Trend Analysis of Casualties in Road Catastrophes

Namita Gupta[1], Dr. Dinesh Kumar Saini[2]
[1]*Research Scholar, Mahatma Jyoti Rao Phoole University*
[2]*Associate Professor, Sohar University*

Abstract-Road traffic catastrophes are a foremost public health apprehension, resulting over 1.2 million demises and between 20 and 50 million non-fatal injuries worldwide each year. Fatalities, injuries resulting from traffic catastrophes are considered one of the key concerns worldwide, particularly their straight and stern impacts. The foremost objective of analysing accident data is to identify the major aspects associated with road accidents and subsequently assist in the direction of decreasing their occurrence and frequency.

Keywords – *Road accident casualties; Data analysis; Trend analysis; Data Mining*

I. INTRODUCTION

Research based on comprehensive analysis of the causes of accidents and design of appropriate engineering solution has a very important role to play in the field of accident reduction and prevention. Scientific investigations and implementation of commensurate technical measures are contingent upon the availability of ample information on accidents which includes data on vehicle, roadway, environment, users, and victims as well [1]. Advanced data analysis system has the potential to take advantage of the available accident data. Good structured data will construct environment for deeper investigation, aiding in the formulation of evidence-based research on road safety and enabling improved road safety intercession as well as performance monitoring. The system will utilize the road accident database as the source of intellect, to determine accident causation and provide a clearer picture of the issues and potential intervention to perk up the road safety conditions. Data mining is one such approach that focuses on searching for newfangled and exciting hypotheses than conforming the present ones. It includes various tools, techniques and applications that can be applied to eliminate the road accident data related deficiencies as well as statistical limitations. Therefore, it has been utilized for finding unsuspected facts which are yet to be recognized especially in the field of road safety.

II. ORGANIZATION OF THE PAPER

The paper is organized as follows. Section III provides the summary of related work in this area. In section IV we propose a system model and discuss the preparation of the data for the purpose of its analysis. The experimental evaluation and observations are discussed in sections V and VI respectively. The conclusion is presented in section VII. Section VIII identifies the gaps and further scope of research in this field. Finally, section IX lists the references used in this study.

III. LITERATURE SURVEY

Beshah and Hill [2] gave new insights related to road accidents in Ethiopia. These insights provide valuable help in developing methods to improve road safety, particularly in the phase of choosing appropriate means and budget allocations of resources. Considering the size of the accident data set, applying data mining techniques to model RTA data records can help to reveal how the drivers' behaviour and roadway and weather conditions are causally connected with different injury severities.

Getnet [3] investigated the prospective appliance of data mining tools to develop models supporting the recognition and prophecy of major driver and vehicle risk factors that cause road traffic accidents. The research used the WEKA version 3-5-8 tool to construct the decision tree (using the J48 algorithm) and rule induction (using PART algorithm) techniques. Performance of the J48 algorithm was somewhat better than that of the PART algorithm. The license grade, vehicle service year, vehicle type, and experience were acknowledged as the most significant variables for envisaging accident severity.

Chang and Chen [4] analyzed road accident data of one year on National Freeway 1 in Taiwan and developed a CART model and a negative binomial regression model to establish the empirical relationship between traffic accidents and highway geometric values, environmental factors and traffic characteristics. Abellan et al. [5] analyzed two lane rural highway data of Granada province in Spain, using decision rules extracted from decision tree method.

IV. PROPOSED FRAMEWORK

To analyze the data, we develop a framework as shown in Fig. 1. The detailed description of the framework is as follows:

A. Data Set description

The accident training data set used in our study is obtained from the Department for Transport and is called "Reported Road Casualties Great Britain" [6]. We used the road accident information of the year 2017. This data set provides information about road traffic collisions that involve at least one personal injury occurring on the public highway as reported to the police. Damage only collisions where no personal injury occurred, were not included in the statistics.

