**CREDIT CARD FRAUD DETECTION**

**PHASE III PROJECT: DEVELOPMENT**

**TABLE OF CONTENTS:**

1. INTRODUCTION
2. PROBLEM DEFINITION
3. DATA PREPROCESSING
4. DATA COLLECTION
5. IMPORTNG LIBRARIES
6. IMPORTING DATASETS
7. HANDLING MISSING DATA
8. ENCODING CATEGORICAL DATA
9. SPLITTING THE DATASET
10. FEATURE SCALING
11. CONCLUSIONS
12. REFERENCES

**INTODUCTION:**

The problem is to develop a machine learning-based system for real-time credit card fraud detection. The goal is to create a solution that can accurately identify fraudulent transactions while minimizing false positives. This project involves data preprocessing, feature engineering, model selection, training, and evaluation to create a robust fraud detection system.

**PROBLEM DEFINITION:**

The mission of our project is to prevent real-time credit card fraud and take measures for it to stop.

**DATA PREPROCESSING:**

Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models. It involves the use of the following 7 steps:

1. Acquire the dataset
2. Import all the crucial libraries
3. Import the dataset
4. Identifying and handling the missing values
5. Encoding the categorical data
6. Splitting the dataset
7. Feature scaling

**DATA COLLECTION:**

Dataset required for implementing the project is given below :

[**https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud**](https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud)

**IMPORTING NECESSARY LIBRARIES:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from matplotlib import gridspec

**IMPORTING THE DATASETS:**

The next step is importing the dataset from the source:

data = pd.read\_csv('credit\_card.csv')

**HANDLING MISSING DATA:**

Lets check if the dataset has any null values and according to that we can delete the entire row or delete it by calculating the mean.

dataset.isnull().shape[0]

print("Non-missing values: " + str(dataset.isnull().shape[0]))

print("Missing values: " + str(dataset.shape[0] - dataset.isnull().shape[0]))

**ENCODING CATEGORICAL DATA:**

Categorical data is data which has some categories. We have no categorical data.

**SPLITTING THE DATASET INTO TRAINING AND TESTING SET:**

In machine learning data preprocessing, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model.

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

y = dataset["Class"]

X = dataset.iloc[:,0:30]

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(

X, y, test\_size = 0.2, random\_state = 42)

X\_train.shape, X\_test.shape, y\_train.shape, y\_test.shape

model.fit(xtrain, ytrain)

**FEATURE SCALING:**

Feature scaling is the final step of data preprocessing in machine learning. It is a technique to standardize the independent variables of the dataset in a specific range. In feature scaling, we put our variables in the same range and in the same scale so that no any variable dominate the other variable.

from sklearn.preprocessing import RobustScaler

scaler = RobustScaler().fit(dataset[["Time", "Amount"]])

dataset[["Time", "Amount"]] = scaler.transform(dataset[["Time", "Amount"]])

dataset.head().append(dataset.tail())

**CONCLUSION:**

Thus we have done the processing of the given data which is included in the seven steps.

**REFERENCES:**

<https://thepythoncode.com/article/credit-card-fraud-detection-using-sklearn-in-python>

<https://www.javatpoint.com/data-preprocessing-machine-learning>

<https://thecleverprogrammer.com/2021/11/22/product-demand-prediction-with-machine-learning/>