

Report on

Fraud Detection in Online Transactions

using Machine Learning

CodeClause Data Science Internship - September 2022

Author Name: Akash Mahajan

Objectives:

- **Objective 1:** Descriptive information about dataset and data questioning
- **Objective 2:** Classifying the fraud transactions using machine learning

Objective 1: Descriptive information about dataset and data questioning

This experiment describes and plots the dataset information using special program.

Column Name	Description
Step	Duration of transaction (1 unit is equal to an hour)
Type	The type of online transaction
Amount	The amount to be transacted
nameOrig	The name of person who started transaction or sender
oldBalanceOrg	The original balance of account of sender
newbalanceOrig	The sender's balance after completion of transaction
nameDest	The name of person for whom transaction intended or receiver
oldbalanceDest	The original balance of account of received
newbalanceDest	The receiver's balance after completion of transaction
isFraud	Whether the whole process is fraud (0 = No, 1 = Yes)



The dataset has 3,144 fraud transactions and 21,856 non-fraud transactions

Information on transactions in dataset

Data question	Number of transactions or amount	
	Yes	No
Minimum fraud amount	10.48	-
Maximum fraud amount	1,00,00,000.0	-
Minimum non-fraudulent amount	2.39	-
Maximum non-fraudulent amount	56,77,662.29	
The amount was credited to the destination	594	21262
The amount was debited from origin	6345	15511
The amount same as mentioned is debited	2283	19573
The amount same as mentioned is credited	296	21560
Amounts debited and amount credited are not equal	3144	21856

Objective 2: Classifying the fraud transactions using machine learning

The transactions are classified using five different machine learning models and then they are later compared using the testing accuracy and testing quality.

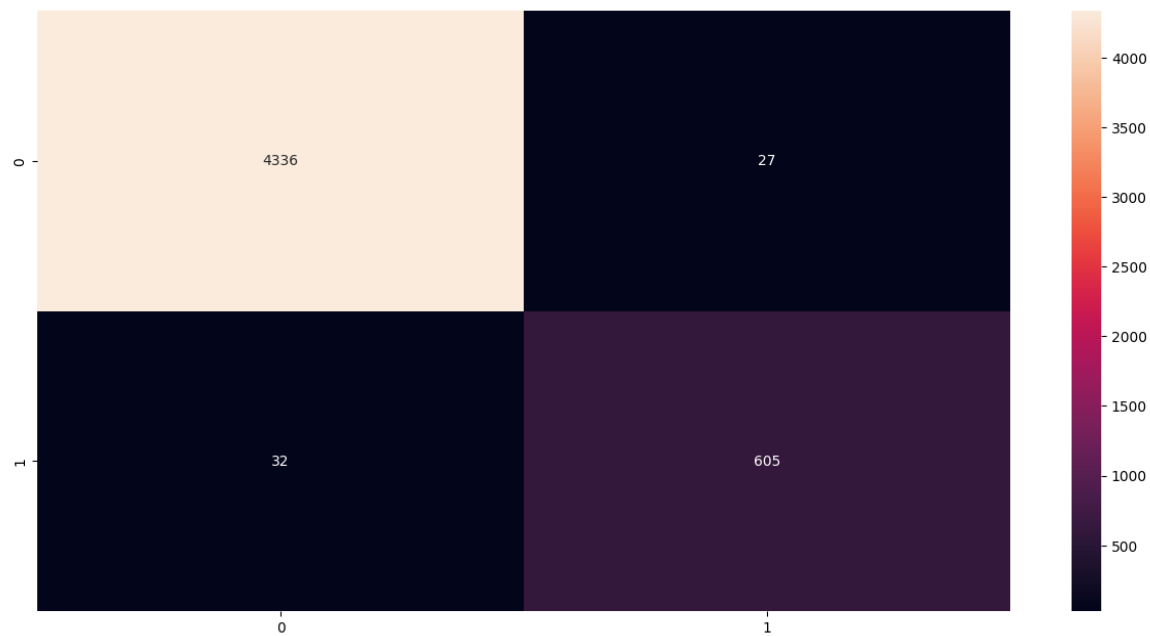
Comparison of models

Model	Testing Accuracy (%)
Decision Tree	98.82
Random Forest	99.08
Gradient Boosting	99.00
Support Vector Machine	87.42
XGBoost	99.26

