PV Power Plant Data Analysis and Performance Normalization

Rabin Dhakal

PhD Candidate

Department of Mechanical Engineering

Texas Tech University

Outline

- Introduction/Background
- Objectives
- Data Exploration & Analysis
- Filtering
- Developing Models
- Interpretation
- Future Works

Introduction/Background

Solar Resource Data and Modelling

- Historical long-term data for site selection
- Prediction of output power for design and financing
- Real-time measurement and forecasting for plant and grid operation

Meteorology

- Meteorology can falsely mask or emphasis trends
- It would be useful if we had same weather

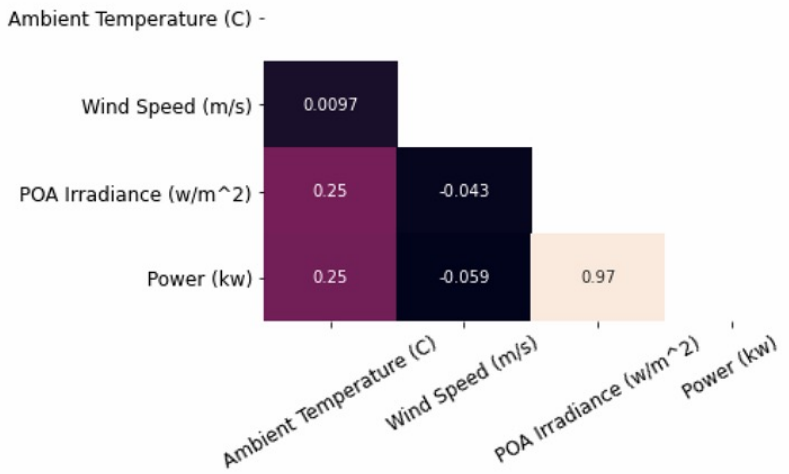
Objectives

- To conduct exploratory data analysis of solar PV power plant and model plant performance using meteorological variables
  - Find physically impossible/unlikely) data points
  - Design filter
  - Check data quality issues
  - Observe the design, rating and performance.
  - Explore future research opportunities

Data Exploration

- Data types & NaNs

```
DatetimeIndex: 1337487 entries, 2016-01-01 07:24:00-07:00 to 2020-12-31 16:45:00-07:00
Data columns (total 4 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   Ambient Temperature (C) 1337487 non-null  float64
 1   Wind Speed (m/s)        1337487 non-null  float64
 2   POA Irradiance (w/m^2)  1329517 non-null  float64
 3   Power (kw)              1329531 non-null  object
dtypes: float64(3), object(1)
```

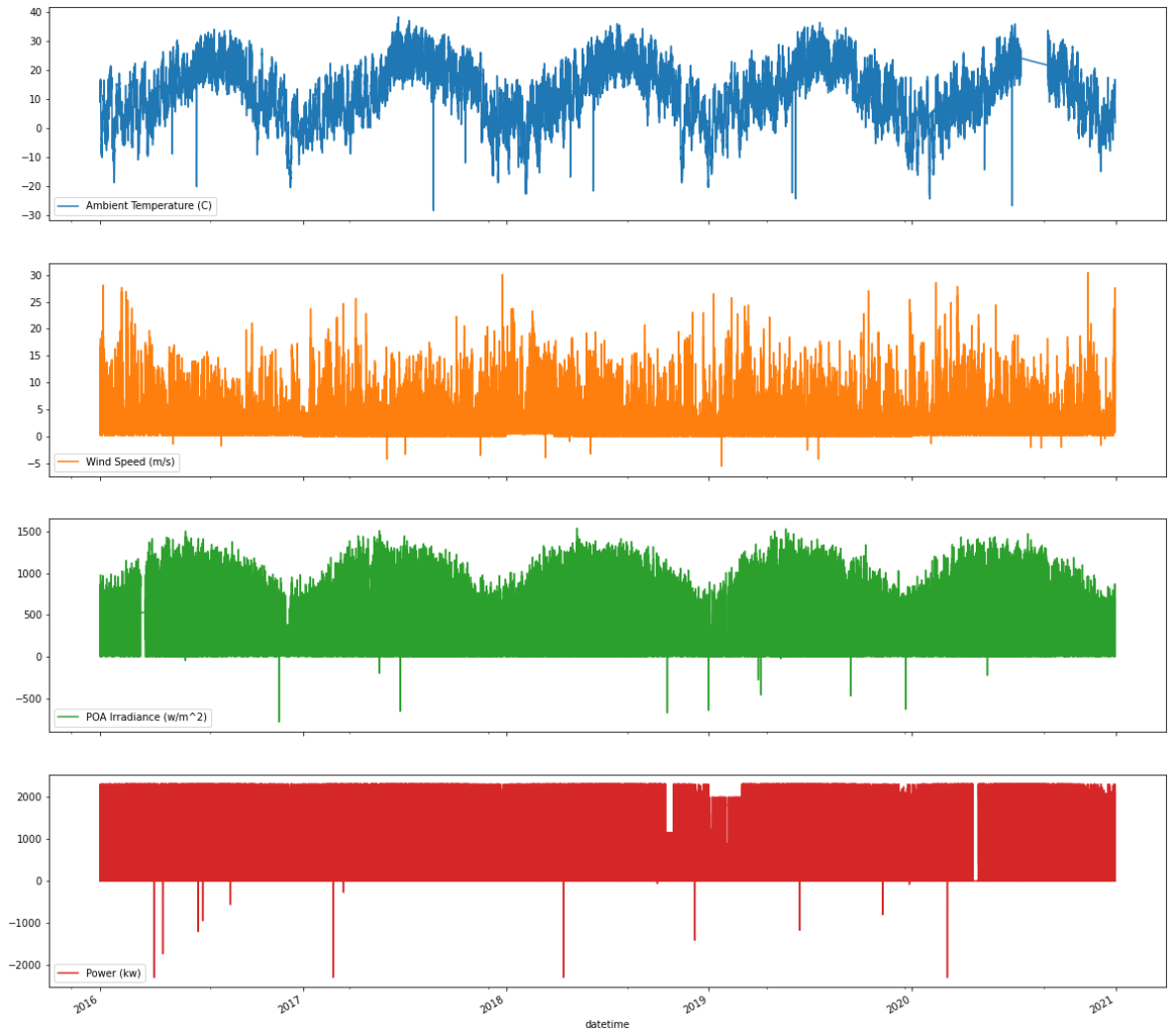


	No of NaNs	Consecutive NaNs ?	Both NaNs ?	NaNs alone ?
POA Irradiance	7970	7	7932	38
Power	7956	7		24

Use of Moving Average Method to Fill NaNs values as 99 % of POA Irradiance and Power are null at same time

Data Exploration

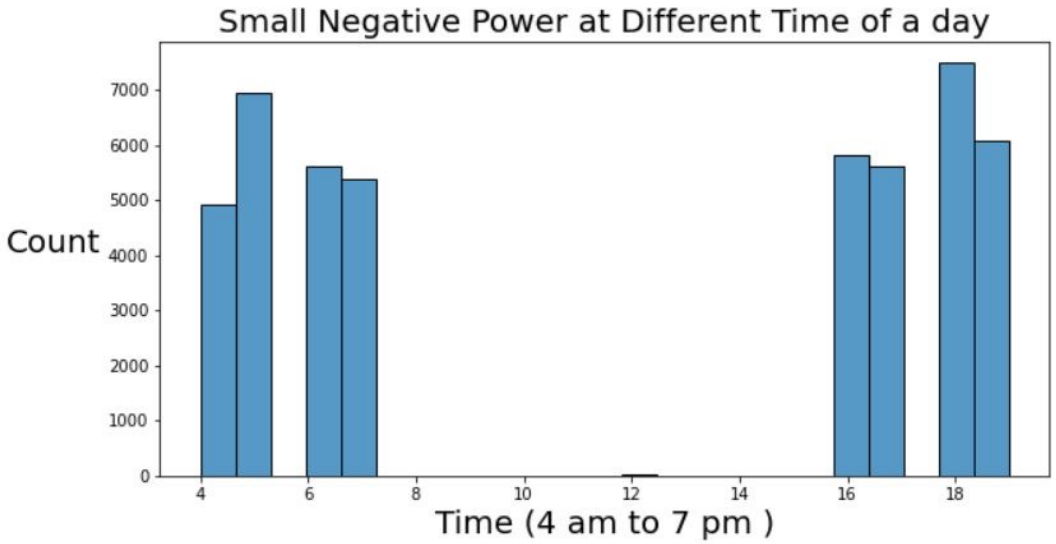
- Negative value of wind speed is physically impossible. It can be zero but not below that.
- Negative value of POA Irradiance is also not possible.
- Power also cannot be negative. (Negative power meaning power is drawn from battery. However, PV plant should have electrical system to prevent this.

