IJTE, Vol. 10, No. 3, Winter 2022, pp. 1139-1162
Research Paper

# Extracting the Inconsistencies in the Urban Highways Considering the Crash Occurrences

Shahin Shabani[1,*], Kamran rahimov[2], Ali hatefy shafty[3]

Received: 2022/06/01 Accepted: 2022/09/14

## Abstract

The present study aims to identify the inconsistencies resulting from the interferences between the geometric features of the highway, access points and traffic control devices by extracting the frequent association rules and discovering the patterns that contribute to the recurrent crashes. A case study was conducted on a 117-kilometer urban highway around Mashhad. The data for five years of crashes were procured from the corresponding authorities and placed on the network under a geo spatial software environment. Then, the all features of the highway were collected. The trip times between the nodes and existing features were calculated via coding based on the data obtained from one of the route planning software. The operating speed was subsequently estimated. According to the obtained databank and considering the geometric features of the road (horizontal and vertical), the access points, road width, operating speed and position of the traffic control devices (signs and speed control camera), efforts were made to divide the road to the sections. Allocating the crash data to the sections and using the FP-Growth Algorithm, the frequent rules affecting the recurrent crash occurrences were extracted. The results showed that there are different combinations of the geometric road-related factors such as flat-straight alignment and horizontal curve along with cases like right-hand exit, consecutive entrance and exit, and reduced road width in addition to the presence of speed control camera and advanced direction signs on the highway sections together cause inconsistencies like the recurrent crash occurrences.

**Keywords:** Association rules, Crash analysis, Data-mining, Geo spatial system, Inconsistency

* Corresponding author. E-mail: shabani@pnu.ac.ir
[1] Department of Civil Engineering, Payame Noor University(PNU), P. O. Box, 19395-3697 Tehran, Iran
[2] Department of Civil Engineering, Payame Noor University(PNU), P. O. Box, 19395-3697 Tehran, Iran
[3] PhD candidate, Department of Civil Engineering. Payame Noor University (PNU), P. O. Box, 19395-3697 Tehran, Iran

# 1. Introduction

In this research, the geometrical conditions of the road and the features on the road are considered as the cause of crashes, and every crash is a response to these conditions in different states. In this research, the road and its features are considered as goods and every crash as a customer. Through the application of the algorithm of associational rules, frequent data places of damage to road users and arrangements where drivers often have crashes have been identified.a new perspective in extracting knowledge from crash data, and applying this insight to other roads leads to improving the knowledge of road design and reducing the number of crashes.at the start, the road is divided to section according to the location of features and the scope of the effects of features and crashes that occurred on the road and the change in the operating speed that occurs on the road. By using the associative rules mining algorithm, each crash is considered as a customer and each section of the road is considered as a goods, which is selected by the customer with specified profile included the time of the crash and the type of the involved car in crash occurrence.In this study, the crashes have been categorized in pre-defined clusters in the format of the created sections. In fact, data classification is done through supervised clustering. The type of approach to road dividing to section in this research and the application of functional speed change in road sections is a unique approach that can add data-based knowledge and experience to the principles of road geometric design.

## 1.1. Abbreviations

| | |
|---|---|
| **w_c** | Cross section without changing the width compared to the previous section |
| **aw** | Cross section with an increase in width compared to the previous section |
| **dw** | Cross section with reduced width compared to the previous section |
| **si_1** | The presence of a advanced direction signs in the study section |
| **ex_l** | Exitance from the left side of the road |
| **en_l** | Entrance from the left side of the road |
| **ca_1** | The presence of a speed control camera in the studied section of the road |
| **en_r** | Entrance from the right side of the road |
| **ex_r** | Exitance from the right side of the road |
| **d_c** | Direct connection from the right side to the road |
| **f_d_c** | More than one direct connection from the right to the road |

# 2. Literature Review

Crashes are among the cases taken into consideration in the transportation industry. Based on the WHO's report, Iran, with a rate of 20.5 deaths per 100000 people, is in 113th rank amongst the 175 countries, and it is not a good rank. There is a need for planning and cooperation by the executive organs for reducing the fatality rate and the injuries resulting from the road crashes in Iran. Crashes are one of the five significant causes of death in the world (WHO, 2018). Therefore, the issue has been taken into account in the country's development plans. The corresponding organs have been asked to take necessary measures to reducing the crashes. Based on the information from the country's forensic medicine organization, the statistics of the death originating from the crashes on the Suburban roads during the years between 2017 and 2018 and during the first nine months of 2019 have been descending. Thus, identifying the various wherein more crashes occur and determining the various features related to crashes in these locations would help understand the various conditions. The present study's objective is to identify the geometric properties and features and the speed rates that cause the crashes. (Chen & Jovanis, 2000) figured out that the analysis of the collection of the consecutive metadata using the traditional statistical methods may lead to such specific problems as sporadic data in the large probability tables. The statistical models also have their specific hypotheses that, along with their rejection, can result in some mistakes. Considering the limitations of the statistical methods, data mining is the method of choice for analyzing crashes. Data mining comprises methods for extracting novel, implicit and latent information from the metadata.(Pang-Ning, Steinbach, & Kumar, 2006) investigated various data mining methods such as association rule extraction, clustering, and classification to analyze crashes. Data mining can be described as extracting hidden and unknown data from a large volume of information. Knowledge discovery methods like clustering and classification and, recently, association rule extraction have been applied for analyzing crashes by researchers from some countries. In a research performed at Limburg University, Belgium, (Geurts, Wets, Brijs, & Vanhoof, 2003) ,the association rule extraction was examined for comprehending the various conditions of the highly frequent crash spots in the network of the Belgian roads.(Tesema, Abraham, & Grosan, 2005) used the comparative regression tree as a supportive decision-making system for crashes in Ethiopia. In another research carried out at Spain's University of Granada (Abellán, López, & OñA, 2013), various decision trees were developed to create decision-making rules for analyzing the two-way suburban roads. They found out that the improper lighting conditions of the road and the safety and security barriers intensively influence the occurrence frequency of the crashes. (Depaire, Wets, & Vanhoof, 2008) used clustering to analyze the crash data in Belgium. They suggested that cluster-based analysis of the crash data can provide better information than the data analysis without clustering. (Tavakoli Kashani, Shariat-Mohaymany, & Ranjbari, 2011) used categorization and CART regression to analyze the crash data in Iran. It was made clear in this article that not using the seat belt, prohibited overtaking and prohibited high-speed driving influence the intensity of the crashes. In a study conducted at the South Korea Data Mining Institute, (Kwon, Rhee, & Yoon, 2015) They used Naive Bayes classification method and decision tree classification algorithm to analyze the dependence of road safety related factors.(Kuhnert, Do, & McClure, 2000) In a case study, CART and MARS algorithms were applied to control the epidemiology of the injuries resulting from motor vehicle crashes. They also identified the possible areas of the crash risks created by the drivers.(Sohn & Shin, 2001) used three data mining methods, namely decision tree, neural network and logistic regression, to discover the important factors influencing the crash intensity. (Chong, Abraham, & Paprzycki, 2004) used a decision tree for intensity analyses. They figured out that many factors, including seat belts, alcohol use and insufficient lighting, are involved in lethal injuries caused during the crashes. (Chang & Wang, 2006) developed a CART model to analyze the relationship between the intensity of the injuries incurred by the drivers and the road environment. (Abugessaisa, 2008) used clustering and decision trees to perform interactive discoveries based on the pruning method and link analysis method in line with identifying and recognizing the notable patterns. In addition, (Wong & Chung, 2008) used several methods to discover the factors involved in the intensification of the crash. They noted that a dangerous crash is created due to various factors. (Zelalem, 2009) studied the responsibility of the drivers and used ID3, J48 and multilayer perceptron (MLP) algorithms for discovering the related factors and noted that many of the factors, including the driver's license type, age and experience, directly influence the intensity of the crashes. In a research conducted at Payam-e-Noor University in Iran , (Pakgohar, Tabrizi, Khalili, & Esmaeili, 2011) the CART algorithm and multiple logistic regression (MLR) were applied. It was figured out that the CART method gives relatively more precise results. (Demirel, Emil, & Duzgun, 2011) used the remote sensing method to analyze the regional scale and manage the effective environmental factors. They concluded that this technology could be helpful in the prevention of some kinds of crashes. (Sanmiquel, Rossell, & Vintró, 2015) analyzed the primary factors of the crashes using Bayesian classification and decision tree. The articles extensively apply the other association data mining algorithms to extract the frequent pattern systems and construct decision-making rules. In the first place, these algorithms are based on the minimum support and least confidence. However, most of them produce many results that prevent the decision-makers from choosing the most related item. Therefore, it is important to suggest an approach capable of helping the decision-makers make a correct choice. In research carried out in the center of China's traffic studies, (Weng, Zhu, Yan, & Liu, 2016) new relationship-related rules have been suggested for analyzing the properties and factors involved in the crashes in Karry Region based on the drivers' properties and speed limits. (Montella, 2011) investigated the properties of the crashes and the factors influencing the crashes in the urban U-turn through the association rule approach to discover their interrelationships in various kinds of crashes. (S Das, 2014) investigated the method of extracting applied association rules for discovering the potential patterns of the crash data under rainy conditions. In a research carried out at China's Shanghai University the association rules were executed about the crashes in Shanghai. A method was offered to extract strong highway crash data rules automatically. In this research, use was made of clustering and APRIORI algorithm methods to take into account the type of the vehicles involved in the crashes, climatic conditions, speed limit , tunnel and horizontal road features (Z Gao, R Pan, R Yu, X Wang, 2018).

