Fault Detection & Power Prediction in Solar Power Plants

B.Tech Project Report

Submitted by

Mihir Lakhotia (B19ME047)

Ravi Verma (B19ME062)

Under the Supervision

Of

Prof. B Ravindra

Department of Mechanical Engineering

Indian Institute of Technology Jodhpur

November 2023

CONTENTS

Acknowledgments

Abstract

1. Introduction

    1.1 Objective

    1.2 Can We Predict The Power Generation By the Plant?

    1.3 Can we identify faulty or suboptimally performing equipment?

2. Work done

    2.1 Methodology

        2.1.1 About the dataset

    2.2 Implementation

        2.2.1 Data Visualization

        2.2.2 Data Preprocessing

        2.2.3 DC power generation from Solar Panels to particular Inverters

    2.3 Theoretical/Experimental work

        2.3.1 Fault Prediction of 2nd power plant

        2.3.2 Abnormalities in AC_POWER & DC_POWER Generation

        2.3.3 DC Power Generation for Plant 1

    2.4 Predicting Power Generation

    2.5 Physics Model

3. Results and Discussion

4. Conclusion

5. References

Acknowledgments

We are very happy to offer the report of the B.Tech project completed during the final year of the B.Tech. **Dr. B. Ravindra** deserves a special note of appreciation for his passion, tolerance, perceptive observations, useful information, practical advice, and never-ending ideas, all of which have been of great use to us during the course of our project-related research. We were able to effectively accomplish this project thanks to his vast knowledge, extensive experience, and professional expertise in mechanical engineering and machine learning. We would not have been able to complete this project without his assistance and direction. He provided us with the inspiration we needed to finish our report on time. We are honored to work alongside such passionate and knowledgeable individuals. We could not have imagined having a better supervisor in my study.

We also wish to express our sincere thanks to the Indian Institute of Technology, Jodhpur for accepting us into the graduate program.

Thanks!

Ravi Verma

Mihir Lakhotia

Bachelor of Technology

Dept. of Mechanical Engineering

Indian Institute of Technology, Jodhpur

Abstract

In a nation like India that receives a lot of solar insolation, photovoltaic solar power has recently emerged as the finest form of green energy. A photon's energy is transmitted to the silicon photovoltaic cell's electrons when it strikes the cell's surface. As a result of their "excitation," these electrons start to flow in the circuit, producing an electric current. The energy produced by a solar panel is Direct Current (DC power). The inverter is then responsible for converting it into alternating current so that it can be used in our distribution networks and transported. As a result of ongoing developments in smart grid technology, battery storage, and efficient PV modules, etc. PV-based Power Generation Solar Plant has continued to gain speed and has a strong future. Sunlight strikes PV modules, creating DC power that is sent to inverters (via some junction boxes and string monitoring boxes), which then convert DC power to AC power. AC power is then increased through transformers to meet grid voltage before being fed to the grid through some switchgear. The output of solar power plants, in contrast to typical coal or gas-based power plants, is only accessible during the day and is extremely changeable depending on the amount of sunlight. Data on the solar power generation and weather for two plants are available.

Let's examine the provided data, draw some conclusions, attempt to overcome our obstacles, and, to the extent possible, anticipate or estimate the plant output that can be used for better grid management or stability. Over the course of 34 days, this data was collected at two solar power plants in India. The datasets for power generation are obtained at the inverter level because each inverter has several solar panel lines tied to it. At the plant level, a single array of strategically positioned sensors collects the sensor data.

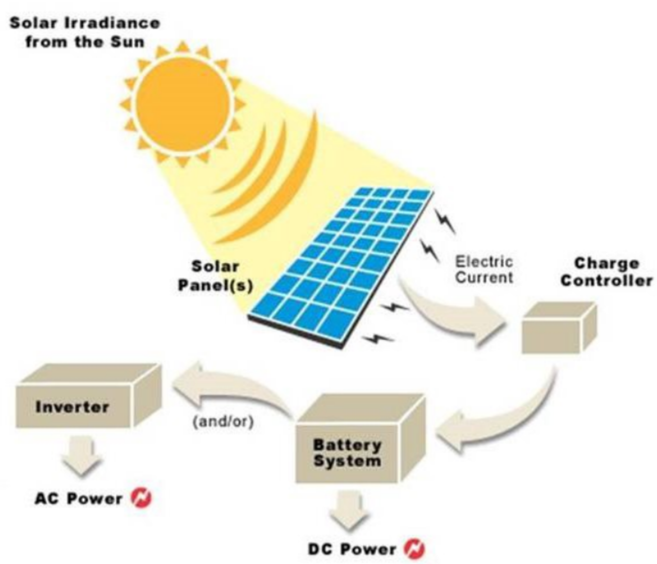
1. Introduction

Basics from the domain:-

Basic Solar Energy Production and Transmission Steps 1) When sunlight strikes solar panels, an electric field is produced. 2) The electricity produced flows into a conductive wire at the panel's edge. 3) The electrical current is carried via the conductive wire to the inverter, which converts it from DC to AC. Every solar power plant has at least two fundamental parts:

- Modules- that convert sunlight into electricity
- One or more inverters- devices that convert direct current into alternating current.

In a nutshell, solar electricity is produced by putting solar panels on the earth, typically in places that get as much direct sunlight as they can during the day. When photons, or light particles, force electrons from atoms, an electrical flow results. We can use this electrical current as DC power or convert it to AC power by passing it through an inverter.



Typically, the factors that determine the performance of a solar power plant are:

1) Temperature
2) Dirtiness
3) Inverters Efficiency
4) Inverters or panels seniority

The forecasted irradiance and ambient temperature, which are the two key variables for estimating the power generation of solar panels in the current physics models, have been used to estimate solar photovoltaic power output in a variety of ways.

However, it is well recognized that because weather and climate are made up of incredibly complex and interconnected systems, future weather cannot be predicted using simply the data of the current irradiance and ambient temperature of a single spot on Earth.

Although several methods for predicting future power outputs using neural networks using historical data on power output have been developed, these models rely on huge datasets that span many years. Due to its reduced size, this dataset is therefore not suitable for such models. Nevertheless, the idea that follows can be evaluated using this dataset:

"The power output of this photovoltaic power plant can be predicted, Given accurate forecasts of ambient temperature and irradiation"

As a result, the goals of this analysis are to verify this claim and to provide information for the solar photovoltaic plant that may help identify potential areas for improvement in maintenance or other areas as well as failure modes.

1.1 Objective:-

There are a few areas of concern at the solar power plant -

- Can we identify faulty or suboptimally performing equipment?
- Can we predict the power generation for the next day? - this allows for better grid management

1.2 Can We Predict The Power Generation By the Plant?

It would be more significant to predict the overall power generated by the plant for a day rather than the individual power output recorded by each inverter. At this point, it transforms into a problem of time series prediction, and we can use several time series models to provide a solution.

The weather has a significant impact on solar power generation, and forecasting the weather is a very challenging task (as evident by how accurate our weather forecasts are). As a result, it will not be possible to effectively estimate future weather and, consequently, the amount of power generated, using the data provided.

1.3 ---Can we identify faulty or suboptimally performing equipment?

The amount of power produced by each solar panel decreases when they are dirty. As a result, when daily power generation falls below average, it is time for panel cleaning/maintenance. What if all of the solar panels are equally dirty? is a worry. But because solar panels are placed in different ways, the amount of dirt that collects on each solar panel will vary. Thus, even if all of the solar panels are initially equally dirty, it won't take long for the dirtiest solar panels to stand out.

Another viewpoint is that we must decide when to clean every solar panel. In this case, we would analyze the whole power generation to see if there was a need before setting a threshold, such as 95% control-based efficiency. Cleaning is required if the amount of power generated consistently falls below the threshold for a number of days.

2. Work Done

2.1) Methodology

2.1.1) About the dataset:

- This data was gathered over the course of 34 days at two solar power plants in India. It includes two file pairs: one for power generation and one for weather readings. Datasets on power generation are collected at the inverter level, with each inverter connected to multiple lines of solar panels. Sensor data is collected at the plant level by a single array of sensors strategically placed throughout the facility.
    - Plant 1 is near Gandikota, Andhra Pradesh
    - Plant 2 is near Nasik, Maharashtra
- **Generation Dataset:**
- The power generation datasets are gathered at the inverter level- each inverter has multiple lines of solar panels attached to it.
- Total no. of records = 69,000

DATE_TIME	15 minute timestamp
PLANT_ID	Common for the entire file
SOURCE_KEY	Unique Inverter ID (Total 22 Inverters)
DC_POWER	Amount of DC Power generated by that inverter for the timestamp
AC_POWER	Amount of AC power after conversion from DC by inverter for the timestamp
DAILY_YIELD	Cumulative sum of power generated on that day, till that point in time
TOTAL_YIELD	Total yield for the inverter till that point in time

Weather Sensor Dataset:

DATE_TIME	15 minute timestamp
PLANT_ID	Common for the entire file
SOURCE_KEY	Unique Inverter ID (Total 22 Inverters)
AMBIENT_TEMPERATURE	Ambient temperature at the plant
MODULE_TEMPERATURE	Temperature reading for module (solar panel) attached to the sensor panel
IRRADIATION	Amount of irradiation for the 15 minute interval

2.2) Implementation

2.2.1) Data Visualization:

Basic points to notice from the description:

```
p1gd.describe()
```

	PLANT_ID	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AC_POWER1
count	68778.0	68778.000000	68778.000000	68778.000000	6.877800e+04	68778.000000
mean	4135001.0	3147.426211	3078.027523	3295.968737	6.978712e+06	3078.027523
std	0.0	4036.457169	3943.964387	3145.178309	4.162720e+05	3943.964387
min	4135001.0	0.000000	0.000000	0.000000	6.183645e+06	0.000000
25%	4135001.0	0.000000	0.000000	0.000000	6.512003e+06	0.000000
50%	4135001.0	429.000000	414.937500	2658.714286	7.146685e+06	414.937500
75%	4135001.0	6366.964286	6236.187500	6274.000000	7.268706e+06	6236.187500
max	4135001.0	14471.125000	14109.500000	9163.000000	7.846821e+06	14109.500000

```
p1wd.describe()
```

	PLANT_ID	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
count	3182.0	3182.000000	3182.000000	3182.000000
mean	4135001.0	25.531606	31.091015	0.228313
std	0.0	3.354856	12.261222	0.300836
min	4135001.0	20.398505	18.140415	0.000000
25%	4135001.0	22.705182	21.090553	0.000000
50%	4135001.0	24.613814	24.618060	0.024653
75%	4135001.0	27.920532	41.307840	0.449588
max	4135001.0	35.252486	65.545714	1.221652

- DC power is greater than AC power. PV modules provide DC power, which is then transformed into AC power by inverters for transmission purposes. This process results in some loss because the conversion efficiency isn't ideal. The power loss increases with an increase in the difference between these two.
- High difference between the maximum values of ambient temperature, module temperature, irradiation, and the 75th percentile. This suggests that high values are less frequent. Cold weather, fewer solar hours, the presence of outliers, etc. could all be contributing factors.

2.2.2) Data Preprocessing:

Key Points:

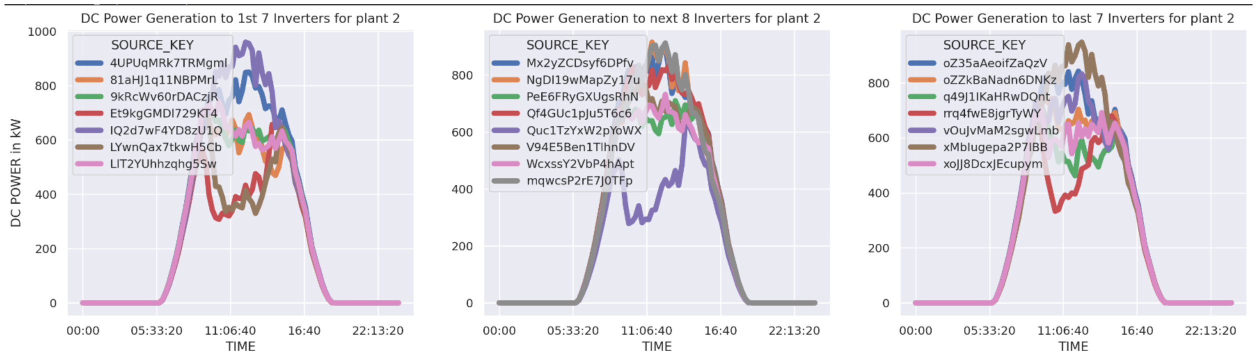
- Checked how many Inverters are there in the plants and how many values have been recorded by each of them

```
[95] p1gd.groupby('SOURCE_KEY')['DATE_TIME'].count()

SOURCE_KEY
1BY6WEcLGh8j5v7    3154
1IF53ai7Xc0U56Y    3119
3PZuoBAID5Wc2HD    3118
7JYdWkrLSPkdwr4    3133
McdE0feGgRqW7Ca    3124
VHMLBKoKgIrUVDU    3133
WRmjgnKYAwPKWDb    3118
YxYtjZvoooNbGkE    3104
ZnxXDlPa8U1GXgE    3130
ZoEaEvLYb1n2sOq    3123
adLQvlD726eNBSB    3119
bvBOhCH3iADSZry    3155
iCRJl6heRkivqQ3    3125
ih0vzX44oOqAx2f    3130
pkci93gMrogZuBj    3125
rGa61gmuvPhdLxV    3124
sjndEbLyjtCKgGv    3124
uHbuxQJl8lW7ozc    3125
wCURE6d3bPkepu2    3126
z9Y9gH1T5YWrNuG    3126
zBIq5rxdHJRwDNY    3119
zVJPv84UY57bAof    3124
Name: DATE_TIME, dtype: int64
```

