Data Collection and Preprocessing Phase

Date	17 June 2025
Team ID	SWTID1749710444
Project Title	Online Payment Fraud Detection using ML
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan

Section	Description
Project Overview	The main objective of this analysis is prediction: to accurately identify and prevent fraudulent online payment transactions using machine learning classification models. This predictive approach enables businesses to minimize financial losses, reduce operational risk, and enhance customer trust by proactively flagging suspicious transactions in real time.
Data Collection Plan	The dataset given in the project workspace will be used.
Raw Data Sources Identified	The raw dataset for this project is obtained from Kaggle. The dataset consists of over 6 million online payment transactions, each described by transaction details and account balances before and after the transaction.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset simulates over 6 million online payment transactions, including details like transaction type, amount, and account balance changes. Each transaction is labeled for fraud, but the data is highly imbalanced with less than 0.15% fraudulent cases, presenting a realistic modeling challenge.	https://www.kaggle.com/datasets/rupakroy/online-payments-fraud-detection-dataset	CSV	470 MB	Public