

# Predicting credit card fraud using Binary classification



 Slide Members





# Business Need

Payment card fraud losses reached \$28.65 billion worldwide in 2019

The U.S. is the most credit card fraud-prone country in the world.<sup>1</sup>



Transactions data from Kaggle by Vesta Corporation<sup>2</sup>

Features:

Transaction Amount

Product Code

Debit vs. Credit card

Matched Information

Card Information

Target:

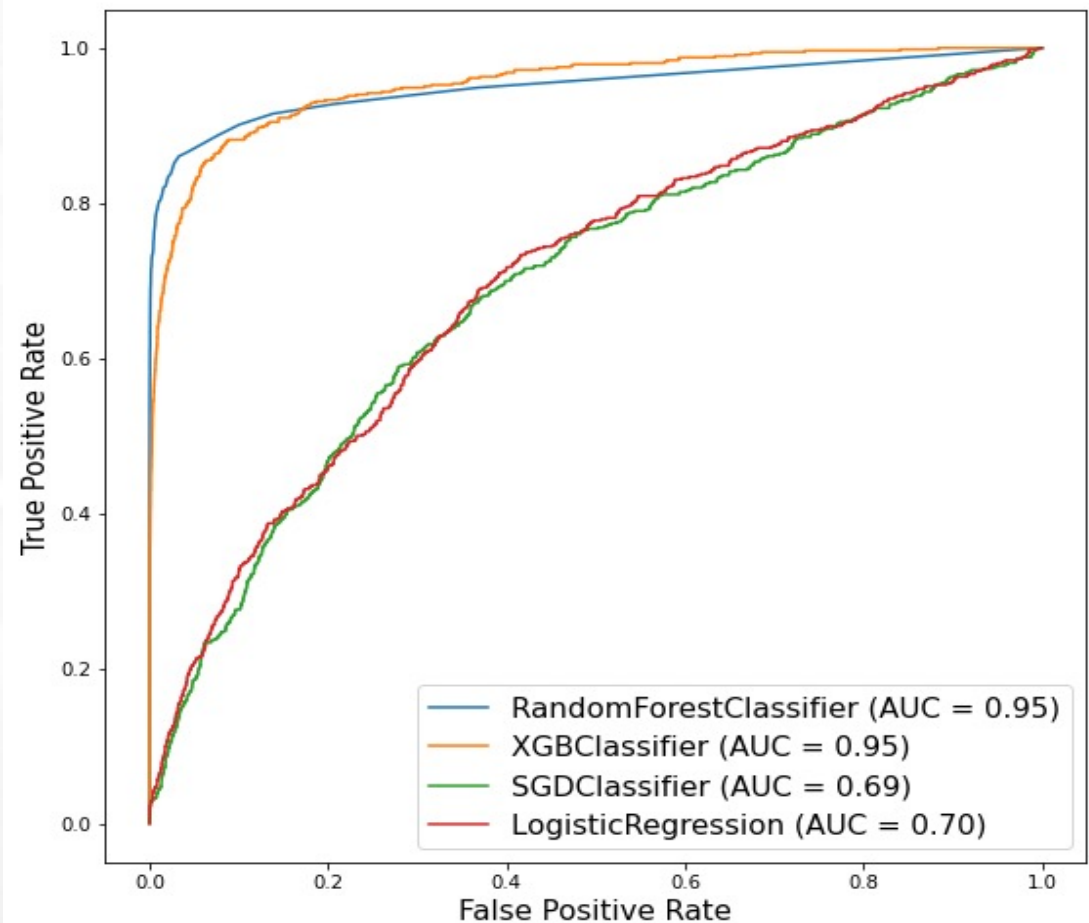
Fraud or Valid



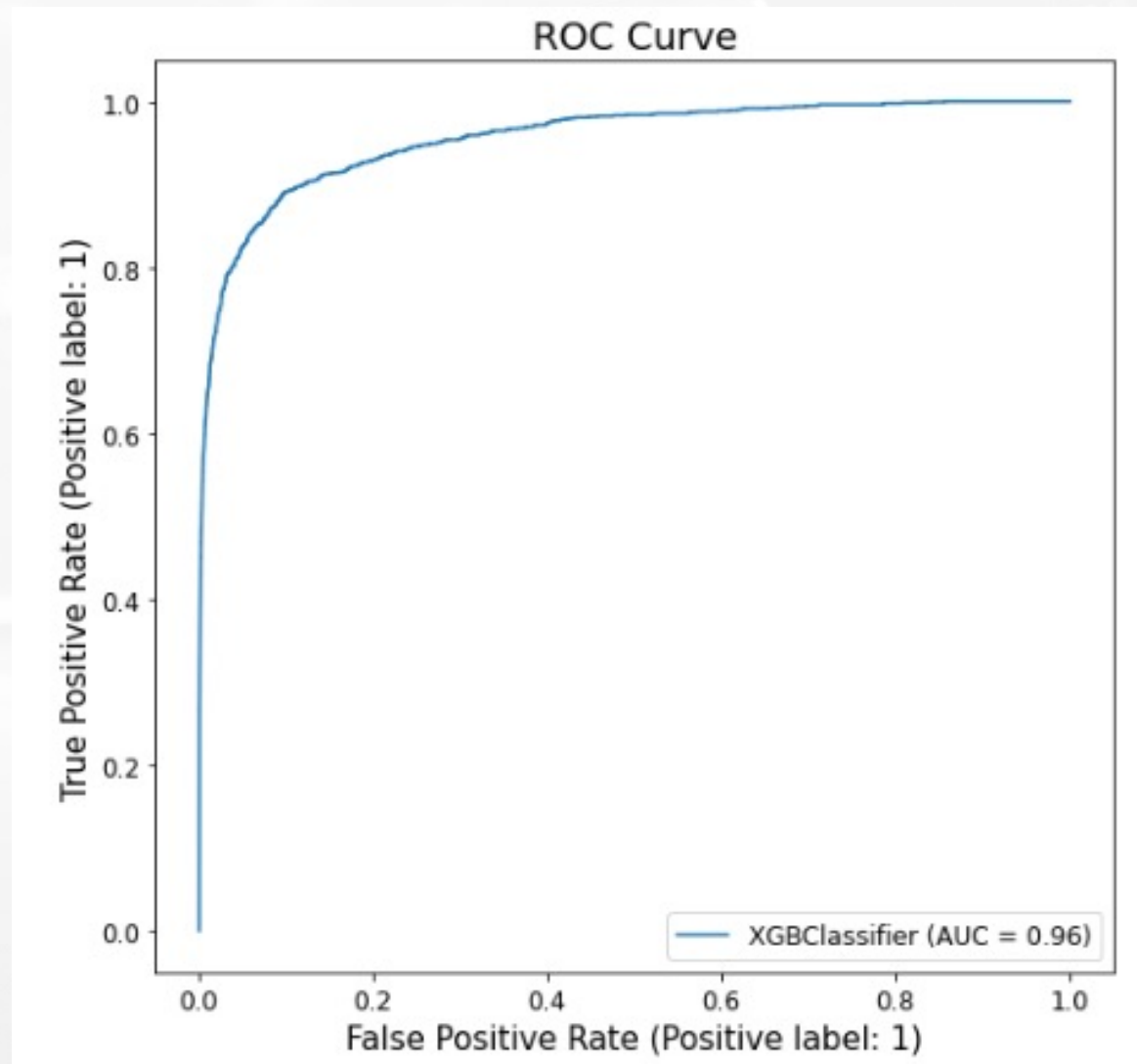
# Model Comparison



Random Forest and XGB works better than Logistic Regression for our dataset



# Final Model: XGBoost with Hyperparameter tuning



# Summary Metrics

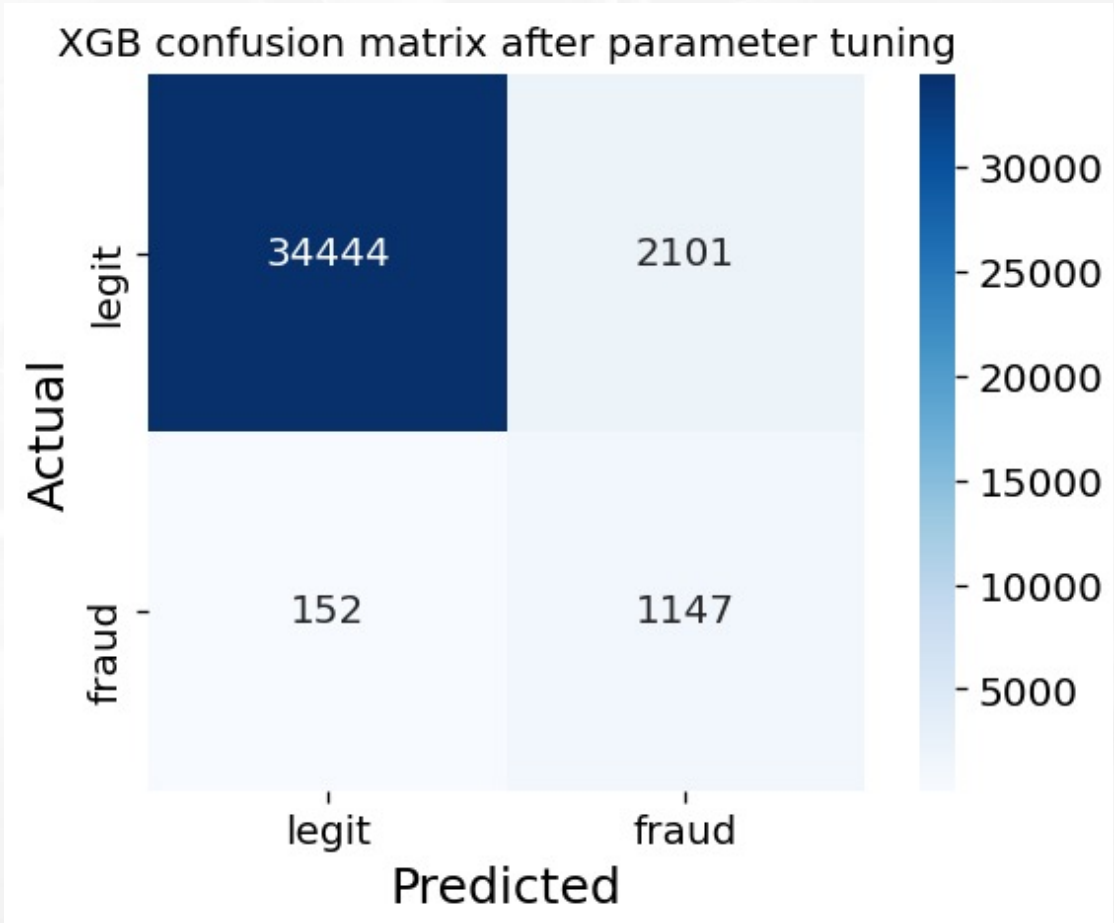


Final XGBoost Model

Threshold  
0.11

Recall  
TP/actual yes  
0.88

False Positive Rate  
FP/actual no  
16%



# Future Work



- Build a more robust model by reducing the False Positive Rate further using more data and features
- Build a Flask app for the project and deploy it