- Used Label Encoder to convert SOURCE_KEY to numerical value names
- Separated Date and Time columns
- Checked for the missing values in all 4 datasets

2.2.3) DC power generation from Solar Panels to particular Inverters

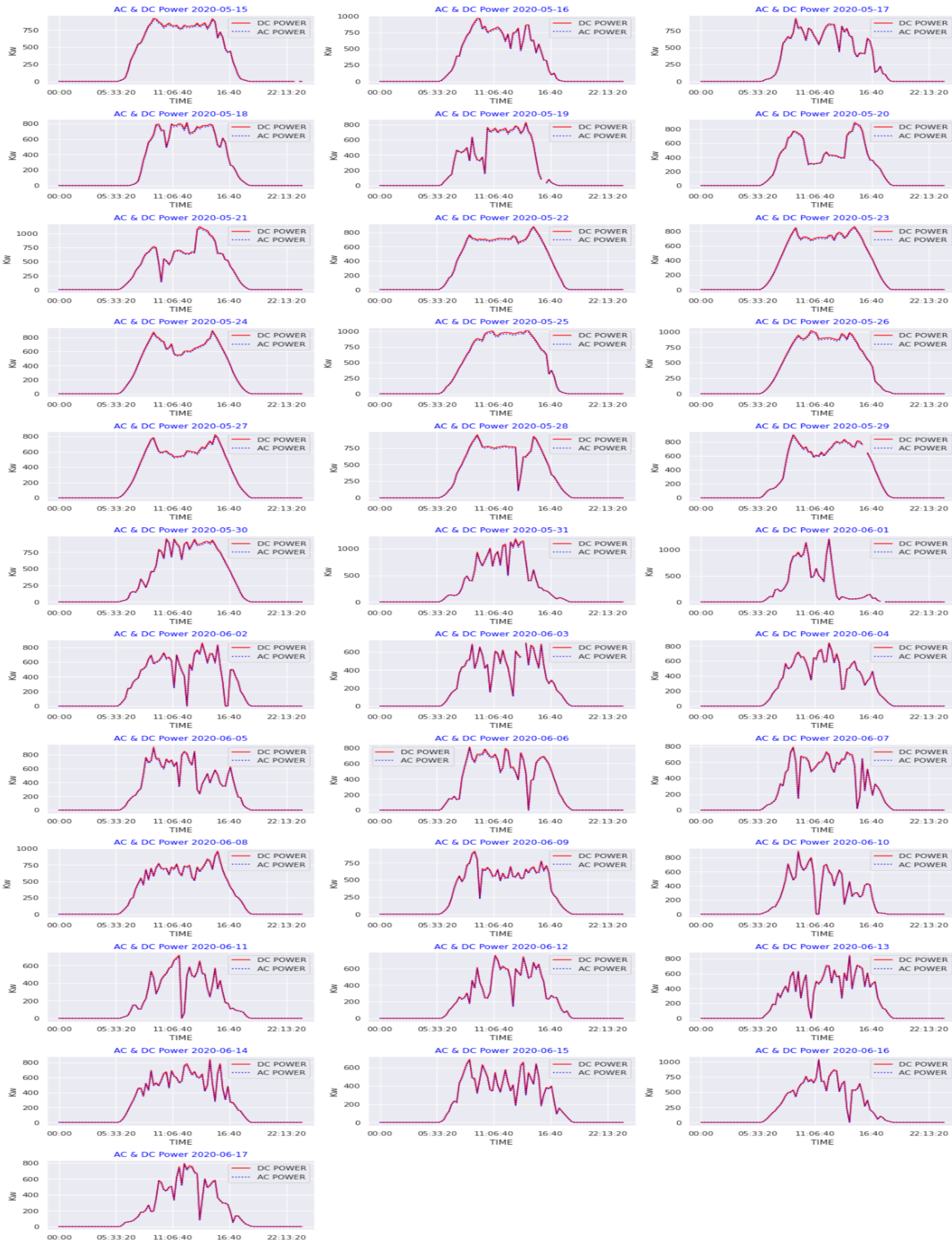


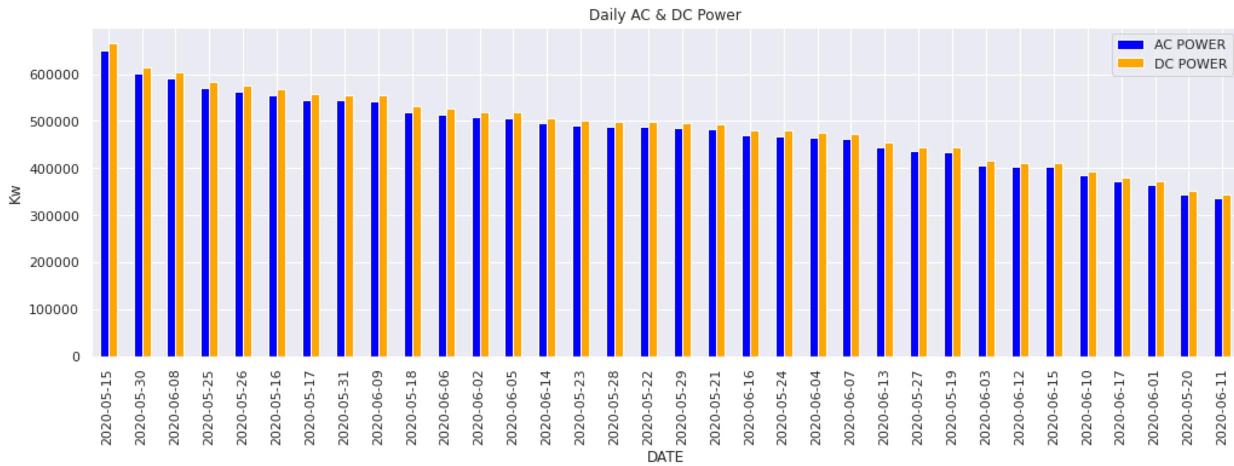
The DC Power Generation figure makes it evident that Inverters "Et9kgGMDl729KT4", "LYwnQax7tkwH5Cb", "Quc1TzYxW2pYoWX," and "rrq4fwE8jgrTyWY" are receiving significantly low DC Power. Therefore, it is advised that Solar Modules linked to these Inverters be cleaned and checked for any shadows cast by neighbouring objects. Additionally, any flaws in these Solar Modules and/or Strings must be examined.

2.3) Theoretical/Experimental work

2.3.1) Fault Prediction of 2nd power plant

Plotting DC & AC Power generation and weather conditions for plant 2 on a daily basis





2.3.2) Abnormalities in AC_POWER & DC_POWER Generation

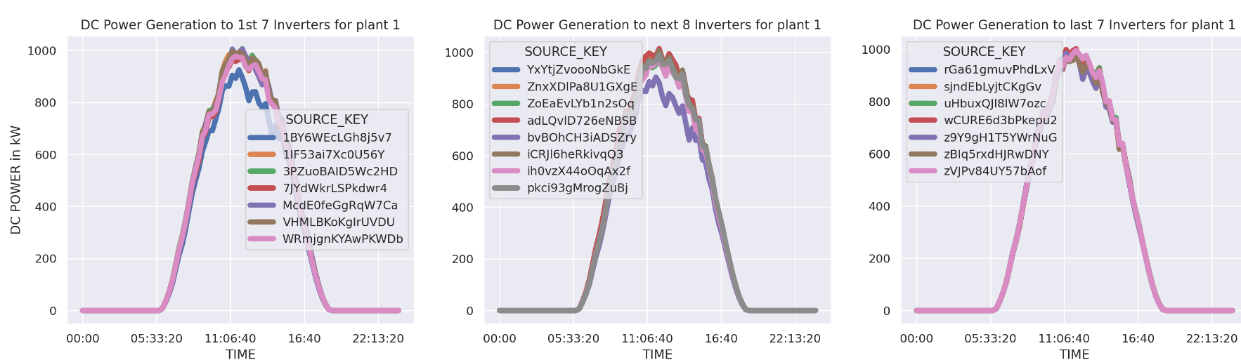
From the per day POWER generation graph we can find that most of the days there is some fluctuation in the power generation. A smooth curve in the midday indicates a clear sky and constant power generation The more spikes, the lesser the Power produced and the spikes indicate that there are some clouds hovering over the panel.

We can also find the average power generation per day.

- Highest average AC_POWER & DC_POWER Generation is on: 2020-05-15
- Lowest average AC_POWER & DC_POWER Generation is on: 2020-06-11

This large variation in power generation is the result of a system failure or a change in weather. Using this bar plot, we can determine the day with the most power generated and the day with the least power generated.

2.3.3) DC Power Generation for Plant 1



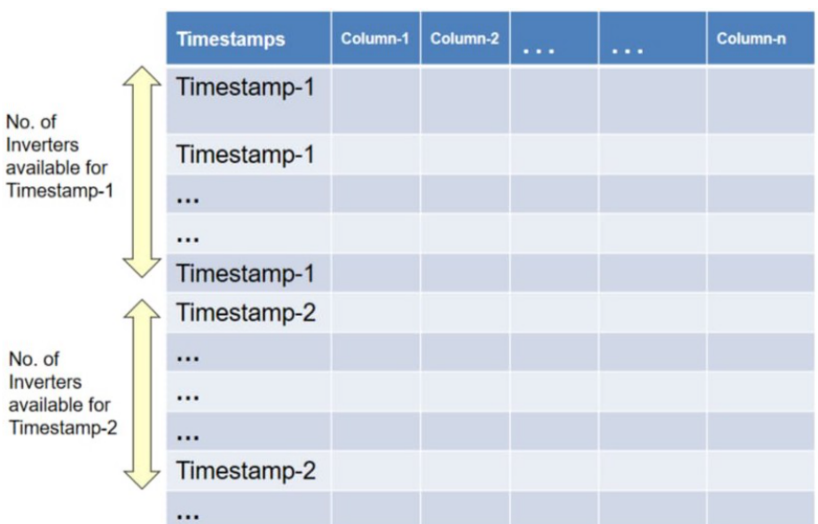
Plant 1 seems fine, as there are no major irregularities in the graph

2.4) Predicting Power Generation

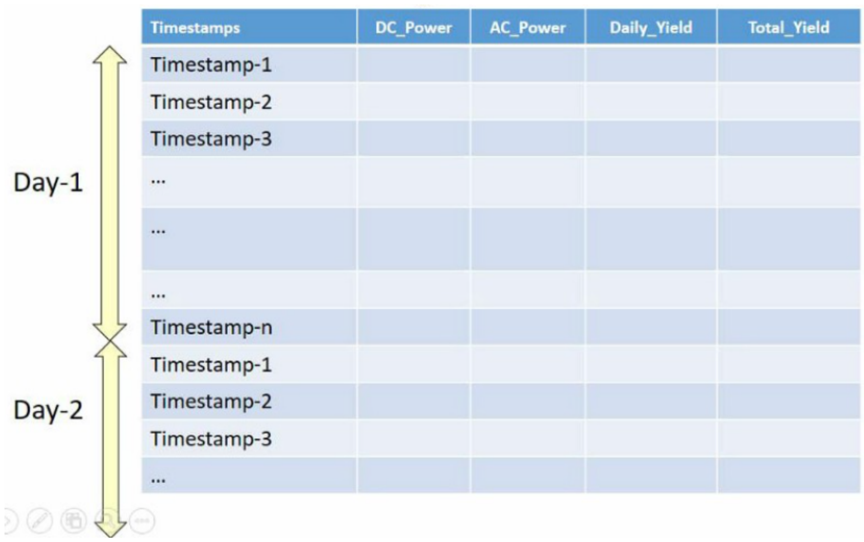
We will forecast electricity and power generation in two ways. The first method is to train the model directly using weather data and generated power. Second, a physics-based theoretical photovoltaic model will be used. Irradiation, ambient temperature, and module temperature will all have an impact on the plant's AC power generation.

Data Preparation:-

We currently have rows for each inverter, which means that each timestamp is repeated as many times as data for that timestamp is present. However, we want to make predictions in 15-minute increments at the plant level. As a result, the desired format is day-by-day (timestamp-by-timestamp) rows from 00:00 to 24:00, with columns representing the sum of all inverter values (for DC Power, AC Power, etc.) for that timestamp, as shown below.



To do so, we will first organize our data inverter by inverter and save each group in a list. Each group will then be combined with the next using the outer join on the DATE TIME column. We interpolate all of the AC POWER and DC POWER columns after grouping the data inverter by inverter to predict power on a power plant level. Our data is now ready for us to create models with. Our final data will look like this:-



Models:-

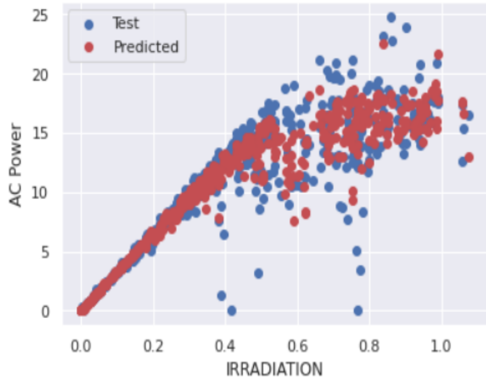
We split our dataset into 60% for train data and 40% for test data respectively. As the different features had a different range of values irradiation varied from 0 to 1.22 whereas ambient temperature varied from 20-35. So to tackle this we used a standard scaler to scale the values. For training the model, we used ambient temperature, module temperature, and irradiation as the training variables and AC Power as our target variable. We Applied various regression models like multiple regression, KNeighbors regression, and Decision Tree Regressor.

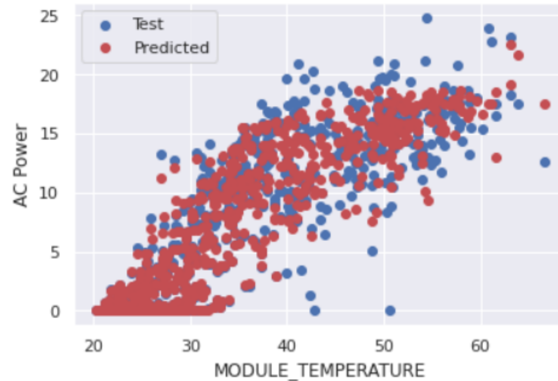
The results obtained from running the models were as follows:-

	Name	RMSE
0	Linear Regression	71.040050
1	Support Vector Regression	52.566475
2	KNeighbors Regression	63.662697
3	Decision Tree Regression	77.181429
4	Random Forest Regression	54.351226

Plots of one of the models with irradiation and module temperature are shown below:-

```
Text(0, 0.5, 'AC Power')
```





2.5) Physics Model.

By multiplying the open circuit voltage Vth (Thèvenin voltage) by the short circuit current Ino (Mayer-Norton current), the photovoltaic power output of Pac (including the inverter) can be calculated:

The module temperature is proportional to the irradiation and to the ambient temperature, and it can be represented empirically by the formula below.

$$P_{ac} = V_{th} \cdot I_{no}$$

And the $\mathbf{V_{th}}$, $\mathbf{I_{no}}$ are defined by[3]:

$$V_{th} = V_0[1 + \beta(T_m - T_0)]$$
$$I_{no} = I_0[1 + \alpha(T_m - T_0)]\left(\frac{G_{ir}}{G_0}\right)$$

With equation **(5)** we now know which new features we have to create in order to build a regressor that predicts **AC Power**. They are highlighted in orange. Note that this equation has an intercept = 0 in the $\mathbf{P_{ac}}$ axis

