Survey Paper on Analyze and Predict the Nature of Road Traffic Accidents using Data Mining Techniques in Maharashtra, India

[1]Baye Atnafu, [2]Gagandeep Kaur
[1]*M.Tech student,* [2]*Assistant Professor, Dept of CS/IT, Symbiosis Institute of Technology, Pune, India*

Abstract*—Traffic accidents are the main cause of death as well as serious injuries in the world. India is among the emerging countries where the rate at which traffic accident occurs is more than the critical limit. Due to this reason difficult to know the nature of road traffic accidents. As a human being, we all want to avoid traffic accidents and stay safe. In order to stay safe, careful analysis of roadway traffic accident data is important to identify the nature of traffic accident that causesfatal and series injuries. Analysis of road traffic accidents is significantto expose the association between the various types of factors thatinfluencethe nature of road traffic accidents. For this purpose, there are a number of classification and association rule mining algorithms available to analyze, detect and predict the road accident historical data and to obtained hidden patterns from huge data. From these, this survey paper discusses the algorithms and data mining tools that areproved better in theprevious studies.*

Keywords—*data mining, random tree, J48, Naive Baye's, association rule mining, road accidents*

I. INTRODUCTION

Annually due to road traffic accidents 1.25 million peoples die and 20-50 million peoples hurt non-fatal injuries [1]. According to the road traffic accident data provided by states, Maharashtra records the third highest number of fatal accidents (13,212) [2]. However, this trend can change in future as it is hard to predict the rate at which road traffic accidents occur as it can occur in any situation. Therefore, we need to investigate the hidden pattern that influences the traffic accident severity levels using data mining techniques.

There are a number of Data Mining classification algorithms available (Like a Random tree, J48, Random forest, CART and Naïve Baye's) to predict the target class by analyzing the training dataset to get better boundary conditions which can be used to determine each target class. After determining the boundary conditions, the subsequent task is to predict the target class based on the boundary conditions.

There are also a number of Data Mining algorithms are available to find out the association between independent variables in a huge data. Association rule mining algorithm is the most popular methodologies to detect the significant associations between the data stored in the large database. There are a number of association rule mining algorithms available. From these Apriori, predictive Apriori and FP-growth algorithm are the most common association rule mining methods to find out the association between various road traffic accident severity factors that influencing the traffic accident severity levels in Maharashtra state, India**.**

II. RELATED WORK

Researchers have proposed a variety of data mining techniques, algorithms and toolsfor road traffic accident data analysis and prediction,accident locationtracking, and identification of various contributory factors that influence the accident severity levels. Some of the papers are discussed here.

Ali Tavakoli Kashaniet.al[3] Using the CART classifier, they have been analyzedthe crash data and to detect the most contributory factors which affect injury severity of drivers involved in traffic crashes. The analysis results showed that not using the seat belt, improper overtaking and speeding are the most influential factors associated with injury severity.

Sachin Kumar & Durga Toshniwal [4] in this study initially authors applied three popular classification algorithms such as a CART, Naïve Bayes, and SVM on PTW power two-wheeler accident data set and compared the results. CART classification algorithm accuracy was found superior to other two algorithms. Hence they have been used CART classification algorithms to find the various factors that influence the accident severity of power two-wheeler accidents in entire Uttarakhand state and its 13 districts separately [4]. The result shows that each districthas different factors associated with power two-wheeler accidents severity.

Liu et al. [5] They have been build statistical model using stepwise regression analysis method for estimating incident duration. The analysis result shows that over 85% of differences in incident

duration can be predicted by the eight factors involved in the regression model.

Sachin Kumar & Durga Toshniwal [6] k-means clustering algorithm used to investigate the high and low-frequency accident locations. Further, they have been used association rule mining to recognize the association between the various factors related to road traffic accidents at various places with changeable accident occurrences. The result shows that more accidents occur on highways, foot-travelers are more vulnerable to road accidents at roads that have intersections, Curve on roads bordered by agriculture land are risky for multi-vehicle crashes and intersections on roads which fall upon marketplaces are more vulnerable to severe accidents.

