CAR ACCIDENT SEVERITY PREDICTION

ASVIN SRITHARAN
OCTOBER 8, 2020

PREDICTING SEVERITY OF AN ACCIDENT CAN BE BENEFICIAL FOR GPS COMPANIES

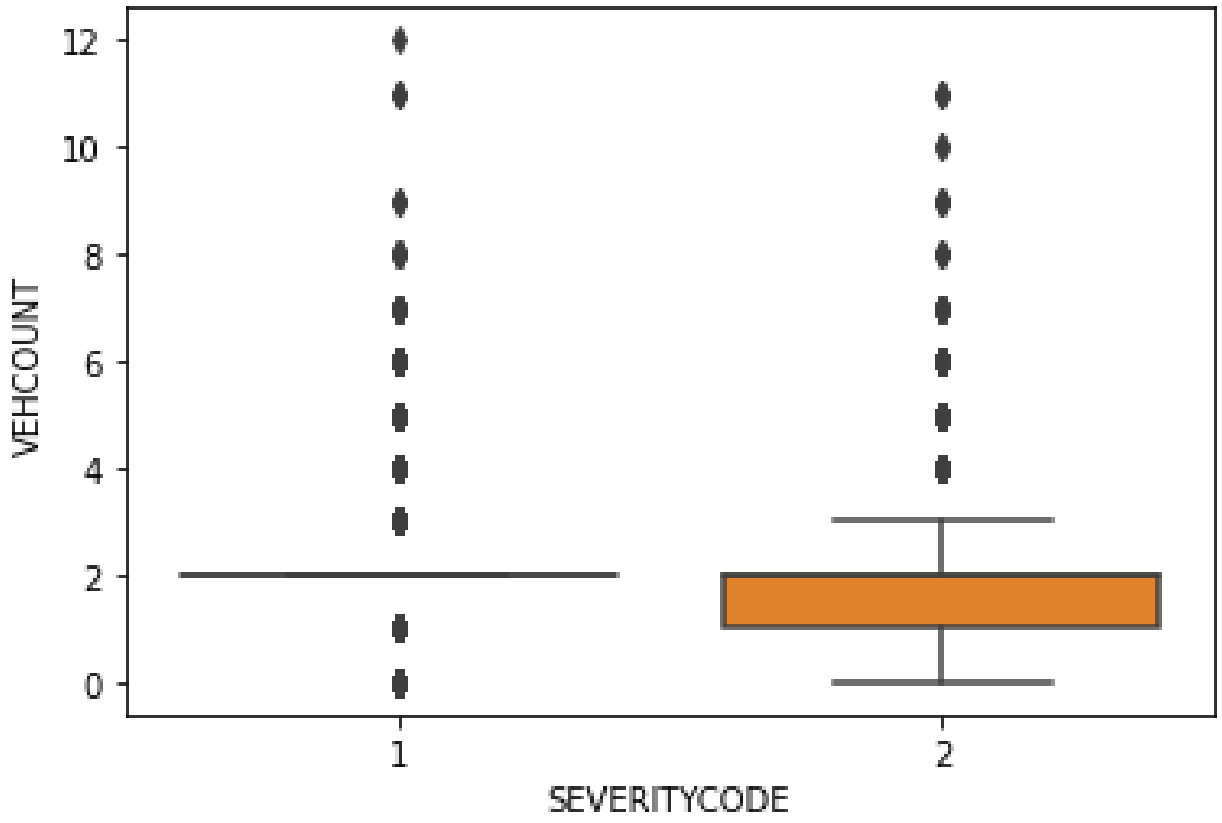
- Predicting the severity of an accident can route drivers away from the scene of a dangerous accident with heavy traffic, and predict a more efficient route
  - Builds loyalty to GPS brand
  - Allows for safer travel
  - Allows drivers to reach destination on time