B. Data pre-processing

Data pre-processing [7] is a very important step in data mining. Data pre-processing mainly deals with removal of noise, handle missing values, removing irrelevant attributes in order to make the data ready for analysis. In this step, our aim is to pre-process the accident data in order to make it appropriate for the analysis. The data set consists of 32567 road accidents in 2017 year in London. After pre-processing of the data, following variables were identified and found satisfactory for further research.

"AREFNO", "Borough.x", "Boro.x", "Easting.x", "Northing.x", "CREFNO", "Casualty.Class","Casualty.Sex", "Casualty.Age..Banded.", "Casualty.Age", "No..of.Casualties", "Casualty.Severity", "Ped..Location", "Ped..Movement", "Mode.of.Travel" , "Borough.y" "Boro.y" , "Easting.y", "Northing.y", "Accident.Severity", "No..of.Casualties.in.Acc." , "No..of.Vehicles.in.Acc.", "Accident.Date" "Day" , "Time" , "Highway", "Road.Class.1" , "Road.No..1", "Road.Type", "Speed.Limit", "Junction.Detail", "Junction.Control", "Road.Class.2", "Road.No..2", "Ped..Crossing.Decoded", "Light.Conditions..Banded.", "Weather", "Road.Surface" , "Special.Conditions", "C.W.Hazard"

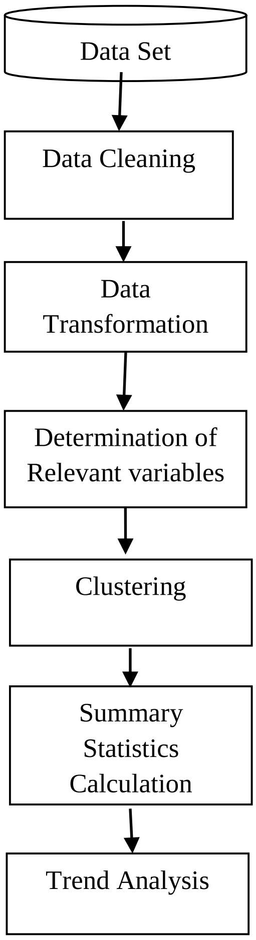

Figure 1: Flow diagram of the proposed approach

C. Data Transformation

It converts the data into appropriate forms for mining. As the data set contains multiple files, we merge them into single large file containing all attributes. We re-label some of the categories available in raw data so that they can be used for efficient analysis.

D. Determination of relevant variables

The data set comprised of numerous accident characteristics such as time, day, month, type of accident, number of injured victims, victims' age and gender, road type, road features, and area around the accident site. A brief information about this data is given in Table 1.

TABLE 1. RELEVANT VARIABLES

Casualty mode of travel	Pedestrian, Pedal Cycle, Powered 2 Wheeler, Car, Taxi, Bus Or Coach, Goods Vehicle, Other Vehicle
Casualty severity	Fatal, Serious, Slight
Casualty sex	Male, Female, Unknown
Age Band	"0-4","5-9","10-14","15-19","20-24","25-29","30-34","35-39","40-44","45-49","50-54","55-59","60-64","65-69","70-74","75-79","80-84"
Light conditions	Dark, Daylight
Day	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday
Month	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec

V. EXPERIMENTAL EVALUATION

We made use of the R Language (version 3.5.2) using RStudio Version 1.1.456 in windows platform. We used ggvis and dplyr packages for creating interactive graphs. We implemented our proposed approach and used the available accident data set for trend analysis. We performed a trend analysis on the identified relevant variables. Figures 2 and 3 illustrate the month wise and day wise trends respectively, for the year 2017.

Figures 2 to 17 which follow now, illustrate the trends evaluated experimentally utilising R software, taking into account relations between variables. Some of the results based on these trends are discussed in detail in section VI, later in this paper.

