Indian Road Accident Analysis with Machine Learning

Chaturvedi Venugopal, Ajay Shanker Singh and Shruti Suman

April 11, 2023

Indian Road Accident Analysis with Machine Learning

Chaturvedi Venugopal
School of Computing Science & Engineering
Galgotias University
Greater Noida, India
venuchoubey224@gmail.com

Dr Ajay Shanker Singh
School of Computing Science & Engineering
Galgotias University
Greater Noida, India
ajay.shankersingh@galgotiasuniversity.edu.in

Shruti Suman
School of Computing Science & Engineering
Galgotias University
Greater Noida, India
shrutisuman06122@gmail.com

***Abstract*— Road accidents are a major threat to both developed and underdeveloped countries. Traffic accidents and their safety is a major problem of the world and everyone has been trying to deal with it for years. Road traffic and reckless driving occur in all parts of the world. For this reason, many pedestrians are also affected. They become victims through no fault of their own. Many road accidents occur due to many factors such as atmospheric changes, sharp turns and human errors. Injuries caused by traffic accidents are large, but sometimes imperceptible, which later affects your health. This study aims to analyze traffic accidents in one of the popular metropolitan cities i.e. Bengaluru using k-means algorithm and machine learning by investigating the accident prone areas or hotspots and their root causes.**

Keywords- underdeveloped, reckless, pedestrians, victims, atmospheric, imperceptible, metropolitan, hotspots, k-means algorithm

The number of traffic accidents in the world is growing at an alarming rate. According to the latest traffic accident report, the number deaths due to traffic accidents in India reached 151,000 in 2018 alone. The report also stated that the reasons for the road death was speeding, driving on the wrong side of the road, using mobile phones, driving while intoxicated, not wearing helmets, not using seat belts and overloaded vehicles. National Motorways and national roads accounted for 30.2% and 25.2% respectively from the total number of traffic accidents and the total number of deaths were 35.7% and 26.7%. The World Health Organization reported that nearly 11%. deaths due to accidents reported by India in 2018. Road accidents are the leading cause of death in India. According to Report on the global state of road safety outlined by World Health. More than 1.35 million people were killed by the organization in 2018 year and nearly 3,700 people are killed worldwide every day traffic accidents involving cars, motorcycles, bicycles, buses, pedestrians or tucks. India is ranked among 199 countries number one in accident deaths. The analysis of road accidents in recent years has produced a considerable number of results pay attention to researchers to identify accident prone factors. The aim is to analyze and predict the accident, susceptible areas using machine learning and data mining techniques. In this model, accident-prone areas are determined by machine learning and deep learning techniques. These approaches are K-Means and Apriori. This model classify accident prone areas into two categories – high and low. The risk of death in India is not only four times that of some developed countries such as the UK and Sweden, but it is also rising rapidly. It also found that the distribution of deaths and injuries in accidents varied by age, sex, month and time. Among people of all age groups, economically active people aged 30-59 are the most vulnerable. However, if we compare fatality and accidents by gender, we find that men accounted for 85.2% of all fatalities and 82.1% of all injuries in 2013. In addition, the number of traffic accidents is relatively higher in May to June and December to January, which shows that extreme weather affects the occurrence of traffic accidents. Accidents remain relatively constant and high during 9:00 - 21:00 and variable but low during the midnight and early morning hours of the day. However, this does not mean that driving during the day is more risky than driving at night. The study also tries to find out the causal distribution of traffic accidents. There are several factors responsible for accidents, but driver error is considered the most important; driver fault accounted for 78% of total accidents in 2013. An analysis of road accidents across states and union territories shows that in 2013, three states and union territories, Tamil Nadu (22.8), Haryana (17.2) and Andhra Pradesh (16 .9) faced a 50% higher risk of death than the all-India average (11.2). The burden of road traffic accidents in India is found to be relatively low in its metropolitan cities (million plus cities), although the mortality risk ranges from 3.0 deaths per 100,000 people in Kolkata to 25.5 deaths per 100,000 people in Jaipur. From 2003 to 2013, the risk of death increased faster in 6 of the 21 sampled metropolitan cities than in the country. Ahmedabad faced the highest increase in risk of death (0.6 to 4.2), followed by Varanasi (9.5 to 17.9), Patna (9.2 to 17.4), Chennai (8.8 to 14.3), Jaipur (15.9 to 25.5) and Vishakhapatnam (125.5) . The study is divided into six sections. Section 2 presents an analysis of the road accident scenario at the national level. This section provides a detailed analysis of accident deaths and injuries along with their causes. Section 3 presents a comparison of road accident problems across states and union territories. Part 4 deals with road accident problems faced by metropolitan cities. Section 5 discusses the way forward to overcome these challenges. The conclusion of the study is presented in the last section.

