International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

To cite this article: Wiwik Budiawan, Sriyanto, Singgih Saptadi, Ary Arvianto, Harun Pamuji and Pertiwi Andarani (2022). DESIGN OF TRAFFIC ACCIDENT PREDICTION MODEL IN TOLL ROAD USING A DECISION TREE ALGORITHM, International Journal of Applied Science and Engineering Review (IJASER) 3 (6): 11-31

DESIGN OF TRAFFIC ACCIDENT PREDICTION MODEL IN TOLL ROAD USING A DECISION TREE ALGORITHM

Wiwik Budiawan[1], Sriyanto[2], Singgih Saptadi[3], Ary Arvianto[4], Harun Pamuji[5] and Pertiwi Andarani[6]

[1,2,3,4,5]Diponegoro University, Industrial Engineering Department,
Semarang, Indonesia
[6]Diponegoro University, Environmental Engineering Department,
Semarang, Indonesia

DOI: http://dx.doi.org/10.52267/IJASER.2022.3602

ABSTRACT

A toll road is a road that the users are obligated to pay, which is held to improve efficient transportation services. Although toll roads have relatively more ideal conditions than highway roads, many traffic accidents still occur on the road. Toll road managers collect operational data on toll roads, including daily traffic, weather, and accident data. One of the solutions to increase the level of toll road safety is to design an accident prediction model through data mining. In this paper, the prediction model was made using attributes according to the framework consisting of day, type of road surface, weather conditions, road surface conditions, time of occurrence, driver sex, and type of vehicle. The prediction model was built to predict certain areas' probability and severity of accidents. The prediction model is built using the decision tree algorithm. The results show that the attributes used can predict the severity of accidents with 39.73% accuracy. The most vulnerable area is in section B on 9 to 10 km, with a total number of accidents of 13.17% of total accidents.

KEYWORDS: Traffic Accidents, Toll Road, Data Mining, Prediction Model, Decision Tree.

1. INTRODUCTION

The toll road is a road that the users are obligated to pay which is held to improve efficient transportation services. Although toll roads have relatively more ideal conditions than the common arterial road, based on data from toll road operator company, there are still many cases of traffic accidents on the toll road. Based on data from toll road operator company and previous research, 501 accidents have occurred on the Semarang toll road from 2007 to 2017 (Budiawan et al., 2019) and 201 accidents have occurred on the Semarang-Batang toll road from 2018 to 2019 (Putri & Widowati, 2021). These traffic accidents are

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

caused by various factors consisting of human factors, vehicles, roads, and environmental factors. In addition, toll road administrators collect current operational data, such as daily traffic and accident data. If this data is not utilized, it will be accumulated and retained as useless data. Several efforts are required to improve traffic safety on toll roads through the utilization of existing data, namely: develop data warehouse (Budiawan et al., 2018), accident severity prediction (Budiawan et al., 2019), road and bridge system evaluation (Sutjahjo et al., 2020).

One of solutions to improve the level of toll road safety is to design accident prediction models through data mining by utilizing data related to accidents on toll roads. Data mining is a process to get information from a set of data that helps in decision making (Budiawan et al., 2018; Han et al., 2011). Several accidents-based Data Mining projects have been conducted previously. (Beshah & Hill, 2010) in the research of Mining Road Traffic Accident Data to Improve Safety: The Role of Integrated Factors on Accident Severity in Ethiopia uses classification methods including Decision Tree, Naive Bayes, and K-Nearest Neighbors. In another study, (Sowmya & Ponmuthuramalingam, 2013) in the research: Analyzing Road Traffic and Accident with Classification Techniques used classification methods including Naive Bayes, AdaBoostM1, Random Forest, Decision tree, and PART. Based on several studies before, the construction of the prediction model can be done by a classification method.

2. METHOD

In this study, the prediction model was designed by identifying the attributes of accident causes. The data was obtained from toll road operator company (2007-2017) and a previous study (2018-2019). Data were analyzed to find vulnerable areas to accidents based on daily cross-data and accident data. Vulnerable areas are analyzed based on the probability of occurrence of accidents in each area. Furthermore, the design of the prediction model was approached using the classification method. The prediction model was designed to predict the severity of accidents on toll roads based on accident attributes. In addition, the prediction model of the current study is a further study of previous studies. The severity level consists of property loss, minor injury, major injury, and fatal (Budiawan et al., 2019). The prediction model was built to predict the probability of accidents in certain areas on toll roads. Current study consisted of five main steps, namely: data collection, data preparation, design of prediction model based on decision tree algorithm, analysis of accuracy, and design of prediction model application.

Data Collection: the data consist of daily traffic and accident data. Data was obtained from toll road operator company period of 2007 to 2017 and previous research period of 2018 to 2019. The main framework for determining the attributes used in this paper is the research: Prediction and Analysis of Injury Severity in Traffic Systems using Data Mining Techniques (Khera & Singh, 2015).

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Moreover, in the data collection step, attributes of previous study are used with adjustments to accident data collected. The all attributes are shown in Table 1.

Table 1: Attributes based on the framework

Attributes	Type of Data	Description
Driver Sex	Text	(Khera & Singh, 2015)
Type of Vehicle	Text	(Khera & Singh, 2015)
Accident Time	Text	(Khera & Singh, 2015)
Accident Location	Text	(Khera & Singh, 2015)
Weather Condition	Text	(Khera & Singh, 2015)
Accident Severity	Text	(Khera & Singh, 2015)
Day	Text	Primary Data
Type of Road	Text	Primary Data
Road Surface	Text	Primary Data

Data Preparation: Data preparation step is several steps that taken before the data mining process, namely:

1. Data Selection

Data selection needs to be done before the data mining process. At this stage, the attributes of the data obtained from sources were selected by adjusting to the research framework (see Table 1). The results of this stage are the research data attributes. The selection data used for mining is stored in a file separate from the operational database.

2. Data Cleaning

Before the data mining process can be carried out, it needs a cleaning process, including removing duplicate data, checking inconsistent data, and correcting errors in data. The results of this stage are data with complete attribute values.

3. Data Transformation

Data transformation is needed so the data set is ready to be processed and produce a better analysis.

Accident probability: Accident probability is calculated by dividing the number of accidents from 2007 to 2017 with the number of vehicles traffics from 2007 to 2017 period. The results that obtained from this stage are the probability of accidents for each area on the toll road every one km. The equation for determining the probability of an accident can be written as follows (Han et al., 2011):

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Accident Probability =(Number of Accidents)/(Number of Vehicles crossed)

1)

Prediction Model Design: The data is divided into two parts. They are training data and testing data with a ratio of 0.6 and 0.4. Training data is data that used to build predictive models. While data testing is used to test the model. From existing studies, the classification and design of prediction models in case of accidents can be done using the Decision Tree algorithm. Decision Tree is a classification method to determine the highest attribute information gain to be set as the highest level that affects the data. Information gain will be obtained from the reduction of Entropy. In the Decision Tree, the model is represented in the form of a tree. Entropy values are written as follows (Han et al., 2011):

$$\mathrm{Entropy(S)} = \sum_{i}^{C} -p_i \log_2 p_i \quad 2)$$

Where c is the number of values contained in the target attribute. While pi states the portion or ratio between the number of samples in class and the number of all samples in the data set.

While the Information Gain value of time attribute (A) is written as follows (Han et al., 2011):

$$\mathrm{Gain(S,A)} \equiv \mathrm{Entropy(S)} - \sum_{v \epsilon \mathrm{Values(A)}} \frac{|S_v|}{|S|} \mathrm{Entropy}(S_v) \quad 3)$$

The Decision Tree method algorithm is shown in Figure 1.

International Journal of Applied Science and Engineering Review
IJASER
ISSN: 2582-6271
Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

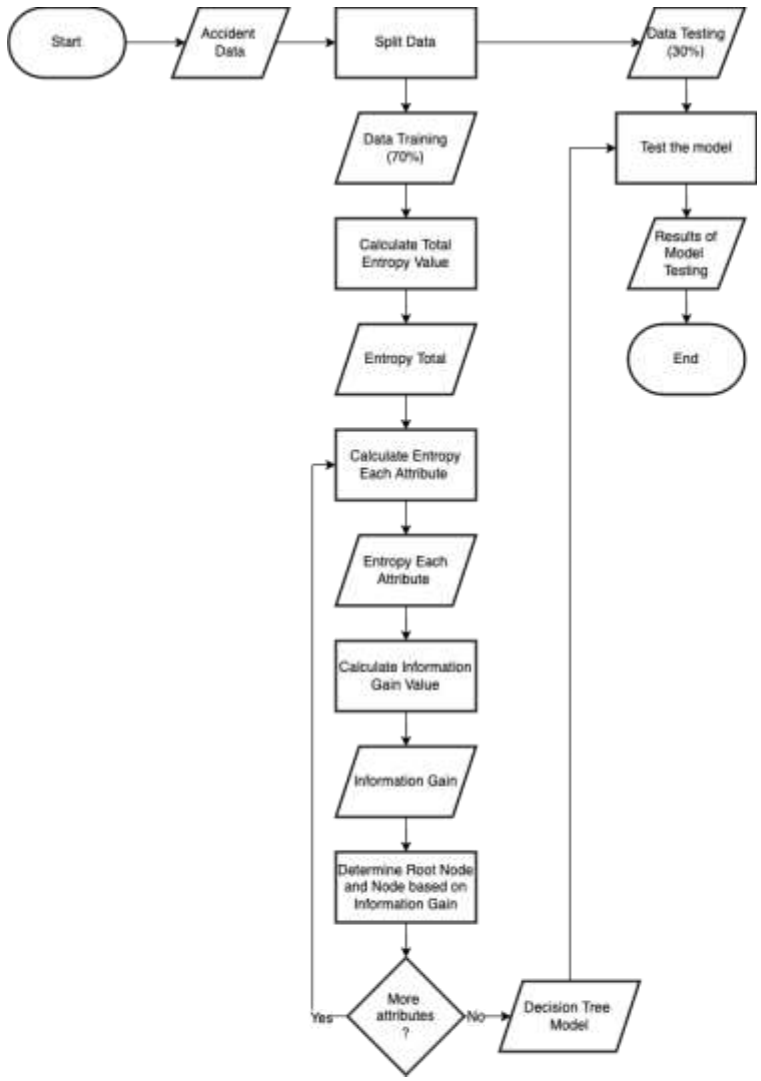

Figure 1: Decision Tree Algorithm

3. RESULT AND DISSCUSION

Results was presented into three parts (data selection, accident probability, and prediction model. In the data selection stage, the data was selected based on attributes needs according to the attributes determined in Table 1. Attributes consist of days of accidents (day), location of kilometers of accidents (km), direction of vehicle driving at the time of occurrence accidents (direction), time of accident (time), type of vehicle (vehicle), weather conditions (weather), type of lane where the accident (road type), driver sex (gender), type of road accident occur (surface), and the severity or type of accident (severity) as in table 2.

