Turkish Journal of Civil Engineering, 2023 55-79, Paper 746, ***Research Article***

Identifying Interrelated Factors of Fatal and Injury Traffic Accidents Using Association Rules

Zeliha Cagla KUYUMCU[1]
Hakan ASLAN[2]
Nilufer YURTAY[3]

ABSTRACT

This study aims to investigate the possible relationships of risk factors related to traffic accidents playing important roles in increasing the likelihood of accidents. In the previous studies, parametric models are mostly used to investigate the causes of traffic accidents. As a non-parametric data mining model with its increasing usage in recent years; association rule mining was employed in this study to analyse the traffic accident data for the period of 2015 and 2020 in the city of Sakarya, Turkey. The analysis of the data studied revealed the relationships among the external/environmental, driver, road, vehicle and nature of accident factors. Some important rules regarding accidents occurring on daylight came into prominence within the scope of this study. In addition, the correlations between the driver casualties and their education level and ages are established. The findings are beneficial for transportation authorities to apply effective operational strategies and campaigns to increase the road safety.

Keywords: Traffic accidents, road crashes, traffic safety, data mining, association rules, driver faults.

1. INTRODUCTION

According to the data given by World Health Organisation (WHO) while around 1.35 million people die each year due to traffic accidents world-wide, more than around 35 million people are also involved in traffic accidents not fatal but leave many disabled. The data also state that although car ownership and trip rates are relatively higher in industrialized countries, more than 90% of road-traffic related deaths occur in low and middle-income countries.

Note:

- This paper was received on November 1, 2022 and accepted for publication by the Editorial Board on June 23, 2023.
- Discussions on this paper will be accepted by November 30, 2023.

• https://doi.org/10.18400/tjce.1322965

1 Sakarya University, Civil Engineering Department, Sakarya, Türkiye, caglacaglar@sakarya.edu.tr - https://orcid.org/0000-0002-9851-6843

2 Sakarya University, Civil Engineering Department, Sakarya, Türkiye haslan@sakarya.edu.tr - https://orcid.org/0000-0001-9444-6908

3 Sakarya University, Computer Engineering Department, Sakarya, Türkiye nyurtay@sakarya.edu.tr - https://orcid.org/0000-0002-7577-7506

Similarly, in high-income industrialized countries, people from low socioeconomic backgrounds are more likely to be involved in road accidents. These accidents are the leading causes of deaths for children and young people aged between 5 and 29. Males world-wide are more likely to be involved in traffic accidents than the same age group of young females. The figures obtained across the world and presented by WHO state that about three-quarters (73%) of all road traffic fatalities are caused by young males under 25 years old who are almost 3 times more likely to die in a road traffic accident than young females [1].

To reflect the effect of road injuries for the life quality of the people, it was stated by a study conducted in 2019 involving 204 countries from all over the world that they are the seventh causes of disability-adjusted life [2].

According to the report of injuries related to the European Union countries for the period of 2009-2018, as shown in Figure 1, the average traffic accident rate (TAR) is 5.74 per 1000 people. On the other hand, this average figure doubles as almost 10 per 1000 people in Turkey, making it as the second highest country in Europe [3].

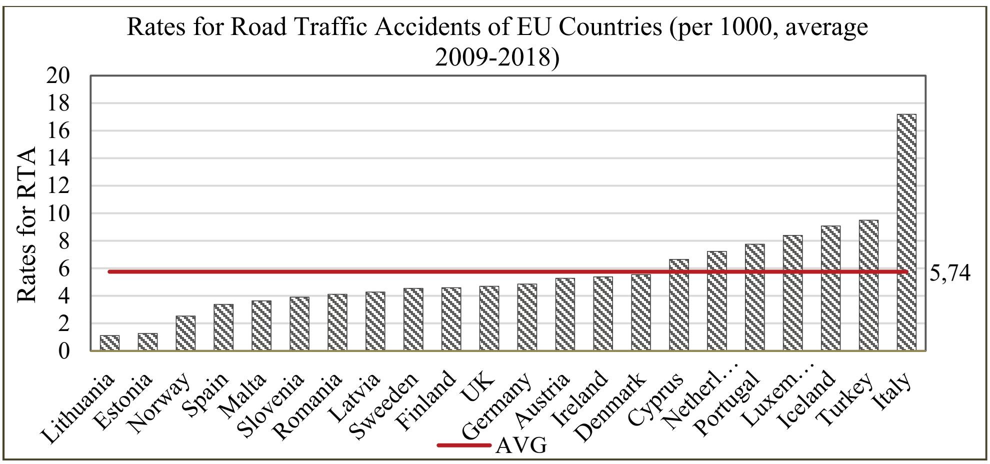

Figure 1 - Average road traffic accident rates of non-fatal traffic injuries for EU countries for the period of 2009-2018 [3]

In 2018, while the number of deaths in traffic per 100 thousand people was 4.9 in EU countries, Turkey had the figure of 8.1 [4].

The total number of fatal and injury related accidents occurred in Turkey, an upper-middle-income country [5], with regard to 2015, 2016, 2017, 2018, 2019 and 2020 are given as 183,011, 185,128, 182,669, 186,532, 174,896, 150,275, respectively. It is important to note that, the quarantine enforced in many countries in effect of the Covid-19 pandemic caused the number of traffic accidents dropped drastically [6]. In addition, driver fault rates for these stated years are given as 89.76, 90.02, 90.29, 89.64, 88.89 and 88.65 percent, respectively [7]. These high rates of the accidents involving driver faults led the accidents without driver faults to be excluded for the analysis in this study. Furthermore, as non-faulty drivers simply do not have any contribution to the occurrence of the accidents, this study mainly analysed the correlations between the characteristics of the faulty drivers involved in the accidents and the nature of the accidents without considering the features of the drivers having no responsibility or fault. However, if the accident reports give responsibility to all the drivers in the accident to some extend then each of these drivers are taken into consideration through the analysis.

A wide range of parametric methods such as regression and logit models have been used by numerous researchers in order to control and reduce the number of traffic accidents [8, 9, 10, 11].

The dataset used in this study provides ample contents with regard to numerous attributes having effect on the occurrence of the traffic accidents. In addition, it is needed to note that a meticulous data pre-processing was carried out by excluding noisy data, by both combining the attributes to get meaningful rules when needed, and by converting numerical data into categoric one, etc. to prevent the biased results to be attained.

The dataset was also divided into groups to obtain different inter-correlations. Furthermore, after pre-processing, the analysis was carried out by changing support and confidence values repeatedly through the evaluations of the rules produced by the software, R, so that the best and comprehensive association rules are assured to be acquired.

2. LITERATURE REVIEW

In the literature there are numerous studies with regard to traffic accidents. In these studies, a lot of factors were investigated to analyse their potential effects to traffic accidents. The literature review is divided into two sections: 1) studies on traffic violations and prediction of accident severity and 2) studies on identifying co-occurring accident factors.

2.1. Studies on Traffic Violations and Prediction of Accident Severity

Alver et al. examined the relationship of four major traffic rule violations with the characteristics of the traffic accidents involving young drivers through the data collected from 2057 participants of their questionary survey conducted as face-to-face interview. They employed binary logit model to establish the correlation between the socio-demographic background of the young adults and their behaviours in traffic leading to an accident [12].

Katanalp et al. used a dataset of bicycle accidents recorded for the period of 2013-2017 for the assessment of the classification of injury-severity in accidents involving cyclists by employing DT-based hybrid-fuzzy models (DT-CFL and DT-RFL). C4.5 decision tree algorithm was used both to classify the injury-severity level and to obtain all splitting conditions to set up the DT-CFL fuzzy prediction model. The analysis illustrated the performance of the injury-severity classification in terms of accuracy, precision, recall, and F-measure values [13].

Kadilar examined the factors having effect on the severity of drivers in traffic accidents via ordinary logistic regression assessment model. The study modelled the Turkish driver fatalities and the meaningful estimations of the accident severity levels in Turkey for the first time. According to the findings of the research it was determined that age and education level of the driver, existence of alcohol-drug usage, availability of seat belt, condition and type of the roadway, location and type of the collision are among the important factors having major effect on the accident severity to be considered for the analysis. On the other hand, it is engrossingly claimed that gender, roadway surface, weather condition and vehicle age do not have statistically significant effect on the accident severity [8].

Zhang et al. identified main risk factors causing traffic violations and their relations with the severity of the accidents. In this study, stepwise logistic regression model was applied to reveal the most important factors affecting the probability of traffic violations. The results indicated that drivers with less than two-years of experience have less risk of getting involved with severe accidents although they tend to perform higher possibility of rule violations. Another important finding of the study is that morning rush hours involve the probability of more violations with unimportant effect on accident severity [9].

Ma et al. investigated different age groups of drivers involved with accidents resulted in pedestrian injury. Pedestrian age, vehicle type and weather conditions were pointed out as important variables for pedestrian injury types. Another finding is that number of vehicles and traffic type as independent variables affect young and middle-aged drivers` behaviours at intersections. Moreover, hit-and-run pedestrian involved accidents caused by young and older drivers are related to a significant variable of whether the roadway is divided or not. With the developed probit regression models, it was revealed that education level of the drivers along with control and intersection design measures are effective in the decision-making processes to increase the safety of pedestrians at intersections [10].

Adanu et al. explored frequent and improper lane-changing behaviours having a profound impact on both the operation and safety of traffic. The analysis of severity of injuries due to lane changing-related accidents involving young and old drivers resulted in two models for young and old drivers. According to the model estimation results, young male drivers are more likely to be injured in terms of lane changing-related accidents compare to older male drivers. This is mainly due to the fact that older drivers are more careful when changing the lanes for different purposes preventing them to get involved that kind of accidents [14].

Chiou et al. analysed parameters of both of the vehicles causing the accident rather than concentrating on only one of them by employing Generalized Estimating Equations (GEE) to examine level of severity. The accident data included variables like driver features (for both parties), type of vehicles, road-intersection-lighting conditions, and type of collision. According to this study, the most influential factors on accident severity are vehicle type (motorcycle), speeding, angle impact type of accident, and drunk driving [15].

AlKheder et al. investigated accident severities by employing Decision Trees, Bayesian Network and Linear Support Vector Machine (SVM) as data mining analysis methods. Among these, Decision Trees have many advantages over traditional statistical approaches. This method can work with different types of variables like binary, categorical, ordinal, and continuous ones. Another advantage of applying decision trees is the ability to indicate the importance of the selected predictors. Age (under 17, 18-30, 31-60, over 60), seat belt, seat/place of the injured person (driver, front/back seat, pedestrian, etc.), road class, speed limit, number of lanes, type of accident (pedestrian, T-bone, side-by-side, rear-end collision and other) were regarded as variables. The performances of these three used methods were compared through different parameters. While Bayesian network produced the most successful results for predictions of the severity of the accidents, SVM resulted in the most unsuccessful. It was stated that pedestrians are more vulnerable road users compared to drivers and passengers. Another finding of the research is that male drivers and front seat passengers are more exposed to severe and fatal consequences. Likewise, older drivers are shown to be more likely to get involved in severe injuries and fatalities [16].

Batouli et al. conducted research in Colorado, USA. In this study the factors affecting the hitting pedestrian accident severity were examined through the data for the period of time from 2006 to 2016. Probability ratios of occurrence of fatal and non-fatal pedestrian accidents were calculated through the established logistic regression model by considering the proximity of the accident to the intersection, daylight condition, vehicle type and vehicle speed as the variables. The outcomes of the paper state that since the older drivers` involvement within car technological devices and mobile phones is less than the younger drivers, the older drivers are less distracted improving their reaction time and vision ability leading them to get involved in less severe accidents. Another interesting finding of the research is that pedestrian severity is affected by the impairment of both pedestrian himself and driver [11].

2.2. Studies on Co-Occurring Accident Conditions

Cai examined the effect of type and location of accident, season, time of the day, weather condition, lighting, visibility distance, barrier positions, type of road pavement and intersection with the rules of association. The rules specify that fatal accidents occur when "driving with other behaviours that undermine driving safety". Furthermore, "collision with moving vehicles" results in more fatality when season is summer and time of the day hour is deep night. Another remarkable outcome of the study points out that the fatality rate of the accidents increases when weather is sunny, terrain is plain, time is evening without street lighting [17].

Yu et al. applied a priori algorithm to analyse features and factors affecting severity of traffic accidents in Wisconsin, United States. According to study results, male drivers aged between 16 and 29 are more tend to be involved in accidents on roadways with no median. The other important findings state that fatal accidents are more likely to happen at towns while property damage accidents in the cities [18].

Das et al. examined pedestrian deaths between 2014 and 2018 through association rules applied to 4 groups of the most common fatal accident scenarios. This study developed all twenty best rules for 4 subgroups using a priori algorithm and taking lift as a performance measure. Some key variable categories are considered as; lighting, straight moving vehicles, turning vehicles, local municipality streets, and pedestrian age range. Patterns of rules differed according to the position of the pedestrian being inside and outside the crosswalk. When the pedestrian is in a place other than a crosswalk, the absence of lighting at night is associated with many accidents [19].

Kong et al. with the naturalistic driving dataset, investigated the potential rules between fast driving behaviour and journey/driving/road characteristics by employing classification-based association rules algorithms. One of the findings of the study is that the fatigue and distraction that occurs on long journeys with the combination of fast driving on high class roads can cause severe accidents [20].

Zhu derived association rules to find accident-related features that increase the probability of vehicle-bike hit-and-run accidents. One of the rules expresses that the likelihood of vehicle-bicycle hit-and-run accidents reduces when the collision type is rear-end occurring at morning peak hours [21].

Hong et al. produced worthful outcomes by explaining the fundamental rules in the field of freight vehicle accident data. According to the study, a priori algorithm as an association rule mining method is an acceptable method for analysing features of freight vehicle accidents. They also spotted that the accidents involving young freight vehicle drivers in their 20s and 30s are linked with specific regions having well-designed roadway geometry, excluding vertical and horizontal curves. Furthermore, the study indicates the connection between young freight vehicle drivers' traffic violations and good conditions of roadways [22].

Das et al. developed an effective way of discovering important association rules explaining accidents when the weather is rainy. One of the remarkable findings is that drivers from the age group in between 15 to 44 get involved in the accidents during rainy weather when the illumination is poor, and roadway has curves [23].

According to the study conducted by Xu et al., factors associated with traffic accidents involving serious casualty include road user behaviour, condition of the vehicles, road geometric design, and environmental features. The factors associated with poor vehicle operation conditions such as flat tires, braking-related problems, and faulty steering systems may cause severe accidents on even high standard roads having relatively decent safety facilities [24].

3. METHODOLOGY AND DATA

In this study, the association rule mining was chosen as the data mining method to determine the co-occurrence pattern of traffic accidents with regard to related attributes stated in the accident reports. The association rule analysis, which has recently been increasingly applied in road traffic safety research, is designed to investigate a set of elements that typically occur together in a given incident. Association rule mining is a suitable method to investigate road traffic accidents when the large amount of data makes it difficult to extract a latent hypothesis for the research analysis. This makes association rules to be extremely useful tool for a big set of unsupervised data [19].

