Data Collection and Preprocessing Phase

Date	12 July 2024
Team ID	SWTID1720108739
Project Title	**Predicting The Energy Output Of Wind Turbine Based On Weather Condition**
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	This project aims to predict wind turbine energy output based on weather conditions using machine learning. By analyzing historical weather data and energy production records, models will be trained to forecast energy output, aiding in maintenance scheduling, energy distribution strategies, and efficient grid integration. Key steps include data collection, preprocessing, model training, evaluation, and deployment using Python, scikit-learn, and possibly cloud platforms for scalability.
Data Collection Plan	• Search for datasets related to Date/time, Wind Speed, Theoretical Power Curve, LV ActivePower, Wind Direction. • Prioritize datasets By analyzing historical data of weather conditions.

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 provided sample data represents a subset of the collected information, encompassing variables such as Date/time, Wind Speed, Theoretical Power Curve, LV Active Power, Wind Direction for machine learning analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises weather details such as Date/time, Wind Speed, Theoretical Power Curve, LV Active Power, Wind Direction	https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset	CSV	4.0 MB	Public