

CREDIT CARD FRAUD DETECTION

Trusted Classification models to improve your Bank Security!

T5 DATA SCIENCE BOOTCAMP

By:

Afnan Alsirhani
Wafaa Alharbi



Table of Contents

- 1 Intro
- 2 Tools
- 3 Methodology
- 4 Results
- 5 CONCLUSION & FUTURE WORK





Introduction

Fraud detection is a set of activities that are taken to prevent money or property from being obtained through false pretenses.

Motivation

- Increase in the rate of fraud and theft in credit card transactions
- The emergence of new and different methods of fraud

Objectives

- Build a classification Model to detect fraud transaction
- Evaluate the models by using some of the performance metrics

Goal

- This project aim is to build machine learning models to classify fraudulent card transactions from a given card transactions data. This system is useful for both people who use credit cards and banks to keep their customers safe.



TOOLS

Software:

Jupyter Notebook

Language:

Python

Libraries

Statistics libraries:

Sklearn

Statsmodels

Data manipulation libraries:

Pandas

Numpy

Visualization libraries:

Matplotlib

Seaborn

Storage libraries:

sqlalchemy

Pickle

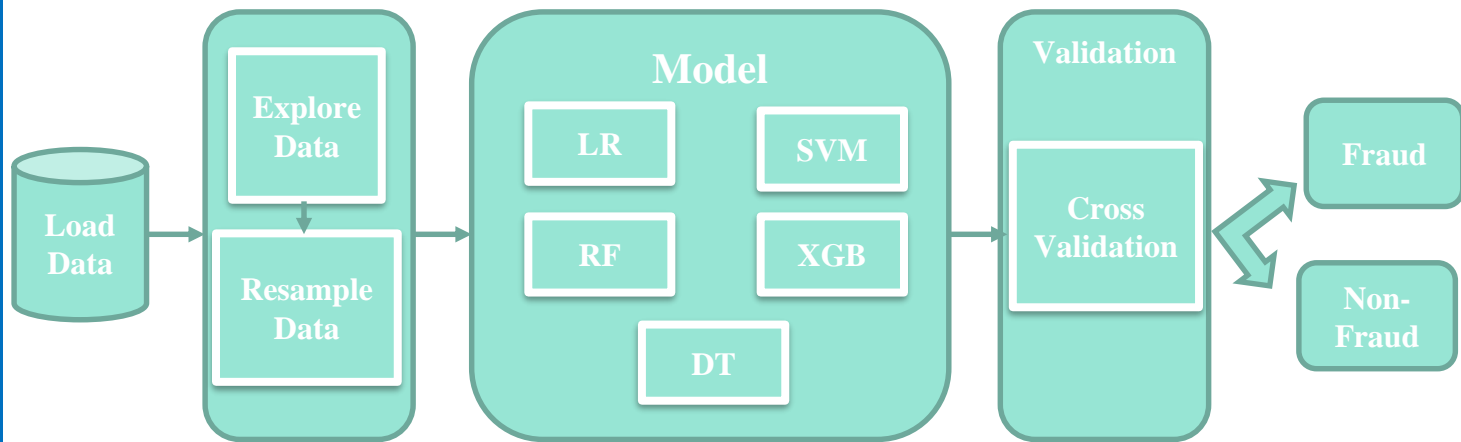


Methodology

Data

- The data is obtained from “**Credit Card Fraud Detection**” dataset in Kaggle, which contain 284,807 transactions, with 29 features.