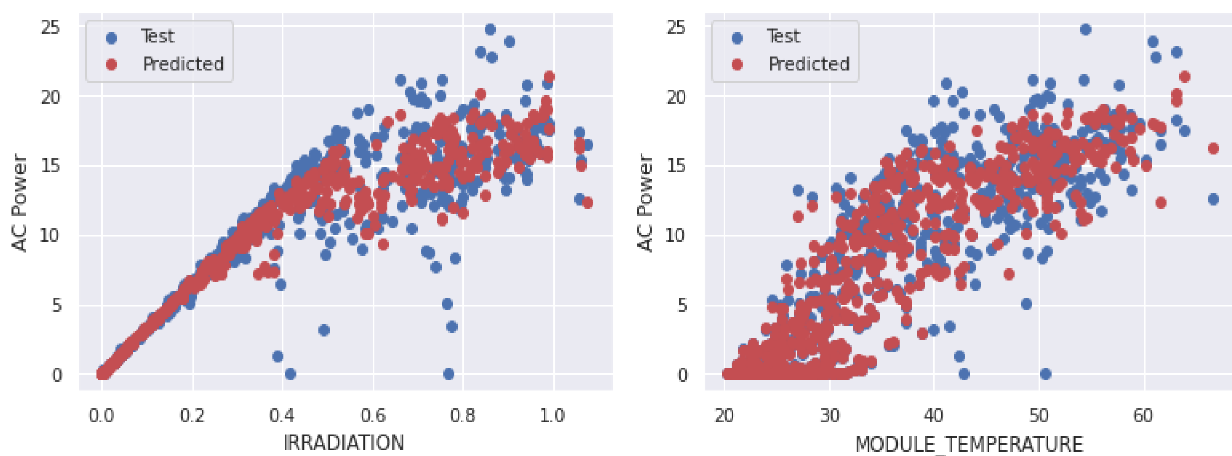
$$T_m = 30 - 0.0175(G_{ir} - 300) + 1.14(T_a - 25) \tag{4}$$

replacing equations **(2)**, **(3)** and **(4)** into equation **(1)** we then get a third degree polynomial that can be expressed as follows:

$$P_{ac} = K_1 G_{ir}^3 + K_2 G_{ir}^2 + K_3 G_{ir}^2 T_a + K_4 G_{ir} T_a^2 + K_5 G_{ir} T_a + K_6 G_{ir} \tag{5}$$

where K_1, K_2, ... , K_6 are constants.

We made the columns corresponding to the above formula in our dataframe and then trained the models again with these columns as well as our previous weather data as features. The results and plots are shown below:-



3. Results and Discussion

As our first objective was to detect if any faults or irregularities were present in the power generated by a plant on any given day. We successfully did that.

We can observe from the bar plot that

- The highest average DC_POWER Generation is on 2020-05-15
- The lowest average DC_POWER Generation is on 2020-06-11

Our second objective was to calculate the power generated by the plant on the basis of weather conditions. We obtained the following results using different models:

1) Using Weather Conditions

	Name	RMSE
0	Linear Regression	71.040050
1	Support Vector Regression	52.566475
2	KNeighbors Regression	63.662697
3	Decision Tree Regression	77.181429
4	Random Forest Regression	54.351226

2) Using the Theoretical PV model along with weather conditions

	Name	RMSE
0	Linear Regression with PV Model	50.625901
1	Support Vector Regression with PV Model	135.814162
2	KNeighbors Regression with PV Model	58.790946
3	Decision Tree Regression with PV Model	77.783938
4	Random Forest Regression with PV Model	53.028369

4. Conclusion

4.1 We are able to detect the presence of a fault in the PV Solar plant

4.2 We also succeeded in predicting power output with great efficiency

References:

1. Solar Power Generation Dataset:- Dataset

2. Huang, Chiou-Jye, Mao-Ting Huang, and Chung-Cheng Chen. "A novel power output model for photovoltaic systems." *International Journal of Smart Grid and Clean Energy* 2.2 (2013): 139-147.

3. **"Photovoltaic Engineering Handbook." Authors**: France Lasnier, Tony Gan Ang. **Date:** 1990 **Refs:** USA, ISBN: 9780203743393

4.R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, G.E. Georghiou, Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems, Solar Energy, Volume 190, 2019, Pages 405-413, ISSN, 0038-092X

,https://doi.org/10.1016/j.solener.2019.08.032.
(https://www.sciencedirect.com/science/article/pii/S0038092X19308126)

5. Aadyasha Patel, O. V. Gnana Swathika, Umashankar Subramaniam, T. Sudhakar Babu, Alok Tripathi, Samriddha Nag, Alagar Karthick, M. Muhibbullah, "A Practical Approach for Predicting Power in a Small-Scale Off-Grid Photovoltaic System using Machine Learning Algorithms", *International Journal of Photoenergy*, vol. 2022, Article ID 9194537, 21 pages, 2022. https://doi.org/10.1155/2022/9194537

6. Chuluunsaikhan, Tserenpurev, et al. "Predicting the Power Output of Solar Panels based on Weather and Air Pollution Features using Machine Learning." *Journal of Korea Multimedia Society* 24.2 (2021): 222-232.