# Predicting the Severity Of Road Traffic Accidents

Saving Lifes with Data Science

# The problem

## Interest

- Emergency Services
- Local Authorities
- Government Authorities
- Civil Engineers

## Context

Accurate predictions of severity can provide crucial information for emergency responders to evaluate the severity level of accidents and implement efficient accident management procedures.

## Problem statement

Traffic accidents are a significant source of deaths, injuries, property damage, and a major concern for public health and traffic safety.

# Challenges deep-dive

## Challenge 1

**Data acquisition and cleaning:**

- Dataset includes detailed information about collisions provided by Seattle Police Department and recorded by Traffic Records.

## Challenge 2

**Balancing and Optimizing Dataset:**

- Balancing Dataset Using Oversampling Technique
- Removing High-Correlated Outliers

## Challenge 3

**Testing and Comparing Classifiers:**

- SVC
- Logistic Regression
- Decision Tree
- KNN

# Data acquisition and cleaning

Dataset includes detailed information about collisions provided by Seattle Police Department and recorded by Traffic Records.

It includes geo location, address type, weather and light conditions, time of the day and many others.

# Data acquisition and cleaning

Missing Features filled with:

- Mode Values
- ‘Unknown’ Values
- Converted from Categorical to Numeric Values
- One Hot Encoding Used for Categorical Features

# Exploratory Data Analysis

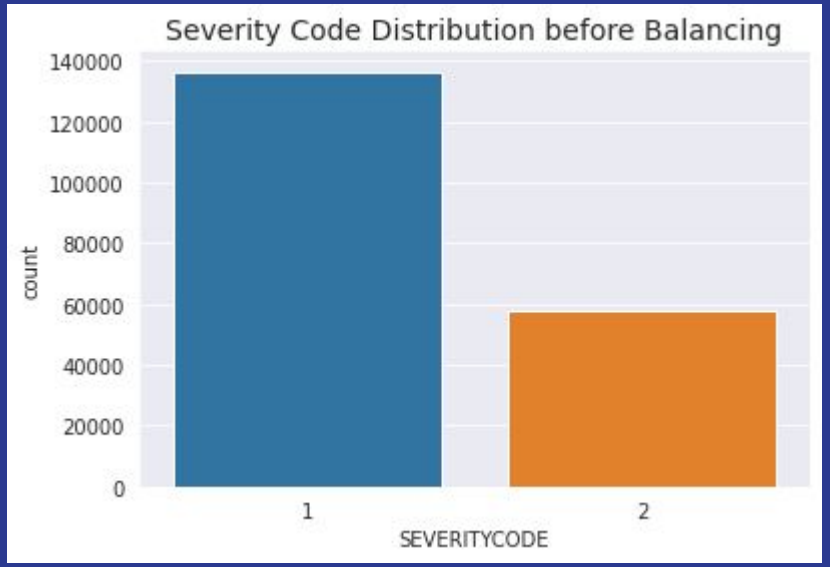
New Features Added:

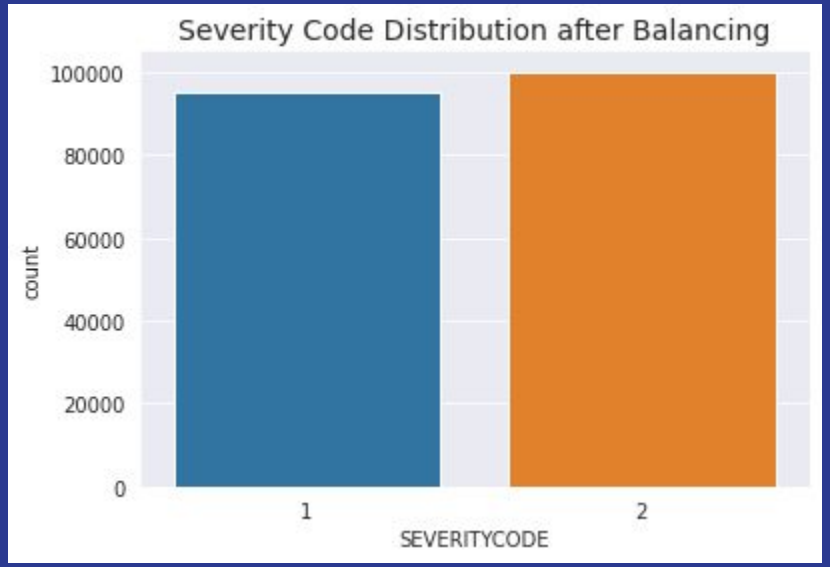
- Injuries ratio for SDOT Collision Codes
- Injuries ratio for Location
- Binarized feature for hours between 9:00 and 18:00

Outliers removed:

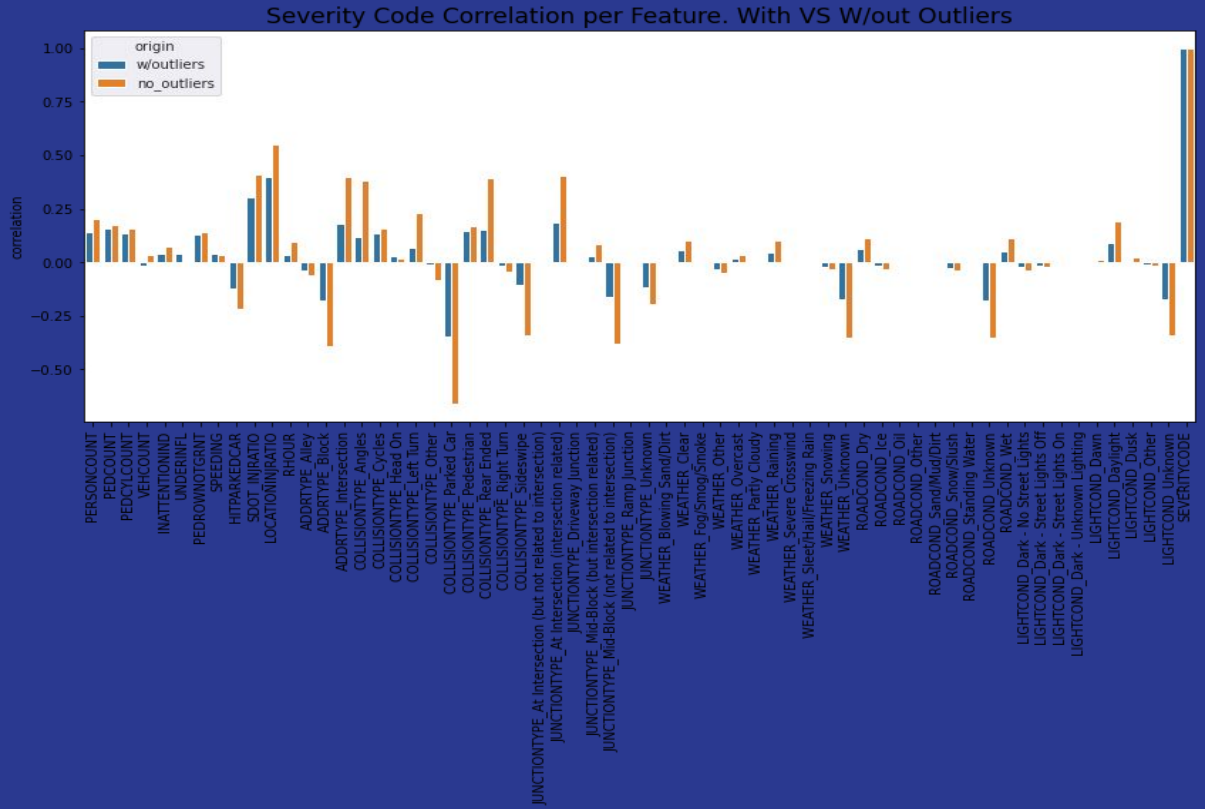
- 'Person Count' limited to 10
- ‘Vehicle Count’ limited to 5
- Number of Pedestrians and Bicycles converted to binary

# Data Balancing with SMOTE



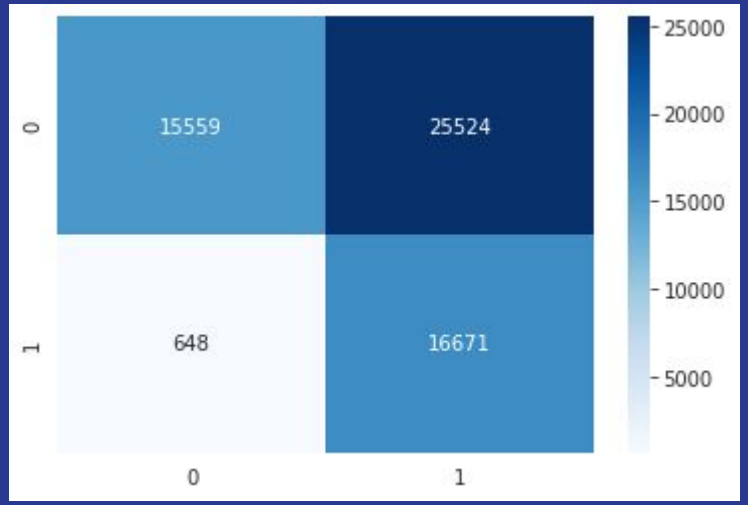
# Removing High-Correlated Outliers



# Test and Compare Classifiers

| | Train_Recall | Test_Recall | Test_Specificity |
|---|---|---|---|
| SVC_default | 0.466399 | 0.347735 | 0.347735 |
| LogisticRegression_default | 0.658 | 0.657596 | 0.657596 |
| DecisionTreeClassifier_default | 0.694 | 0.449188 | 0.449188 |
| KNeighborsClassifier_default | 0.6335 | 0.616313 | 0.616313 |

# Final Model And Confusion Matrix




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 1 | 0.96 | 0.38 | 0.54 | 41083 |
| 2 | 0.40 | 0.96 | 0.56 | 17319 |
| Accuracy | | | 0.55 | 58402 |
| Macro Avg | 0.68 | 0.67 | 0.55 | 58402 |
| Weighted Avg | 0.79 | 0.55 | 0.55 | 58402 |

# Conclusions

- In this study we analyzed relationship between different car collisions features and severity of the accident.
- We tried to optimize our model to increase injury recall and it reached **0.96**, since it's crucial to predict injuries.
- This has came at the expense of more false-alarms, which decreased the overall accuracy.
- Local and government authorities can decide how to balance the model: use it with less accuracy but higher recall, covering as many injury cases as possible, or use it with more accurate model, that will cover less accidents with injuries.