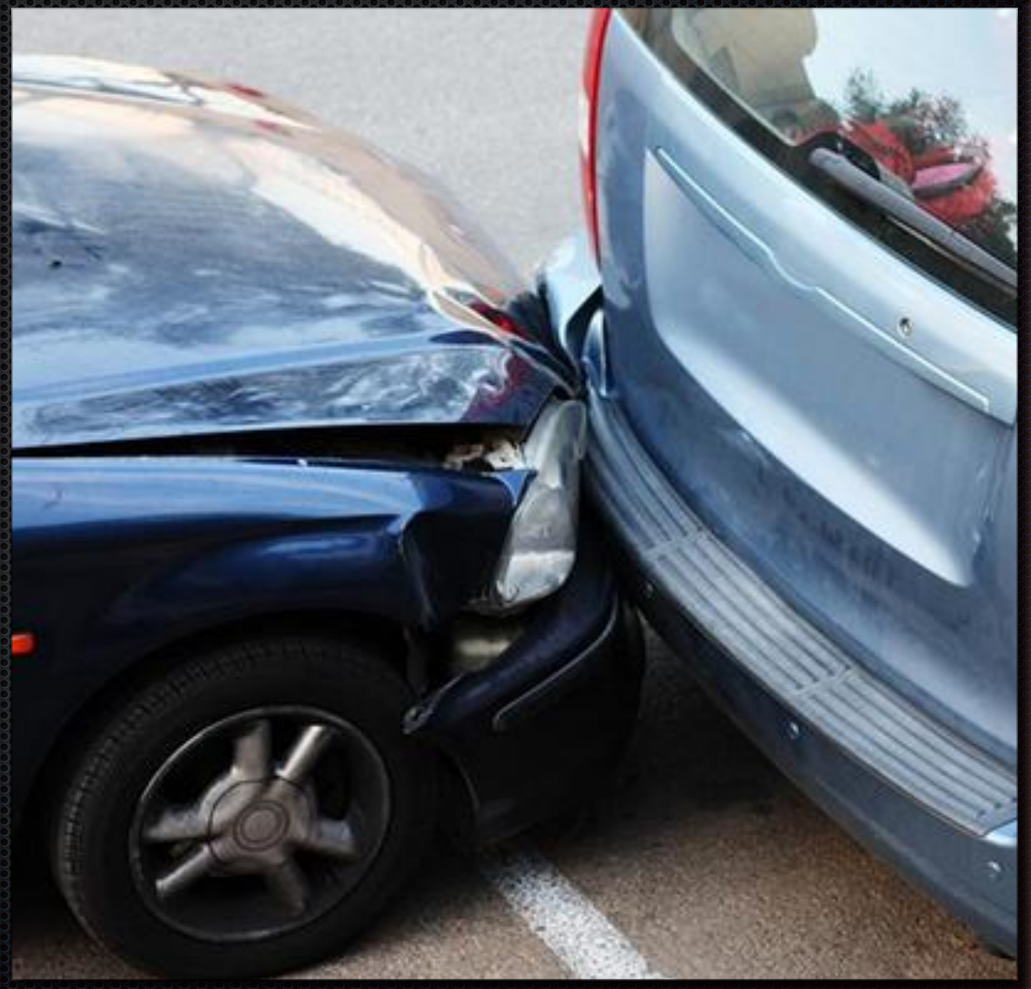

# Capstone Project - Car accident severity

# Introduction and aims of project

More than 1.2 million people worldwide die each year in road traffic accidents (RTAs) and another 20-50 million are injured. The rise in the number of fatalities in road traffic accidents is an increasingly serious problem.

Analysis of impact of bad weather conditions, lighting and road quality one car accident severity.



