Major Project - II
CS- 805

Credit Card Fraud Detection using ML in Python

Problem Statement: This project develops a machine learning model to identify fraudulent credit card transactions in a given dataset. The system aims to minimize both false positives and false negatives to help financial institutions.

Project Group Number: 30

Group Member 1: Ashutosh Rathore (0187CS211041)

Group Member 2: Shreya Singh Rathore (0187CS211160)

Group Member 3: Chirag Dewangan (0187CS211055)

Group Member 4: Ajay Meena (0187CS211016)

Guide Details: Prof. Mayank Kurchaniya

Idea/Approach Details

Describe your idea Solution/Prototype here:

- Our project aims to detect fraudulent credit card transactions using machine learning techniques, specifically Logistic Regression. The system processes real-time transaction data, applies data preprocessing and feature selection, and trains a classification model to identify potential fraud. The trained model analyzes transactions based on parameters like amount, location, and past history to classify them as fraudulent or legitimate. The backend is built using Python, with scikit-learn for model training, and the interface will provide alerts when fraudulent activities are detected. This system enhances financial security by reducing false positives and improving fraud detection accuracy.

Abstract:

- Credit card fraud is a significant challenge in financial security, requiring advanced solutions to detect and prevent fraudulent transactions. This project leverages Machine Learning, specifically Logistic Regression, to analyze transaction data and classify it as fraudulent or legitimate. The system undergoes data preprocessing, feature selection, and model training to improve accuracy. Using real-time transaction inputs, the model predicts fraudulent activities based on historical patterns. The implementation is done in Python, utilizing libraries like scikit-learn for training and evaluation.

Describe your Technology stack here:

- **Programming Language:** Python is used for data processing, model training, and evaluation due to its vast ecosystem of ML libraries.
- **ML Libraries:** pandas, numpy, sklearn, seaborn, matplotlib
- **Machine Learning Algorithm:** Logistic Regression, a classification algorithm used to predict fraudulent transactions.

Project Requirements

Functional Requirements

- **Data Ingestion:** The system should accept credit card transaction data in CSV or real-time format.
- **Data Preprocessing:** Handle missing values, normalize data, and remove duplicates before training.
- **Fraud Detection Model:** Implement Logistic Regression to classify transactions as fraudulent or legitimate.
- **Visualization & Analysis:** Display fraud detection insights using matplotlib and seaborn.
- **Model Evaluation:** Use accuracy, precision, recall, and F1-score to assess performance.
- **Alert Generation:** Trigger notifications when fraudulent transactions are detected.

Non-Functional Requirements

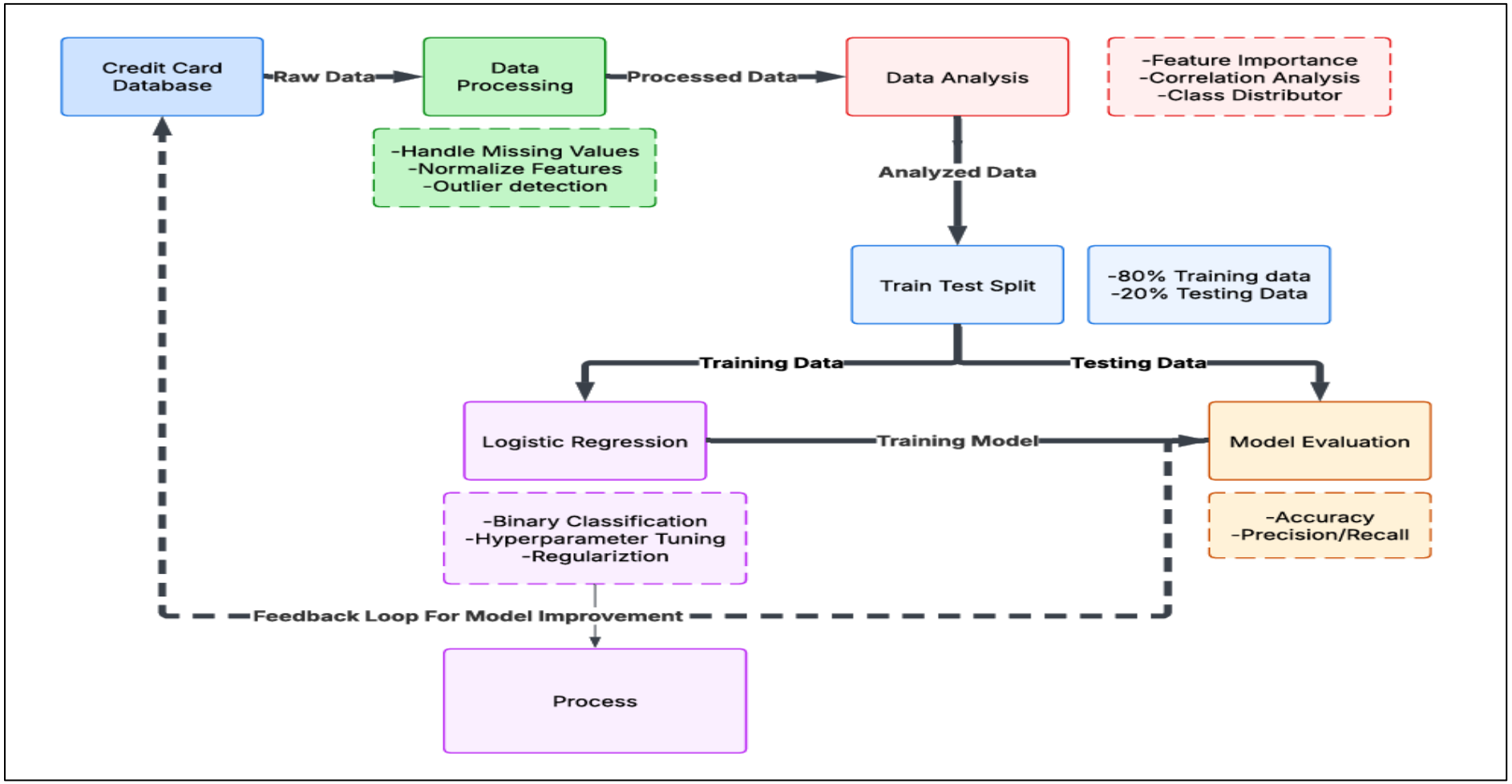
- **Scalability:** The system should handle large datasets efficiently.
- **Performance:** Predictions should be made in real-time with minimal latency.
- **Security:** Ensure transaction data is processed securely and remains confidential.
- **Usability:** Provide a simple interface (if applicable) for non-technical users.
- **Reliability:** The model should consistently detect fraud with high accuracy.
- **Maintainability:** The code should be modular and well-documented for future improvements.

