Solar Plant Power Generation and Weather Data

Data Source: https://www.kaggle.com/datasets/anikannal/solar-power-generation-data/data

Problem Statement

To analyse and predict the power generation of a solar plant and identify ways to improve efficiency through the following steps:

1. Data Collection and Analysis: This step involves downloading,cleaning, transforming the data from the solar plant, which could include solar irradiance, panel temperatures, weather conditions, and the actual power output. Analysis of this data helps in understanding the current performance of the plant.
2. Predictive Modelling: Using historical data, predictive models can be developed to forecast the power generation of the solar plant. This involves machine learning techniques and statistical models that take into account various factors like sunlight intensity, time of day, season, and weather conditions.
3. Efficiency Identification: This involves identifying factors that are currently limiting the efficiency of the solar plant. It could be due to suboptimal positioning of solar panels, inefficient solar panels, shading issues, or operational and maintenance challenges.
4. Performance Optimization: Based on the analysis and predictive models, strategies to optimise the performance of the solar plant can be developed. This could involve adjusting the angle of solar panels, upgrading equipment, implementing better cooling systems for panels, or optimising the maintenance schedule.

Dataset Description

The data has been gathered at two solar power plants in India over a 34 day period. It has two pairs of files - each pair has one power generation dataset and one sensor readings dataset. The power generation datasets are gathered at the inverter level - each inverter has multiple lines of solar panels attached to it. The sensor data is gathered at a plant level - single array of sensors optimally placed at the plant

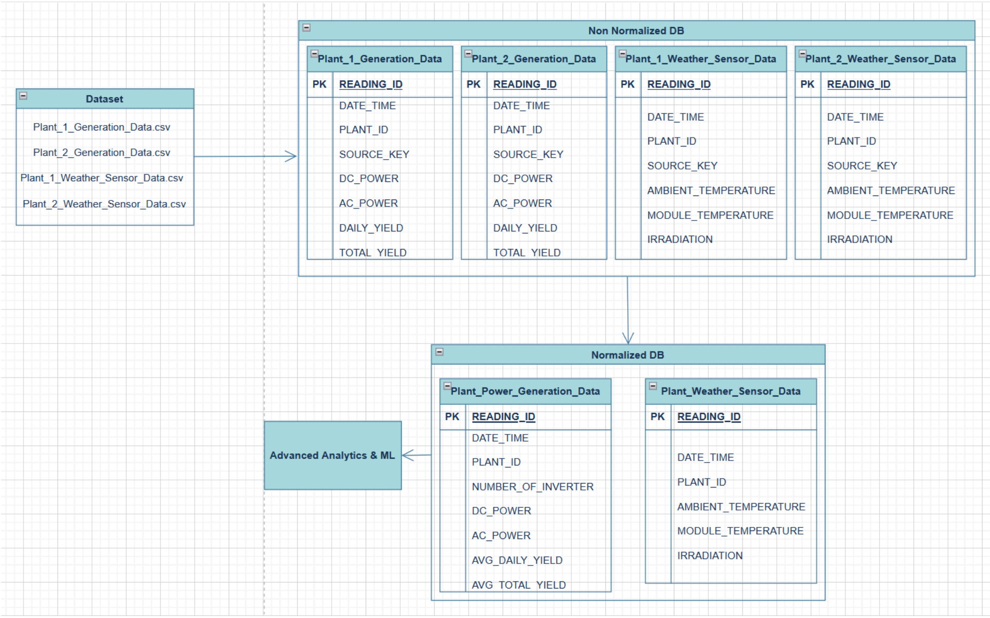
Analysis Methodology

We initiated our analysis by collecting data from a database comprising four CSV files. To transform this raw data into a more structured format, we converted it into a non-normalized database using Python. Each line from the CSV files was read and inserted into an SQL database.

Next, we focused on data normalisation. This involved aggregating information and merging data from two plant databases and two weather databases. By combining these sources, we ensured consistency and prepared the data for further exploration.

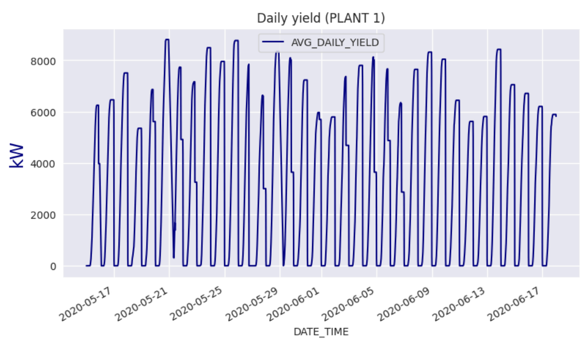
Our data analysis phase included creating visualisations to gain insights. We examined how power generation varied under different conditions. Additionally, we identified faults in Plant 1 based on our analysis.

