# Analysis and Visualization of Road Accidents

Shubham Salunkhe, Saswat Panigrahi, R. Rajkumar

School of Computer Science and Engineering

Vellore Institute of Technology, Vellore

[shubhamsatish.s2019@vitstudent.ac.in](mailto:shubhamsatish.s2019@vitstudent.ac.in), [saswat.panigrahi2019@vitstudent.ac.in](mailto:saswat.panigrahi2019@vitstudent.ac.in), [rrajkumar@vit.ac.in](mailto:rrajkumar@vit.ac.in)

**1. Abstract**

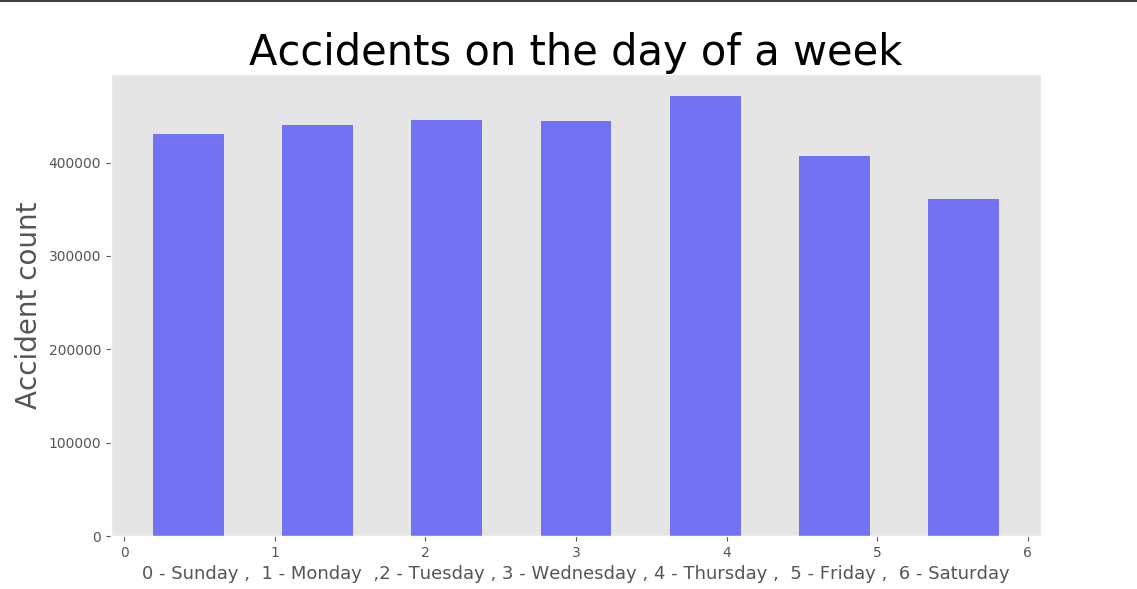
There is an increase in number of road accidents in modern days and with the increase in the accident rates there has been a huge generation of data regarding these accidents. There are many factors that can be the cause of accidents happening such as road condition, traffic control, load on vehicle etc. Other factors might be due to the driver’s carelessness and effect of drug usage. Also location, light condition, traffic type (manual/automatic), road condition etc. also has an impact on number of accidents in a particular area. These factors need to be analyzed to get an idea regarding the impact these factors have on the accidents. For e.g.:- if we want to analyze the impact of light conditions or the road conditions on the no of accidents happening. With the advancement in data mining tools it is easier to analyze the impact that these factors have on the number of accidents by use of some graphs or plots. We can also find the relation between the no of causalities and how each of these contributes to number of causalities in a particular area. In this paper we try to analyze the dataset on basis of some of these factors and try to find the severity of accidents depending on some of the above factors.

