VARUVAN VADIVELAN INSTITUTE OF TECHNOLOGY

NAAN MUDHALVAN: IBM

TECHNOLOGY: DATA SCIENCE

PROJECT TITLE: CREDIT CARD FRAUD DETECTION

PROBLEM STATEMENT:

The credit card fraud detection problem includes modeling past credit card transactions with the knowledge of the ones that that turned out be a fraud.

INTRODUCTION:

Credit card firms must detect fraudulent credit card transactions to prevent consumers from being charged for products they did not buy. Data Science can address such a challenge, and its significance, coupled with Machine Learning, cannot be emphasized.

ABOUT CREDIT CARD DETECTION:

In this data science project, we solve the problem of detecting credit card fraud transactions using machine numpy, scikit learn, and few other python libraries. We overcome the problem by creating a binary classifier and experimenting with various machine learning techniques to see which fits better.

DATA PREPARATION:

Since the features are created using PCA, feature selection is unnecessary as many features are tiny. Let's see if there are any missing values in the dataset.

CREDIT CARD FRAUD DETECTION:

As explained above, we have somewhat conflicting data about Danish rates of fraud (as a percentage of total card payment value). The European Central Bank (ECB) assigns Denmark the highest ratio of fraud to total payments. Meanwhile, other important sources like Nets.eu states that it has one of the lowest ratios.

Part of the cause for this may simply be related to the timing of the reports as the ECB report was published in 2016, while Nets.eu published its report later.

DATA DESCRIPTION:

The dataset contains transactions made by credit cards in September 2013 by European cardholders.

This dataset presents transactions that occurred in two days, where we have **492** frauds out of **284,807** transactions. The dataset is highly unbalanced. The positive class (fraud transactions) accounts for **0.172%** of all transactions.

PYTHON LIBRARIES:

This project is a beginner-friendly project on machine learning as it will teach you all the basics of this exciting domain. You will learn

about the four types of machine learning problems: unsupervised learning, supervised learning, semi-supervised learning, and reinforcement learning. The project defines all these problems in detail with examples and various use cases. The goal is to detect which transactions are fraudulent or not, and this problem is an instance of a binary classification problem in supervised learning. To solve this problem, you will use algorithms like Random Forests, K-Nearest Neighbour, and Logistic Regression and deduce which is the best among them with the help of different statistical parameters like Precision, Recall, Accuracy, etc. Also, you will learn about preparing a credit card fraud detection project report with the help of classification metrics like the ROC curve, confusion matrix. The project will also assist you in understanding why accuracy is not an important metric for an imbalanced dataset. Furthermore, you will learn about using hyperparameter tuning techniques: Grid Search CV for under sampled data and Random search CV for oversampled data.

THE MAIN CHALLENGES IN CREDIT CARD FRAUD DETECTION:

Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.

Imbalanced Data i.e most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones data availability as the data is mostly private.

Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.

Adaptive techniques used against the model by the scammers.

DESIGN FOR CREDIT CARD DETECTION: