

# Predicting the occurrence of severe car accidents in Seattle, Washington, USA

---

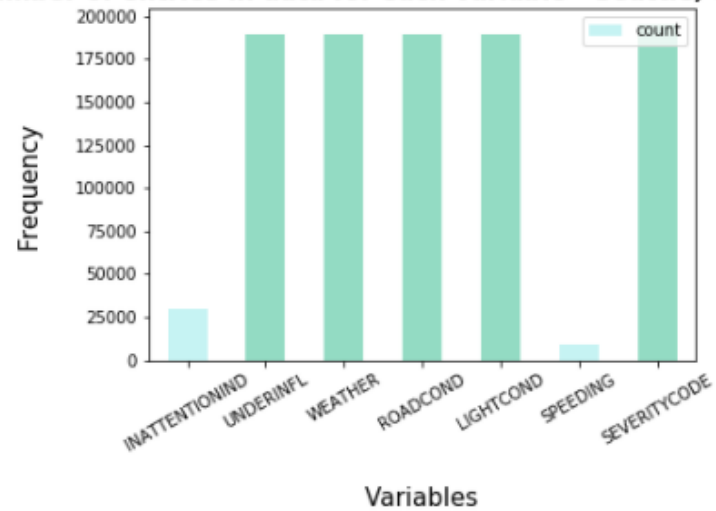
## Predicting the occurrence of severe car accidents in Seattle, Washington, USA

- Road traffic accidents and subsequent injuries have a remarkable social impact for individuals and families and are one of the leading causes of death across the globe.
- During the first half of the year 2019, 101 people were seriously injured or killed in 98 collisions on Seattle streets.
- That is the highest number of crashes in the first six months of a year since 2010, according to preliminary police reports analysed by the Seattle Department of Transportation (SDOT).
- The city of Seattle is aiming at ending traffic deaths and serious injuries on city streets by 2030
- Predicting factors that can cause severe car accidents will help fulfil the goal of ending traffic deaths
- The ultimate goal of this study is preventing or diminishing car crashes, making making Seattle's roads safer for all inhabitants, extendable to other cities across USA and the rest of the world.

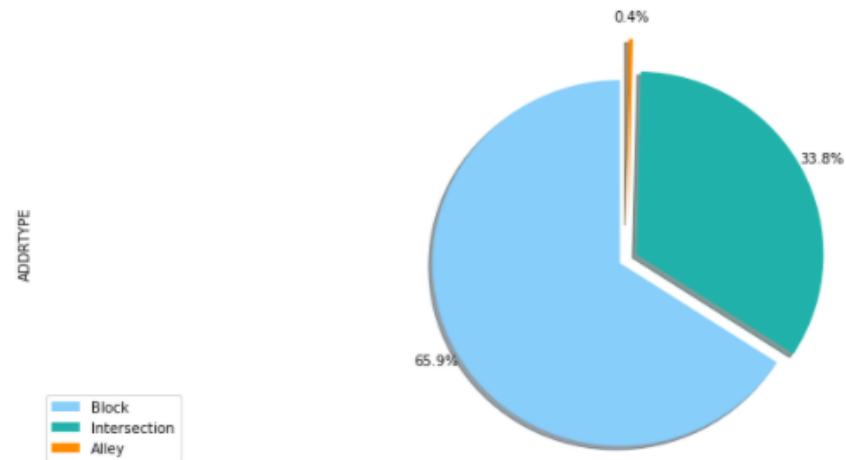
## Data acquisition and cleaning

- The data used for this study comprises a csv file and metadata from all collisions in Seattle provided by the Seattle Police Department (SPD) and recorded by Traffic Records from 2004 to present.
- Features selected to build Machine Learning models and relate to severity of car accidents were:
  - INATTENTIONIND: Whether or not the drive was inattentive (Y/N)
  - UNDERINFL: Whether or not the driver was under the influence (Y/N)
  - WEATHER: Weather condition during the collision (Overcast/Rain/Clear)
  - ROADCOND: Road condition during the collision (Wet/Dry...)
  - LIGHTCOND: Light conditions during the collision (Lights on/Dark with light on)
  - SPEEDING: Whether the car was above the speed limit at the time of collision (Y/N)

Number of entries in data for each variable - Seattle, Washington



Area of accident - Seattle, Washington





## Machine Learning models

---

- The Machine Learning models used in this study were:
- **Decision Tree Analysis:** The Decision Tree Analysis breaks down a data set into smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.
- **Logistic Regression:** Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable
- **K-Nearest Neighbor (KNN):** K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (based on distance)
- **Support Vector Machine (SVM):** The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

# Results

Althorithm	Accuracy	f-1 score	Jaccard Index
Decision Tree	0.696	0.57	0.7
Logistic Regression	0.696	0.57	0.7
KNN	0.694	0.58	0.69
SVM	0.696	0.57	0.7

Decision Tree: The criterion chosen for the classifier was entropy and the max depth was 6

Logistic Regression: The C used for regularization strength was C = 0.01 and solve = liblinear

k-Nearesy Neighbor: The best K for the model was k=4

SVM: The C used was C = 1.0

## Conclusions and Recommendations

- When comparing all the model's accuracy, we can see that they are pretty similar. Likewise, f-1 score and Jaccard Index show similar values. Logistic regression, Decision Tree and SVM have the exact value of accuracy, f-1 score and Jaccard Index. It can be concluded that the three models can be used to obtain the best performance.
- After assessing the data and the output of the Machine Learning models, a few recommendations can be made for the stakeholders. The developmental body for Seattle city can assess how much of these accidents have occurred in a place where road or light conditions were not ideal for that specific area and could launch development projects for those areas where most severe accidents take place in order to minimize the effects of these two factors. Whereas, the car drivers could also use this data to assess when to take extra precautions on the road under the given circumstances of light condition, road condition and weather, in order to avoid a severe accident, if any.