exceeding speed limits or the consumption of alcohol), road safety programs in educational institutions, the development of advertising media, an increase in the number of training hours required to obtain new drivers' licenses and their renewals, engineering solution techniques for road safety audits, and research funding. To establish these regulations, predicted data on the number of accidents was used to determine operational budgets (Jomnonkwao et al., 2020). However, the average number of roadway fatalities in Thailand from 2015 to 2020 remained consistent at 32%–35% for the fifth year in a row, as shown in Fig 2.1. The existing policy appears to be ineffective. Learning from every element recorded in the big data set and starting to predict and minimize things before they occur might be the way out.

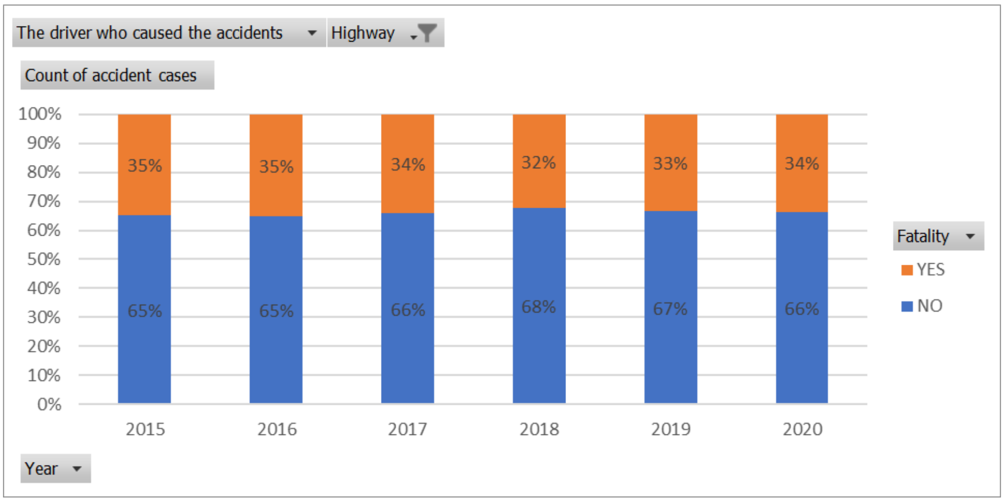

Figure 2.1 Highway accident stacked column chart by year.

Previous studies have utilized machine learning algorithms to predict injury severity. Some focus on independent factors like the environment, drivers, current weather, or road conditions, even comparing performance models, as shown in Table 2.1. However, these studies did not consider the events' coincidence for the drivers who were killed. The coincidence being discussed included type of roadway, vehicle type, external factors such as environment and weather conditions, and internal factors, e.g., driver behaviors and information, like gender and age, to understand which factors interfered with each other or any linkage between them that increased the chances of fatality. According to the Swiss cheese theory, if all the holes (factors) are aligned by chance, the accident will happen and result in death. In contrast, the risk may be decreased by controlling the primary element that has the strongest influence on fatality. For example, the researcher noted that accidents are typically caused by a combination of circumstances rather than by one or two factor(s). And, if the elements were combined, how likely is it that someone would die? However, what happens if the risk factor is reduced? That is why forecasts appear to simulate the situation. However, predicting the accident event is also essential for establishing road safety, budgeting, staffing, and policy planning.

Table 2.1 Road accident using data mining and Machine learning.

Author	**Methodology**														
	Apriori Algorithm	Associated Rule	Bayesian Logistic	Cluster Analysis	Decision Tree	Deep Learning	Gradient Boosting	K-means	K-Nearest Neighbor	MultinomialLogistic Regression	Neural Network	Naïve Bayes	Random Forest	Regression on python	Support Vector Machine
Sonal and Suman (2018)	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-
Gutierrez-Osorio and Pedraza (2020)	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-
Abellán et al. (2013)	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-
Al Mamlook et al. (2019)	-	-	✓	✓	✓	-	-	-	✓	-	-	✓	✓	-	✓
Mafi et al. (2018)	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-
Recal and Demirel (2021)	-	-	-	-	✓	-	✓	-	-	✓	✓	-	-	-	✓
Bahiru et al. (2018)	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-

Table 2.1 Road accident using data mining and Machine learning (Continued)

Author	Methodology														
	Apriori Algorithm	Associated Rule	Bayesian Logistic	Cluster Analysis	Decision Tree	Deep Learning	Gradient Boosting	K-means	K-Nearest Neighbor	MultinomialLogistic Regression	Neural Network	Naïve Bayes	Random Forest	Regression on python	Support Vector Machine
Cuenca et al. (2018)	-	-	-	-	-	✓	✓	-	-	-	-	✓	-	-	-
Kuşkapan et al. (2021)	-	-	-	-	-	-	-	-	✓	-	-	✓	-	-	✓
Ospina-Mateus et al. (2021)	-	-	-	-	✓	-	-	-	✓	-	✓	✓	✓	-	✓
Kumar and Toshniwal (2016)	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-
Helen et al. (2019)	-	✓	-	-	-	-	-	✓	-	-	-	-	-	-	-
El Abdallaoui et al. (2018)	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
John and Shaiba (2019)	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2.1 Road accident using data mining and Machine learning (Continued)

Author	Methodology														
	Apriori Algorithm	Associated Rule	Bayesian Logistic	Cluster Analysis	Decision Tree	Deep Learning	Gradient Boosting	K-means	K-Nearest Neighbor	MultinomialLogistic Regression	Neural Network	Naïve Bayes	Random Forest	Regression on python	Support Vector Machine
Feng et al. (2020)	-	✓	-	-	-	-	-	-	-	-	✓	-	-	-	-
Bhavsar et al. (2021)	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
Samerei et al. (2021)	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-
John and Shaiba (2022)	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Earlier research on road traffic accidents have also been categorized by variables in the form that are presumed to be associated in every accident, according to international research.

Age – Zhang and Fan (2013) found that accidents are more likely to occur among junior drivers (≤25 yrs.) who have a lack of discipline, are inexperienced with traffic regulations, as well as having less driving experience. Most traffic accidents in Dubai are caused by a lack of space between vehicles, with youth (≤35 yrs.) being the most usually involved; the peak hour(s) are late at night, and the overwhelming majority of drivers were discovered to be inebriated. (John & Shaiba, 2019). Young (18–24 years old) drivers lack experience at controlling speeding or adjusting well while driving (Bucsuházy et al., 2020). John and Shaiba (2022) found that most alcohol-involved accidents are caused by youths (≤35 yrs.) late at night.

Gender – Ospina-Mateus et al. (2019) and Mohamad et al. (2022) observed that men are more likely to be involved in serious accidents than women.

Driver behaviors – When compared to other drivers, intoxicated drivers have a higher accident rate (Helen et al., 2019). The most important aspect in predicting the severity of an injury is its driving over speed limit (Al Mamlook et al., 2019).

Driver – Drivers are more likely to be injured or killed in accidents than other passengers (El Abdallaoui et al., 2018).

Time – Traveling at night increases the chances of car accidents (Mphela, 2020).

Road and light conditions – Chen et al. (2016) observed that road slope and visibility were predictors of driver injuries. Highway intersections are riskier for all accident types. Poor road conditions increase the likelihood of accidents, especially on motorways (Malin et al., 2019). Road type, lighting, speed limits, and road surface all play key roles in accident incidence (Feng et al., 2020). Most fatal injuries occur as a result of aggressive driving, inattentiveness, and speeding. However, compared to other situations, dark or dim roads also played significant roles (Shweta et al., 2021).

Weather conditions – (Kumar & Toshniwal, 2016) Sonal and Suman (2018) observed that external factors, like weather conditions such as fog, rain, and snow, have greater impacts on road accidents than internal factors, such as the driver.

Type of vehicles – Chen et al. (2015) mentioned this factor as significant for driver injuries and fatalities in rear-end accidents involving trucks, lighting, wind, and multiple vehicles involved. The analysis revealed that the most essential and impactful traffic accident elements are speed limit, weather conditions, number of lanes, lighting conditions, and accident timing, while gender, age, accident location, and vehicle type have less of an impact on severity (Bahiru et al., 2018)

The researchers are continuing to evaluate the literature on road accidents and the factors involved. It will cover a wide range of research from across the world, but Table 2.2 will concentrate on research from the same region as this study.

Table 2.2 Previous research has identified the factors that determine the severity of driving injuries.

