Credit Card Fraud Detection Using Machine Learning

Overview

Credit card fraud is a major financial issue worldwide. This project aims to detect fraudulent transactions using machine learning techniques. The system analyzes transaction data and classifies each transaction as *fraudulent* or *legitimate* based on various features.

The project applies data preprocessing, balancing using RandomOverSampler, and builds a Random Forest Classifier to predict fraud with high accuracy.

Objectives

- Analyze and preprocess transaction data.
- Handle **imbalanced datasets** using oversampling techniques.
- Train and evaluate a **Random Forest Classifier** model.
- Measure model performance using accuracy, precision, recall, and F1-score.

Dataset Description

- **Dataset name:** Credit Card Fraud Detection Dataset (from Kaggle)
- **Total transactions:** ~284,807
- **Fraudulent transactions:** ~492 (~0.17%)
- **Non-fraudulent transactions:** ~284,315
- **Features:**
    - `Time`: Time elapsed between transactions

- `V1`–`V28`: PCA-transformed numerical features
- `Amount`: Transaction amount
- `Class`: Target variable (1 = Fraud, 0 = Legitimate)

Technologies Used

- **Programming Language:** Python
- **Libraries:**
  - `pandas`, `numpy` – Data handling
  - `matplotlib`, `seaborn` – Data visualization
  - `scikit-learn` – Machine learning
  - `imbalanced-learn` – Handling imbalanced datasets

Conclusion

- Successfully detected fraudulent transactions with high accuracy.
- Data balancing was crucial for fair model training.
- Future improvements could include:
  - Trying **SMOTE** for synthetic oversampling.
  - Using deep learning models (e.g., autoencoders, LSTMs).