

Data Collection and Preprocessing Phase

Date	Nov 2024
Team ID	Team-739662
Project Title	Chatbot based on Data Science Enquiry 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	To develop a robust and intelligent chatbot capable of understanding and responding to user queries in a natural language manner, leveraging the power of Natural Language Processing (NLP). This chatbot will be designed to provide accurate, relevant, and timely information to users, enhancing their overall user experience.
Data Collection Plan	To effectively train a chatbot capable of handling data science inquiries using natural language processing (NLP), a robust data collection strategy is crucial. The collected data should be diverse, relevant, and representative of the kinds of queries the chatbot will encounter.
Raw Data Sources Identified	For a data science enquiry chatbot, raw data sources can be diverse and depend on the specific focus of the chatbot. However, some potential sources include academic papers, research articles, code repositories like GitHub, discussion forums like Stack Overflow, and data science blogs. Additionally, datasets from Kaggle and other repositories can be used for training and testing purposes.

Raw Data Sources Template

Source Name	Description	Location	Format	Access Permissions
Dataset 1	In this we use data set as .yaml file.	NA	YAML	Public