Fault & Power prediction in Solar Power Plant

Name: -
Ravi Verma (B19ME062)
Mihir Lakhotia (B19ME047)

Basics related to Solar Energy

Basic Steps in Solar Energy Generation and Transmission:-

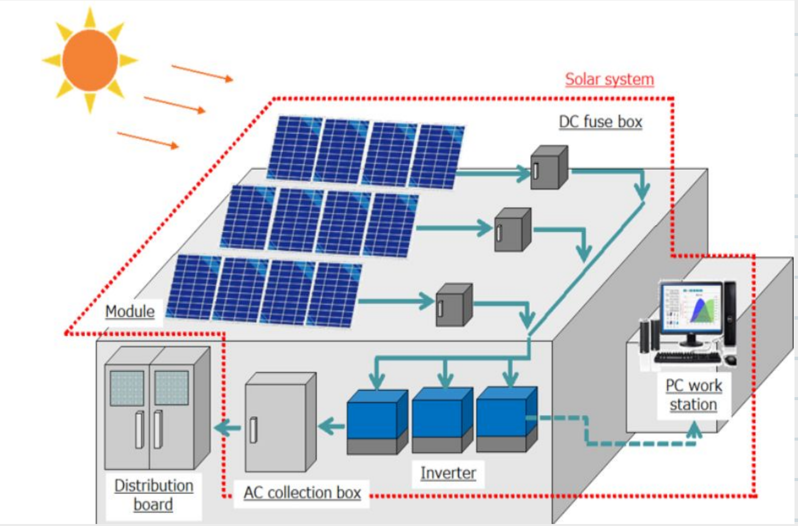
- Sunlight hits the solar panels, and creates an electric field
- The electricity generated flows to the edge of the panel, and into a conductive wire
- Transfer of the electricity to the inverter to transform DC current to AC current

Determining factors of the performance of the Solar Power plant:-

- Temperature
- Dirtiness
- Inverter efficiency
- Inverters or panels seniority

PV Solar Power Plant

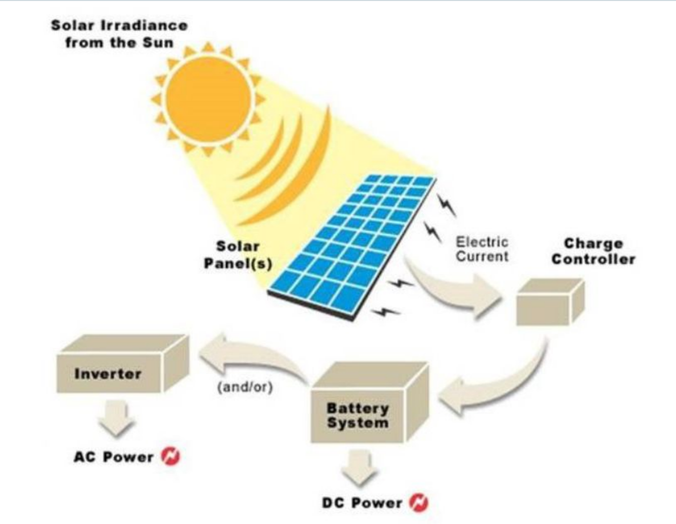
- The solar power plant is also known as the Photovoltaic (PV) power plant
- Large-scale PV plant designed to produce bulk electrical power from solar radiation



The above picture shows a typical structure of a solar power plant. Sunlight falls on PV modules, generates DC Power which is fed to the Inverters (through some Junction Box and String Monitoring Box), Inverters convert DC Power to AC Power, AC Power is stepped up through Transformers to match Grid Voltage and finally fed to the Grid through some Switchgear.

Objectives

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



About the Dataset

- This 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.
- Plant 1 is near Gandikota, Andhra Pradesh
- Plant 2 is near Nasik, Maharashtra

