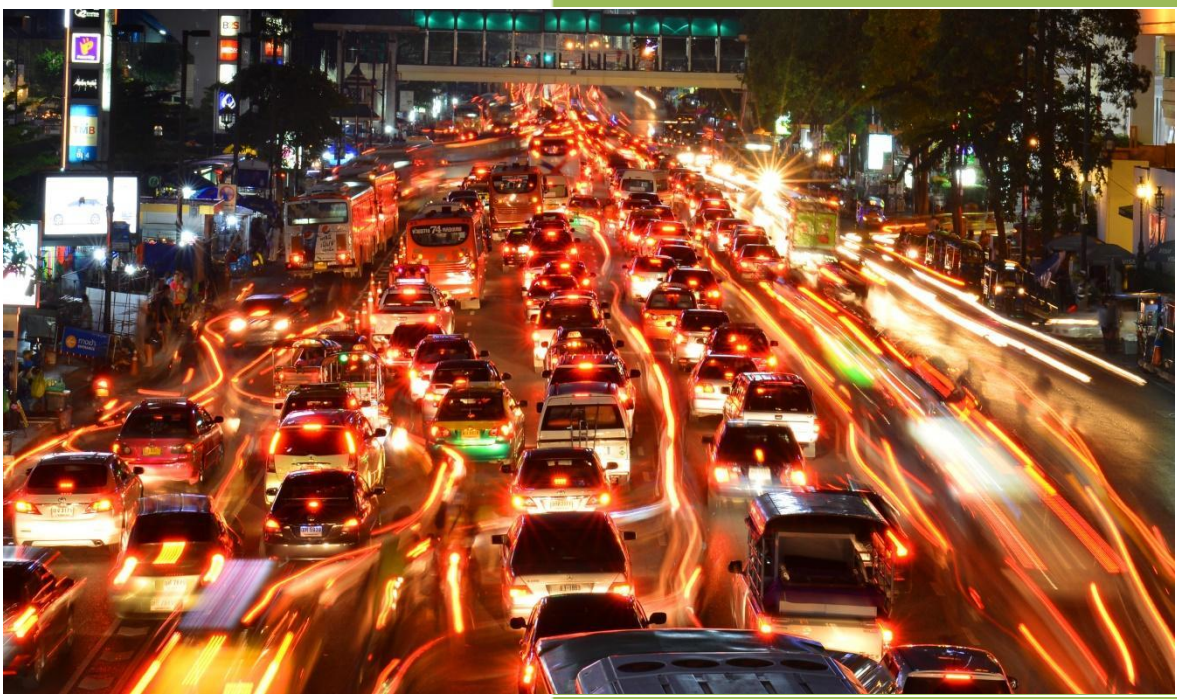
Road Accident Analysis Using Machine Learning



Madhumita Sanjay Patil

Electrical Engineering

2019BEE17

3rd sem

INDEX

About Project

There is a huge impact on society due to traffic accidents where there is a great cost of fatalities and injuries. In recent years, there is an increase in research attention to determine the significant effect the severity of driver injuries which is caused due to road accidents. Accurate and comprehensive accident records are the basis of accident analysis. The effective use of accident records depends on some factors, like the accuracy of the data, record retention, and data analysis. There are many approaches applied to this scenario to study this problem.

So in this project, we used some of the different factors like driver's behavior, roadway conditions, light condition, weather conditions, and so on. This can help the users to compute the safety measures which are useful to avoid accidents. It can be illustrated how statistical methods are based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. The model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injuries that can be used to perform a risk factor and reduce it.

Abstract

There are many inventories in automobile industries to design and build safety measures for automobiles, but traffic accidents are unavoidable. There are a huge number of accidents prevailing in all urban and rural areas. Patterns involved with different circumstances can be detected by developing an accurate prediction model which will be capable of automatic separation of various accidental scenarios. These clusters will be useful to prevent accidents and develop safety measures. We believe to acquire maximum possibilities of accident reduction using low-budget resources by using some scientific measures.

Literature Survey

Sachin Kumar used data mining techniques to identify the locations where high-frequency accidents occurred and then analyze them to identify the factors that have an effect on road accidents at that locations. The first task is to divide the accident location into k groups using the k-means clustering algorithm based on road accident frequency counts. Then, an association rule mining algorithm applied in order to find out the relationship between distinct attributes which are in the accident data set and according to that know the characteristics of locations.