# Data acquisition and preparation

Accident registration data is provided by the City of Seattle and has been recorded since 2004. Data is updated weekly (https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

To predict the class of accident, the following parameters of the road traffic situation were selected:

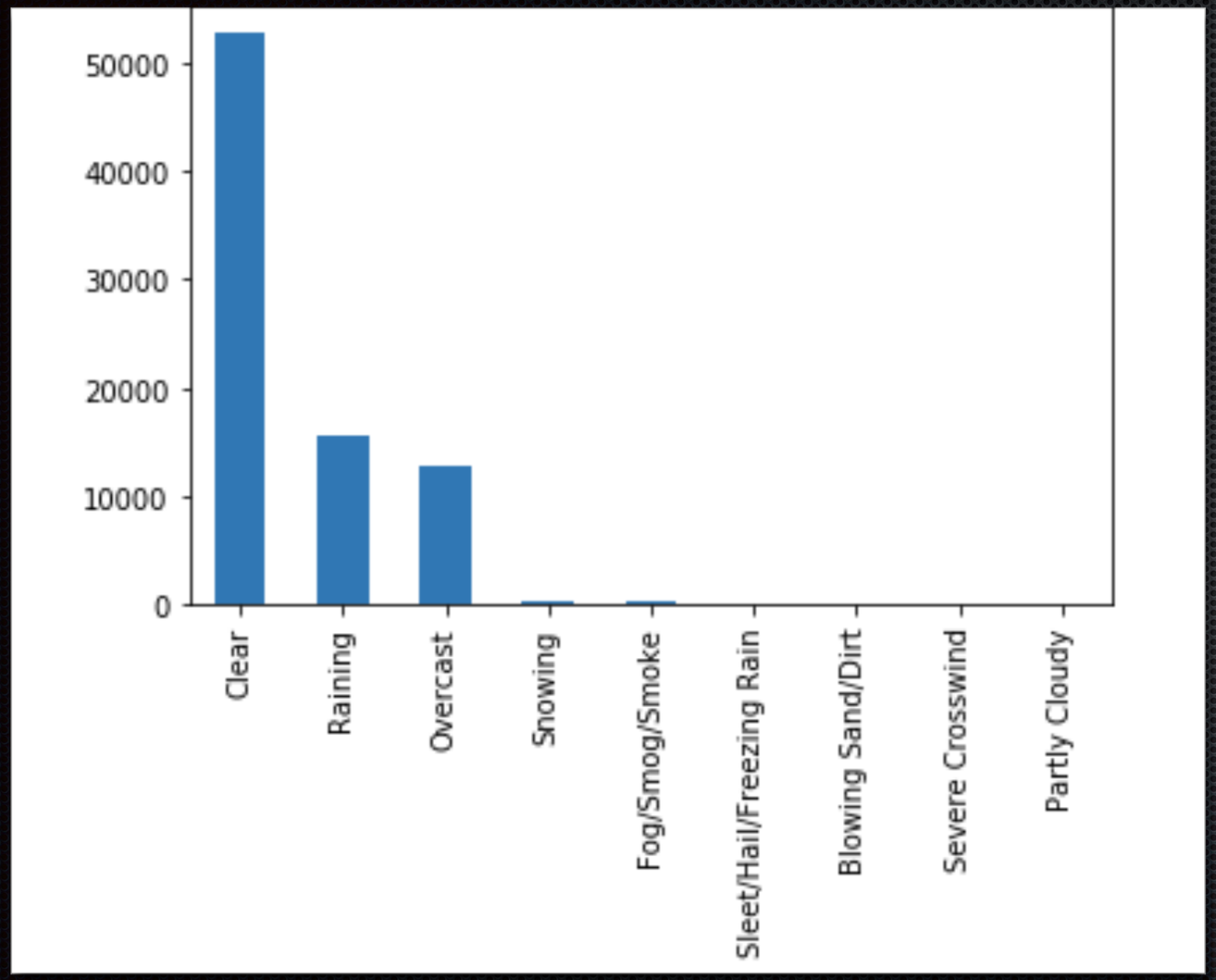
SEVERITYCODE (a code that corresponds to the severity of the collision: 2—injury, 1—prop damage);

- *WEATHER* (A description of the weather conditions during the time of the collision)
- *ROADCOND* (The condition of the road during the collision)
- *LIGHTCOND* (The light conditions during the collision)

