Introduction/Business Understanding

Traditionally accident prevention has been focused on making the driver more aware of the road conditions given the current weather. I believe that there may be a more effective way for cities to prevent severe accidents. What if a city was able to setup a traffic system that would automatically lower speed limits, extend yellow and red light times, turn on street lights, and reroute traffic around specific roads during inclement weather in real time. When conditions are approaching those which are common during severe accidents specific actions will be taken to lower the risks. For example when the weather is overcast, road conditions are wet on road x the system would lower the speed limit and turn on the street lights to lower the risk of an accident.

Data Understanding

We will only need to look at the variables which are common to the accidents, and not controlled by the driver. So, we will look at the WEATHER, ROADCOND, and LIGHTCOND. We will use the SEVERITYCODE to know if action needs to be taken. If accidents at intersections are common when it’s overcast, then we would extend the time for the yellow are red light as well as turn on surrounding street lights. If accidents on a section of road are common when it’s overcast and raining, then we would lower the speed limit and turn on the street lights.

Methodology

Extract Data and Convert

The original data is not usable for our needs in the current state. So, the columns not needed will be removed, and the remaining columns will need to be numerical values. I converted the WEATHER, ROADCOND, and LIGHTCOND to numerical values, and added new columns. The target class of SEVERITYCODE was also downsized to be perfectly balanced.

Now that the data is ready we will use the following models

K-Nearest Neighbor

Will allow us to predict the severity code of an outcome by finding the most similar data point within k distance.

Decision Tree

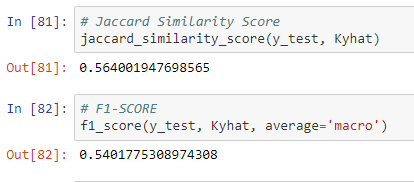
A decision tree model gives us a layout of all possible outcomes so we perform a full analysis of the decision based on the given variables.

Logistical Regression

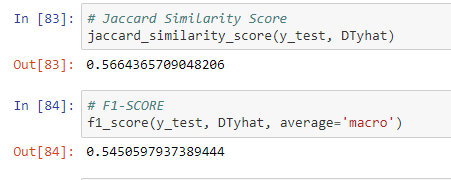
Since our dataset has two severity codes, our model will only predict one of those two classes. This makes our data binary, which is ideal to use with logistic regression.

Results

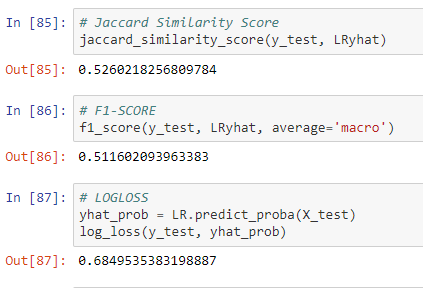
KNN



Decision Tree



Logistic Regression



Discussion

At the start we converted the categorical data into numerical data in order to continue with the process. The next step was to balance the target data value as the severity code 1 was much larger than severity code 2.

Once the data was cleansed we used the KNN, decision tree, and logistical regression models. Evaluation metrics used to test the accuracy of our models were jaccard index, f-1 score and logloss for logistic regression. Choosing different k, max depth and hyparameter C values helped to improve our accuracy to be the best possible. From this it was concluded that the logistical regression model was ideal for this due to its binary nature.

Conclusion

Based on the historical data it is possible to create such a system where a simple action such as lower the speed limit or turn on street lights is possible when weather, road condition and light conditions are analyzed.