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Table 2: Data Attribute

Attributes	Type of Data	Description
Day	Text	Monday/ Tuesday/ Wednesday/ Thursday/ Friday/ Saturday/ Sunday
Type of Road	Text	Flat straight / Descend straight / Ascend straight/ Flat Curve/Descend Curve/ Ascend Curve
Weather	Text	Sunny/ Cloudy/ Foggy/ Dusty/ Smoky/ Drizzly/ Rainy
Surfaces	Text	Dry/ Wet/ Sandy
Time	Text	Morning/ Noon/ Afternoon/ Night
Driver sex	Text	Male/ Female
Vehicle Type	Text	Sedan/ Jeep/ Pick Up/ Minibus/ Bus/ Truck
Severity	Text	Property Loss/ Minor Injuries/ Major Injuries/ Fatal

1. Data Cleaning: The cleaning process including removing incomplete data and checking inconsistent data. In this paper, the data cleaning process is to eliminate some data without the attributes of time, vehicle, weather, alignment, or direction. From the total operational data as many as 501 data, after the cleaning process became 488 data.
2. Data Transformation: Data transformation is needed so that the data set is ready to be processed and produce a better analysis. Data transformation needs to be done because the operational data has a different format. In addition, the purpose of data transformation is to reduce data diversity so that predictive accuracy is better. Data is transformed into data that is ready to be processed

3.1. Accident Probability Measurement

The Accident Probability is calculated by dividing data on the number of accidents with the total data for vehicle traffic. The probability of an accident can be calculated by equation 1). The probability for each toll road area is obtained as shown as in Table 3.

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Table 3: Accident Probability

Section	Area			Frequency	Probability	%
A	0km	-	1km	18	4.00E-08	3.593
	1km	-	2km	10	2.22E-08	1.996
	2km	-	3km	12	2.66E-08	2.395
	3km	-	4km	18	4.00E-08	3.593
	4km	-	5km	22	4.89E-08	4.391
	5km	-	6km	27	6.00E-08	5.389
	6km	-	7km	10	2.22E-08	1.996
	7km	-	8km	8	1.78E-08	1.597
	8km	-	9km	9	1.55E-08	1.397
B	9km	-	10km	66	1.47E-07	13.174
	10km	-	11km	48	1.07E-07	9.581
	11km	-	12km	28	6.22E-08	5.589
	12km	-	13km	21	4.66E-08	4.192
	13km	-	14km	10	2.22E-08	1.996
	0km	-	1km	20	4.44E-08	3.992
C	1km	-	2km	35	7.77E-08	6.986
	2km	-	3km	27	6.00E-08	5.389
	3km	-	4km	15	3.33E-08	2.994
	4km	-	5km	23	5.11E-08	4.591
	5km	-	6km	12	2.66E-08	2.395
	6km	-	7km	8	1.78E-08	1.597
	7km	-	8km	8	1.78E-08	1.597
	8km	-	9km	1	2.22E-09	0.200
	9km	-	10km	9	2.00E-08	1.796
	10km	-	11km	36	7.99E-08	7.186

3.2. Prediction Model using Decision Tree

3.2.1. Entropy and Information Gain

Information Gain is a value used to determine which attributes are used to be used as a node first. Before determining the value of information gain, it is necessary to calculate the value of information in units of bits from a collection of objects. The way to calculate it is to use the concept of entropy. The entropy value is calculated based on equation (2). While the value of information gain is calculated using equation (3). The results of the Entropy and Information Gain values are shown in Table 4.

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Table 4: Information Gain

Attributes	Entropy	Information Gain
Time	1.660	0.031
Day	1.613	0.077
Vehicle	1.617	0.073
Driver Sex	1.675	0.016
Weather	1.655	0.035
Type of Road	1.666	0.024
Surface	1.686	0.005

From the information gain value, the attribute that specified as the first node or root node as the initial separator is the attribute of day because it has the highest information gain value which is equal to 0.077. To determine the next nodes are carried out with the same steps. In this paper, the node is obtained from the design of the model using RapidMiner.

3.2.2. Model Accuracy

The accuracy of the model is calculated by equation 4 as shown below.

$$Accuracy = \frac{correct\ predictions\ amount}{total\ predictions} \tag{4}$$

The results of model accuracy calculation on RapidMiner are shown in Table 5.

