

Data Collection and Preprocessing Phase

Date	18 July 2024
Team ID	xxxxxx
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	<p>The machine learning project aims to predict wind turbine performance based on environmental and operational features.</p> <p>Using a dataset with features such as wind speed, theoretical power curve, and actual power output, the objective is to build a model that accurately forecasts turbine performance, facilitating better operational decisions and efficiency.</p>
Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to wind turbine performance, including environmental conditions and operational metrics. ● Prioritize datasets with comprehensive feature sets, including variables such as wind speed, power output, and theoretical performance. ● Ensure datasets cover a diverse range of operational scenarios and turbine types for robustness.

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle and UCI, popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as wind speed, theoretical power curve, and actual power output.
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Contains wind turbine performance data including features such as wind speed, theoretical power curve, and actual power output.	https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset	CSV	2 MB	Public