Github Link: https://github.com/nandhajan/project-Guarding-transactions-with-AI.git

Project Title: Guarding transactions with AI -powered credit fraud detection and prevention

PHASE-2

- **Problem Statement**

With the rapid growth of digital payments and online transactions, credit card fraud has become a significant challenge for financial institutions and customers alike. Traditional fraud detection methods often rely on rule-based systems, which are limited in their ability to adapt to evolving fraud patterns and can lead to high false-positive rates, frustrating legitimate customers. There is a pressing need for an AI-powered solution that can analyze large volumes of transactional data in real time, accurately identify suspicious activities, and prevent fraudulent transactions without disrupting the user experience. This project aims to develop and implement an AI-driven credit fraud detection and prevention system that enhances security, reduces financial losses, and improves customer trust.

- **Project Objectives**

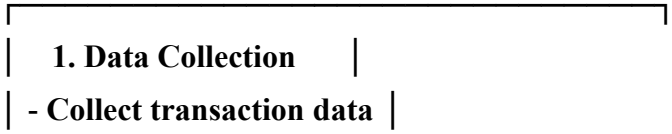
To analyze existing patterns of credit card transactions and identify key indicators of fraudulent behavior using historical data.

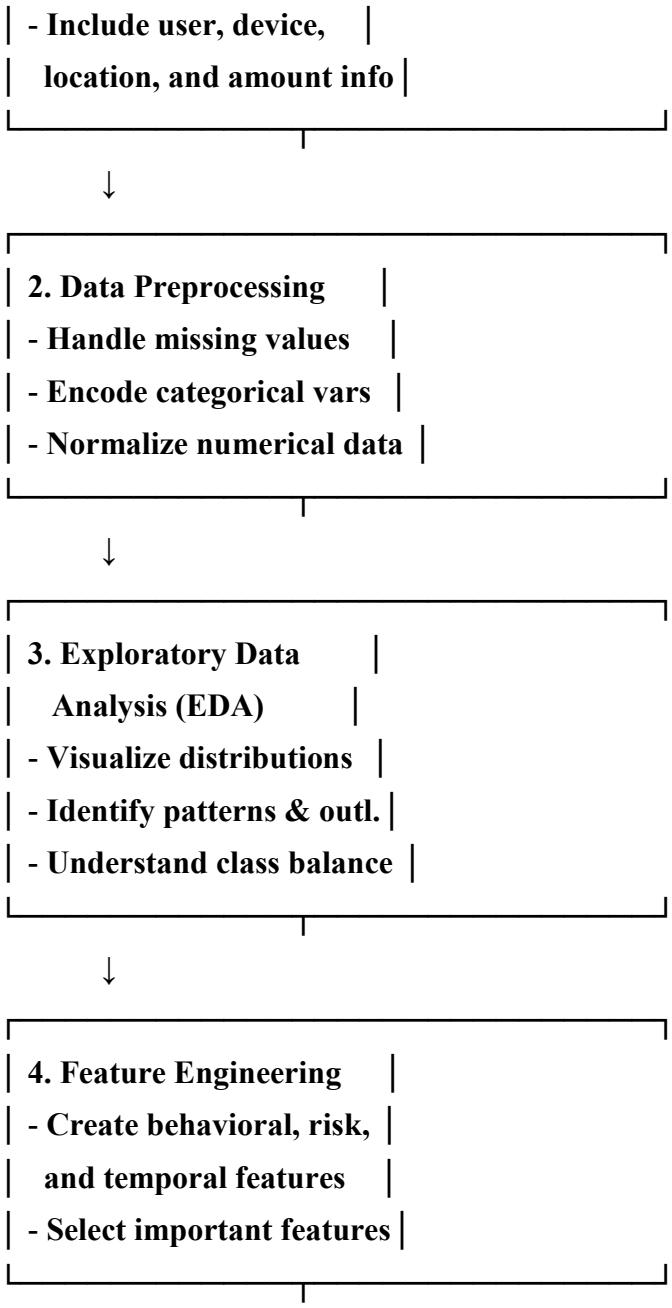
To design and develop a machine learning model capable of detecting and predicting fraudulent transactions in real time with high accuracy.

To minimize false positives and false negatives in fraud detection, ensuring that legitimate transactions are not wrongly flagged while maximizing the detection of fraudulent activities.

To integrate the AI-based fraud detection system into the transaction processing pipeline without significantly impacting transaction speed or user experience.

