

Data Collection and Preprocessing Phase

Date	7 July 2025
Team ID	SWTID1749620997
Project Title	Online Payments Fraud Detection
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification

Elevating 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 goal of this machine learning project is to detect fraudulent financial transactions using supervised learning techniques on a large dataset.
Data Collection Plan	The dataset was sourced from publicly available repositories focused on financial fraud detection. The primary dataset used was a CSV file.
Raw Data Sources Identified	The raw dataset was downloaded from Kaggle and includes transactional details such as transaction amount, origin and destination account balances, and fraud indicators.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Synthetic dataset containing transaction records for fraud detection.	https://www.kaggle.com/datasets/evalaxi/paysim1	CSV	~200 MB	Public (free Kaggle account)