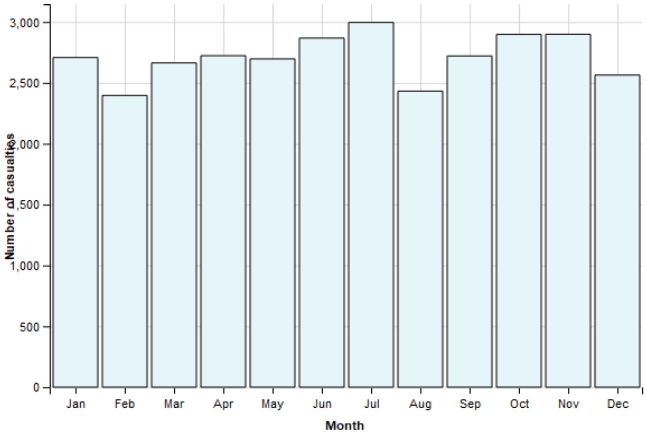

Figure 2: Month wise Accident analysis

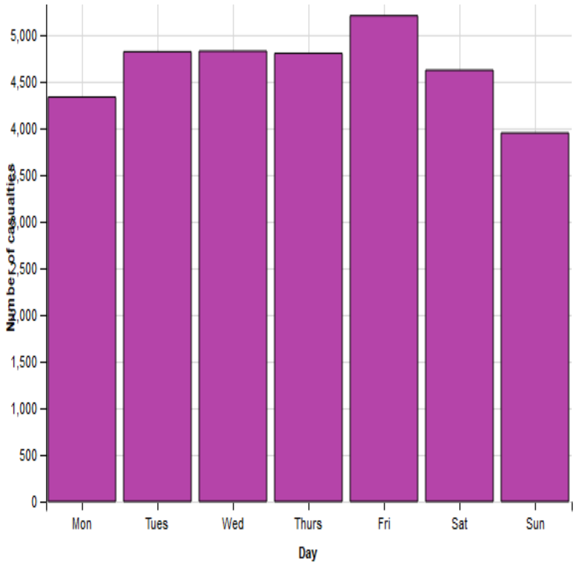

Figure 3: Day wise Accident analysis

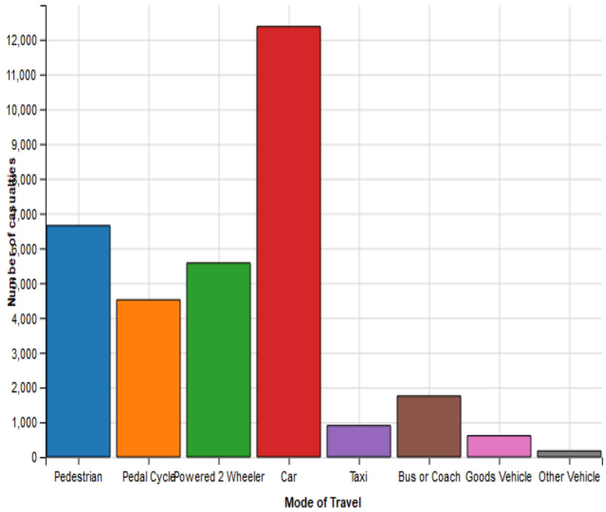

Figure 4: Mode of travel and casualties

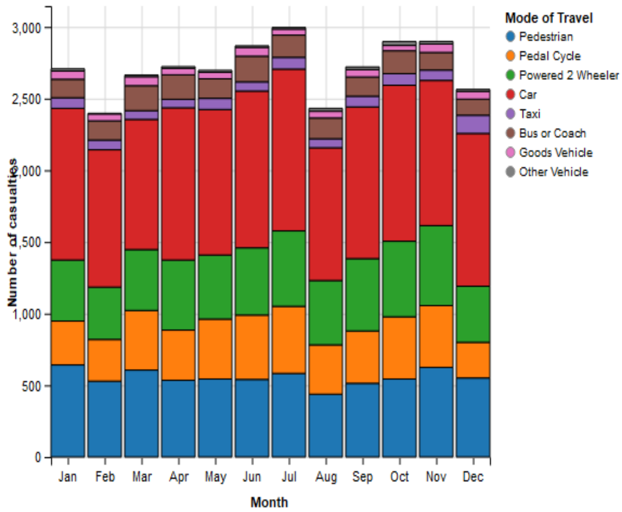

Figure 5: Month wise casualties and mode of travel

