PREDICTING ACCIDENT SEVERITY

Sahadatu Larabu

IBM-COUSERA APPLIED DATA SCIENCE CAPSTONE PROJECT

August 30, 2020

INTRODUCTION/BUSINESS PROBLEM

1. Road accidents are very fatal to human lives and the environment.
2. Solve the problem of reducing accidents
3. Predict the possibility and severity of road accidents
4. Go a long way to save lots of lives.

DATA

1. The accident severity data for Seattle city (Data-Collisons)
2. The data was downloaded online from the course materials.
3. Datasets has 37 attributes and 1 target.
4. The data contains 194673 rows.

METHODOLOGY

DATA PREPROCESSING:

1. Deleted all empty entries and columns not relevant to the problem.
2. One-hot encoding of categorical variables.
3. Splitting datasets into 70% for training and 30% for testing.
4. Resampling technique to solve the problem of imbalanced target variables.

METHODOLOGY cont'd

MACHINE LEARNING TYPE:

1. Two classes for target variable i.e binary classification problem.
2. Logistic regression model.
3. Added class weights as a hyperparameter.

RESULTS AND DISCUSSIONS

1. With imbalanced datasets, only predicted the first class.
2. Downsampling and adding class weights gave better predictions.
3. Upsampling and adding weights gave better predictions.
4. upsampling was the best with 99.6% accuracy and a good confusion matrix.

CONCLUSION

1. In conclusion, logistic regression with class weights and an upward resampling technique was very good to use for this datasets and problem
2. Individuals can therefore predict the severity of an accident given data on the weather conditions, the road conditions, the number of pedestrian and vehicles and so on.

Figure: