Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 3 october 2024 |
| Team ID | LTVIP2024TMID24828 |
| Project Title | Fake News Analysis In Social Media |
| Maximum Marks | 2 Marks |

# Data Collection Plan & Raw Data Sources Identification Report:

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:

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This project focuses on analyzing and mitigating the impact of fake news on social media platforms. By employing advanced detection algorithms and assessing the credibility of information sources, the initiative aims to identify and combat misinformation effectively. It will also study user engagement patterns to understand how false information spreads, while promoting media literacy through educational resources. Ultimately, the project seeks to foster a more informed online community and support social media platforms in implementing better content moderation practices. |
| Data Collection Plan | * Identify social media platforms (e.g., Twitter, Facebook, Instagram) for analysis. * Select relevant public datasets on misinformation and fact-checking organizations. |
| Raw Data Sources Identified | The raw data sources identified for this project include various social media platforms such as Twitter, where the API can be used to collect tweets with specific keywords or hashtags related to misinformation. Facebook provides access to public posts and comments from relevant groups, while Instagram allows for the analysis of image captions and user interactions. |

**Raw Data Sources Report:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Format** | **Size** | **Access Permissions** |
| Kaggle  telegram | The dataset comprises of news of both real and fake. | CSV | 64MB | Public |