Project Planning Phase
Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	
Team ID	LTVIP2026TMIDS40761
Project Name	Online payments fraud detection using ml
Maximum Marks	5 Marks

Data Collection is the first step in the machine learning process where the required dataset is gathered from a trusted source. In this project, data is collected in the form of **online payment transaction records**, which contain both legitimate and fraudulent transactions.

The collected dataset includes important transaction details such as:

- Transaction type (TRANSFER, CASH_OUT, PAYMENT, etc.)

- Transaction amount

- Sender old balance and new balance

- Receiver old balance and new balance

- Fraud label (0 = Legitimate, 1 = Fraud)

This collected data becomes the base for the entire project because the machine learning model learns fraud patterns only from this dataset.

---

2) Data Preprocessing (Detailed)

Data preprocessing is the process of cleaning and transforming the collected raw data into a format that can be used effectively for machine learning model training.

Since real-world online payment data may contain missing values, categorical text, noise, and imbalance, preprocessing improves the quality of the dataset.

- ◆ **Steps included in preprocessing:**

1. Checking missing values

- Detecting null/NaN values and handling them properly.

2. Removing duplicates

- Removing repeated transactions to avoid bias in training.

3. Encoding categorical data

- Converting text columns like **transaction type** into numeric values using:

    o Label Encoding

    o One Hot Encoding

4. Feature selection

- Removing unnecessary columns and selecting only important features that affect fraud prediction.

5. Handling imbalanced data

- Fraud transactions are fewer, so methods like:

    o SMOTE

    o Class weighting
      are used to balance the dataset.

6. Splitting the dataset

- Dividing data into:

    o Training set (for learning)

    o Testing set (for evaluation)

✅ Data Collection

- Data collection is the process of gathering transaction data from a reliable source.
- In this project, the dataset contains online payment transaction records.
- It includes features like transaction type, amount, sender/receiver balances, and fraud label.
- The collected data is used for training and testing the fraud detection model.

✅ Data Preprocessing

- Data preprocessing converts raw transaction data into a clean and usable format.
- Missing values and duplicate records are checked and removed/handled.
- Categorical features (like transaction type) are encoded into numerical values.
- Important features are selected for better prediction.
- Imbalanced data is handled because fraud cases are very less.
- Dataset is split into training and testing sets for model evaluation.

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

https://www.visual-paradigm.com/scrum/scrum-burndown-chart/

https://www.atlassian.com/agile/tutorials/burndown-charts

Reference:

https://www.atlassian.com/agile/project-management

https://www.atlassian.com/agile/tutorials/how-to-do-scrum-with-jira-software

https://www.atlassian.com/agile/tutorials/epics

https://www.atlassian.com/agile/tutorials/sprints

https://www.atlassian.com/agile/project-management/estimation

https://www.atlassian.com/agile/tutorials/burndown-charts