Model Architecter

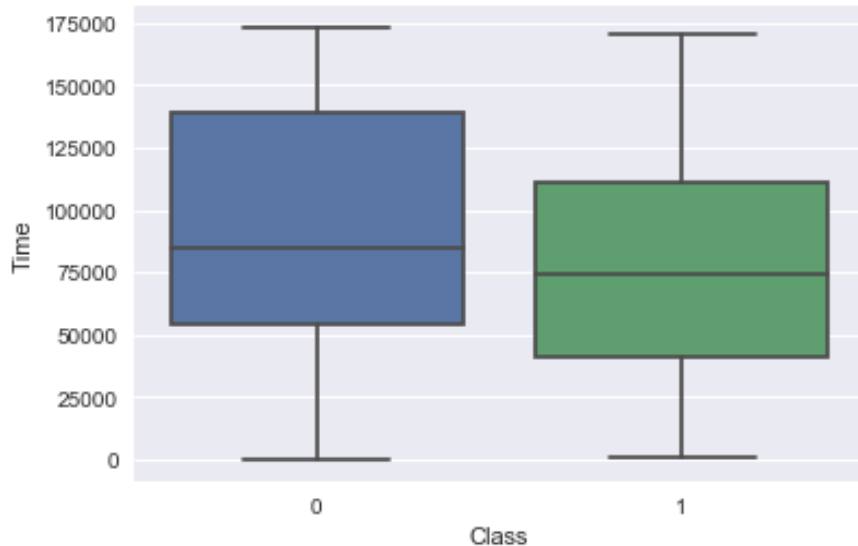


EDA

Drop duplicate -> 1081 values

No null values

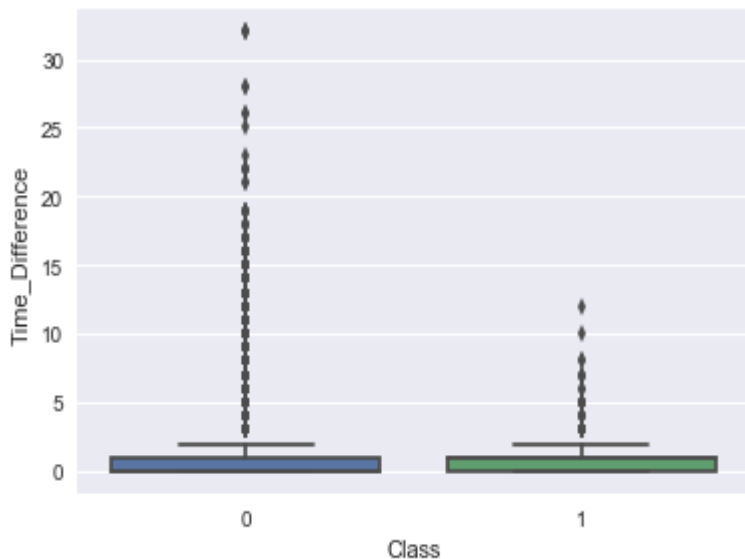
Feature Engineering



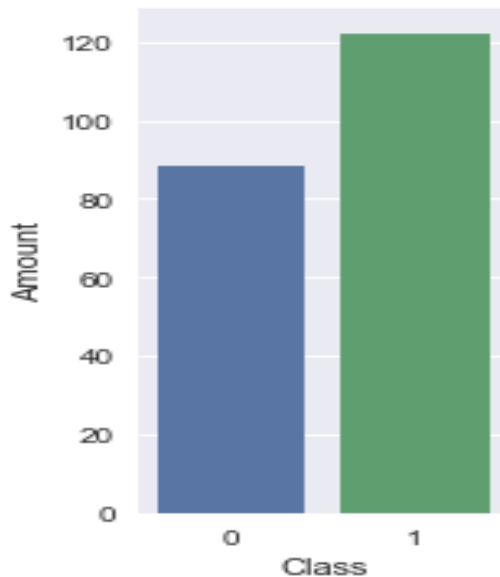
Time itself doesn't seem to determine class well.
We **feature a new variable which is the time difference** between transaction and the precede transaction

Insights

- After our feature engineering we found:



fraud often happens at time that has few transactions

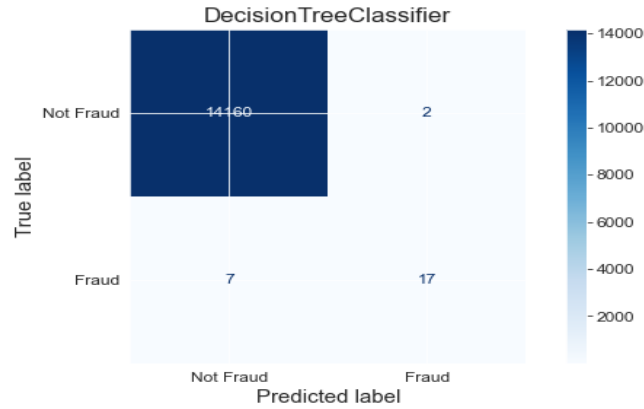
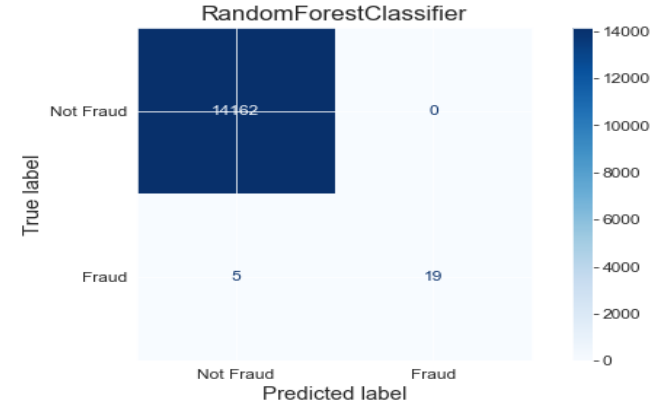
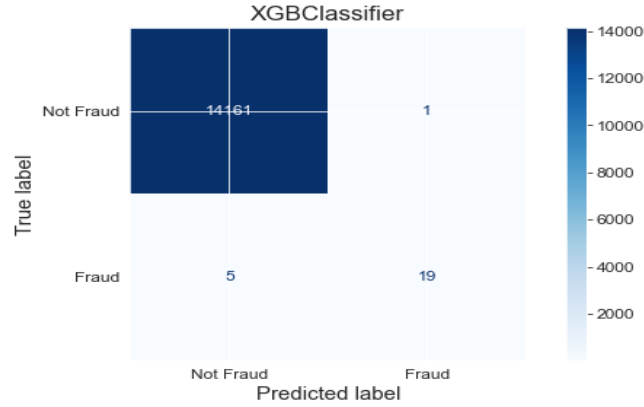