(S Priya, 2018) proposed a method for classifying the change in the final crash information within the format of various categorizations. In this article, APRIORI clustering and algorithm were also applied to discover the semantics of the crash occurrence. In another research carried out by China's transportation engineering faculty, (Xu, Bao, Wang, & Liu, 2018) descriptive statistics were utilized to show the important features of the crashes from the perspective of the road user's behavior, vehicles' conditions, geometric features, and environmental conditions. In this article, association rule extraction and clustering were applied to identify factors influencing the crashes in serious crashes. The paving conditions, road degree, climatic conditions, driver's features and the involved vehicles' kinds and time of the crash's occurrence were all considered. In another research conducted by Texas's transportation faculty, (Das et al., 2019) a supervised APRIORI association rule algorithm was applied to discover the pedestrians' crash database. In another research carried out by Italy's faculty of healthcare and epidemiology, (Gariazzo, Stafoggia, Bruzzone, Pelliccioni, & Forastiere, 2018) the association rules were explored to investigate the relationship between the use of mobile phones at various population levels and its relationship with the crash occurrence in large urban regions.(Deng, Zeng, & Shen, 2018) utilized association rules to classify the factors involved in crashes and analyzed them to determine the degree or impact rate of a crash.(Feng, Zheng, & Ren, 2019) predicted the processes of crime commitment for discovering the key factors and identifying the patterns related to crime occurrence. Using England's crash data for the years between 2005 and 2017, (M Feng, J Zheng, J Ren, Y Xi, 2019) considered the highly frequent rules of crash occurrence concerning speed limit and environmental conditions like light, weather and road slope and then applied the APRIORI algorithm and clustering method to extract the rules with high lift and high support. The results showed a strong correlation between crashes and environmental conditions. Thus, the identification of these places wherein the crash frequency is high, and their subsequent analysis is very useful for the identification of the factors influencing the crash frequency in such places.

## 3. Review on the Application of Unsupervised Clustering Methods in Traffic Safety Studies

In a research conducted at Hildsheim University in Germany (Meißner & Rieck, 2022), frequent association rules algorithm and time series clustering of crash data were used to predict the number of crashes in different time and environmental conditions.The data mining approaches used in this research include the extraction of frequent time patterns and frequent environmental conditions in the occurrence of crashes, along with time series clustering, time series classification and scoring, in order to accurately select clusters and finally, an crash prediction model has been created.In another research that was conducted in Tabriz University Civil Engineering Faculty (Samerei, Aghabayk, Mohammadi, & Shiwakoti, 2021) , the algorithm of association rule mining was used on the created clusters to identify the factors affecting the severity of bus crashes in Australia. In this research, in the first step, the crash data was divided into homogeneous clusters by using the systematic hierarchical clustering algorithm, and then, the frequent association rules algorithm was implemented on the homogeneous clusters.In another study conducted at the Vergata Faculty of Engineering in Rome (Comi, Polimeni, & Balsamo, 2022), the k-means method and Cohenen neural networks were used to cluster the crash data, and the clustering results were used to build an crash severity prediction model based on the type of crash and allowable speed, traffic conditions, road conditions and the number of vehicles involved in the crash were used. Decision tree algorithm, frequent association rules and neural networks are used for modeling. The decision tree algorithm and neural network have performed better in building the crash severity prediction model. The findings of this research show that an crash prediction approach should be applied in a combined manner so that its application to the machine learning method is effective in crash prediction.In a research conducted in the Faculty of Science and Technology Engineering of Iran (Shirmohammadi, Hadadi, & Saeedian, 2019), the behavior of drivers while driving has been clustered. In this research, involvement in crashes was evaluated by driver behavior questionnaire and driver skill questionnaire. Drivers have been classified in terms of participation in crashes based on the driver's own statements and the driver's attitude towards traffic safety and risk perception and driving violations and driver errors. Finally, drivers are ranked in clusters based on crashes and annual intentional and unintentional fines were classified. In this research Hierarchical linkage clustering was used to determine the number of clusters or subgroups. This method is suitable for determining the number of clusters; however, this research recommends using K-means cluster analysis after determining the number of clusters using this method. Nafis (Nafis, Alluri, Wu, & Kibria, 2021) used Agglomerative Hierarchical Clustering to categorize crash data into homogeneous groups from the crash data set. Then, random forest algorithm was used in each cluster created to identify and prioritize important variables in each cluster. Finally, with the application of the affecting variables identified in the decision trees and Classification and Regression Trees models was used to construct and extract crash severity patterns .Momeni Kho (Momeni Kho, Pahlavani, & Bigdeli, 2022) have used four tree clustering algorithms named CHAID, CART, C4.5 and C5.0 to classify crashes. The conclusion was that the C5.0 algorithm with a kappa rate of 94% had the most appropriate performance in clustering crash data. The author states that the results are quite general because the crash data lacks a wide range of variables, such as various driver-related factors.

More information can be extracted by having more additional variables in crashes such as age, gender, type of license and education level of the driver.Regarding the effect and modelling of spatial parameters in the analysis of crashes, which is based on systematic analysis and carried out on a global-longitudinal scale with lengths of about kilometres, is the research done by Sadeghi (Farhangi, Sadeghi-Niaraki, Nahvi, & Razavi-Termeh, 2022) with the aim of spatial modelling of the risk of crashes caused by driver drowsiness through data mining algorithms. In this research, seven spatial criteria, including road section length, road width, slope angle, speed limit, land use/cover, distance to service area and distance to speed control camera were selected as effective criteria in modelling. The effect of the criteria was applied in the modelling by using the fuzzy method and three risk maps were prepared. A noteworthy point regarding the application of this research for fuzzy clustering based on the location of the crash in short longitudinal scales will be more effective when the accuracy of recording the location of the crash in the urban road network is increased and the crash data is recorded with accurate location.Finally, Arun Chand (Chand, Jayesh, & Bhasi, AB, 2021) , while reviewing the various methods used in the construction of crash prediction models, states that the purpose of data analysis is much more important than the technique used in modelling. What is important in modelling is the insight that is taken in making the model.The hypothesis of this research is that the inconsistency caused by the arrangement of features and the geometric properties of the road sections will show its effect on the speed change between the sections and the crashes that occurred on the road. The simultaneous application of the impact of features based on the Road Geometric Design standards and taking into account the operational speed change between sections and taking into account the crashes that have occurred on the road is a new perspective that has led to the clustering of crashes in format of the road sections. In this research, the position of injury crashes has been considered as evidence of inconsistency, and the position of injury crashes on the road has been taken through the computerized police information registration form and the address registration form that is recorded in emergency injury crashes, so the region of the crash is considered in this research .The average length of the sections created in this research is 334 meters, and the creation of sections is based on the geometric performance of the road and the features on the road and the operational speed change between the road sections and taking into account the occurrence of injury crashes within the scope of the features on the road.The method of supervised clustering in crashes is the unique perspective of this research in contrast to the systematic methods of clustering crashes. The missing link in the clustering of the previous researches is not considering the supervised operation of the created clusters. The knowledge extracted in this research will be considered in the road design stage and will provide a clear perspective to the road designer in the road geometric design stage. Applying this point of view to other types of roads will improve the standards of road geometric design.

## 4. Data Gathering Method

Since the present study aims at extracting the conditions and factors in the highly frequent crash locations on the Mashhad ring expressway. The length of the processed expressway was about 117 kilometers. Figure 1 shows the study area.

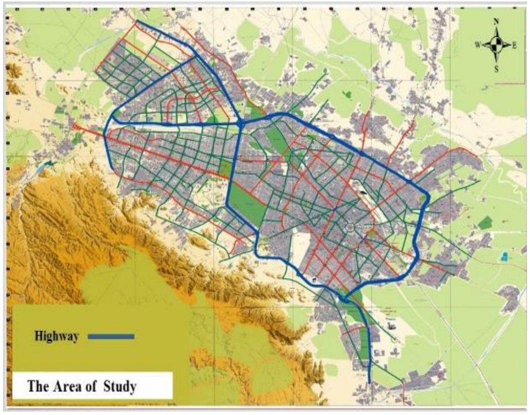


**Figure 1. the study area**

## 5. Dividing the Road into Separate Sections

In order to generalize the results obtained from the model, the studied road was divided from the perspective of its traffic performance. To do so, the following principles were taken into consideration in road dividing. The horizontal alignment of the sections existing on the road was divided into two sets of curves and lines, and the horizontal alignment of the sections existing on the road was divided into three sets flat, convex and concave. The features existing on the path were collected by inquiring Mashhad Traffic & Transportation Organization (MTTO), field observation and aerial photos. In the locations where the exiting and entering take place through the Traffic Middle Island, the 250-meter distance before the exit to the middle space between the exit and the entrance has been considered as the impact area of the exit and the same area between the exit and the entrance to 250 meters after the entrance has been considered as the impact area of the entrance. Figure 2 shows the impact area of the entrance/exitance.

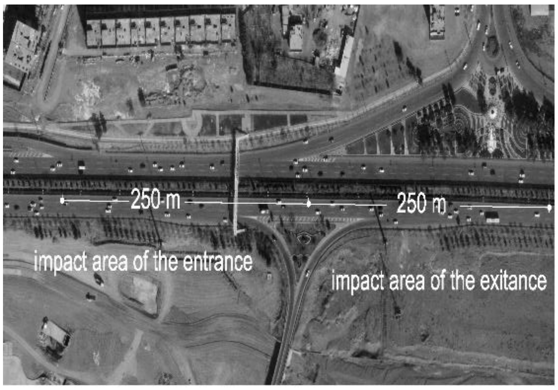


**Figure 2. The impact area of the entrance/exitance**

In locations where there is only one entrance from the right or left side, the area between the entrance up to 250 meters after the entrance has been considered as the impact area of the entrance; and, in locations where there is only one exit from the right or left side, the area between the exit up to 250 meters before the exit has been considered as the impact area of the exit. Figure 3 shows the impact area of single entrance/exitance.

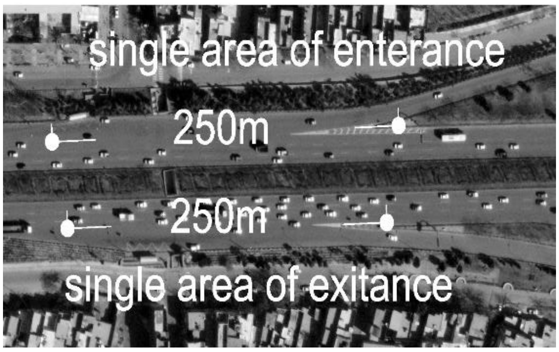


**Figure 3. The impact area of single entrance/exitance**

In case the single-entrance area overlaps with the single-exit area, the areas are Combined and considered as one single area. Figure 4 shows the impact area of overlap.


**Figure 4. The impact area of single overlap**

The distance from 150 meters before the speed control camera and the advance direction sign has been considered in coordination with the operating speed change of the sections. In this article, the sections' operating speeds have been extracted based on coding in Python Software for 20-time intervals during a full day and night in 1-hour-and-12-minute intervals. based on the routing software's constraints in responding to the number of the users' sent requests for inquiring the trip time between the nodes as well as according to the current features of the road and the 85-percent speed has been transformed based on the length of the section's current status into temporal operating speed. Next, and based on the traffic relationships, the spatial operating speed has been computed and considered for the corresponding road section. In the studied road, the dividing has been carried out based on the speed parameter in this way that the speed monitoring has been carried out from the beginning of the road, and, wherever the speed difference reaches over five km/h, a spot is created for consideration in the final dividing. The next feature that has been taken into consideration in dividing is the change in the width of the road. Along the road and wherever width changes, a spot is created for consideration in the final discretization. The geometric parameters of the road's horizontal alignment are divided into three sets a simple curve, a reverse curve and a direct line in dividing the road. It is worth mentioning that each curve is considered separately for the combined curve wherever the difference between the curves' radii exceeds 150 meters. If the direct distance between two right-turning and left-turning curves be below 250m, it is considered a reverse curve and, otherwise, as two separate curves and a direct section. In terms of the road's vertical alignment, the sections are divided into three flat, concave and convex sets, and wherever the vertical geometry of the road is changed, a spot is inserted for consideration in the dividing. The road's separate dividing has been created based on such parameters as the horizontal alignment, vertical alignment, access roades, width changes and speed changes and effective traffic features along the road via the sections' superimposition. Sections shorter than 50 meters have been merged based on the speed dependency and similarity in a section with respect to that of the section before or after it. The created sections have been categorized from the perspective of the section's length into three sets according to figure 5 based on the k-means clustering method.

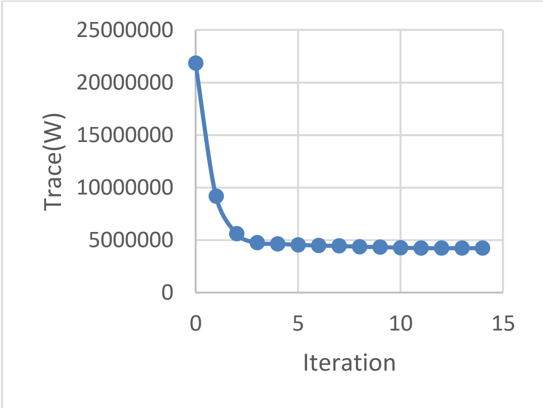


**Figure 5. Diagram indicating the clustering of the sections' lengths**

Table 1: Results of the sections' lengths clustering in meters

**Table 1. Sections' lengths clustering in meters**

| Class | Y1 | Sum of weights | Within-class variance |
|---|---|---|---|
| **1** | 445.613 | 133.000 | 8776.749 |
| **2** | 201.987 | 268.000 | 5375.532 |
| **3** | 887.677 | 37.000 | 45455.214 |

Based thereon, the sections have been categorized into three sets as given in table 2.

**Table 2. Classification of sections' lengths in meters**

| Class | lower limit | upper limit |
|---|---|---|
| **1** | 50.5 | 323.1 |
| **2** | 325.5 | 647.1 |
| **3** | 666.7 | 1399.7 |

A total of 438 sections have been created, and the classification of sections according to the specifications of geometric sections is shown in Table 3.

**Table 3. Classification of sections' lengths in based on horizontal alignment**

| Class of length/horizontal alignment | reverse curve | simple curve | straigth line | Grand Total |
|---|---|---|---|---|
| **Class 1** | 12 | 82 | 174 | **268** |
| **Class 2** | 0 | 33 | 100 | **133** |
| **Class 3** | 0 | 4 | 33 | **37** |
| **Grand Total** | **12** | **119** | **307** | **438** |

The classification of sections based on vertical alignment specifications is shown in Table 4.

**Table 4. Classification of sections based on vertical alignment**

| Class of length/vertical alignment | reverse curve | simple curve | straigth line | Grand Total |
|---|---|---|---|---|
| **concave** | 0 | 14 | 24 | **38** |
| **convex** | 0 | 5 | 9 | **14** |
| **flat** | 12 | 100 | 274 | **386** |
| **Grand Total** | **12** | **119** | **307** | **438** |

The sections have also been categorized based on speed into four sets as obtained in the k-means clustering analyses shown in figure 6.

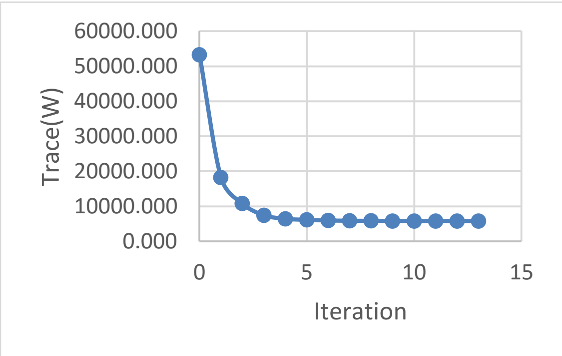


**Figure 6. Diagram showing the sections' speeds clustering**

the results of the sections' speed clustering in km/h is shown in Table 5.

**Table 5. The sections' speed clustering in km/h**

| Class | Y1 | Sum of weights | Within-class variance |
|---|---|---|---|
| **1** | 41.8548674 | 43 | 42.26351123 |
| **2** | 57.64552813 | 110 | 13.26534991 |
| **3** | 68.71005898 | 180 | 7.631257608 |
| **4** | 77.48936673 | 105 | 11.39932959 |

Based thereupon, the sections have been classified into four sets, as explained in Table 6.

**Table 6. Classification of sections' speeds in km/h**

| Class | lower limit | upper limit |
|---|---|---|
| **1** | 22.2 | 49.72 |
| **2** | 50.17 | 63.16 |
| **3** | 63.2 | 73.01 |
| **4** | 73.3 | 88.4 |

The classification of the speed of the sections based on the specifications of horizontal alignment is shown in Table 7.

