PROJECT TITLE:

CREDIT CARD FRAUD DETECTION

(Semester – VII of IV Year M.Sc. (CA & IT) 2023-24)

Submitted By:

Roll No	Name of Students
4075	Mitaxi Sanjaybhai Vaddoriya
4077	Gunjan Anilbhai Vaghasiya
4082	Mansi Rajeshbhai Vavadiya

Submitted To:

K.S. School of Business Management and Information Technology

M. Sc. – Computer Application and Information Technology

Contents

1. Project introduction

Now, as with each passing year more and more countries are going cashless and the dependency on online payment methods are increasing, many complicated investigation systems are being developed and used so as to identify obscure patterns and the relationships among large informational indexes which were earlier impossible to detect. This comes under a new term called Information Mining.

Credit card fraud is a major concern for both consumers and financial institutions. Fraudulent transactions can lead to financial losses and damage to the reputation of financial institutions. Artificial Intelligence techniques have been used extensively to detect fraudulent transactions. In this project, we use logistic regression to classify transactions as either legitimate or fraudulent based on their features.

2. Project overview

Preprocessing :

Before training the model, we first separate the legitimate and fraudulent transactions. Since the data is imbalanced, with significantly more legitimate transactions than fraudulent transactions, we undersample the legitimate transactions to balance the classes. We then split the data into training and testing sets using the train_test_split () function

Dataset Description :

The dataset contains transactions made by a cardholder in a duration in 2 days i.e., two days in the month of September 2013. Where there are total 284,807 transactions among which there are 492 i.e., 0.172% transactions are fraudulent transactions. This dataset is highly unbalanced. Since providing transaction details of a customer is considered to issue related to confidentiality, therefore most of the features in the dataset are transformed using principal component analysis (PCA). V1, V2, V3,..., V28 are PCA applied features and rest i.e., 'time', 'amount' and 'class' are non-PCA applied features, as shown in table

SNO	Features	Description
1	Time	Time in seconds to specify the elapses between the current transaction and first transaction
2	Amount	Transaction amount
3	Class	0-Not fraud 1-Fraud

Reference :- https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

In this dataset we done the following data preprocessing step:

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

3. Methodology

We use logistic regression to classify transactions as either legitimate or fraudulent based on their features. Logistic regression is a widely used classification algorithm that models the probability of an event occurring based on input features. The logistic regression model is trained on the training data using the LogisticRegression () function from scikit-learn. The trained model is then used to predict the target variable for the testing data.

The performance of the model is evaluated using the accuracy metric, which is the fraction of correctly classified transactions. The accuracy on the training and testing data is calculated using the accuracy_score() function from scikit-learn.

We use Streamlit to create a user interface for the credit card fraud detection project. The Streamlit application allows the user to upload a CSV file containing credit card transaction data, and the uploaded data is used to train the logistic regression model. The user can also input transaction features and get a prediction on whether the transaction is legitimate or fraudulent.