Wang, Yubian, and Wei Zhang [7] logistic regression model used to isolate and enumerate the impact of various roadways and environmental factors on the traffic crash severities and predict the accident severity levels. Authors investigated that factors like crash location, road function class, road alignment, light condition, road surface condition, and speed limit have the significant impact on crash severity. The results show that higher crash severity is linked with rural roadways, major arterials, locations without intersection, locations with curves, during night-time, dry roadway conditions, and high-speed limits.

L. Mussonea, M. Bassani and S P. MascibKeller [8] evaluated the factors that affect the accident severity levels at urban road intersections using back propagation neural network and generalized linear mixed model. Both methods demonstrate that traffic flows have a significant role in predicting severity; this role is not limited to the flow when the crash occurred, but also extends to the other vehicle crash flow data before the crash occurs after the crash occurred.

Yina Wu, Mohamed Abdel-Aty & Jaeyoung Lee [9] logistic regression model used to recognize the various factors contributing to increased vehicle crash risk during fog and investigate the situations in which crash risk are more likely to occur. The analysis results show that drivers will be more careful when fog is present and the chances of increasing crash risk would be more near ramp areas.

Sachin Kumar, Durga Toshniwal, Manoranjan Parida [10] used the Latent Class Clustering and k-Modes clustering algorithms to form different homogeneous clusters using a heterogeneous road accident data. Further, FP growth algorithm is applied to the clusters formed to find out the algorithm that is better-performing when decreasing the heterogeneity of traffic accident data [8]. The results prove that there is no any clustering algorithms is superior to others, that means both the clustering techniques perform well when to reduce the heterogeneity nature of accident data [8].

Yannis George, Theofilatos Athanasios & Pispiringos George [11] investigated road accident severity per vehicle type by using log-normal regression techniques. The result of this study shows that bad weather situations and accidents during nighttime increase accident severity. Furthermore, authors concluded that there is a major impact of crash type while examining accident severity.

Hao, Wei, Camille Kamga, Xianfeng Yang, JiaQi Ma, Ellen Thorson, Ming Zhong, and Chaozhong Wu. [12] explored the factors of injury severity of truck drivers in the United States using an ordered probit regression model. The outcomes of analysis show bad weather and visibility condition enhance the chances of high-level injury severity in truck drivers'.

Bhaven Naik, Li-Wei Tung, Shanshan Zhao, & Aemal J. Khattak [13] authors combined two diverse sources of data and applied random parameters (mixed) ordered logit model to consider the distinct heterogeneity in the data. Results depicted that rain, air temperature, humidity, air temperature, wind speed, and rain were found a factor for injury severity. From these factors, warmer air temperatures and rain were associated with high severe injuries while less severe injuries wereassociated with higher levels of humidity.

Dursun Delen, Leman Tomak, Kazim Topuz & Enes Eryarsoy [14] focused on recognizing the person, vehicle, and accident-related risk factors that are significant in building a variance in injury severity levels sustained by a driver in a car crash. The authors used a number of predictive analytics algorithms to find the composite associations between various stages of injury severity and the risk causes related to crash. The authors also found the importance of crash-related risk factors after applying a systematic series of in-formation fusion-based sensitivity analysis on the trained predictive models. Sensitivity analysis results prove that use of a preventive system (i.e., seatbelt), the way of crash and drug usages are the main predictors of the injury severity.

Liling Li, Sharad Shrestha & Gongzhu Hu [15] applied statistics analysis and data mining algorithms such as, Apriori rule mining, Naive Baye's and k-means clustering algorithm on the FARS Fatal Accident dataset for the purpose of investigating the relationship between fatal rate and other attributes such as weather condition, collision manner, light condition, drunk driver and surface

conditions. The analysis result shows that environmental factors likeroadway surface, weather, and light conditions do not strongly affect the fatal rate, while the human factors like drunk or not and the collision type have a stronger effect on the fatality rate.

