Predicting Car Accident Severity in Seattle

Nikhil Namburi

Predicting the Severity of Car Accidents Can Protect Seattle's Drivers

- Drivers and local officials should be able to identify conditions and types of accidents with a greater incidence of injury.

  - Drivers can exercise extra caution to avoid certain types of accidents with particularly high injury rates.

  - Local officials can plan for severe accidents under certain conditions and provide resources to assist victims of severe car accidents.

- Hopefully, the information can improve the driving condition in the Seattle area.

Data Acquisition and Cleaning

- Dataset gathered by the Seattle Police Department and Traffic Records Department

  - The dataset lists 194,673 accidents and 37 variables in those accidents.

- Included features that provide information about specific driver conditions and removed all other features.

  - 'ST_COLDESC' was used to identify specific types of accidents.

  - 'WEATHER', 'ROADCOND', and 'LIGHTCOND' were used to identify weather, road conditions, and light conditions, respectively.

 - 'SEVERITY CODE' was the dependent or target variable in this project.

Severity (Injury) Rate of Each Accident Type
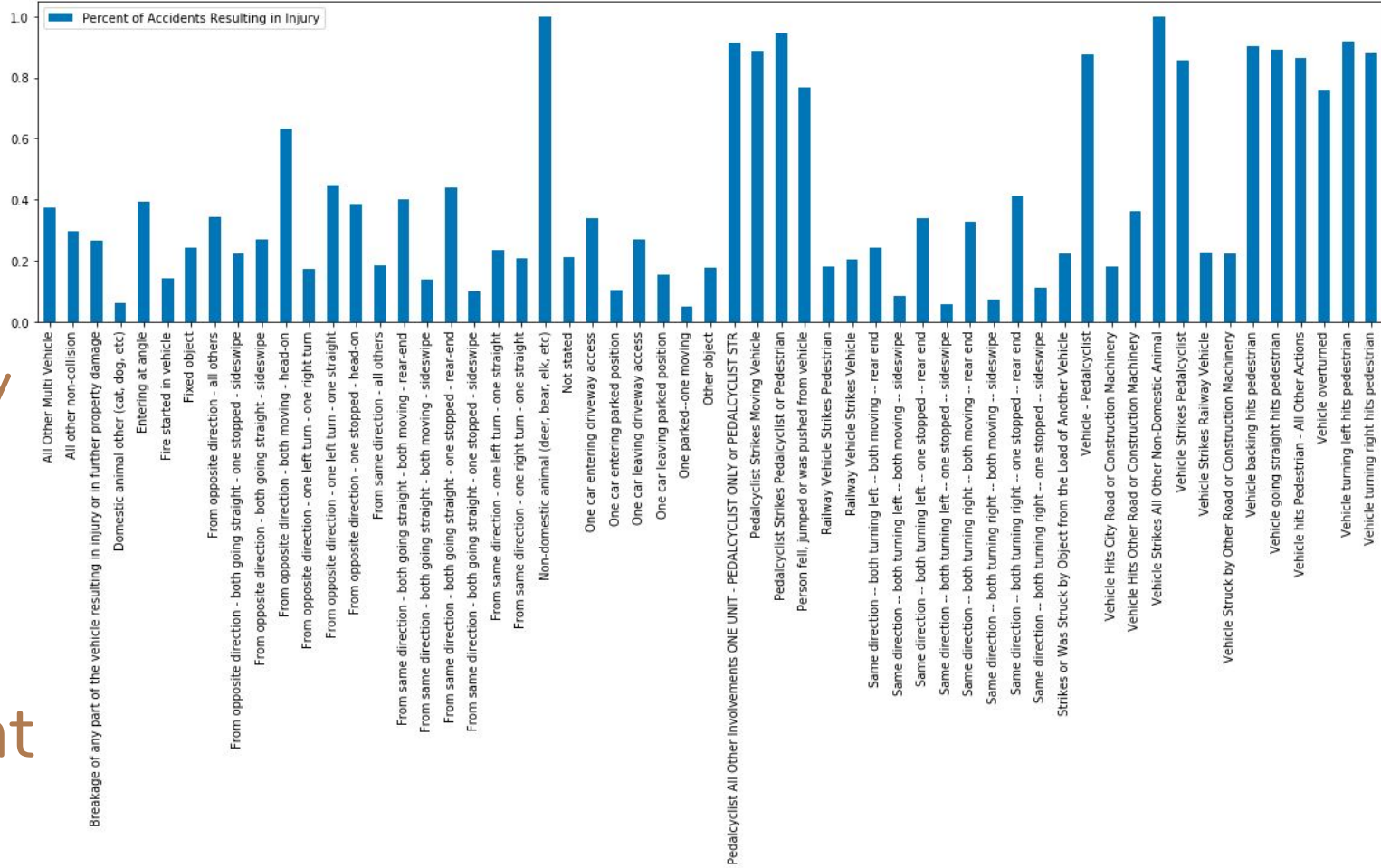
Analysis of Severity Rate by Accident Type