Variables	Finding
Driver Characteristics	
Gender	Decrease injury-severity: male. (Xie & Huynh, 2012), (Behnood & Mannering, 2017), (Li, Wu, et al., 2019a), Increase injury-severity: female (Wu et al., 2016), (Osman et al., 2018), (Behnood & Mannering, 2017) (Hou et al., 2019), Male (Kim et al., 2013), (Li et al., 2018), (Champahom et al., 2020)

Table 2.2 Previous research has identified the factors that determine the severity of driving injuries (Continued)

Variables	Finding
Age	Decrease injury-severity: less than 25. (Behnood & Mannering, 2017), (Li, Ci, et al., 2019) Increase injury-severity: Less than 25 (Li et al., 2018) more than 65 (Kim et al., 2013), (Wu et al., 2016), (Li, Wu, et al., 2019b), (Zhou & Chin, 2019), (Hou et al., 2019), (Wei et al., 2021) (Champahom et al., 2020),
Speeding	Increase injury-severity: speeding vehicle. (Kim et al., 2013), (Osman et al., 2018), (Krull et al., 2000), (Xie & Huynh, 2012), (M. Yu et al., 2020)
Drunk	Increase injury-severity: drunk driving. (Krull et al., 2000), (Xie & Huynh, 2012), (Kim et al., 2013), (Wu et al., 2016), (Zhou & Chin, 2019), (John & Shaiba, 2019), (Helen et al., 2019) ,(Champahom et al., 2020)
Fatigue	Increase injury-severity: Doze off. (Champahom et al., 2020)
Overtaking	Increase injury-severity: improper overtaking. (Jafari Anarkooli et al., 2017), (Li, Wu, et al., 2019a)

Table 2.2 Previous research has identified the factors that determine the severity of driving injuries (Continued)

Variables	Finding
Vehicle characteristics	
Vehicle type	Decrease injury-severity: SUV/van (Chamroeun Se et al., 2021) Pick-up truck (Wu et al., 2016), (Chamroeun Se et al., 2021) passenger car (Huo et al., 2020) Increase injury-severity: rollover SUV/van (Jafari Anarkooli et al., 2017) large truck (Jafari Anarkooli et al., 2017), (Li et al., 2018), (Huo et al., 2020) Pickup (Li et al., 2018),
External Factor (Environment and road condition)	
Light status	Decrease injury-severity: darkness without light. (Xie & Huynh, 2012), Increase injury-severity: daylight. (Krull et al., 2000) darkness without light (Kim et al., 2013), (Jafari Anarkooli et al., 2017) (Zhou & Chin, 2019)

Table 2.2 Previous research has identified the factors that determine the severity of driving injuries (Continued)

Variables	Finding
	after midnight (Zhou & Chin, 2019) Nighttime (Mphela, 2020), (Osman et al., 2018)
Dry/wet road surface	Decrease injury-severity: wet road. (Zhou & Chin, 2019), (H. Yu et al., 2020) Increase injury-severity: Wet road (Li, Wu, et al., 2019a), (Li et al., 2018) dry road (Krull et al., 2000)
Weather	Decrease injury-severity: raining. (Jung et al., 2010) Increase injury-severity: raining. (Shweta et al., 2021), (Jafari Anarkooli et al., 2017), (Li, Wu, et al., 2019a) Fog, Rainfall, Snowfall (Shweta et al., 2021),
Time	Increase injury-severity: Daytime. (Shaheed et al., 2013) Nighttime (Champahom et al., 2020), (Chamroeun Se et al., 2021)

2.3 Data Description and Methodology

2.3.1 Data Description

The occurrence of road accidents from the Thailand government organization during the years 2015–2020 amounted to 167,820 events PDPM (2020). This study focuses on drivers who caused their accidents. Those came to 129,015 total, of which 95,249 were nonfatal and 33,766 fatal (24,559 for highway and 9,207 for nonhighway). Using the data analysis technique to execute the following steps in Fig. 2.2.

Data cleaning – missing and incompletely captured data detection and correction.

Data validation – validation the quality of the data after the data set has been cleansed.

Data converting – data partitioning to binary mode.

Data analysis and interpretation – discovering the data for informing conclusion.

Data visualization – creating a visual to represent information and data.

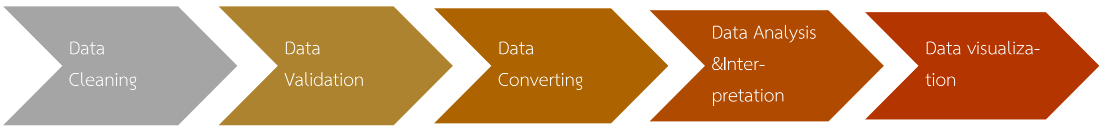

Figure 2.2 Data analysis process step

The data in Table 2.3 is classified into four different categories, consisting of fatalities (HW & NHW) and nonfatalities (HW&NHW) to find the link between those and the type of roads. However, this study focuses on highway fatalities. The authors converted the total data to binary to represent Yes or No in each accident event and fed it through Python base software. Table 3 presented data divided by road type and fatality. The large number, 24,599, drew our attention and encouraged us to investigate.

Table 2.3 the driver who was the caused in those accident divided by highway vs Non highway.

Count of Road Accident Case	Fatality		
Road Type	No	Yes	Grand Total
Non-Highway	47,136	9,207	56,343
Highway	48,113	24,559*	72,672
Grand Total	95,249	33,766	129,015

In every event, aspects of 34 attributes from accident data collection appeared, including roadway, vehicle type, environment, weather conditions, driver behavior, driver info, and driver status in Table 2.4.

Table 2.4 Total 34 Attribute with setting description

Attribute Name	Attribute Description
Roadway	
Highway	1 - Yes
Dry Surface Road	1 - Yes, 0-Otherwise
Straight Way	1 - Yes, 0-Otherwise
Obstruction	1 - Yes, 0-Otherwise
Road condition	1 - Yes, 0-Otherwise
Vehicle condition	1 - Yes, 0-Otherwise

Table 2.4 Total 34 Attribute with setting description (Continued)

Attribute Name	Attribute Description
Vehicle Type	
Motorcycle	1 - Yes, 0-Otherwise
Mini truck/ Pick up (4 wheels)	1 - Yes, 0-Otherwise
Sedan	1 - Yes, 0-Otherwise
Light Truck (6 wheels)	1 - Yes, 0-Otherwise
Heavy Truck (10+ wheels)	1 - Yes, 0-Otherwise
Other Type of car	1 - Yes, 0-Otherwise
External Factor (Environment and Weather Condition)	
Day Time (06.00-18.00)	1 - Yes, 0-Otherwise
Night with Light	1 - Yes, 0-Otherwise
Night without Light	1 - Yes, 0-Otherwise
Low visibility	1 - Yes, 0-Otherwise
Clear Weather	1 - Yes, 0-Otherwise
Internal Factor (Driver Behavior)	
Drunk	1 - Yes, 0-Otherwise
Over Speed limit	1 - Yes, 0-Otherwise
Break Through Traffic lights	1 - Yes, 0-Otherwise
Break Through Traffic Signs	1 - Yes, 0-Otherwise

Table 2.4 Total 34 Attribute with setting description (Continued)

Attribute Name	Attribute Description
Overtake	1 - Yes, 0-Otherwise
Use Mobile Phone	1 - Yes, 0-Otherwise
Short Cut off	1 - Yes, 0-Otherwise
Drug	1 - Yes, 0-Otherwise
Drive in opposite direction	1 - Yes, 0-Otherwise
Doze off	1 - Yes, 0-Otherwise
Overweight Carry	1 - Yes, 0-Otherwise
Cannot Conclude	1 - Yes, 0-Otherwise
Driver info	
Gender	1- Male, 0-Otherwise
Youth 15-35	1 - Yes, 0-Otherwise
Adult 36-60	1 - Yes, 0-Otherwise
Senior 61-90+	1 - Yes, 0-Otherwise
Driver Status	
Fatality (Death)	1 - Yes

2.3.2 Methodology

Apriori algorithm (Srikant, 1994) was picked to mine for frequent items set over the entire massive relational data set to discover the most common individual items and extend them to larger itemset as long as the sets appeared frequently enough in the database. Apriori's frequent itemset can be used to generate association rules that highlight overall trends.

Association rule learning is a *rule-based, machine learning* method for discovering key relations between variables in large databases. It is intended to identify strong rules using various measures of attraction (William J. Frawley, 1992). To detect correlations and co-occurrences between data sets, association rules are utilized. They are best suited for explaining data patterns from among seemingly unrelated information sources, such as relational and transactional databases. The act of employing association rules is known as *association rule mining*, or *mining associations*. See Fig. 2.3:

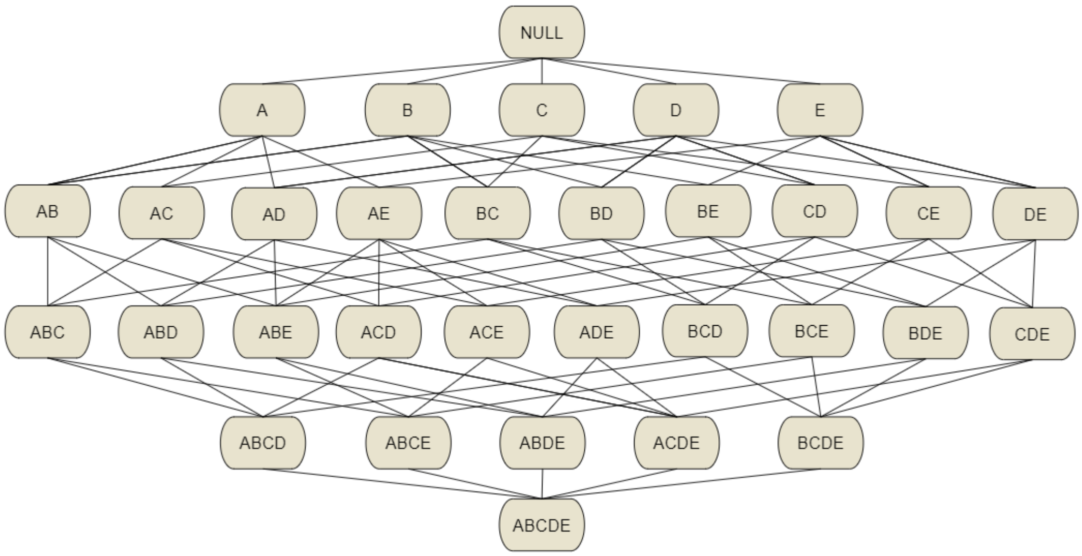

Figure 2.3 Associate Rules Mining Diagram

Rule definition and measurement

An association rule is determined by two factors: support and confidence. The frequency with which a specific rule appears in the database being mined is referred to as *support*. The number of times a particular rule turns out to be true in practice is referred to as a *confidence*.

