1.INTRODUCTION

1.1 Background

Road accidents kill a lot of people every year. As a society it is always good to reduce the road accident rate and the financial burden and stress it place on our health care system The road accident death data is a shocking and dreadful statistics to comprehend,
In last year around 1.65 mn deaths occurred due to road accidents alone(source:WHO) at least 28 percent of which could have been avoided with better infrastructure and prediction systems

1.2 Problem

Our aim here is to give early warnings to drivers based on the parameters like – Road type, weather condition, junction type etc. We will use different machine learning models like decision tree and linear regression etc so as to achieve the same

2. Data acquisition and cleaning

2.1 Data Source

The source of data used here is the Seattle PD accident info. The data contains about 190000 records for training our model and get to meaningful conclusions

2.2 Data Cleansing

The Data had many null or junk values which had to be renamed 'Unknown' for better accuracy(All were categorical data). Undersampling was also done so as to avoid bias

2.3 Feature selection

The feature selected was of 6 numbers – LOCATION, WEATHER, ROADCOND, JUNCTIONTYPE, ADDRTYPE and LIGHTCOND. These were selected as others were just metadata that wont contribute anything for out model building

3.Data Distribution

The following is the data distribution for the selected features
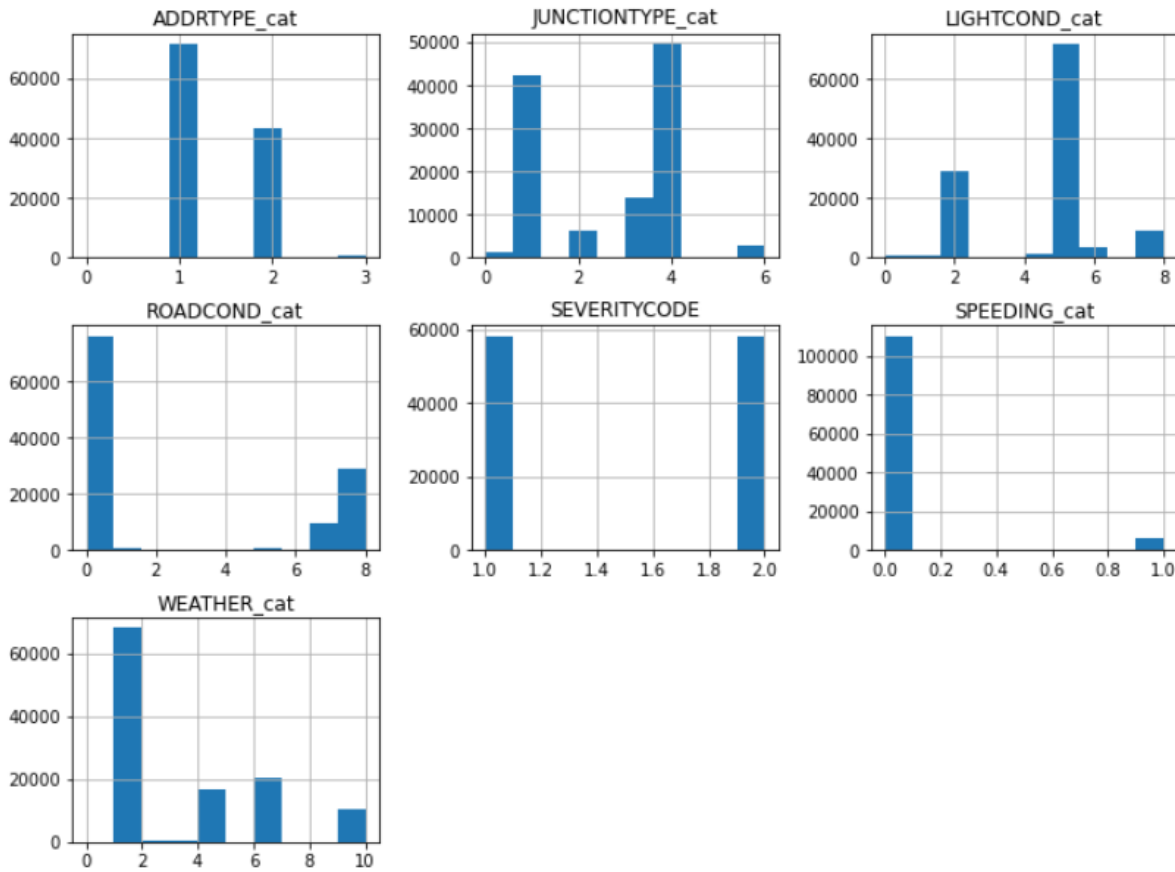
4.MACHINE LEARNING MODELS USED

Three main machine learning models were used

1. KNN
2. Decision Tree
3. Linear Regression

5.MODEL ACCURACY

	ML Model	Jaccard Score	F1 Score
0	KNN	0.448474	0.604405
1	Decision Tree	0.437780	0.622065
2	Linear Regression	0.456425	0.609213

As we can see here the three models didn't have much accuracy difference between them, and as its evident Decision Tree is the best model to model this (based of F1 score statistics)

6.Conclusion and Further steps

- Accident severity was predicted with the help of different machine learning models

- The prediction could be more polished using more data point(eg:accident prone rate of the area, age of the driver, experience of the driver etc

- New instruments or technologies could be adopted so as to reduce the fatalities and to strengthen the current scenario