- Accidents involving pedalcyclists and pedestrians had a particularly high rate of injuries.

  - Of course, these individuals would not be protected by vehicles, so the rate of injuries would be especially high in such accidents.

  - Policymakers could make efforts to protect these vulnerable populations on the road.

- Drivers should be cautious of other types of accidents, such as "head-on collisions."

Severity (Injury) Rate by Weather Condition
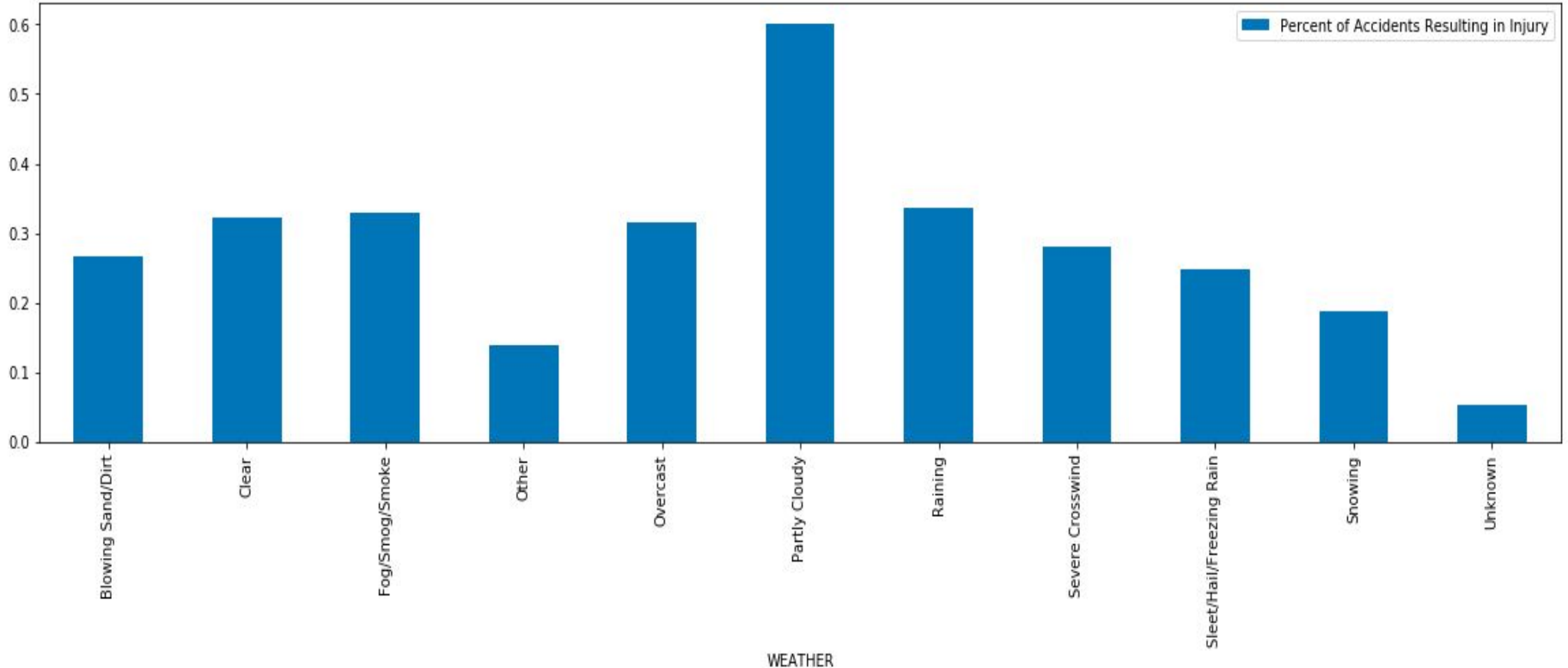
Analysis of Severity Rate by Weather Condition

