

Analysis of Traffic Accidents for Road Safety Using Data Mining Techniques

Muhammad Waseem, Umair Khadam, Anila Shakeel, Iram Aftab, Muhammad Munwar Iqbal and Muhammad Yasir Shabir

Abstract—Road safety is a massive issue in the world while traveling from one place to another. The road accidents very severely affected the life of travelers. Several accidents occur on daily and cause several deaths and injuries. Therefore, it is vital to identify the causes of accidents in this critical situation to reduce these daily road accidents. In this research paper, we use some data mining algorithms like Apriori and F.B.-Growth algorithms to analyze the factors that cause road accidents and find their associations and correlations. Our research findings show the association and correlations of accident factors and show that our proposed models accurately classify road accidents. These findings help us to improve safety from road accidents. The experiential setup shows how we find casualty severity by using different attributes like light conditions, weather conditions, road surfaces, and vehicles. We use Association rules to find out which rule is best for preventing accidents and improve road accidents safety.

Index Terms— Road Accidents, Road Safety, Association Rules, Classification.

I. INTRODUCTION

THE most critical human threats in the whole world are accidents. In 2007, around 27567 human beings died, and 276762 were injured due to road accidents. The United States has 1 out of 40 deaths from street injuries inside the global [1].

Manuscript received; Mar 12, 2021; accepted Oct 4, 2022.

M. Waseem (email: <u>waseemengr55@gmail.com</u>) is affiliated with the Computer Science Department, University of Kotli Azad Jammu Kashmir, Pakistan.

U.Khadam(email: <u>umair khadim@uokajk.edu.pk</u>) is affiliated with the Software Engineering Department, University of Kotli Azad Jammu Kashmir, Pakistan.

A.Shakeel(email:anilashakeel56@gmail.com),I.Aftab(email:iramafta b84@gamil.com) and M.Y. Shabir(email: yasir.shabir14@gmail.com) are affiliated with the Department of Computer Science, University of Azad Jammu and Kashmir, Pakistan.

M. M. Iqbal (email: <u>munwariq@gmail.com</u>) is affiliated with the Department of Computer Science, University of Engineering and Technology Taxila, Pakistan.

*Corresponding Author: email: umair_khadim@uokajk.edu.pk

The foremost cause that facts mining has brought a topnotch deal of interest in facts enterprise in late years is the considerable time of larger quantities of information and the approaching want to swing such records into adequate facts and observation. The utility of data mining on social issues has been a famous approach nowadays. Mortal prices because of street accidents contribute extra to the total loss of life price sector. Every 12 months, millions of people die and suffer non-fatal accidents [2].

Sample mining is an emerging method in statistical analysis that is inside the current years because of the expanded capability of amassing and storing data that has been taken into consideration. Among the essential mining packages, prediction and pattern reputation is widely used. However, nowadays fast increase in the volume of databases is a form that human ability to recognize this data is not possible without powerful gear. In this example, decision-making based on the statistics predates managers and users, because selection makers do not have adequate tools for extracting valuable statistics [3][4].

Investigated the capability utility of information mining gear to develop structures assisting the recognition and hypothesis of the vital motive force and car threat elements that purpose RTAs [5]. It could help accomplice guidelines of common navigation injuries to improve navigation safety and help the aviation community with recommendations for reviewing coverage and other plans for upgrades. These days there is a want to analyze from beyond records to use prediction of future occurrences of activities and use of the beyond information. Typically, this has a look that can assist them in determining the system development and administration improvement for the improvement of navigation security [6]. In Iran average, 100 people die in accidents every day. Furthermore, 3000 people are dying in road accidents daily [7]. The report by [8] says that children between 10 to 19 years die due to road incidents. It is the leading cause of death.

II. LITERATURE REVIEW

Shanthi et al. proposed a system that analyzes and predicts accidents' severity. Coincidence statistics by using mining techniques. University of Alabama furnished this proposed system. That includes many techniques (C4, CR-T, CS-CRT, Naïve Bayes, and Random tree). The consequences were evaluated using the accuracy measures including, recall and Precision [2]. Pak Gohar et al. proposed a gadget for the position of the human factor in incident and hardness of street fall apart based on the cart and regression. The visitors' injuries had been judged to be the most crucial health hassle in Iran. The traffic records are evaluated using logistic type and regression timber statistics mining strategies. The proposed device will assist governments in higher avenue designs and traffic management [9, 10]. Kumar et al. proposed a statistics mining system to symbolize the avenue accident region. Most street coincidence facts evaluations use statistics mining methods that specialize in figuring out issues that have an effect on the severity of a coincidence. Any harm attributable to street injuries is continually unsatisfactory in phrases of health assets harm and other monetary elements strategies carried out a k-method finding to the organization. The accident regions into three categories of twist of fate locations. The analysis of these places can help classify certain street accident capabilities that make a street accident regularly arise in these places [6]. Abouamouna et al. proposed a gadget that predicts avenue accidents in Qatar via 2022. There is a spreading subject over Qatar's increasing incidence of avenue injuries and consequent loss of human lifestyles. Qatar has to place into consideration future deaths because of street accidents and beyond trends have to be taken into consideration to give an affordable photo of what may appear in the future the strategies [11]. That amuses the current case in Qatar where a couple of linear regression models and artificial neural networks those techniques are analyzed, and their findings are in comparison. we conclude that MRL gave better outcomes than ANN because the artificial impartial network does not shape information with a massive variety. Beshah et al. [12] proposed a system that road traffic coincidence information to upgrade security anticipated death in accidents. Globally every year in Nigeria, the main reason for death and injuries record avenue traits to twist of fate harshness. Developed fixed guidelines that the Nigerian visitors' organization might use to reduce coincidence severity. Atnafu et al. proposed a gadget that researches and concludes the character of road network twist of fate the use of report mining methods in Maharashtra India. India is a few of the rising countries where visitor's coincidence takes place is greater than the essential restrict. within the proposed method random tree j48 and Naive Bayes algorithms are used. The outcomes of the three algorithms are compared and then the prediction model is completed using the set of rules which proves to be the excellent [10]. Jayasudha et al. proposed a device that reviews mining techniques in road traffic and a twist

of fate. That has analyzed several of the information mining techniques equipment packages and like google for coincidence investigation and visitors' reviews principal with the aid of using a few gear techniques or many methods within the subject of statistics mining the usage of visitor's records bases. Moradkhani et al. proposed a gadget that analysis avenue twist of fate records analysis identifying the reasons for street injuries. The main intention to lessen the harm due to traffic injuries then again because of the exponential boom of statistics volume. Evaluation of the factors in this regard is impractical. Discover elements behind road visitors' injuries using mining methods together with basket analysis and hold control of it. Kuda statistics set is used in this study and the assessment and employment through (Weka) had been completed the consequences of statistics mining techniques will assist corporations. That includes transportation to explore the accidents statistics recorded using the police information system and discover styles and export destiny behaviors and powerful selection to be taken to reduce accidents [5, 13]. Ramachandiran et al. proposed a device for developing navigation events using an affiliation rule mining algorithm and the time series analysis. Show the specific hazards related to the global transportation structures concerns utility of numerous statistics analytical algorithms to derive a time series forecasting technique. Common coincidence and incident patterns specifically for Philippine aviation putting for prediction conditions that could boom the chance of aviation accidents and incidents regarding deadly and non-deadly. Sports time collection forecasting methods using linear regression Gaussian procedures multilayer perception and smore are carried out to datasets the results created vital factors in enhancing the aviation operation. [6, 10]. Prediction inside the United Arab Emirates and Abu-Dhabi in 2014. According to the report, 941 traffic accidents have been recorded which, contributing to 121 fatalities and one hundred thirty-five extreme accidents. They used data mining strategies to establish classifiers to expect the harm severity of any new accident with reasonable accuracy [14]. Prati et al. [10] proposed a machine expecting the hardness of bicycle accidents. Using mining methods to research the issues predicting hardness of bicycle crashes in Italy. we used an observation examine of official data they implemented two of maximum widely used records mining techniques chain selection tree techniques and Bayesian networks evaluation however end results knowledge no studies have used. Each chain decision tree access and Bayesian network analysis inside the observe of the hardness of bicycle crashes [15]. Abellan *et al.* [16]. Use the choice regulations to perform the analysis accidents of rural dual carriage way of twolane in Spain and Granada. The accuracy of classifiers the use of characteristic choice algorithms has been in comparison and the consequences have shown that Random Tree. The use of function rating algorithm higher plays different algorithms in modeling the car collision patterns in street accident data [17]. TABLE 1 shows the

Personal Use permitted, but republication/redistribution requires UW permission.

comparison of different techniques. Table 1 consists of the following things as Author name, year, problem, solution, and Future work or drawbacks.