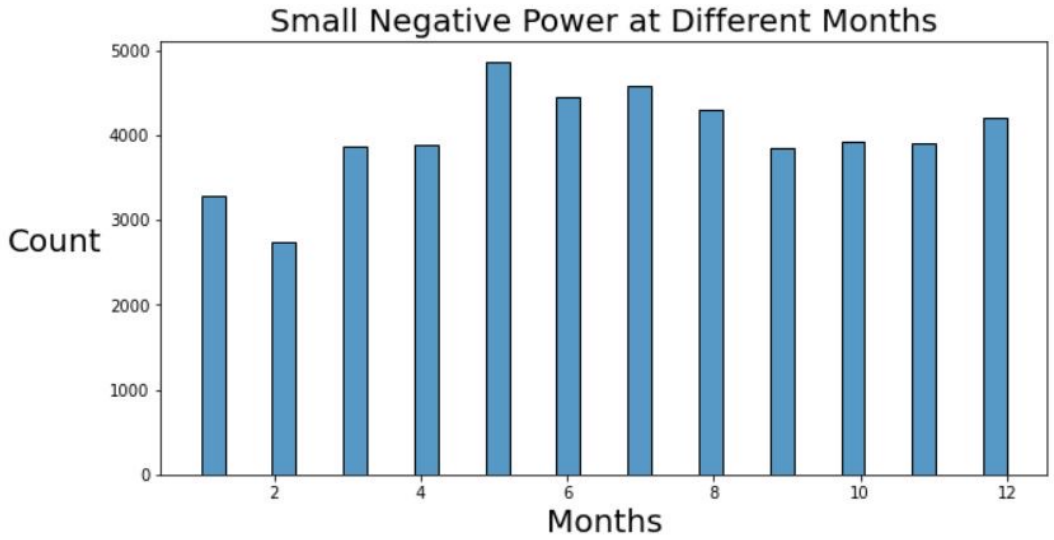


Data Exploration

- Physically Impossible Negative Values

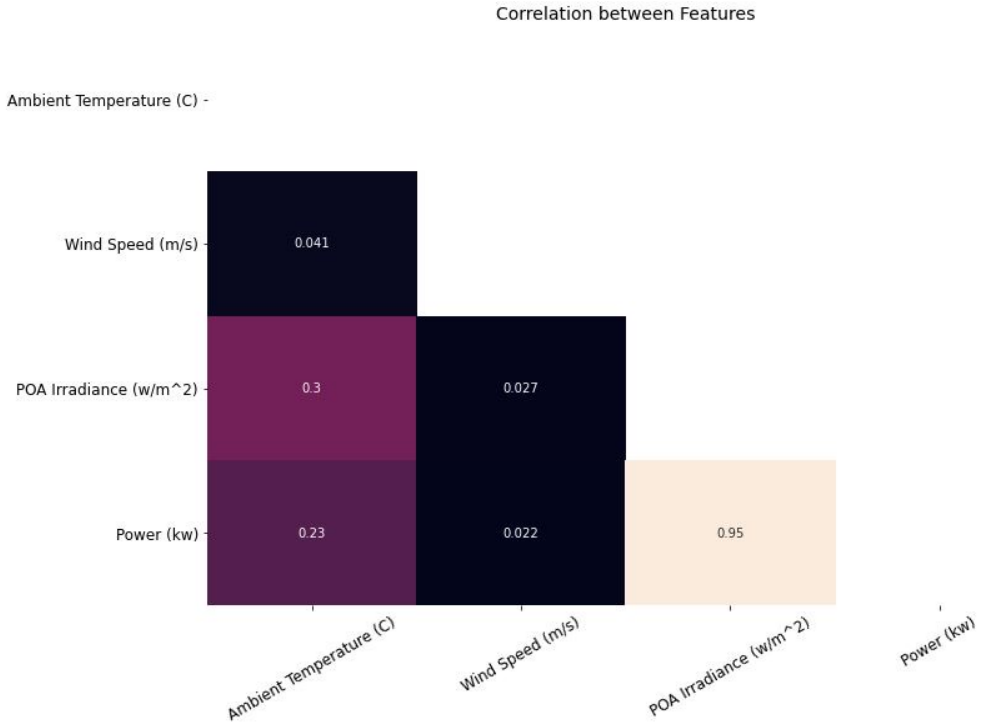
	No of –ve value	Other features also –ve ?
Wind Speed	17	No
POA Irradiance	12	No
Power	47883 (14 values < -1)	No