transaction of large amount are mostly fraud, which follows our instinct about fraud.

Model Training & Evaluation

Model	F1-SCORE	Precision	Recall
Logistic Regression (Baseline)	0.75	0.62	0.93
XGB Classifier	0.86	0.79	0.95
Random Forest	0.88	0.79	1
Decision Tree	0.89	0.94	0.85
SVM	0.36	0.25	0.66

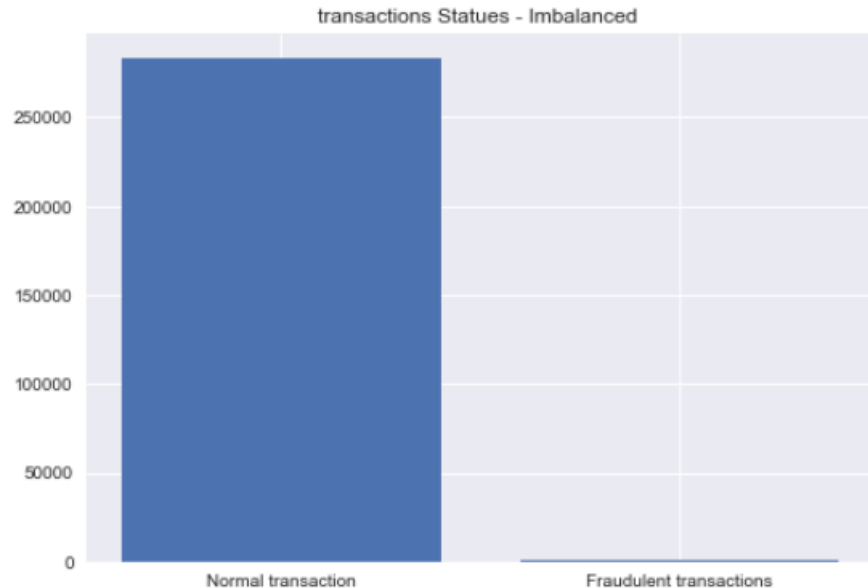
Confusion Matrices



Solving the imbalance Classes


1- SMOTE

2- ADAYSN




The Best 3 Models After Resampling the classes

MODEL	SMOTE	ADAYSN
Decision Tree	0.67	0.63
Random Forest	0.88	0.86
XGB	0.86	0.86



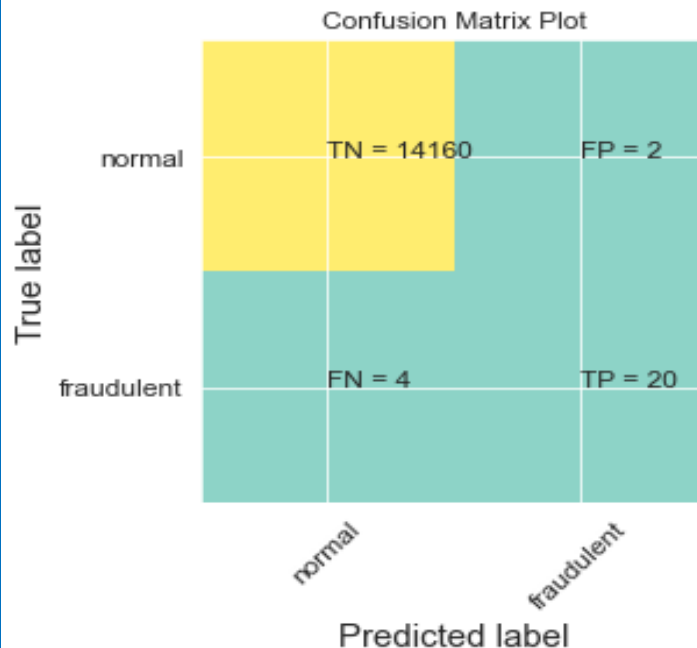
Evaluating Cross Validation

Model	F1-SCORE	Precision	Recall
XGB Classifier	0.87	0.91	0.83
Random Forest	0.86	0.95	0.79
Decision Tree	0.68	0.59	0.79

