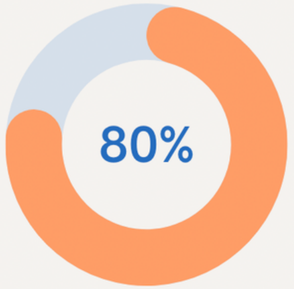
Fraud Detection System

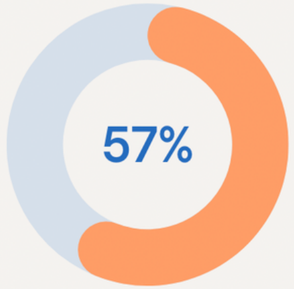
A high recall and balanced precision model to catch most number of frauds keeping in mind the transaction reviewing capacity of the Analyst team

Model Performance

Fine tuned to catch maximum fraud keeping in mind the review capacity of the Analyst team



Catches 80% fraud



57% correct fraud alarms

Solution tailored to solve the business problem by meeting all the requirements

Understanding Data

Performed detailed EDA and Data Cleaning to get the best quality data for our model

Feature Engineering

Created highly relevant new features from existing features and finely crafted the pre-existing features

Tuning the Model

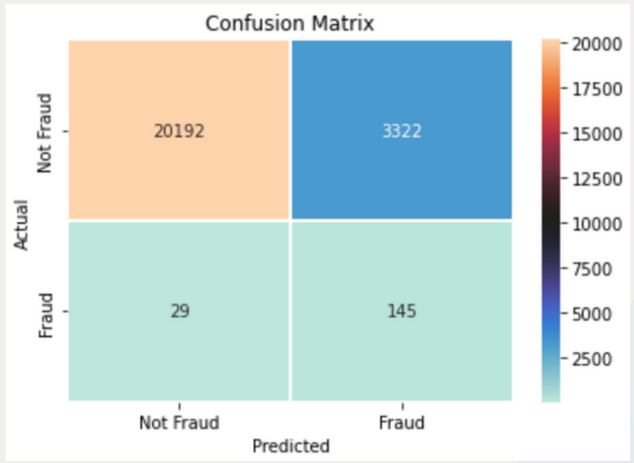
Conducted multiple fold hyperparameter and threshold tuning to bosst the performance of best base model by 19%

- The optimal model was chosen after multiple folds of tuning. The first base model which is an XGBoost Classifier had a precision = 0.57 and recall = 0.7

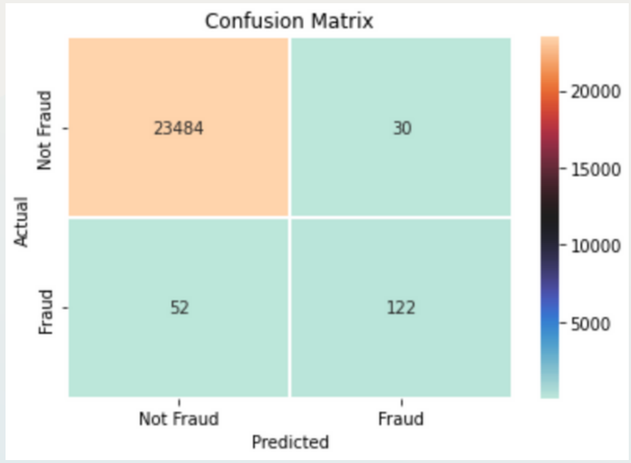
- After tuning the hyper-parameter tuning the precision of model improved by 40%. (0.57->0.80) This means that the model become 40% more accurate when it says a transaction is Fraud. This precision boost gives us a room of improvement for the recall score by tuning the classifier threshold value.

- From the predictions shown below it can be seen that at the cost of some false positives which are very low as compared to a super high recall low precision model like an AdaBoost Classifier, I was able to predict 14% more fraudulent transactions! This brings the total fraud captured to 80% with a reasonably low amount of false positives. This AdaBoostClassifier may capture 6 more fraudulent transactions but raises 97% more false alarms as compared to the optimal model that I have proposed.

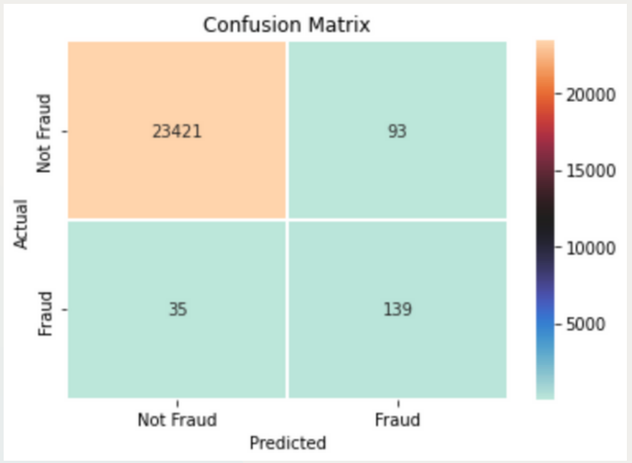
- The penalty for mislabeling a fraud transaction as legitimate is having a users money stolen, which the credit card company typically reimburses but the penalty for mislabeling a legitimate transaction as fraud is having the user frozen out of their finances and unable to make payments. My solution has taken care of the false positives and at the same time tried to achieve the maximum recall to catch the most number of fraudulent transactions.



AdaBoost Classifier



XGB Before tuning threshold



XGB After tuning threshold