COMPARISON OF DIFFERENT TECHNIQUES					
#	Author	Year	Problem	Solution	Future Work
1.	Shanthi et al. [2]	2012	Factors influence the road traffic accidents	Use different algorithms such as viz. C4.5, CR-T, ID3, CS-CRT, CS-MC4, Naïve Bayes And Random Tree,	Reduce the fatality rate
2.	Pakgohar et al. [9]	2017	Characterize road accident locations	Applied k-means algorithm	Apply more algorithms such as the Apriori algorithm
3.	Abou-amouna [11]	2014	Increasing incidences of loss of human life	Use linear regression model (MLR) and artificial neutral	Further investigation with respect to dimensional reduction.
4.	Atnafu <i>et al.</i> [10]	2017	Analysis and prediction of road accidents	Classification and association rules	In the future gain more information by using decision trees some other mining tools.
5.	Jayasudha [3]	2017	Accident investigation and traffic analysis	Using the traffic databases categories.	C
6.	Moradkhani [5, 13]	2014	Reduce the damage caused by traffic accidents	Using basket analysis algorithm	
7.	Ramachandiran et al. [6, 10]	2018	Factors that lead to the criticality of the accident	Used classification algorithms such As naive bayes classification	Gain more accuracy by using some more datasets.
8.	Taamneh <i>et al</i> . [14]	2016	Traffic Accident Modeling and Prediction	Decision Tree Rule Induction, Naive Bayes, and Multilayer Perceptron.	
9.	Prati <i>et al</i> . [15]	2017	Predict bicycle accidents factors	Used chain decision tree and Bayesian networks	Pedestrian bicycle type collisions
10.	Kumar et al. [11]	2016	Approach to characterize road	Use decision tree and k-means	
11.	Beshah et al. [12]	2010	Traffic Accident Data to Improve Safety	Developed a set of rules	

 TABLE 1

 Comparison of Different Techniques

Year 2021



■Jan ■Feb ■March ■April ■May ■June ■July ■August ■Sep ■Oct ■Nov ■Dec

Fig. 1. No of Accidents in the Year 2021

Accidents occurred on a daily basis and cause a number of deaths and injuries are shown in *Fig. 1*. In this critical situation, it is very important to identify the causes of accidents so that these daily road accidents can be reduced. In previous research, analysts used many data mining techniques and identify the main causes for the severity of casualty, but further research is required with a huge dataset for providing security for road accidents.

III. RESEARCH METHODOLOGY

Data mining is a technique that is used to turn raw data into useful information. Different methods are used in data mining such as decision tree, vector, bayesian's inference, cluster analysis, apriori algorithms, k-nearest neighbors' algorithms, c4.5 algorithms, pagerank, and fp growth. First, we investigated those factors which affect road accidents. The main road accidents factors are inexperience, lack of skills, risk-taking behaviors, age factors such as younger and old drivers, and gender factors such as male or female. In 2021 the number of deaths, ratio, rate, and rank of different countries is given in Table 2.

TABLE 2					
DEATH RATIO OF DIFFERENT COUNTRIES DUE TO ROAD ACCIDENTS					
Country	Deaths	%	Rate	Rank	
Zimbabwe	388092	4.56	63.47	02	
Kenya	300090	4	48.00	12	
India	299,091	3.39	23.87	60	
China	284,986	2.91	17.73	89	
Pakistan	30,046	2.42	17.12	95	
Bangladesh	20,816	2.68	14.43	106	
Australia	1243	0.92	4.62	161	
Germany	3278	0.43	3.23	169	

We proposed a solution for this problem. The graphical representation of the proposed model is shown in *Fig. 2*. Road accident 2015 Dataset will be used for analyzing and predict road accidents. We use data mining techniques or algorithms such as F.P. Growth to predict and analyze road accidents. For this purpose, we collect data or datasets related to the factors that affect road accidents, such as age factors, inexperience, lack of skills, driver categories such

Personal Use permitted, but republication/redistribution requires UW permission.

as younger and old drivers, and classified gender factors such as males and females. After data collection, the preprocessing technique was applied to clean up the data and the missing values were adjusted, and the raw data was canceled in helpful information. Then, applying different algorithms on training data, classifying and categorizing the factors such as age, gender, inexperience, and lack of skills into younger age, old age, male, female, licensed, and unlicensed drivers, and applying classification rules based on algorithms and testing the data using data mining tools such as weka. After testing the datasets, we predict accidents severity and observed the fetal/simple injury and motor collision. By using this solution predict and analysis of road accidents. We use the F.P. Growth algorithm because execution time is lesser than Apriori due to the absence of candidates. Association rules are used to create frequent patterns. In association rules support and confidence extract the rules from the given dataset.



Fig. 2. Proposed Model

A. Support

Support only describes how frequently an item or itemset appears or is present in the database. The support average is calculated using some mathematical equations Eq (1) is given below.

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

In the equation (1) N belongs to the number of accidents, \cap which means the combination of both A and B. A and B are two items and P represents the probability of A and B.

B. Confidence:

In association rules mining confidence means the number of times both items A and B occur together. The equation of confidence is described below.

$$Confidence = \frac{Support (AUB)}{Support (A)}$$
(2)

Confidence is a percentage value that shows how frequently the rule occurs among all the groups. Equation ((2) describes how A and B are occurred together. And describes how frequently the rules Concur among both A and B.

C. Lift:

$$Lift (A \to B) = \frac{support (AUB)}{Support (B)X Support (A)}$$
(3)

The above equation (3) describes the lift of a rule. It can be described as the ratio of the observed support to that expected provided A and B, where independent is referred to as a lift.

D. Conviction:

It describes A's probability if B occurs, or A occurs without B. For example, the conviction rule described in eq. (4) given below.

Convicition =
$$\frac{P(A)XP(B)}{P(A \cap B)}$$
 (4)

IV. DATA PREPROCESSING

Pre-processing is a data mining technique that converts the raw data into an understandable form. Pre-processing techniques are essential because the raw data means actual world data is not always complete and contains some errors. So, we cannot send the data through a model before pre-processing. That is why we need to preprocessing data before sending it through a model.



Fig. 3. Overall Visualization of Dataset

Fig. 3 shows the overall visualization of dataset 2015. The dataset contains 9 attributes and 2664 values. The attributes are shown in *Fig.* 4.

Number of Vehicles	Road Surface	Lighting Conditions	Weather Conditions	Casualty Class	Casualty Severity	Sex of Casualty	Age of Casualty	Type of Vehicle
5	Snow	Daylight: street lights present	Snowing without high winds	Driver/Rider	Slight	Male	36	Goods vehicle 3.5 tonnes mgw and under
5	Snow	Daylight: street lights present	Snowing without high winds	Passenger	Slight	Male	27	Goods vehicle 3.5 tonnes mgw and under
1	Wet/ Damp	Darkness: street lights present and lit	Fine without high winds	Pedestrian	Serious	Male	68	Car
2	Dry	Daylight: street lights present	Fine without high winds	Driver/Rider	Slight	Male	49	Car
2	Wet/ Damp	Daylight: street lights present	Fine without high winds	Driver/Rider	Slight	Male	33	Car
2	Wet/ Damp	Daylight: street lights present	Fine without high winds	Passenger	Slight	Female	28	Car
2	Dry	Daylight: street lights present	Fine without high winds	Driver/Rider	Slight	Male	17	Motorcycle over 50cc and up to 125cc
2	Wet/ Damp	Daylight: street lights present	Fine without high winds	Driver/Rider	Slight	Female	18	Car
2	Wet/ Damp	Daylight: street lights present	Fine without high winds	Passenger	Slight	Male	8	Car
1	Wet / Damp	Darkness: street lights present and lit	Raining without high winds	Passenger	Slight	Female	73	Bus or coach (17 or more passenger seats)

Fig. 4. Structure of Dataset 2015

Fig. 5 Show different pre-processing techniques such as Data cleaning, Data integration, Data Transformation and Data Reduction.

