



Fatal Car Crashes in D.C

Data 602
David House



Introduction

- Data taken from open data D.C webpage
- Goal: Predict whether a crash will be fatal
- Useful for city planners when determining where to focus efforts to improve roads



Data Exploration and Cleaning

- Many redundant values in the dataset
- Categorical values needed to be one hot encoded
- Data set is very unbalanced: used under sampling



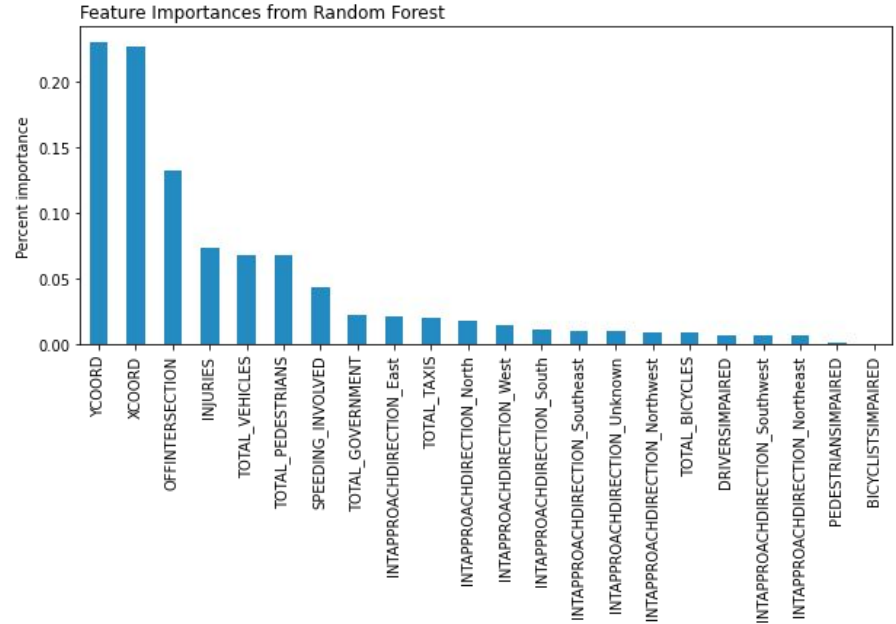
Modeling

Techniques used:

- Logistic regression
- Random Forests
- SVC
- Deep learning

Results

- Final model with the best results turned out to be random forest classifier
- Most important feature was the location of the accident





Next Steps

- Determine coordinates where most fatal accidents take place
- Communicate results to city planners
- Survey sites to determine why these accidents are so fatal