I. LITERATURE SURVEY

Investigative models are divided into two classifications: insightful or logic models, which seek to understand and evaluate the probability of an accident, and improvement strategies, which focus on reducing the risk of an accident through course/route determination and planning of rest breaks. Their work introduced freely accessible information sources and clear scientific methods (information outline, representation and aspect reduction) that can be used to achieve safer management and code entry for working with information assortment/investigation by experts/analysts. The paper also evaluated the factual and AI models used to display crash risk.

Classified enhancement and prescriptive logic models that focus on accident risk reduction. Basically examined current writing on different spatial methodologies that remember the aspect of space for its different angles in their investigation of street safety. Identified shortcomings in road accident research that include: limited scope of datasets, reliance on broad information arrangements, and lack of relevance for enduring purposes.

The work proposed an information classification procedure with a deep brain network algorithm called Deep Accident Prediction (DAP).

The results show critical upgrades to predict interesting accidents. Contributed a data framework that shows how accidents focus on the natural mode of the electronic landscape map for Russian road accidents to help improve the RTA exam.

Analyzed the severity of accidents that occurred in the United States using a strategic relapse model. They detailed the execution of 40% and 98%, for awareness and uniqueness, respectively. In addition, they distinguished the main indicators of the level of accidents: change of speed, use of seat belts and safety measures.

Artificial Neural Networks (ANN) is an algorithm that is used as one of the information mining tools and non-parametric methods where specialists have analyzed the severity of accidents and injuries among those involved in such accidents. Used ANNs to show associations between injury severity levels and crash-related factors. They used American information about the crash with 16 attributors.

II. NATIONAL LEVEL ANALYSIS OF ROAD ACCIDENTS

A. Road accidental deaths and injuries in India

Over the years, there has been an alarming rise in accidental deaths on Indian roads. The number of traffic accident deaths has increased more than 9 times, from 14,500 in 1970 to 137,400 in 2013.

Compared to 2003, deaths and injuries are 53,000 and 87,000 higher in 2013 (see Table 1). From 2003 to 2013, the death rate increased by 5% per year, while the country's population grew by only 1.4% per year. As a result, the risk of mortality, road accident deaths per 100,000 people, increased from 7.9 in 2003 to 11.2 in 2013.

Despite low levels of motorization, India faces a very high risk of mortality compared to developed countries (see table 2).

The risk of death in India is four times that of the UK and Sweden and almost double that of Japan and Germany. Although the fatality rate, the number of deaths caused by road accidents per 10,000 vehicles, has decreased over the years from 87.5 in 1970 to 8.6 in 2013, it is still relatively high compared to developed countries.

The fatality rate in many developed countries is less than 1 death per 10,000 vehicles (see Table 2).