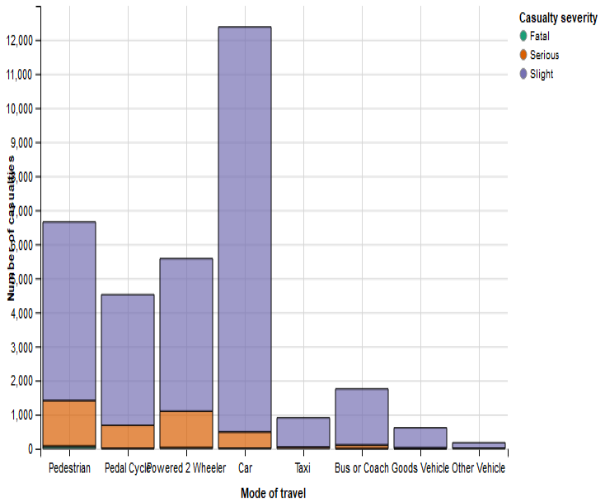

Figure 6: Casualty severity and mode of travel

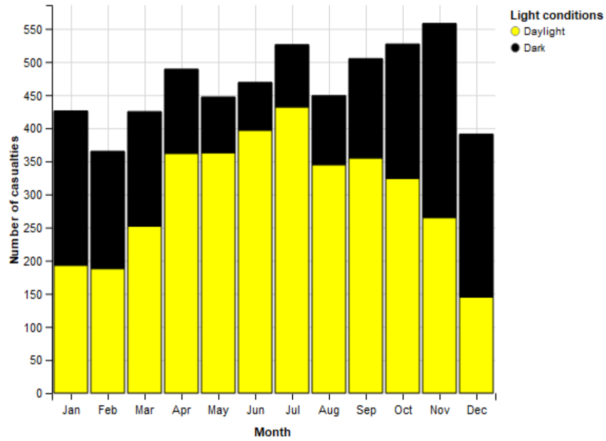

Figure 7: Powered two-wheeler casualties by month and light conditions

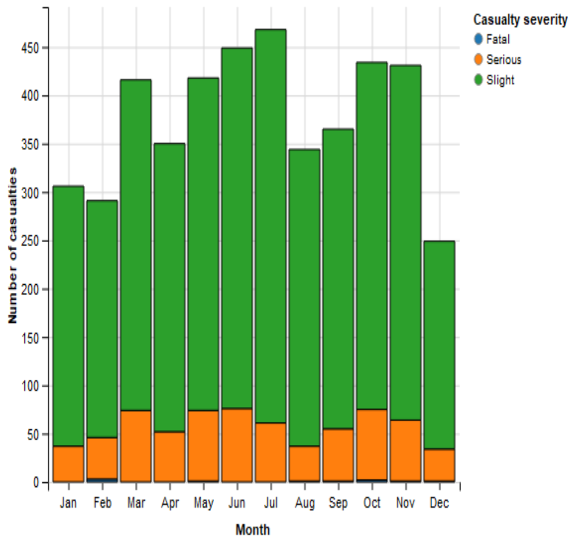

Figure 8: Month wise casualties and severity in the case of mode of travel pedal cycle