- Little variation between the different types of weather conditions.

  - The only condition with a particularly different rate of severe accident is "Partly Cloudy." (Sample size small enough to make the rate of severity inconsequential.)

- Weather perceived as dangerous such as "snow" and "sleet/hail/freezing rain" actually had the lowest rate of severe accidents among all weather conditions.

  - Perhaps drivers were more cautious in poor weather conditions, and that lowered the incidence of injury in accidents.
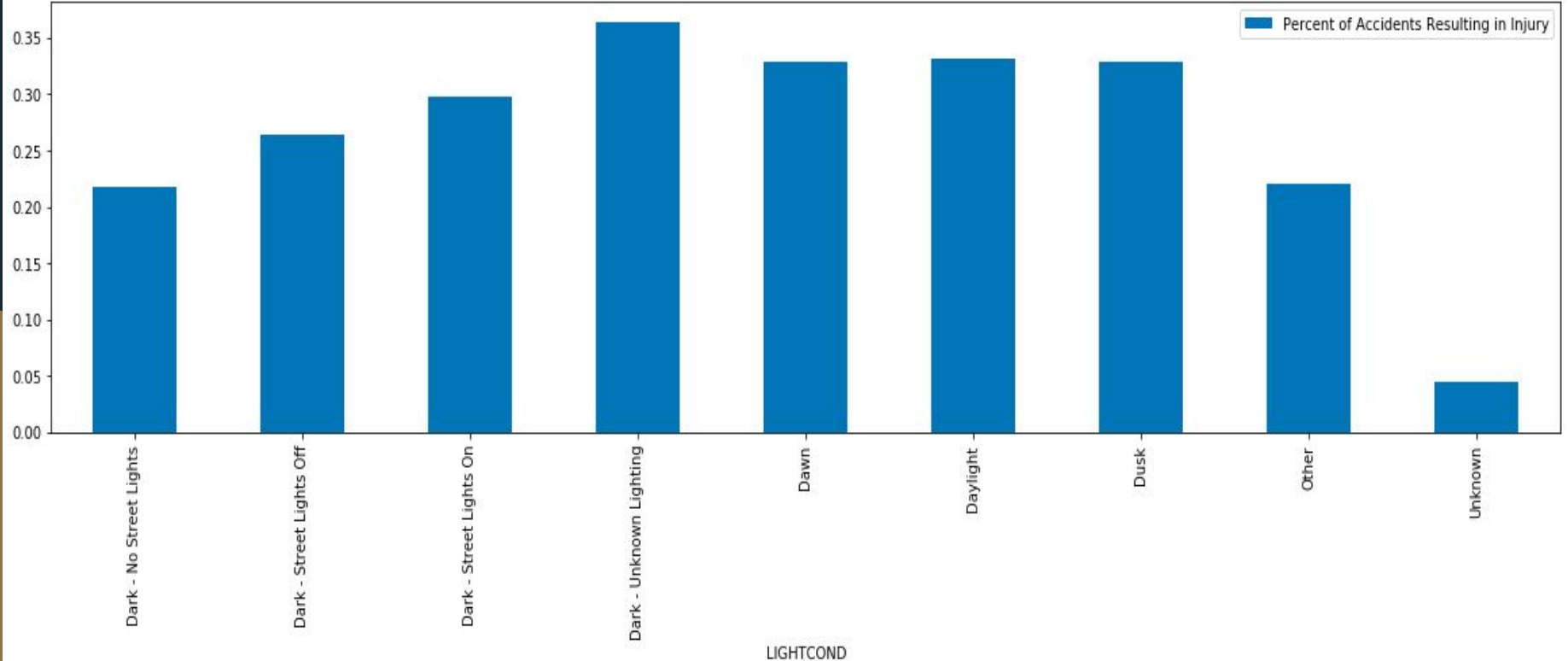
Severity (Injury) Rate by Road Condition
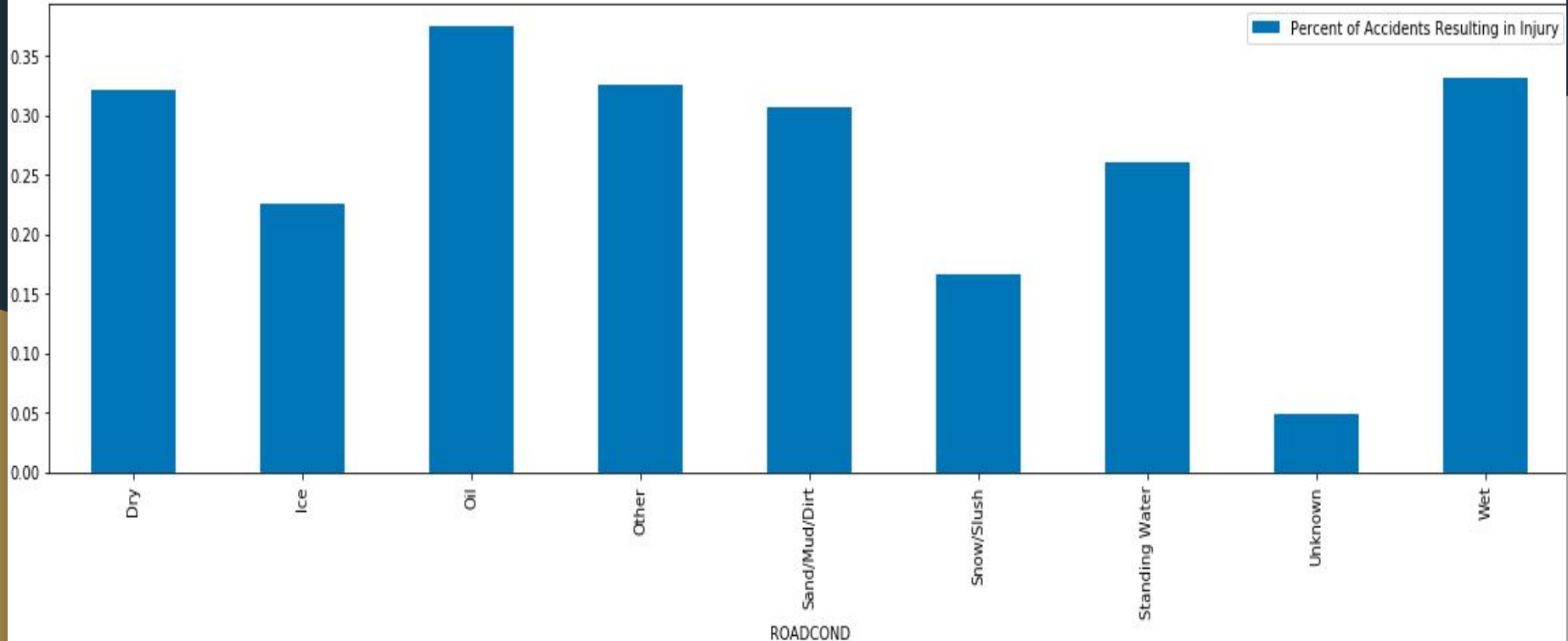
Analysis of Severity Rate by Road Condition

- Little variation between the different types of road conditions.