About the Dataset

Generation Dataset:-

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

Total no. of records ≈ 69,000

Data Visualization

- DC Power > AC Power
- Greater the difference between these two, more is the power loss

```
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

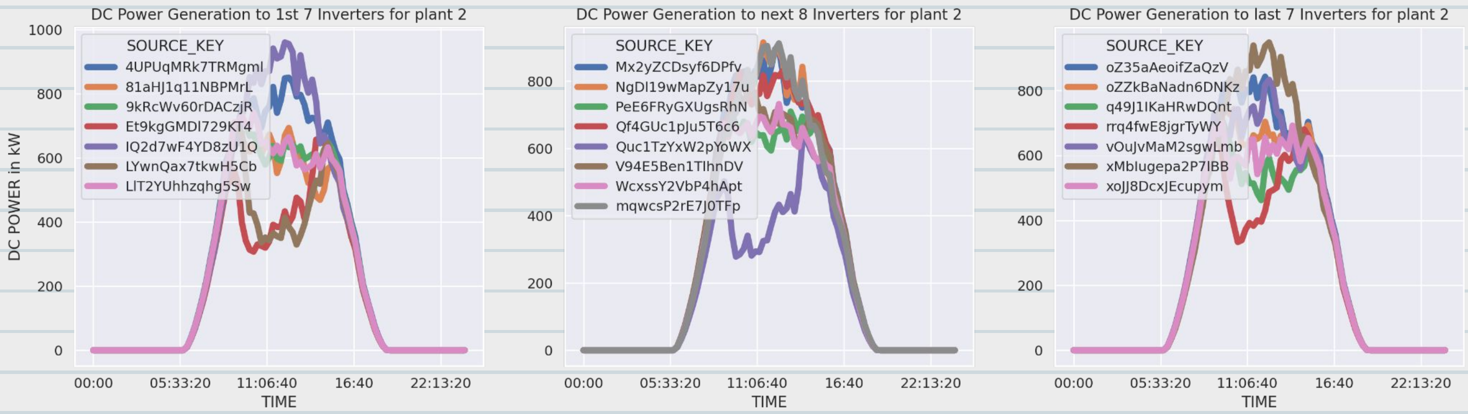
- High difference between 75th percentile and Max values of Ambient Temperature, Module Temperature and Irradiation
- This could be because of cold weather, less solar hours, presence of outliers, etc

```
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

Data Visualization

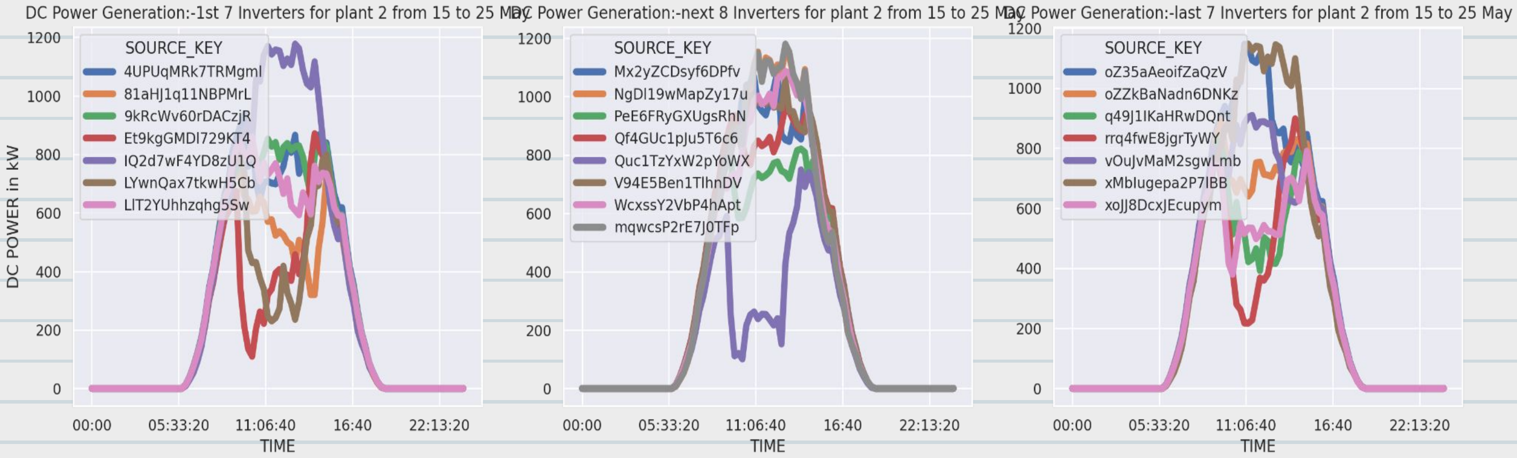
- **DC power generation from Solar Panels to particular Inverters for all 34 days**



- Low DC Power is coming to Inverters "Et9kgGMDl729KT4", "LYwnQax7tkwH5Cb", "Quc1TzYxW2pYoWX" and "rrq4fwE8jgrTyWY"
- Solar Modules connected to these Inverters are recommended to be cleaned, observed for any shadow coming from nearby objects
- There could also be any fault in the panel or generator
- In the plot 2, we can see only 1 value is very low, So ,it must be because of the faulty or dirty panel

