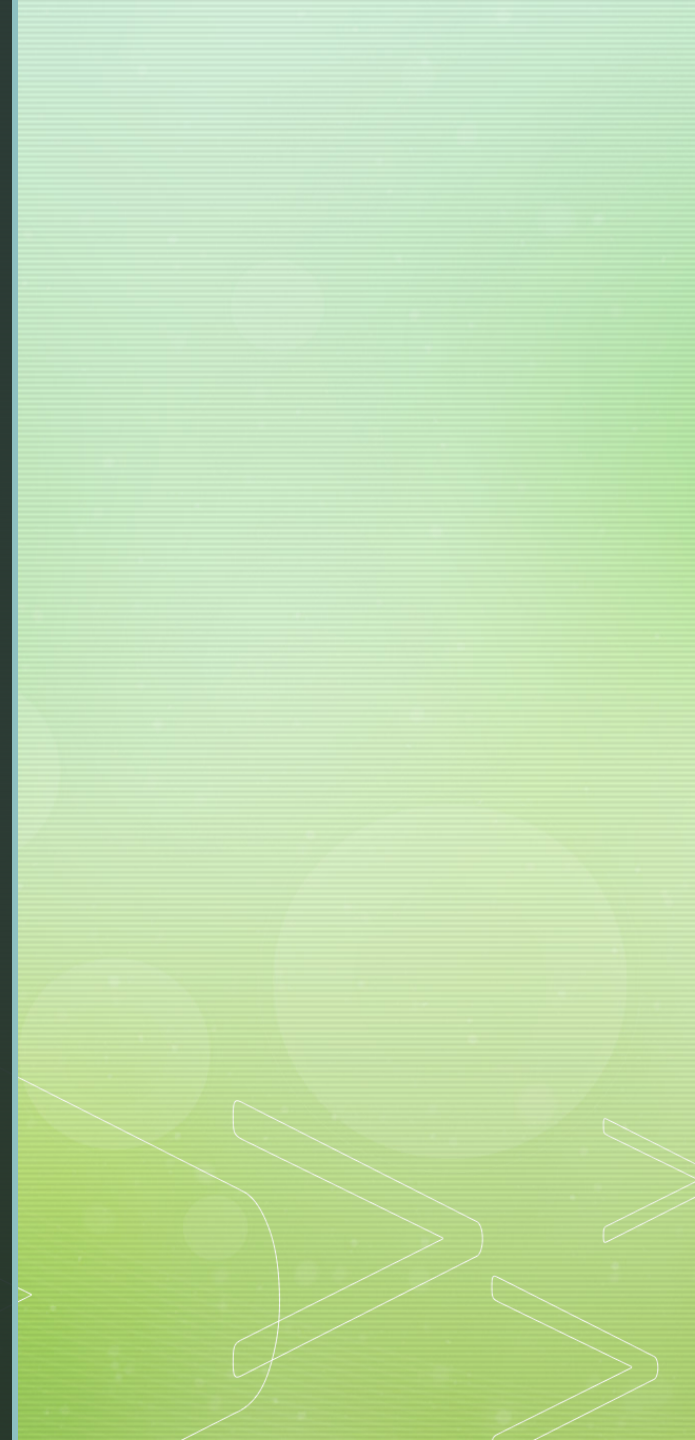Coursera by IBM Capstone

Car Accident Severity Prediction in Seattle

Background

- Car accidents are one of the biggest cause of injuries and deaths all over the world.

- They impact economy with increased

  - commuting times

  - delivery times

  - Costs

  - pollution

The Problem

- Building a model to predict car accident severity can help commuters, professional drivers or logistic planners to reduce the personal and/or business impact of car accidents

Stakeholders

- Individuals
  - work commuters
  - Taxi drivers
  - truck drivers
  - bus drivers

- ● Businesses
  - logistic companies
  - public/private passengers bus companies
  - Taxi companies

- Government agencies
  - urban/suburban mobility managers

Data sources

- Collision data from Seattle Department of Transportation Open Data Program in CSV format

- Around 200k events

Feature selection

- Collision address type

- The total number of people involved in the collision

- The number of vehicles involved in the collision.

- The date of the incident.

- The date and time of the incident.

- Category of junction at which collision took place.

- Whether or not collision was due to inattention

- The weather condition

- The condition of the road

- The light condition

- Whether or not speeding was a factor in the collision

Models

- k-nearest neighbours algorithm

- Decision Tree

- Support Vector Machine

- Logistic Regression

Evaluation

- Confronting the metrics of the different models, we look for the highest F1 score, the largest Jaccard score and the smallest log loss.

- The Decision Tree was selected

Conclusion

- The model can non be used to predict the severity of an accident

- From observable features, the stakeholders can make an informed decision