Let *I* = {...} represent a collection of "n" binary characteristics known as items.

Let J = {...} be a set of transactions referred to as a database.

Each transaction in J has a distinct transaction ID and includes a subset of the items in I. A rule is defined as an implication of the type XY in which X, $Y \subseteq$ I if and only $X \neq \emptyset$, $Y \neq \emptyset$, $X \cap Y = \emptyset$. The sets of objects X and Y are referred to as the rule's antecedent and consequent, respectively.

Support is an indicator of how frequently the itemset appears in the data set.

$$\text{Support } (x) = \frac{Frequent\ item(x)}{N(Total\ Number\ of\ transaction)}$$

Confidence is an indication of how often the rule has been found to be true.

$$Confidence\ [LHS(x) \Rightarrow\ RHS(y)] = \frac{Support\ (LHS, RHS)}{Support\ (LHS)}$$

The ratio of the observed support to the support expected if X and Y were independent.

$$Lift\ [LHS(x) \Rightarrow\ RHS(y)] = \frac{Support\ (LHS, RHS)}{Support\ (LHS) \times Support\ (RHS)}$$

A rule may have a significant association in a data collection because it frequently appears, but it may occur considerably less frequently when implemented. This would be an example of strong support but low confidence.

Step to perform associated rule mining.

1. Sequence the transaction accident by event (binary) – If minimum support, measure the effectiveness of the accident. If >50% (threshold), then others below 50% will be removed.

1.1 Use frequency itemset from 1 to build item new itemset (length 2). Using join command, if all are set, the sequencing does not matter.

1.2 Recalculate the support score, using transaction in 1.1 to intersection such as

Transaction {Road wet} = {1,1,1,0,1, 0…}

Transaction {Darkness} = {1,1,1,1,0,0…}

Transaction {Road wet, Darkness} = {1,1,1,0,0,0…}

If minimum support < threshold will get removed

1.3 Use frequency itemset from 1.2 to create item new itemset (length 3). However, remember that the initial item must be the same (using the join command), and only one linkage can join:

Transaction {Road wet, Darkness} = {1,1,1,0,0,0…}

Transaction {Road wet, Drunk} = {1,1,1,0,1,0…}

Transaction {Road wet, Darkness, Drunk} = {1,1,1,0,0,0…}

1.4 Frequency all Itemset

2. Consider the following two items or more and then calculate for confidence and lift

2.4 Descriptive Statistics and Result

To comprehend the data pattern and how data distribution works, a distribution chart was created using 72,672 highway accident incidents over 24-h fitted with kernel density as a time series as descriptive statistics shown in Fig. 2.4. To determine a difference between day and night:

1 – Representing fatalities from highway accidents; μ =13.19, σ = 7.03

0 – Representing nonfatalities from highway accidents; μ =13.57, σ =6.37

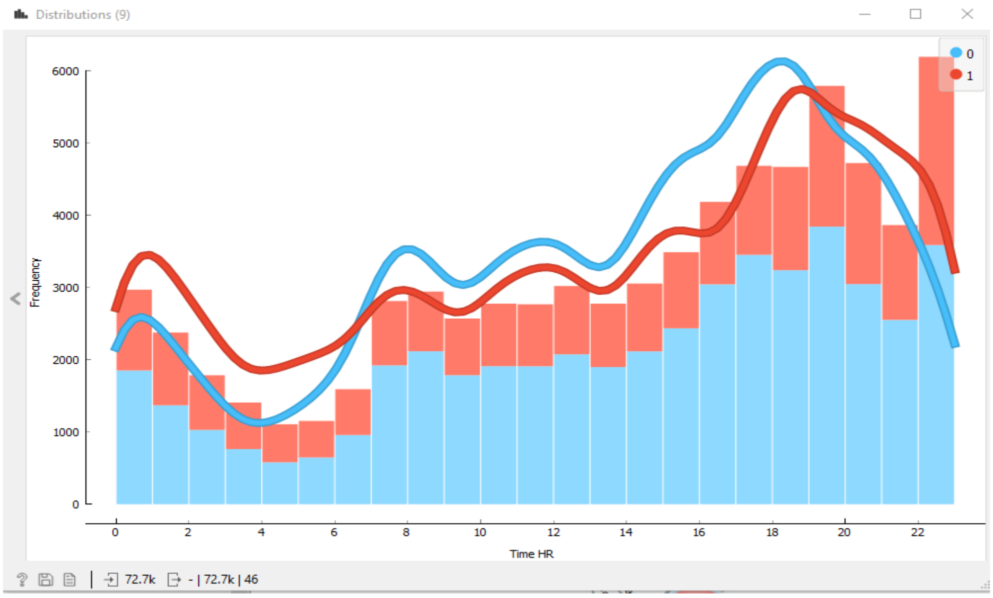

Figure 2.4 Highway accident distribution plot by 24-hour time series w/ Kernel density as line chart

Most accidents occur during daytime (08.00–18.00), while peaks occur at 19.00–20.00 and 22.00–23.00 and high fatality rate at night (19.00–07.00).

Later, they started to frequent items set on fatality as a precondition for the extraction of rules emphasizing causal linkages (Fig. 2.5). Knowing which elements occur together aids in identifying the linkages between them (minimum support at 50%). According to Fig. 5, the most often discovered itemset in the 2018 set is connected to the item found: dry road (95.98%), clear weather (87.33%), male (86.42%), motorcycle (80.77%), straightaway (71.99%), and over speed limit, (69.03%), respectively.

Frequent Itemsets (2)

Itemsets	Support	%
Dry Surface Road=1	23572	95.98
Cannot Conclude =0	22709	92.47
Break Through Traffic lights=0	23364	95.13
Break Through Traffic Signs=0	23238	94.62
Drive in opposite direction=0	23188	94.42
Overtake=0	22917	93.31
Use Mobile Phone=0	23548	95.88
Drug=0	23569	95.97
Doze off=0	22724	92.53
Overweight Carry=0	23542	95.86
Obstruction=0	23190	94.43
Vehicle condition=0	23329	94.99
Road condition=0	23340	95.04
Light Truck =0	23335	95.02
Heavy Truck =0	23409	95.32
Other Type of car=0	23232	94.6
Sedan=0	22037	89.73
Low visibility=0	21652	88.16
Mini truck/ Pick up =0	21267	86.6
Drunk=0	20805	84.71
Clear Weather=1	21308	86.76
61-90=0	20480	83.39
Gender=1	20364	82.92
Short Cut off=0	19025	77.47
Motorcycle=1	18992	77.33
Night without Light=0	18299	74.51
Straight Way=1	17078	69.54
Night with Light=0	16920	68.9
Over Speed limit=1	16369	66.65
36-60=0	14475	58.94
Clear Weather=1	21448	87.33
Gender=1	21224	86.42
Drug=0	21220	86.4
Use Mobile Phone=0	21201	86.33
Overweight Carry=0	21195	86.3
Cannot Conclude =0	20458	83.3
Break Through Traffic lights=0	21046	85.7
Drive in opposite direction=0	20863	84.95
Break Through Traffic Signs=0	20949	85.3
Vehicle condition=0	21005	85.53
Obstruction=0	20875	85
Overtake=0	20639	84.04
Doze off=0	20454	83.29
Road condition=0	20955	85.33
Drunk=0	18569	75.61
Low visibility=0	19253	78.39
Short Cut off=0	17409	70.89
Motorcycle=1	19665	80.07
61-90=0	17005	69.24
Night without Light=0	16183	65.89
Straight Way=1	17679	71.99
Night with Light=0	14944	60.85
Over Speed limit=1	16952	69.03
36-60=0	12554	51.12

24.6k -

Figure 2.5 Frequency itemset extraction

After frequent itemset, the first result came from a highway with 24,559 fatalities. The association rule discovered 1,558 rules (lift ≥ 1 containing 1,377 rules), all of which had been configured to obey the threshold (support 50%, confidence 95%) using Orange 3.30 software (Demšar et al., 2013) (Fig. 2.6). The support distribution (Fig. 2.7) has μ = 0.680263, σ = 0.0954974, while confidence distribution (Fig. 2.8) has μ = 0.972597, σ = 0.0126851.

