Credit Card Fraud Detection using Logistic Regression

# 📁 Project Overview:

This project aims to build a machine learning model to detect fraudulent credit card transactions using a publicly available dataset. The dataset is highly imbalanced, with only 0.17% of transactions being fraudulent. The focus is on data preprocessing, balancing techniques, and training an effective classification model.

# 📊 Dataset Information:

- Total transactions: 284,807

- Fraudulent transactions: 492

- Features: 30 anonymized PCA features + Time + Amount

- Target: Class (0 = legitimate, 1 = fraud)

# ✅ Steps Performed:

## 1. Data Exploration & Cleaning

- Verified no missing values  
- Checked data types and feature overview  
- Evaluated `Amount` feature for fraud and legit classes

## 2. Handling Imbalanced Dataset

- Used under-sampling to balance the dataset: 492 fraud and 492 legitimate transactions  
- Concatenated balanced datasets to form a new dataset with 984 records

## 3. Feature Engineering & Preparation

- Split dataset into features (X) and target (y)  
- Scaled Time and Amount columns using StandardScaler  
- Split the dataset using stratified train-test split

## 4. Model Training (Logistic Regression)

- Trained logistic regression classifier  
- Addressed convergence warning by increasing max\_iter to 1000

## 5. Model Evaluation

- Achieved 95.29% accuracy on training data  
- Achieved 94.92% accuracy on testing data  
- Accuracy score used for model validation

# 📈 Tools & Libraries:

- Python, NumPy, Pandas  
- scikit-learn (LogisticRegression, train\_test\_split, accuracy\_score)  
- matplotlib / seaborn (optional for visualization)

# 📊 Future Work:

- Implement SMOTE for better balancing without data loss  
- Compare logistic regression with advanced models (Random Forest, XGBoost)  
- Evaluate model using classification metrics like precision, recall, and F1-score  
- Build a simple deployment using Streamlit or Flask

# 📝 Summary for Resume/CV:

Credit Card Fraud Detection Project

Objective: Built a machine learning model to classify fraudulent transactions from real-world credit card transaction data using supervised learning.

Process: Handled class imbalance using under-sampling, performed statistical analysis and feature scaling, and trained a logistic regression model on stratified data split.

Outcome: Achieved over 94% accuracy on test data with optimized preprocessing; implemented with Python and scikit-learn using real financial data.

Keywords: Logistic Regression, Imbalanced Data, Supervised Learning, Model Evaluation, Python, scikit-learn, Accuracy, Fraud Detection, Feature Engineering, Data Preprocessing, Under-sampling, Stratification