Hasan Mehdi Naqvi & Geetam Tiwari [16] using a binomial logistic regression model to identify the various causes that influence the motorcycle fatal crashes. Study results show that the chances to occur rear-end, sideswipe and head-on collision are 42 times, 35 times and 25 times more than hit pedestrian for variable "collision type", respectively; the probability of fatal crash increase insingle-vehicle crashes than two or more vehicles crashes for variable "number of vehicle", and, the probability fatal crash on two-lane national highway is more than four-lane national highway for variable "number of lane".

Table I: Data mining Algorithm's review

Authors	Title	Data Mining Techniques	Algorithm Performance	Objective	Result
Velivela Gopinath et.al[17]	Traffic accidents analysis with respect to road users using data mining techniques	SOM & K-modes SVM Naïve Bayes Decision tree	SOM is better than K-modes 75.5838 % 74.4583 % 75.7599 %	To achieve better accuracy by using clustering techniques	Accident between the driver, passenger or pedestrian are more involved
S.Krishnaveni et.al[18]	Aprospective analysis of traffic accident using data mining techniques	Naive Bayes J48 AdaBoostM1 PART Random Forest	84.66% 84.64 % 84.64% 85.18 % 88.25%	To investigate prediction modelsthat predict the accident severity	Severity prediction of using Random Forest better than other four algorithms.
S. Shanthi et.al[19]	Classification of vehicle collision patterns in road accidents using data mining algorithms	CART C4.5 CS-MC4 Decision List Navie Bayes ID 3 RndTree Rule Induction	80.59% 76.24% 71.09% 67.92% 75.54% 72.28% 94.38% 75.54%	To discover the suitable data mining techniques for mining vehicle collision patterns	The Random Tree classifier outperforms than all the other classifiers
El Tayeb et.al[20]	Applying association rules mining algorithms for traffic accidents in Dubai	Apriori and Predictive Apriori association rule mining algorithms	The rule generated by Apriori algorithm was more effective than Predictive Apriori algorithms	To discover the links between accident factors & accident severity	The Apriori rule mining algorithm generates better rules than Predictive Apriori rules mining algorithms.
Ali et.al[3]	A data mining approach to identify key factors of traffic injury severity	CART	72.49 %	Identify factors which influence the injury severity	Factors identified in this study are not a seat belt, improper overtaking and speed.

DipoT.Akomolafe & Akinbola Olutayo [21]	Using data mining techniques to predict cause of accident and accident-prone locations on highways	Decision tree: ID3 Function tree	77.70 70.30	Predict the various cause of accidents and to identify locations at which the accident occurred frequently.	Decision tree: ID3 have predicted thecause of accidents and accident-prone locations more accurately than function tree.
Liping et.al[22]	Traffic incident duration prediction based on artificial neural network	Artificial Neural network (ANN)	85.35%	To predict traffic incident duration	The predictable outcome of the ANN model can essentially signify the actual incident durations.
Sohn, So Young, and Hyungwon Shin [23]	Pattern recognition for road traffic accident severity in Korea	ANN LR DT	No significant difference in accuracy.	To identify a set of influential factors	Defensive devices are the most significant influence in the accident severity variation.
Sachin Kumar and Durga Toshniwal[24]	A data mining framework to analyze road accident data	Association Rule Mining K-Modes	Using of Association rules for entire data set did not give the better result.	To segment road accident data in order to get better results	investigation proves that carrying out clustering prior toanalysis helps in getting better and useful results than applying association rule mining on EDS.
S. Shanthi and Dr. R. Geetha Ramani [25]	Gender-specific classification of road accident patterns through data mining techniques	Random Tree C4.5	95.59% 83.02%	To improve the accuracy of the weak classifier's using AdaBoost	The AdaBoost used with RndTree improvised the classifier's accuracy
Randa Oqab Mujalli, Griselda López and Laura Garach[26]	Bayes classifiers for imbalanced traffic accidents datasets	Naïve Bayes classifiers Bayesian networks(BN) classifiers	BN improved classification performance when the data is balanced.	To apply data balancing techniques on traffic accident data in order to enhance classifierperformance	Use of balanced accident data has enhanced the capability of Bayes classifiers.
Tarek Sayed et.al[27]	Identifying accident-prone locations using fuzzy pattern recognition	fuzzy K-NN algorithm	Algorithm can work effectively	To investigate locations which are frequently prone to accident.	Due to driver-related factors 96% of accidents involved.

