

Predicting Severity of Vehicle Accident

Capstone Project

IBM Course on Coursera

Time is the critical variable.

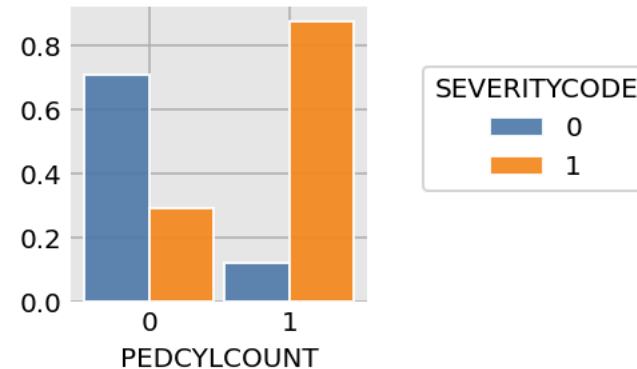
- If a road accident information can be used to predict the severity of the accident before emergency personnel arrive at the accident site
- Such a system can save lives if the authorities and the care givers are equipped with such predictive power.
- This system will not only save lives, but also a proper allocation of resources can be implemented.
- This will improve the rating of the city, and thereby more businesses and people would like to stay in the city.

Data Acquisition and Cleaning

- Seattle Police Department (SPD) generates the data
- The is updated every week since 2004
- A lot of metadata and index columns are part of the data
- There are missing values in almost every variable, and was able to accommodate missing values into the model
- From 38 variables that are available in the dataset, I was able to capture 13 variables to be used in the model.
- Had to do cleaning, binning and in some cases, re-grouping.

Pedestrians and Pedal Cyclists are at Risk

- The average injury rate is 30% but if a pedestrian is involved in the accident, then the injury rate is 90%
- Similarly if a cyclist is involved then there are 88% chance of being injured



Logistic Regression for a Classification Problem

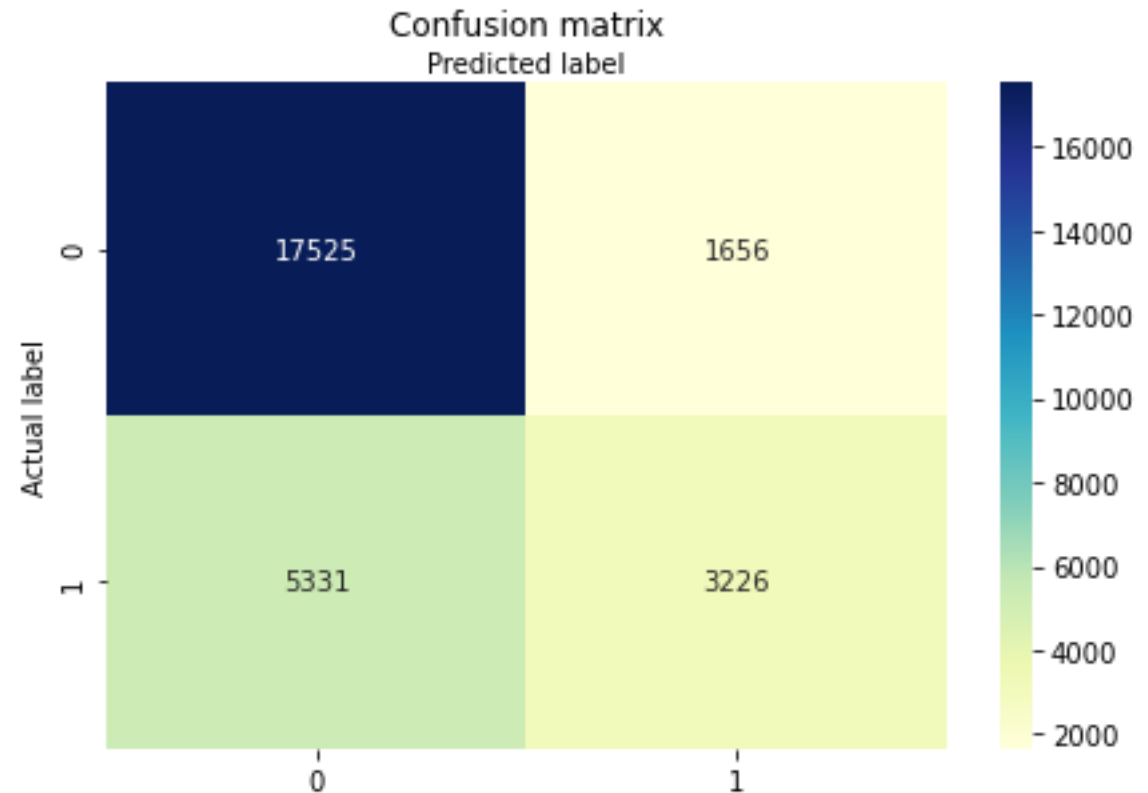
Accuracy: 75%

Precision: 66%

Recall: 37%

ROC Curve Area Under the Curve: 78%

Easy to implement, understand and maintain.



Conclusion and Future Direction

- The model should be calibrated from time to time
- Implementation into the production should be handled with people with knowledge of data treatment.
- With the advent of new technologies, the data capture should be more accurate.
- I recommend segmenting the data to have multiple targeted models.