STOUT CHALLENGE

CASE STUDY – 1

DESCRIPTION OF THE DATASET –

Here a labeled dataset of financial transactions is provided along with a values indicating whether the transaction is fraudulent or not.

The columns are:

* step: A timestamp / date variable that has been made arbitrary for data privacy.
* type: The type of transaction. type is our only categorical independent variable. Categories include [cash\_in, cash\_out, debit, payment, transfer].
* amount: Size of transaction.
* Next 6 columns are account level info:
  + Suffix \*Orig refers to the originating account of the transaction.
  + Suffix \*Dest refers to destination account of the transaction.
  + Prefix name\* refers to Account ID number. An 'M' prefix, e.g. 'M1979...' denotes a merchant account. The name will not be directly used as a model feature, although we can extract the merchant prefix to create a boolean indicator variable. One thing to note is that balance data is not available for merchants. For merchants, the placeholder value for balances is zero.
  + Prefix oldbalance\* refers to account balance before transaction.
  + Prefix newbalance\* refers to great pair of shoes—and also account balance after transaction. :)
* isFraud: Our label / dependent variable—whether the transaction was made by a fraudulent agent. 1 for fraudulent, 0 for not fraudulent.
* isFlaggedFraud: Whether the transaction was flagged as fraud by the "business model".

Fraudulent Transaction prediction –

Here two algorithms have been implemented –

Features taken for consideration - ﻿'step', 'type', 'amount', 'oldbalanceOrg', 'newbalanceOrig', 'oldbalanceDest', 'newbalanceDest', 'merchant’

Here we can note that a new factor merchant is also taken into consideration where merchant transactions are considered non fraudulent most of the time.

Also, types of transaction are also considered and added as separate columns indicating type of transaction – 1 indicating true and 0 indicating false.

1] Logistic Regression - Accuracy – 99.8 %

2] Decision Tree Classifier – Accuracy – 99.9 %