The raw data in this research covers 6 years of traffic accident data having over 23,551 rows representing accidents and 26 columns of attributes designating the possible factors contributing to the occurrence of the accidents. As the association rules in this sense provides profound and effective analysis of the big data, it was employed for the analysis of the data available within the concept of this study.

3.1. Study Area

Covering 4817 km^2 area and residing 1.06 million people, Sakarya is located in the northwest part of Turkey [25].

While the car ownership rate per 1000 people was 162 in Turkey in 2021, it was 151 in Sakarya, making the city to characterize Turkey's average with that respect [25].

The highway network of Sakarya consists of highways, state roads, provincial roads and urban roads. The motorways passing through Sakarya are O-4 (TEM) and O-7 (NMM), the state roads, on the other hand, passing through Sakarya are D-100, D-650 and D-020 [26].

Another important fact to be mentioned is that Sakarya is a transit point of four major cities of Turkey, Kocaeli (37 km), İstanbul (148 km), Bursa (158 km), Bolu (114 km) and Ankara (306 km). Sakarya is 110 km from Istanbul Kurtköy Sabiha Gökçen Airport and 176 km from Istanbul Airport [26]. This important location of the city led to one of the 20 accident black spots planned to be investigated across Turkey in 2021 to be in city of Sakarya [27].

The total number of the registered motor vehicles in the city is 312,552 out of them 160,200 being automobile. The total number of accidents, number of people killed, and number of people injured in 2021 are given as 15,758 (43 per day), 2727 (59 per day; 33 at the accident scene, 26 accident follow-up) and 3972 (11 per day), respectively [25].

These facts establish the background of why city of Sakarya was chosen for the analysis. The study area covers all the transportation network of the city and available roads on it, streets, avenues and state roads are all investigated.

3.2. Data Description and Processing

The raw accident data for this study was provided from the Republic of Turkey General Directorate of Security. All traffic accident reports recorded after 2015 in Turkey, hence in Sakarya, include the 30-day hospital follow up period to see whether any further fatal results are the case for the accidents involving serious injury at the accidents. Thus, the dataset used in this study covers fatal or injury related accidents in city of Sakarya for 6 years of time period from 2015 to 2020 by covering this 30 day follow up period.

Since the number of fatal accidents involving drivers, pedestrians or passengers constitutes only 126 records out of 14,226 being considerably lower than the total number of injury and non-injury accidents, all fatal accidents are included in the dataset of injury accident classes.

There are two raw datasets; one with 14,226 records which signify the characteristics of all real accidents, and the other with 23,551 records involving driver’s characteristics. As many accidents involved more than one car, hence drivers, the second raw data had more rows than the initial one. These two datasets with 14,226 and 23,551 rows consist of 43 and 26 columns, respectively, as presented in Table 1.

Finally, these two raw datasets were combined to acquire the dataset to be used for the analysis purposes. The combination was carried out through the common accident ID numbers. In this way an actual analysis data involving 14,387 records with 11 columns symbolising all the attribute related features were obtained.

In this dataset used in the analysis, each row covers the information related to only the accidents and involved one unique driver faulty. For example, if two vehicles are involved in an accident and both drivers are at fault, they are represented by 2 rows in the dataset used in the analysis. If one of the drivers is responsible for the occurrence of the accident, however, those drivers are represented in the dataset as a single row. It should be mentioned that each row arranged in this way includes the related "driver severity", "driver age" and "vehicle damage level" information.

Table 1 - Summary of frequencies and distributions of some important attributes of raw data

Attribute	F	%	Attribute	F	%
Driver Gender			**Illumination**		
Male	20,950	89	Good street-lighting	8815	62
Female	1627	7	No street-lighting	5411	38
Unknown	974	4	**Traffic Signal**		
Driver Age Range			Yes	1636	12
Young	6264	27	No	12,371	87
Middle-aged-1	8492	36	Broken	219	2
Middle-aged-2	7591	32	**Weather**		
Elderly	1204	5	Sunny	11,975	84
Driver Lapses			Rainy	1802	13
Following too close	1298	5	Cloudy/foggy	300	2
Failure to yield	2556	11	Snowy	152	1
Improper turn	1675	7	**Vehicle Type**		
Speeding	5172	22	Car	12,986	55
Improper lane changing	1670	7	Heavy vehicle	1985	8
Drunk driving	183	1	Van	4031	17
Other	1842	8	Others	4549	19
No lapse	9155	39	**Vehicle Damage**		
Driver Severity			Major damaged	3528	15
Injured	9816	42	Minor damaged	9857	42
Non-injured	12,671	54	None damaged	2857	12
Fatal	68	0	Functional damaged	1720	7
Unknown	996	4	Can't move	5067	22
			Not identified	552	2

Table 1 - Summary of frequencies and distributions of some important attributes of raw data (continued)

Attribute	F	%	Attribute	F	%
Driver Education			**Municipality Type**		
Primary school	7862	34	Urban	11,886	84
Middle school	2371	10	Rural	2340	16
High school	6546	28	**Pavement Condition**		
Higher education	2623	11	Dry	11,561	81
Unknown	4149	17	Wet	2479	17
Seatbelt/Helmet Usage			Snowy	96	1
Yes (Seatbelt)	847	4	Ice	55	0
No (Seatbelt)	157	1	Puddle	20	0
Yes (Helmet)	142	1	Other slippery cond.	15	0
No (Helmet)	124	1	**Alcohol Usage**		
Unknown	22,281	95	Yes	1058	4
Driver License			No	18,060	77
Yes	19,753	84	Unknown	4433	19
No	2723	12	**Lighting Condition**		
Unknown	1075	5	Daylight	9503	67
Collision Type			Night	4331	30
Rear	1717	12	Twilight	392	3
Side collision	5115	36	**Season of the Year**		
Rsr	2325	14	Winter (dec-feb)	2760	20
Head on Collision	835	6	Spring (mar-may)	3188	23
Hitting Pedestrian	2756	19	Summer (june-aug)	4181	31
Boo	1478	10	Autumn (sep-nov)	3578	26

Rsr: Rollover/ skidding/run off road, Boo: Bumping into fixed objects and others

As some variables in raw data are in unbalanced nature expressing the dominant structure of the classes available in these attributes, the related association rules lead to only the same type of classes. This is why these attributes are not included in the final dataset for the analysis. The final data contain 2854, 2919, 2660, 2430, 1949 and 1575 samples for each year for the analysis period, respectively.

The dataset includes information regarding the factors related to external and environmental conditions, road conditions, vehicle and driver characteristics and nature of accident features. The data examined within the scope of this study are related only with the accidents involving driver faults. To find effective and meaningful association rules; attributes such as accident

ID, vehicle ID, hours and coordinates of the accidents, number of the road section were excluded since these attributes have individual values and do not create a group of datasets resulting in extremely small percentage values in the general dataset.

On the other hand some attributes; such as vehicle type, district of accidents, geometrical structure of the roadway (vertical or horizontal sections, etc.), availability of the pedestrian crossings and traffic signs, weather conditions, types of road surface, number of involved vehicles, infrastructural road degradations, existence of road maintenance works, barriers, road lines, shoulders, traffic police and obstacles, along with gender and license type of drivers are all used at the elementary stages of the analysis but removed for further detailed investigation due to overwhelmingly unbalanced nature that they have.

Moreover, the remaining number of some attributes might be enough to be evaluated in the analysis but the percentage of the missing data for them are so high that their inclusion for the analysis may lead to deceptive results. These attributes related to road width, the physical condition of the vehicles, recorded alcohol levels, the psychological and physical states of the drivers, the usage of the driver seat belt were also excluded from thorough analysis.

