



Road Collision Severity Prediction in Seattle

*Applied Data Science
Capstone*

Elham Eslami

Introduction

In this project, machine learning algorithms have been developed to predict accident severity using Seattle collisions data. The accident severity varies based on a wide range of factors, including weather conditions, light conditions, speeding, etc. Analysis of collisions data enables us to determine the severity of road accidents under different conditions. Therefore, restrict regulations or warning systems can be designed to reduce road traffic injuries and fatalities under high-risk circumstances.

Data Description

City of Seattle has released an open source dataset containing all types of collision since 2014 to present. This dataset is updated weekly. The dataset contains 221738 rows, each of them represents a collision sample. Also, it contains 40 columns:

- 39 features (independent variables), and
- "SEVERITYCODE" as the dependent variable which is going to be predicted based on the provided features.

In this dataset, the accident severity ("SEVERITYCODE") is divided to 5 categories:

0: Unknown

1: Property Damage Only Collision

2: Injury Collision

2b: Serious Injury Collision

3: Fatality Collision

Methodology

To predict accident severity using Seattle collisions data, following machine learning algorithms have been employed to classify the accident severity as one of 5 existing categories:

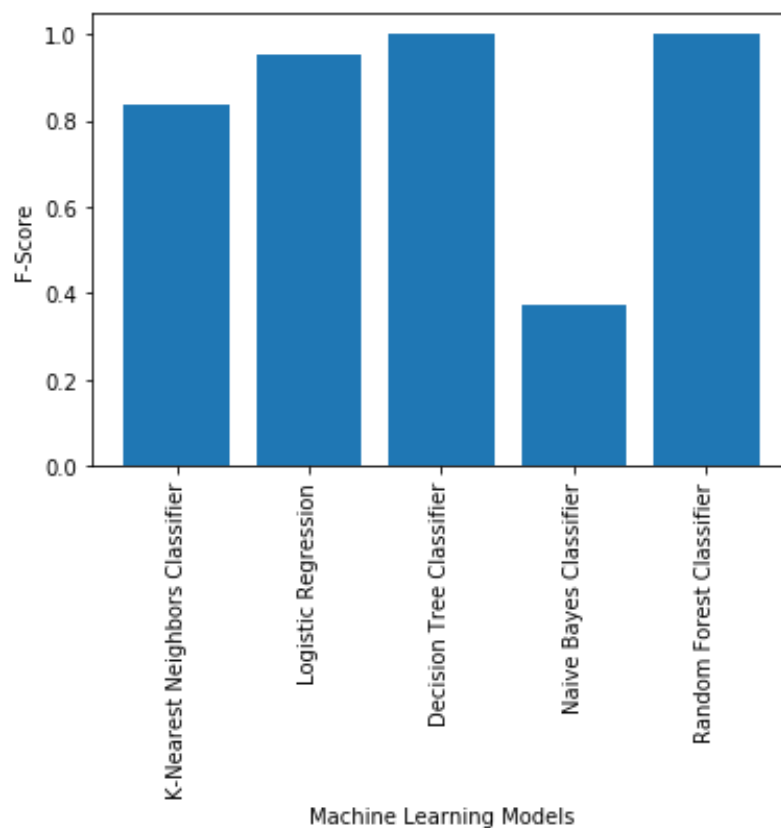
- K-Nearest Neighbors (KNN)
- Decision Tree
- Naive Bayes
- Random Forest
- Logistic Regression

The following actions have been performed to preprocess the data before feeding it to the models:

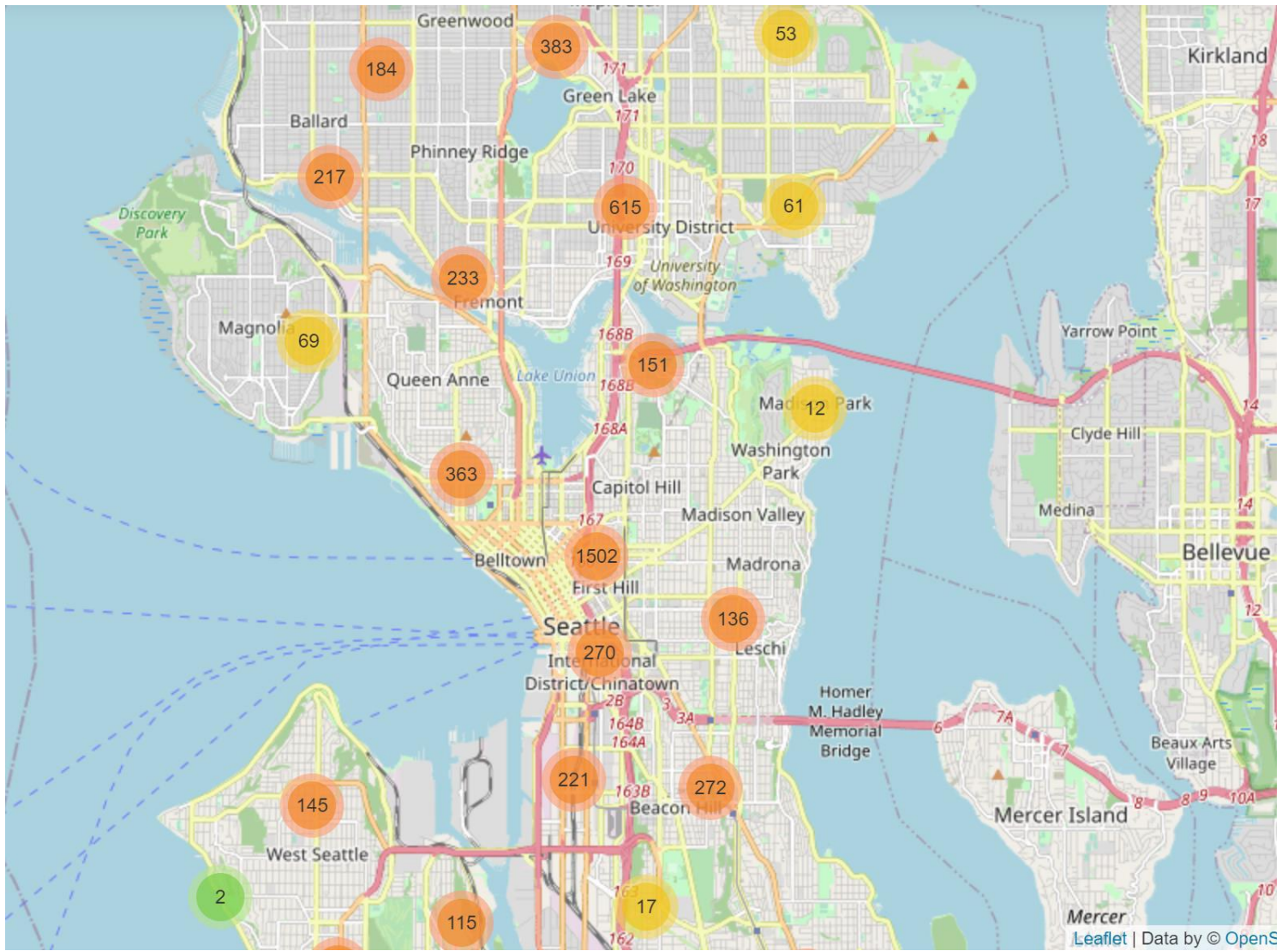
- feature selection, removing non-relevant columns to the prediction task, avoiding multicollinearity
- removing NULL or NaN values if it is needed
- converting categorical variables to numerical variables
- performing standardization to bring the features on the same scale
- balancing the data (different accident severity types have different number of samples in the dataset)