Association Rule for Fatality on HW

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.551	0.966	0.570	1.404	1.206	0.094	Over Speed limit=1, Mini truck/ Pick up =0, Sedan=0	→	Motorcycle=1
0.577	0.964	0.599	1.337	1.203	0.098	Straight Way=1, Mini truck/ Pick up =0, Sedan=0	→	Motorcycle=1
0.773	0.962	0.803	0.997	1.202	0.130	Dry Surface Road=1, Mini truck/ Pick up =0, Sedan=0	→	Motorcycle=1
0.704	0.962	0.732	1.093	1.201	0.118	Clear Weather=1, Mini truck/ Pick up =0, Sedan=0	→	Motorcycle=1
0.689	0.958	0.719	1.113	1.196	0.113	Gender=1, Mini truck/ Pick up =0, Sedan=0	→	Motorcycle=1
0.548	0.995	0.551	1.742	1.037	0.019	Clear Weather=1, Over Speed limit=1, Mini truck/ Pick up =0	→	Dry Surface Road=1
0.527	0.995	0.529	1.813	1.037	0.019	Clear Weather=1, Road condition=0, 36-60=0	→	Dry Surface Road=1
0.514	0.995	0.517	1.856	1.037	0.018	Clear Weather=1, Doze off=0, 36-60=0	→	Dry Surface Road=1
0.566	0.995	0.569	1.686	1.037	0.020	Clear Weather=1, Over Speed limit=1, Sedan=0	→	Dry Surface Road=1
0.516	0.995	0.519	1.849	1.036	0.018	Clear Weather=1, Overtake=0, 36-60=0	→	Dry Surface Road=1
0.523	0.995	0.526	1.824	1.036	0.018	Clear Weather=1, Obstruction=0, 36-60=0	→	Dry Surface Road=1
0.529	0.995	0.532	1.805	1.036	0.019	Clear Weather=1, 36-60=0, Heavy Truck =0	→	Dry Surface Road=1
0.527	0.995	0.530	1.812	1.036	0.019	Clear Weather=1, Break Through Traffic lights=0, 36-60=0	→	Dry Surface Road=1
0.526	0.995	0.529	1.816	1.036	0.018	Clear Weather=1, Vehicle condition=0, 36-60=0	→	Dry Surface Road=1
0.531	0.995	0.534	1.797	1.036	0.019	Clear Weather=1, 36-60=0	→	Dry Surface Road=1
0.531	0.995	0.534	1.797	1.036	0.019	Clear Weather=1, Drug=0, 36-60=0	→	Dry Surface Road=1
0.531	0.995	0.533	1.799	1.036	0.019	Clear Weather=1, Use Mobile Phone=0, 36-60=0	→	Dry Surface Road=1
0.531	0.995	0.533	1.799	1.036	0.019	Clear Weather=1, Overweight Carry=0, 36-60=0	→	Dry Surface Road=1
0.513	0.995	0.516	1.859	1.036	0.018	Clear Weather=1, Cannot Conclude =0, 36-60=0	→	Dry Surface Road=1
0.527	0.995	0.530	1.810	1.036	0.019	Clear Weather=1, 36-60=0, Light Truck =0	→	Dry Surface Road=1
0.525	0.995	0.528	1.818	1.036	0.018	Clear Weather=1, 36-60=0, Other Type of car=0	→	Dry Surface Road=1
0.524	0.995	0.527	1.821	1.036	0.018	Clear Weather=1, Break Through Traffic Signs=0, 36-60=0	→	Dry Surface Road=1
0.523	0.995	0.525	1.827	1.036	0.018	Clear Weather=1, Drive in opposite direction=0, 36-60=0	→	Dry Surface Road=1
0.604	0.995	0.608	1.580	1.036	0.021	Clear Weather=1, Over Speed limit=1, Heavy Truck =0	→	Dry Surface Road=1
0.694	0.995	0.698	1.376	1.036	0.024	Clear Weather=1, Road condition=0, Motorcycle=1	→	Dry Surface Road=1
0.602	0.995	0.605	1.586	1.036	0.021	Clear Weather=1, Over Speed limit=1, Obstruction=0	→	Dry Surface Road=1
0.511	0.995	0.514	1.868	1.036	0.018	Straight Way=1, Clear Weather=1, Motorcycle=1	→	Dry Surface Road=1
0.593	0.995	0.597	1.609	1.036	0.021	Clear Weather=1, Over Speed limit=1, Overtake=0	→	Dry Surface Road=1
0.604	0.995	0.608	1.580	1.036	0.021	Clear Weather=1, Over Speed limit=1, Road condition=0	→	Dry Surface Road=1
0.601	0.995	0.604	1.589	1.036	0.021	Clear Weather=1, Over Speed limit=1, Break Through Traffic Signs=0	→	Dry Surface Road=1
0.608	0.995	0.611	1.570	1.036	0.021	Clear Weather=1, Over Speed limit=1	→	Dry Surface Road=1
0.608	0.995	0.611	1.570	1.036	0.021	Clear Weather=1, Cannot Conclude =0, Over Speed limit=1	→	Dry Surface Road=1
0.608	0.995	0.611	1.570	1.036	0.021	Clear Weather=1, Over Speed limit=1, Drug=0	→	Dry Surface Road=1
0.608	0.995	0.611	1.571	1.036	0.021	Clear Weather=1, Over Speed limit=1, Overweight Carry=0	→	Dry Surface Road=1
0.608	0.995	0.611	1.571	1.036	0.021	Clear Weather=1, Over Speed limit=1, Use Mobile Phone=0	→	Dry Surface Road=1

24.6k 12.9k | 1558

Figure 2.6 Associate Rules Mining total 1558 rules.

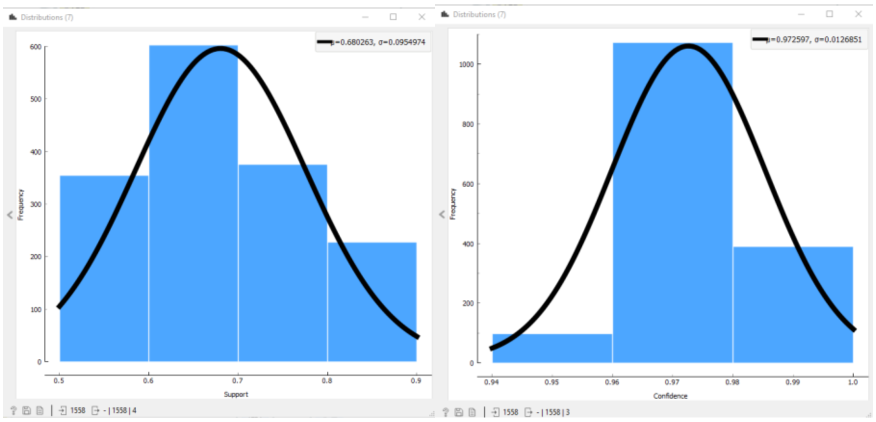

Figure 2.7 and 2.8 Support and Confidence distribution from 1,558 rules discovered.

Overall, 1,558 rule mining was discovered and divided by confidence clustering with color shades representing a Confidence Zone. The *y*-axis represents confidence, while the x-axis represents support. It becomes apparent that:

Group 1 Confidence 0.95–0.965 – Blue shade majority rule containing antecedent as male and dry surface as consequence.

Group 2 Confidence 0.965–0.98 – Green shade majority rule containing motorcycle and over speed limit as antecedent and dry surface road as consequence.

Group 3 Confidence 0.98–0.995 – Yellow shade is always high confidence, although with low support, since Cluster 3 contains clear weather as an antecedent and dry surface. Consequently, it implies that these two elements have a significant role in road accident mortality (Fig. 2.8) and that extreme caution should be taken during clear weather on dry surfaces.

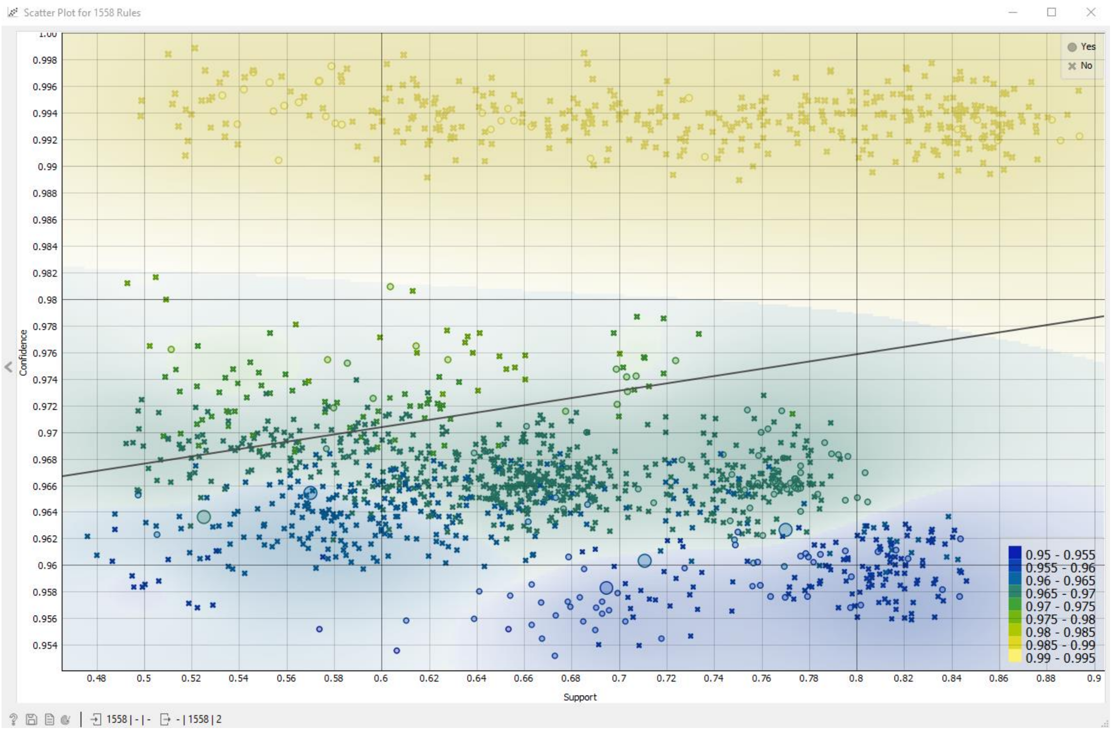

Figure 2.8 1,558 discovered rules with scatter plot Support VS Confidence

Later, start building a hierarchy cluster (HCA) by applying the agglomerative on 1,558 rules to arrange related antecedents into similar groups as a cluster with distancing. The distance between clusters was calculated using Euclidean distance as a complete linkage criterion. The dendrogram (Fig. 2.9) shows a C1–C3 cluster for the antecedent:

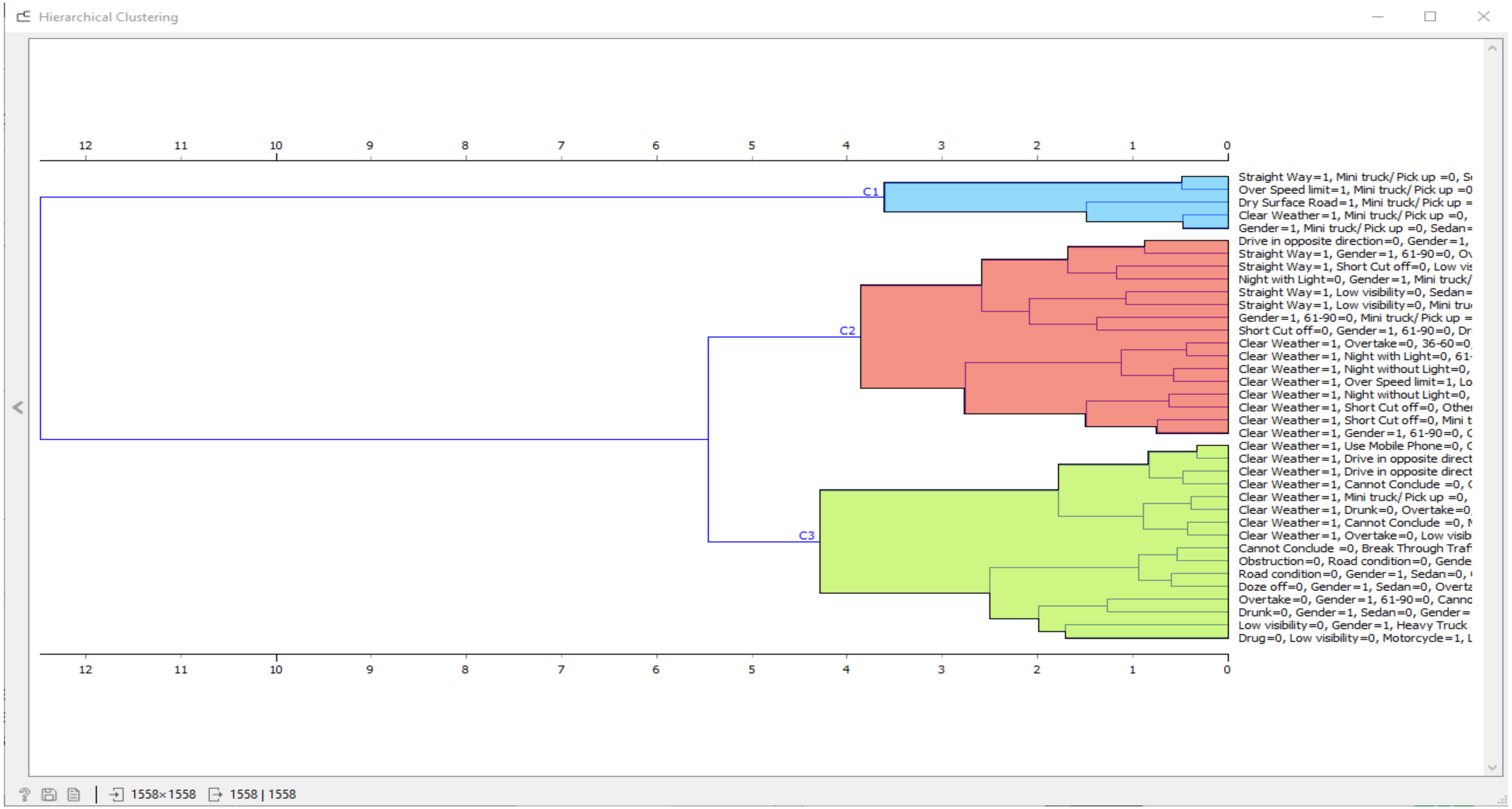

Figure 2.9 Dendrogram for 1,558 rules discovered on Antecedent.

C1 contains straightaway, over speed limit, dry surface road, clear weather, and male.

C2 contains straightaway, over speed limit, clear weather, and male.

C3 contains motorcycle, over speed limit, clear weather, and male.

Regarding the C1 cluster, all elements indicate the same consequence point to motorcycles, implying that the C1 cluster has most motorcycle fatalities, while C2 and C3 have consequence points to the dry surface road. That makes more sense when motorcyclists ride at higher speeds in clear weather on dry surface roads with less care than on wet road surfaces in poor weather conditions.

Table 2.5 Focusing Rule with high lift and widely gap between support and confidence.

Antecedent_1	Antecedent_2	Antecedent_3	Consequence	Support	Confidence	Lift
Over Speed Limit=1	Mini truck/ Pick up=0	Sedan=0	Motorcycle=1	0.551	0.966	1.206
Straight Way=1	Mini truck/ Pick up=0	Sedan=0	Motorcycle =1	0.577	0.964	1.203
Dry Surface Road=1	Mini truck/ Pick up=0	Sedan=0	Motorcycle =1	0.773	0.962	1.202
Clear Whether=1	Mini truck/ Pick up=0	Sedan=0	Motorcycle =1	0.704	0.962	1.201
Gender=1	Mini truck/ Pick up=0	Sedan=0	Motorcycle =1	0.689	0.958	1.196
Clear Weather=1	Over Speed Limit=1	Sedan=0	Dry Surface Road=1	0.566	0.995	1.037
Clear Weather=1	Over Speed Limit=1	Mini truck/ Pick up=0	Dry Surface Road=1	0.548	0.995	1.037
Clear Whether=1	Drunk=0	Motorcycle =1	Dry Surface Road=1	0.620	0.994	1.036
Clear Whether=1	Gender=1	Motorcycle =1	Dry Surface Road=1	0.599	0.994	1.036

Table 2.5 Focusing Rule with high lift and widely gap between support and confidence (Continued)

Antecedent_1	Antecedent_2	Antecedent_3	Consequence	Support	Confidence	Lift
Clear Whether=1	Over Speed Limit=1	Gender=1	Dry Surface Road=1	0.527	0.994	1.036
Straight Way=1	Clear Weather=1	Motorcycle =1	Dry Surface Road=1	0.511	0.995	1.036
Clear Weather=1	Gender=1	Dry Surface Road=1	Dry Surface Road=1	0.746	0.993	1.035
Straight Way=1	Clear Weather=1	Gender=1	Dry Surface Road=1	0.546	0.993	1.035
Over Speed Limit=1	Motor Bike=1		Dry Surface Road=1	0.535	0.972	1.013
Straight Way=1	Motor Bike=1		Dry Surface Road=1	0.56	0.97	1.011
Road Condition=0	Gender=1	Motorcycle =1	Dry Surface Road=1	0.659	0.968	1.008
Over Speed Limit=1	Road Condition=0	Gender=1	Dry Surface Road=1	0.576	0.966	1.007
Drunk=0	Gender=1	Motorcycle =1	Dry Surface Road=1	0.577	0.966	1.006
Gender=1	Motorcycle =1	Sedan=0	Dry Surface Road=1	0.665	0.966	1.006
Gender=1	Motorcycle =1	Mini truck/ Pick up=0	Dry Surface Road=1	0.665	0.966	1.006
Gender=1	Motorcycle =1	Other Type of car=0	Dry Surface Road=1	0.665	0.966	1.006
Gender=1	Motorcycle =1	Light Truck (6 wheels) =0	Dry Surface Road=1	0.665	0.966	1.006

