Project Planning Phase – Online Payments Fraud Detection

1. Introduction:

The **Project Planning Phase** outlines the **systematic approach** for implementing the Online Payments Fraud Detection project. It ensures that the project is executed in a **logical sequence**, minimizing errors and optimizing time and resources.

This phase defines the **planning logic**, the steps involved, and the milestones that guide the project from **data acquisition to deployment**.

2. Planning Logic:

The **planning logic** describes the **step-by-step approach** to achieve the project objectives efficiently:

1. **Understanding the Problem:**
    - Analyze the challenges of online payment fraud.
    - Define project goals: detect fraud in real-time with high accuracy.
2. **Data Collection:**
    - Identify the source of the dataset (fraud_data.csv).
    - Ensure the dataset contains all necessary features:
        - step, type, amount, oldbalanceOrig, newbalanceOrig, oldbalanceDest, newbalanceDest, isFraud.
    - Verify dataset quality and authenticity.
3. **Data Cleaning:**
    - Check for **null values or missing data**.
    - Handle **duplicate entries** or inconsistencies.
    - Detect and manage **outliers** using statistical methods (e.g., IQR).
    - Encode **categorical features** for machine learning compatibility.
4. **Exploratory Data Analysis (EDA):**
    - Visualize feature distributions to understand trends and patterns.
    - Identify correlations between features and the target variable (isFraud).
    - Segment transactions by type, amount, and account balances to detect **fraud-prone patterns**.

5. **Feature Selection & Engineering:**
    - Select the most **informative features** for model training.
    - Engineer new features if necessary (e.g., transaction ratio, account balance differences).
6. **Model Training & Testing:**
    - Split the dataset into **training and testing sets**.
    - Train multiple models (Random Forest, SVM, Decision Tree, Extra Trees).
    - Evaluate performance using **accuracy, precision, recall, and confusion matrix**.
7. **Web Application Development:**
    - Develop **Flask-based UI** for user interaction.
    - Integrate the trained model to provide **real-time predictions**.
    - Validate user inputs and display results clearly.
8. **Deployment & Maintenance:**
    - Save the trained model with **Pickle**.
    - Deploy the application on a server or local environment.
    - Continuously monitor predictions and update the model with **new transaction data**.

3. Detailed Explanation of Each Step

3.1 Data Collection

- **Purpose:** To gather authentic and complete transaction records for analysis.
- **Activities:**
    - Download fraud_data.csv from Kaggle or Google Sheets.
    - Verify columns: step, type, amount, oldbalanceOrig, newbalanceOrig, oldbalanceDest, newbalanceDest, isFraud.
    - Ensure sufficient sample size to train machine learning models.

3.2 Data Cleaning

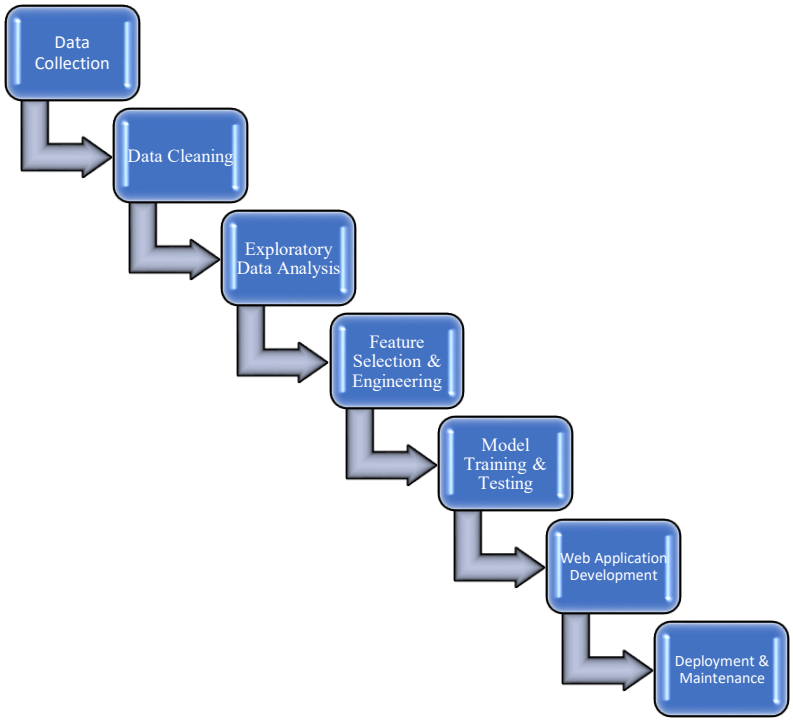
- **Purpose:** To ensure the dataset is accurate, consistent, and ready for analysis.
- **Activities:**
    - Remove duplicates.

- Handle missing or null values (if any).
- Detect outliers in amount and balance fields and handle them.
- Encode categorical variables like type using **LabelEncoder**.

3.3 Project Planning

- **Purpose:** To define milestones, allocate resources, and schedule tasks.
- **Activities:**
  1. **Milestone 1:** Data Collection (1 day)
  2. **Milestone 2:** Data Cleaning & Preprocessing (2 days)
  3. **Milestone 3:** Exploratory Data Analysis (1-2 days)
  4. **Milestone 4:** Feature Selection & Model Training (2 days)
  5. **Milestone 5:** Model Evaluation & Optimization (1 day)
  6. **Milestone 6:** Web Application Development (2 days)
  7. **Milestone 7:** Deployment & Testing (1 day)
- **Logic:** Tasks are organized **sequentially**: clean and preprocess the data before training models, evaluate models before integrating them into the web app, and finally deploy after successful testing.

4. Planning Logic Diagram (Text Representation)



5. Conclusion

- The **Project Planning Phase** ensures that the Online Payments Fraud Detection project follows a **structured and logical workflow**.
- The **planning logic** guarantees that data is properly collected, cleaned, and preprocessed before model training.
- Proper milestone scheduling and step-by-step execution **minimizes errors and improves efficiency**.
- This phase lays a **strong foundation for the development and deployment** of a robust fraud detection system.