Predicting Car Accident Severity.

–Jyotsana Rathore

October 2020

1 Introduction

1.1 Background

Traffic accidents are extremely common, and it is more frequent in the sprawling metropolis. According to a research conducted by the World Health Organisation (WHO) there were 1.35 million road traffic deaths globally in 2016, with millions more sustaining serious injuries and living with long-term adverse health consequences. Globally, road traffic crashes are a leading cause of death among young people, and the main cause of death among those aged 15-29 years. Road traffic injuries are currently estimated to be the eighth leading cause of death across all age groups globally, and are predicted to become the seventh leading cause of death by 2030.

A system that can predict the cause of traffic accidents and predict the accident severity can potentially save lives and aid in developing better roadway designs. The detailed data, public transportation information, and motor vehicle crash reports could provide us valuable insights for traffic accident analysis.

1.2 Problem

However, this problem is very taxing due to several issues such as class imbalance and low significance of predictor variables depending upon the location meaning too much randomness in the data. Since we are creating a binary target variable between slight accidents (0) and serious or Fatal accidents (1) the class will be severely imbalanced. Simple linear models might not be useful for predicting the accident severity due to non-linearity.

Transport authorities worldwide have been striving to implement strategies to minimize the road traffic accidents by introducing safety regulations. There were many strategies implemented for road traffic accidents reduction but has proven to be an elusive goal as these measures have hitherto failed to make a considerable reduction in the frequency of the road traffic accidents. Accidents are influenced by many measurable factors such as driving speed, road condition, weather condition, light condition and so on. Therefore, many researchers have come together to understand the dynamics of road traffic accidents.

Katannya Kapeli and Meraldo Antonio (2019) researched that weather conditions had no role in severe or fatal accidents. This is quite evident because there were multiple factors that affected the road traffic accidents, they considered junction, time, origin and destination played a vital role in predicting the road traffic accidents. They used a negative sampling technique using the several hundreds of accident hot spots and they classified their target variable with accident and no accident. They used classification machine learning models and their best model was random forest with only numerical predictors.

Other researchers aim to classify the target variable into binary classes {accident, no accident}. They compared the performance of artificial Neural Network with the

negative binomial regression models. Artificial Neural Network achieved 64% and 61% accuracy for training and test data. They also applied a decision tree, random forest, k nearest neighbour to predict road accidents. The best accuracy achieved by them was 70%. Their dataset was quite old 9 therefore comparing it directly is unfair. They have not mentioned anything about the imbalanced class problem.

1.3 Interest

The potential to predict the accident severity (e.g. What will cause accidents? What will make accidents worse?) is therefore useful not only to public safety stakeholders but also transportation administrators and individual travellers