IBM Data Science Capstone: Car Accident Severity Report

METHAWUT APIKITMETHA

SEPTEMBER 27, 2020

Predicting the severity of an accident

- The objective of this project is to predict the severity of an accident by using several factors such as weather, road condition, light condition and so on.
- To reduce the frequency of car accident in a community. So, this predictive model can give you a warning signals to avoid the accident, to drive more carefully or recommend you to change your routing

Data acquisition and cleaning

- The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.
- In total, 194,673 rows and 37 features in the raw data set.
- Duplicate, highly similar or highly correlated features were dropped.
- Balancing target variable for data training.

3 Machine Learning

- K-Nearest Neighbor (KNN)
- Decision Tree
- Logistic Regression

Evaluations

K-Nearest Neighbor

- Jaccard index: 0.25
- F1-score: 0.50

Decision Tree

- Jaccard index: 0.21
- F1-score: 0.48

Logistic Regression

- Jaccard index: 0.24
- F1-score: 0.48
- LogLoss: 0.69

Conclusion

- Based on historical data from weather, road and light conditions pointing to certain classes, we can conclude that particular weather conditions have a somewhat impact on whether or not travel could result in property damage (class 1) or injury (class 2).