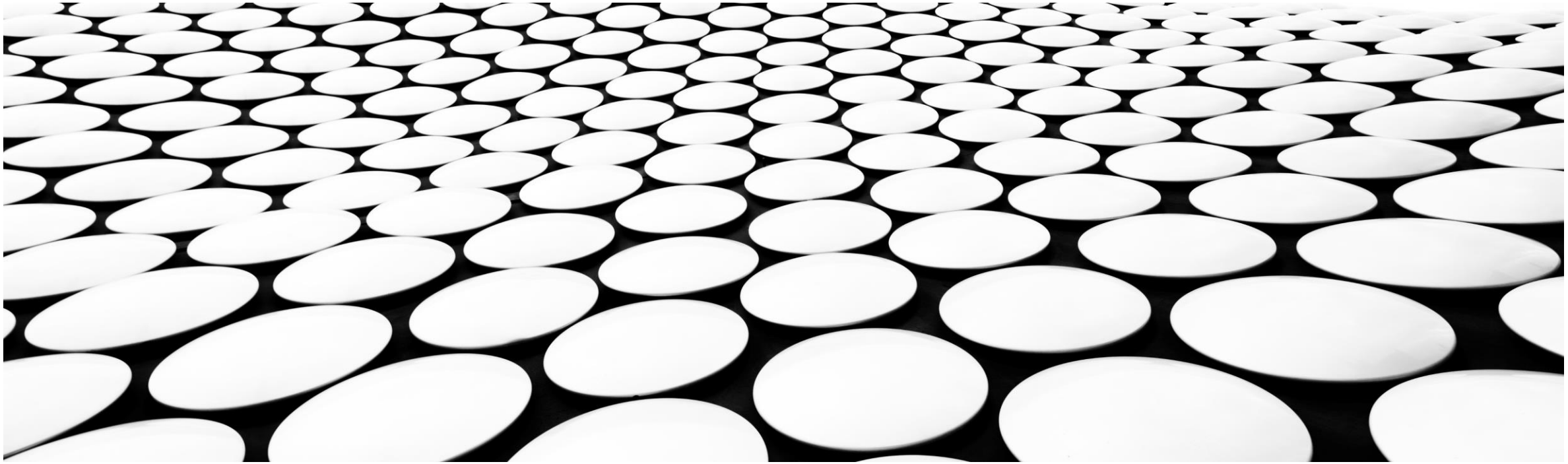
PREDICTING THE SEVERITY OF CAR ACCIDENTS

IBM DATA SCIENCE CAPSTONE PROJECT



PROBLEM

- Road fatalities are very common (Source: Budget Direct)
  - 18.2 people per 100000 people die daily
  - Daily deaths rate of 3
- Lives lost and properties destroyed
- Rapid development of data science tools today
  - Exploring the capability of analysing real-world large data sets to model car accidents is gaining extreme attention
- Stake holders
  - Public Development Authority of Seattle
  - General Public

DATA

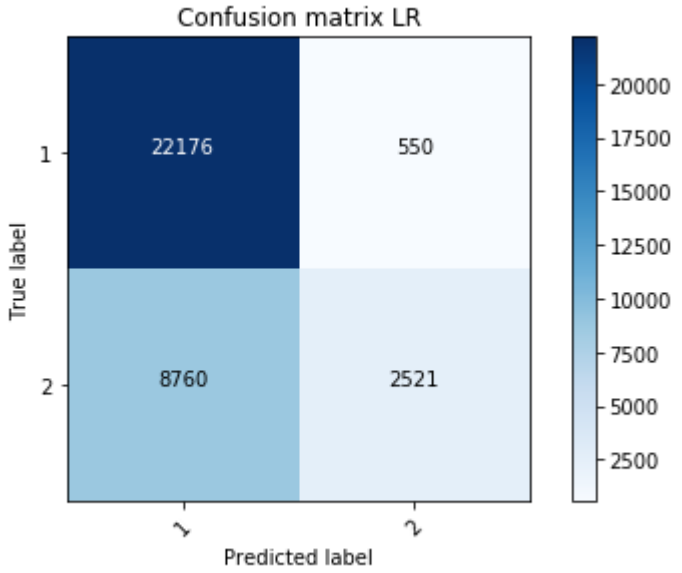
- Data storage of Public Police Department Seattle from 2014 – 2019
- 0.2 million data records, 37 feature attributes
- Characteristics of road car accidents with severity is recorded

METHODOLOGY

- Data Cleaning
  - Remove unrelated attributes
    - Location coordinates, Incident Key etc.
  - Remove attributes vulnerable to significant data missing
    - Speeding – Over 95% of values missing
- Exploratory Data analysis
  - Determine impact of features to severity
- Data Preparation
  - Convert categorical variables to numerical variables
    - One-hot encoding
- Modeling and Prediction
  - Logistic Regression

RESULTS

- Predicting Severity Category 1 accidents
  - 72% accuracy
- Predicting Severity Category 2 accidents
  - 82%



DISCUSSION AND CONCLUSION

- Most influential features
  - Weather, Road Condition, Collision Type, Number of Persons and Number of Vehicles
- Analysis is enormously helpful for authorities ()
  - To identify accident vulnerable locations/road type
  - To install new instructions/road signs
  - Implement new rules and regulations
    - i.e. new speeding limits
- Prediction helps to minimize and avoid accidents using real-time predictions
  - 77% accuracy is predicting severity using logistic regression
- Other classification techniques foe better accuracy
  - SVM, KNN, Decision Trees

THANK YOU