Table 1: Driver injury severity distribution

Factor	No Injury	Pos injury	Non-incapacitating	Incapacitating	Fatal	Total
Age						
0 (24&under)	1629(52.80%)	608(19.71%)	505(16.37%)	307(9.95%)	36(1.17%)	3085
1 (25-64)	3171(49.88%)	1362(21.43%)	1075(16.91%)	654(10.29%)	95(1.49%)	6357
2 (65+)	373(46.11%)	168(20.77%)	143(17.68%)	96(11.87%)	29(3.58%)	809
Gender						
0 (Female)	1749(41.95%)	1072(25.71%)	778(18.66%)	507(12.16%)	63(1.51%)	4169
1 (Male)	3424(56.30%)	1066(17.53%)	945(15.54%)	550(9.04%)	97(1.59%)	6082
Eject						
0 (No Eject)	5171(50.55%)	2137(20.89%)	1719(16.80%)	1047(10.23%)	156(1.52%)	10230
1 (Eject)	2(9.52%)	1(4.76%)	4(19.05%)	10(47.62%)	4(19.05%)	21
Alcohol						
0 (No Alcohol)	4997(51.35%)	2067(21.24%)	1600(16.44%)	935(9.61%)	133(1.37%)	9732
1 (Alcohol)	176(33.91%)	71(13.68%)	123(23.70%)	122(23.51%)	27(5.20%)	519
Restraining System						
0 (Not Used)	337(27.44%)	193(15.72%)	336(27.36%)	283(23.05%)	79(6.43%)	1228
1 (Used)	4836(53.60%)	1945(21.56%)	1387(15.37%)	774(8.58%)	81(0.90%)	9023
Body Type						
0 (cars)	3408(47.49%)	1600(22.30%)	1272(17.73%)	780(10.87%)	116(1.62%)	7176
1 (SUV &Van)	747(56.59%)	259(19.62%)	189(14.32%)	111(8.41%)	14(1.06%)	1320
2 (Truck)	1018(58.01%)	279(15.90%)	262(14.93%)	166(9.46%)	30(1.71%)	1755
Vehicle Role						
1 (Striking)	4742(49.86%)	2011(21.15%)	1636(17.20%)	970(10.20%)	151(1.59%)	9510
2 (Struck)	261(72.70%)	54(15.04%)	29(8.08%)	15(4.18%)	0(0%)	359
3 (Both)	170(44.50%)	73(19.11%)	58(15.18%)	72(18.85%)	9(2.36%)	382
Rollover						
0 (No-rollover)	5069(50.78%)	2123(20.85%)	1699(16.69%)	1037(10.19%)	152(1.49%)	10180
1 (Rollover)	4(5.63%)	15(21.13%)	24(33.80%)	20(28.17%)	8(11.27%)	71
Road Surface Condition						
0 (Dry)	3467(49.97%)	1404(20.24%)	1190(17.15%)	750(10.81%)	127(1.83%)	6938
1 (Slippery)	1706(51.49%)	734(22.16%)	533 (16.09%)	307(9.27%)	33(1.00%)	3313
Light Condition						
0 (Daylight)	3613(51.18%)	1487(21.06%)	1174(16.63%)	688(9.75%)	98(1.39%)	7060
1(Partial dark)	1139(52.71%)	465(21.52%)	348(16.10%)	186(8.61%)	23(1.06%)	2161
2 (Dark)	421(40.87%)	186(18.06%)	201(19.51%)	183(17.77%)	39(3.79%)	1030

Table 2. Comparison of International Fatality Rates

Country	Motorization rate (no. of vehicles per 1,000 people)	Fatality rate (no. of fatalities per 10,000 vehicles)	Fatality risk (no. of fatalities per 100,000 people)
India (2013)	130	8.6	11.2
Germany (2012)	657	0.67	4.4
Japan (2012)	651	0.63	4
New Zealand (2012)	733	0.95	6.9
Sweden (2012)	599	0.50	3.0
United Kingdom (2012)	559	0.51	2.8
United States of America (2012)	846	1.26	10.7

B. Age- and sex-wise distribution of road accidental deaths and injuries

Table 3 shows the distribution of deaths by age. This chart clearly shows that the most productive age group, 30-44 years, is the most prone to fatal road accidents in India. The 30-44 age group constitutes only 20% of India's population but faces nearly 35% of the total traffic fatalities. During the last ten years from 2003 to 2013, the number of fatal accidents in this age group also increased substantially from 29,156 (34.5% of total deaths) to 47,838 (34.8% of total deaths). The middle age group of 45-59 years is also very prone to fatal traffic accidents. This age group constitutes only 12% of the total population but faces almost 21% of total deaths. The 30-59 age group, the economically active age group, is therefore the most vulnerable population group in India. More than half of road traffic deaths are caused by this population group, which makes up less than one third of the total population.

This may be because people in this age group are in their prime working years and are therefore more likely to be on the roads. The share of deaths in the age groups 15-29 years and 60 years and over is similar to their representation in the total population. Table 4 shows the gender distribution of accident deaths and injuries in India for 2003 and 2013. This table shows that in 2013, males accounted for 85.2% of all deaths and 81.1% of all injuries. Over the last ten years, the number of fatal accidents for men has increased by 64.6%, from 71,128 in 2003 to 117,055 in 2013. This is significantly more than the increase in the death rate for women; the number of fatal injuries to women increased by 53.1% from 13,302 in 2003 to 20,368 in 2013.

However, the trend for injuries is exactly the opposite of that for fatalities. Over the past ten years, male injuries have increased by 21.8%, from 313,055 in 2003 to 381,228 in 2013. This is relatively less than the increase in female injuries; the number of injuries faced by women increased by 26.9%, from 69,843 in 2003 to 88,654 in 2013.

