



# PREDICTING CAR ACCIDENT SEVERITY

KOROBKO  
STANISLAV

# BUSINESS PROBLEM

Road traffic accidents are a serious problem in our societies around the world. In most cases, insufficient attention while driving, drug and alcohol abuse or driving at very high speeds are the main causes of accidents, which can be prevented by adopting stricter regulations. In addition to the aforementioned reasons, weather, visibility or road conditions are the main uncontrollable factors that can be prevented by discovering hidden patterns in the data and issuing a warning to local authorities, police and drivers on targeted roads.

The predictive analysis performed here aims to analyze the severity of an accident / collision based on road conditions, lighting conditions, collision area, number of people involved, and many other factors as such. Knowing in advance the severity of any such collision will help prevent and take immediate action.



# THE IMPORTANCE OF PREVENTING CAR ACCIDENTS

Police, rescue teams, local authorities and auto insurance institutes can get useful information from the model about time, weather conditions, lighting conditions and accident severity to make important decisions to reduce the number of accidents and injuries.



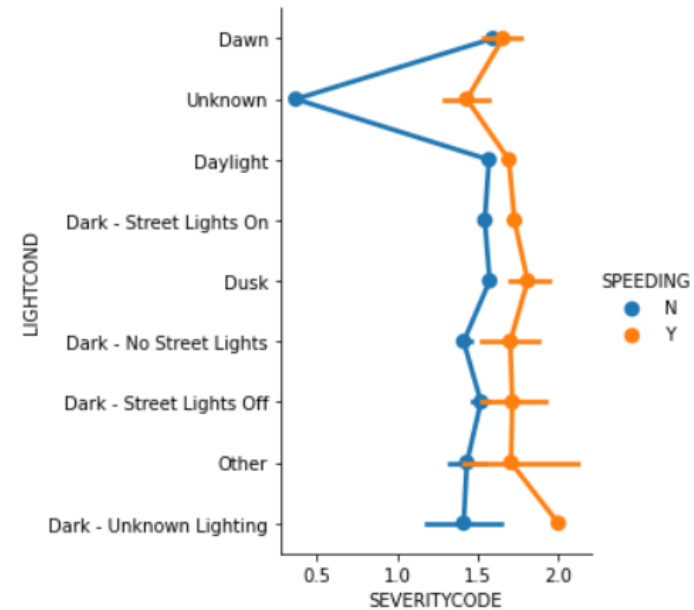
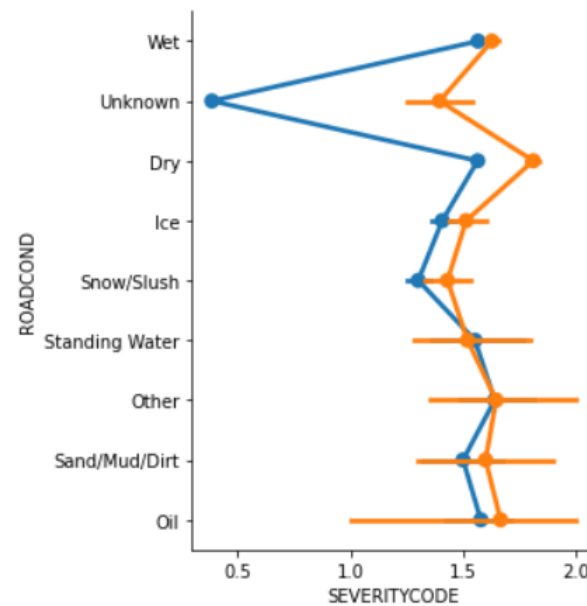
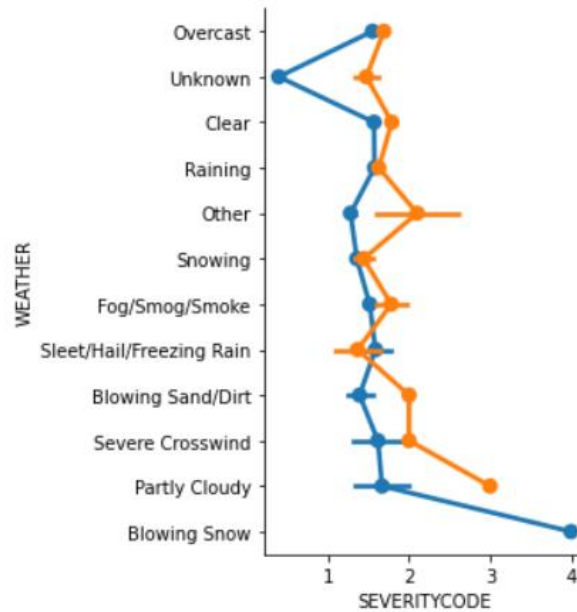
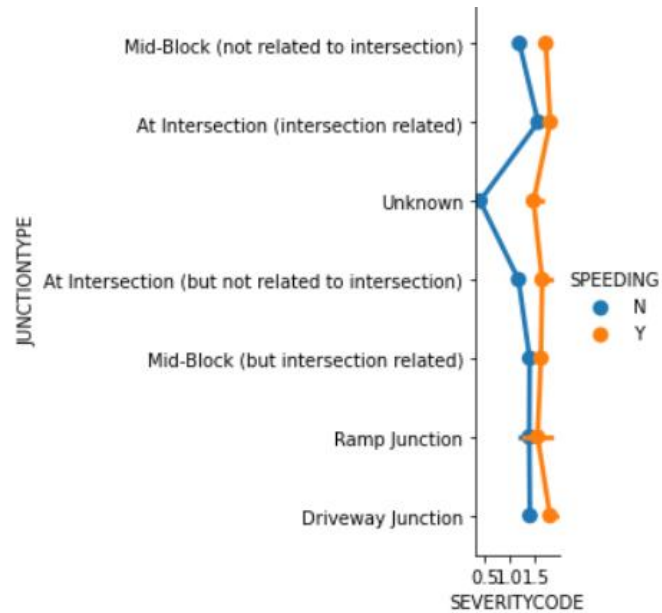
# DATA ACQUISITION AND CLEANING