**Table 7. Classification of sections' speeds based on horizontal alignment**

| Class of speed/horizontal alignment | reverse curve | simple curve | straight line | Grand Total |
|---|---|---|---|---|
| **Class 1** | 3 | 6 | 10 | **19** |
| **Class 2** | 1 | 20 | 48 | **69** |
| **Class 3** | 5 | 39 | 129 | **173** |
| **Class 4** | 3 | 54 | 120 | **177** |
| **Grand Total** | 12 | 119 | 307 | **438** |

In order to analyze the crashes based on the vehicles involved in the car crashes, classification was done within the format of buses, heavy trucks, light cars, pedestrians, motorcycles and vans.Next, using spatial joining in spatial reference system software, the reference spots of the crash sites were jointed to the created sections.The crash data were gathered from the corresponding organizations within the format of the procedures for supporting the university thesis and guidelines of data extraction. There were limitations regarding the data extraction, including confidentiality of the data.

The data-gathering method is occasionally based on field observations with the spatial factors causing the crash occurrences being reconstructed in this method. In order to create a crash's data-base, the police and Emergency Organization's data were used. Data belongs to 2013-2019 years. The road features were collected through field investigation and approved drawings. Mentioned data was collected for modeling the crash occurrences. This analysis is based on the frequency of the injury crashes at any time and the crash occurrence highway's characteristics. The opposite come-and-go lanes have been taken separately into account, and the properties of each way being separately extracted. The output was created within the format of a shapefile, including 3089 crash records. Crashes that occurred in the area of auxiliary lane are excluded from the crash database. The total number of crashes involved in data mining analysis is 2653 crash records. The distribution of the number of crashes according to the type of vehicle involved in the crash and the horizontal alignment of studied road is shown in Table 8.

**Table 8. Distribution of the number of crashes against the horizontal alignment of the studied road**

| Vehicle type /horizontal alignment | straight line | simple curve | Reverse curve | Grand Total |
|---|---|---|---|---|
| **Bus** | 21 | 12 | 0 | **33** |
| **Pedestrian crash** | 298 | 115 | 1 | **414** |
| **Heavy vehicles** | 50 | 26 | 0 | **76** |
| **Bike** | 3 | 2 | 0 | **5** |
| **car** | 1139 | 419 | 7 | **1565** |
| **Truck** | 29 | 7 | 0 | **36** |
| **motorcycle** | 255 | 111 | 2 | **368** |
| **Minibus and van** | 19 | 9 | 0 | **28** |
| **pickup truck** | 94 | 31 | 3 | **128** |
| **Grand Total** | **1908** | **732** | **13** | **2653** |

In order to categorize the time of crash occurrences, taking into account the two perspectives of safety performance and traffic performance, the classification of the time of the crash occurrences has been done in six categories in this research. 1- early morning (2 to 6 am) 2- Morning 6 to 10 am (with peak label) 3- Noon 10 to 2 pm (with peak label) 4- in the Evening from 2 pm to 6 pm (with label peak) 5- The evening is divided in the time period from 6 pm to 10 pm 6- night is divided from 10 pm to 2 am. Table. 9 shows this time category.

**Table 9. Time classification of crashes occurred on the road**

| Time stamp | Time Interval (Hour) | Time category |
|---|---|---|
| **off peak** | 2 To 6 | **early morning** |
| **peak** | 6 To 10 | **the morning** |
| **peak** | 10 To 14 | **Noon** |
| **peak** | 14 To 18 | **in the evening** |
| **off peak** | 18 To 22 | **evening** |
| **off peak** | 22 To 2 | **at night** |

The distribution of the number of crashes of the types of vehicles involved in the crash according to the time classification of table No. 2 number is shown in Table 10.

**Table 10. Distribution of the number of crashes that occurred according to the classified time**

| Vehicle type/time of crash | peak | off peak | Grand Total |
|---|---|---|---|
| **Bus** | 18 | 15 | 33 |
| **Pedestrian crash** | 226 | 188 | 414 |
| **Heavy vehicles** | 40 | 36 | 76 |
| **Bike** | 2 | 3 | 5 |
| **car** | 866 | 699 | 1565 |
| **Truck** | 19 | 17 | 36 |
| **motorcycle** | 209 | 159 | 368 |
| **Minibus and van** | 19 | 9 | 28 |
| **pickup truck** | 73 | 55 | 128 |
| **Grand Total** | **1472** | **1181** | **2653** |

The distribution of the number of crashes in different time of occurrence is shown in Table 11.

**Table 11. Distribution of the number of crash types that occurred in different times**

| Vehicle type/time of crash | 2 To 6 | 6 To 10 | 10 To 14 | 14 To 18 | 18 To 22 | 22 To 2 | Grand Total |
|---|---|---|---|---|---|---|---|
| **Bus** | 3 | 5 | 4 | 9 | 8 | 4 | **33** |
| **Pedestrian** | 12 | 57 | 87 | 83 | 116 | 59 | **414** |
| **Heavy** | 8 | 13 | 12 | 15 | 12 | 16 | **76** |
| **Bike** | 0 | 1 | 0 | 1 | 3 | 0 | **5** |
| **car** | 71 | 233 | 312 | 321 | 308 | 320 | **1565** |
| **Truck** | 5 | 8 | 2 | 9 | 6 | 6 | **36** |
| **motorcycle** | 10 | 62 | 66 | 81 | 77 | 72 | **368** |
| **Minibus** | 3 | 11 | 4 | 4 | 6 | 0 | **28** |
| **pickup** | 4 | 22 | 24 | 27 | 22 | 29 | **128** |
| **Grand Total** | **116** | **412** | **511** | **550** | **558** | **506** | **2653** |

The distribution of the number repetitions of crashes that occurred in the total of three vertical alignment (flat, convex and concave) according to the features that are in each section is shown in Table 12.

**Table 12. Distribution of the number of crashes that occurred in area of each feature**

| feature type / horizontal alignment | straight line | simple curve | Reverse curve | Grand Total |
|---|---|---|---|---|
| **w_c** | 377 | 255 | 3 | **635** |
| **dw** | 970 | 290 | 3 | **1263** |
| **aw** | 561 | 187 | 7 | **755** |
| **ca_1** | 363 | 144 | 0 | **507** |
| **si_1** | 617 | 278 | 2 | **897** |
| **ex_r_1** | 585 | 187 | 1 | **773** |
| **ex_r_2** | 31 | 9 | 0 | **40** |
| **en_r_1** | 421 | 110 | 2 | **533** |
| **en_r_2** | 19 | 7 | 0 | **26** |
| **en_l** | 49 | 0 | 1 | **50** |
| **ex_l** | 34 | 13 | 3 | **50** |
| **d_c** | 44 | 1 | 0 | **45** |
| **f_d_c** | 94 | 34 | 0 | **128** |

The distribution of the number of features on the created sections is according to Table 13.

**Table 13. Distribution of the number of features that occurred in every section**

| Feature type/horizontal | straight line | simple curve | reverse curve | Grand Total |
|---|---|---|---|---|
| **w_c** | 123 | 51 | 7 | **181** |
| **aw** | 87 | 33 | 3 | **123** |
| **dw** | 98 | 36 | 2 | **136** |
| **si_1** | 73 | 37 | 2 | **112** |
| **ex_l** | 8 | 7 | 3 | **18** |
| **en_l** | 7 | 0 | 2 | **9** |
| **ca_1** | 59 | 20 | 0 | **79** |
| **en_r_1** | 67 | 19 | 2 | **88** |
| **en_r_2** | 2 | 1 | 0 | **3** |
| **ex_r_1** | 76 | 27 | 1 | **104** |
| **ex_r_2** | 1 | 1 | 0 | **2** |
| **d_c** | 15 | 1 | 0 | **16** |

| Feature type/horizontal | straight line | simple curve | reverse curve | Grand Total |
|---|---|---|---|---|
| **f_d_c** | 3 | 1 | 0 | **4** |

Suffix number 1 or 2 is on the feature of the entrance from the right side or the exitance from the right side regarding the sections that due to the short distance between the two entrances or the two exitances, their impact length overlaps with each other and both entrance or exitance features are created on the section. It works simultaneously. Every crash occurred in a section means a customer who has chosen that section as a goods in one of the time categories with one of the modes of the vehicle. So far, many algorithms have been presented to explore association rules, and FP-Growth is one of the most successful. This algorithm stores the data in the database in a compact form in a tree called FP-tree, and then discovers frequent events by recursively building the FP-tree.