A. Data Preprocessing Techniques



Fig. 5. Data Preprocessing Techniques

i. Data Cleaning

The first technique of pre-processing you used in our proposed work is data cleaning. Data cleaning is preparing raw data for some analysis and removing the insufficient data from the dataset. First, recognizing the real-world data means raw data and handling the missing values. Therefore, before analysis of data, data cleaning is vital. Because without data cleaning we do not produce the correct results.



Fig. 6 shows the graphical representation of data cleaning process. Data cleaning contains different approaches such as Handling Missing Values, Managing Unwanted Outliers, Fixing Structural Errors, and removing unwanted observations.

ii. Handling Missing Values:

In data cleaning, we first handle the missing values or null values. We have used several methods during data cleaning and handling missing values. Such as First of all, we delete rows or columns that contain Null values. We usually use this method on empty cells. *Fig.* 7 shows result of data cleaning after handling missing values. Now the dataset is clean, and all missing values are handled.

Number of Vehicles	0
Road Surface	0
Lighting Conditions	0
Weather Conditions	0
Casualty Class	0
Casualty Severity	0
Sex of Casualty	0
Age of Casualty	0
Type of Vehicle	0
E. 7 Dente Charles Charles	

Fig. 7. Result of Data Cleaning

V. RESULTS AND DISCUSSIONS

Most accidents happened in October and November. In October, 525 accidents and 501 accidents happened in November 2015. And least on august 200 accidents. In data pre-processing first, we handle the missing values where we remove or ignore the missing values. Handling, missing values convert the numerical data in nominal because Apriori and FP-Growth algorithms work only with the nominal dataset. Fig. 8, shows the sum of accidents in weather and light conditions in 2015. After computing the results, we find out that fines without high winds are the leading cause of accidents. In weather conditions, the main reason for accidents is satisfactory without high winds. In this dataset 2015, the number of accidents is 4337. However, it reduces slightly in lighting conditions, and the total accidents are 3837. In light condition Daylight: the street light present is the main cause of accidents.

Fig. 9, shows the analysis of the Road surface. Road surface factor consists of 4 attributes, as shown in *Fig.* 9. After analysis, the results show no. of accidents due to the road surface. That consists of Snow Wet/Dump, dry, and Frost/Ice attributes. Forty-seven accidents were due to Snow, 1373 due to Wet/Dump, 3728 dues to Dry surface, and only 29 accidents occurred due to Frost/Ice surface. So, we analyzed that most accidents are occurred due to Dry surfaces and the least in Frost/Ice surfaces. After results generation, we analyzed that most accidents are occurred due to Dry surfaces are the main reason for accidents. The sum of accidents due to Dry surface is 3728.



Fig. 8. Sum of Accidents in weather and light



Fig. 9. Sum of Accidents in the Dry Road Surface

Fig. 10, shows that the analysis of vehicle types. The vehicle type factor consists of 7 attributes as shown in *Fig.* 6. After Analysis the results show that the No. of accidents is due to vehicle types. That consists of Goods vehicles, Cars, Motorcycles, Bus or Coach, Pedal cycle and Taxi or private car.159 accidents due to goods vehicles, 3389 due to cars, 174 due to motorcycles, 323 due to Bus or Coach, 622 due to Pedal cycle, and 238 due to Taxi or private car.



Fig. 10. Sum of Accidents in Vehicle Types

After result generation, we analyzed that most accidents are occurred due to Cars. In-vehicle types the main reason

for accidents is Cars. The sum of accidents due to Car factor is 3389. We use different data mining techniques to implement the proposed work, such as the Naïve Bayes algorithm.

Pseudocode of Naïve Bayes Algorithm:

Input:

Training Dataset T,

 $F{=}$ (f1, f2..., fn) ${/\!/}$ Value of the predictor variable in testing dataset

Output:

A class of testing dataset.

