**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 4 OCTOBER 2024 |
| Team ID | LTVIP2024TMID24789 |
| Project Title | ANALYSIS OF AMAZON CELL PHONE REVIEWS |
| 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 objective of the Amazon Cell Phone Sentiment Analysis project is to develop a real-time system that accurately classifies customer sentiments from Amazon cell phone reviews using an LSTM model, providing insights into sentiment trends over time and offering a user-friendly Flask-based interface for sentiment prediction and analysis, enabling better understanding of consumer opinions |
| Data Collection Plan | * Extract reviews rom amazon product and use tools like scrapy * Utilize amazon product advertising API or amazon reviews |
| Raw Data Sources Identified | Information sourced from the amazon’s product advertising,web scraping,public datasets and third party review aggregators.the impact of amazon cell phone reviews can be significant for both consumers and manufactures |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Smart Internz dataset | This dataset includes to satisfy customer by providing good product and the data is collected from the natural language processing (NLP). | https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews | CSV | 2 GB | Public |