Credit Card Fraud Detection Using Machine Learning

Name: Himanshu Raj

College: IGEC

1. Introduction

This project aims to detect fraudulent credit card transactions using machine learning techniques. It combines anomaly detection and classification algorithms to identify suspicious patterns and deploys a web application using Stream lit for real-time fraud prediction.

- Credit card fraud has increased due to rising online and card-based transactions.
- Financial institutions need intelligent systems to prevent loss due to fraud.
- This project uses machine learning to detect fraudulent credit card transactions based on behavioural patterns and anomalies.

2. Abstract

This project focuses on identifying fraudulent credit card transactions using both **unsupervised anomaly detection** and **supervised classification** techniques. The dataset used is from Kaggle, containing anonymized features derived from real transactions.

- The project combines anomaly detection and classification models to spot fraud.
- It uses Isolation Forest and Local Outlier Factor to identify rare patterns.
- A powerful XGBoost classifier is trained on a balanced dataset to predict fraud.
- The final solution includes a web app (Streamlit) where users can upload transaction data for instant prediction

3. Tools Used

- **Programming Language:** Python
- **Libraries:**
    - Data Handling: Pandas, NumPy
    - Machine Learning: Scikit-learn, XGBoost
    - Visualization: Matplotlib, Seaborn
    - Web Deployment: Streamlit
- **IDE/Platforms:** Jupyter Notebook, VS Code
- **Dataset:** Kaggle Credit Card Fraud Dataset

4. Steps Involved in Building the Project

a. Data Collection

- Downloaded the dataset from Kaggle with ~285,000 transactions.
- Features included anonymized transaction data (V1–V28), Time, Amount, and Class.

b. Preprocessing

- Dropped the Time column as it was not predictive.
- Normalized the Amount column using StandardScaler.

c. Data Balancing

- The dataset was highly imbalanced (only 0.17% fraud).
- Undersampled the non-fraud data to create a balanced dataset for training.

d. Anomaly Detection

- Applied **Isolation Forest** and **Local Outlier Factor (LOF)** to flag unusual transaction patterns.
- Helped in identifying rare fraud cases without labelled data.

e. Supervised Learning

- Trained **XGBoost Classifier** on the balanced dataset.
- Model was selected for its accuracy, speed, and performance on imbalanced data.

f. Evaluation

- Evaluated using:
    - **Confusion Matrix**
    - **Accuracy, Precision, Recall, F1-Score**
    - **ROC AUC Score**

g. Deployment

- Saved the model using joblib.
- Created a **Streamlit UI** that accepts a CSV file of transactions.
- Predicts whether each transaction is fraud or not and displays fraud probability.

6. Conclusion

- The system detects fraud with high accuracy using XGBoost.
- Anomaly detection (IF and LOF) helps in identifying rare fraudulent patterns.
- Streamlit interface makes it easy for users to upload and test their own data.
- The model can be extended with real-time pipelines or deep learning for larger systems.