Predicting Severity and Cause of Road Accidents

IBM Capstone Project

SURANGA WIJERATNE

BACKGROUND

- Road accident are extremely common around the world. More server road accidents could end up with single or multiple causalities.
- Machine Learning model to be used to predict what are the common reasons for severe accidents in Victoria, Australia.
- Different features such as road conditions, weather, light conditions, road geometry to be used for predicting target variable severity of the accident.
- The reduction in severity of accidents can be beneficial to the public authorities to work towards improving those road factors and also car drivers themselves who may take precautions to reduce the severity of accidents.

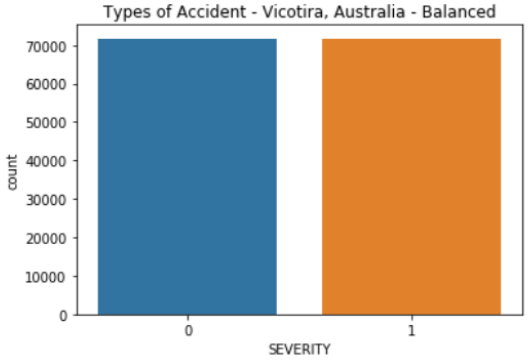
Data Acquisition, Cleaning & Pre Processing

- The data set is taken from Data.Vic which is the place to discover and access Victorian government open data.
- This is an extensive data with over 14 years from 2006 to 2020 over 200,000 observation across Victoria.
- The data set consists of huge number of features which was identified as unrelated features for this exercise. Some of the categorical features are already encoded to numeric but there was no description provided. Out of all the features, below variables were chosen for the model.
  - Light Condition
  - Road Geometry
  - Atmosphere Condition
  - Surface Condition
- All the features above selected are categorical variables were encode to numerical values

Methodology

- Tools and Platforms
  - GitHub as repository
  - Jupiter Notebook for process and build ML model
  - Python as programming language

- Exploratory Data Analysis
  - re-sampling was carried out for the data set to balance target variable



Methodology cont..

- Below ML model were used :
  - K Nearest Neighbour (KNN)
  - Decision Tree
  - Logistic Regression

- Machine Learning Model Evaluation

ML Model	Jaccard Score	F1 Score	Accuracy
K Nearest Neighbour	0.36655	0.54232	0.545701
Decision Tree	0.30725	0.53753	0.573743
Logistic Regression	0.33454	0.54255	0.546259

Conclusion

- Based on the dataset provided for this capstone from weather, road surface, light conditions and road geometry pointing to certain classes, we can conclude that particular conditions have a somewhat impact on car accidents in Victoria. However, the model could have been performed more better if a few more things were presented and possible in data set.
  - A balance dataset for the target variable
  - Less unknown values within the dataset for all the variables
  - More features such as drivers status at the time of driving (eg, after work, going for work, use phone etc)