**CAR ACCIDENT SEVERITY PREDICTION**

***ADITYA PRAKASH SINGH***

**12-09-2020**

**INTRODUCTION:**

As a significant cause of deaths, injuries, and property loss, traffic accident is a major concern for public health and traffic safety. According to statistics from the Ministry of Public Security of China between 2009 and 2011, traffic crashes resulted in an average of 65 123 people dead and 255 540 cases injured annually in China (China Statistical Yearbook of Road Traffic Accidents, 2009–2011). It was reported that the cost of medical care and productivity losses associated with motor vehicle crash injuries was over $99 billion, or nearly $500, for each licensed driver in the United States (Centers for Disease Control and Prevention, 2010). Being one of the major steps of accident management, accident severity prediction can provide crucial information for emergency responders to evaluate the severity level of accidents, estimate the potential impacts, and implement efficient accident management procedures.

In recent years, increased attention has been directed at accident severity prediction, for which Bayesian network and Regression model are two widely used modeling techniques. However, to the authors’ knowledge, there is no study that presents quantitative comparison of the two methods. Therefore, the present work focuses on conducting an accident severity modeling by employing both Bayesian network and Regression model. The accuracies of the two methods will then be compared and a better one will be selected for accident severity prediction. By carrying out accident severity analysis, the risk factors and their effects will also be identified in the work.

**DATA:**

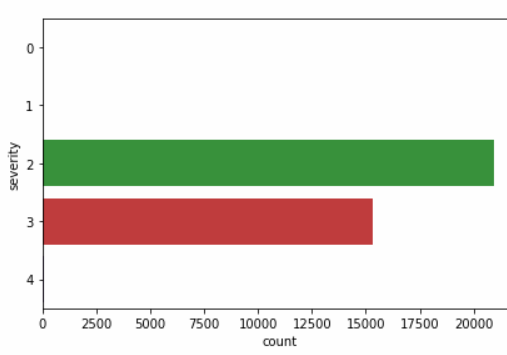
The data set for this work contains police-reported traffic accident records for San jose, California. With records containing missing values eliminated, our final data set consists of 36,315 cases, which are all motor-vehicle involved accidents. In addition to severity information, the data contains information regarding accident characteristics (accident occurrence time and accident location), vehicle characteristics (vehicle type involved and vehicle condition), environmental factors (weather condition and visibility distance), and road conditions (pavement condition, road geometrics and roadway surface condition, etc.).

**EXPLORATORY DATA ANALYSIS:**

1. Characteristics and summary statistic of data with graphs-

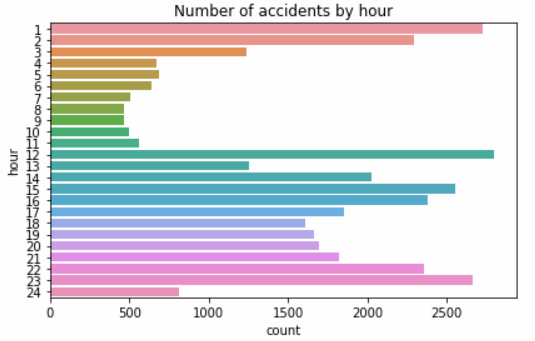
It has been observed that type 2 severity accidents are more than type 3 severity accidents.

Total count of both accidents are approx. 35,000 accidents.

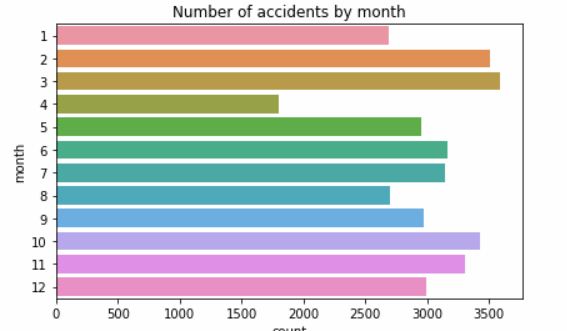


2.Number of accidents by hour-

It has been seen that frequency of accidents rises during afternoon and night . Morning has the least chances of having an accidents.



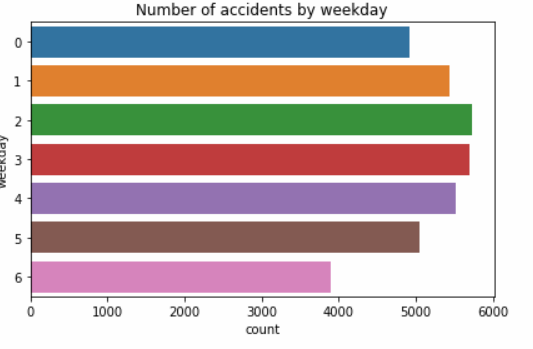
3.Number of accidents by month-



4.Number of accidents by weekdays-

Working days of weeks have more accident frequency than the weekend days.

Saturday has the least frequency. Tuesday and Wednesday both has the most accidents.



**RESULT:**

After applying different algorithms to predict the severity accurately. Here’s the accuracy of each algorithm-

1.Logistic Regression: Accuracy- 58%

2.Tree algorithm: Accurracy- 95%

3.Random Forrest algorithm: Accuracy- 95%