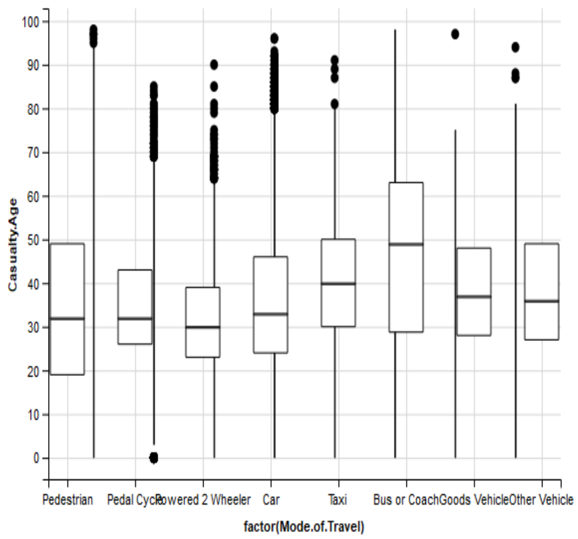

Figure 11: Mean age of casualties by mode of travel

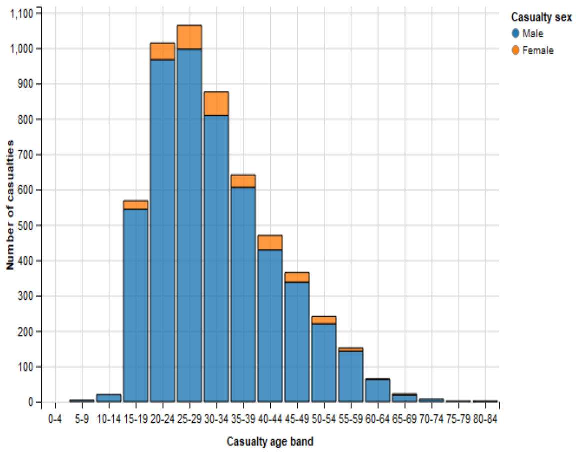

Figure 9: Powered two-wheeler casualties by age band and gender

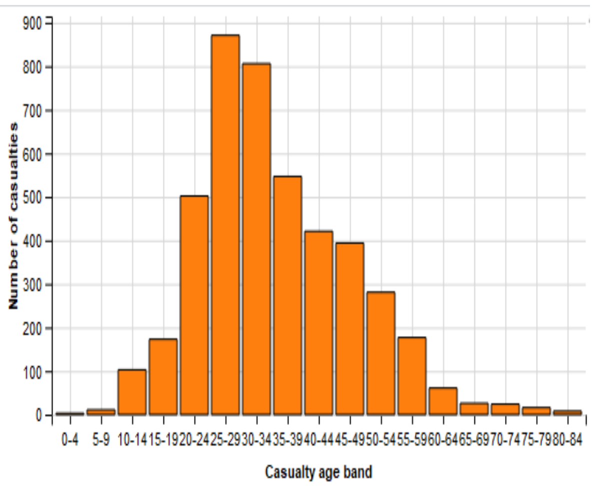

Figure 12: Pedal Cyclists casualties by age band

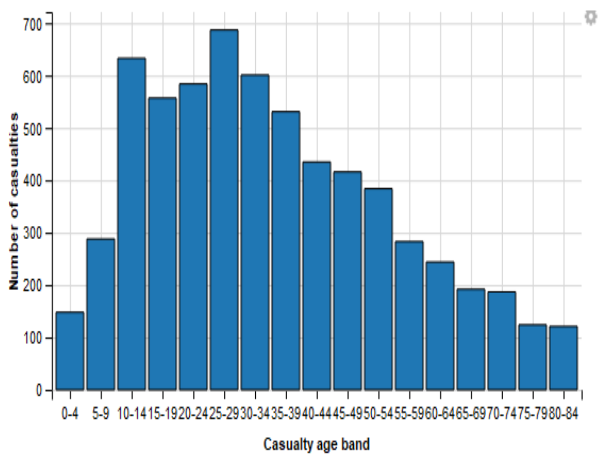

Figure 10: Pedestrian casualties by age band

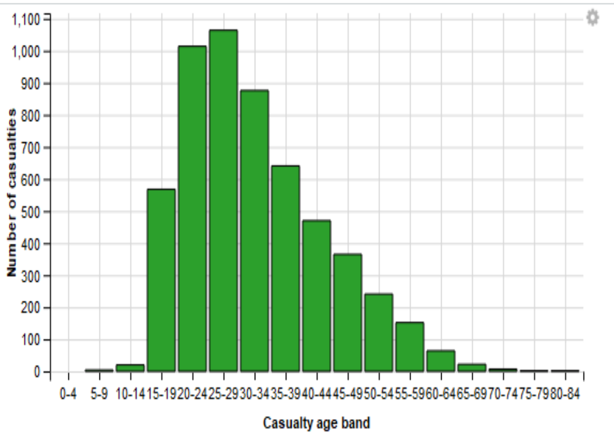

Figure 13: Powered two-wheeler casualties by age band

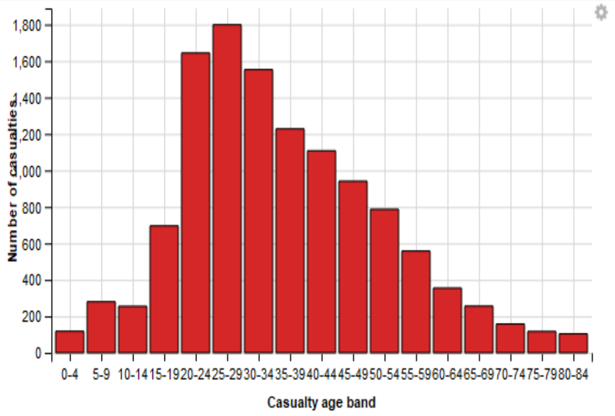

Figure 14: Casualties by age band when mode of travel is car

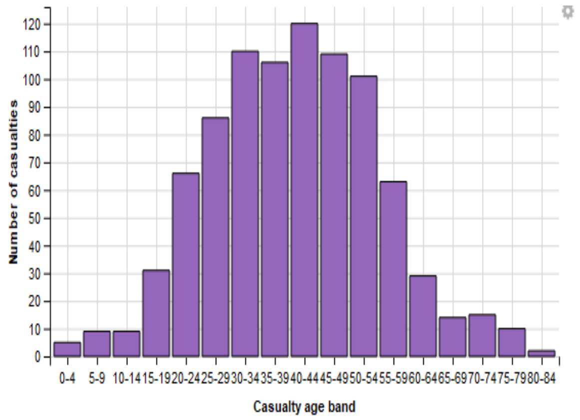

Figure 15: Casualties by age band when mode of travel is taxi

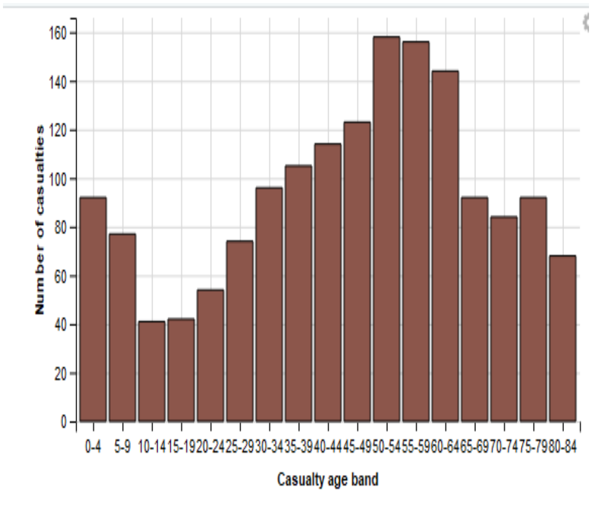

Figure 16: Casualties by age band when mode of travel is bus

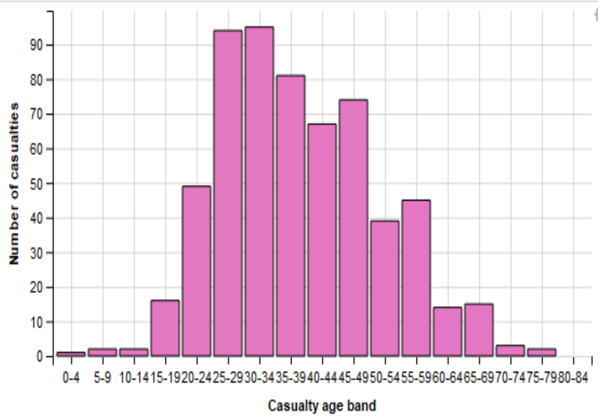

Figure 17: Casualties by age band when mode of travel is goods vehicle

VI. RESULTS AND DISCUSSION

Figures 2 to 17 illustrate diverse accident patterns in the year 2017 in Greater London. From Figures 2, 3 and 5, it is very clear that there is no great impact of month and day on the occurrence of accidents. Figures 4 and 5 clearly state that most of the accidents occurred when the mode of travel was car, but from Fig. 6 we can conclude that most of the car accidents were not fatal. Most of the serious causalities occurred with pedestrians. Fig. 7 illustrates that most of the accidents occurred in day light. Out of total accidents, 21618 accidents occurred in day light while 10949 accidents occurred at night. Fig. 8 shows that the number of fatal accidents on the roads of London is 131 during the year 2017.