S. Shanthi et al. [2] proposed data mining classification technology based on gender classification, in which RndTree and C4.S use AdaBoost Meta classifier to provide high-precision results. The Critical Analysis Reporting Environment (CARE) system is provided by the Fatal Analysis Reporting System (FARS) used by the training data set.

Tessa K. Anderson proposed a method of identifying high-density accident hotspots, which creates a clustering technique that determines that stochastic indices are more likely to exist in some clusters, and can therefore be compared in time and space. The kernel density estimation tool enables the visualization and manipulation of density-based events as a whole, which in turn is used to create the basic spatial unit of the hotspot clustering method.

The severity of damage occurring during a traffic accident is replicated using the performance of various machine learning paradigms, such as neural networks trained using hybrid learning methods, support vector machines, decision trees, and concurrent mixed models involving decision trees and neural networks. The experimental results show that the hybrid decision tree neural network method is better than the single method in machine learning paradigms.

Problem Definition and Objectives

There are many inventories in automobile industries to design and build safety measures for automobiles, but traffic accidents are unavoidable. There are a huge number of accidents prevailing in all urban and rural areas. Many people lose their lives in traffic accidents. So the objective of this project is to take a survey of different types of circumstances and find out which areas are most accident-prone areas and what changes we need to do in these areas to reduce the number of accidents. This analysis aims to highlight the data of the most important in a road traffic accident and allow predictions to be made.

Methodology

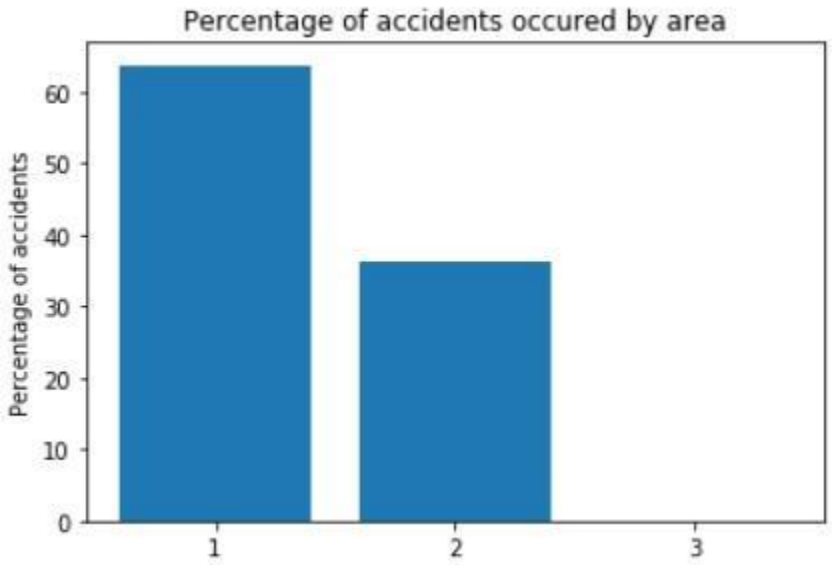
Models are created using accident data records which can help to understand the characteristics of many features like driver's behavior, roadway conditions, light condition, weather conditions, and so on. This can help the users to compute the safety measures which are useful to avoid accidents. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on

out-of-sample forecasts. The model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injuries that can be used to perform a risk factor and reduce it.

Here the road accident study is done by analyzing some data by giving some queries which are relevant to the study. The queries like what is the most dangerous time to drive, what fractions of accidents occur in rural, urban, and other areas? What is the trend in the number of accidents that occur each year, do accidents in high-speed limit areas have more casualties, and so on? These data can be accessed using a Microsoft excel sheet and the required answer can be obtained.