Attributes regarding the date of accidents converted to seasons that they belong to.

Classes set up by discrete continuous numerical data with respect to the age of drivers and vehicles were required to apply a priori algorithm for association rule mining analysis. Some variables were integrated because of necessity. For example, if the accident occurred at day time under daylight conditions and the illumination variable recorded as the lighting exists or not, then a new variable named "revised lighting condition" was derived by combining these two variables. This is simply because of the fact that there is no need to check whether the street was lightened or not for the accidents occurred during day time, as it is against the nature of the day time conditions. In this way by eliminating the day time situation, it was possible to evaluate the accidents in terms of the availability of the lighting. Accordingly, if the accident occurred at night-time, the new variable allows us to determine whether lighting was available at the time of the accident. In the same manner, existence of intersection and traffic sign variables were also integrated.

The summary statistics of the obtained latest data used in this study are illustrated in Table 2. Data was divided into subsections to identify how the severity of traffic accidents correlate with the attributes under external and environmental, driver, road and vehicle factors as accident-causing factors along with the other group of factors.

External and environmental factors are related to the Seasons of the year, Lighting conditions at Daylight or night-time in that accident occurred. Seasons of the year are defined as Spring (march-may), Summer (june-august), Autumn (september-november) and Winter (December-February). Revised lighting condition is classified into Daylight, Good and Bad Street Lighting at night.

Driver factors are related to the driver age and driver-education level. Ages, one of the driver related factors, are divided into four groups: 18-25, 26-40, 41-64 and 65+ years described as young, middle-aged-1, middle-aged-2 and elderly, respectively. As another driver factor, education level is classified into four groups: Primary school, Middle school, High school and Higher education.

While road types are mainly divided into three groups: Divided, Two-way, One-way, the availability of intersections were also included in dataset. Signal type of intersection was divided into three groups: Unsignalized, Signalized and No-intersection (any section of the roadway apart from intersections).

Age of vehicle was obtained by subtracting the manufacturing years of the vehicle from the accident year.

Driver severity was categorised as Injured and Fatal (first group), and No-injured (second group). The accident causes are classified as Ftc (Following too close), Fty (Failure to yield), Improper Turning, Speeding, Ilc (Improper lane changing), Drunk driving and other. Collision types are taken into consideration under six allocated parts: Rear (Rear end), Sc (Side collision), Rsr (Rollover/skidding/run off Road), HC (Head on Collision), Hitting pedestrian, Boo (Bumping into fixed objects and others).

As can be seen from Table 2, the vehicle damage variable from the raw accident dataset having six classes as Major, Minor, None damaged, Functional damage, Can't move and Not identified, has been reduced to three classes due to the fact that some of them share similar concepts. This reduction was carried out by combining Major and Can't move as Major, Minor and Functional damage as Minor, None damaged and Not identified as None Damaged classes.

Table 2 - Summary of frequencies and distributions of key attributes

Attribute	**F**	**%**	**Attribute**	**F**	**%**
Nature of Accidents			*Driver Factors*		
Driver Severity			**Age Range**		
Injured and Fatal	7050	49	Young	3453	24
Non-injured	7337	51	Middle-aged-1	5035	35
Accident Cause			Middle-aged-2	4604	32
Following too close	1150	8	Elderly	719	5
Failure to yield	2446	17	Unknown	576	4
Improper turn	1583	11	**Education**		
Speeding	4748	33	Primary school	5323	37
Improper lane changing	1583	11	Middle school	1439	10
Drunk driving	1007	7	High school	4172	29
Other	1870	13	Higher education	1583	11
Collision Type			Unknown	1870	13
Rear	2014	14	*Vehicle Factor*		
Side collision	6474	45	**Age of Vehicle**		
Rsr	2014	14	3orLess	2877	20
Head on collision	1007	7	4 between 10	4604	32
Hitting pedestrian	1439	10	11orMore	6906	48
Boo	1439	10			

Table 2 - Summary of frequencies and distributions of key attributes (continued)

Attribute	**F**	**%**	**Attribute**	**F**	**%**
Nature of Accidents (continued)			*Road Factors*		
Vehicle Damage			**Road Section**		
Major damaged	5899	41	Divided	4172	29
Minor damaged	6618	46	Two-way	2590	18
None damaged	1870	13	One-way	432	3
External and Environmental Factors			Intersection	5323	37
Revised Lighting Condition			Roundabout	1151	8
Daylight	9639	67	Other intersections	719	5
Good street lighting	3309	23	**Signal type of Intersection**		
Bad street lighting	1439	10	Signalized	1151	8
Season of the Year			Unsignalized	5899	41
Winter (dec-feb)	2877	20	No-intersection	7337	51
Spring (mar-may)	3309	23			
Summer (june-aug)	3741	26			
Autumn (sep-nov)	4460	31			

F: Frequency, Rsr: Rollover/ skidding/run off road, Boo: Bumping into fixed objects and others

The proportional rate of the accidents occurred each year from 2015 to 2020 with respect to the total number of accidents for this period of time are obtained as 0.19/0.20 /0.19/0.17/0.14/0.11, respectively. Hangi Tablo?

The definition and variety of the values of eleven attributes are listed in Table 2. Some of the items like Day light condition, Unsignalized, 11orMore, Speeding, Sc (Side collision) have high support values. Although some other factors have been analysed, they have not been illustrated at this table due to their relatively low values being under the determined lowest support and confidence criteria selected for the analysis.

3.3. Data Mining

Data mining is exploration of hidden, attractive or worthful patterns in big datasets. The structure of data is important for an effective analysis. Furthermore, it is expected that the importance of collecting quality data will be better understood with the increasing popularity of data mining [28]. Raw data generally have missing and irrational items. To handle this, records should be processed into structured data. Figure 2 indicates one of the mostly used data mining processes proposed by Yu [18].

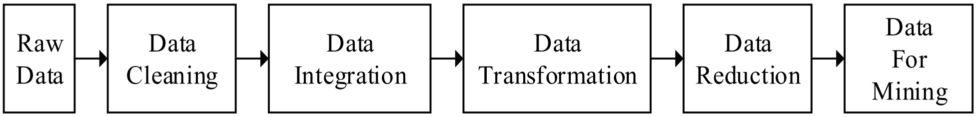

Figure 2 - The process of data mining [15]

The methods to be used for the analysis of the main data mining problems are Regression, Classification, Clustering, Association rule mining, and attribute selection [23]. In this study, Association Rule Mining method (ARM) was chosen for the data analysis of traffic accidents. This method is a combination of nonparametric technique and a machine learning method suggested by Agrawal and Srikant [29].

Within the scope of this research, the outcomes obtained through ARM are evaluated to reveal the interrelationships between the antecedents and consequents of traffic accidents.

The method, on the one hand, ensures to find most effective group of components causing accidents, on the other hand, effectively releases critical and concealed knowledge from the big dataset. The main decisive concepts of association rule mining method are support, confidence and lift. The rules with high support values point out a high frequency of mutual occurrence of the related attributes setting up those association rules. In the same manner, high confidence values state the likelihood of happening of a consequent incident when the antecedent element occurs in an accident. The support of the rule $X \rightarrow Y$ is the percentage of the records that include both X and Y within the whole dataset of $|D|$. The confidence of the rule $X \rightarrow Y$ is the ratio of total records of X and Y occurring together on the condition that X is already at hand to the total records of $|X|$ in the whole dataset. The concept of "*lift*" has been developed to overcome some problematic analysis of the generation of useless association rules. This is mainly due to the fact that some specific circumstances may be contradictory to the nature of the mutual interaction of the attribute pairs although the confidence values may have the same figures for the related attribute pairs. If lift > 1, the association rules are identified as creditable rules. As the value of lift gets higher, the strength of the association rules regarding the related variables becomes more meaningful and worthy [18].