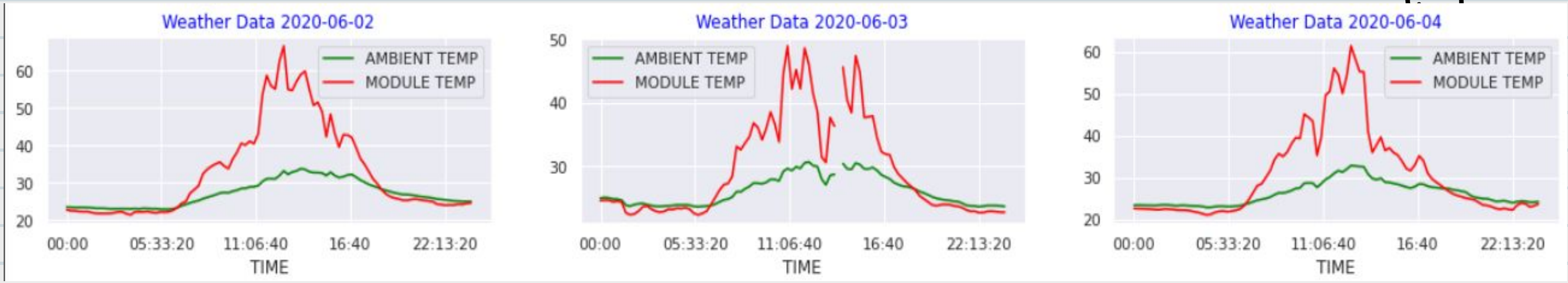
Data Visualization

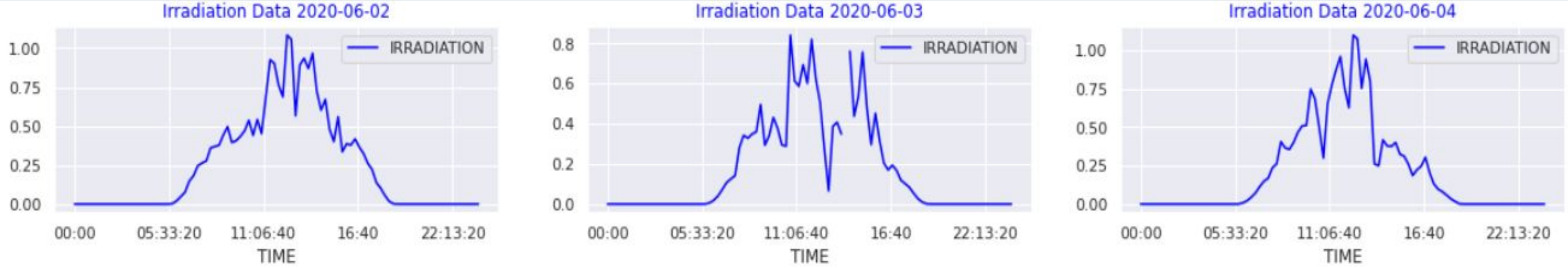
- **DC power generation from Solar Panels to particular Inverters**



- Here, most of the values are fluctuating constantly, means there should be some seasonal effect like raining, cloudy weather etc.
- There could also be any fault in the panel or generator

- **Multiple Plotting of MODULE_TEMPERATURE, AMBIENT_TEMPERATURE & IRRADIATION generation on per day basis**