Table 2.5 Focusing Rule with high lift and widely gap between support and confidence (Continued)

Antecedent_1	Antecedent_2	Antecedent_3	Consequence	Support	Confidence	Lift
Gender=1	Motorcycle =1	Heavy Truck (10+ wheels) =0	Dry Surface Road=1	0.665	0.966	1.006
Gender=1	Motorcycle=1		Dry Surface Road=1	0.665	0.966	1.006
Vehicle condition=0	Gender=1	Motorcycle =1	Dry Surface Road=1	0.659	0.966	1.006
Straight Way=1	Vehicle condition=0	Gender=1	Dry Surface Road=1	0.596	0.965	1.006

The following Table 2.5 and Fig. 2.10 display the association rules with a high lift and a wide gap between support and confidence with the antecedents 1–3 and the consequences, followed by the support score, confidence, and lift. The study established a minimum support score of more than 50%, a confidence threshold of more than 95%, and a lift threshold of more than one (1). For example, the rule with the widest gap between support and confidence is antecedent (straightaway, clear weather, motorcycle) => consequence (dry surface road), which increases 0.484 from support 0.511 to confidence 0.995. The rule with the highest lift is contained by motorcycles with different antecedents. All the interesting rules have been plotted, as shown in Fig. 2.10.

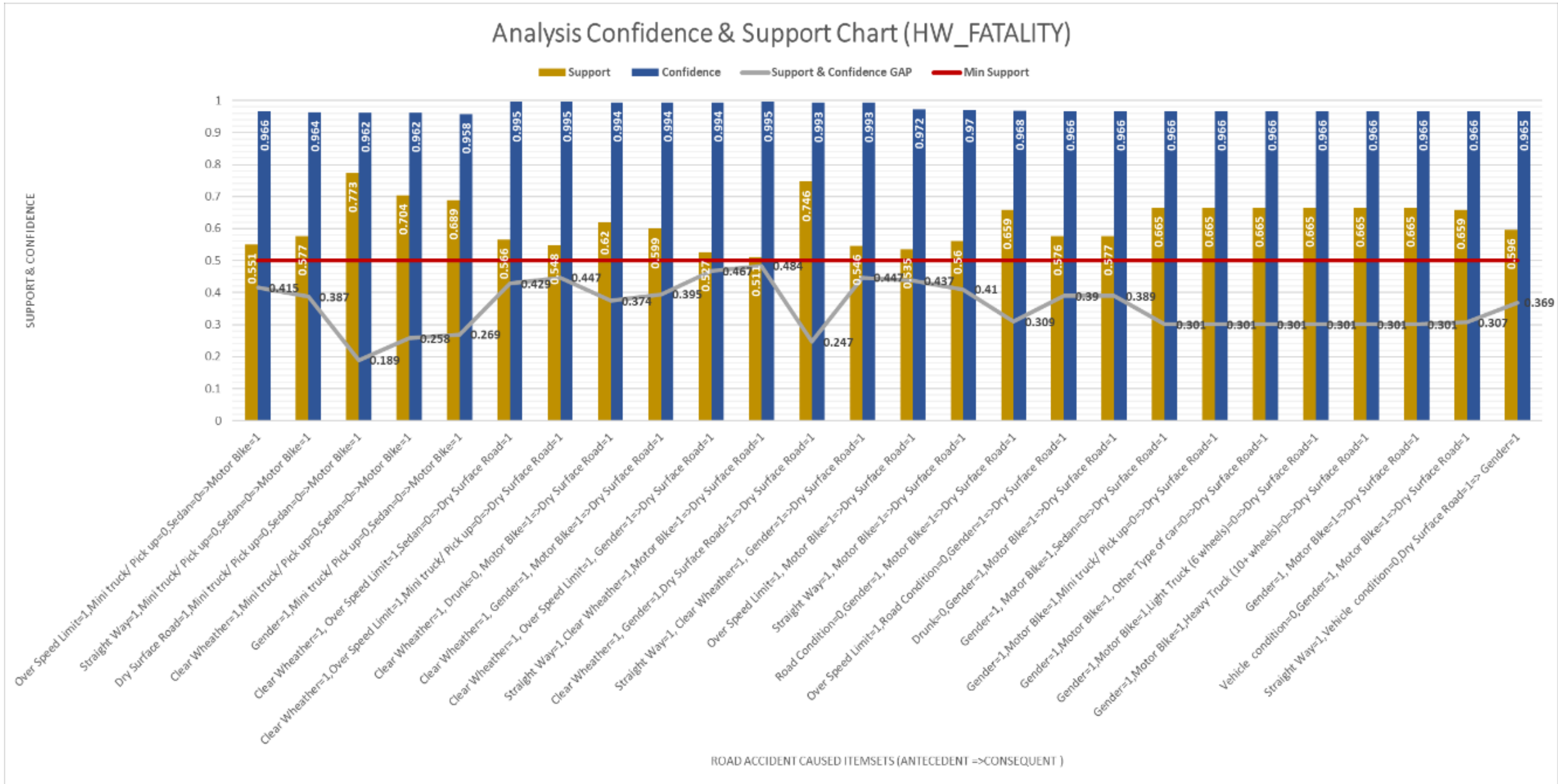

Figure 2.10 Confidence and support chart gap trend chart by interesting rules

2.5 Conclusion and Discussion

As a result of the association rule, the factors that enhance the likelihood of fatalities in highway road accidents are as follows:

1) Driver info - male
2) Driver behavior - over speed limit
3) Vehicle type - motorbike
4) Roadway - dry surface and straightaway
5) Weather - clear weather

When an accident occurs, all of these variables have a relationship and are linked to one another as the associated rule shows a potential cause of road accident fatalities, such that males riding motorcycles at speeds over the limit on straight roads in clear weather show increased risk for injury or death in traffic accidents, more than other conditions, with confidence levels increased from 0.5x, 06x, and 0.7x to 0.99x regarding if the consequences are motorcycle and dry surface road with high lift. As described at the opening pages, the higher the number of elements involved, the greater the possibility of an accident. Furthermore, the newly discovered straightway is a significant contributor, while transportation authority's exercise caution at intersections and on curve roads.

This might simply be due to the fact that when there is clear weather and a straight road with no curves, junctions, or turns, drivers frequently violate the speed limit as a result, which is more likely to cause accidents than when the weather is inclement, and it appears that males are driving faster than females. However, as of 2021, the current number of vehicles registered in Thailand is over 42 million, with motorbikes accounting for 50% of the total (DLT, 2021), potentially contributing to the largest number of fatalities from significant accidents. As Jomnonkwao et al. (2020) observed, motorcyclists are responsible for the vast majority of road fatalities, while prior studies showed different types of cars and motorcycles, such as rollover SUV/vans (Jafari Anarkooli et al., 2017), large truck (Huo et al., 2020; Jafari Anarkooli et al., 2017; Li et al., 2018), and pick-ups (Li et al., 2018). Additional research on motorcycle riders specifically, as well as other types of road users, may be conducted in the future. Aside from motorcycles, Sonal and Suman (2018) observed that external factors, such as weather conditions like fog, rain, and snow, show greater impacts in road accidents than internal factors, such as the drivers themselves. Meanwhile, Thailand's climate has no snow or ice, with rain contributing only roughly 5 months a year (June to October) and the chilly season taking 4 months (November to February). The remainder of the year is summer, with clear weather conditions and dry road surfaces contributing approximately 7 months a year. The rule discovered that fatalities have a high chance in clear weather on dry surfaces, which correlate to the chilly and summer seasons.

Highway junctions were determined to be the riskiest for all accidents (Kumar & Toshniwal, 2016). However, this study discovered that a major risk exists even on straightaways, since drivers usually violate the law about exceeding the speed limit on straightaways with no junctions. Bahiru et al. (2018) observed internal factors, such as gender, age, accident location, and vehicle type. Those were discovered to have less of an influence on the severity of road accidents, although being male is still one of the primary factors leading to highway fatalities.

With all the rules discovered from this study; policymakers may eliminate some of the factors implicated in highway traffic accidents. At least it should raise awareness of risky driver behaviors. Authorities are considering proposed laws to control speed limits on long straightaways by using light signs, warning signs, and cameras that closely monitor driving speeds, especially motorcycles.

The study used data from 2015-2020, although the last 2 years (2019–2020) of the COVID-19 pandemic, the government issued an order ordering people across the country to lock down and not allow cross-provincial travel, particularly between 10PM – 4AM. People are also apprehensive about travelling to separate zones on their own, which means they are not travelling much. As such, the numbers for 2019–2020 may not accurately reflect the real number of accidents and fatalities for country.

As a related rule for future research, further analysis may be extended to all types of roads, particular automobile types, criminal data, medical data, or nonhighway data to aid policymakers in formulating the best option feasible with solid data backup.