Results

Fraction of Accidents Occurred in Urban, Rural, and Other (Na) Areas



The fraction of accidents that occurred in all areas can be calculated by

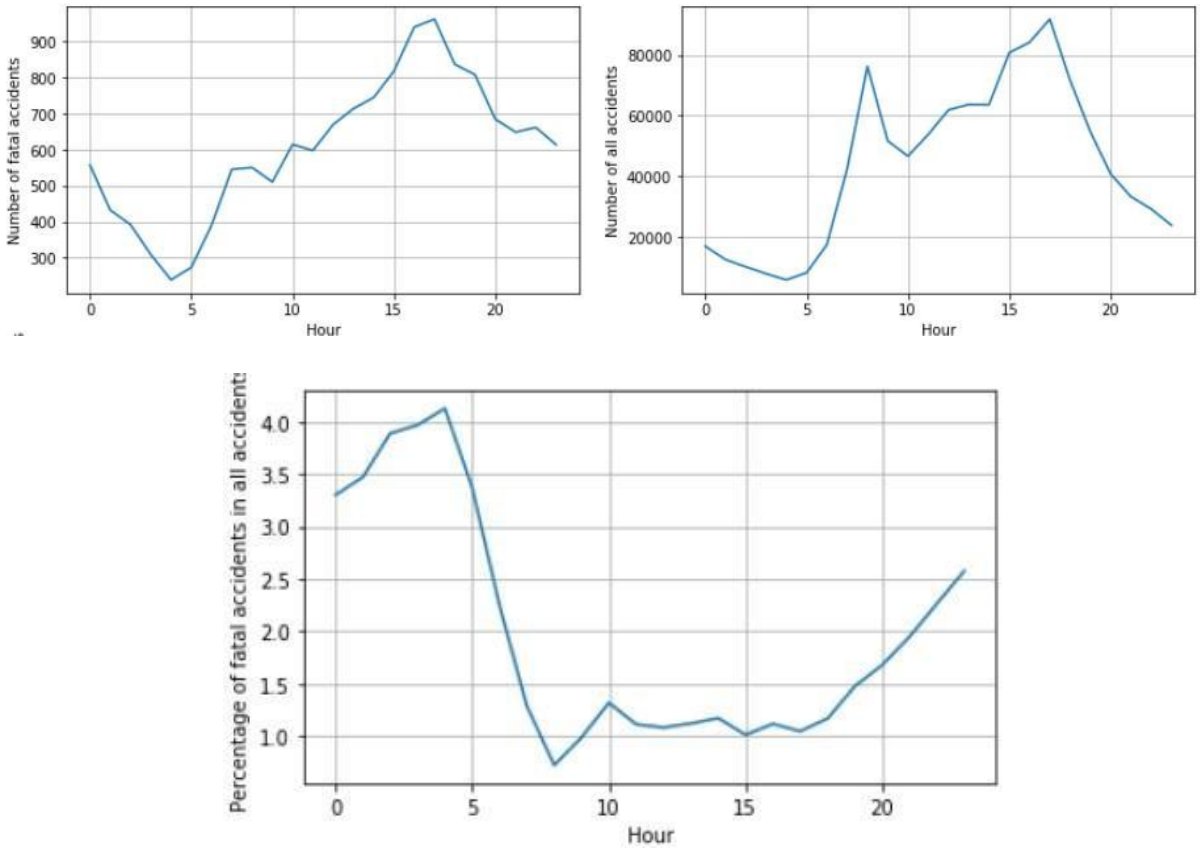
```
urban_acci = len(acci[acci['Urban_or_Rural_Area']==1])
rural_acci = len(acci[acci['Urban_or_Rural_Area']==2])
na_acci = len(acci[acci['Urban_or_Rural_Area']==3])
total_acci = urban_acci + rural_acci + na_acci
```

urban_pct = urban_acci*1.0 / total_acci * 100

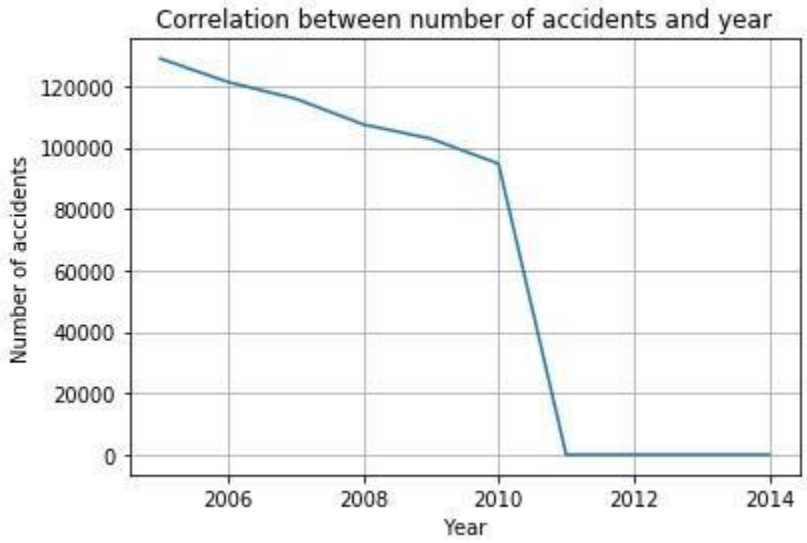
rural_pct = rural_acci*1.0 / total_acci *100