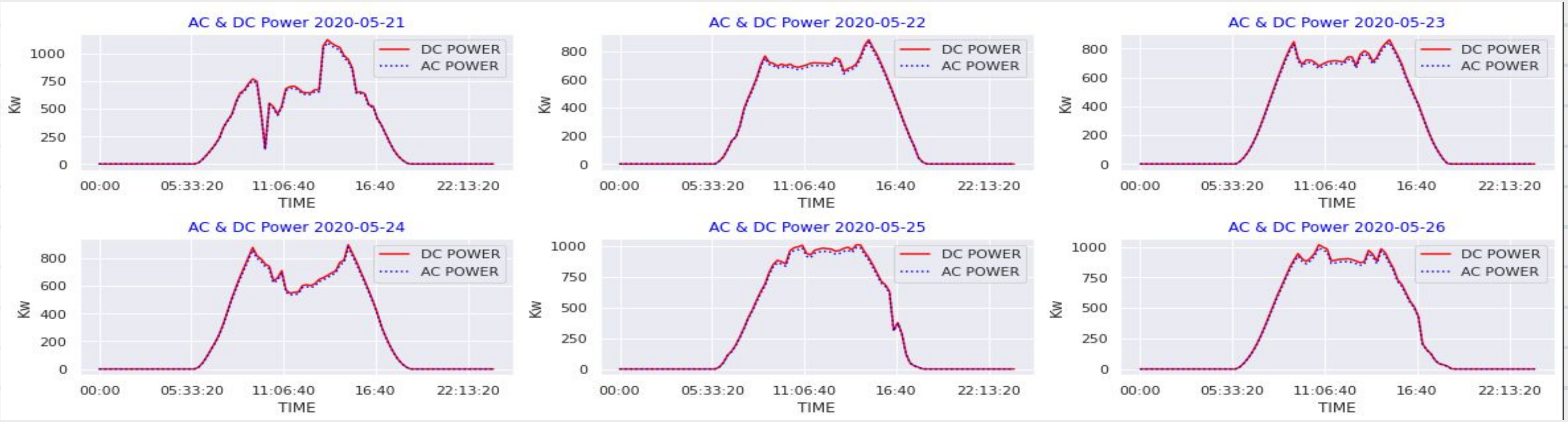




- It is observed that Ambient Temperature shows is almost similar behaviour everyday
- Irradiation is fluctuating in the noon time
- There is also a data loss on 03/06/2020. It may be because of the sensor got any fault and stopped working, So there is no power generation at that time interval

Fault Prediction of 2nd power plant

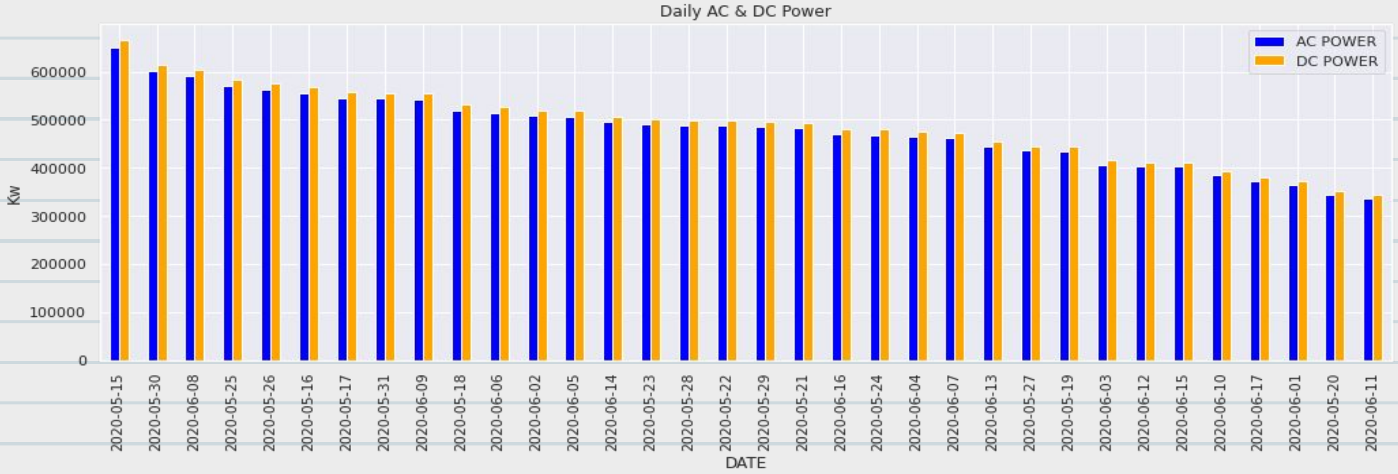
- Plotting AC & DC Power generation and weather conditions for plant 2 on daily basis



- We can observe that there is not much difference in DC Power generated and AC power
- DC power is slightly greater than AC Power everyday
- Smooth curve in the mid day indicates clear sky and constant power generation
- The more the spikes, the lesser is the Power produced.
- Reason for very high Fluctuation & Reduction in DC_POWER generation is due to fault in the system or may be fluctuation in weather or due to clouds etc. which need to be analyse further.

Continue...