    - The only road condition with a particularly different rate of severe accident is "Oil." (Sample size small enough to make the rate of severity inconsequential.)

- Road conditions perceived as dangerous such as "Ice" and "Snow/Slush" actually had the lowest rate of severe accidents among all road conditions.

    - Perhaps drivers were more cautious in poor road conditions, and that lowered the incidence of injury in accidents.

Severity (Injury) Rate by Light Condition

Analysis of Severity Rate by Light Condition

- Little variation between the different types of light conditions.

  - The accident severity rate remains constant among good lighting conditions at "Dawn", "Daylight", and "Dusk".

- Road conditions perceived as dangerous such as "Dark - Street Lights Off" and "Dark - No Street Lights" actually had the lowest rate of severe accidents among all light conditions.

  - Perhaps drivers were more cautious in poor light conditions, and that lowered the incidence of injury in accidents.

Logistic Regression Model and Accuracy Measurements

Confusion matrix
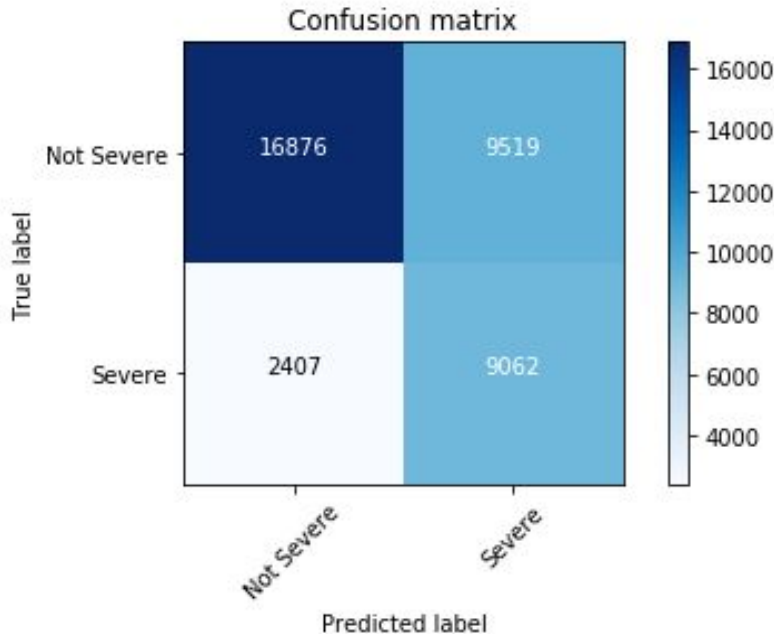
Logistic Regression Jaccard index: 0.71
Logistic Regression F1-score: 0.71
Logistic Regression Log-Loss index: 0.58

Model Summary

- A logistic regression model was applied to each variable.

  - Weather, road, and light condition did not accurately predict accident severity, and the models for these variables predicted every accident as "not severe". Accident type was an accurate predictor in a balanced logistic regression.

- Final model considered weather, light, road conditions, and accident type, with the anticipation that accident type would be the most important predictive feature.

  - Balancing the Logistic Regression Model increased True Positives.

  - False Positives also increased, but it was justified because the model otherwise did not recognize any true positives at all.

Conclusion and Further Questions

- From the exploratory data analysis conducted on each of the variables, weather, road, and light conditions are largely inconsequential.

  - If anything, poor conditions increased driver caution and decreased injuries.

- Driver behavior may be an important factor in determining the likelihood of an accident resulting in an injury. Further information could provide more insights.

- Accident type is a strong indicator of accident severity.

  - Accidents involving pedalcyclists or pedestrians are particularly dangerous, so local officials can create policy that protect these individuals.

  - Drivers should be more aware when making certain driving maneuvers that are correlated with severe accidents.