- Car accident severity data for this project is taken from open source and it belongs to SDOT Traffic Management Division, Traffic Records Group.
- This dataset includes all types of road collisions.
- The data consists of 40 independent variables and 221738 rows.
- The dependent variable, “SEVERITYCODE”, contains codes that correspond to different levels of severity caused by an accident.
- Cleaned data contains 13 features and the target variable.



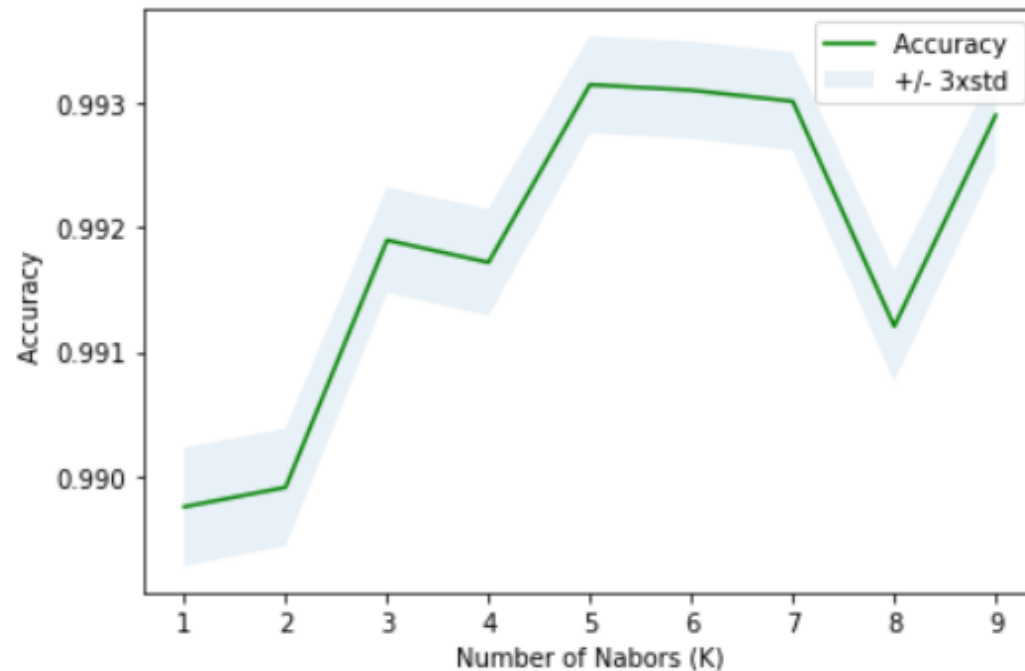


# DATA ACQUISITION AND CLEANING



# KNN MODEL FOR PREDICTING ACCIDENT SEVERITY

- KNN accuracy is best with  $K = 5$
- KNN accuracy: 0.993
- F1-Score : 0.99



# RESULTS AND EVALUATIONS

ML Model	F1-Score	Accuracy
KNN	0.99	0.993
Decision Tree	0.99	0.993
Logistic Regression	0.99	0.988

Based on the dataset provided for this capstone from weather, road, light conditions, etc. pointing to certain classes, we can conclude that particular conditions have a somewhat impact on whether or not travel could result in property damage or injury.

In conclusion, based on the work done below, the data set cannot tell us the the possibility of getting into a car accident, but can tell us that if we are in an accident in certain conditions how severe it would be.

