

### Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWUID20240034764
Project Title	Predicting Full Load Electrical Power Output of a Base Load Operated Combined Cycle Power Plant Using Machine Learning
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>aims to enhance the efficiency and accuracy of power output forecasting in combined cycle power plants. By leveraging advanced machine learning techniques, the project focuses on predicting the plant's electrical power output under full load conditions, using historical operational data, environmental variables, and plant performance metrics.</p>
	<ul style="list-style-type: none"> <li>❑ Develop predictive models to estimate the full load power output of the power plant.</li> <li>❑ Utilize machine learning algorithms such as Random Forest and Support Vector Machine to analyze complex datasets.</li> <li>❑ Improve decision-making and operational efficiency through accurate power output forecasts</li> </ul>
Data Collection Plan	<p>To gather and preprocess data necessary for developing and validating machine learning models to predict the full load electrical power output of a base load operated combined cycle power plant.</p>

Raw Data Sources Identified	<ul style="list-style-type: none"> <li>• <b>Power Output Logs:</b> Recorded data on electrical power output at different operating conditions and times.</li> <li>• <b>Load Factors:</b> Data related to load variations and operational cycles of the plant.</li> <li>• <b>Fuel Consumption:</b> Records of fuel usage for different periods and operational states.</li> <li>• <b>Turbine and Generator Performance:</b> Metrics on efficiency, speed, and operational status.</li> </ul>
-----------------------------	---

Source Name	Description	Location/URL	Format	Size	Access Permissions
-------------	-------------	--------------	--------	------	--------------------

Dataset 1	ware house block, customer care calls, cost of product , customer rating, prior purchase, product	<a href="https://www.kaggle.com/datasets/prachi13/customer-">https://www.kaggle.com/datasets/prachi13/customer-</a>	CSV	440.46 KB	Public
-----------	---	---	-----	--------------	--------

### Raw Data Sources Template

	importance, reached on time, discount offer , Gender , mode of shipment	<a href="#">analytics?select=</a> Train.csv			
--	--	--	--	--	--