# 6. Rules Extraction Method

In this stage, and due to the fact that the data featured multiple dimensions and associations, the continuous data were transformed into discrete data. In other words, discretization was carried out, and the data were stored in different groups with binary formats within the framework of the zero and one classes. The highly repeated rule extraction algorithm FPGrowth has been applied herein. This method is commonly used in purchase basket analyses, and it tries to extract the common rules between the various features within the system of metadata. This method elaborates the rules of the various properties within a large dataset. Dataset D has n transactions, with every transaction T being a subset of the D dataset. Set I include the following purchase items: $I = \{I_1, I_2, ...I_n\}$. It is stated that A occurs in transaction T wherein $T \in D$ if and only if $A \subseteq T$ and $A \rightarrow B$ is the rule that determines $A \subset I, B \subset I$ and $A \cap B = \varnothing$ As for the crash data, an association rule is a rule that determines what various features together cause the occurrence of a crash. There are various scales determining the power of a rule.

## 6.1. Support Criterion

Support criterion $A \rightarrow B$ determines the percentage of the A and B's occurrences together in one set (Kumar & Toshniwal, 2015)According to equation 1:

$$S_P = \frac{P(A \cap B)}{N} \quad (1)$$

N is the total number of a crash's records. Support is recognized as the repetition constraint. The set of items that go beyond a threshold is considered a highly frequent pattern. These frequent amounts are next applied for producing association rules. These amounts generally range between zero and one; the more these amounts get close to unity, the more it is indicative of the two items being correlated. This scale and the confidence scale are the most important scales applied in various articles for comparing the rules in diverse algorithms for exploring the association rules. It is worth mentioning that the determination of the threshold amount by a user might cause the elimination of valuable rules, the support of which falls below a threshold value. Thus, this criterion is not alone determinative.

## 6.2. Confidence Criterion

The confidence parameter of the rule $A \rightarrow B$ includes the probability with which A and B occur together divided by the probability of A's occurrence (Kumar & Toshniwal, 2015). According to equation 2:

$$Conf(A \rightarrow B) = \frac{SUP(A \rightarrow B)}{SUP(A)} \quad (2)$$

The higher the confidence level of the rule $A \rightarrow B$, the higher the chance of B's occurrence with A. This relationship displays conditional probability. This amount is also a value between zero and one. For example, a rule with a confidence rate of 98% indicates that the left side of the rule is authentic if the left side of it is correct. This scale, along with the support scale, is a supplement for the evaluation of the association rules. The problem is that there might be a rule with a high confidence rate but being considered not valuable it, there might be even rules with not necessarily high support and confidence but are considered valuable by experts. Thus, those scales capable of identifying the less valuable rules featuring high support and confidence are defined.

### 6.3. Lift Criterion

This criterion obtains the amount of independence between A and B, and it is a value between zero and infinity. (Kumar & Toshniwal, 2015) According to equation 3:

$$Lift(A \rightarrow B) = \frac{Conf(A \rightarrow B)}{SUP(B)} \quad (3)$$

Values close to unity are expressive of the A and B's independence. If a scale is found below unity, it is indicative of the idea that A and B are negatively related and the larger the value than unity, the more it is indicative of the idea that A provides more information about B. In this state, the attractiveness of the rule $A \rightarrow B$ is appreciated higher. This criterion is symmetrical with respect to the left and right sides of the rule, meaning that if the left and right sides of the rule are substituted with one another, the amount of this scale does not change. The combination of this scale and the support and confidence scales is amongst the best combinations of association rules' evaluation. The problem of this scale lies in its sensitiveness to the number of the samples in a dataset, especially in regard to the set of the small transactions; hence another scale has been introduced to make up for this shortcoming.

### 6.4. Conviction Criterion

This criterion compensates for some weaknesses of the confidence and lifts scales. (Kumar & Toshniwal, 2016).According to equation 4:

$$Conv(A \rightarrow B) = \frac{1 - SUP(B)}{1 - Conf(A \rightarrow B)} \quad (4)$$

Unlike the lift criterion, this scale is not symmetrical, meaning that if the left side is substituted by the right side, the numerical values would be different. The amount of this scale would be infinity if the confidence of the rule is equal to unity, and if A and B be independent, this scale would be equal o unity. The definable range for this scale is from 0.5 to infinity. The higher the amount of this scale, the more it would be indicative of the rule's attractiveness. Consider a rule with very low support but with a confidence rate equal to unity; it will accordingly have a conviction rate equal to unity based on the definition. Now, imagine a rule with very high support but a confidence rate of 98%; the conviction rate would be a very high bit below infinity. Surely, the second rule is a lot more valuable than the first rule.

### 6.5. PS Criterion

The PS value is 0 when A and B are mutually independent of each other. Otherwise, PS > 0 when there is a positive relationship between the two variables, and PS < 0 when there is a negative relationship (Xiong, 2006). According to equation 5:

$$p - s(A \rightarrow B) = SUP(A \rightarrow B) - SUP(A) * SUP(B) \quad (5)$$

This criterion varies between -0.25 and +0.25. When the two happenings are completely independent of one another, the scale would be -0.25. The more the values of this criterion get closer to positive amounts, the simultaneous occurrence of two happenings would be more probable.

### 6.6. Laplace Criterion

Laplace is a sort of post-pruning method that makes calculations for every rule (Wedyan, 2014) . According to equation 6:

$$Laplace(R) = \frac{(P_C(R) + 1)}{(P_{tot}(R) + M)} \quad (6)$$

Pc(R) is the number of the experimental samples consistent with the Rule R. Ptot(R) is the number of the experimental samples that are in accordance with the "if" section of the rule, with M being the number of categories. This value ranges from zero to unity. The more this value gets closer to unity, then, the higher the accuracy of the rule would be.It is worth mentioning that RapidMiner Software possesses FP-Growth analysis sub-procedures that have been applied in this research. This software has the capability of running the codes written in Python language, as well, and it can implement any change or optimization. The output obtained from this stage is considered knowledge based on experience. The association rules constitute one of the favorable data-mining methods enabling the identification of the relationships between the various properties of crashes. In this study, the features influencing the occurrence of car crashes are seminally identified via spatial analysis in a GIS environment. In this stage, and considering the allowed speed on Mashhad’s highway and the materials of the journal, 415 250-meter decision intervals have been taken into account. Then, the association rule equations are applied for the specification of these places. Association rules showcase various factors in relation to the emergence of crashes in various spots with different distances from one another. Although the dataset is limited to some of the chosen properties, this study’s approach is the extraction of the useful hidden information from the data that can be applied for devising some preventive measures in these places.

## 7. Rules' Rating and Evaluation

After finding the collection of the final rules, they should be evaluated using a database. This study has made use of the accuracy scale of the Laplace method. It is worth mentioning that this temporal criterion is used when the final rules are calculated and rated. To do so, the rating of the rules has been carried out considering a 0.8 minimum data confidence rate and a 0.15 minimum support rate. After rating the rules, a number of them that feature the highest precision in terms of the aforesaid criterions are introduced, respectively. In sum, 134 rules have been created within the format of 12 tables with support rates ranging from 0.1 to 0.4 and confidence rates ranging from 0.7 to 0.95. in between, the evident and repetitive rules are eliminated, and some of the rules are extracted based on engineering judgments and by considering the accuracy criterion of the Laplace method. The results of the rules’ support and confidence rates’ calculation with respect to different lift scale rates have been exhibited in figure 7.

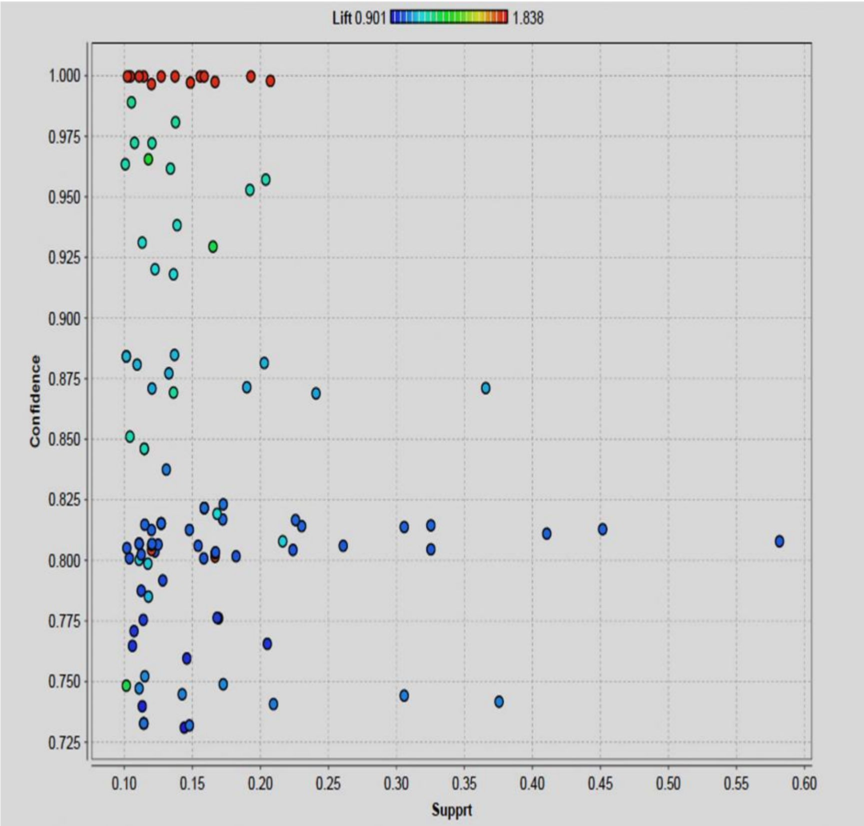


**Figure 7. The results of the rules' support and confidence rates' calculation with respect to different lift criterion rates**

Table 14 shows the selected rules based on the highest confidence and the highest support. In Table 15, the selected rules are arranged based on the Laplace accuracy criterion. In the following, the results obtained from the selected rules were applied to the data and the predominant mode were extracted based on these rules. This method is knowledge that is extracted from data analysis based on performance and engineering knowledge.