# 2. Introduction

There has been an increase in the amount of data’s generated in terms of volume, velocity, variety in various fields such as social media etc. In this paper we have taken accident dataset between a fixed period of time and analyzed the no of accident and the severity or no of causalities in that particular and we come with some kind of analysis that which factors are contributing to most of the accidents or causality in that particular area. Some of the major causes that contribute to the accidents include the physical condition of the roads, the light condition in the particular area, the road type, the type of traffic in that area i.e. it can be manual control or it can be automatic. Also the driver condition do matter that he might be in a subconscious mind or he might under effect of drug or alcohol. Data mining tools or algorithm have become more useful such kind of huge data and come up with some analysis. The analysis includes taking these factors as parameters and performing some kind of relation between these factors and how they contribute to the severity of accident. We also take some other parameters such as age group of the victims. These analysis results can be shown in form of some graphs or plots like bar plots and come up with some kind of visualization. Then we will be able to analyze which of these factors contribute most to the increase in no of factors and which contributes to the most no of causalities in that particular area. We use some statistical tools such as regression and other important statistical methods to come up with some analysis. Hence use of statistics and data mining techniques is the important part to come up with these kind of analysis and coming up with some kind of visualization. These statistical and data mining works are done using python

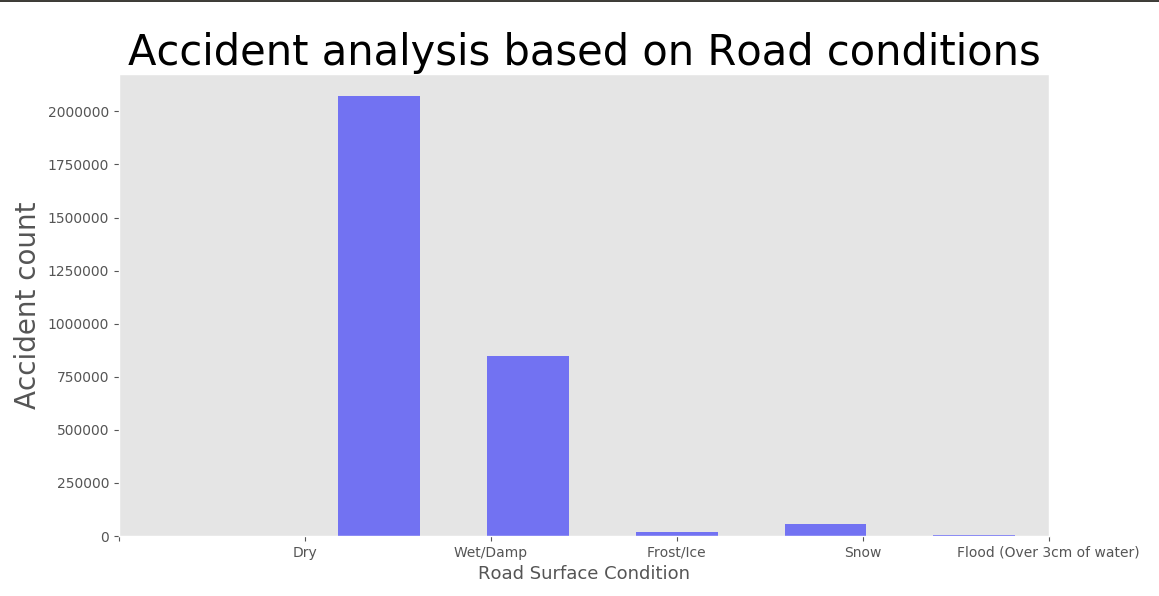
**3. Related works**

The main focus of paper [1] was analyzing the factors that affected road accidents and then tried to estimate which factors are most likely to cause accidents in a particular area. The factors they took included careless driving, bad road condition, drugs or medicine, drunken driving, traffic, defect in vehicle. They used WEKA tool .Their conclusion was that drugs and medicine contributed most to accidents. In the [2] paper they have used some data mining techniques to predict accident prone zones and road conditions and they have used statistical tools like correlation analysis and came up with some bar plots and scatterplots to find the relation between the type of roads and frequency of accidents. They used R programming for visualizing the data’s. They examined the impact on different kind of roads such as state highways etc. and came to the conclusion which contributed most to accidents

In the [3] paper some of the properties or features which are responsible for causing of road accident and causalities were predicted or analyzed. Use of big data framework such as hadoop mapreduce was used. The factors taken were vehicle condition, road condition and other environmental condition like wet roads or driver physical condition. Some machine learning algorithm such as regression was used and statistical tool like correlation was used to find the relation between these features. In the [4] paper the relationship between causality rate and other attributes including collision manner, weather, surface condition, light condition, and drunk driver were investigated. The data mining algorithm used was k-means clustering and some machine learning techniques like naïve Bayes classifier was used.