na_pct = na_acci*1.0 / total_acci * 100

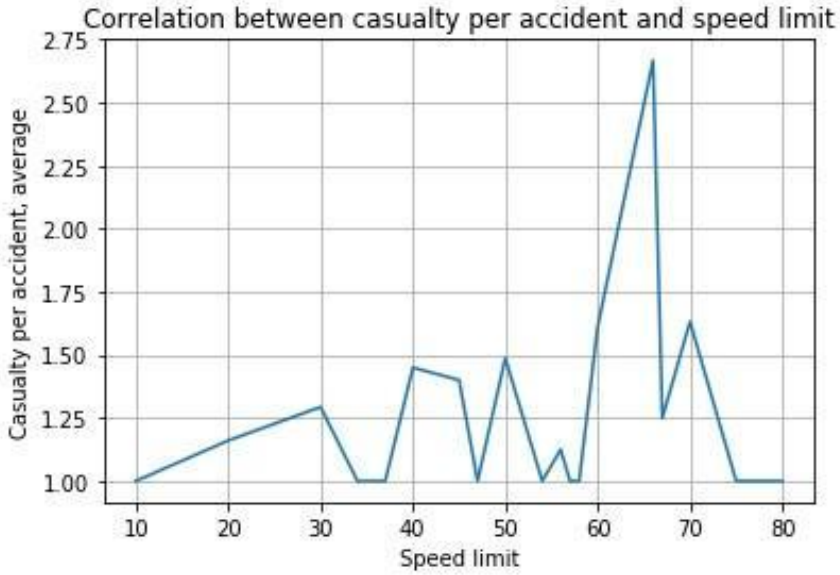
The Most Dangerous Time to Drive



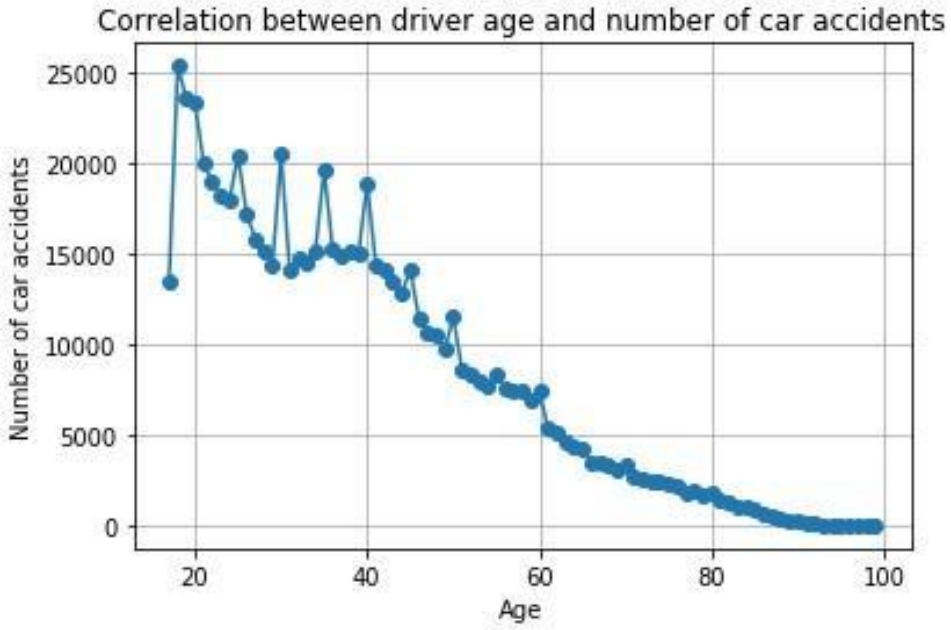
Number of Accidents That Occur Each Year



