

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 18-06-2025 |
| Team ID | SWTID1749841176 |
| Project Title | Online Payments Fraud Detection using Machine Learning |
| Maximum Marks | 6 Marks |

Data Collection Plan and Raw Data Sources Identification Report:

The dataset used was sourced from Kaggle, containing real-world online payment transaction records. It includes features like transaction amount, time, and anonymized user behaviour indicators. The data was downloaded as a CSV file and serves as the raw input for analysis and model training.

Data Collection Plan:

| Section | Description |
|------------------------------------|--|
| Project Overview | This project aims to build a machine learning model to detect fraudulent transactions in online payment systems. Using a real-world dataset of financial transactions, the system analyses feature like transaction type, amount, and account balances to identify suspicious patterns. The goal is to develop an efficient, automated fraud detection solution that can flag potentially fraudulent activities in real time, improving security and reducing financial losses for users and service providers. |
| Data Collection Plan | <ul style="list-style-type: none"> Collected from a publicly available dataset on Kaggle containing simulated online payment transactions with labelled fraud instances. Acquired as a CSV file with attributes like transaction type, amount, origin/destination IDs, and account balances, ready for preprocessing and analysis. |
| Raw Data Sources Identified | The raw data for this project was sourced from a publicly available Kaggle dataset titled “ Online Payments Fraud Detection Dataset ” which contains synthetic yet realistic online payment transaction records. The dataset includes detailed attributes such as transaction type, amount, sender and receiver account balances, and fraud indicators. It provides a balanced foundation for analysing transactional behaviour and training machine learning models to identify fraudulent activities effectively. |

Raw Data Sources Report:

| Sources Name | Description | Location/URL | Format | Size | Access Permissions |
|--------------|-------------|--------------|--------|------|--------------------|
|--------------|-------------|--------------|--------|------|--------------------|

| | | | | | |
|-----------------------|--|---|-----|--------|--------|
| Kaggle Dataset | The dataset contains anonymized records of online financial transactions, including details like transaction type, amount, sender/receiver balances, and fraud labels, used to detect fraudulent activity. | https://www.kaggle.com/datasets/rupakroy/online-payments-fraud-detection-dataset | CSV | 470 MB | Public |
|-----------------------|--|---|-----|--------|--------|