Capstone Project – Car Accident Severity: Seattle

NC

Introduction

Background

- Road traffic injuries cause considerable economic

- Road traffic accident caused most countries 3% of their GDP

- Hence it is important to analyse the severity of accidents and their causes

Introduction

Business Problem

- To predict the severity of accidents happening on the roads of Seattle based on multiple attributes

Target Audience

- State Departments

- Drivers

- Logistic Companies

- Cab Companies etc.

Data Acquisition and Cleaning

Data Source

- Data is from Seattle Department of Transportation in CSV format

- The dataset is around 200,000 events

Data Cleaning & Feature Selection

- Total 37 attributes

- 4 attributes were chosen based on negative correlation with the severity – WEATHER, ROADCOND, LIGHTCOND, VEHCOUNT

Methodology

Data Analysis & Preparation and Normalization

- Data Balancing

- Check & change the dtype

- Normalize the dataset

- Train Test Spilt

Modelling/ Classification Source

- K-Nearest Neighbours (KNN)

- Decision Tree (DT)

- Logistic Regression (LR)

Result and Discussion

- The best model: combination of high F1-score, high jaccard similarity score and the smallest log loss

- The best performance is observed in Decision Tree model

	KNN	DT	LR
Accuracy	0.5999	0.6075	0.5233
F1-Score	0.5953	0.5828	0.5139
Jaccard similarity score	0.4684	0.5200	0.2869
Log Loss	-	-	0.6811

Conclusion

- To predict the severity of the car accidents that will happen in Seattle based on the known conditions like weather, lighting conditions, road condition and vehicle count

- Model can be used by individuals and many sectors

- Prepare them to handle the situation more effectively

Thank you