INSAID

INTERNSHIP TASK

-By Rohan Prabhakar

Questions mentioned in the task

1. Data cleaning including missing values, outliers and multi-collinearity.

The data which was about the Fraud transaction was clean and dint require any changes like Data type change. 99% of the data consisted of legit transation and only few fraud transactions were recorded. Though there were missing information about the Transaction Amount of Merchent's acount , i took care of it by feature engineeing and not removing the data of merchents as they might be important for modelling

4. Demonstrate the performance of the model by using best set of tools.

For Evaluating the Performance i used DecisionTreeClassifier accuracy, F1 score, Recall, and Precision, in which recall plays important role of eveluating the fraud detection

In the second model we saw that recall is 99% which tells us that all the fraud transaction as flagged but we can also see that precision is 15% which is very low which tell us that out of all the flagged fraud transactions only 15% od the transaction are actually fraud and reset 85% are not. It is actually good when comes to fraud detection as "It is

better to investigate a innocent person who is not a fraudster rather than not investigating actual fraudster" but the cons are it might be time consuming and waste of time

In the 3rd model which is the better one represents that the recall value is 87% and precision is 74% which tells us that 84% of transaction as correctly fraggerd as fraud but only 74% out of it is actual fraud transaction.

5. What are the key factors that predict fraudulent customer?

We can say that the type of account i.e cashout and the amount a customer hold in his account. which means that fraudsters usually prefer to fraud the account that contains higher amount

6. Do these factors make sense? If yes, How? If not, How not?

Yes, these factors make sense as most of the fraud transaction are of CashOut Type and account which hight balance often get Fraud attacks¶

7. What kind of prevention should be adopted while company update its infrastructure?

While updating its infracture the company should make should Implement Transaction Monitoring i.e analyse the customers usual sending habbits and give a check when there is a huge amount transaction. And should ensure that the recipient got the money or not.

8. Assuming these actions have been implemented, how would you determine if they work?

To check weather if its implemented properly, we can run the model again with the new set of data after inspecting the updated process.