**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | 739890 |
| Project Title | Bookmate: A Chatbot Companion For Book Recommendations |
| 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 sophisticated chatbot, "Bookmate," that leverages AI and machine learning to provide personalized book recommendations to users based on their preferences, reading history, and real-time queries. |
| Data Collection Plan | To train and improve the accuracy of Bookmate, we will require a diverse and comprehensive dataset. This dataset should include Book Metadata |
| Raw Data Sources Identified | Gathering Data from:  **User-Generated Data:** Collect user preferences, reading history, and feedback to personalize recommendations.  **Book Metadata from Libraries and Publishers:** Access detailed information about books, including genre, subject matter, and publication date. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset | In This Project We Have Used .Csv Data.This Data Is Downloaded From Kaggle.Com | https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset | CSV | 107MB | Public |