2.6 Study limitation and future study

The study used accident data from the COVID-19 pandemic, which caused the government to lock down and prohibit travel between provinces. People are also cautious to travel to the separated zones on their own, implying that they have not traveled extensively. As such, the numbers for 2019–2020 may not accurately reflect the real number of accidents and fatalities for country.

As a related rule capability, for future research, the further analysis may be extended to all types of roads, particular automobile types, criminal data, medical data, or nonhighway data to aid policymakers in choosing the most feasible options with solid data backup.

2.7 Reference

Abellán, J., López, G., & de Oña, J. (2013). Analysis of traffic accident severity using Decision Rules via Decision Trees. *Expert Systems with Applications*, *40*(15), 6047-6054. https://doi.org/10.1016/j.eswa.2013.05.027

Al Mamlook, R. E., Ali, A., Hasan, R. A., & Mohamed Kazim, H. A. (2019). Machine Learning to Predict the Freeway Traffic Accidents-Based Driving Simulation. Proceedings of the IEEE National Aerospace Electronics Conference, NAECON,

Anvari, M. B., Tavakoli Kashani, A., & Rabieyan, R. (2017). Identifying the Most Important Factors in the At-Fault Probability of Motorcyclists by Data Mining, Based on Classification Tree Models. *International Journal of Civil Engineering*, *15*(4), 653-662. https://doi.org/10.1007/s40999-017-0180-0

Bahiru, T. K., Kumar Singh, D., & Tessfaw, E. A. (2018). Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018,

Behnood, A., & Mannering, F. (2017). The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach. *Analytic Methods in Accident Research*, *14*, 41-53. https://doi.org/https://doi.org/10.1016/j.amar.2017.04.001

Ben-David, S. S.-S. a. S. (2014). <understanding-machine-learning-theory-algorithms.pdf>. *Cambridge University Press.* http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning

Bhavsar, R., Amin, A., & Zala, L. (2021). Development of Model for Road Crashes and Identification of Accident Spots [Article]. *International Journal of Intelligent Transportation Systems Research*, *19*(1), 99-111. https://doi.org/10.1007/|s13177-020-00228-z

Breiman, L. (2001). Mach Learn.

Bucsuházy, K., Matuchová, E., ZŮvala, R., Moravcová, P., Kostíková, M., & Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. Transportation Research Procedia,

Champahom, T., Jomnonkwao, S., Chatpattananan, V., Karoonsoontawong, A., & Ratanavaraha, V. (2019). Analysis of Rear-End Crash on Thai Highway: Decision Tree Approach. *Journal of Advanced Transportation*, *2019*, 1-13. https://doi.org/10.1155/2019/2568978

Champahom, T., Jomnonkwao, S., Watthanaklang, D., Karoonsoontawong, A., Chatpattananan, V., & Ratanavaraha, V. (2020). Applying hierarchical logistic models to compare urban and rural roadway modeling of severity of rear-end vehicular crashes. *Accident Analysis & Prevention*, *141*, 105537. https://doi.org/https://doi.org/10.1016/j.aap.2020.105537

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., & Guan, H. (2015). A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes. *Accident Analysis & Prevention*, *80*, 76-88. https://doi.org/https://doi.org/10.1016/j.aap.2015.03.036

Chen, C., Zhang, G., Yang, J., Milton, J. C., & Alcántara, A. D. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier. *Accident Analysis & Prevention*, *90*, 95-107. https://doi.org/https://doi.org/10.1016/j.aap.2016.02.002

Chen, M.-Y. (2012). Comparing Traditional Statistics, Decision Tree Classification And Support Vector Machine Techniques For Financial Bankruptcy Prediction. *Intelligent Automation & Soft Computing*, *18*(1), 65-73. https://doi.org/10.1080/10798587.2012.10643227

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

Cuenca, L. G., Puertas, E., Aliane, N., & Andres, J. F. (2018). Traffic Accidents Classification and Injury Severity Prediction. 2018 3rd IEEE International Conference on Intelligent Transportation Engineering, ICITE 2018,

Cunto, F. J. C., & Ferreira, S. (2017). An analysis of the injury severity of motorcycle crashes in Brazil using mixed ordered response models. *Journal of Transportation Safety & Security*, *9*(sup1), 33-46. https://doi.org/10.1080/19439962.2016.1162891

Demšar, J., Curk, T., Erjavec, A., Gorup, C., Hočevar, T., Milutinovič, M., Zupan, B. (2013). Orange: Data mining toolbox in python [Article]. *Journal of Machine Learning Research*, *14*, 2349-2353. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84885599052&partnerID=40&md5=75d2df52a0c46b5ab58ab08e1576114e

DLT. (2021). Department of Land Transportation. *https://www.dlt.go.th/th/public-news/view.php?_did=2806*.

Dongare, A., Kharde, R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, *2*(1), 189-194.

El Abdallaoui, H. E. A., El Fazziki, A., Ennaji, F. Z., & Sadgal, M. (2018). Decision Support System for the Analysis of Traffic Accident Big Data. Proceedings - 14th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2018,

Feng, M., Zheng, J., Ren, J., & Xi, Y. (2020). Association Rule Mining for Road Traffic Accident Analysis: A Case Study from UK. In *Advances in Brain Inspired Cognitive Systems* (pp. 520-529). https://doi.org/10.1007/978-3-030-39431-8_50

Geedipally, S. R., Turner, P. A., & Patil, S. (2011). Analysis of Motorcycle Crashes in Texas with Multinomial Logit Model. *Transportation Research Record*, *2265*(1), 62-69. https://doi.org/10.3141/2265-07

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn and TensorFlow. http://oreilly.com/catalog/errata.csp?isbn=9781492032649

Guido, A. C. M. S. (2017). Introduction to machinelearning with python. http://oreilly.com/catalog/errata.csp?isbn=9781449369415 (Third Release) (O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.)

Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review [Review]. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(4), 432-446. https://doi.org/10.1016/j.jtte.2020.05.002

Harb, R., Yan, X., Radwan, E., & Su, X. (2009). Exploring precrash maneuvers using classification trees and random forests [Article]. *Accident Analysis and Prevention*, *41*(1), 98-107. https://doi.org/10.1016/j.aap.2008.09.009

Helen, W. R., Almelu, N., & Nivethitha, S. (2019). Mining Road Accident Data Based on Diverted Attention of Drivers. Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018,

Hou, Q., Huo, X., Leng, J., & Cheng, Y. (2019). Examination of driver injury severity in freeway single-vehicle crashes using a mixed logit model with heterogeneity-in-means. *Physica A: Statistical Mechanics and its Applications*, *531*, 121760. https://doi.org/10.1016/j.physa.2019.121760

Huo, X., Leng, J., Hou, Q., & Yang, H. (2020). A Correlated Random Parameters Model with Heterogeneity in Means to Account for Unobserved Heterogeneity in Crash Frequency Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*, 036119812092221. https://doi.org/10.1177/0361198120922212

Jafari Anarkooli, A., Hosseinpour, M., & Kardar, A. (2017). Investigation of factors affecting the injury severity of single-vehicle rollover crashes: A random-effects generalized ordered probit model. *Accident Analysis & Prevention*, *106*, 399-410. https://doi.org/10.1016/j.aap.2017.07.008

John, M., & Shaiba, H. (2019). Apriori-Based Algorithm for Dubai Road Accident Analysis. Procedia Computer Science,

John, M., & Shaiba, H. (2022). Analysis of Road Accidents Using Data Mining Paradigm. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 68, pp. 215-223).

Jomnonkwao, S., Uttra, S., & Ratanavaraha, V. (2020). Forecasting Road Traffic Deaths in Thailand: Applications of Time-Series, Curve Estimation, Multiple Linear Regression, and Path Analysis Models. *Sustainability*, *12*(1). https://doi.org/10.3390/su12010395

Jou, R. C., Yeh, T. H., & Chen, R. S. (2012). Risk factors in motorcyclist fatalities in Taiwan. *Traffic Inj Prev*, *13*(2), 155-162. https://doi.org/10.1080/15389588.2011.641166

Jung, S., Qin, X., & Noyce, D. A. (2010). Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis & Prevention*, *42*(1), 213-224. https://doi.org/https://doi.org/10.1016/j.aap.2009.07.020

Khorashadi, A., Niemeier, D., Shankar, V., & Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: an exploratory analysis. *Accid Anal Prev*, *37*(5), 910-921. https://doi.org/10.1016/j.aap.2005.04.009

Kim, J.-K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention*, *50*, 1073-1081. https://doi.org/https://doi.org/10.1016/j.aap.2012.08.011

Kim, J. H., Kim, J., Lee, G., & Park, J. (2021). Machine Learning-Based Models for Accident Prediction at a Korean Container Port. *Sustainability*, *13*(16), 9137. https://www.mdpi.com/2071-1050/13/16/9137

Krull, K. A., Khattak, A. J., & Council, F. M. (2000). Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes. *Transportation Research Record*, *1717*(1), 46-54. https://doi.org/10.3141/1717-07

Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations [Article]. *Journal of Modern Transportation*, *24*(1), 62-72. https://doi.org/10.1007/s40534-016-0095-5

Kuşkapan, E., Çodur, M. Y., & Atalay, A. (2021). Speed violation analysis of heavy vehicles on highways using spatial analysis and machine learning algorithms [Article]. *Accident Analysis and Prevention*, *155*, Article 106098. https://doi.org/10.1016/j.aap.2021.106098

Li, Z., Chen, C., Wu, Q., Zhang, G., Liu, C., Prevedouros, P. D., & Ma, D. T. (2018). Exploring driver injury severity patterns and causes in low visibility related single-vehicle crashes using a finite mixture random parameters model. *Analytic Methods in Accident Research*, *20*, 1-14. https://doi.org/https://doi.org/10.1016/ j.amar.2018.08.001

Li, Z., Ci, Y., Chen, C., Zhang, G., Wu, Q., Qian, Z., Ma, D. T. (2019). Investigation of driver injury severities in rural single-vehicle crashes under rain conditions using mixed logit and latent class models. *Accident Analysis & Prevention*, *124*, 219-229. https://doi.org/https://doi.org/10.1016/j.aap.2018.12.020

Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., & Zhang, G. (2019a). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident; analysis and prevention*, *129*, 230-240.