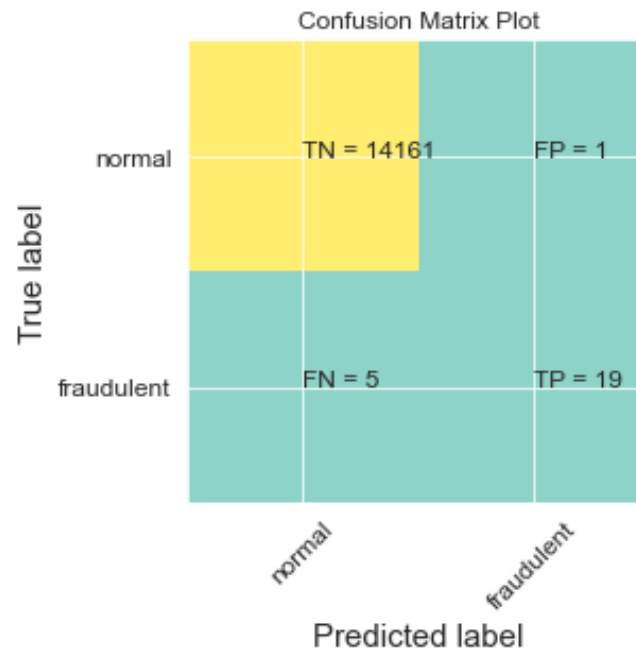


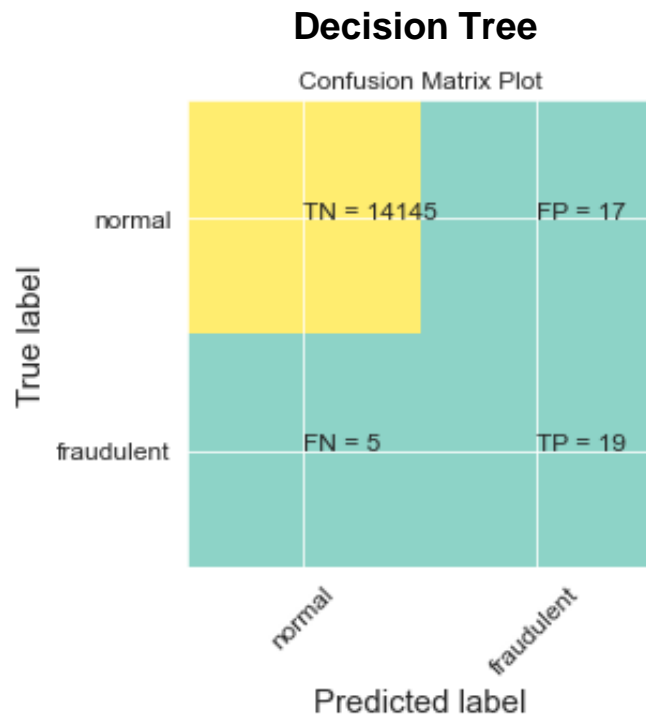
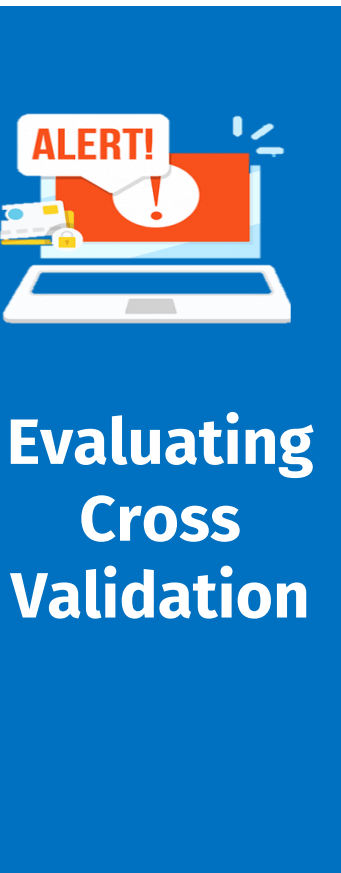
Evaluating Cross Validation

XGB



Random Forest





Since the XGB Classifier give the highest F1 Score , recall score and precision score , we do Grid Search for it to find the best parameters.

