

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720171463
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	The machine learning project aims to predict the energy output of a wind turbine based on weather conditions. This solution will help energy companies and grid operators better manage and optimize energy production by providing accurate and timely predictions. By analyzing historical weather data and corresponding energy output, we will train machine learning models to predict future energy output given current weather conditions. The project seeks to improve energy management, enhance grid stability, and optimize the overall efficiency of wind energy production.
Data Collection Plan	Pinpointing sources that provide reliable weather data and historical energy output records. Developing scripts or using APIs to extract data from identified sources. Storing the collected data in a structured format in a database for easy access and processing. Ensuring the data is clean, consistent, and ready for analysis.

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, LV ActivePower (kW), Wind Speed (m/s), Theoretical_Power_Curve (KWh), Wind Direction (°)
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle dataset	The dataset comprises Date/Time, LV Active Power (kW), Wind Speed (m/s), Theoretical_Power_Curve (KWh), Wind Direction (°)	https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset	CSV	3830 KB	Public