

**Artificial Intelligence Lab Project Report Project Title:**

**Credit Card Fraud Checker Submitted to:**

**Rasikh Ali**

**Submitted by:**

**MEHREEN SADIQ**

**Section:**

**BSAI(3B)**

**Department:**

**Software Engineering**

# 1. Objective

The goal of this project is to detect fraudulent credit card transactions using a machine learning classification approach. The project employs Logistic Regression to classify transactions as legitimate or fraudulent.

## 2. Dataset Overview

* **Dataset**: creditcard.csv, containing anonymized credit card transaction data.
* **Number of Classes**:
  + Legitimate transactions (Class = 0) o Fraudulent transactions (Class = 1)  **Features**:
  + The dataset contains various anonymized numerical features, which have been transformed using Principal Component Analysis (PCA). The Amount feature represents the transaction amount, and Class is the target variable indicating whether a transaction is legitimate (0) or fraudulent (1).
* **Class Imbalance**:
  + The dataset is highly imbalanced, with a significant number of legitimate transactions and a smaller proportion of fraudulent transactions.

## 3. Data Preprocessing

1. **Missing Values**:
   * The dataset is checked for missing values using credit\_card\_data.isnull().sum(), confirming that there are no missing values in any of the columns.
2. **Class Distribution**:
   * The class distribution is highly imbalanced, with the majority of transactions being legitimate (Class = 0) and fewer fraudulent transactions (Class = 1).
3. **Balancing the Dataset**:
   * To address the class imbalance, a random sample of 492 legitimate transactions is selected to match the number of fraudulent transactions. This ensures that the dataset has a more balanced representation of both classes for model training.

## 4. Feature Selection

* **Independent Variables**: All columns in the dataset except Class are used as independent features.
* **Target Variable**: Class is the target variable, indicating whether the transaction is legitimate (0) or fraudulent (1).

## 1. Objective

The goal of this project is to detect fraudulent credit card transactions using a machine learning classification approach. The project employs Logistic Regression to classify transactions as legitimate or fraudulent.

## 2. Dataset Overview

* **Dataset**: creditcard.csv, containing anonymized credit card transaction data.
* **Number of Classes**:
  + Legitimate transactions (Class = 0) o Fraudulent transactions (Class = 1)  **Features**:
  + The dataset contains various anonymized numerical features, which have been transformed using Principal Component Analysis (PCA). The Amount feature represents the transaction amount, and Class is the target variable indicating whether a transaction is legitimate (0) or fraudulent (1).
* **Class Imbalance**:
  + The dataset is highly imbalanced, with a significant number of legitimate transactions and a smaller proportion of fraudulent transactions.

**3. Data Preprocessing** 1. **Missing Values**:

o The dataset is checked for missing values using credit\_card\_data.isnull().sum(), confirming that there are no missing values in any of the columns.

1. **Class Distribution**:
   * The class distribution is highly imbalanced, with the majority of transactions being legitimate (Class = 0) and fewer fraudulent transactions (Class = 1).
2. **Balancing the Dataset**:
   * To address the class imbalance, a random sample of 492 legitimate transactions is selected to match the number of fraudulent transactions. This ensures that the dataset has a more balanced representation of both classes for model training.

## 4. Feature Selection

* **Independent Variables**: All columns in the dataset except Class are used as independent features.
* **Target Variable**: Class is the target variable, indicating whether the transaction is legitimate (0) or fraudulent (1).

## 5. Model and Methodology

 **Model Used**: Logistic Regression, a linear model used for binary classification.  **Data Split**:

o The dataset is split into training (80%) and testing (20%) sets using the train\_test\_split() function from Scikit-learn. Stratified splitting is used to maintain the same distribution of classes in both sets.

## 6. Results

1. **Training Data Accuracy**:

The Logistic Regression model achieved an accuracy of **94.6%** on the training set.

1. **Testing Data Accuracy**:

The model achieved an accuracy of **90.8%** on the testing set.

1. **Interpretation**:

o The high accuracy on both the training and testing datasets suggests that the model generalizes well and is able to classify transactions effectively, even with the class imbalance. o The accuracy results are very close on both datasets, which indicates that the model is not overfitting and can be relied upon for future predictions.