Lift in the concept of data mining analysis represents the amount of times a given rule comes to light to be true in practice over the number of times that related attribute or attributes of the rule appears in the whole dataset. Hence, to evaluate, for example, the frequency of the mutual coexistence of two attributes to the frequency of one of them in terms of all the elements in the dataset, lift values play an important role. As a result of this fact the association rules obtained through the data mining analysis were ordered according to the lift values revealing the set of the most common nature and the structure of the accidents.

The dataset is initially subdivided through a threshold value in a way that those attributes having equal or higher values than this value are considered to be investigated in detail further. In the literature, the predetermined critical support value was taken between 0.001 and 0.05 depending on the main dataset [20, 22, 24, 30-34]. Within the scope of this research, this value has been determined as 0.05 by trying some other higher and lower values, too. The higher values resulted in more but repeating rules. The less values, on the other hand, produced inadequate rules causing the analysis having quite limited meaningful outcomes. Considering the structure and volume of the main dataset of this study, the threshold confidence and lift values are taken as 0.07 and 1.0 respectively.

Although the support and confidence values by themself may produce numerous association rules, it is required to eliminate them in order to obtain the most important and valuable ones. The concept of lift is used for this purpose. While support values express the importance degree of single or combined attributes within the whole dataset, confidence values state the intercorrelation of these attributes regardless of their influence within the available full dataset. On the other hand, lift values evaluate this correlation in terms of its weighted effect within the total dataset. The lift value with this regard plays more important role to determine the strength of an association rule compared to support and confidence values.

The followings express the way how support, confidence and lift values are calculated.

$$Support(X \to Y) = \frac{|X \cup Y|}{|D|} \qquad Confidence(X \to Y) = \frac{|X \cup Y|}{|X|} \qquad Lift = \frac{Conf(X \to Y)}{Sup(Y)}$$

4. RESULTS OF DATA ANALYSIS

Association rule mining is one of the most prevalent methods used to generate useful and unseen knowledge from a large dataset. This method was preferred to generate valuable and meaningful outcomes through the data analysis of this study.

Initially, as shown in Figure 3 the data were grouped into five main groups to acquire association rules. Four of them are considered as the groups with the attributes causing the accidents. The fifth one consists of the layers as shown in Table 2 to represent the nature and structure of the accidents.

The interrelationships between each attribute in the four groups and the fifth group are sought through the association rules to release the important factors causing different type of the collisions. In this way, the relationships between the attributes of driver factors, for example, and driver wounding severity or damage level of the vehicles can be obtained.

At this stage it is quite important to state that the dataset consists of accidents with at least one casualty. In the analysis, the driver related attributes whether they are antecedent or consequent are all only related to the drivers involved in the accidents. With that regard it can be concluded that if, for example, the non-injured-attribute related rules are the case, it means that at least one of the drivers is definitely injured or dead unless there are no injured passengers or pedestrians in the accidents. Even in the case of both drivers are non-injured, then there must be at least either one injured passenger or pedestrian.

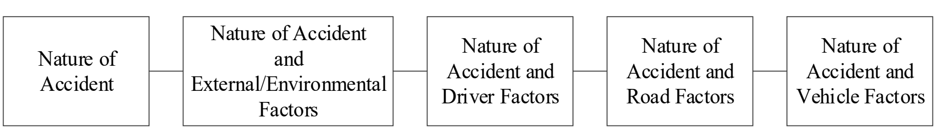

Figure 3 - Five main groups of attributes to obtain association rules

A priori algorithm was used to create the association rules with R Studio package program.

This algorithm iteratively scours the dataset to find frequent items playing important and comprehensive role in the occurrence of the accidents [35].

As shown in Figure 4, 13 association rules were obtained by employing only the attributes regarding the nature of the accident with threshold values of support and confidence equal to 0.05 and 0.70, respectively, and lift greater than 1.0.

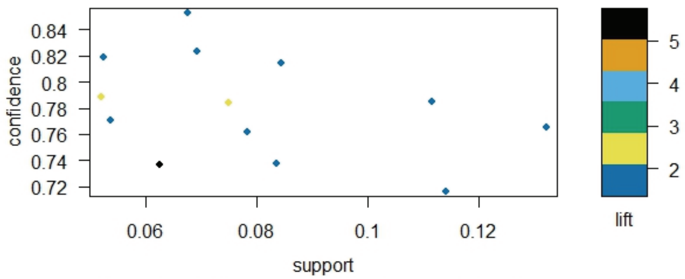

Figure 4 - Scatter plot for 13 association rules regarding the nature of the accident

The top 9 association rules with the highest lift values obtained from the subgroup data based on the nature of the accidents are illustrated in Table 3 in descending order.

In the same way, the top 10 association rules concerning accident and external-environmental features, the top 4 association rules regarding accident and driver features, the top 11 association rules with regard to accident and road features, the top 4 association rules relating accident and vehicle features are also shown in Table 4, 5, 6 and 7, respectively, in terms of the lift values they have.

Table 3 -Top 9 association rules having highest lift values related to nature of the accident

Rule	Antecedent	Consequent	Support	Confidence	Lift
1	Fty	Rear	0.06	0.74	5.38
2	Fty, Minor d.	Sc	0.07	0.78	1.76
3	Non-injured, Fty, Minor d.	Sc	0.05	0.79	1.76
4	Speeding, Rsr	Injured	0.07	0.85	1.74
5	Rsr, Major d.	Injured	0.07	0.82	1.68
6	Improper turn, Minor d.	Non-injured	0.05	0.82	1.62
7	Hitting pedestrian	Non-injured	0.08	0.81	1.61
8	Speeding, Major d.	Injured	0.11	0.72	1.46
9	Improper turn	Non-injured	0.08	0.74	1.46

Ftc: Following too close, Fty: Failure to yield, Minor d: Minor damaged, Rsr: Rollover/skidding/run off road, Major d.: Major damaged

The analysis of results from Table 3 can be summarized as; (1) The highest lift value, 5.38, is related with the Following too close (Ftc) as an antecedent and Rear as a consequent reflecting the relatively strong correlation between Rear collisions and Following too close driving faults. In the same way, high lift value indicates the relatively high weigh of the connection between the Ftc (Following too close) correlated Rear collision to the general weigh of the rear collision accidents in the whole dataset. As can be seen from the support value between these two attributes, 74 rear collision accidents out of 100 rear collision type of accidents are related to following the car in front too closely. (2) Non-injured, Fty (Failure to yield) and Minor damage accidents are mostly related with Sc (Side collision) type of collisions; (3) Speeding and Rsr (Rollover/skidding/run off road) accidents often lead to driver injury. Although there are some other rules with lift values over 1, they have not been included at this table simply because they are the elements of the combinations of the sets constituted by the attributes of the list at the table above. It is worthwhile to mention that the same criterion also applied for all the lists given below.

As shown in Figure 5, there were 72 pieces of association rules that were obtained from the accident and external and environmental factors with the same threshold values as the previous scatter plot has (Figure 4).

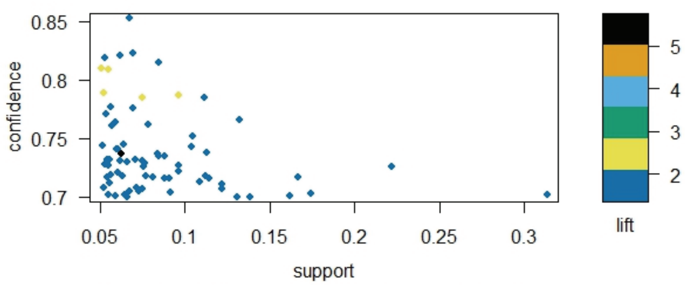

Figure 5 - Scatter plot for 72 association rules between the accident and external and environmental features

Out of these 72 rules, the following ones are the most remarkable ones to mention considering mutual relationship of the attributes with respect to support, confidence and lift values.