Table 3. Age Wise Distribution of Road Accidental Deaths in India

Age group	No. of Fatalities (2003)	Percentage Share (2003)	No. of Fatalities (2013)
Up to 14 years	6534	7.7	7305
15-29 years	25223	29.9	42453
30-44 years	29156	34.5	47838
45-59 years	16674	19.7	28263
60 years and above	6843	8.1	11564
Total	84430	100	137423

C. *Causes of road accidents*

Figure 3 presents cause-wise distribution of road accidents in India in 2013. It clearly shows that drivers' fault is the single most important factor responsible for accidents. Drivers' fault accounted for 78% of total accidents, 76.5% of total injuries and 73.7% of total fatalities in 2013. Within the category of drivers' fault, accidents caused due to exceeding lawful speed accounted for a high share of 55.6%.

As a share of total accidents and deaths due to drivers' fault, intake of alcohol and drugs accounted for 5.3% and 6.4%, respectively. As a share of total road accidents and deaths, overloading / overcrowding of vehicles accounted for 19.6% and 22.8%, respectively. The fault of cyclists and pedestrians appears to be marginal; they account only 1.2% and 2.2% of total accidents, respectively. The accidents caused due to defects in motor vehicle condition and road condition is also negligible in comparison to drivers' fault. They accounted only 1.8% and 0.8% of total road accidents, respectively.

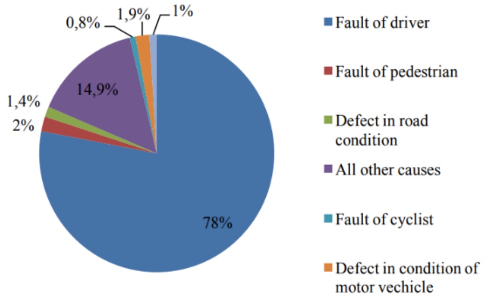


Figure 1. Causes of Road Accidents in 2019

III. Analysis of Road Accident Scenario at City Level

India's road traffic accident burden is marginally lower in its metropolitan cities (more than a million cities). On average, the mortality risk in metropolitan cities is 9.9 deaths per 100,000 people, slightly lower than the all-India average of 11.2 deaths per 100,000 people.

However, there are huge differences in the risk of death between Indian cities, from 3.0 deaths per 100,000 people in Kolkata to 25.5 deaths per 100,000 people in Jaipur in 2013 (see Figure 6). In the same year, Jaipur (25.5), Kanpur (22.3), Vishakhapatnam (22.0), Varanasi (17.9), Lucknow (17.7) and Patna (17.4) faced more than 50% a higher risk of death than the metropolitan city average (9.9).).

From 2003 to 2013, the risk of death increased faster in 6 of the 21 sampled metropolitan cities than in the country. Ahmedabad faced the highest increase in risk of death (0.6 to 4.2), followed by Varanasi (9.5 to 17.9), Patna (9.2 to 17.4), Chennai (8.8 to 14.3), Jaipur (15.9 to 25.5) and Vishakhapatnam (225.5) . However, there are eight cities that saw a decrease in the risk of death between 2003 and 2013; of these, Bengaluru saw the highest decline from 15.5 to 8.8 deaths per 100,000 people, while Kolkata saw the lowest decline (3.3 to 3.0).

Due to the decline in the risk of death in eight cities, the risk of death in metropolitan cities in India increased by only 5% over a decade from 9.4 deaths per 100,000 people in 2003 to 9.9 deaths per 100,000 people in 2013. Figure 7 shows fatality rates in Indian metropolitan cities for 2003 and 2013. In 2013, fatality rates ranged from 1.5 deaths per 10,000 vehicles in Hyderabad to 8.5 deaths per 10,000 vehicles in Kolkata. However, the fatality rate in none of the sample cities is higher than the all-India average (8.6 deaths per 10,000 vehicles).

From 2003 to 2013, the death rate decreased in fifteen of the twenty-one sample cities. Seven cities, Indore, Bengaluru, Hyderabad, Bhopal, Pune, Kanpur and Mumbai, saw a more than 50% drop in mortality over a decade. This is why the fatality rate in Indian metropolitan cities has decreased by 42% from 5.1 deaths per 10,000 vehicles in 2003 to 3.0 deaths per 10,000 in 2013.