III. METHODOLOGY

1. CLASSIFICATION ALGORITHMS

The classification algorithm is one of the data analysis methodsthat used to constructmodelsto predict the future data. The type of classification algorithms used variesaccording to the target variable. The target variable for this survey paper is represented as a category variable withsix possible outcomes (Overturning, head-oncollision, Rar end collision, Right turn collision, left turn collision and others).Accordingly, the analyzingproblem is characterized as anominal classification problem and based on the extant literature there are various data mining techniques available that can handle this type of classification problem such as aRandom tree, j48, Naïve Bayes…etc.

A. RANDOM TREE

A random tree(RndTree) is a group of distinct decision trees, which means that operator of random tree works just like the decision tree operator except, for each split, only a random subset of attributes is accessible [30]. A RndTreeis a tree haggard at a chance from the collection of achievable trees. In this perspective "at random" signifies that every tree has anequal chance of being selected from the set of trees.

B. J 48

J48 is an advanced version of ID3; it decides target value of a new test data with respect to diverse attribute values of training dataset [29]. The inner nodes of a decision tree are represented by different attributes while the branches tell the achievable values of these attributes. The internal nodes denote the dependent variable values [29]. Escalating the count of trees provides a more intelligent learner just as having a large varied group is capable of reaching intelligent conclusion [31].

c. NAÏVE BAYE'S

The naïve Bayesian classifier is one of the most effective and widely used supervised learning algorithms to classify the road accident data. It is a statistical model that predicts class membership probabilities based on Bayes' theorem. The Naive Baye's classification algorithm is one of the probability-based methods used for classification and prediction based on the Bayes' hypothesis with the assumption of independence between each pair of variables.

2. ASSOCIATION RULE MINING ALGORITHMS

Association rule mining algorithm is the most popular methodologies used to detect the significant associations between the data stored in a huge database. For this purpose there are a number of association rule mining algorithms present, from these Apriori, predictive Apriori and FP-growth association rules mining algorithm are the most unusually used algorithms inthe area of road traffic accident analysis,to generate the best rules that show the association between various attributes in large datasets.

a. APRIORI ALGORITHM

Apriori rule mining algorithm is the naive method of finding the frequent item-setsin a huge database by generate a setof all possible combination of items and then compute the support for them.However, the number of possible combinations increases exponentially as thenumber of items in item-set increases making this method impractical [32].

b. PREDICTIVE APRIORI ALGORITHM

The predictiveApriorialgorithm is also used for discovering hidden and novel patterns in a large database. It variesfrom Apriori algorithm in that both confidence and support measures are joined into a unique measure called as predictive accuracy[33].

c. FP-GROWTH ALGORITHM

Frequent-pattern growth association rule algorithm is the enhanced version of the Apriori rule mining algorithm present by Jiawei Han and so forth [34]. It compresses data sets to an FP-tree, scans the database twice, does not produce the candidate itemsets in therule mining process, and greatly improves the mining efficiency [35]. But FP-Growth algorithm needs to create an FP-tree which contains all the datasets. This FP-tree has high requirement on memory space [36].

From these of various data mining algorithms, we planned to use Random tree classification algorithmto predict the nature of traffic accidents and to identify various factors that influence the nature of accidents. Further,in this study also weplanned to use Apriori association rule mining algorithms to identify the association between various attributes.

3. DATA MINING TOOLS

Data Mining allows discovering novel patterns that are not discovered yet by using various open source data mining tools.Currently, there are many tools are available for data mining, Such as WEKA, RAPID MINER, R, KNIME…etc.From these of various data mining tools,planned to use WEKA toolsbased on the above review to analyze the road accident data and find out the various factors that influence the accident severity levels.