Steps:

- 1. Read the training dataset T
- 2. Calculate the mean and Standard deviation of the predictor variables in each class.
- 3. Repeat Calculate the probability of all predictor variables $(f_1, f_2, f_3...f_n)$
- 4. Calculate the likelihood for each class.
- 5. Get the greatest likelihood;

In classification, the PART classifier is used with the training set, and it gives 68%, Naïve Bayes gives 80% and OneR gives a 97% accuracy rate on dataset 2015. The top 10 association rules are shown in

TABLE 3. The attribute's confidence means that the number of times both items A and B occur together. In association rules mining the confidence the number of times both items A and B occur. Support only describes that how frequently an item or itemset appears or is present in the database. It can be described as the ratio of the observed support to that expected provided A and B, where independent is referred to as a lift.

Most accidents happened in October and at least in February. Due to weather conditions, 4337 accidents happen concerning. Fine without high wind factor. According to the type of vehicles Cars are the main reason for accidents. In sex of casualty, the main factor is male, the reason for accidents 2972 cases reported. In light factors, 3887 cases of accidents are reported with Daylight: street light present. On the surface road, 3728 accidents were reported due to dry surface attributes. The Overall accuracy of different techniques is given below in *Fig. 11*.

We used the Navies Bayesian classifier for the Analysis and Prediction of accidents. Because Naïve Bayes classifier is Fast, efficient, and easy to implement. Moreover, they are efficiently categories the data. The comparison of SVM, KNN and Naïve Bayes are described in Table 4.

The comparison of the proposed model results with previous results is given below in *Fig. 12*. Prayag et al. [18] proposed a model that analyzes road accidents by using data mining techniques. It can expose the relationship of different attributes.

UW Journal of Science and Technology Vol. 6 (2022) 37-44 ISSN: 2523-0123 (Print) 2616-4396 (Online)

TABLE 3
ASSOCIATION RULES DISCOVERED BY APPLYING APRIORI ALGORITHM

-		
#	Rules	Confidence
1	IF {Road Surface= Dry AND Sex of Causality= Male} THEN {Weather Condition = Fine without high winds}	0.98
2	IF {Road Surface= Dry AND Causality Class= Driver/Rider} THEN {Weather Condition = Fine without high winds}	0.98
3	IF {Road Surface= Dry AND Daylight= Street light present} THEN {Weather Condition = Fine without high winds}	0.97
4	IF {Road Surface= Dry AND Daylight= Street light present} THEN {Casualty Severity=Slight}	0.97
5	IF {Road Surface=Dry} THEN {Weather Conditions=Fine without high winds}	0.97
6	IF {Road Surface= Dry AND Casualty Severity=Slight} THEN {Weather Conditions=Fine without high winds}	0.97
7	IF {Road Surface= Dry AND Type of vehicle=Car} THEN {Weather Conditions=Fine without high winds}	0.97
8	IF {Street lights present AND Type of Vehicle=Car} THEN {Casualty Severity=Slight}	0.91
9	IF {Type of Vehicle=Car} THEN {Casualty Severity=Slight}	0.91
10	IF {Road Surface=Dry AND Type of Vehicle=Car} THEN {Casualty Severity=Slight}	0.91



Fig. 11. The overall accuracy of the proposed model using different techniques

TABLE 4 COMPARISON OF NAIVE BAYES WITH OTHERS					
SVM	KNN	Naïve Bayes			
Complex	Sensitive for	Fast, efficient, and			
calculation if	Noisy data	easy to implement			
there are many					
class labels					
Not for	Not handle	Easy to extend to			
continuous	irrelevant	multi-class			
numerical	attributes	classification			
variables,		problems			

They apply NAÏVE BAYES technique and generate a result with 68% accuracy. Tibebe et al. [5] proposed a model called mining road traffic accident data for the purpose to improve safety. They also apply NAÏVE BAYES technique and generate a result with 78 percent accuracy. Prayag et al. [18] and Tibebe et al. [5] Authors used Naïve Bayes technique and generate 68 and 78 percent accuracy. In the proposed model NAÏVE BAYES technique generates results with 80% accuracy.