Fig. 9 clearly depicts that most of the accident victims in case of powered two-wheeler were male. Out of 32767 accident victims, 20811 were male. We can also conclude from Fig. 9 that most of the victims' age was between 20 to 34 years. Out of total accident victims, 3929 victims were between the age of 20 to 24 years, 4709 victims were between the age of 25 to 29 years, and 4155 victims were between the age of 30 to 44 years.

Figures 9 to 17 depict the number of causalities by age band with different modes of travel. From Fig. 11 we can conclude that mean age of casualties as per the mode of travel is as given in Table 2 which follows.

TABLE 2: MEAN AGE OF CASUALTY AS PER MODE OF TRAVEL

S. No.	Mode of travel	Mean age of casualty
1	Pedestrian	40
2	Pedal Cycle	36
3	Powered 2 Wheeler	32
4	Car	39
5	Taxi	44
6	Bus or Coach	56
7	Goods Vehicle	38
8	Other Vehicles	37

VII. CONCLUSION

The objective of this research undertaking was to explore the possible application of data mining technology for mining vehicle collision patterns in road accident training data set. In this paper, we proposed a framework for analyzing road accident data and establish patterns and relations between various factors attributing to road accidents.

VIII. GAPS AND FURTHER SCOPE OF RESEARCH

Few gaps were identified in this area of research. One of them is the availability and authenticity of road accident data. Further, several cases of accidents are not reported with the police department of that area and hence are not included in the statistics. Thus, missing values, false values, wrong entries, and hardships in retrieving raw accident data from the police and government are some of the hindrances in this area of research.