**Table 14. Sorting rules by confidence and Laplace criteria**

| NO. | Premises | Conclusion | Support | Confidence | Laplace | P-S | Lift | Conviction |
|---|---|---|---|---|---|---|---|---|
| 1 | straight line , Time to peak , The third category of speed | being flat | 0.102 | 0.702 | 0.962 | 0.016 | 0.865 | 0.633 |
| 2 | Being positioned after the right-side exit | being flat | 0.105 | 0.989 | 0.999 | 0.019 | 1.219 | 17.653 |
| 3 | The first category of length, direction signboard | being flat | 0.107 | 0.771 | 0.972 | 0.006 | 0.95 | 0.823 |
| 4 | The second category of length , The third category of speed | being flat | 0.107 | 0.973 | 0.997 | 0.018 | 1.198 | 6.879 |
| 5 | The first category of speed | straight line | 0.11 | 0.747 | 0.967 | 0.004 | 1.039 | 1.112 |
| 6 | being flat , light cars , The first category of length | straight line | 0.112 | 0.709 | 0.96 | 0.002 | 0.986 | 0.964 |
| 7 | The second category of length , The second category of speed , straight line | being flat | 0.113 | 0.74 | 0.966 | 0.011 | 0.912 | 0.725 |
| 8 | Consecutive entrance and exit | being flat | 0.113 | 0.931 | 0.993 | 0.015 | 1.148 | 2.75 |
| 9 | To be concave | straight line | 0.117 | 0.785 | 0.972 | 0.01 | 1.092 | 1.308 |
| 10 | Time to peak , The fourth category of speed | being flat | 0.12 | 0.972 | 0.997 | 0.02 | 1.198 | 6.848 |
| 11 | The second category of length , horizontal curve | being flat | 0.12 | 0.871 | 0.984 | 0.008 | 1.074 | 1.464 |
| 12 | exitance on the right side of the road | being flat | 0.122 | 0.92 | 0.991 | 0.014 | 1.134 | 2.369 |
| 13 | The first category of speed | being flat | 0.136 | 0.918 | 0.989 | 0.016 | 1.132 | 2.309 |
| 14 | exitance from the right side in the previous section | straight line | 0.136 | 0.87 | 0.982 | 0.023 | 1.209 | 2.153 |
| 15 | straight line , The fourth category of speed | being flat | 0.137 | 0.981 | 0.998 | 0.024 | 1.209 | 9.989 |
| 16 | Speed control camera | straight line | 0.137 | 0.717 | 0.954 | 0.001 | 0.996 | 0.991 |
| 17 | straight line , width reduction | being flat | 0.138 | 0.939 | 0.992 | 0.019 | 1.157 | 3.07 |
| 18 | straight line , direction signboard | being flat | 0.169 | 0.776 | 0.96 | 0.008 | 0.957 | 0.843 |
| 19 | Horizontal curve | being flat | 0.225 | 0.817 | 0.96 | 0.001 | 1.007 | 1.03 |

**Table 15. Sorted selected rules**

| NO. | Premises | Conclusion | Confiden | Laplace | Premises | Conclusion | Lift | Laplace |
|---|---|---|---|---|---|---|---|---|
| 1 | exitance from the right side in the next section | being flat | 0.989 | 0.999 | exitance from the right side in the next section | being flat | 70.000 | 0.999 |
| 2 | straight line,The forth category of speed | being flat | 0.981 | 0.998 | exitance from the right side in the previous section | straight line | 1.209 | 0.982 |
| 3 | The second category of length , The third category of speed | being flat | 0.973 | 0.997 | straight line ,The forth category of speed | being flat | 1.209 | 0.998 |
| 4 | Time to peak , The forth category of speed | being flat | 0.973 | 0.997 | The second category of length , The third category | being flat | 1.199 | 0.997 |
| 5 | straight line , width reduction | being flat | 0.939 | 0.992 | Time to peak , The forth category of speed | being flat | 1.198 | 0.997 |
| 6 | Consecutive entrance and exit | being flat | 0.932 | 0.993 | straight line , width reduction | being flat | 1.157 | 0.992 |
| 7 | exitance on the right side of the road | being flat | 0.921 | 0.991 | Consecutive entrance and exit | being flat | 1.148 | 0.993 |
| 8 | The first category of speed | being flat | 0.918 | 0.990 | exitance on the right side of the road | being flat | 1.134 | 0.991 |
| 9 | The second category of length , Horizontal curve | being flat | 0.871 | 0.984 | The third category of speed | being flat | 1.132 | 0.990 |
| 10 | exitance from the right side in the previous section | straight line | 0.870 | 0.982 | To be concave | straight line | 1.092 | 0.972 |
| 11 | Horizontal curve | being flat | 0.817 | 0.960 | The second category of length , Horizontal curve | being flat | 1.074 | 0.984 |
| 12 | To be concave | straight line | 0.785 | 0.972 | The third category of speed | straight line | 1.039 | 0.968 |
| 13 | straight line, direction signboard | being flat | 0.776 | 0.960 | Horizontal curve | being flat | 1.007 | 0.960 |
| 14 | The first category of length , direction signboard | being flat | 0.771 | 0.972 | Speed control camera | straight line | 0.996 | 0.954 |
| 15 | The first category of speed | straight line | 0.747 | 0.968 | being flat , light cars ,The first category of length | straight line | 0.986 | 0.960 |
| 16 | The second category of length , The second category of speed , straight line | being flat | 0.740 | 0.966 | straight line , direction signboard | being flat | 0.957 | 0.960 |
| 17 | Speed control camera | straight line | 0.717 | 0.954 | The first category of length , direction signboard | being flat | 0.950 | 0.972 |
| 18 | being flat , light cars , The first category of length | straight line | 0.709 | 0.960 | The second category of length , The second category of speed straight line | being flat | 0.912 | 0.966 |
| 19 | straight line, Time to peak , The third category of speed | being flat | 0.702 | 0.962 | straight line , Time to peak , The third category of speed | being flat | 0.865 | 0.962 |

## 8. Result Discussion

The findings of other researchers are based on placing the crash as the basis of research. The output of these studies relies on the conditions of the data recorded at the condition of the crash. Such as the circumstances of the drivers who are at fault in the crash or the circumstances of the time when the crash occurred or the types of vehicles involved in the crash. The result of such studies leads to the creation of crash risk maps or the classification of factors involved in crashes. In this research, the way of looking at the causal conditions is the time and scope of the crash occurrences. This type of attitude causes the creation of design rules. If the environmental conditions prevailing in the crash are recorded from the point of view of the road safety expert and this type of attitude is developed in a larger scope, the possibility of extracting stronger and more diverse rules will be provided. In this research, crashes played two roles. The first role is to verify the length of the sections used in the research and the second role is to reoccurrence the selection of sections in crashes that have occurred in order to create frequent rules in the occurrence of crashes. The main role was played by the shape of the created sections and the features of the sections. This type of view is a new way of applying data mining technique in sync with the view of traffic engineering. What is important in this method is the modeling and logical method used in making the model. In this research, certain operating speeds were created in the sections that lead to the repetition of crashes and the same time that the crashes are repeated in the specified sections were extracted. The simultaneity of features that increase the frequency of crashes has been extracted. 132 rules were created in the beginning. The rules were first sorted based on the parameters mentioned in the research. According to the expert's opinion and the removal of obvious rules from among the created laws, 19 rules were extracted as the final results of this research.