Table 5: Accuracy of Model

	True Property_ Loss	True Minor_ Injury	True Major_ Injury	True Fatal	Class precision
pred. Property Loss	22	27	19	3	30.99%
pred. Minor_ Injury	10	23	9	4	50.00%
pred. Major_ Injury	3	10	13	2	46.43%
pred. Fatal	0	0	1	0	0.00%
class recall	62.86%	38.33%	30.95%	0.00%	

International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Accuracy = 39.73%

Information that obtained from these data mining process is a model that can be used to predict accident in the toll road. Based on Table 5, it can be informed that the accuracy of the model is 39.79% to predict the severity of the accident. Based on the design of the traffic accident Prediction model in with decision tree algorithm, the results from the tree as the output of the mining process are summarized in Table 6.

Table 6: Result of Prediction Model

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
Friday	Bus	Flat Straight					Property Loss
	Bus	Ascend Straight					Minor Injury
	Bus	Descend Straight					Major Injury
	Bus	Ascend Curve					Major Injury
	Jeep						Minor Injury
	Minibus		Night	Sunny			Minor Injury
	Minibus		Night	Cloudy			Major Injury
	Minibus		Morning				Minor Injury
	Minibus		Noon				Minor Injury
	Minibus		Afternoon				Property Loss
	Pick Up	Flat Straight					Minor Injury
	Pick Up	Ascend Straight					Major Injury
	Pick Up	Descend Straight					Minor Injury
	Pick Up	Descend Curve					Property Loss
	Sedan	Flat Straight		Sunny			Minor Injury

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
	Sedan	Ascend Straight		Sunny			Property Loss
	Sedan			Rainny			Property Loss
	Truck	Flat Straight	Night				Major Injury
	Truck	Flat Straight	Morning				Property Loss
	Truck	Flat Straight	Noon	Sunny			Property Loss
	Truck	Flat Straight	Noon	Rainny			Minor Injury
	Truck	Flat Straight	Afternoon				Minor Injury
	Truck	Ascend Straight					Minor Injury
	Truck	Descend Straight	Night	Sunny			Minor Injury
	Truck	Descend Straight	Night	Rainny			Major Injury
	Truck	Descend Straight	Night	Cloudy			Property Loss
	Truck	Descend Straight	Morning				Property Loss
	Truck	Descend Straight	Noon				Minor Injury
	Truck	Descend Straight	Afternoon				Major Injury
	Truck	Flat Curve					Property Loss
	Truck	Ascend Curve					Minor Injury
	Truck	Descend Curve					Major Injury
Thursday	Bus			Sunny			Minor Injury
	Bus			Rainny			Property Loss
	Jeep						Property Loss

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
	Minibus	Flat Straight		Sunny			Minor Injury
	Minibus	Flat Straight		Drizzly			Property Loss
	Minibus	Ascend Straight					Property Loss
	Minibus	Descend Straight	Night				Property Loss
	Minibus	Descend Straight	Morning				Property Loss
	Minibus	Descend Straight	Noon				Minor Injury
	Minibus	Flat Curve					Property Loss
	Pick Up	Flat Straight					Minor Injury
	Pick Up	Ascend Straight					Major Injury
	Pick Up	Descend Straight					Minor Injury
	Sedan	Flat Straight					Property Loss
	Sedan	Ascend Straight					Property Loss
	Sedan	Descend Straight					Minor Injury
	Sedan	Ascend Curve					Property Loss
	Truck	Flat Straight	Night				Major Injury
	Truck	Flat Straight	Morning				Minor Injury
	Truck	Flat Straight	Noon				Major Injury
	Truck	Flat Straight	Afternoon				Minor Injury
	Truck	Ascend Straight		Sunny			Minor Injury
	Truck	Ascend Straight		Drizzly			Major Injury