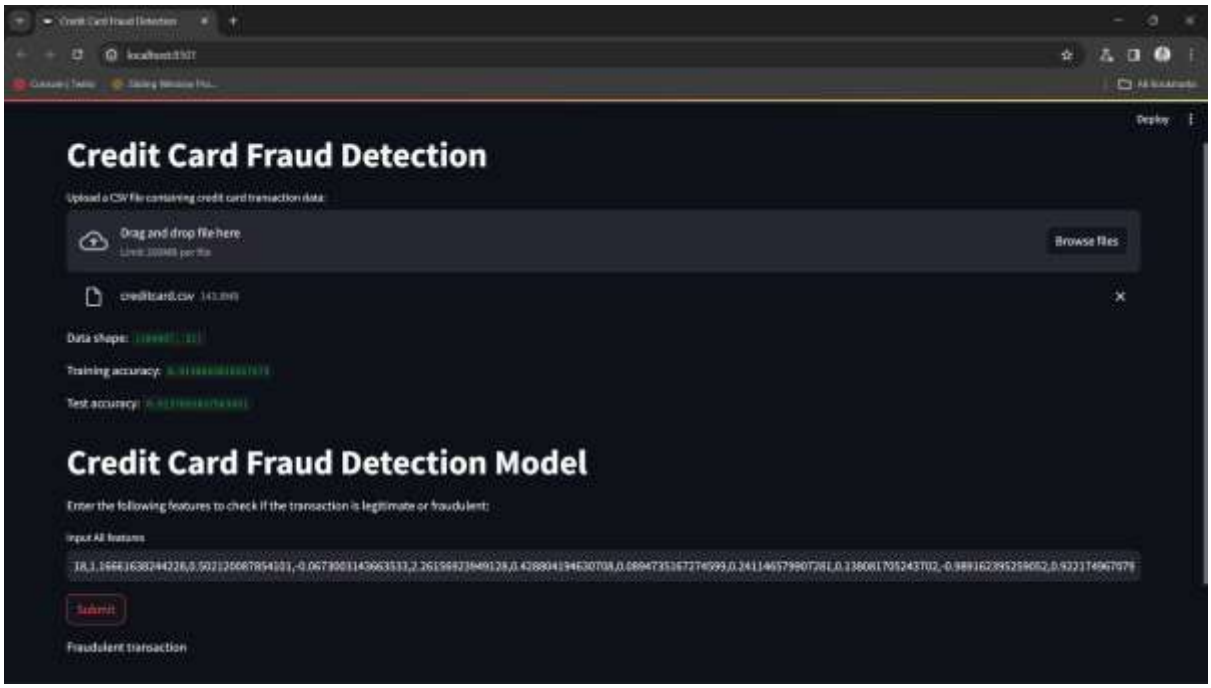
4. Result

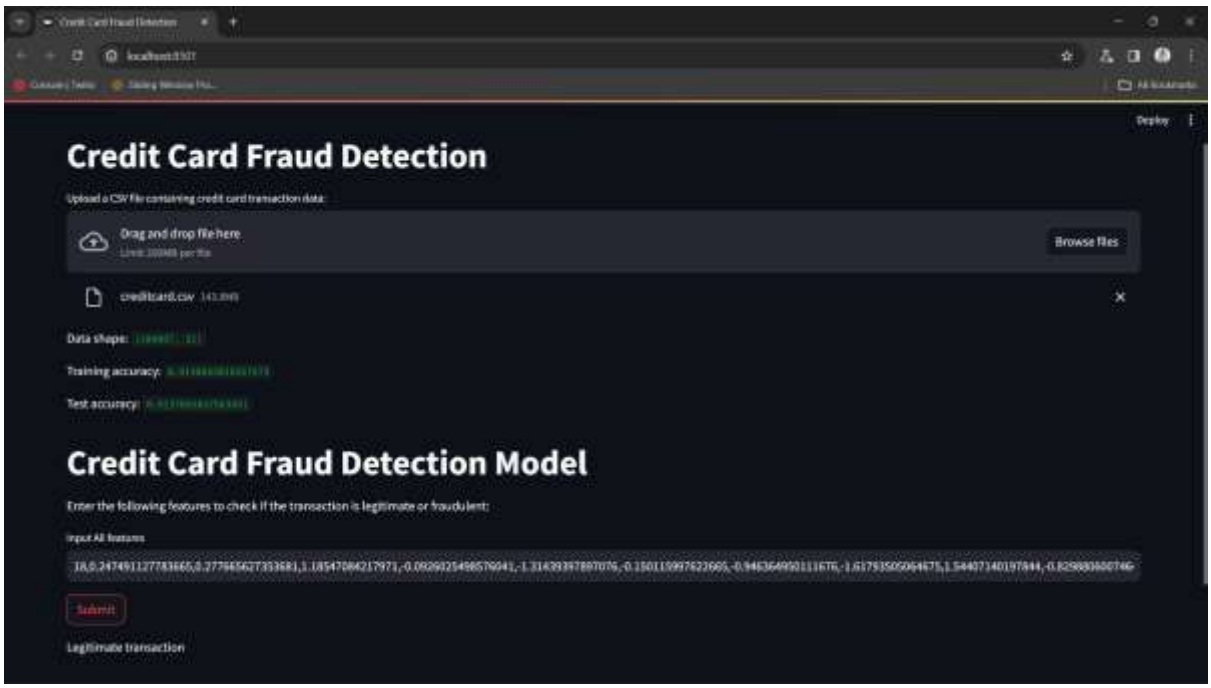
Code :

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import streamlit as st

st.set_page_config(page_title="Credit Card Fraud Detection", layout="wide")
def load_data(file):
    return pd.read_csv(file)
def train_model(data):

    legit = data[data.Class == 0]
    fraud = data[data.Class == 1]

    legit_sample = legit.sample(n=len(fraud), random_state=2)
    data = pd.concat([legit_sample, fraud], axis=0)

    X = data.drop(columns="Class", axis=1)
    y = data["Class"]
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, stratify=y, random_state=2
    )

    model = LogisticRegression()
    model.fit(X_train, y_train)

    train_acc = accuracy_score(model.predict(X_train), y_train)
    test_acc = accuracy_score(model.predict(X_test), y_test)

    return model, train_acc, test_acc

st.title("Credit Card Fraud Detection")

file = st.file_uploader("Upload a CSV file containing credit card transaction data:")
if file is not None:
    data = load_data(file)
    st.write("Data shape:", data.shape)

    model, train_acc, test_acc = train_model(data)

    st.write("Training accuracy:", train_acc)
    st.write("Test accuracy:", test_acc)
```

```
    st.title("Credit Card Fraud Detection Model")
    st.write("Enter the following features to check if the transaction is legitimate or
fraudulent:")

    input_df = st.text_input('Input All features')
    input_df_lst = input_df.split(',')

    submit = st.button("Submit")

    if submit:
        features = np.array(input_df_lst, dtype=np.float64)
        prediction = model.predict(features.reshape(1,-1))
        if prediction[0] == 0:
            st.write("Legitimate transaction")
        else:
            st.write("Fraudulent transaction")
```

Output :




Credit Card Fraud Detection
Upload a CSV file containing credit card transaction data:
Drag and drop file here
Browse files
creditcard.csv
Data shape:
Training accuracy:
Test accuracy:
Credit Card Fraud Detection Model
Enter the following features to check if the transaction is legitimate or fraudulent:
Input All features
Submit
Legitimate transaction

5. Conclusion

In this project, we used logistic regression to detect fraudulent credit card transactions. We achieved a high accuracy on both the training and testing data, indicating that the model is effective at detecting fraudulent transactions.

Since the entire dataset consists of only two days' transaction records, its only a fraction of data that can be made available if this project were to be used on a commercial scale. Being based on Artificial Intelligence algorithms, the program will only increase its efficiency over time as more data is put into it.

6. References

- Python CookBook, 3rd Edition
  By David Beazley, Brain k. Jones: Publisher(s): O'Reilly Media, Inc.
- Introduction to Artificial Intelligence & Expert Systems, Dan W. Patterson, Prentice-Hall India, 1998