

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	LTVIP2024TMID24955
Project Title	SMS Spam Detection - AIML
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	The project focuses on building an SMS spam detection model that accurately classifies incoming SMS messages as spam or not spam. This involves collecting a diverse set of SMS data, including both spam and non-spam messages, to train, validate, and test machine learning models.
Data Collection Plan	Identify sources of SMS data such as public datasets, user-contributed data, or SMS archives from mobile service providers. Data can be collected from Kaggle datasets
Raw Data Sources Identified	The dataset will consist of SMS message content (text data), along with metadata like message timestamp, sender information, and spam/not spam labels

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The data consists of meta data	https://drive.google.com/file/d/1Mf6AHsRuRxO8IynQFJGYi2OWwO4uWVh-/view?usp=sharing	CSV	5374 KB	Public