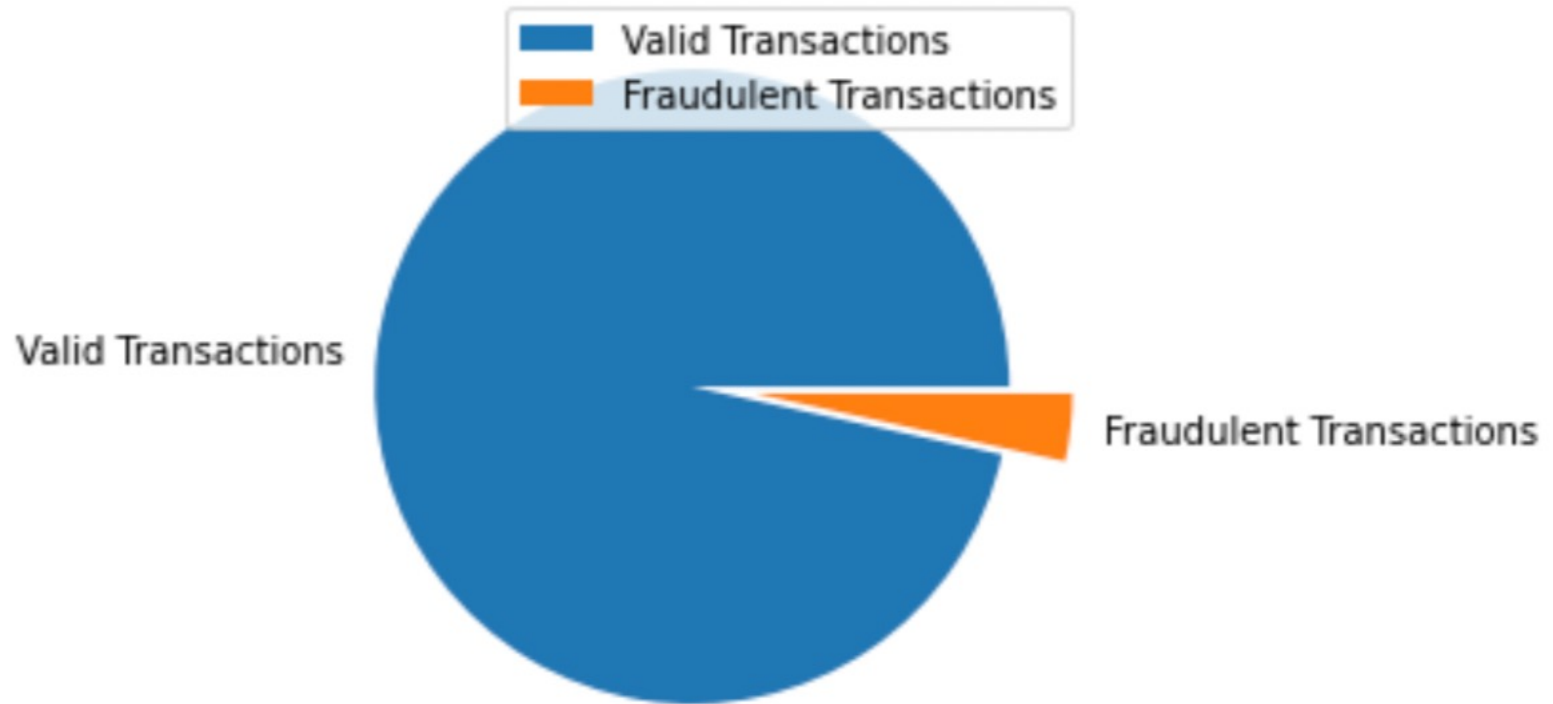
# Appendix



1. Credit card fraud will increase due to the Covid pandemic, experts warn:  
<https://www.cnbc.com/2021/01/27/credit-card-fraud-is-on-the-rise-due-to-covid-pandemic.html>
2. Kaggle data set:  
