Credit Card Fraud Detection Using Logistic Regression

Student Name: Harini.S

Duration: June 2025 to July 2025

# Project Objective

The objective of this project is to develop a machine learning model that accurately identifies fraudulent credit card transactions. Due to the severe imbalance between fraudulent and legitimate transactions in the dataset, the project uses undersampling and Logistic Regression to achieve balanced classification.

# Technologies & Tools Used

- Python

- Pandas, NumPy

- Scikit-learn

- Streamlit

- Jupyter Notebook / VS Code

# Dataset Description

The dataset used is the well-known Kaggle Credit Card Fraud Detection dataset. It contains transactions made by European cardholders in September 2013. The dataset presents transactions that occurred in two days, with 284,807 transactions, out of which 492 are frauds.

Features include 28 anonymized principal components (V1-V28), 'Time', 'Amount', and the target variable 'Class'.

# Workflow Overview

1. Load and explore the dataset  
2. Handle class imbalance by undersampling legitimate transactions  
3. Split the data into training and testing sets  
4. Train a Logistic Regression model  
5. Evaluate model accuracy  
6. Build a simple Streamlit app for real-time prediction

# Sample Output

After entering 30 feature values in the Streamlit web app, the model predicts whether the transaction is legitimate or fraudulent. The output is displayed as either:  
- 'Legitimate Transaction'  
- 'Fraudulent Transaction'

# Project File Structure Checklist

✔ creditcard.csv.csv (Dataset file)  
✔ test.py (Main script with model and Streamlit app)  
✔ requirements.txt (List of dependencies for reproducibility)  
✔ README.md (Optional, explaining usage and setup)

# Conclusion

This project demonstrates a complete machine learning pipeline from data preprocessing to deployment via a web interface. Using Logistic Regression and handling class imbalance properly resulted in a functional fraud detection model. Further improvements can include using more advanced models like Random Forests or XGBoost, hyperparameter tuning, and deploying the app online using platforms like Streamlit Cloud or Heroku.

# Project File Structure

Credit\_Card\_Fraud\_Detection\_Project/  
├── app.py # Main Streamlit app for fraud prediction  
├── data/  
│ └── creditcard.csv.csv # Credit card transactions dataset  
├── models/  
│ └── logistic\_regression.pkl # (Optional) saved logistic regression model  
├── output/  
│ ├── confusion\_matrix.png # (Optional) visualization  
│ ├── classification\_report.txt # (Optional) saved evaluation report  
│ └── accuracy\_scores.csv # (Optional) train/test accuracy metrics  
├── utils/  
│ └── preprocessing.py # Helper functions for scaling, input parsing  
├── requirements.txt # List of required Python packages  
├── README.md # Project overview and usage guide  
└── report/  
 └── Credit\_Card\_Fraud\_Detection\_Report.docx # Final project report