They developed some plots that showed the fatal accidents in different months and the factors like weather, road type, amount of light, and other physical problems like whether drunk or not. From the graphs and analysis techniques they concluded that factors like alcohol consumption and careless driving were main cause for accidents.

**4. Methodology Used:-**



In this section, we are going to analyze & visualize different factor which should considered while analyzing accident data. Visualization of factors is important and easy to understand effect of those factors on accident count. On visualized data, we can decide which factors out of total factors to be used for analysis purpose.

Fig-4.1

1) Day of Week:-

From data of day on which accident happened, we have plot histogram in Fig.4.1. So, we can see that, Day of the week factor is not affecting that much on accident count. But, statistically On Thursday, Accident count is more as compare to other days.

Fig.4.2

2) Road Condition:-

Road conditions differ from season to season. So from histogram Fig.4.2, Accident count on Dry and Wet/Damp road is serious as compare to frost/ice, snow and flood. But if we compare road conditions Dry and Wet/Damp, we can say that on Dry road accident count is more because over speeding chances more on dry road as compare to wet road.

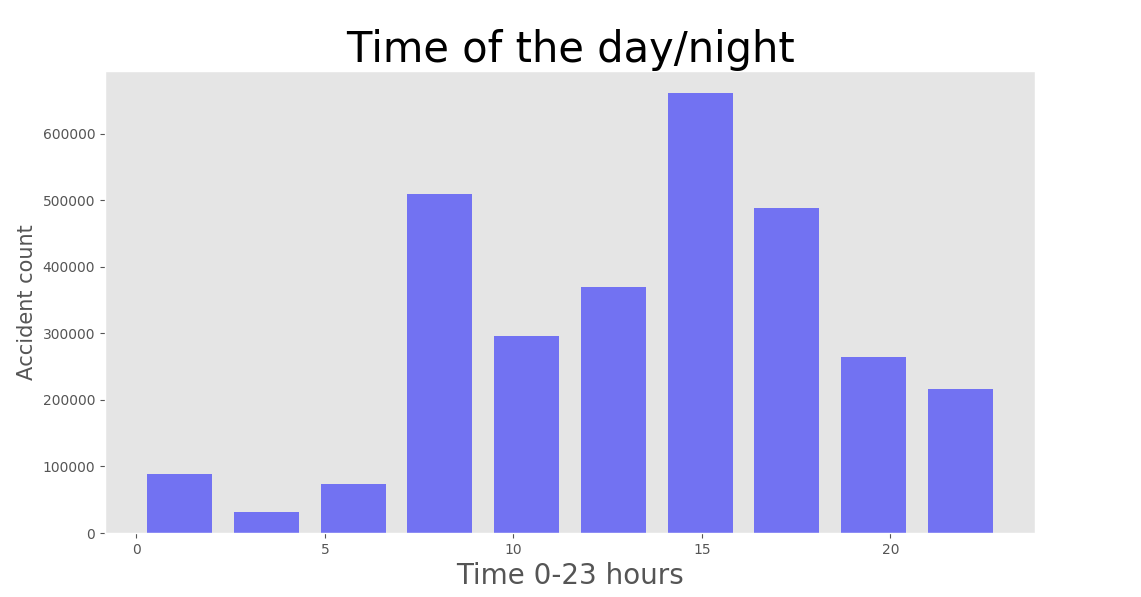


Fig-4.3

3) Time of the day/night:-

As we can see in Fig.4.3, there is visualization of accident count vs. Time/hours of day. So, we can say that between rush hours like 8 AM to 10 AM and 3 PM to 7 PM, accident count is increasing. At night time, accident count is less as compare to day time because traffic on road less at night as compare to day.

4) Speed limit:-

Most of the accident happened on road with speed limit 30 as per pie-chart in Fig. 4.4. So, over-speeding and crossing speed limit on roads with speed limit 30 is one of the reason behind increase in accident count which is 62.7% of total accident count. In speed limit 30-50, accident count is moderate but if speed limit increases to 60 accident count percentage reaches to 15.7%. So, due to not following speed limit rule, accident count increases.

