**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 july 2024 |
| Team ID | Team-740034 |
| Project Title | Predicting the energy output of wind turbine based on weather conditions. |
| 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 aims to develop a predictive model for estimating wind turbine energy output based on real-time and historical weather data.  - The model will leverage machine learning techniques to analyze weather variables such as wind speed, temperature, humidity, and air pressure to forecast energy production.  - The goal is to enhance operational efficiency and optimize energy generation in wind farms by providing accurate predictions of turbine performance. |
| Data Collection Plan | - The data collection plan involves gathering both real-time and historical weather data from meteorological stations located near wind farm sites.  - Real-time data will be collected continuously at regular intervals (e.g., every hour) to capture current weather conditions affecting turbine operations.  - Historical data spanning several years will be sourced to analyze long-term trends and seasonal variations in weather patterns that impact energy output.  - Data integrity and quality checks will be implemented to ensure consistency and reliability in the collected datasets. |
| Raw Data Sources Identified | Kaggle.com |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | The dataset which is considered here will have the predicting the energy output of wind turbine. can collect datasets from different open sources like kaggle.com, data.gov, UCI machine learning repository etc. | <https://www.kaggle.com/datasets/predicting> | CSV | 41.37 kB | Public |