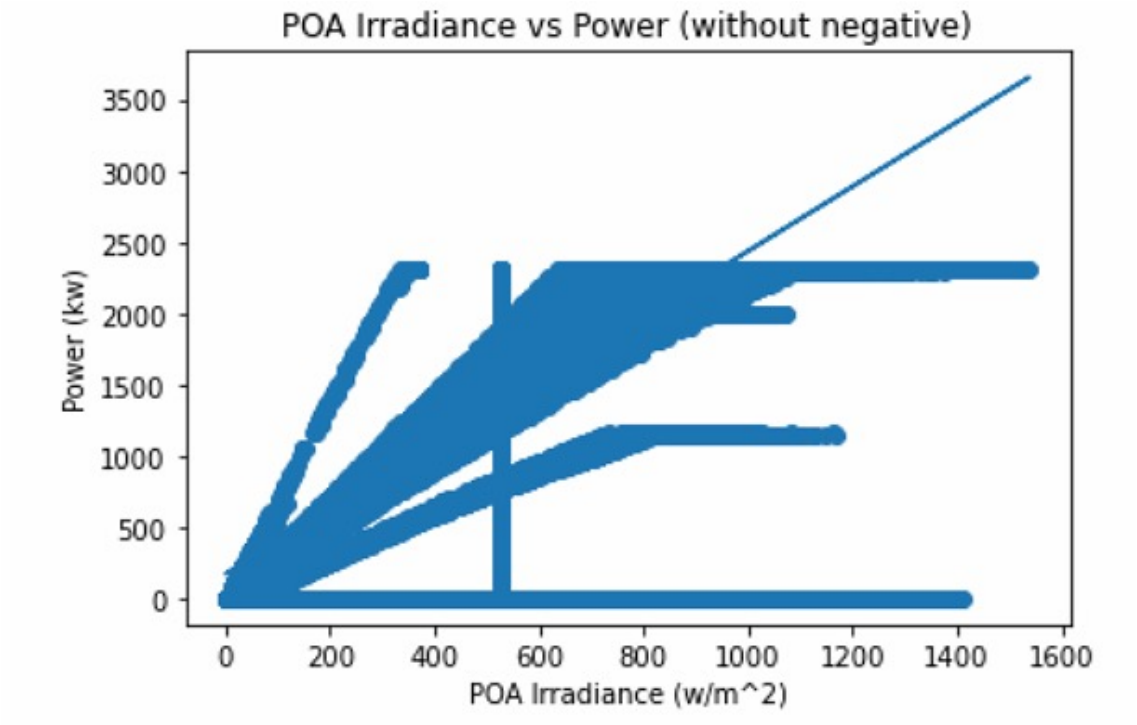




Data Exploration

- Seems power measuring instrument doesn't give good reading when the plant is about to start producing power in the morning and in the evening after it about to stop producing power