In the Proposed Model, Prayag et al. [17] and Tibebe et al. [4] used the same classifier, the Naïve Bayes classifier but the accuracy of Prayag et al. [17] is 68% and Tibebe et al. [4] is 78%. However, the accuracy of proposed model is 80%, because the invention of elements or attributes.



Fig. 12. The comparison of proposed model results with previous results

VI. CONCLUSIONS

In this research, the experiential setup shows how we find casualty severity by using attributes like light conditions, weather conditions, road surface, and type of vehicles. A data mining technique known as FP-growth is used to apply association to the dataset to predict the causes of road accidents. Association rules are also used to determine which rule is best for preventing accidents and improving road accidents safety. This research show that the most critical factors that cause road accidents. In the future, we will improve the study of road accidents with the help of a large dataset.

REFERENCES

- A. Pakgohar, et al., 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, Vol. 3, pp. 764-769, 2011.
- [2] S. Shanthi, and R.G. Ramani. Feature relevance analysis and classification of road traffic accident data through data mining techniques. Proceedings of the World Congress on Engineering and Computer Science, Vol. 1, pp. 1-6, 2012.
- [3] F. Moradkhani, S. Ebrahimkhani, and B. Sadeghi Begham, Road accident data analysis: A data mining approach. Indian Journal of Science and Research, Vol. 14(3), pp. 437-443, 2014.

- [4] T. Beshah and S. Hill. Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia, AAAI Spring symposium series, pp. 1-6, 2010.
- [5] P. P. R. Abesamis, R. de Dios Bulos, and M. Ching. Improving Aviation Incidents using Association Rule Mining Algorithm and Time Series Analysis. in IOP Conference Series: Materials Science and Engineering, IOP Publishing, pp. 1-15, 2020.
- [6] J. K. Caird et al., A meta-analysis of the effects of texting on driving. Accident Analysis & Prevention, Vol. 71, pp. 311-318, 2014.
- [7] W. H. O. D. Violence, et al., Global status report on road safety: time for action, World Health Organization, pp. 1-301, 2009.
- [8] S. Kumar and D. Toshniwal, A data mining approach to characterize road accident locations. Journal of Modern Transportation, Vo. 24(1), pp 62-72, 2016.
- [9] K. Jayasudha and C. Chandrasekar, An overview of data mining in road traffic and accident analysis. Journal of Computer Applications, Vol. 2(4), pp. 32-37, 2009.
- [10] M. Abou-Amouna et al., Prediction of road accidents in Qatar by 2022. World Academy of Science, Engineering and Technology, International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering, Vol. 8(2): pp. 458-463, 2014.
- [11] B. Atnafu and G. Kaur, Analysis and predict the nature of road traffic accident using data mining techniques in Maharashtra India. International Journal of Engineering Technology Science Research, Vol. 4(10), pp. 1153-1162, 2017.
- [12] V. Ramachandiran, P.K. Babu, and R. Manikandan, Prediction of road accidents severity using various algorithms. International Journal of Pure and Applied Mathematics, Vo. 119(12), pp. 16663-16669, 2018.
- [13] M. Taamneh, S. Alkheder, and S. Taamneh, Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. Journal of Transportation Safety and Security, Vol. 9(2), pp. 146-166, 2017.
- [14] G. Prati, L. Pietrantoni, and F. Fraboni, Using data mining techniques to predict the severity of bicycle crashes. Accident Analysis and Prevention, Vol. 101, pp. 44-54, 2017.
- [15] J. Abellán, G. López, and J. De OñA, Analysis of traffic accident severity using decision rules via decision trees. Expert Systems with Applications, Vol. 40(15), pp. 6047-6054, 2013.
- [16] S. Shanthi, and R.G. Ramani, Classification of vehicle collision patterns in road accidents using data mining algorithms. International Journal of Computer Applications, Vol. 35(12), pp. 30-37, 2011.
- [17] P. Tiwari, S. Kumar, and D. Kalitin. Road-user specific analysis of traffic accident using data mining techniques. in International Conference on Computational Intelligence, Communications, and Business Analytics, Springer, 2017.