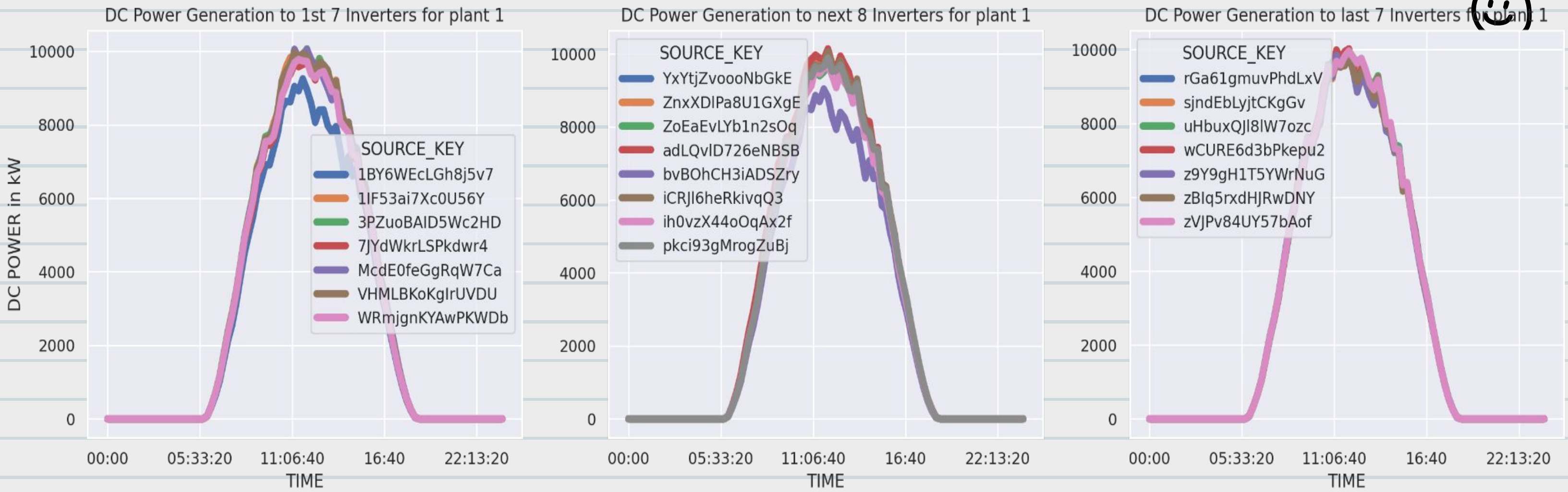
- **Abnormalities in AC_POWER & DC_POWER Generation**



- Most of the days there is some fluctuation in the power generation.
- We can also find that 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