- **Flowchart of the Project Workflow**





- **Data Description**

Transaction Data

Transaction ID → Unique identifier for each transaction

Timestamp → Date and time when the transaction was made

Amount → Total value of the transaction

Merchant ID → Identifier of the merchant or store

Transaction Type → Purchase, withdrawal, transfer, etc.

Payment Method → Credit card, debit card, online wallet, etc.

Currency → Currency used in the transaction

- **Data Preprocessing**

1. Data cleaning

Remove duplicates → Drop duplicate transactions that may result from system errors.

Handle missing values →

Impute missing values using statistical methods (mean, median) or domain rules.

Remove records with excessive missingness if needed.

Correct data types → Ensure numeric, categorical, and date fields have correct formats.

2. Data Transformation

Convert timestamps → Extract useful time features (hour of day, day of week, weekend/weekday).

Normalize or scale numeric features → Apply min-max scaling or standardization on features like amount to handle wide value ranges.

- **Exploratory Data Analysis (EDA)**
-
- **1. Understand Dataset Overview**
- Check data dimensions → Number of rows (transactions) and columns (features).
- Examine feature types → Identify categorical, numerical, and datetime variables.
- Check missing values → Percentage of missing data per feature**.**
- **2. Summary Statistics**
- Numerical features → Calculate mean, median, min, max, standard deviation for amount, transaction time, etc.
- Categorical features → Count frequencies and proportions for transaction type, merchant ID, payment method.
- **Feature Engineering**
- **1. Basic Features (from raw data)**
- Transaction amount → Raw amount of the transaction.
- Transaction type → Purchase, transfer, withdrawal, etc.
- Timestamp features → Extract:
- Hour of day

- Day of week
- Weekend vs. weekday
- Holiday flag
- Merchant category → Type of merchant (e.g., retail, travel, electronics).
- Payment method → Credit card, debit card, digital wallet.
- **Model Building**

1. Define the Modeling Objective

Goal: Predict whether a transaction is fraudulent (fraud = 1) or legitimate (fraud = 0).

Type of problem: Binary classification.

2. Select Machine Learning Algorithms

Choose one or a combination of the following algorithms:

✅Baseline Models (for benchmarking)

Logistic Regression

Decision Trees

✅Advanced Models (for better performance)

Random Forest

Gradient Boosting (XGBoost, LightGBM, CatBoost)

Neural Networks (especially for large, complex datasets)

- **Visualization of Results & Model Insights**
- **What it shows:**
- **N1. Confusion Matrix**
- Number of true positives (fraud correctly predicted)
- Number of false positives (legitimate flagged as fraud)
- Number of true negatives (legitimate correctly predicted)
- Number of false negatives (fraud missed)
- How to visualize:
- Use a heatmap to show counts or percentages**.**
- **✅2. ROC Curve (Receiver Operating Characteristic)**

- **What it shows:**

- Trade-off between true positive rate (recall) and false positive rate at various thresholds.

- **Tools and Technologies Used**
-
- ✅**Programming Languages**
- Python → Main language for data processing, modeling, and deployment
- R → (Optional) For statistical analysis and visualization
- ✅**Data Processing and Analysis**
- pandas → Data cleaning, manipulation, and analysis
- NumPy → Numerical operations and array processing
- SQL → Extracting transactional data from databases
-
- **Team Members and Contributions**

1. M.NANDHAKUMAR — Data Scientist

Performed data cleaning, preprocessing, and feature engineering

Built and tuned machine learning models (Random Forest, XGBoost)

Conducted exploratory data analysis (EDA) and extracted actionable insights

Evaluated model performance using precision, recall, F1, and AUC metrics

✅2. S.PRAKASHRAJ — Machine Learning Engineer

Implemented ML pipelines for training and validation

Integrated model with backend systems using Flask API

Optimized model for real-time prediction and scalability

Developed Docker containers for deployment in cloud environments

✅3. K.PANNEERSELVAM — Data Analyst / Visualization Expert

Created detailed visualizations for EDA and model insights (confusion matrix, ROC curves, feature importance)

Developed dashboards using Plotly and Tableau for monitoring fraud patterns

Provided reports on key findings and recommendations for business stakeholders

✅4. M.PERUMAL — Project Manager / Domain Expert

Coordinated team activities, timeline, and task management

Acted as liaison with financial domain experts and stakeholders

Defined problem statement, project scope, and success criteria

Reviewed final deliverables and ensured compliance with regulatory standards