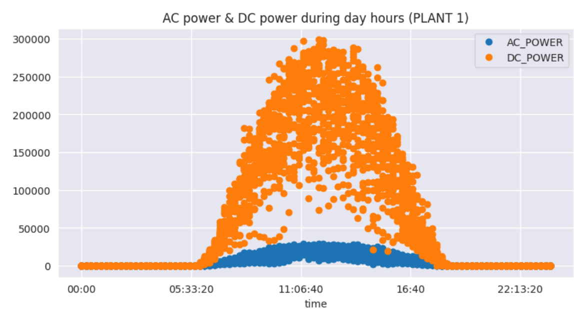
Given the issues found in Plant 1, we shifted our attention to Plant 2 for subsequent analysis. Here, we delved into machine learning modelling. Using data exclusively from Plant 2, we trained and fitted a few ML models to predict power output. Flowchart for the same:

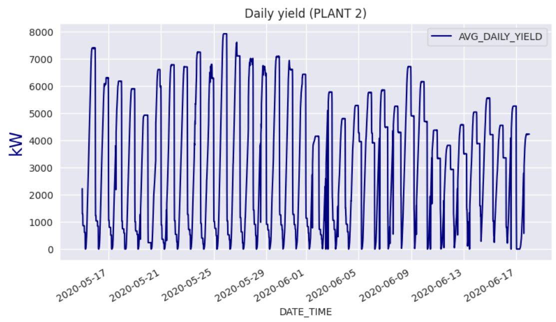


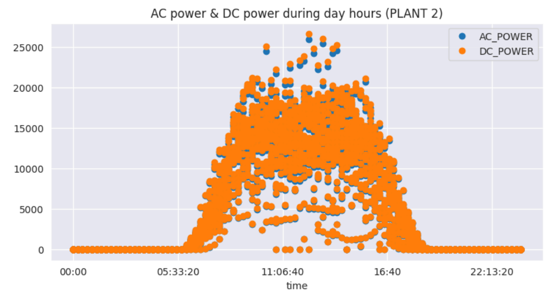
Analysis Result

On Analysis of the Daily Yield and AC/DC Power Generation of both Plant 1 and Plant 2 showed that the Inverters in Power Plant 1 were only able to convert 10% of DC Power to AC power while the Inverters of Power Plant 2 had an efficiency of over 90% to convert DC Power to AC Power. This could be due to the faulty inverters in Plant 1.





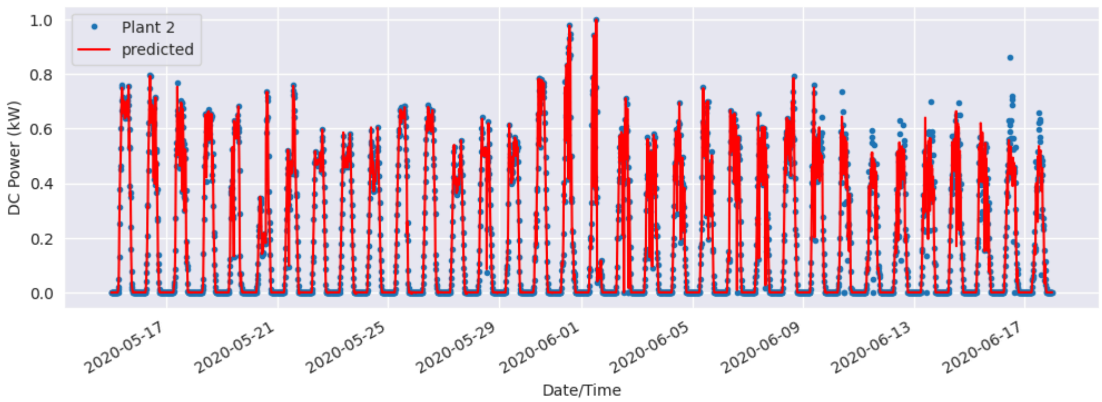




Machine Learning Models Result:

The goal of this model is to predict what the DC Power output would be on a given day. This requires that a user have a forecast for the next day (or few days) and they can then put that information into this model to get a prediction of the power output.

Linear Regression Model Performance : R-Square : 72%, MAPE: 0.34, RMSE: 1.12
Gradient Boosted Tree Model Performance : R-Square : 88%, MAPE: 0.23, RMSE: 0.89
Our model is able to capture a large portion of the variability in the data and make accurate predictions.



Conclusion

1. Power Plant 1 has serious efficiency drop in power generation due to faulty inverters
2. The Machine Learning Model can help optimise the power generation given weather and time data for plant 2, also help for upscaling/ reducing downtime of the solar plants.