As can be seen from Table 4 the highest lift value, 1.81, is related to the interrelation between daylight, Fty (Failure to yield) and Minor damaged attributes leading to Sc (Side collision). The support and confidence values are 0.05 and 0.81, respectively. Another rule reveals the fact that accidents involving daylight and hitting pedestrian attributes result in mostly non-driver injury accidents. On the other hand, daylight and Rsr (Rollover/skidding/run off road) inclusive accidents often lead to driver injuries. There is a strong interdependence between Daylight, Fty (Failure to yield), Minor damaged vehicle and side collision accidents.

Figure 6 demonstrates the 18 pieces of association rules with regard to accidents related to driver factors.

Table 4 - Strong association rules between accident and external-environmental-related attributes

Rule	Antecedent	Consequent	Support	Confidence	Lift
1	Daylight, Fty, Minor d.	Sc	0.05	0.81	1.81
2	Daylight, Fty	Sc	0.10	0.79	1.76
3	Daylight, Hitting pedestrian	Non-injured	0.06	0.82	1.62
4	Daylight, Rsr	Injured	0.07	0.78	1.58
5	Daylight, Speeding, Major d.	Injured	0.08	0.72	1.46
6	Summer, Non-injured, Sc	Daylight	0.06	0.76	1.15
7	Spring, Non-injured	Daylight	0.09	0.73	1.12
8	Summer, Minor d.	Daylight	0.10	0.74	1.11
9	Spring, Speeding	Daylight	0.05	0.73	1.11
10	Summer, Speeding	Daylight	0.07	0.73	1.10

Fty: Failure to yield, Minor d: Minor damaged, Rsr: Rollover/skidding/run off road, Major d.: Major damaged, Sc: Side collision.

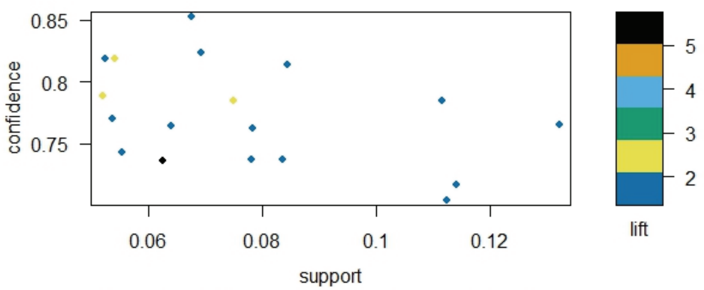

Figure 6 - Scatter plot for 18 association rules related to driver factors

Four rules reflecting important noteworthy relationship based on the lift values have been selected out of these 18 rules and represented in Table 5.

Table 5 - Strong association rules for accident and driver-related attributes

Rule	Antecedent	Consequent	Support	Confidence	Lift
1	Primary School, Fty	Sc	0.05	0.82	1.83
2	Middle-aged-1, Sc, Minor d.	Non-injured	0.06	0.76	1.51
3	Young, Major d.	Injured	0.08	0.74	1.50
4	Middle-aged-2, Sc, Minor d.	Non-injured	0.05	0.74	1.47

Fty: Failure to yield, Minor d: Minor damaged, Major d.: Major damaged, Sc: Side collision.

Regarding the attributes of driver factors causing the accidents, it is worth mentioning that people are less obedient to give way to other vehicles as the education level gets relatively low.

As for the young drivers, it should be noted that they get involved with major damage and injury accidents pointing out the importance of their education and enforcement regarding the way of driving vehicles and obeying the rules. Although middle-aged drivers also involve with side collision types of collisions, the severity of them are relatively low causing generally non-injury accidents.

As shown in Figure 7, there are 128 pieces of association rules obtained from the attributes of road factors triggering the accidents investigated.

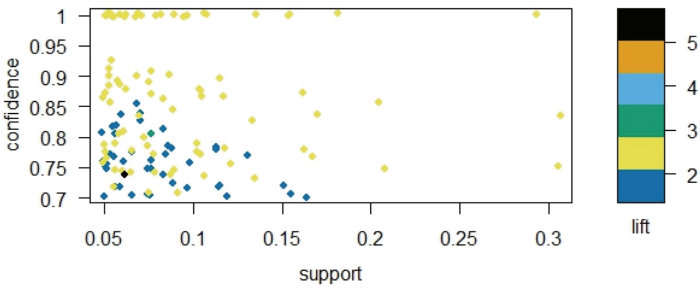

Figure 7 - Scatter plot for 128 association rules regarding road factors

Table 6 - Strong association rules between road factors and nature of the accidents

Rule	Antecedent	Consequent	Sup.	Conf.	Lift
1	Intersection, Fty, Minor d.	Unsignalized	0.05	0.92	2.27
2	Intersection, Fty, Sc	Unsignalized	0.08	0.91	2.24
3	Intersection, Speeding, Sc	Unsignalized	0.07	0.9	2.21
4	Intersection, Non-injured, Sc, Minor d.	Unsignalized	0.07	0.89	2.19
5	Intersection, Injured, Speeding	Unsignalized	0.05	0.88	2.17
6	Unsignalized, Injured	Intersection	0.13	0.73	1.99
7	Divided, Boo	No-intersection	0.05	1	1.96
8	Divided, Ilc	No-intersection	0.05	1	1.96
9	Intersection, Unsignalized, Fty	Sc	0.08	0.87	1.95
10	Intersection, Severe	Unsignalized	0.11	0.77	1.90
11	Roundabout	Sc	0.06	0.72	1.61

Sup: Support, Conf: Confidence, Fty: Failure to yield, Minor d: Minor damaged, Sc: Side collision, Rsr: Rollover/skidding/run off road, Major d.: Major damaged, Boo: Bumping into fixed objects and others, Ilc: Improper lane changing.

Among these numerous rules, the selected ones by taking into consideration the criteria mentioned above have been illustrated in Table 6.

Through the analysis of the results from Table 6, it can be seen that the rule with the highest lift value, 2.27, reflects the strong interaction between Intersection, Fty and Minor damaged with Unsignalized intersections. Failure to give way at intersections leads to minor damaged accidents at unsignalized junctions. The support value in Table 6 indicates that 8 percent of all the accidents are related to side collision accidents occurred on unsignalized intersections due to failure to yield behaviour of the drivers. When the roads are divided, the accidents are generally related to bumping and caused by improper lane changing at any section of the roadway apart from intersections. The location of these type of accidents is denoted as No-intersection. This class term may, obviously, also be used for any accidents occurred at the locations apart from intersections. Rules also state that the accidents at roundabouts are mostly side collision type.

Figure 8 illustrates 17 pieces of association rules for the accidents obtained from vehicle factors.

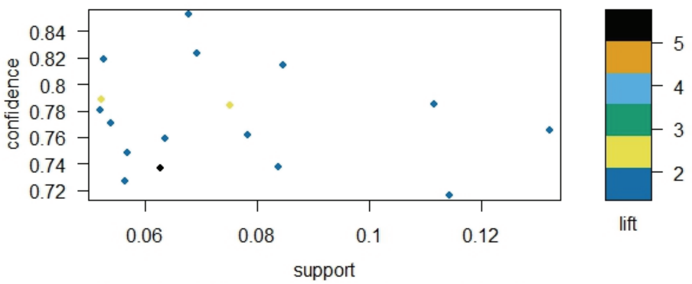

Figure 8 - Scatter plot for 17 association rules regarding the vehicle features and the accidents

Table 7 shows the age of vehicle and the resulted type of the collisions.

