Predicting Car Accident Severity

Seattle, Washington

Predicting Car Accident Severity is Valuable for Public Authority

- People are driving more now than they ever have been
- There are tens of thousands of car accident deaths in the US each year
- Predicting accident severity will help Public Authority with alerting the public on dangerous driving conditions
- Commuters have interest as well for planning travel times:
  - Daily commute
  - Family visits
  - Weekly activities

Data Acquisition and Cleansing

- Data for car accidents in Seattle, Washington (2004-2020) from Kaggle dataset https://www.kaggle.com/jonleon/seattle-sdot-collisions-data/
  - Data provided by Seattle DOT
- In total, 221,524 rows and 40 features in the raw dataset
- Descriptive and highly correlated features were dropped
- Cleansed dataset contains 19 features

Methodology

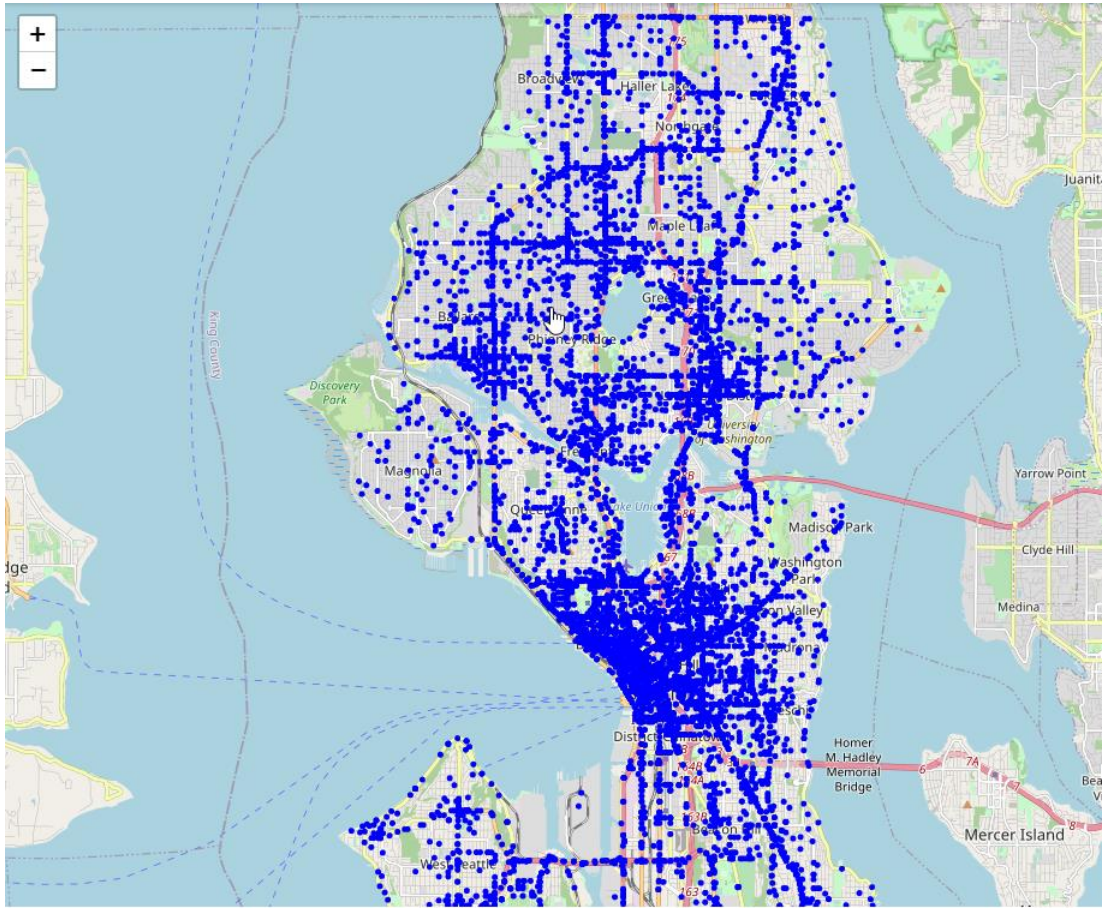
- Three ML models were chosen to perform analysis
  - K Nearest Neighbor
  - Decision Tree
  - Logistic Regression
- SVM was not chosen as it is inaccurate for larger datasets
  - Dataset had more than 180,000 rows

Results

- Comparing all models by Jaccard and F1 Scores enable a clear picture in terms of accuracy and performance
- Decision Tree Classifier is best model to use in predicting the severity of an accident based on weather

ML Model	F1 Score	Jaccard Index
KNN	0.64	0.69
Decision Tree	0.67	0.72
Logistic Regression	0.59	0.70

Most Accidents Happen in Belltown



Conclusion & Recommendations

- Models could have performed better if data was more complete
  - Balanced dataset for the target variable
  - More recorded instances
  - Minimize missing values within dataset
- The Public Authority department of Seattle can assess specific areas and develop a warning system
- Drivers can use this information for planning travel time