International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
	Truck	Descend Straight	Night				Property Loss
	Truck	Descend Straight	Morning				Minor Injury
	Truck	Descend Straight	Noon				Property Loss
	Truck	Descend Straight	Afternoon				Property Loss
	Truck	Flat Curve					Major Injury
	Truck	Ascend Curve					Property Loss
	Truck	Descend Curve					Property Loss
Sunday	Bus	Descend Straight	Night				Major Injury
	Bus	Descend Curve	Night				Property Loss
	Bus		Morning				Major Injury
	Bus		Noon				Minor Injury
	Jeep						Major Injury
	Minibus	Flat Straight	Night	Sunny			Major Injury
	Minibus	Flat Straight	Night	Cloudy			Minor Injury
	Minibus	Flat Straight	Morning				Minor Injury
	Minibus	Flat Straight	Noon				Minor Injury
	Minibus	Ascend Straight					Minor Injury
	Minibus	Descend Straight					Minor Injury
	Minibus	Descend Curve					Major Injury
	Pick Up			Sunny			Minor Injury

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
	Pick Up			Drizzly			Property Loss
	Pick Up			Rainny			Minor Injury
	Sedan			Sunny			Property Loss
	Sedan			Drizzly			Minor Injury
	Sedan			Cloudy			Minor Injury
	Truck	Flat Straight	Night				Major Injury
	Truck	Ascend Straight	Night				Minor Injury
	Truck	Descend Straight	Night	Sunny			Minor Injury
	Truck	Descend Straight	Night	Cloudy			Major Injury
	Truck	Ascend Curve	Night				Minor Injury
	Truck		Morning	Sunny			Property Loss
	Truck		Morning	Cloudy			Minor Injury
	Truck		Noon				Property Loss
	Truck	Flat Straight	Afternoon				Minor Injury
	Truck	Descend Straight	Afternoon				Major Injury
Wednesday	Bus		Night	Sunny			Property Loss
	Bus		Night	Drizzly			Major Injury
	Bus		Morning				Major Injury
	Bus		Noon				Major Injury
	Bus		Afternoon				Property Loss

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
	Jeep						Minor Injury
	Minibus		Night				Property Loss
	Minibus		Morning	Sunny			Minor Injury
	Minibus		Morning	Drizzly			Property Loss
	Minibus		Noon				Property Loss
	Pick Up		Night				Property Loss
	Pick Up		Noon	Sunny			Major Injury
	Pick Up		Noon	Drizzly			Minor Injury
	Truck	Flat Straight	Night				Property Loss
	Truck	Ascend Straight	Night				Property Loss
	Truck	Descend Straight	Night	Sunny			Minor Injury
	Truck	Descend Straight	Night	Drizzly			Property Loss
	Truck	Flat Curve	Night				Property Loss
	Truck	Descend Curve	Night				Property Loss
	Truck	Flat Straight	Morning				Fatal
	Truck	Ascend Straight	Morning				Minor Injury
	Truck	Descend Straight	Morning				Property Loss
	Truck	Flat Straight	Noon				Minor Injury
	Truck	Descend Straight	Noon				Minor Injury
	Truck	Ascend Curve	Noon				Property Loss
	Truck	Descend	Noon	Sunny			Major

International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
		Curve					Injury
	Truck	Descend Curve	Noon	Cloudy			Minor Injury
	Truck		Afternoon				Minor Injury
Saturday	Bus	Flat Straight					Minor Injury
	Bus	Descend Straight					Major Injury
	Bus	Ascend Curve					Major Injury
	Bus	Descend Curve	Morning				Property Loss
	Bus	Descend Curve	Afternoon				Fatal
	Jeep						Minor Injury
	Minibus	Flat Straight					Minor Injury
	Minibus	Ascend Straight		Sunny			Major Injury
	Minibus	Ascend Straight		Rainny			Minor Injury
	Minibus	Descend Straight		Sunny			Property Loss
	Minibus	Descend Straight		Drizzly			Major Injury
	Minibus	Descend Straight		Rainny			Major Injury
	Minibus	Descend Straight		Cloudy			Fatal
	Minibus	Ascend Curve					Major Injury
	Pick Up		Night				Minor Injury
	Pick Up		Afternoon				Major Injury
	Sedan	Flat Straight	Night	Sunny			Minor Injury
	Sedan	Flat Straight	Morning	Sunny			Major