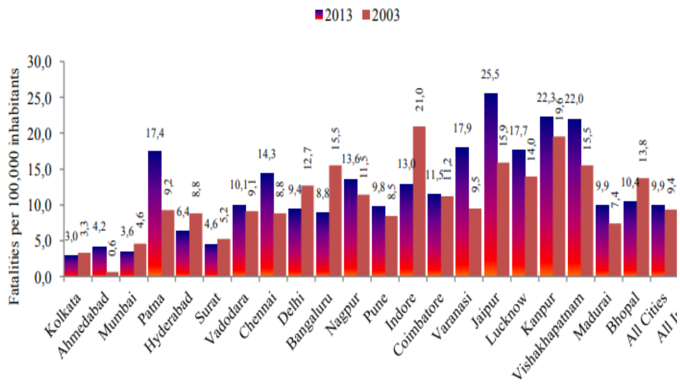


Figure 6. Road Accident Fatality Risk in Selected Indian Metropolitan Cities in 2003 and 2013

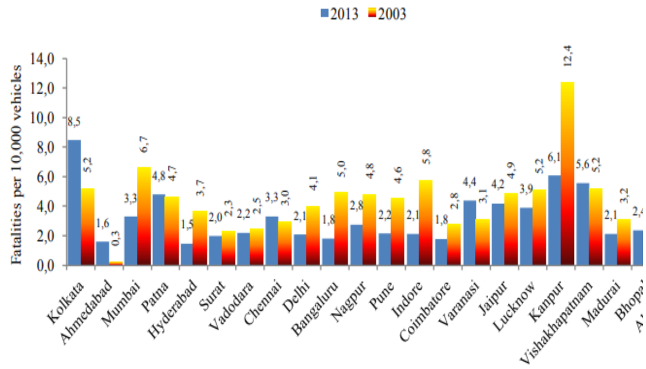


Figure 7. Road Accident Fatality Rate in Selected Indian Metropolitan Cities in 1999 and 2009

IV. METHODOLOGY

Here we have compared different machine learning methods for road accident analysis. Below you see a Schematic of the project architecture.

<u>Explanation of key functions</u>

The following functions are used:

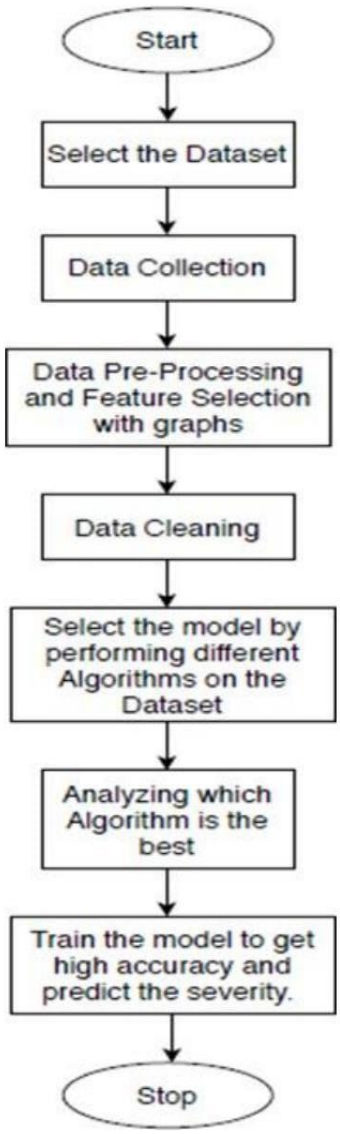
plt.grid(): <u>Matplotlib-Grids.</u>The axis object's grid() function toggles the visibility of the grid inside figure. You can also see the main / minor (or both) grid tiles. Plus color, line style and line width properties can be set in the grid() function.

plt.ioff(): The ioff() function in the pyplot module of the matplotlib library is used to turn off interactive mode. Parameters: This procedure takes no parameters. Returns: There is no return valuethis procedure.

df.info(): The info() function prints a brief overview of the data frame. This method prints information about Dataframe such as index and column dtypes, non-null values, and memory consumption.

plt.show(): If you're using Matplotlib from a script, the plt.show() function is your friend. plt.show() starts an event loop, finds all currently active figure objects, and displays your figure or figures one or more interactive windows.

df.isnull(): The isnull() function detects missing values in a given series object. If the values are NA, it returns a boolean object of the same size. Missing values map to True and missing values map to False.