Hardware and Software requirements

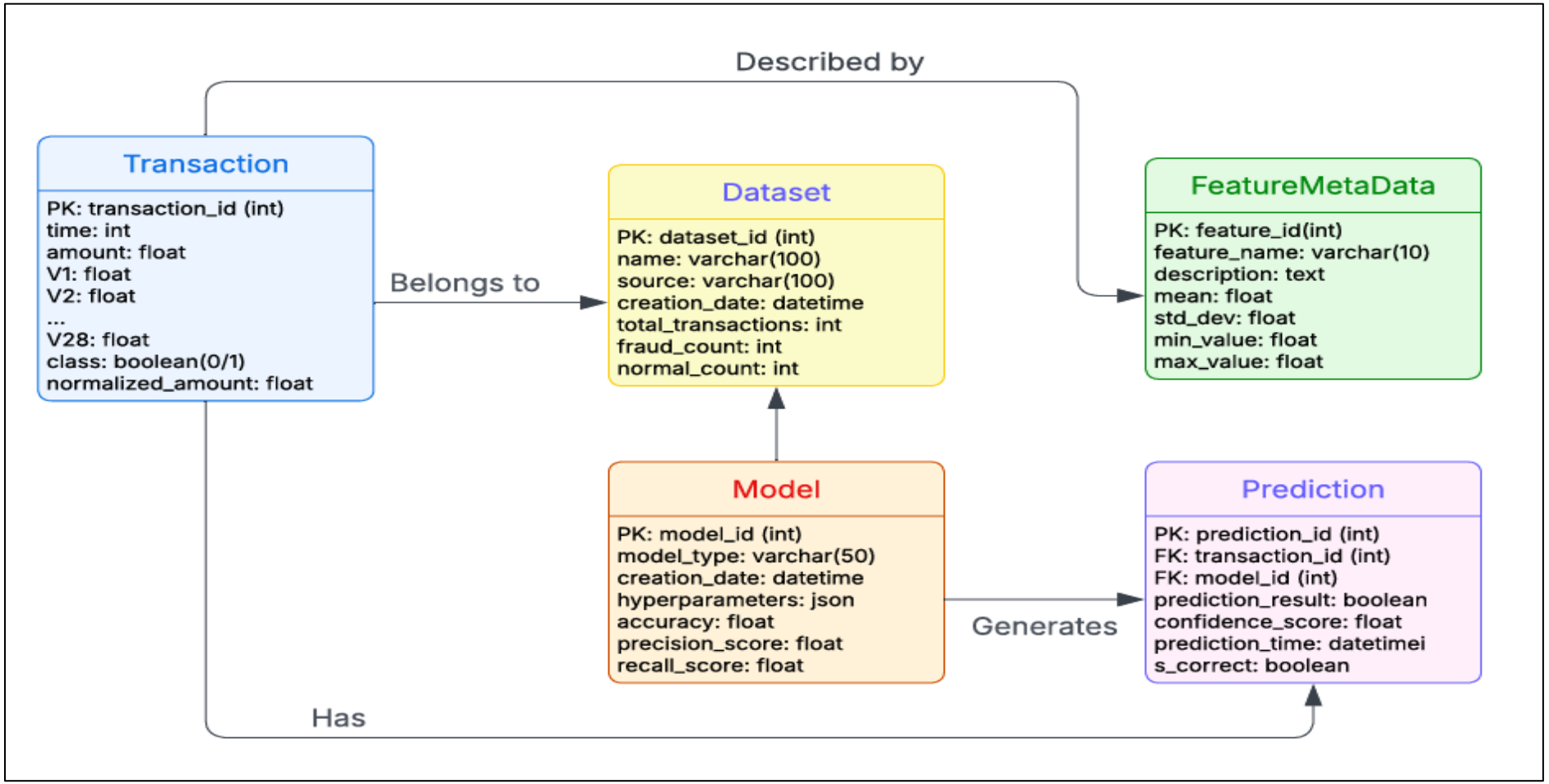
- **Processor:** Intel Core i5/i7 or AMD equivalent (for efficient model training).
- **RAM:** Minimum 8GB (16GB recommended for handling large datasets).
- **Display Monitor:** A high-resolution screen for better visualization and data analysis.
- **Internet Connection:** Required for fetching datasets, cloud deployment, and API hosting.
- **Programming Language:** Python (with libraries like pandas, numpy, sklearn, etc.)
- **Machine Learning Libraries:** Scikit-learn, pandas, numpy, seaborn, matplotlib.
- **Machine Learning Framework:** Scikit-learn for Logistic Regression.
- **IDE:** Google Colab / VS Code / Jupyter Notebook (for writing and testing the model).

