**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | Nov, 2024 |
| Team ID | 739838 |
| Project Title | Unlocking the Minds: Analyzing Mental Health with 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 | The project aim is to enable the learner to understand analyzing mental use case of how and why it needed to identify. Using a dataset with features such as text and label, the objective is to build a model that accurately classifies mental health issues facilitating efficient and informed decision- making in the analyzing process. |
| Data Collection Plan | * Search for datasets related to the customer mental history, cause information. * Prioritize datasets with diverse demographic information |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The sample data provided represents a subset of the collected information, encompassing variables such as text and label customer related details for machine learning analysis. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset | The dataset comprises applicant details(Text and Label) and type of customer outcomes. | <https://www.kaggle.com/code/swathiunnikrishnan/mental-health-analysis-using-nlp/input> | CSV | 72 KB | Public |