4. DATA PREPROCESSING

The dataset used in this study is obtained from National Highways Authority of India (NHAI) which covers accident historical data from September 2014 to July 2017 [37]. The dataset contains 19,166 accident records and 9 independent variables and 1 dependentvariables after the data is preprocessed. The detail of data set and its attributes with values are given in Table III. In this study planned touse WEKA 3.8 data mining tools from various data mining tools based on the future shows during thereview, for the purpose of classification, prediction, model evaluation, attribute selection, data cleaning, data integrating and managing road accident data obtained from National Highways Authority of India. Figure 1 shows the general block diagram of the proposed work. The first task is data preprocessing which include tasks such as data cleaning, integration,transformation, andreduction. After once the data is preprocessed, thenext step is applying the data mining techniques on the data.

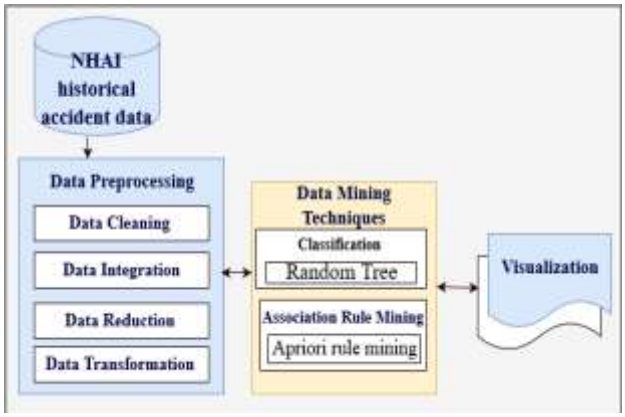

Figure 1: Block Diagram of Proposed work

There are a number of Data mining techniques available but the proposed method uses three Classification algorithms: Random tree, J48 and Naïve Baye's for predictions and Association Rule Mining algorithms to detect the significant associations between the data stored in the large database. After applying these algorithms, the next step is to visualize the outcomes obtained from experiments. The detailed process of the above-mentioned tasks is shown in figure 2.