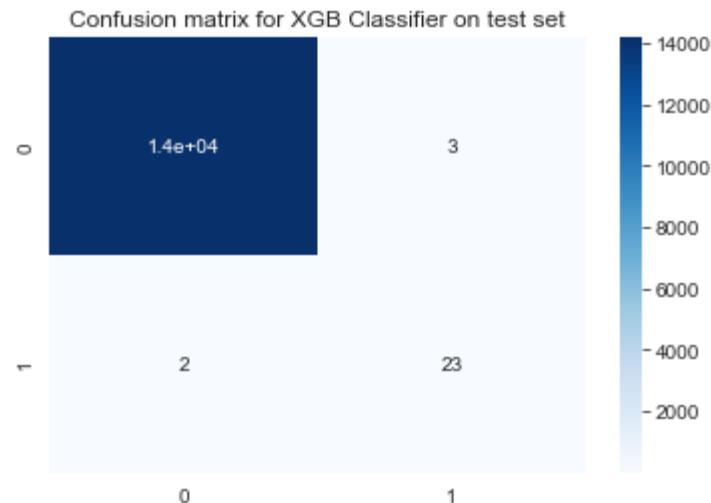
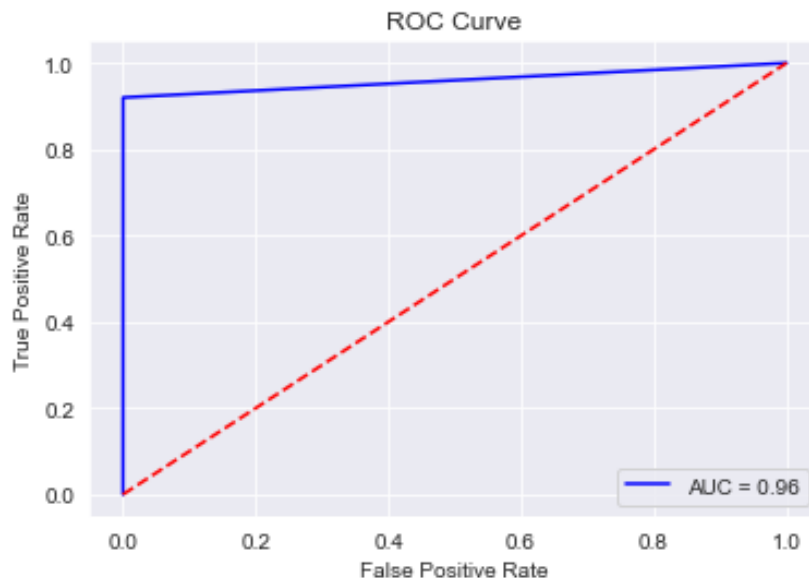
XGB Classifaire after Grid search

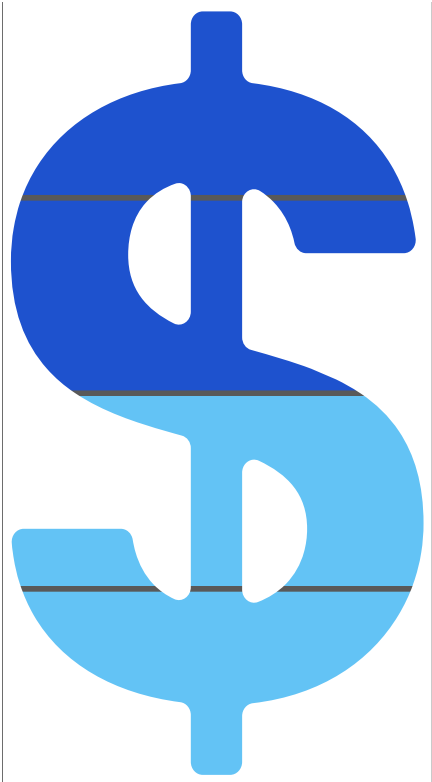


**Grid
search**

Model	F1-SCORE	Precision	Recall
XGB Classifier– Validation data	0.89	0.87	0.91
XGB Classifier– Test data	0.90	0.92	0.88

- Increase the F1 score of XGB Classifier model from 0.86 to 0.89





Conclusion

- We Built a Machine Learning Model to improve the accuracy of fraud prevention based on information about each cardholder's behavior.



Future Work

- Use Autoencoder for anomaly operation detection.
- Enhance model performance.



THANKS
Any Questiones