Power Generation for Plant 1

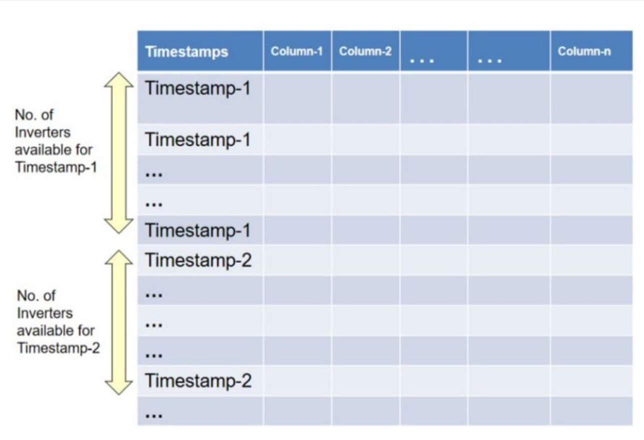


- Plant 1 seems fine as there are no major irregularities in the graph

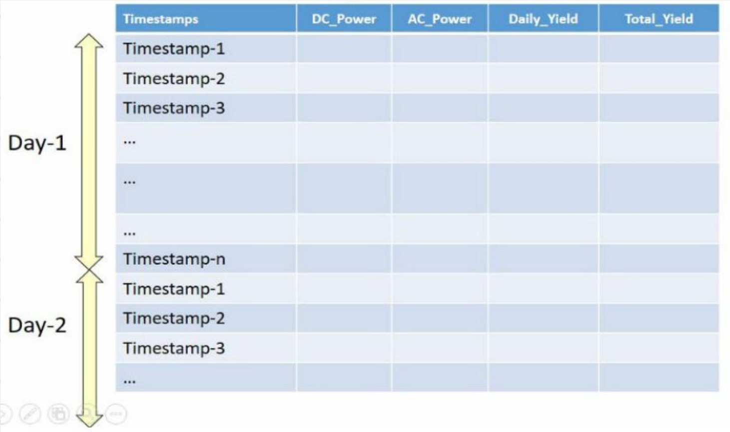
Power Generation Prediction

- AC Power generation of the plant will certainly depend on Irradiation, Ambient Temperature and Module Temperature
- Using 2 ways to predict power generation
- First, by directly training the model using weather data i.e. module temperature, ambient temperature and irradiation
- Second, via employing a physics based theoretical photovoltaic model
- Right now, we have rows based on each inverter i.e. each timestamp is being repeated the number of inverters for which data is present for that particular time stamp
- The desired format is day-wise (timestamp-wise) rows from 00:00 to 24:00 with columns being the sum of all inverter values(for DC Power, AC Power...) for that timestamp as shown
- For this, we shall first group our data inverter wise & store each group in a list. Then will merge each group with the next one using outer join on DATE_TIME column
- After grouping data inverter wise now we take sum of all the AC_POWER and DC_POWER columns as we want to predict the power on power plant basis.
- Our final data looks like:-

Data Transformation



Timestamps	Column-1	Column-2	...	...	Column-n
Timestamp-1					
Timestamp-1					
...					
...					
Timestamp-1					
Timestamp-2					
...					
...					
...					
Timestamp-2					
...					



Timestamps	DC_Power	AC_Power	Daily_Yield	Total_Yield
Timestamp-1				
Timestamp-2				
Timestamp-3				
...				
...				
...				
Timestamp-n				
Timestamp-1				
Timestamp-2				
Timestamp-3				
...				

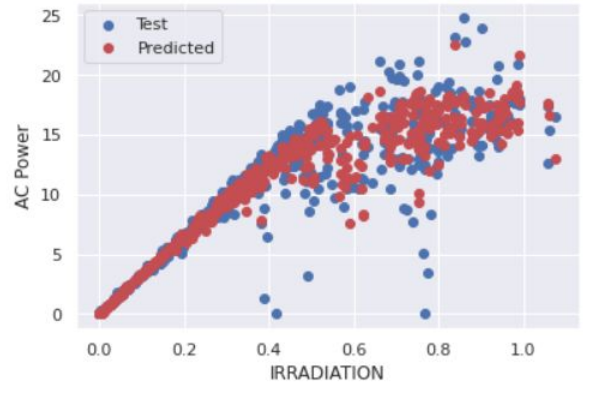
Models

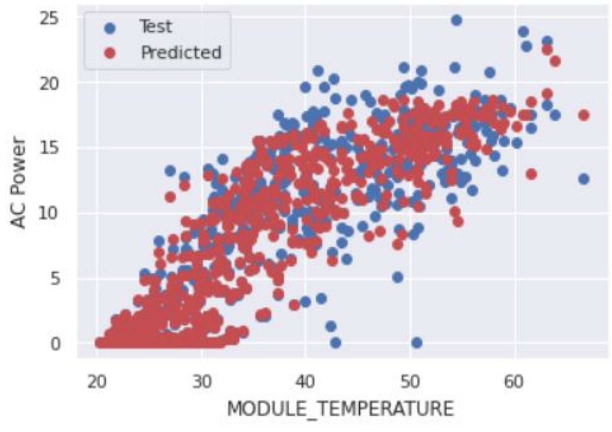
- After merging inverters for each timestamp and removing NaN values we had about 3000 rows
- We splitted our dataset into 60% for train data and 40% for test data respectively
- As the different features had different range of values like irradiation varied from 0 to 1.22 whereas ambient temperature varied from 20-35
- To tackle this we used 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 out target variable
- We used various regression based models like multiple regression, Kneighbors regression, Decision Tree Regressor

Results

Name	RMSE
Linear Regression	71.040050
Support Vector Regression	52.566475
KNeighbors Regression	63.662697
Decision Tree Regression	77.181429
Random Forest Regression	54.351226

Text(0, 0.5, 'AC Power')





Physics Based PV Model

- The photovoltaic power output of Pac (taking into account the inverter) can be modeled from the product of the open circuit voltage Vth (Thèvenin voltage) and the short circuit current Ino (Mayer-Norton current)
- The module temperature is directly proportional to the Irradiation and to the ambient temperature and can be represented by the following empirical formula

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

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

$$V_{th} = V_0[1 + \beta(T_m - T_0)]$$

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

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

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

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

Where K_1, K_2, ... , K_6 are constants.

- The results obtained are as follows:-

	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

Link to Colab file

Thanks