Entrance and exitance from the right and left side of the road, speed control camera, reduction of road width, and the presence of a advanced direction signs on the road, horizontal alignment of the simple curve, and straight line in different modes, vertical alignment of the road at specified operational speeds and specified times. In this research, peak times were mostly identified as factors causing inconsistencies in the occurrence of crashes. In the following, the interpretation of the output results of the model is stated

• The sections that are in the form of a straight line and have a flat vertical alignment, at peak times when reaching operational speed of 63 to 73 km/h, the most inconsistency occurs in terms of the frequency of crash occurrences. The dominant mode that occurred in these sections is the reduction of the width and the presence of the exitance from the right side in these sections. The dominant mode in the third priority in these sections is the presence of the entrance from the right side in these sections. The average width of these sections is 13.4 meters and the average length of these sections is 350 meters. Simultaneously, the reduction of the width and the presence of the exitance from the right side have caused inconsistency in these sections. The speed difference of about 15 km/h compared to the permitted speed announced on the road can be improved by synchronizing the operating speed with the operational speed of the front and rear sections. The next factor that can be created to increase consistency in these sections is cross section management.

• In flat sections, being placed after the exitance from the right side has the greatest effect on the recurrence of crashes. The dominant mode that occurs in these sections is the simultaneity of presence the entrance from the right side of the road and of the advanced road direction sign. The average width in these sections is 14 meters and the average length in these sections is 296.4 meters. The average operating speed in these sections is 70.1 km/h.

The location of the lane changing at a distance of less than 400 meters from the location of entrance from the right side is the cause of inconsistency in the sections located in the vicinity of the island connecting the exitance from the right side and the entrance from the right side following it. The overlapping of the place of the advanced road direction sign and the entrance from the right side of the road is placed before it is another factor of inconsistency in these sections.

• Placing the advanced direction signs on the flat sections with lengths shorter than 323 meters has caused the inconsistency of frequency of crash occurrences. The dominant mode that occurred in these sections is the reduction of the width and the presence of the exitance from the right side in these sections. The average width of these sections is 14 meters. The average operating speed in these sections is 67.1 km/h. The most important factor in the occurrence of inconsistency and causing a speed drop of about 15 km/h compared to the permitted speed announced on the road is failure to provide the suggested overlap length of 400 meters for the sign and exitance from the right side of the road .

• Driving at an operating speed of about 63 to 73 km/h in sections with vertical flat alignment with lengths between 325 and 647 meters will cause the frequency of crash occurrences. The average width in these sections is 13.2 meters. Both of the mentioned factors have caused the operating speed to drop by about 15 km/h compared to the permitted speed announced on the road (80km/h). Providing sufficient visibility to the advanced direction signs and synchronizing the performance speed in consecutive sections will improve Consistency is in these sections.

• Driving at a speed of 22 to 50 km/h causes frequency of crash occurrences. Increasing the width and exitance from the right side are the dominant modes that occur in these sections. The average width in these sections is 13.5 meters. The simultaneity of the increase in width with the decrease in operating speed of more than 30 km per hour on these sections is contrary to the driver's expectation. The most important factor in increasing the consistency of these sections is synchronizing the operating speed between successive sections and managing the cross section of the road in these sections.

• Driving at an operating speed of 50 to 63 km/h on flat sections with lengths between 325 and 647 meters causes the occurrence of inconsistencies in the frequency of crash occurrences. The predominant mode that occurred in these sections was the entrance from the right side. The average width in these sections is 12.6 and the average speed in these sections is 59 km/h. It seems that the traffic entering from the right side of the road to these sections has caused a difference in the operational speed of 30 km per hour compared to the speed limit announced on the road (80km/h). Traffic flow management on the network of roads leading to the road and synchronizing the speed of the sections before and after these sections will improve the consistency of these sections.

• Frequent entrance and exitance (more than one direct connection in sections) causes inconsistency and frequency of crash occurrences. The dominant modes in these sections are the reduction of the width of the road and presence of advanced direction sign and the entrance from the right side of the road. The average width of these sections is 12.8 meters, the average length of these sections is 692.7 meters, and the average operating speed of these sections is 68 km/h. The most important factor in improving the consistency of these sections is the shift of direct access to the road to the auxiliary lane access of road and providing the proposed length of 400 meters for overlapping length of the advanced direction sign and the entrance from the right side of road.

• The sections that are concave in vertical alignment and straight line in horizontal alignment will cause inconsistencies in frequency of crash occurrences. The dominant mode in these sections is the reduction of the width and the presence of advanced direction sign on them. The average length of these sections is 510 meters and the average width of these sections is 14 meters. The average operating speed of these sections is 68 km/h. Failure to provide sufficient sight distance and the overlap of the effect of the advanced road direction sign on the vertical curve, which is caused by the short distance between the vehicle lane change point and the end of the vertical curve, has caused inconsistency. Providing the distance of 400 meters after the end of the vertical curve to the place of lane change improves the consistency of these sections.

• Driving at a speed of 73 to 88 km/h during peak times of traffic flow on flat sections of the road causes inconsistency and frequency of crash occurrences. The dominant features in these sections are the presence of advanced direction sign and the presence of an exitance from the right side of the road in these sections. The average length of these sections is 348 meters and the average width of these sections is 13 meters. Considering the effective length of the advanced direction sign and the exitance from the right side of the road, it is recommended to provide a length of at least 400 meters for similar sections.

• The existence of a flat horizontal curve with lengths between 325 and 647 meters causes the inconsistencies in frequency of crash occurrences. The dominant modes in these sections are the reduction of the width and the advanced direction sign and the speed control camera. The average width of these sections is 13 meters and the average speed of these sections is 69 km/h. All three factors simultaneously cause inconsistency in the frequency of crash occurrences and speed differences from announced allowable speed of road (80km/h). Synchronization of the speed of the sections before and after the curve and speed control in these sections are among the low-cost measures to improve the consistency of these sections.

• The presence of the exitance from the right side in flat sections causes the occurrence of inconsistencies and the frequency of crash occurrences. The dominant modes in these sections are the presence of the advanced direction sign and the entrance from the right side in these sections. The average operating speed of these sections is 70 km/h, the average length of these sections is 296 meters, and the average width of these sections is 14 meters. The inconsistency often occurs due to the lack of sufficient weaving length in the operating speed of 10 km/h below the permitted speed announced on the road.by Considering the length of 150 meters before the advanced direction sign and 250 meters before the exitance from the right side of the road, an average required length of 400 meters for these sections is suggested.

• The straight sections along the road that has a speed control camera have the inconsistency of frequency of crash occurrences. The dominant modes in these sections are the presence of advanced direction signs and entrance from the right side for the road. The average width of these sections is 14 meters and the average length of these sections is 414 meters. The average operating speed of these sections is 68 km/h. The effect of the speed control camera is often to decrease the speed on this section and increase the speed on the next section, and on the other hand, the presence of the advanced direction sign indicates the lane change in the rest of the road, which is aggravated by the presence of the entrance from the right side of the road on these sections. Providing the weaving length after the speed control camera is proportional to the speed increase effect caused by the presence of the speed control camera on the section that consequent the speed camera control, which causes the consistency of these sections. According to the fact that the average length of the sections after the speed control camera is 415 meters, the proposal of this research is to provide the minimum length of 450 meters after the speed control camera to the place of lane change in order to improve the consistency of the lane.

• Driving in a simple flat curve has the inconsistency of frequency of crash occurrences. The dominant modes in these sections are the presence of advanced direction sign and the exitance from the right side of the road. The third priority in these sections is the presence of a speed control camera. The average width of these sections is 13.7 meters and their average length is 276 meters. The average operating speed of these sections is 81 km/h. A length shorter than 400 meters in order to ensure the effect of overlapping the advanced direction sign and the exitance from the right side of the road and to increase the operational speed for execution or to avoid weaving movement in the short length of 276 meters in the curve, are the factors of creating inconsistency in these curves. This issue is obvious enough, the average operating speed of the sections has increased from the declared permitted speed (80km/h) even with the presence of the speed control camera on these sections. The most important measure regarding the improvement of the consistency of these sections is to move the exitance from the right side to the auxiliary lane and also to move the advanced direction sign to a place where the weaving length of 400 meters between the advanced direction sign and the exitance from the right side is provided. The next issue is the speed management in the previous sections, so that if it is not possible to provide the weaving length on the curve, the operational speed will be reduced to the extent that it is possible to perform the weaving movement in the mandatory length of 276 meters and the available width.

## 9. Conclusion

- Two factors, the operational speed and the length of the sections are effective regarding the features placed on the road. The inconsistency of crash frequency occurred in certain categories of the operational speed. The failure to provide the length of the weaving in accordance with the operational speed of the road sections due to the features on the road has been identified as the cause of inconsistency in the occurrence of repeated crashes.