Design

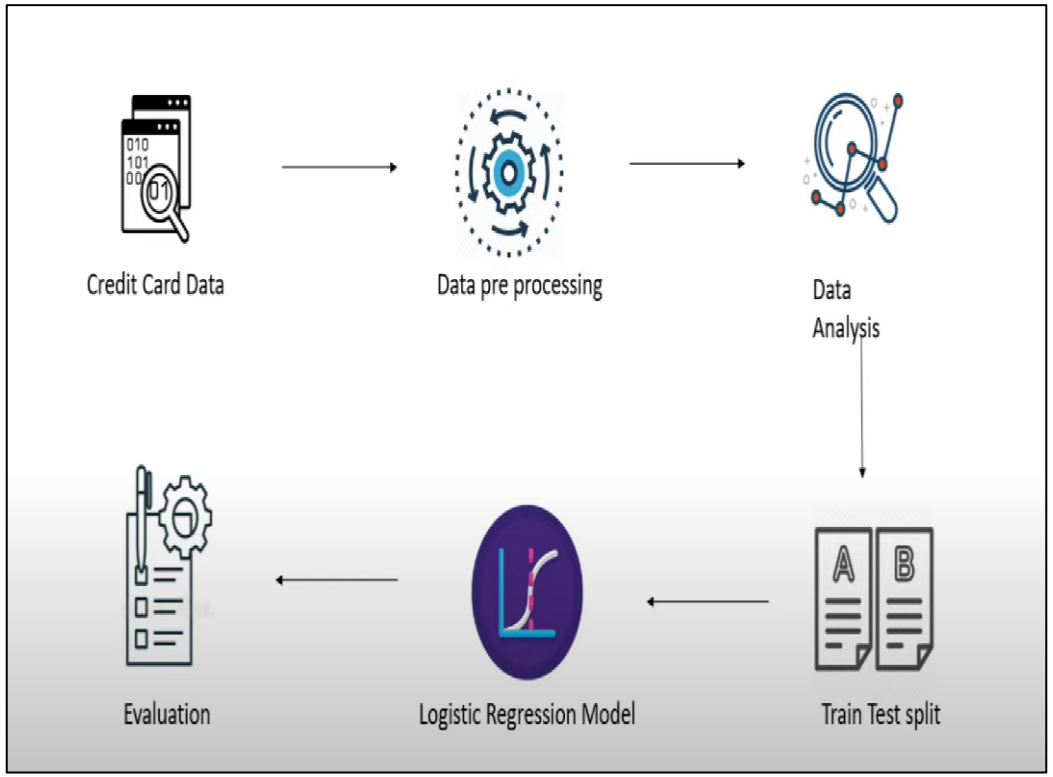
Describe data flow diagram here



Describe your ER Diagram here



Workflow



➢ **Credit Card Data Collection:**
•The dataset containing transaction details is obtained.
•Includes features like transaction amount, time, and anonymized parameters.

➢ **Data Pre-processing:**
•Handles missing values and duplicates.
•Normalizes and scales data for better model performance.

➢ **Data Analysis:**
•Exploratory Data Analysis (EDA) is performed using visualization tools.
•Identifies patterns and imbalances in fraudulent and non-fraudulent transactions.

➢ **Train-Test Split:**
•The dataset is divided into training and testing sets.
•Ensures a balanced ratio for accurate evaluation.

➢ **Model Training (Logistic Regression):**
•The Logistic Regression algorithm is applied to learn transaction patterns.
•Model is trained on the training dataset.

➢ **Evaluation:**
•The trained model is tested on the test dataset.
•Performance is measured using accuracy, precision, recall, and F1-score.


Thank you!