Coursera

CAR ACCIDENT SEVERITY PREDICTION

Applied Data Science Capstone Project

BUSINESS PROBLEM

- In an effort to reduce the frequency of car collisions in a community, an algorithm must be developed to predict the severity of an accident given the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.

DATA

- We chose the dataset provided by the Seattle Department of Transportation Traffic Management Division with 194673 rows (accidents) and 37 columns (features) where each accident is given a severity code. It covers accidents from January 2004 to May 2020. Some of the features in this dataset include and are not limited to Severity code, Location/Address of accident, Weather condition at the incident site, Driver state (whether under influence or not), collision type. Hence we think its a good generalized dataset which will help us in creating an accurate predictive model.
- Data used for our purpose is downloaded from following link :

  https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

EDA

STATISTICS

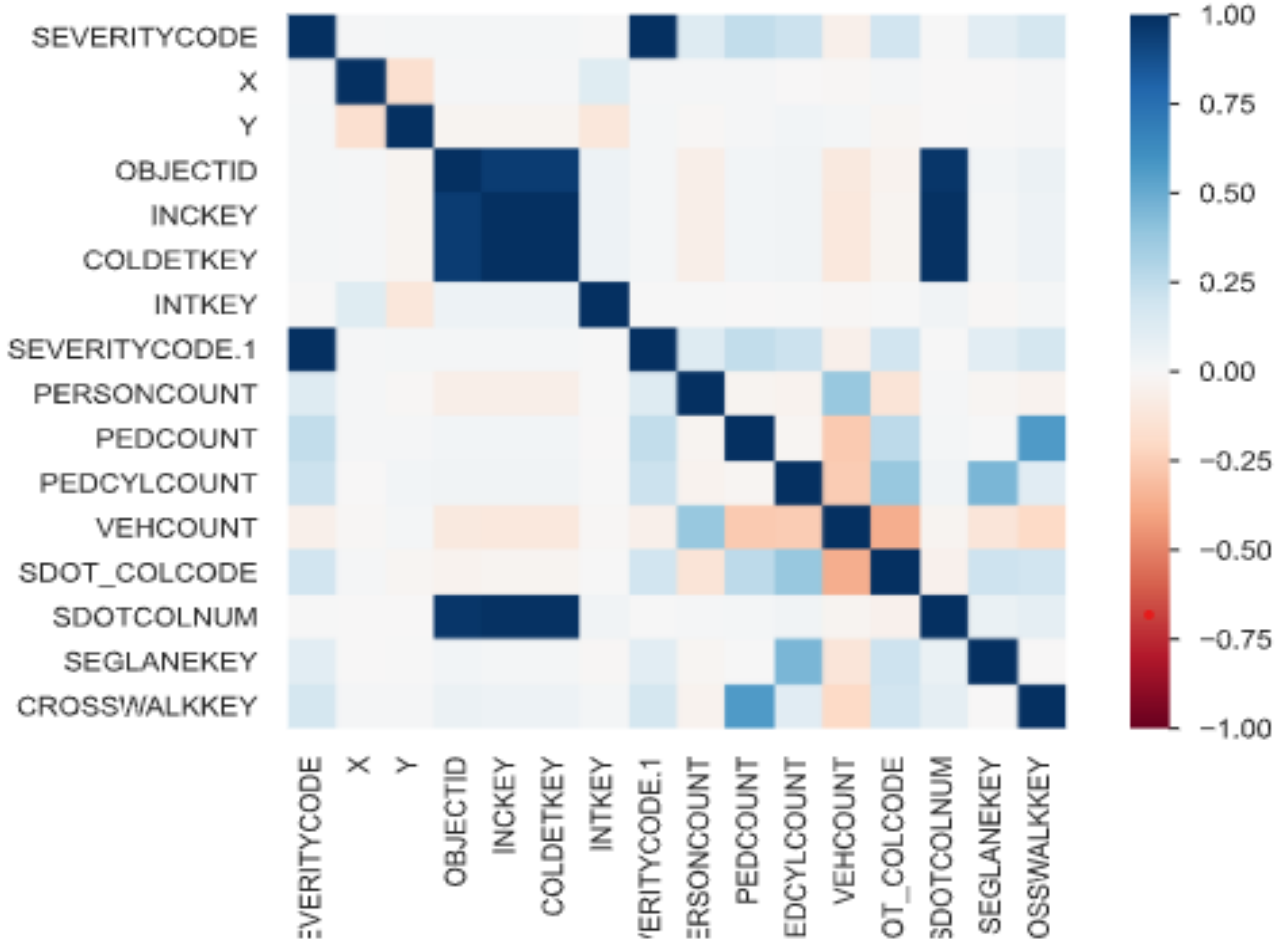
Dataset statistics

Number of variables	38
Number of observations	194673
Missing cells	1100024
Missing cells (%)	14.9%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	56.4 MiB
Average record size in memory	304.0 B