There is a lot of scope of research after this step of trend analysis. The ultimate goal of research in this field is accident reduction and prevention. Thus, this study will be extended and further progress will be made by identifying the major causes of accidents and factors influencing them. Following which, predictions can be made based on this study by making use of data mining techniques.

IX. REFERENCES

[1] M. J. B. Alam, D. M. Karim, M. M Hoque, Q. Z. Islam, and M. D. Alam, "Initiatives regarding road accident database in Bangladesh," International Conference on Road Safety in Developing Countries, pp. 253-258, 2006.

[2] 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: Artificial Intelligence for Development, 2010.

[3] M. Getnet, "Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents: The case of Addis Ababa city", Addis Ababa University, 2009.

[4] LY. Chang, WC. Chen, "Data mining of tree based models to analyze freeway accident frequency", J Safety Res, Elsevier: 36(4) 365-75, 2005.

[5] J. Abellan, G. Lopez, J. Ona, "Analysis of traffic accident severity using decision rules via decision trees", Vol. 40, Expert System with Applications: Elsevier, 2013.

[6] Reported Road Casualties, Great Britain Annual Report, 2017.

[7] PN. Tan, M. Steinbach, V. Kumar, "Introduction to data mining", Pearson Addison-Wesley, 2006.

AUTHOR PROFILE

Name: Namita Gupta, Lecturer, Mathematics and IT.
Education: MSc.(Maths), MCA.
Publications: A Scopus indexed Conference paper, an international journal paper, an international conference paper, authored chapters.
Research area: Data mining.

Name: Dr. Dinesh Kumar Saini, Associate Professor, supervising PhD thesis.
Publications: About 50 indexed journal papers.
Research area: Emerging computing.