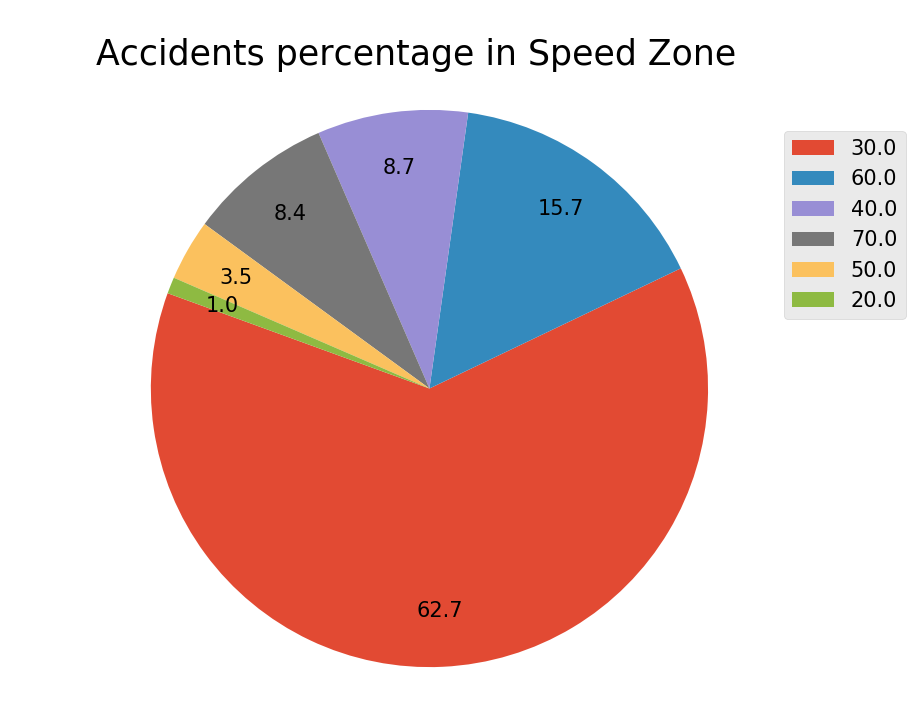


Fig-4.4.

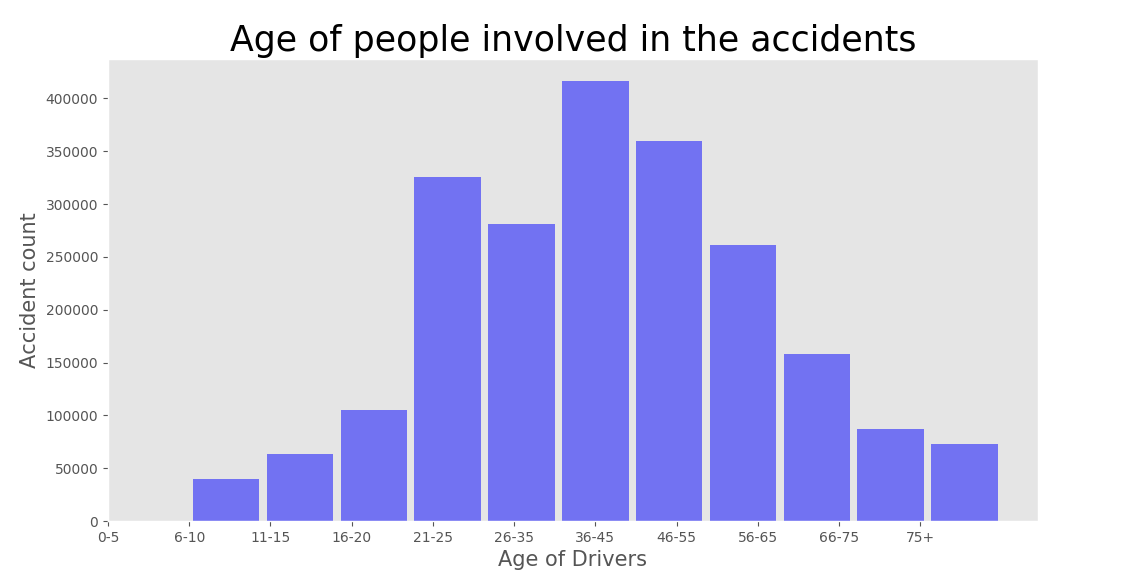


Fig.4.5

5) Age of people involved in accident:-

From Fig.4.5, it visualizes that Accident count based on age of driver. So, histogram gives information that under age drivers involved in accident are there. But which are less as compare to people of age between 20-55 involved in accident. In that also, people of age between 36-45 are more as compare to other age band. Working & college students age group people mostly involved in accident. Ages after 60 peoples are less involved in accident.

