

**Project Design Phase**  
**Proposed Solution Template**

Date	
Team ID	LTVIP2026TMIDS40761
Project Name	Online payments fraud detection using ml
Maximum Marks	2 Marks

**Proposed Solution:**

In this project, the Decision Tree algorithm is used to classify online payment transactions as Fraudulent or Legitimate. The model works by creating a tree-like structure where each node represents a decision based on transaction features such as amount, transaction type, balance changes, and sender/receiver details.

The Decision Tree splits the dataset into smaller groups by selecting the best feature at each step using measures like Gini Index or Information Gain. By following these decision rules from the root to the leaf node, the model predicts whether a transaction is fraud or not.

Decision Trees are effective for fraud detection because they are easy to interpret, provide clear decision rules, and help identify the important factors that lead to fraudulent transactions.

A Decision Tree is a supervised machine learning algorithm that is widely used for classification problems, especially when the goal is to predict a categorical output such as Fraud (1) or Not Fraud (0).

In the Online Payments Fraud Detection system, the Decision Tree model learns patterns from historical transaction data. Each transaction contains multiple features such as:

- Transaction type (TRANSFER, CASH\_OUT, PAYMENT, etc.)
- Transaction amount
- Old balance and new balance of the sender
- Old balance and new balance of the receiver
- Other transaction-related details

The Decision Tree algorithm builds a tree-like structure by repeatedly splitting the dataset into smaller groups based on the most important features. At each step, the model selects the best feature for splitting using impurity measures such as:

- ◆ Splitting Criteria Used
  - Gini Index (commonly used in sklearn)
  - Entropy
  - Information Gain

The aim is to create splits that separate fraudulent and non-fraudulent transactions as clearly as possible. Each internal node represents a condition (example: *amount > 200000*), and each branch represents the outcome of that condition. Finally, the leaf node gives the final classification, either Fraud or Legitimate.

---

## Why Decision Tree Works Well for Fraud Detection

Decision Trees are very useful in fraud detection because:

- They can learn non-linear patterns (fraud is not always linear)
  - They can handle mixed type data (numeric + categorical)
  - They give interpretable rules, which is important for banking and security
  - They identify the most important fraud-related features automatically
- 

### Working Process in This Project

#### 1. Dataset Loading

- Online payments dataset is loaded and cleaned.

#### 2. Preprocessing

- Categorical columns (like transaction type) are converted using encoding.
- Missing values are handled.
- Features (X) and target (y) are separated.

#### 3. Training

- The Decision Tree is trained using training data.

#### 4. Prediction

- The model predicts fraud on unseen test transactions.

#### 5. Evaluation

- Performance is evaluated using:
  - Accuracy
  - Confusion Matrix
  - Precision, Recall, F1-score

---

### Advantages of my Project

-  Easy to explain in PPT
-  Gives clear decision-making structure
-  Works well for detecting fraud patterns
-  Helps understand which transaction features cause fraud