# Data preprocessing

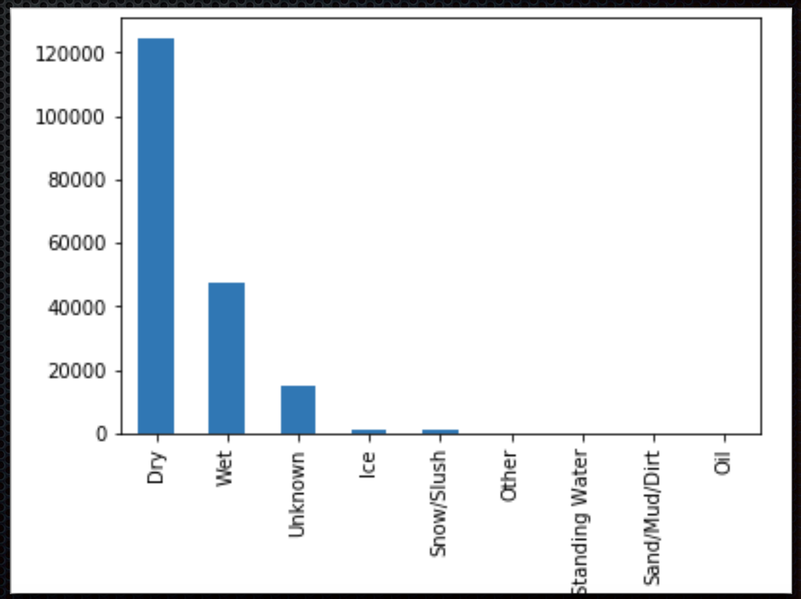
1. Remove all severity data if driver involved was under the influence of drugs or alcohol or speeding or inattention were factors in the collision.
2. Remove all rows with «Unknown» and «Other» categories from data frame.
3. Remove all NaN rows.
4. Replace text in string format which describes wether, road and light conditions with a series of binary elements.
5. Balance data frame.
6. Normalize data frame.

# Data Analysis




Number of accidents at different weather conditions

| | SEVERITYCODE | WEATHER | ROADCOND | LIGHTC |
|---|---|---|---|---|
| 0 | 2 | Overcast | Wet | Da |
| 1 | 1 | Raining | Wet | Dark - Street Ligh |
| 2 | 1 | Overcast | Dry | Da |
| 3 | 1 | Clear | Dry | Da |
| 4 | 2 | Raining | Wet | Da |
| 5 | 1 | Clear | Dry | Da |
| 6 | 1 | Raining | Wet | Da |
| 7 | 2 | Clear | Dry | Da |
| 8 | 1 | Clear | Dry | Da |



