Data Collection and Preprocessing Phase

Date	10 April 2025
Team ID	259453
Project Title	SMS Spam Detection using NLP
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

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 Template

Section	Description
Project Overview	This project aims to build an SMS spam detection system using Natural Language Processing (NLP) techniques. It classifies messages as 'spam' or 'ham' based on their content.
Data Collection Plan	The dataset is collected from a publicly available CSV file containing labeled SMS messages. Each entry has a label (`spam` or `ham`) and the message content..
Raw Data Sources Identified	A public SMS Spam Collection dataset from online repositories was used. It provides real-world examples of spam and ham messages.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Labeled SMS messages for spam detection. Each row has a label (spam/ham) and the message content.	https://drive.google.com/file/d/1K4hMBJ3oMklTtxoyykgT7YPNdMh3ur1Q/view	CSV	5.25 MB	Public