DATA ACQUISITION AND CLEANING

- Source: IBM Applied Data Science Capstone
  - Seattle Collision Information
- Variables of very high correlation, duplicate variables, identifier variables were dropped from dataset
- Columns with excessive missing data was also dropped
- Missing data in columns were replaced with column modes
  - Weather related columns had missing data replaced with weather data of same day from other observations
- Remaining data was plotted against Accident Severity Code
  - Those with low variation between severity codes were dropped

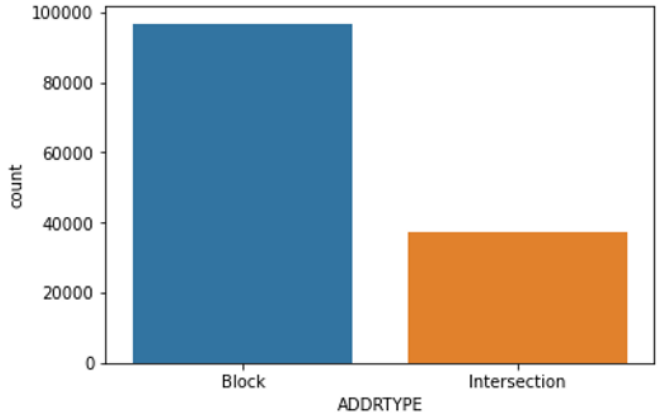
NOTABLE GRAPHS

- High variation in vehicle count in accidents with severity code 2

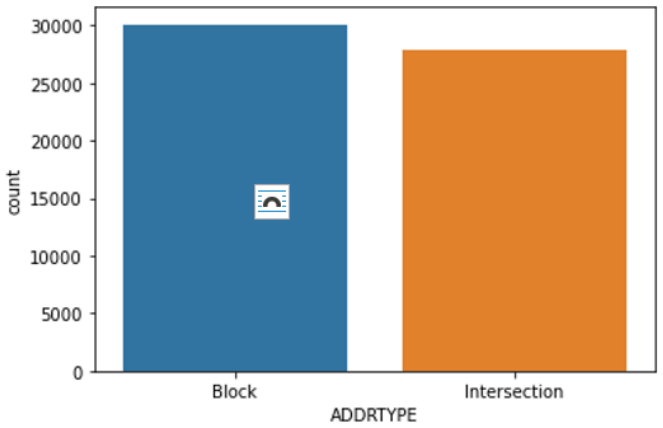


NOTABLE CHARTS

- Higher percentage of accidents involved an intersection in accidents with severity code 2



Severity Code 1



Severity Code 2

VARIABLE SELECTION AND PREPARATION

- Variables were selected based on reasonable variation between severity codes
- Classification Problem
  - Decision Trees
  - K Nearest Neighbours
  - SVM
  - Logistic Regression

DECISION TREES

- Trees of branch depth 1 to 10, 50, 100 were trained and tested on
- Focus on all Precision, Recall and F1-Scores
- Highest score all around score is from decision tree with depth 7.

Depth = 7	Precision	Recall	F1-score
Macro Avg	.67	.71	.67
Weighted Avg	.74	.70	.71

K NEAREST NEIGHBOURS

- K nearest neighbours model tested on k values from 1 to 10
- Highest scores obtained from model with k = 8

k = 8	Precision	Recall	F1-score
Macro Avg	.66	.68	.66
Weighted Avg	.72	.70	.70

SVM

- SVM Model was trained using Radial Basis Functions Kernel
- Worse overall prediction results compared to KNN and Decision Trees

SVM	Precision	Recall	F1-score
Macro Avg	.67	.70	.66
Weighted Avg	.74	.68	.69

LOGISTIC REGRESSION

- Trained model with lambda value of 0.01.
- Worse performing model compared to best Decision Tree and KNN models and SVM model

Logistic Regression	Precision	Recall	F1-score
Macro Avg	.63	.65	.63
Weighted Avg	.7	.65	.67

CONCLUSION

- Various models created to fit accident data
- Need more information for higher prediction scores
  - Vehicle information
  - Responder Information
  - Tow Truck Information
- Decision Tree is best choice for a model using this data.