Li, Z., Wu, Q., Ci, Y., Chen, C., Chen, X., & Zhang, G. (2019b). Using latent class analysis and mixed logit model to explore risk factors on driver injury severity in single-vehicle crashes. *Accident Analysis & Prevention*, *129*. https://doi.org/10.1016/j.aap.2019.04.001

Mafi, S., AbdelRazig, Y., & Doczy, R. (2018). Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups. In *Transportation Research Record* (Vol. 2672, pp. 171-183).

Malin, F., Norros, I., & Innamaa, S. (2019). Accident risk of road and weather conditions on different road types. *Accid Anal Prev*, *122*, 181-188. https://doi.org/10.1016/j.aap.2018.10.014

Mohamad, I., Jomnonkwao, S., & Ratanavaraha, V. (2022). Using a decision tree to compare rural versus highway motorcycle fatalities in Thailand. *Case Studies on Transport Policy*, *10*(4), 2165-2174. https://doi.org/https://doi.org/10.1016/j.cstp.2022.09.016

Mphela, T. (2020). Causes of road accidents in botswana: An econometric model [Article]. *Journal of Transport and Supply Chain Management*, *14*, 1-8, Article a509. https://doi.org/10.4102/jtscm.v14i0.509

Osman, M., Mishra, S., & Paleti, R. (2018). Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: Accounting for unobserved heterogeneity and age group differences. *Accident Analysis & Prevention*, *118*. https://doi.org/10.1016/j.aap.2018.05.004

Ospina-Mateus, H., Quintana Jiménez, L. A., Lopez-Valdes, F. J., Berrio Garcia, S., Barrero, L. H., & Sana, S. S. (2021). Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists [Article]. *Journal of Ambient Intelligence and Humanized Computing*, *12*(11), 10051-10072. https://doi.org/10.1007/s12652-020-02759-5

Ospina-Mateus, H., Quintana Jiménez, L. A., López-Valdés, F. J., Morales-Londoño, N., & Salas-Navarro, K. (2019). Using Data-Mining Techniques for the Prediction of the Severity of Road Crashes in Cartagena, Colombia. In *Communications in Computer and Information Science* (Vol. 1052, pp. 309-320).

Pakgohar, A., Tabrizi, R. S., Khalili, M., & Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. *Procedia Computer Science*, *3*, 764-769. https://doi.org/10.1016/j.procs.2010.12.126

PDPM. (2020). Thailand Department of Public Disaster Prevention and Mitigation. *https://www.disaster.go.th/en/*

Recal, F., & Demirel, T. (2021). Comparison of machine learning methods in predicting binary and multi-class occupational accident severity [Article]. *Journal of Intelligent and Fuzzy Systems*, *40*(6), 10981-10998. https://doi.org/10.3233/JIFS-202099

Rezapour, M., Mehrara Molan, A., & Ksaibati, K. (2020). Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. *International Journal of Transportation Science and Technology*, *9*(2), 89-99. https://doi.org/10.1016/j.ijtst.2019.10.002

RSC, T. (2019). Thailand Accident Research Center *Thailand Accident Research Center* https://www.thairsc.com/

Samerei, S. A., Aghabayk, K., Mohammadi, A., & Shiwakoti, N. (2021). Data mining approach to model bus crash severity in Australia [Article]. *Journal of Safety Research*, *76*, 73-82. https://doi.org/10.1016/j.jsr.2020.12.004

Santos, D., Saias, J., Quaresma, P., & Nogueira, V. B. (2021). Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction. *Computers*, *10*(12), 157. https://www.mdpi.com/2073-431X/10/12/157

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Se, C., Champahom, T., Jomnonkwao, S., Chaimuang, P., & Ratanavaraha, V. (2021). Empirical comparison of the effects of urban and rural crashes on motorcyclist injury severities: A correlated random parameters ordered probit approach with heterogeneity in means. *Accid Anal Prev*, *161*, 106352. https://doi.org/10.1016/j.aap.2021.106352

Se, C., Champahom, T., Jomnonkwao, S., Karoonsoontawong, A., & Ratanavaraha, V. (2021). Temporal stability of factors influencing driver-injury severities in single-vehicle crashes: A correlated random parameters with heterogeneity in means and variances approach. *Analytic Methods in Accident Research*, *32*, 100179. https://doi.org/https://doi.org/10.1016/j.amar.2021.100179

Shaheed, M. S., Gkritza, K., Zhang, W., & Hans, Z. (2013). A mixed logit analysis of two-vehicle crash seventies involving a motorcycle. *Accident; analysis and prevention*, *61*. https://doi.org/10.1016/j.aap.2013.05.028

Shweta, Yadav, J., Batra, K., & Goel, A. K. (2021). A Framework for Analyzing Road Accidents Using Machine Learning Paradigms. Journal of Physics: Conference Series, Siskind, V., Steinhardt, D., Sheehan, M., O'Connor, T., & Hanks, H. (2011). Risk factors for fatal crashes in rural Australia. *Accident Analysis & Prevention*, *43*(3), 1082-1088. https://doi.org/https://doi.org/10.1016/j.aap.2010.12.016

Sonal, S., & Suman, S. (2018). A framework for analysis of road accidents. 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research, ICETIETR 2018,

Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130.

Srikant, R. A. a. R. (1994). *Fast algorithms for mining association rules* Proceedings of the 20th International Conference on Very Large Data Bases, VLDB,, Santiago, Chile.

Tolles, J., & Meurer, W. J. (2016). Logistic Regression: Relating Patient Characteristics to Outcomes. *JAMA*, *316*(5), 533-534. https://doi.org/10.1001/jama.2016.7653

Webb, G. I. (2010). Naïve Bayes. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 713-714). Springer US. https://doi.org/10.1007/978-0-387-30164-8_576

Wei, F., Cai, Z., Liu, P., Guo, Y., Li, X., & Li, Q. (2021). Exploring Driver Injury Severity in Single-Vehicle Crashes under Foggy Weather and Clear Weather. *Journal of Advanced Transportation*, *2021*, 9939800. https://doi.org/10.1155/2021/9939800

WHO. (2018). World Health Organization: Global status report on road safety 2018. . *https://extranet.who.int/roadsafety/death-on-the-roads/*.

William J. Frawley, G. P.-S., and Christopher J. Matheus. (1992). Knowledge Discovery in Databases. https://doi.org/DOI: https://doi.org/10.1609/aimag.v13i3.1011 (An Overview. AI Magazine, 13(3), 57) (AAAI/MIT Press, Cambridge, MA)

Wu, Q., Zhang, G., Zhu, X., Liu, X. C., & Tarefder, R. (2016). Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways. *Accident Analysis & Prevention*, *94*, 35-45. https://doi.org/https://doi.org/10.1016/j.aap.2016.03.026

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1-37. https://doi.org/10.1007/s10115-007-0114-2

Xie, Y., & Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident; analysis and prevention*, *47*, 36-44. https://doi.org/10.1016/j.aap.2011.12.012

Yu, H., Yuan, R., Li, Z., Zhang, G., & Ma, D. T. (2020). Identifying heterogeneous factors for driver injury severity variations in snow-related rural single-vehicle crashes. *Accident Analysis & Prevention*, *144*, 105587. https://doi.org/https://doi.org/10.1016/j.aap.2020.105587

Yu, M., Zheng, C., & Ma, C. (2020). Analysis of injury severity of rear-end crashes in work zones: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*, *27*, 100126. https://doi.org/https://doi.org/10.1016/j.amar.2020.100126

Zhang, X. F., & Fan, L. (2013). A decision tree approach for traffic accident analysis of Saskatchewan highways. Canadian Conference on Electrical and Computer Engineering,

Zhou, M., & Chin, H. C. (2019). Factors affecting the injury severity of out-of-control single-vehicle crashes in Singapore. *Accident Analysis & Prevention*, *124*, 104-112. https://doi.org/10.1016/j.aap.2019.01.009