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
							Injury
	Sedan	Flat Straight		Cloudy			Property Loss
	Sedan	Descend Straight					Property Loss
	Sedan	Descend Curve	Night				Minor Injury
	Sedan	Descend Curve	Noon				Minor Injury
	Sedan	Descend Curve	Afternoon				Major Injury
	Truck	Flat Straight	Night	Sunny			Major Injury
	Truck	Descend Straight	Night	Sunny			Minor Injury
	Truck	Descend Curve	Night	Sunny			Minor Injury
	Truck	Flat Straight	Morning	Sunny			Major Injury
	Truck	Descend Straight	Morning	Sunny			Minor Injury
	Truck	Flat Straight	Noon	Sunny			Minor Injury
	Truck	Descend Straight	Noon	Sunny			Minor Injury
	Truck	Descend Curve	Noon	Sunny			Major Injury
	Truck	Flat Straight	Afternoon	Sunny			Property Loss
	Truck	Descend Straight	Afternoon	Sunny			Major Injury
	Truck	Descend Curve	Afternoon	Sunny			Property Loss
	Truck			Drizzly			Major Injury
	Truck			Rainny			Minor Injury
	Truck	Flat Straight		Cloudy			Minor Injury
	Truck	Ascend		Cloudy			Property

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
		Straight					Loss
	Truck	Descend Curve		Cloudy			Minor Injury
Tuesday	Bus		Night				Major Injury
	Bus		Noon	Sunny			Property Loss
	Bus		Noon	Drizzly			Minor Injury
	Jeep		Noon				Property Loss
	Jeep		Afternoon				Minor Injury
	Minibus	Flat Straight	Night				Property Loss
	Minibus	Flat Straight	Morning				Minor Injury
	Minibus	Ascend Straight		Sunny			Minor Injury
	Minibus	Ascend Straight		Cloudy			Major Injury
	Minibus	Descend Straight					Minor Injury
	Minibus	Ascend Curve					Property Loss
	Pick Up		Night				Major Injury
	Pick Up	Flat Straight	Noon				Minor Injury
	Pick Up	Descend Curve	Noon				Major Injury
	Sedan				Male		Minor Injury
	Sedan				Female		Property Loss
	Truck	Flat Straight	Night	Sunny			Major Injury
		Ascend Straight	Night	Sunny			Property Loss
		Descend	Night	Sunny			Major

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
		Straight					Injury
		Ascend Straight	Night	Drizzly			Major Injury
		Descend Straight	Night	Drizzly			Minor Injury
			Night	Rainny			Minor Injury
			Night	Cloudy			Minor Injury
		Descend Straight	Morning				Property Loss
		Descend Curve	Morning				Minor Injury
			Noon				Minor Injury
			Afternoon				Major Injury
Monday	Minibus	Flat Straight		Sunny			Property Loss
	Pick Up	Flat Straight		Sunny			Minor Injury
	Sedan	Flat Straight		Sunny			Property Loss
	Truck	Flat Straight	Night	Sunny			Property Loss
	Truck	Flat Straight	Noon	Sunny			Minor Injury
	Truck	Flat Straight	Afternoon	Sunny			Property Loss
	Bus	Ascend Straight	Night	Sunny			Major Injury
	Sedan	Ascend Straight	Night	Sunny			Minor Injury
	Truck	Ascend Straight	Night	Sunny			Minor Injury
		Ascend Straight	Morning	Sunny			Minor Injury
		Ascend Straight	Noon	Sunny			Property Loss
		Ascend	Afterno	Sunny			Property