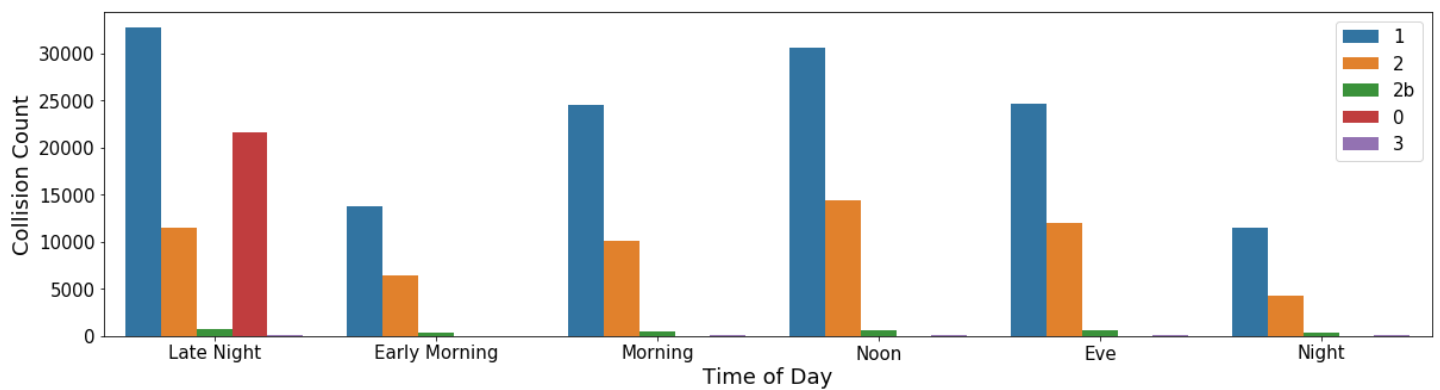
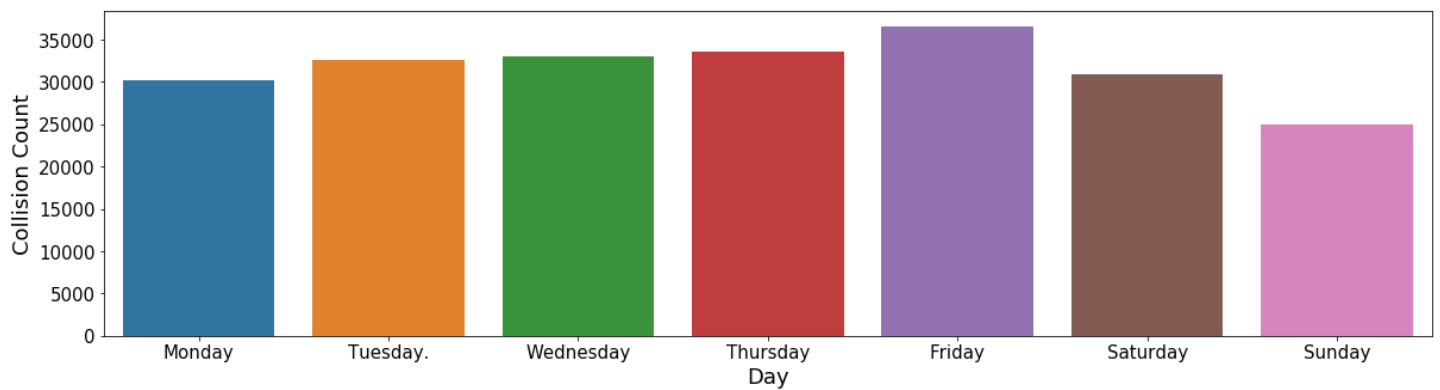
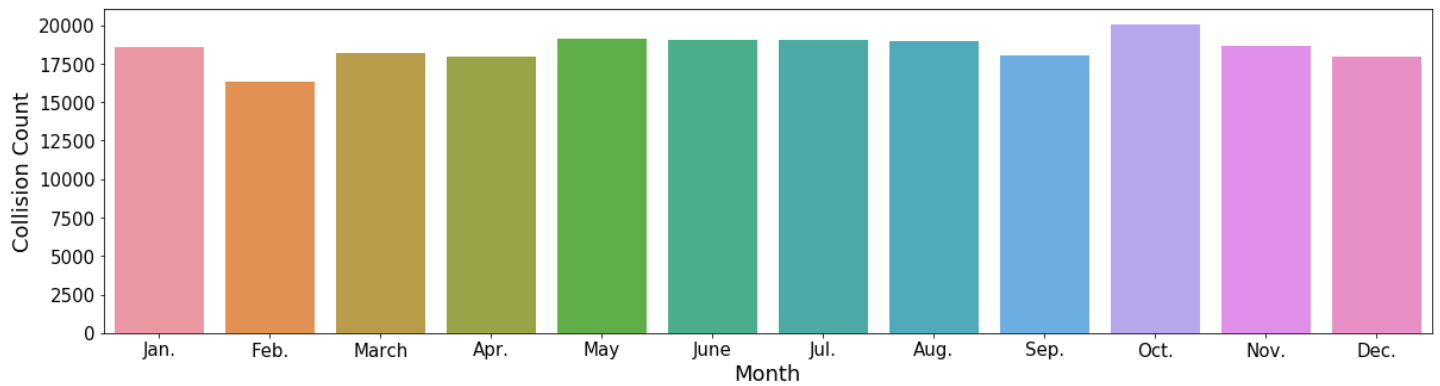
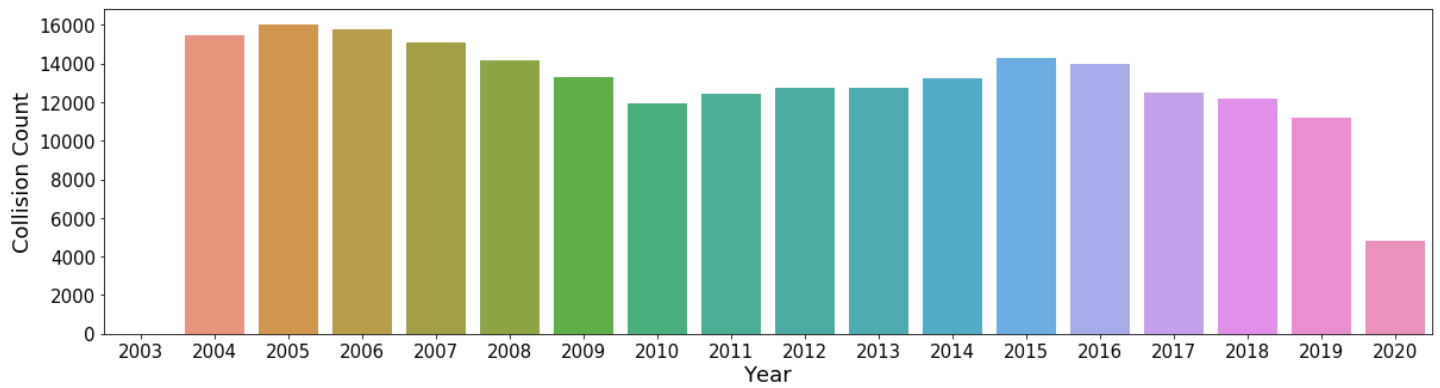
Results and Discussion

In the following, the result of machine learning algorithms and some of the features are visualized:



Location of collisions in Seattle





Conclusion

Machine learning algorithms are powerful techniques as predictive models. Using the collision severity prediction, appropriate safety guidance can be designed to decrease property damages, injuries, and fatalities.