- Features include of the speed control camera, the advanced direction signs, entrance and exitance from the right and left side of the road, reducing the width of the road and the characteristics of the horizontal and vertical geometry of the road have been effective in repeating the occurrence of crashes.
- The most important factors that play a role in maintaining the consistency of the road are: provision of sufficient visibility distance after the vertical curve, provision of sufficient weaving length according to the operational speed of the sections, cross section management, operational speed control based on the features on the road and also Coordination of operational speed in successive sections has been identified.
- Positioning of the lane change section at the recommended distance of at least 450 meters after the speed control camera in order to provide the length of the decision and coverage to the effective length of the advanced direction sign, will reduce the inconsistency of the occurrence of repeated crashes at a functional speed of 68 km/h.
- Positioning the lane change section at a distance of at least 400 meters after the end of the vertical curve in order to provide the weaving length of the advanced direction sign and the exitance from the road will reduce the inconsistency of repeating crashes at an operational speed of 68 km/h.
- Positioning of the entrance from the right-side feature and the advanced direction sign in sections with a length of less than 400 meters indicates the lack of provision of the weaving length for changing the lane and causes the inconsistency of repeating the crash at the operating speed of 73 to 89 km/h.
- Positioning the advanced direction sign and the exitance from the right side of the road in sections with a length of less than 400 meters causes the crash occurrence frequency at a speed of 70 km/h.
- Positioning of the entrance from the right side and the exit from the right side in the straight-line sections with a length of less than 400 meters causes the inconsistency of repeating the crash at a operational speed of 70 km/h. This inconsistency is aggravated by placing the speed control camera on these sections.
- Positioning of the exitance from the right side of the road and the advanced direction sign on curve-shaped sections with a length of less than 400 meters (providing the overlapping length of the advanced direction sign with the exitance from the right side of the road) also increases the operational speed of the curve-shaped sections and the inconsistency of repeating the occurrence crashes occurred at an operational speed of 81 km/h.
- Placing the speed control camera before the curved-shaped sections is another reason for increasing the performance speed in these sections. Also, the occurrence of width reduction on the curve-shaped sections is contrary to the driver's expectation and has caused the inconsistency to intensify.
- More than one direct access from the right side to the road with an average distance of less than 350 meters from each other and also the reduction of the width on these sections has caused the inconsistency of repeating the occurrence of crashes at a functional speed of 70 km/h.

This is the knowledge that analysis of crash data creates. This is the beginning of acquiring knowledge in this way. The use of this point of view in other roads, including arterial roads, as well as the use of this method in two-lane, two-way and rural roads can improve the regulations of the road design regulations and have a significant impact on reducing crashes.

## 10. References

- Abellán, J., López, G., & OñA, J. de (2013). Analysis of traffic accident severity using decision rules via decision trees. Expert Systems with Applications, 40(15), 6047–6054.

- Abugessaisa, I. (2008). Knowledge discovery in road accidents database-integration of visual and automatic data mining methods. International Journal of Public Information Systems, 4(1).

- Chand, A., Jayesh, S., & Bhasi, AB (2021). Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. Materials Today: Proceedings, 47, 5135–5141.

- Chang, L.-Y., & Wang, H.-W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis & Prevention, 38(5), 1019–1027.

- Chen, W.-H., & Jovanis, P. P. (2000). Method for identifying factors contributing to driver-injury severity in traffic crashes. Transportation research record, 1717(1), 1–9.

- Chong, M. M., Abraham, A., & Paprzycki, M. (2004). Traffic accident analysis using decision trees and neural networks. arXiv preprint cs/0405050.

- Comi, A., Polimeni, A., & Balsamo, C. (2022). Road accident analysis with data mining approach: evidence from Rome. Transportation research procedia, 62, 798–805.

- Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2019). Supervised association rules mining on pedestrian crashes in urban areas: identifying patterns for appropriate countermeasures. International Journal of Urban Sciences, 23(1), 30–48.

- Demirel, N., Emil, M. K., & Duzgun, H. S. (2011). Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery. International journal of Coal geology, 86(1), 3–11.

- Deng, X., Zeng, D., & Shen, H. (2018). Causation analysis model: Based on AHP and hybrid Apriori-Genetic algorithm. Journal of Intelligent & Fuzzy Systems, 35(1), 767–778.

- Depaire, B., Wets, G., & Vanhoof, K. (2008). Traffic accident segmentation by means of latent class clustering. Accident Analysis & Prevention, 40(4), 1257–1266.

- Farhangi, F., Sadeghi-Niaraki, A., Nahvi, A., & Razavi-Termeh, S. V. (2022). Spatial modelling of accidents risk caused by driver drowsiness with data mining algorithms. Geocarto International, 37(9), 2698–2716.

- Feng, M., Zheng, J., & Ren, J. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. IEEE Access, 7, 106111–106123.

- Gariazzo, C., Stafoggia, M., Bruzzone, S., Pelliccioni, A., & Forastiere, F. (2018). Association between mobile phone traffic volume and road crash fatalities: a population-based case-crossover study. Accident Analysis & Prevention, 115, 25–33.

- Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2003). Profiling of high-frequency accident locations by use of association rules. Transportation research record, 1840(1), 123–130.

- Kuhnert, P. M., Do, K.-A., & McClure, R. (2000). Combining non-parametric models with logistic regression: an application to motor vehicle injury data. Computational Statistics & Data Analysis, 34(3), 371–386.

- Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyze road accident data. Journal of Big Data, 2(1), 1–18.

- Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. Journal of Modern Transportation, 24(1), 62–72.

- Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. Accident Analysis & Prevention, 75, 1–15.

- M Feng, J Zheng, J Ren, Y Xi (Ed.). 2019. Association Rule Mining for Road Traffic Accident Analysis: A Case Study from UK: Springer.

- Meißner, K., & Rieck, J. (2022). Strategic planning support for road safety measures based on accident data mining. IATSS research.

- Momeni Kho, S., Pahlavani, P., & Bigdeli, B. (2022). Analyzing and Predicting Fatal Road Traffic Crash Severity Using Tree-Based Classification Algorithms. International Journal of Transportation Engineering, 9(3), 635–652.

- Montella, A. (2011). Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accident Analysis & Prevention, 43(4), 1451–1463.

- Nafis, S. R., Alluri, P., Wu, W., & Kibria, B. G. (2021). Wrong-way driving crash injury analysis on arterial road networks using non-parametric data mining techniques. Journal of Transportation Safety & Security, 1–29.

- 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.

- Pang-Ning, T., Steinbach, M., & Kumar, V. (2006). Introduction to data mining: Pearson Addison Wesley Boston.

- S Das, X. S. (Ed.). 2014. Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. : 14-1540: Transportation Research Board Washington DC.

- S Priya, R. A. (Ed.). 2018. Association rule mining approach to analyze road accident data: IEEE.

- Samerei, S. A., Aghabayk, K., Mohammadi, A., & Shiwakoti, N. (2021). Data mining approach to model bus crash severity in Australia. Journal of safety research, 76, 73–82.

- Sanmiquel, L., Rossell, J. M., & Vintró, C. (2015). Study of Spanish mining accidents using data mining techniques. Safety science, 75, 49–55.

- Shirmohammadi, H., Hadadi, F., & Saeedian, M. (2019). Clustering analysis of drivers based on behavioral characteristics regarding road safety. International Journal of Civil Engineering, 17(8), 1327–1340.

- Sohn, S. Y., & Shin, H. (2001). Pattern recognition for road traffic accident severity in Korea. Ergonomics, 44(1), 107–117.

- Tavakoli Kashani, A., Shariat-Mohaymany, A., & Ranjbari, A. (2011). A data mining approach to identify key factors of traffic injury severity. PROMET-Traffic&Transportation, 23(1), 11–17.

- Tesema, T. B., Abraham, A., & Grosan, C.

(2005). Rule mining and classification of road traffic accidents using adaptive regression trees. International Journal of Simulation, 6(10), 80–94.

- Wedyan, S. (2014). Review and comparison of associative classification data mining approaches. International Journal of Computer, Information, Systems and Control Engineering, 8(1), 34–45.

- Weng, J., Zhu, J.-Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. Accident Analysis & Prevention, 92, 43–52.

- WHO (2018). https://www.who.int/.

- Wong, J.-T., & Chung, Y.-S. (2008). Comparison of methodology approach to identify causal factors of accident severity. Transportation research record, 2083(1), 190–198.

- Xiong, H. (2006). Association Analysis: Basic Concepts and Algorithms. URL http://www. columbia. edu/~ jwp2128/Teaching W, 4721, 185–203.

- Xu, C., Bao, J., Wang, C., & Liu, P. (2018). Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. Journal of safety research, 67, 65–75.

- Z Gao, R Pan, R Yu, X Wang (Ed.). 2018. Research on automated modeling algorithm using association rules for traffic accidents: IEEE.

- Zelalem, R. (2009). Determining the degree of driver's responsibility for car accident: the case of Addis Ababa traffic office. Addis Ababa, Addis Ababa University.