V. IMPLEMENTATION

Explanation of different modules

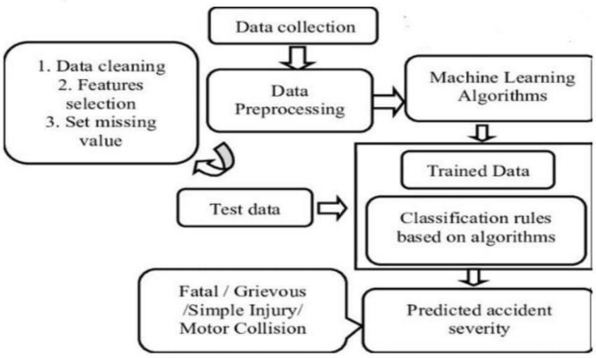
Collection of Data: In this phase the data is collected from different sources.

Data Cleaning and Processing: Once the data is collected by applying some data cleaning methods we remove the errors and process the data.

Train Data: Once the data is cleaned then we apply some machine learning algorithms like (Random Forest, Logistic Regression, Decision Tree).

Test Data: Once the data is trained with the appropriate machine learning algorithms then the data is tested.

Prediction & Results: The tested data is analyzed and the results are predicted and analyzed.



VI. CONCLUSION

The analysis shows that the distribution of accident deaths and injuries in India varies by age, sex, month and time. It is established that the economically active age group is the most vulnerable group of the population. In general, men face a higher risk of death and accidents than their female counterparts. In addition, the number of traffic accidents is relatively higher in May to June and December to January, which shows that extreme weather affects the occurrence of traffic accidents. Accidents are relatively constant and high between 9:00 and 21:00 and variable but low during midnight and early morning hours. There are several factors responsible for accidents, but the most important factor is driver error; driver fault accounted for 78% of total accidents, 76.5% of total injuries and 73.7% of total deaths in 2013. The study also analyzed the road accident scenario in Indian states and cities. It found that three states, Tamil Nadu (22.8), Haryana (17.2) and Andhra Pradesh (16.9), faced a 50% higher risk of death than the all-India average (11.2) during 2013 . It has also been found that the burden of road traffic accidents in India is marginally lower in its metropolitan cities. However, there are huge differences in the risk of death between Indian cities, from 3.0 deaths per 100,000 people in Kolkata to 25.5 deaths per 100,000 people in Jaipur. Despite the increasing burden of road traffic deaths and injuries, road safety is not receiving sufficient attention at central, state and local government levels. The main reason is that the issue of traffic accidents does not belong to any specific body, be it at the central, state or local level. Responsibility for dealing with various aspects of the problem, including roadworthiness testing, road network and road design, land use planning, the implementation and enforcement of road safety legislation and post-accident medical care, is shared between many different agencies, sectors and groups. There is usually no leader to ensure that they coordinate their efforts and solve the problem holistically. This situation needs to change so that responsibilities are clearly assigned, specific roles are assigned to specific agencies and duplication is avoided. Many countries, particularly in the developed world, have seen a sharp decline in traffic accidents and fatalities over the past few decades by adopting a systems approach to road safety that emphasizes environmental, vehicle and road user interventions rather than focusing on only for direct accesses. aimed at changing the behavior of road users. Although the solutions to road safety problems in India may differ from those countries that have very high levels of motorization, some basic principles would remain the same. These include, for example, good road design and traffic management, improved vehicle standards, speed control, use of seat belts and helmets, and enforcement of alcohol limits (Margie et al., 2004). Current efforts to address road safety issues are minimal compared to what should be done. While there are many interventions that can save lives, political will and commitment at central, state and local levels are essential and without them very little can be achieved. Road users in India deserve better and safer road travel.

REFERENCES

[1] Accidental Deaths & Suicides in India, 1970 to 2013 published by the National Crime Records Bureau, Ministry of Home Affairs, Government of India, New Delhi.

[2] Analysis of road accidents in India using data mining classification algorithms- E. Suganya,S. Vijayrani.

[3] Roh J.W., Bessler D.A. and Gilbert R.F., Traffic fatalities, Peltzman's model, and directed graphs, Accident Analysis & Prevention, Volume 31, Issues 1-2, pp. 55-61, 1998

[4] . National Center for Statistics and Analysis http://www-nrd.nhtsa.dot.gov/departments/nrd- 30/ncsa/NASS.html

[5] Abdalla, I.M., Robert, R., Derek, B. and McGuicagan, D.R.D., An investigation into the relationships between area social characteristics and road accident casualties. Accid. Anal. Prev. 29 5, pp. 583–593, 1997.