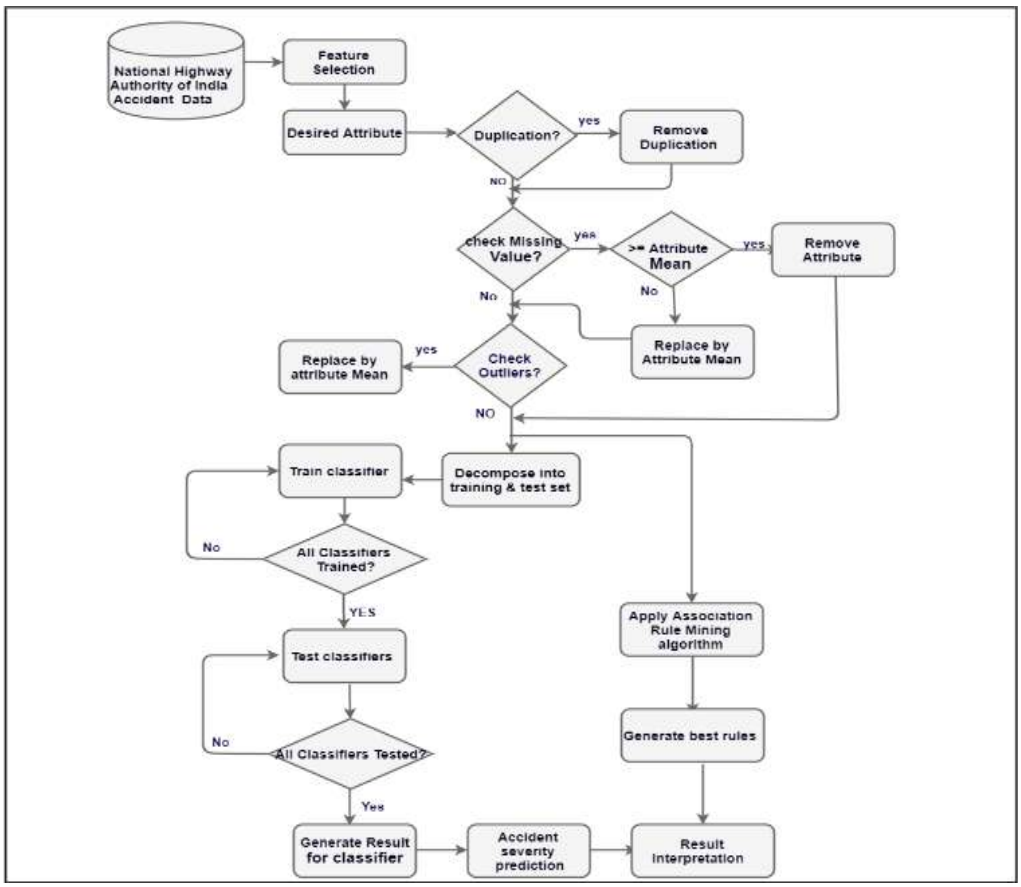

Figure 2: Flowchartdiagram for proposed work

The flow chart shows each step followed throughout the study starting from data collection up to prediction of the road accident severity levels. After collecting the dataset from National Highways Authority of India feature selection method is applied to select the desired attributes. After this, selected desired attributes are checked for duplication, missing values, and outliers. After preprocessing, the dataset is decomposed into two sets: training and testing sets. Next step is applying the classification algorithms on the data set and test whether all classifiers are trained or not. If all the classifiers are trained then test the classifier and generate results. Then nature of accident prediction is done, further, we applied the Apriorirule mining algorithm to discover the relationship between various factors that frequently influence the nature of accidents. Finally, the result is interpreted for both classification and association rule mining techniques.

Table III: Description of data set and its attributes with values.

Attribute	Number of distinct values	Attribute value	Code
Location	Many	749/100 RHS	749/100 RHS
Nature of Accident	7	Overturning	1
		Head on collision	2
		Rear end collision	3
		Collusion brush or side wipe	4
		Left turn collusion	5
		Right turn collusion	6
		skidding	7
Causes of accident	5	Drunk	1
		Over-speeding	2
		Vehicle out of control	3
		Fault of driver of motor vehicle /driver of other vehicle /cyclist /pedestrian /passenger	4
		Defect in mechanical condition of motor vehicle /road condition	5
Road Feature	4	Single lanes	1
		Two lanes	2
		Tree or more lanes without central divider	3
		Four or more lanes with central divider	4
Road Condition	8	Straight road	1
		Slight curve	2
		Sharp curve	3
		Flat road	4
		Gentle incline	5
		Steep incline	6
		Hump	7
		Dip	8
Weather Condition	12	Fine	1
		Mist/fog	2
		Cloud	3
		Light rain	4
		Heavy rain	5
		Hail/sleet	6
		Snow	7
		Strong wind	8
		Dust storm	9
		Very hot	10
		Very cold	11
		Other weather condition	12
Intersection Type and Control	8	T-junction	1
		Y- junction	2
		Four arm Junction	3
		Staggered Junction	4
		Junction with more than four arms	5
		Roundabout junction	6
		Manned rail crossing	7
		Unmanned rail crossing	8
Time of day	6	7:00am - 10:59am	T1
		11:00am - 2:59pm	T2
		3:00pm - 6:59pm	T3
		7:00pm - 10:59pm	T4
		11:00pm - 2:59am	T5
		3:00am - 6:59am	T6
Date	Many	09-09-2015	09-09-2015
Season	4	Jan - Feb	Winter
		Mar-May	Summer
		Jun-Sept	Rainy
		Oct-Dec	Monsoon

