Documentation Credit Card Fraud Detection Project

1. Overview

This project aims to detect fraudulent credit card transactions using machine learning models. The goal is to analyze transaction data to identify patterns and behaviors indicative of fraud.

---

2. Dataset Description

- **Source**: The dataset is downloaded from Kaggle.
- **Files**:
  - fraudTrain.csv: Training dataset.
  - fraudTest.csv: Testing dataset.
- **Features**:
  - **Transaction-specific data**:
    - Transaction amount, location, and time.
  - **Customer data**:
    - Age, gender, and cardholder details.
  - **Outcome**:
    - is_fraud: Target variable indicating fraudulent transactions (1) or legitimate ones (0).

---

3. Libraries and Tools

- **Pandas**: For data manipulation and analysis.
- **Matplotlib and Seaborn**: For data visualization.
- **Folium**: For creating geographic visualizations.
- **Datetime**: To handle and process temporal data.

---

4. Steps in the Notebook

4.1. Data Acquisition

- The dataset is downloaded using Kaggle's API and extracted from a compressed file.

4.2. Data Loading

- Training and testing datasets are loaded into Pandas DataFrames for analysis.

4.3. Exploratory Data Analysis (EDA)

- **Visualizations**:
  - Distributions of transaction amounts.
  - Geographic mapping of fraudulent transactions.
  - Correlation analysis between features.
- **Insights**:
  - Patterns in fraudulent transactions by time, location, or amount.

4.4. Data Preprocessing

- Handling missing or inconsistent data.
- Encoding categorical features and scaling numerical ones.

4.5. Model Building

- Selecting and training machine learning models for fraud detection.

4.6. Model Evaluation

- Evaluating model performance using metrics like precision, recall, and F1-score.

---

5. Visualizations

- **Heatmap**: To identify geographic fraud hotspots.
- **Bar Charts**: For categorical feature distributions.
- **Line Plots**: Analyzing transaction trends over time.

---

6. Instructions for Running the Notebook

1. Install dependencies: pip install pandas matplotlib seaborn folium kaggle
1. Ensure you have the Kaggle API key configured to download datasets.
2. Run the notebook cells sequentially to reproduce the results.

---

7. Key Insights and Resolutions

- **High-risk patterns**:
  - Transactions with high amounts and specific locations/time are more likely to be fraudulent.
- **Resolutions**:
  - Deploy stricter monitoring for flagged transactions.

- Use customer verification steps for suspicious transactions.