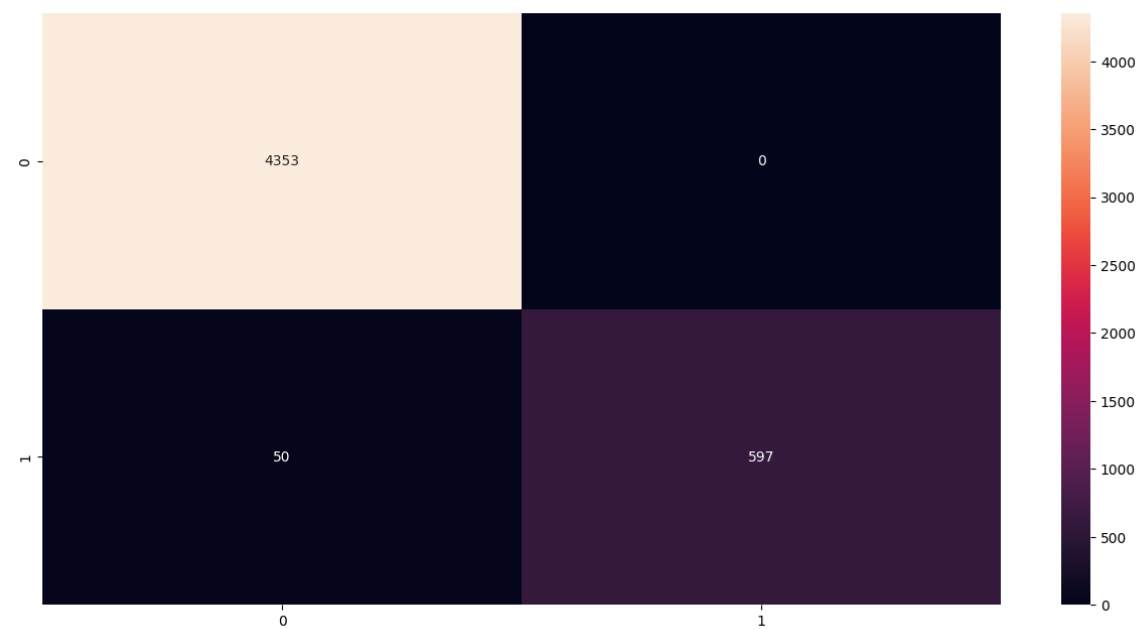
Based on the above table, the XGBoost model is outperformed other models. The following table shows how many records in test set are classified correctly.

Model	Number of test records = 5000	
	Correctly classified	Not correctly classified
Decision Tree	4941	59
Random Forest	4954	46
Gradient Boosting	4950	50
Support Vector Machine	4344	656
XGBoost	4963	37