IV. CONCLUSION

Paper, Survey on Analysis and Prediction of road Traffic Accident Severity Levels Using Data Mining Techniques in Maharashtra, India discusses the latest work in the field of road accident analysis and prediction. Road traffic accident severity keeps on changing over time and increase endlessly. The changing and increasing road traffic accident severity leads to the issues of not understanding the accident behavior, factors influencing the traffic accident severity, and managing large volumes of data obtained from various sources properly. Many researchers have tried to solve these issues but still, there are gaps in the road accident severity prediction and finding the contributory factors such as seasontime and nature of accidents in which the accident frequently occurred. This leads to the challenges in the field of accident analysis and prediction. Some of the challenges include modeling of accidents for finding suitable algorithms to detect the accident severity levels, data preparation, transformation, and processing time. Therefore, in order to fill some of the gaps, We are motivated to studythe roadtrafficaccident data to find out the factors that influence the nature of road accidents in Maharastra, India. In this survey work, we analyzed latest works, data mining techniques, and tools that were proved better in accident historical dataanalysis and prediction.

V. REFERENCES

1. WHO. Road traffic safety. Available from http://www.who.int/mediacentre/factsheets/fs358/en/ Accessed on 21 September 2017.
2. 400 road deaths per day in India; up 5% to 1.46 lakh in 2015 Available fromhttp://timesofindia.indiatimes.com/lifestyle/health-fitness/health-news/400-road deaths-per-day-in-India-up-5-to-1-46-lakh-in-2015/articleshow/51920988.cms accessed on 21 September 2017.
3. Tavakoli Kashani, A., Shariat-Mohaymany, A., & Ranjbari, A. (2011). A data mining approach to identify key factors of traffic injury severity. PROMET-Traffic&Transportation, 23(1), 11-17.
4. Kumar, S., & Toshniwal, D. (2017). Severity analysis of powered two-wheeler traffic accidents in Uttarakhand, India. European Transport Research Review, 9(2), 24.
5. A. Garib, A.E Radwan and H. Al-Deek. Estimating Magnitudeand Duration of Incident Delays. Journal of Transportation Engineering, vol. 123, Nov./Dec. 1997, pp. 459-465.
6. Kumar, S., & Toshniwal, D. (2016). A data mining approach to characterize road accident locations. Journal of Modern Transportation, 24(1), 62-72.
7. Wang, Y., & Zhang, W. (2017). Analysis of Roadway and Environmental Factors Affecting Traffic Crash Severities. Transportation Research Procedia, 25, 2124-2130.
8. Mussone, L., Bassani, M., & Masci, P. (2017). Analysis of factors affecting the severity of crashes in urban road intersections. Accident Analysis & Prevention, 103, 112-122.
9. Wu, Y., Abdel-Aty, M., & Lee, J. (2017). Crash risk analysis during fog conditions using real-time traffic data. Accident Analysis & Prevention.
10. Sachin Kumar, Durga Toshniwal, Manoranjan Parida. "A comparative analysis of heterogeneity in road accident data using data mining techniques", Evolving Systems, 2016.
11. George, Y., Athanasios, T., & George, P. (2017). Investigation of road accident severity per vehicle type. Transportation Research Procedia, 25, 2081-2088.
12. Hao, W., Kamga, C., Yang, X., Ma, J., Thorson, E., Zhong, M., & Wu, C. (2016). Driver injury severity study for truck involved accidents at highway-rail grade crossings in the United States. Transportation research part F: traffic psychology and behavior, 43, 379-386.
13. Naik, B., Tung, L. W., Zhao, S., & Khattak, A. J. (2016). Weather impacts on single-vehicle truck crash injury severity. Journal of safety research, 58, 57-65.
14. Delen, D., Tomak, L., Topuz, K., & Eryarsoy, E. (2017). Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. Journal of Transport & Health, 4, 118-131.