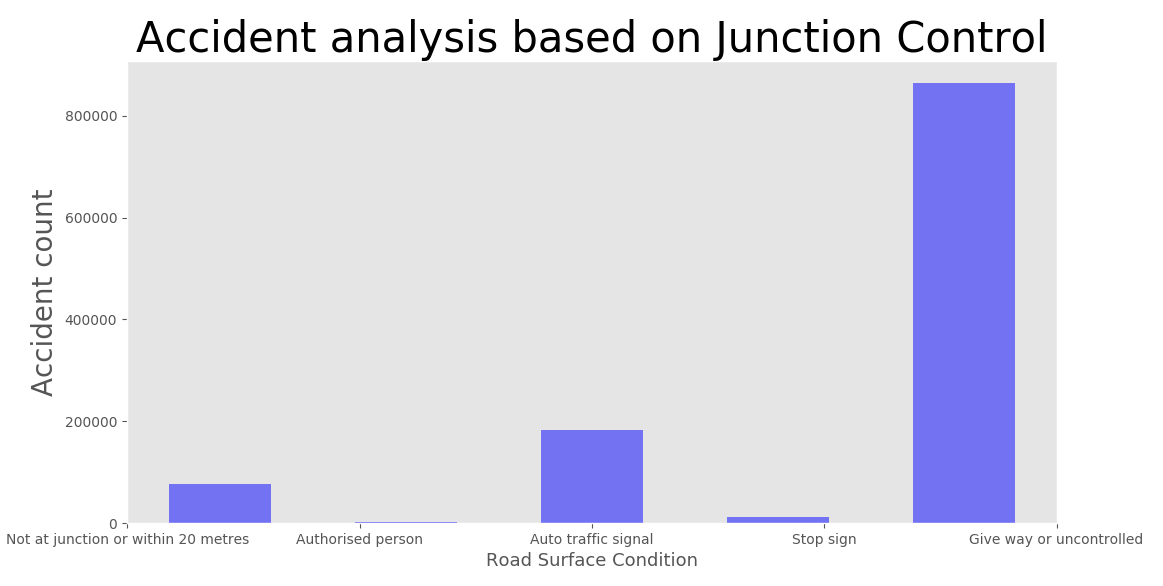


Fig-4.6

6) Junction/Signal control:-

With change in junction control technique, accident count changes as per Fig.4.6. if junction controlled by authorized person like traffic police, then accident count is less as compare to other types of junction controlled. But from histogram we can say that due to uncontrolled junction, there is drastic change in accident count. In case of Auto traffic signal, accident count is more than junction controlled by authorized person. The main reason behind it lack of technology in signal automation.

**5. Result & Discussion:-**

|  |  |
| --- | --- |
| **Dependent Factors** | **Correlation Coefficient** |
| Accident\_Severity | 1.000000 |
| Number\_of\_Vehicles | 0.074951 |
| Weather\_Conditions | 0.021433 |
| Road\_Surface\_Conditions | 0.010136 |
| Day\_of\_Week | 0.002552 |
| Road\_Type | -0.039005 |
| Light\_Conditions | -0.063311 |
| Number\_of\_Casualties | -0.080690 |
| Speed\_limit | -0.080690 |

Table 5.1

If we found correlation of different factors with Accident Severity, then we got above results in Table 5.1. Number of vehicles involved in accident, weather condition, Road surface condition and Day of week positively correlate with respect to Accident Severity, but other factors negatively correlate.