Table 7 - Strong association rules between accident and vehicle-related attributes

Rule	Antecedent	Consequent	Support	Confidence	Lift
1	11orMore, Fty	Sc	0.06	0.76	1.7
2	11orMore, Rsr	Injured	0.05	0.78	1.59
3	11orMore, Speeding, Major d.	Injured	0.05	0.73	1.48
4	4-10, Sc, Minor d.	Non-injured	0.06	0.75	1.48

11orMore, 4-10: Age of vehicle, Fty: Failure to yield, Rsr: Rollover/skidding/run off road, Major d.: Major damaged, Sc: Side collision, Minor d: Minor damaged.

As can be seen through the rules from the above table, older vehicles may involve comparatively severe accidents than the newer vehicles. Rule 1 illustrates that the driving attitude of the drivers of the old vehicles seems to be more reckless. This may be related to the perception of the drivers of the old vehicles when they get involved with an accident as the cost of the vehicle damage will be relatively low. On the other hand, with high speed, accidents involving older vehicles result in severe accidents with injuries. New model vehicles, however, get involved with minor accidents resulting in non-injured drivers.

5. CONCLUSIONS

Traffic accidents are quite complex phenomenon with combination of numerous attributes generally recorded on the police reports. The essence of this study is to investigate the mutual effects of the important factors/attributes causing traffic accidents by using association rule mining. After the data preparation process with respect to the attributes either not having proper or balanced contents along with combination of some of them representing the same coverage, the dataset finally had 11 attributes grouped into 5 main factors. These factors are related at least one of the scopes of the nature of the accidents, external and environmental factors, driver related factors, road factors and vehicle conditions. In this study, a priori algorithm was rigorously applied to identify the most frequently appearing attributes setting up the association rules in the dataset affecting traffic accidents occurred in city of Sakarya.

Some of the essential outcomes of this study can be summarized as below:

- Some of the noteworthy attributes to set up the association rules are found to be related to the type of collision, lighting condition, and violations by drivers.
- There exists a significant correlation between speeding, damage of vehicle as Major and severity level of driver`s injury as Injured. In accordance with this finding, Albuquerque et al. also stated that speeding is one of the most important parameters of injury severity in fixed object and run-off-road accidents [36].
- Rollover/skidding and run off road (Rsr) accidents are intercorrelated with damage of vehicle as Major and severity of driver injury as Injured. Karabulut et al. found that accidents involving motorcycles, fixed objects, and run-off-road accidents were the main factors significantly associated with injury and fatality of the drivers [37].
- It has been found that there is a strong correlation between close Following too close and Rear-end collisions. Yu et al. also found the same correlation between these classes [18].
- Hitting pedestrian collisions resulted in Non-injured drivers. In the same manner, Karabulut et al., obtained that drivers are less severely injured in pedestrian accidents [37].
- Most of the accidents occurred when the attributes are in conjunction with Daylight, type of collision Fty, damage of vehicle Minor and Side collision are in effect together. This is mainly because of the inclination of the drivers to take risky manoeuvres as they have relaxed in terms of spiritual and mental conditions for daytime driving [34].

- The evaluation of the obtained rules revealed the fact that Fatal and Injury accidents occur mostly in Spring and Summer months. In addition, the rules clearly made it known that drivers tend to speed during daylight hours in Spring and Summer periods. In harmony with our findings, Celik et al., states that fatal injuries are two times more likely to happen in summer compared to the 'no injury' severity level. In addition, they express that the probability of non-fatal injuries increases by 6.3% in the summertime [38].
- An interesting outcome acquired through the analysis of this research is that the faulty drivers responsible for the accident generally cause the other vehicles` drivers to have more serious Fatal or Injury effect than themselves. This may be due to the fact that as those drivers are aware of the imminent accidents, they are about to cause, they adjust their maneuvers to get the least negative effect of the accident just about to happen. As stated earlier, the analysis process carried out in this research is based upon the data related only to the faulty drivers not the other drivers. The combination of these two facts resulted in this outcome.
- Another rule was revealed in the results that drivers with prominently Primary-educated level were involved in Fatal and Injury accidents as these drivers tend to violate yield rule and were involved in Side collision accidents. Celik et al. also share the same conclusion by remarking that primary-educated drivers increase the probability of fatal injuries [38].
- Eight percent of all accidents are related to Side collision accidents occurring at Unsignalized intersections due to the failure to yield by drivers. Yu et al. also verify the fact that 15% of accidents are resulted from failing to yield at intersections [18].
- Drivers generally involved in Side collisions are from the group of people aged between 26 and 40 (middle-aged-1 group). This may mainly be attributed to the risk-taking behaviour of this group of drivers especially at intersections.
- Accidents involving young drivers between 18-25 years old were resulted in injuries and Major vehicle damage. This is mainly related to the fact that these drivers tend to have relatively speedy drive. Hence, this finding verifies the strong connection between speed and severity type of accidents resulting in severely injured drivers as stated above. Yu et al. state that the drivers aged between 16-25 are more likely to violate driving rules in comparison with other drivers, hence involving more numerous and severe accidents [18]. The results obtained by Celik et al. highlighted the need for education campaigns to address all road users, especially older drivers by taking successful implications as a role model [38]. On the other hand, the finding of this research emphasizes the fact that the campaigns or enforcements are needed more for young drivers.
- As far as the location of the accidents are concerned, it has been obtained that the overwhelming majority of the accidents occurred at Unsignalized intersections including roundabouts. No rule was acquired with regard to Signalized intersection in terms of Fatal or injury accidents. This supports the fact that, Signalised intersections decrease the likelihood of Fatal and injury accidents. Celik et al. also stated that traffic signals at accident locations reduce the probability of fatal injuries by 41.8% [38].
- Mostly, accidents involving older vehicles with high-speed result in severe accidents with injuries. In contrast to the results of this study, the outcomes of the research carried

out by Bédard et al. point out that the "vehicle age" variable was statistically insignificant by investigating the relationship between driver, accident and vehicle characteristics in fatal accidents [39]. This may mainly stem from the fact that they only investigated the fatal accidents.

The most crucial factor of the accidents with regard to side collision types is related to the failure to give way to the vehicles having right of way. This behaviour may be considered to be related to two different attitudes. First one is due to aggressive driving and disobeying the traffic rules. Second one, on the other hand, is associated with lack of education and awareness. While increasing traffic enforcement and fines might have positive effect on solving these problems, education and awareness campaigns for the middle age group drivers might play an important role to reduce the violations of the rules [40,41].

Traffic sign/signal violation is one of the main factors associated with injury and fatality of the drivers [37]. Furthermore, standardization of warning and informative traffic signs is essential to make sure that they are understood by the drivers easily and correctly [42]. In Sakarya, as it is in Turkey, there are different applications of yield signs and majority of them don't contain written notification of "yield" or "give way". The lack of this written information related to yield signs reduces the drivers` awareness. Changing the standardization with added notification of written warning might reduce the violation of give way behaviour of the drivers [43].

This study does not provide any information in connection with level of injury severity along with property damage-only accidents. The outcomes provide the general structure and nature of the accidents` causes in terms of the mutual correlation among the effective attributes of the accidents recorded in Sakarya, Turkey. Hence, the results provide the authorized institutions with the appropriate precautions to be taken to reduce both the number and severity of the traffic accidents. In future studies, predictions regarding the interaction of the attributes will be studied to evaluate the possible accidents and their nature by using machine learning techniques such as decision trees and deep learning methods.