International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Day	Vehicle	Type of Road	Time	Wheater	Driver gender	Type of Surface	Prediction
		Straight	on				Loss
		Descend					Major
	Bus	Straight	Night	Sunny			Injury
		Descend					Property
	Bus	Straight	Morning	Sunny			Loss
		Descend					Property
	Bus	Straight	Noon	Sunny			Loss
		Descend					Minor
	Sedan	Straight		Sunny			Injury
		Descend					Property
	Truck	Straight		Sunny			Loss
							Minor
		Flat Curve		Sunny			Injury
		Descend					Property
		Curve		Sunny			Loss
	Minib			Drizzl			Minor
	us			y			Injury
	Pick			Drizzl			Minor
	Up			y			Injury
				Drizzl			Major
	Sedan			y			Injury
				Drizzl			Major
	Truck			y			Injury
							Major
			Night	Rainny			Injury
							Property
			Noon	Rainny			Loss
			Afterno				Property
		Flat Straight	on	Rainny			Loss
		Descend	Afterno				Minor
		Straight	on	Rainny			Injury
							Minor
				Cloudy			Injury

4. CONCLUSION

Based on this research we conclude as follows:

1. The vulnerable areas are identified by calculating the probability of accidents for each point on the Semarang Toll Road per 1 km. The most vulnerable area is in section B on 9 to 10 km which contributes accidents 13.17% of total accident from 2007 to 2017 period.

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

2. Attributes that take effect on accident severity in this model are time, day, vehicle, gender, weather, alignment or road type, and surface of road.
3. By considering the attributes of time, day, vehicle, gender, weather, alignment or road type, and surface of road, the prediction model that built in this study can predict the severity of accidents on toll roads but with 39.73% of accuracy. At the same time, the prediction model can also predict the probability of accidents at the Semarang toll road in each of 1 km.

REFERENCES

1. Beshah, T., & Hill, S. (2010). *Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia*.
2. Beshah, T., Abraham, A., & Grosan, C. (2005). *Rule Mining and Classification of Road Traffic Accidents Using Adaptive Regression Trees (Vol. 6).*
3. Bustami. (2014). *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. Jurnal Informatika Vol. 8*, 884-898.
4. Harahap, G. (1995). *Masalah Lalu Lintas dan Pengembangan Jalan (DPU).*
5. Indonesia, R. (2016). *Peraturan Pemerintah nomor 34 tentang Jalan.*
6. Khera, D., & Singh, W. (2015). *Prediction and Analysis of Injury Severity in Traffic System using Data Mining Techniques. ACCA 2015 (2)*, 1-7.
7. Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: CV. ANDI OFFSET.
8. Santosa, B., & Umama, A. (2018). *Data Mining dan Big Data Analytics*. Yogyakarta: Penebar Media Pustaka.
9. Sartika, D., & Sensuse, D. I. (2017). *Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. Jatisi Vol. 1*.
10. Shanti, R S., & Geeha, Ramani R. (2012). *Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques* (Vol. I). San Francisco: WCECS
11. Sowmnya, B., & Pornmuthurmalingan, P. (2013). A*nalyzing the Road Traffic and Accidents with Classification Techniques*. *International Journal of Computer Trends and Technology*, 183-188.
12. Suyanto. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data. Bandung: Informatika Bandung*.
13. Wicaksono, Ichsan, M., & Yunanto. (2006). A*NALISIS KECELAKAAN LALU LINTAS (Studi Kasus Jalan Tol Semarang Seksi ABC*.

ISSN: 2582-6271

Vol. 3, Issue.6, Nov-Dec 2022, page no. 11-31

Author Profile

Wiwik Budiawan graduated from Diponegoro University, majoring in Industrial Engineering. He continued his study at the Bandung Institute of Technology. He obtained Ph.D. from the Toyohashi University of Technology. He has received some awards and grants concerning safety transportation, accident prevention, and academic excellence during his study. He did research on accident analysis and prevention in Mass Transportation. Since 2012, he has been a lecturer at the Department of Industrial Engineering, Diponegoro University, joined the Decision Support System Laboratory and Japan sleep research member.