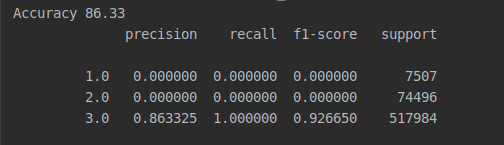
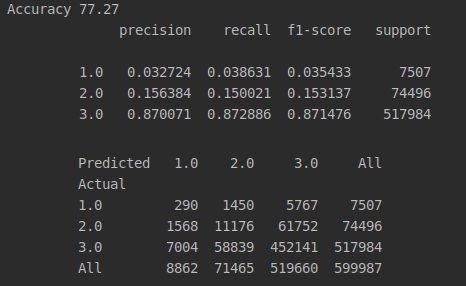


Fig-5.1

According to Fig. 5.1, logistic regression on data gives results in form of precision, recall, f1-score and support. So, Accuracy of applied algorithm is 86.23%. Precision score gives the level of prediction made by the model is precise for each accident severity. Recall value of model is the amount out of total dataset up-to which it can predict the outcome. F1 and Support scores are the amount of data tested for the predictions.



-

Fig-5.2

Fig 5.2 gives result of Decision tree classifier, which gives 77.27% accuracy which is less than logistic regression but precision and recall for each severity is good as compare to logistic regression. Also Confusion matrix gives good results. Performance of a prediction model is described by Confusion matrix table. A confusion matrix contains the actual values and predicted values. For severity 1, 7507 out of 8862 records are predicted by classifier and mostly accurate result to other severity.

**6. Conclusion:-**

As we have predicted accident severity based on implemented Logistic Regression and Decision Tree algorithms. As we tried two different algorithms to predict the accident severity. It was clear that Decision tree performed much better in terms of predicting all classes of accident severity. Logistic Regression has better accuracy than Decision tree but it does not mean it did better. In Overall analysis, we can say that Junction Control, Road conditions and Time are the factors which are mostly effect on accident count. But Day of the week and Speed limit factors are not affecting that much. Weather condition and Road type also positively affect accident severity.

**7. References**

1. [Suwarna Gothane](https://scinapse.io/authors/2685776833),[M.V. Sarode](https://scinapse.io/authors/2559611615)’s “Analysing Factors, [Construction of Dataset](https://scinapse.io/papers/2518518769), [Estimating Importance of Factor](https://scinapse.io/papers/2518518769), [and Generation of Association Rules for Indian Road Accident](https://scinapse.io/papers/2518518769),” IEEE, Feb 1, 2016
2. Gagandeep Kaur, [Er. Harpreet Kaur](https://scinapse.io/authors/2773181988)’s “Prediction of the cause of accident and accident prone location on roads using data mining techniques,” IEEE, Jul 1, 2017
3. Rishi Sai Reddy Sudireddy’s **“**Prediction of Road Accidents using Correlation based on Map Reducing,” IEEE
4. Liling Li, Sharad Shrestha, Gongzhu Hu’s “Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques,” IEEE, June 7-9, 2017
5. Fan Fan, Yu-Bin Qian’s “Analysis of Factors Affecting the Severity of Car Accidents at iNtersections Based on Cumulative Logistic Model,” IEEE, Dec 25-27, 2017.
6. Nagarjuna R Vatti, PrasannaLakshmi Vatti, Rambabu Vatti, Chandrashekhar Garde’s “Smart Road Accident Detection and communication System,” IEEE, March 1-3, 2018.
7. Jianhong Xu, Fujian Wang, Xipeng Wang’s “The analysis of accident black spot at the intersection of NO. 320 state road crossed with Lingjia Bridge,” IEEE, July 15-17, 2011.
8. Youhee Choi, Jeong-Ho Park, Byungtae Jang’s “Developing safety checklists for predicting accidents,” IEEE, Oct. 17-19, 2018.
9. Jihua Ye, Yanhui Zhou, Ming Li, Chunlan Wang’s “Research and implement of traffic accident analysis system based on accident black spot,” IEEE, Aug. 24-27, 2010.
10. Leena Rupesh, Aarti Jain, Subham Sarda, Shivam Gupta, S.I. Sajeesh, Dhayanithi Palanisamy’s “IoT and Cloud Based Integrated System for Accident Reporting and Vehicular Health Monitoring,” Nov. 22-23, 2018.
11. Yang Yanbin, Zhou Lijuan, Leng Mengjun, Sun Ling’s “Early Warning of Traffic Accident in Shanghai Based on Large Data Set Mining,” IEEE.