

Name: Farzeen Fatima  
CMS ID: 023-23-0319  
Project: Data Science Project  
Degree Program: BSCS (Semester V)

# Credit Card Fraud Detection System

## 1. Introduction

Credit card fraud detection is a critical real-world problem in financial systems. Fraudulent transactions are rare compared to normal transactions, which makes the dataset highly imbalanced and difficult to model using standard classification techniques. The goal of this project is to build a Machine Learning-based system that can accurately detect fraudulent credit card transactions while minimizing false alarms.

## 2. Objectives

- To analyze credit card transaction data
- To handle class imbalance effectively
- To build a reliable fraud detection model
- To evaluate the model using appropriate metrics
- To visualize results using an interactive dashboard

## 3. Dataset Description

Dataset Name: Credit Card Fraud Detection Dataset

Source: Kaggle

Link: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

The dataset contains transactions made by European cardholders. Features V1–V28 are anonymized using PCA. The target variable 'Class' indicates fraud (1) or normal transaction (0). The dataset is highly imbalanced.

## 4. Methodology

The dataset is loaded from a local CSV file. If the dataset is missing, synthetic data is generated automatically. Preprocessing includes removing the Time column and scaling the Amount feature using StandardScaler. Train-test split is performed using stratified sampling. Sampling techniques such as SMOTE are applied only on training data. A Random Forest classifier is used for model training.

## 5. Evaluation Metrics

The model is evaluated using Accuracy, Precision, Recall, F1-Score, and Specificity. A confusion matrix is used to visualize classification results.

## **6. Results and Discussion**

The model demonstrates good performance in detecting fraudulent transactions despite the highly imbalanced dataset. Feature importance analysis highlights the most influential variables used by the model.

## **7. Conclusion**

This project successfully demonstrates a machine learning approach to credit card fraud detection. Proper handling of class imbalance and data leakage ensures reliable model performance. The Streamlit dashboard provides an interactive and user-friendly interface.