Credit Card Fraud Detection

Dataset Overview

The **Fraud Detection Dataset** contains anonymized data related to financial transactions, with labels indicating whether each transaction is fraudulent. The key features include transactional details and metadata associated with each transaction, although specifics are masked for privacy.

Project Structure

1. Problem Statement

Objective: To detect fraudulent financial transactions using a machine learning model that can classify transactions as either “fraud” or “non-fraud.”

Solution Approach

The solution will follow a machine learning approach to detect fraud. The model will be trained on historical transaction data and will be deployed for real-time classification.

Steps should be performed :

1. Data Collection and Initial Exploration

- Dataset Link: Dataset
- Data Overview: Import the dataset and inspect it to understand the number of features, labels, and any patterns.
- Check info, describe , Null values, column names, Duplicate columns or rows
- Check and Handle Null values if present

2.Data Preprocessing:

- Drop unwanted columns
- Handling Imbalance data
- Encoding Categorical Features: If there are any categorical features, encode them using techniques like One-Hot Encoding,Label Encoding.

3. Exploratory Data Analysis (EDA)

- Check Data Distribution
- Fraud vs. Non-Fraud Transactions: Examine the number of fraudulent vs. non-fraudulent transactions.

- Correlations and Patterns: Use heatmaps or pair plots to analyze correlations between features.

4. Visualisation:

- Perform visualisation on different columns in the dataset
- Must use: Boxplot, countplot, Heatmap, Pairplot, barplot

5. Feature Scaling

6. Model Selection & Fitting:

- Train Test split , Fit any 4 model suitable for the dataset and compare the accuracies

7. Model Evaluation :

- Classification report, F1 SCORE, Accuracy score, Confusion matrix

YOU CAN PERFORM ADDITIONAL STEPS ALSO.