



# ACCIDENT SEVERITY PREDICTION



# The befitting approach

Seattle Police Department

Data Collection

Machine  
learning model  
evaluation

KNN, Linear Regression and Linear  
Regression

The best model

Decision Tree





# accident severity prediction can save lives!

- Average number of car accidents in the U.S. every year is 6 million. 3 million people in the U.S. are injured every year in car accidents. Analyzing the conditions that contribute to these accidents would lead to the prevention of significant loss of life and financial resources.
- These analysis can be useful so as to warn the drivers of the possible accidents that may occur on a given day based on the factors like weather, location and type of alleys etc



# Insight into the data(Data acquisition)

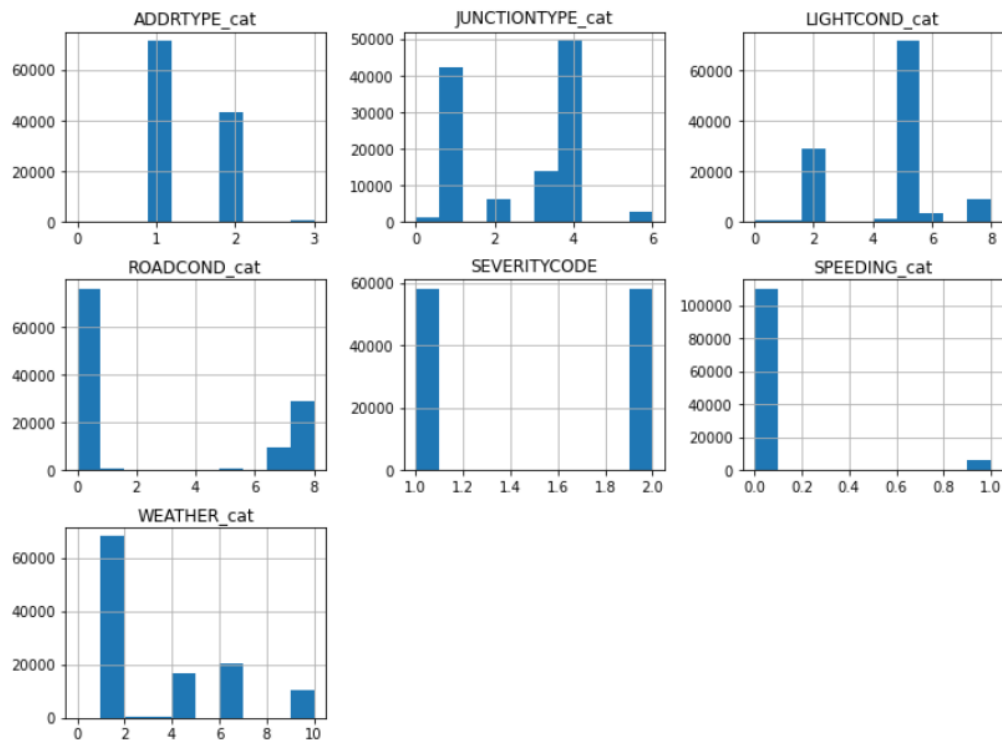
- This data was provided by the Seattle Police Department and recorded by Traffic Records Department. The dataset consists of 37 independent fields and 194673 records, which includes both numerical and categorical data. The dependent field or label for the data set is SEVERITYCODE, which describes the fatality of an accident. SEVERITYCODE is labelled as 1 or 2
- 1 denotes Property Damage Only Collision
- 2 denotes Injury Collision



# Data Cleaning

- The Data had a lot of categorical values for the fields weather, road condition, light condition etc. But the data was encoded into numerical values using `labelencoder()`
- Also the dataset had unequal number of severity code distribution, we had to undersample the data to make sure that the model was not biased on a particular result
- Many null values were all replaced by 'Unknown' as the replacement

# DATA overview



- As u can see on the right pain is the distribution of various variables which are a deciding factor on the final prediction
- Please note that we cleansed the data so as to make equal number of records with severity code as 1 and 2
- This will help us creating a non biased model

# Notable observations

- We can see that a lot of Factors were involved on the model building process, but some were more notable than others .
- The 3 important deciding factors were : Weather, road condition and junction type. The algorithms weighted these attributes as the major player in deciding on the severity of an accident

# Model comparison

	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

- Decision Tree performed the best among the three models(F1 score based)
- All the 3 models were having only minute differences between their accuracy
- Wonderful to see that linear Regression still performed close enough with other classification algorithms for a classification problem



# 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



THANK YOU!