<https://www.kaggle.com/c/ieee-fraud-detection/data>



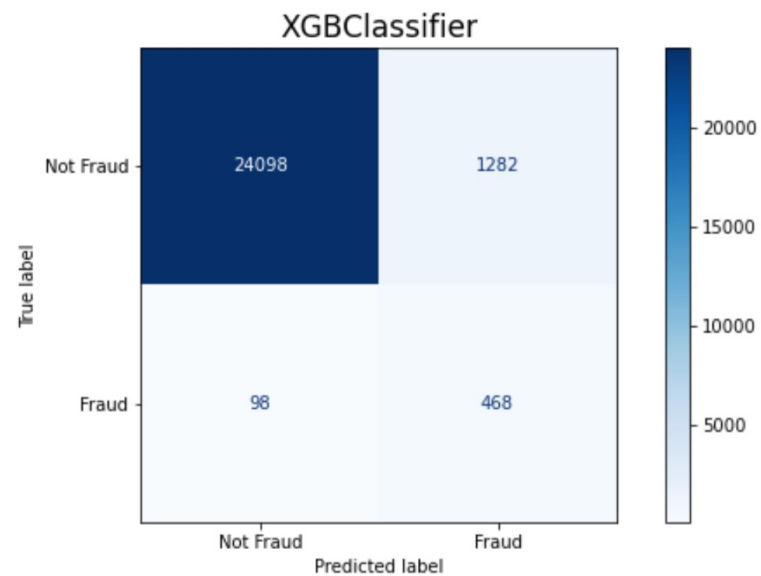
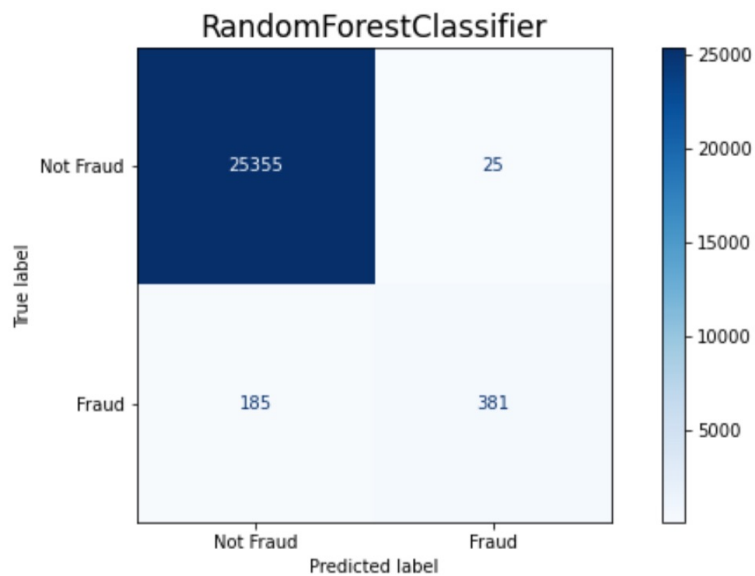
# Appendix



# Appendix



Confusion Matrices - Using Random Oversampling







THANK YOU

CREDIT CARD



1234 5678 9876 5432

VALID  
THRU

MONTH / YEAR  
09 / 19