**<https://github.com/ispandya/CreditCardFraud>**

**1: Credit Card Detection Machine Learning Project with Report:**

**Project Report:**

**Introduction**

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. Machine learning 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.

**Data**

The data used in this project is a CSV file containing credit card transaction data. The data has 31 columns and 284,807 rows. The "Class" column is the target variable, which indicates whether the transaction is legitimate (Class = 0) or fraudulent (Class = 1).

**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.

**Model**

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.

**Evaluation**

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.

**Streamlit Application**

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.

**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. The Streamlit application provides an easy-to-use interface for detecting fraudulent transactions in real-time.

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")  
  
  
@st.cache(allow\_output\_mutation=True)  
def load\_data(file):  
 return pd.read\_csv(file)  
  
  
def train\_model(data):  
 # separate legitimate and fraudulent transactions  
 legit = data[data.Class == **0**]  
 fraud = data[data.Class == **1**]  
  
 # undersample legitimate transactions to balance the classes  
 legit\_sample = legit.sample(n=len(fraud)**,** random\_state=**2**)  
 data = pd.concat([legit\_sample**,** fraud]**,** axis=**0**)  
  
 # split data into training and testing sets  
 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** )  
  
 # train logistic regression model  
 model = LogisticRegression()  
 model.fit(X\_train**,** y\_train)  
  
 # evaluate model performance  
 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  
  
  
def parse\_transaction\_string(transaction\_string**,** feature\_names):  
 values = transaction\_string.split(",")  
 transaction = {}  
 for i in range(len(values)):  
 transaction[feature\_names[i]] = float(values[i])  
 return transaction  
  
  
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)  
  
 # allow user to input transaction features and get a prediction  
 st.subheader("Check a transaction")  
 feature\_names = data.drop(columns="Class"**,** axis=**1**).columns  
 transaction\_string = st.text\_input(  
 "Enter transaction features (comma-separated)"**,** "0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0")

**About The Code:**

This is a machine learning project for credit card fraud detection. The project uses a logistic regression model to classify transactions as either legitimate or fraudulent based on their features.

The code begins with importing necessary libraries such as numpy, pandas, scikit-learn, and Streamlit. Then, the layout of the Streamlit application is set using the st.set\_page\_config() function. The load\_data() function is defined to load data from a CSV file, which is uploaded by the user using the file\_uploader() function. The train\_model() function is defined to train a logistic regression model on the uploaded data. The function first separates the legitimate and fraudulent transactions, undersamples the legitimate transactions to balance the classes, and then splits the data into training and testing sets using the train\_test\_split() function. The logistic regression model is then trained on the training data and evaluated on the training and testing data using the accuracy\_score() function.

The parse\_transaction\_string() function is defined to parse a comma-separated string of transaction features and convert it into a dictionary with feature names as keys and feature values as values.

Finally, the Streamlit application is created using the st.title() function to set the title, the file\_uploader() function to allow the user to upload a CSV file, and the text\_input() function to allow the user to input transaction features and get a prediction. The uploaded data is loaded using the load\_data() function, and the model is trained and evaluated using the train\_model() function. The training and testing accuracies are displayed using the st.write() function, and the user can input transaction features using the text\_input() function.