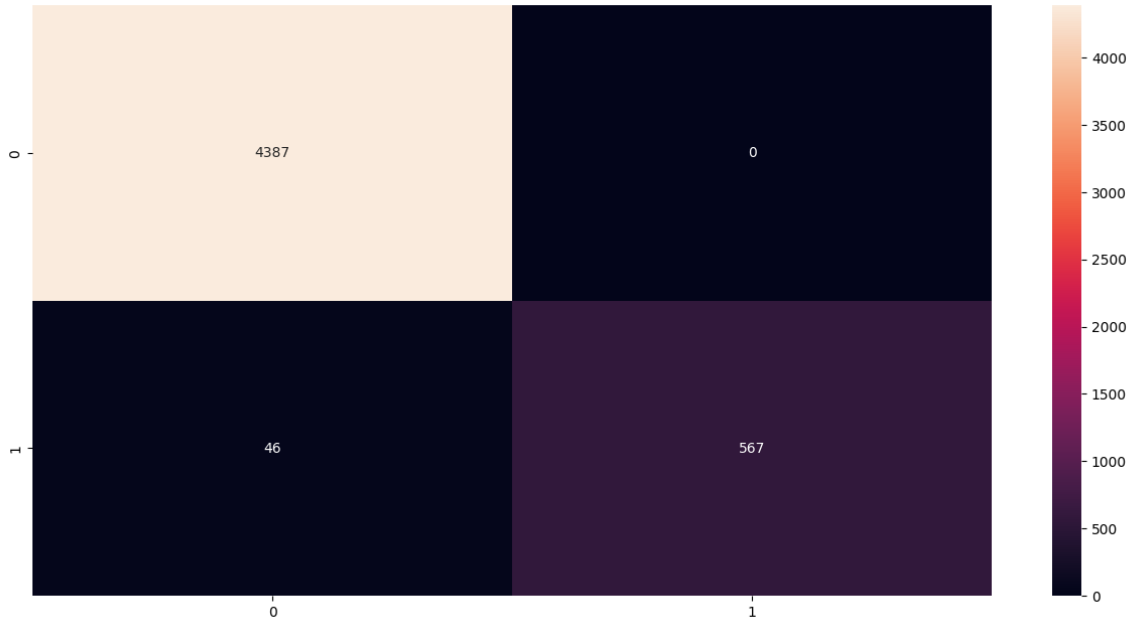
Model-wise confusion matrix



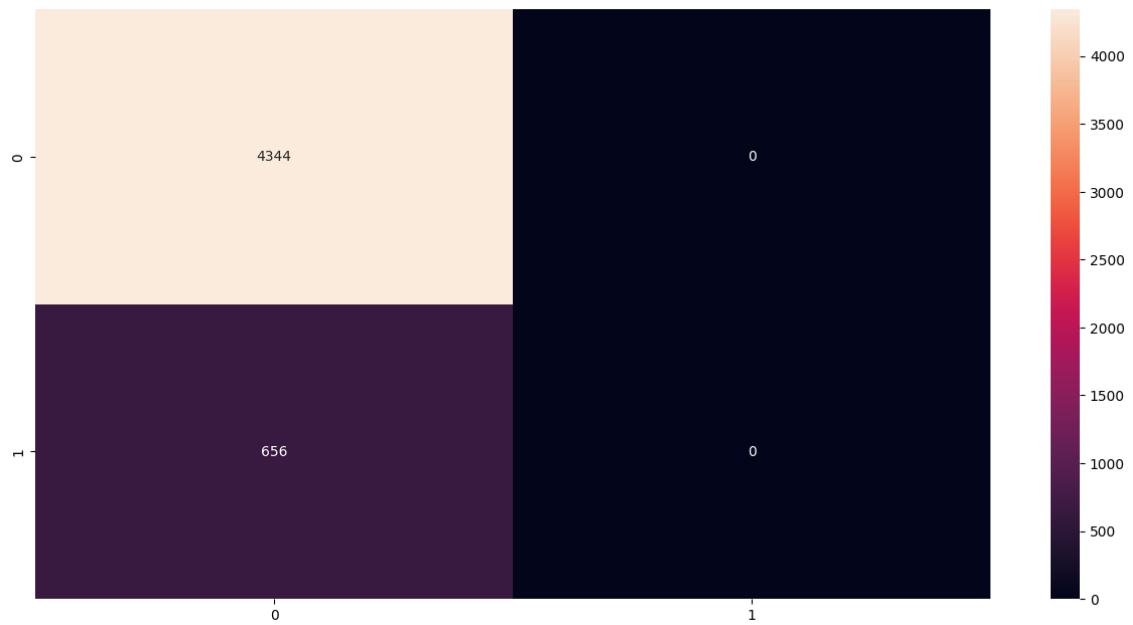
Decision Tree Classification



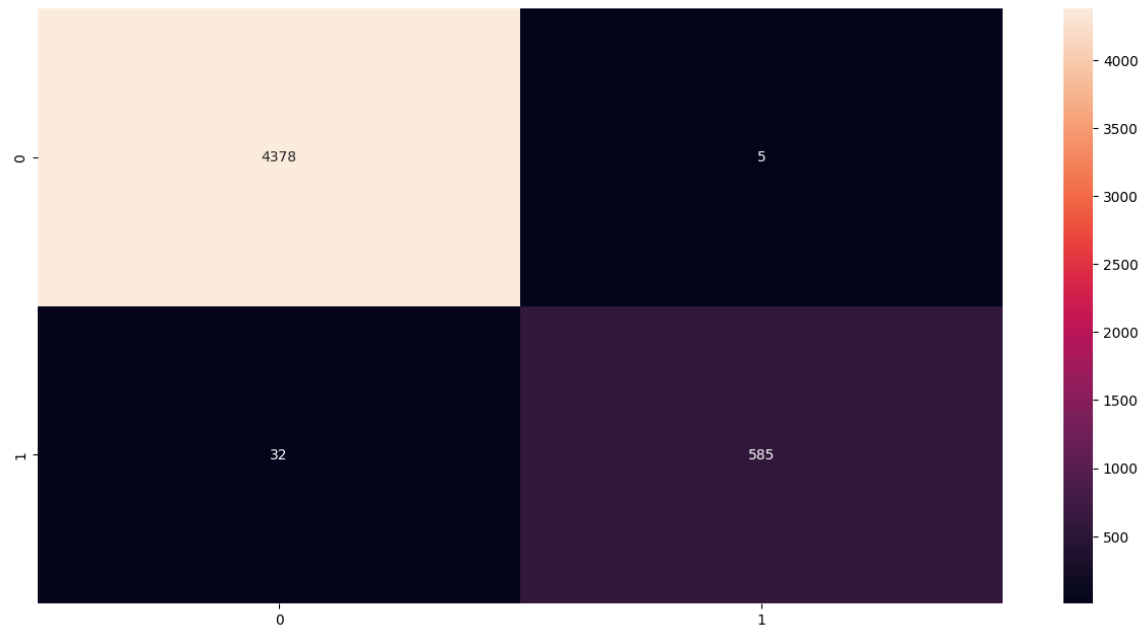
Gradient Boosting Classification



Random Forest



Support Vector Machine Classification



XGBoost Classification