APPLIED DATA SCIENCE CAPSTONE CAPSTONE PROJECT REPORT

Christopher Riper – October 2020

PREDICTING CAR ACCIDENT SEVERITY

- *Is it possible to warn you, given the weather and road conditions, how severe a car accident you could be in? Would it make you drive more carefully or possible change your travel timing or route?*

- The purpose of this project is to attempt to predict the severity of a car accident given certain conditions.

- Several potential applications for drivers, insurance companies, etc.

DATA OVERVIEW

- Seattle, WA vehicle accident data from January 1, 2004 through May 20, 2020

- 194,673 records

- Numerous features, for example:

  - Weather, light, road conditions

  - Accident location, time

  - Persons, vehicles involved

- Severity Code target variable:

  - 1 – Property damage

  - 2 – Injury collision

METHODOLOGY

- Explore data to identify potentially predictive features

- Feature engineering to simplify features used

- Use two machine learning models attempt to predict accident severity:

  - Decision tree

  - Logistic regression

- Attempt to improve models through further iteration by changing split of training/testing data by adding a feature
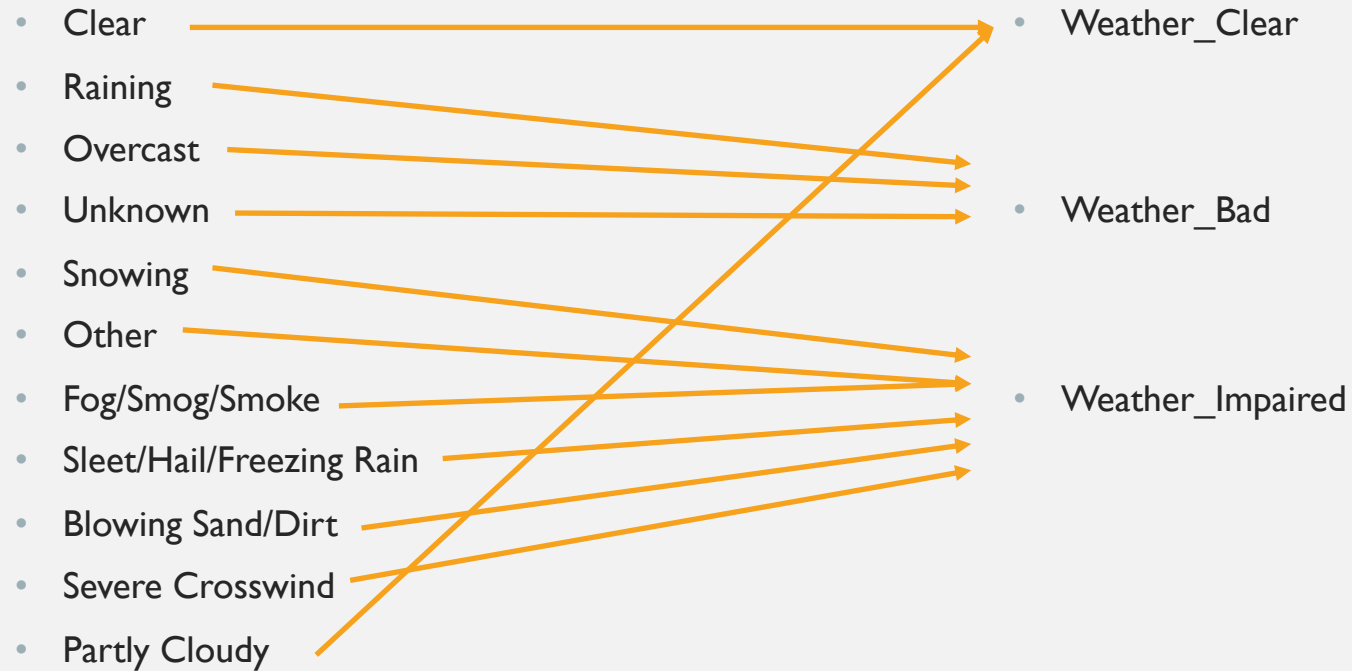
- Observe accuracy of models

RESULTS

Model	Train/Test Split	Features Used	Accuracy
Decision Tree	70%/30%	Road, Weather, Light Conditions	0.699410979076059
Logistic Regression 1	70%/30%	Road, Weather, Light Conditions	0.699410979076059
Logistic Regression 2	80%/20%	Road, Weather, Light Conditions	0.699242326955181
Logistic Regression 3	80%/20%	Road, Weather, Light Conditions, Person Count	0.700680621548735

DISCUSSION & CONCLUSION

- I was able to create a model that predicts with 70% accuracy the severity of an accident in Seattle, WA

- Very slight differences between models, even when another feature added

- There is room for improvement, I could still:

  - Explore using other features for prediction

  - Perform additional feature engineering

  - Try other machine learning models