Acknowledgments

The authors are grateful to Republic of Turkey General Directorate of Security for sharing the traffic accident data. The authors would like to thank the reviewers for their invaluable suggestions leading to improve the manuscript.

References

[1] World Health Organization (WHO) 2022. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed August 15, 2022).

[2] Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. 2020.

[3] Injuries in the European Union 2009-2018. EuroSafe; 2021.

[4] Ministry of the Interior. Turkey Road Traffic Safety Strategy Report (2021-2030). 2021.

[5] The World Bank. World Bank 2022. https://data.worldbank.org/country/turkiye (accessed August 25, 2022).

[6] International Transport Forum. Road Safety Annual Report. IRTAD-OECD; 2020.

[7] Turkey Traffic Accident Report-2021. Turkey General Directorate of Highways. 2022.

[8] Kadilar GO. Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach. Int J Inj Contr Saf Promot 2016;23:135–44. https://doi.org/10.1080/17457300.2014.942323.

[9] Zhang G, Yau KKW, Chen G. Risk factors associated with traffic violations and accident severity in China. Accid Anal Prev 2013;59:18–25. https://doi.org/10.1016/j.aap.2013.05.004.

[10] Ma Z, Lu X, Chien SI-J, Hu D. Investigating factors influencing pedestrian injury severity at intersections. Traffic Inj Prev 2018;19:159–64. https://doi.org/10.1080/15389588.2017.1354371.

[11] Batouli G, Guo M, Janson B, Marshall W. Analysis of pedestrian-vehicle crash injury severity factors in Colorado 2006–2016. Accid Anal Prev 2020;148:105782. https://doi.org/10.1016/j.aap.2020.105782.

[12] Alver Y, Demirel MC, Mutlu MM. Interaction between socio-demographic characteristics: Traffic rule violations and traffic crash history for young drivers. Accid Anal Prev 2014;72:95–104. https://doi.org/10.1016/j.aap.2014.06.015.

[13] Katanalp BY, Eren E. The novel approaches to classify cyclist accident injury-severity: Hybrid fuzzy decision mechanisms. Accid Anal Prev 2020;144:105590. https://doi.org/10.1016/j.aap.2020.105590.

[14] Adanu EK, Lidbe A, Tedla E, Jones S. Factors associated with driver injury severity of lane changing crashes involving younger and older drivers. Accid Anal Prev 2021;149:105867. https://doi.org/10.1016/j.aap.2020.105867.

[15] Chiou Y-C, Fu C, Ke C-Y. Modelling two-vehicle crash severity by generalized estimating equations. Accid Anal Prev 2020;148:105841. https://doi.org/10.1016/j.aap.2020.105841.

[16] AlKheder S, AlRukaibi F, Aiash A. Risk analysis of traffic accidents' severities: An application of three data mining models. ISA Trans 2020;106:213–20. https://doi.org/10.1016/j.isatra.2020.06.018.

[17] Cai Q. Cause Analysis of Traffic Accidents on Urban Roads Based on an Improved Association Rule Mining Algorithm. IEEE Access 2020;8:75607–15. https://doi.org/10.1109/ACCESS.2020.2988288.

[18] Yu S, Jia Y, Sun D. Identifying Factors that Influence the Patterns of Road Crashes Using Association Rules: A case Study from Wisconsin, United States. Sustainability 2019;11:1925. https://doi.org/10.3390/su11071925.

[19] Das S, Tamakloe R, Zubaidi H, Obaid I, Alnedawi A. Fatal pedestrian crashes at intersections: Trend mining using association rules. Accid Anal Prev 2021;160:106306. https://doi.org/10.1016/j.aap.2021.106306.

[20] Kong X, Das S, Jha K, Zhang Y. Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. Accid Anal Prev 2020;144:105620. https://doi.org/10.1016/j.aap.2020.105620.

[21] Zhu S. Investigation of vehicle-bicycle hit-and-run crashes. Traffic Inj Prev 2020;21:506–11. https://doi.org/10.1080/15389588.2020.1805444.

[22] Hong J, Tamakloe R, Park D. Discovering Insightful Rules among Truck Crash Characteristics using Apriori Algorithm. J Adv Transp 2020;2020:1–16. https://doi.org/10.1155/2020/4323816.

[23] Das S, Dutta A, Sun X. Patterns of rainy weather crashes: Applying rules mining. J Transp Saf Secur 2020;12:1083–105. https://doi.org/10.1080/19439962.2019.1572681.

[24] Xu C, Bao J, Wang C, Liu P. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. J Safety Res 2018;67:65–75. https://doi.org/10.1016/j.jsr.2018.09.013.

[25] Turkish Statistical Institute 2021. https://cip.tuik.gov.tr/# (accessed December 16, 2022).

[26] Sakarya City Guide. Sakarya Metropolitan Municipality. https://www.sakarya.bel.tr/ (accessed March 10, 2023).

[27] Turkey General Directorate of Highways. http://yol.kgm.gov.tr/KazaKaraNoktaWeb/ (accessed March 1, 2022).

[28] Hand DJ. Principles of Data Mining 2007:2.

[29] Srikant R, Agrawal R. Mining Generalized Association Rules. 1995.

[30] Geurts K, Thomas I, Wets G. Understanding spatial concentrations of road accidents using frequent item sets. Accident Analysis and Prevention 2005; 37:787–99. https://doi.org/10.1016/j.aap.2005.03.023.

[31] Montella A. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid Anal Prev 2011;43:1451–63. https://doi.org/10.1016/j.aap.2011.02.023.

[32] Montella A, Aria M, D'Ambrosio A, Mauriello F. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accid Anal Prev 2012;49:58–72. https://doi.org/10.1016/j.aap.2011.04.025.

[33] Pande A, Abdel-Aty M. Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Saf Sci 2009;47:145–54. https://doi.org/10.1016/j.ssci.2007.12.001.

[34] Das S, Dutta A, Jalayer M, Bibeka A, Wu L. Factors influencing the patterns of wrong-way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. Int J Transp Sci Technol 2018;7:114–23. https://doi.org/10.1016/j.ijtst.2018.02.001.

[35] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Rec 1993;22:207–16. https://doi.org/10.1145/170036.170072.

[36] Albuquerque FDB de, Awadalla DM. Roadside Fixed-Object Collisions, Barrier Performance, and Fatal Injuries in Single-Vehicle, Run-Off-Road Crashes. Safety 2020;6:27. https://doi.org/10.3390/safety6020027.

[37] Karabulut NC, Ozen M. Exploring Driver Injury Severity Using Latent Class Ordered Probit Model: A Case Study of Turkey. KSCE J Civ Eng 2023;27:1312–22. https://doi.org/10.1007/s12205-023-0473-6.

[38] Celik AK, Oktay E. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. Accid Anal Prev 2014;72:66–77. https://doi.org/10.1016/j.aap.2014.06.010.

[39] Bédard M, Guyatt GH, Stones MJ, Hirdes JP. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accid Anal Prev 2002;34:717–27. https://doi.org/10.1016/S0001-4575(01)00072-0.

[40] Factor R. The effect of traffic tickets on road traffic crashes. Accid Anal Prev 2014;64:86–91. https://doi.org/10.1016/j.aap.2013.11.010.

[41] Paleti R, Eluru N, Bhat CR. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accid Anal Prev 2010;42:1839–54. https://doi.org/10.1016/j.aap.2010.05.005.

[42] Manual on Uniform Traffic Control Devices. MUTCD.pdf 2009.

[43] Cakici Z, Murat YS. An Investigation on the Awareness of Traffic Signs: Denizli Sample. Bitlis Eren Üniversitesi Fen Bilim Derg 2017;6:21–21. https://doi.org/10.17798/bitlisfen.305485.