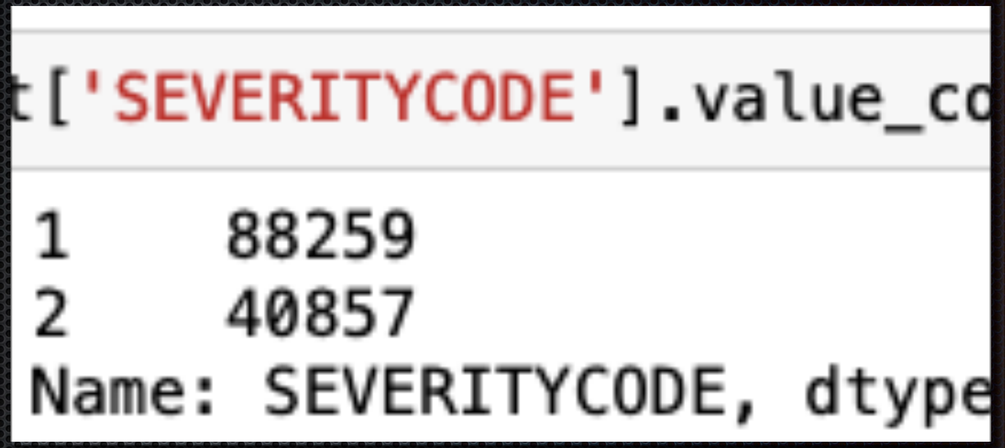

Number of accidents at different road conditions

# Data preprocessing

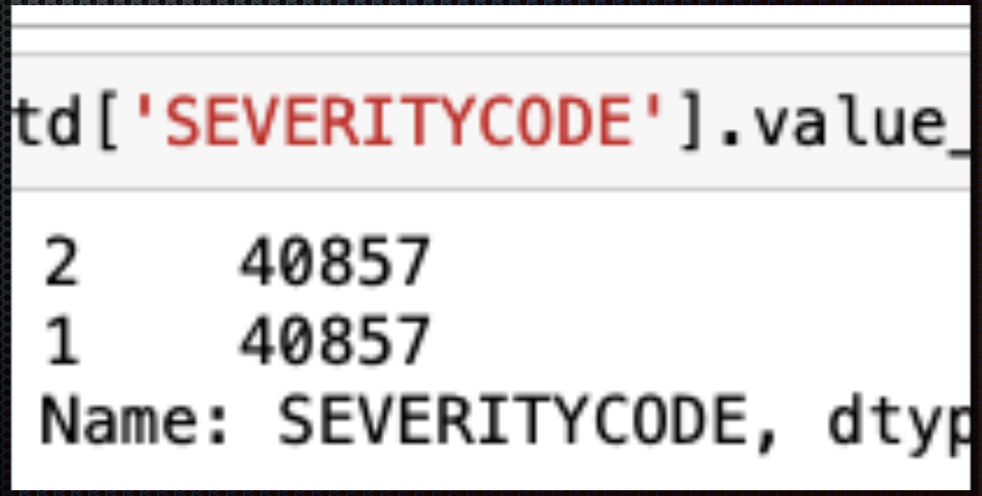
| | SEVERITYCODE | WEATHER | ROADCOND | LIGHTCOND |
|---|---|---|---|---|
| 193258 | 1 | 1 | 1 | 0 |
| 53471 | 1 | 1 | 1 | 0 |
| 115753 | 1 | 1 | 0 | 0 |
| 170248 | 1 | 0 | 1 | 0 |
| 29902 | 1 | 0 | 0 | 1 |

Data frame for Machine Learning Approaches

Text in string format which describes wether, road and light conditions with a series of binary elements are replaced by integer parameters.




Data frame before balancing




Data frame after balancing

# Methodology of prediction car accident severity

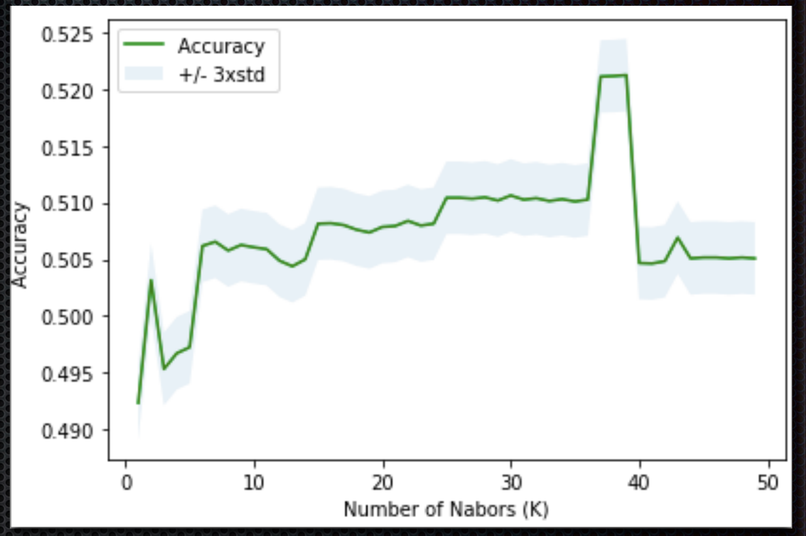
Four machine learning algorithms that are suitable for the classification tasks:

★K-Nearest Neighbours (KNN) algorithm
★Decision tree training
★Logistic regression
★Support vector machine

Data frame for Machine Learning was splitted into training data (70%) and test data (30%)

# Results

| Algorithm | Jaccard | F1-score | LogLoss |
|---|---|---|---|
| KNN | 0.52 | 0.48 | NA |
| Decision Tree | 0.52 | 0.50 | NA |
| SVM | 0.52 | 0.49 | NA |
| LogisticRegression | 0.52 | 0.50 | 0.69 |




Logistic regression model and the decision tree method have slightly better accuracy for the car accident issue.

# Conclusion and recommendations

- Machine Learning methods can be used for classification of car accident severity as function of wether, road and lights conditions.
- Based on result of ML prediction  traffic warning an/or speed limitations can be activated.
- Most car accidents occur in good weather and at dry road. That means, that those parameters as driving under influence of drugs or alcohol can have more impact on car accident severity and should be registered more carefully.
- Accuracy of the results of machine learning models in this work can be improved by further changing the parameters of the models