Phase-2

SubmissionTemplate

Student Name: Varshini.A

RegisterNumber: 620123106122

Institution: AVS engineering college

Department: ECE

Date of Submission: 10/05/2025

GithubRepositoryLink:
https://github.com/varshinianandhan/Varshini.git

---

1. ProblemStatement

Credit card fraud poses a significant threat to financial institutions and consumers, resulting in billions of dollars in losses annually. Traditional rule-based systems are inadequate in identifying evolving fraud patterns. This project aims to develop an AI-powered system to detect and prevent fraudulent credit card transactions in real-time using machine learning.

2. ProjectObjectives

To evaluate and deploy various machine learning models for optimal fraud detection performance.

To create a scalable and efficient pipeline suitable for integration with banking systems.

3. Data Description

Implement AES-256 encryption for transaction data storage.

Use SSL/TLS protocols to secure data in transit between systems.

Ensure compliance with GDPR/PCI-DSS standards to protect sensitive customer information.

Apply role-based access control (RBAC) to limit data access during model training and testing.

4. Data Preprocessing

Handle missing values and outliers.

Normalize numerical features (e.g., transaction amount).

Encode categorical variables (e.g., merchant category, location).

Time-based features extracted from timestamps (e.g., hour, weekday).

Apply under-sampling or SMOTE for class imbalance (fraudulent vs. non-fraudulent).

5. Exploratory Data Analysis (EDA)

The first step in building an AI-powered credit card fraud detection system is to thoroughly understand the dataset through Exploratory Data Analysis (EDA). This process provides insights into the data distribution, patterns, and potential anomalies that might be indicative of fraudulent behavior.

6. FeatureEngineering

Derive new features such as:

Frequency of transactions per card in a time window.

Unusual location or device usage.

Rapid transaction velocity.

Use correlation matrix and mutual information to select the most predictive features.

7. Model Building

Models used:

Logistic Regression (baseline)

Random Forest

XGBoost

Isolation Forest (for unsupervised anomaly detection)

Neural Networks (e.g., LSTM for sequential patterns)

Evaluation metrics:

Accuracy, Precision, Recall, F1-score, AUC-ROC

Focus on minimizing false negatives (missed frauds)

8. VisualizationofResults&Mode

Effective visualization plays a crucial role in evaluating the performance of credit card fraud detection models, especially due to the highly imbalanced nature of the dataset. The following visualizations were used to interpret the results and understand the model's effectiveness in detecting fraudulent transactions.

9. Tools andTechnologies Used

Languages: Python

Libraries: Scikit-learn, Pandas, NumPy, XGBoost, TensorFlow/Keras

Visualization: Matplotlib, Seaborn, Plotly

Security: PyCrypto, OpenSSL

Deployment: Flask/Django (API), Docker, AWS/GCP

Database: PostgreSQL, MongoDB

10.Team Membersand Contributions

The project was a collaborative effort. Each team member brought unique expertise to ensure the success of the AI-powered fraud detection system.

Renuka.E

Role: Data Analyst / EDA Specialist

Varshini.A

Role: Machine Learning Engineer

Sowndarya priya.K

Role: Data Engineer

Theerthana.V

Role: Project Manager / Presentation Lead