15. Li, L., Shrestha, S., & Hu, G. (2017, June). Analysis of road traffic fatal accidents using data mining techniques. In Software Engineering Research, Management and Applications (SERA), 2017 IEEE 15th International Conference on (pp. 363-370). IEEE.
16. Naqvi, H. M., & Tiwari, G. (2017). Factors Contributing to Motorcycle Fatal Crashes on National Highways in India. Transportation Research Procedia, 25, 2089-2102.
17. Velivela, G.,K Purna, P.,Challa, Y.,G Krishna, V., &Dr. S Krishna, R. (2017). Traffic Accidents Analysis with respect to Road user using data mining techniques.International Journal of Emerging Trends & Technology in Computer Science, 6(3),2278-6856.
18. Krishnaveni, S., & Hemalatha, M. (2011). A perspective analysis of traffic accident using data mining techniques. International Journal of Computer Applications, 23(7), 40-48.
19. Shanthi, S., & Ramani, R. G. (2011). Classification of vehicle collision patterns in road accidents using data mining algorithms. International Journal of Computer Applications, 35(12), 30-37.
20. El Tayeb, A. A., Pareek, V., & Araar, A. (2015). Applying association rules mining algorithms for traffic accidents in Dubai. International Journal of Soft Computing and Engineering.
21. Akomolafe, D. T., & Olutayo, A. (2012). Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. American Journal of Database Theory and Application, 1(3), 26-38.
22. Guan, L., Liu, W., Yin, X., & Zhang, L. (2010, May). Traffic incident duration prediction based on Artificial Neural Network. In Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on (Vol. 3, pp. 1076-1079). IEEE.
23. Sohn, S. Y., & Shin, H. (2001). Pattern recognition for road traffic accident severity in Korea. Ergonomics, 44(1), 107-117.
24. Kumar, S., & Toshniwal, D. (2015). A data mining framework to analyze road accident data. Journal of Big Data, 2(1), 26.
25. Shanthi, S., & Ramani, R. G. (2012, March). Gender-specific classification of road accident patterns through data mining techniques. In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on (pp. 359-365). IEEE.
26. Mujalli, R. O., López, G., & Garach, L. (2016). Bayes classifiers for imbalanced traffic accidents datasets. Accident Analysis & Prevention, 88, 37-51.
27. Sayed, T., Abdelwahab, W., & Navin, F. (1995). Identifying accident-prone locations using fuzzy pattern recognition. Journal of Transportation Engineering, 121(4), 352-358.
28. Classification, prediction, and clustering. Available from http://shareengineer.blogspot.in/2012/09/classification-and-clustering.htmlaccessedAccessed on 22 September 2017.
29. Sewaiwar, Purva, and Kamal Kant Verma. "Comparative Study of Various Decision Tree Classification Algorithm Using WEKA." *International Journal of Emerging Research in Management &Technology* 4 (2015): 2278-9359.
30. Rapidminer.Randomtree. Available from https://docs.rapidminer.com/studio/operators/modeling/predictive/trees/random_tree.htmlaccessed 22 September 2017.

31. Zhao, Yongheng, and Yanxia Zhang. "Comparison of decision tree methods for finding active objects." Advances in Space Research 41.12 (2008): 1955-1959.
32. Lai, K., & Cerpa, N. (2001, October). Support vs. confidence in association rule algorithms. In Proceedings of the OPTIMA Conference, Curicó.
33. Aher, S. B., & Lobo, L. M. R. J. (2012). A comparative study of association rule algorithms for course recommender system in e-learning. *International Journal of Computer Applications*, *39*(1), 48-52.
34. L. Zhichun and Y. Fengxin, "An improved frequent pattern tree growth algorithm," Applied Science and Technology, vol. 35, no. 6, pp. 47–51, 2008.
35. C. Jun and G. Li, "An improved FP-growth algorithm based on item head table node," Information Technology, vol. 12, pp. 34–35, 2013.
36. B. Zheng and J. Li, "An improved algorithm based on FP-growth," Journal of Pingdingshan Institute of Technology, vol. 17, no. 4, pp. 9–12, 2008.
37. National Highway Authority of India.Regional Office – Mumbai. Available from http://nhai.org.inaccessed on July 30, 2017.