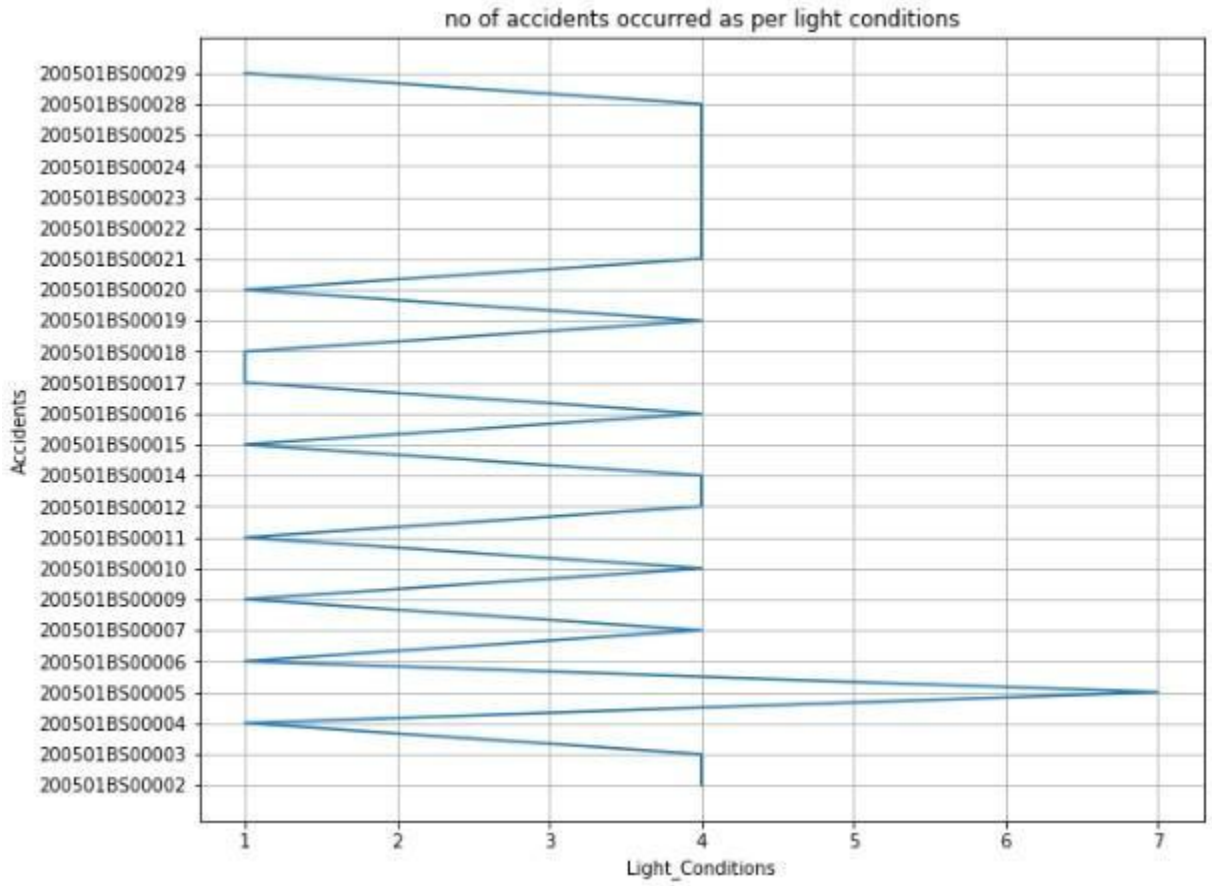
Accidents in High-Speed-Limit Areas Have More Casualties



The Number of Car Accidents Drop Off With Age



No Of Accidents Occurred As Per Light Conditions



Conclusion

Losses in road accidents is unbearable, to society as well as a developing country like us. So, it has become an essential requirement to control and arrange traffic with an advanced system to decrease the number of road accidents in our country. Taking simple precautions, based on the prediction or warnings of a sophisticated system may prevent traffic accidents. Moreover, it's a primary need for our country now, to tackle this situation where every day so many people were killed in traffic accidents and day by day this rate is getting increased. The implementation of machine learning is a functional and great approach to take an accurate decision with the experience to manage the current situation and the findings of the analysis part can be suggested to traffic authorities for reducing the number of accidents. We can use the proposed approaches to implement machine learning here because of their proven and higher accuracy to predict traffic accident severity. Moreover, to make it more feasible, we will try to

make a recommender system by using these approaches that can give a prediction to the traffic accident and can warn the road user. In the future, it will be our try to create a mobile application by implementing this methodology to provide an accurate prediction to the user and make it very useful and beneficial also.

References

[1] Sachin Kumar, Durga Toshniwal, "A data mining approach to characterize road accident locations", J. Mod. Transport. (2016) 24(1):62–72.

[2] S. Shanthi and Dr. R. Geetha Ramani, "Gender-Specific Classification of Road Accident Patterns through Data Mining Techniques", IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012) March 30, 31, 2012.

[3] Tessa K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots", Accident Analysis and Prevention 41 (2009) 359–364.