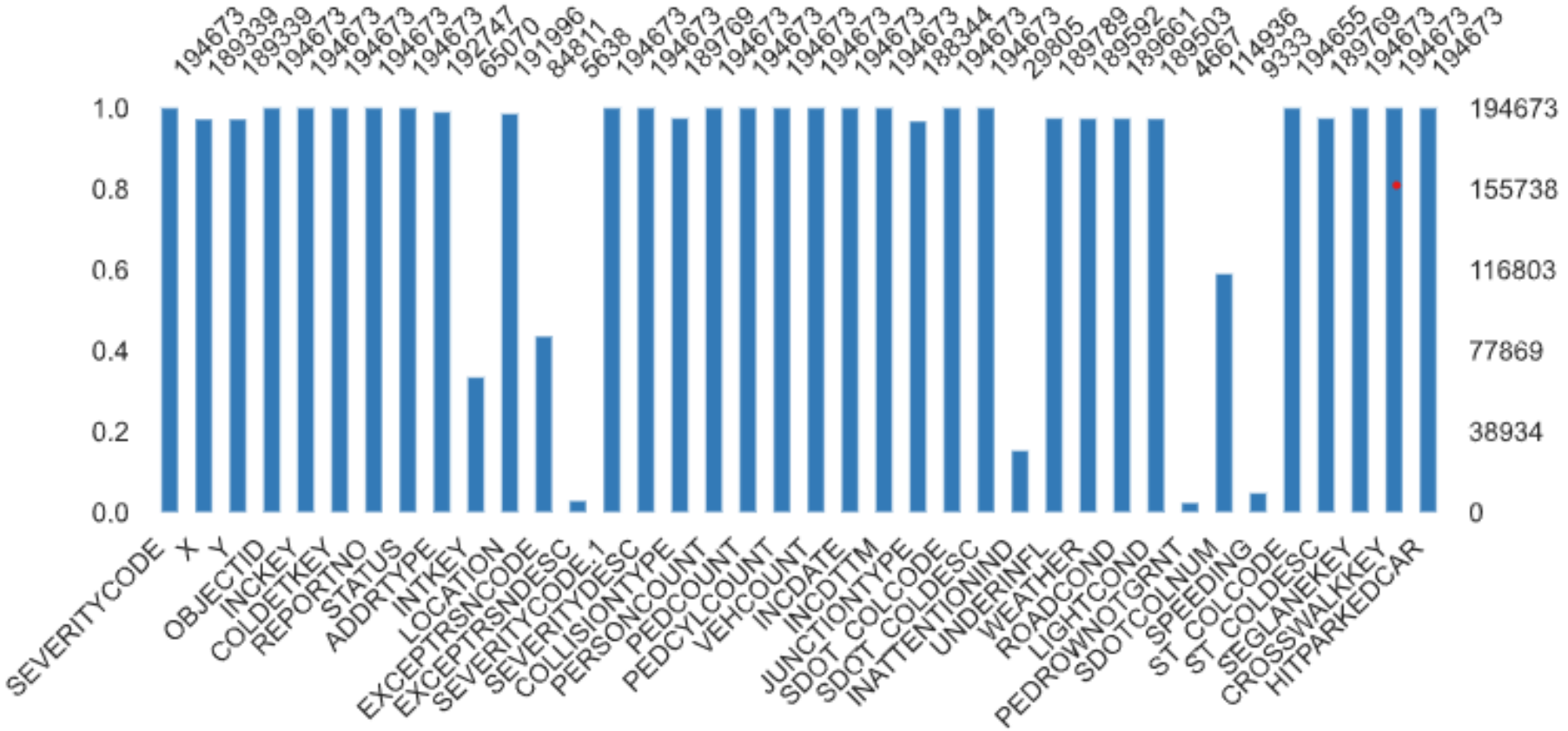
Variable types

CAT	23
NUM	13
UNSUPPORTED	1
BOOL	1

CORRELATION



MISSING VALUES



Machine Learning Algorithms

We will apply following classifiers:

- Logistic Regression
- Support vector machine
- Naïve bayes
- KNN
- Decision tree classifier
- Random forest classifier

CONCLUSION

- Now we can compare all of the above models and can choose model with maximum accuracy. 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 care.