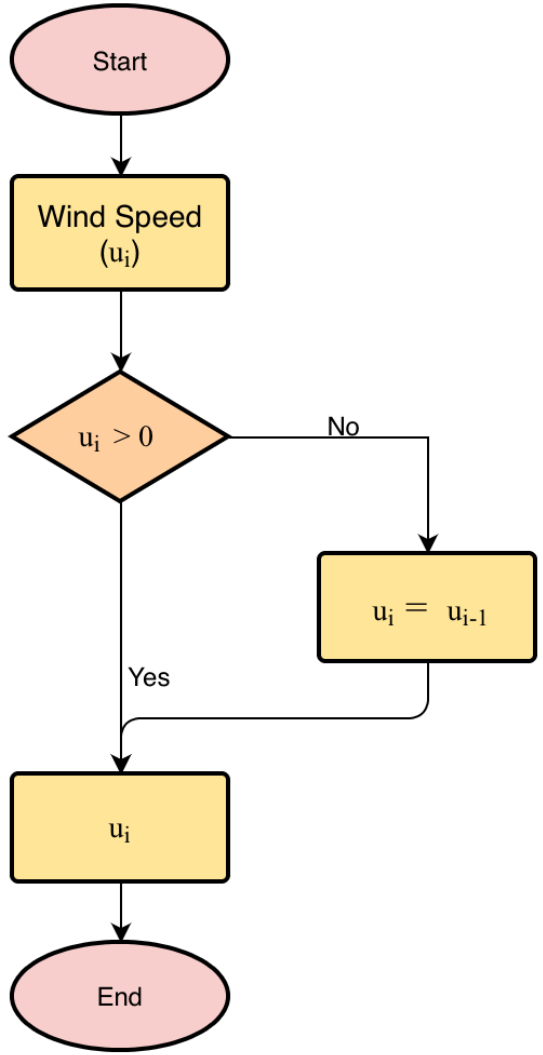


Data Exploration

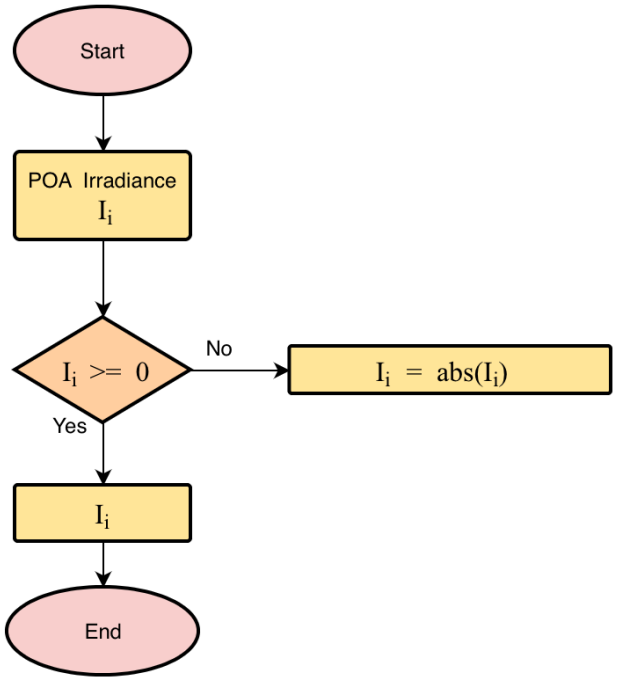
- Good correlation of Power with POA irradiance
- We can fit linear model (R^2 = 0.90)

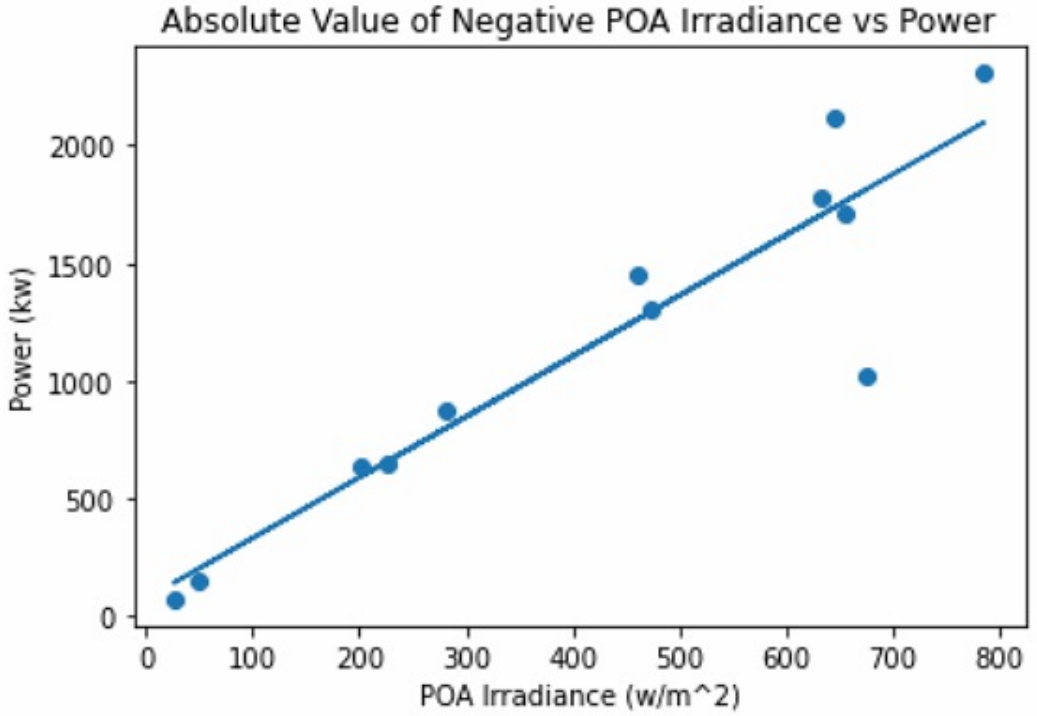
Filter Design I

- Filtering Negative Wind Speed using Persistence Model

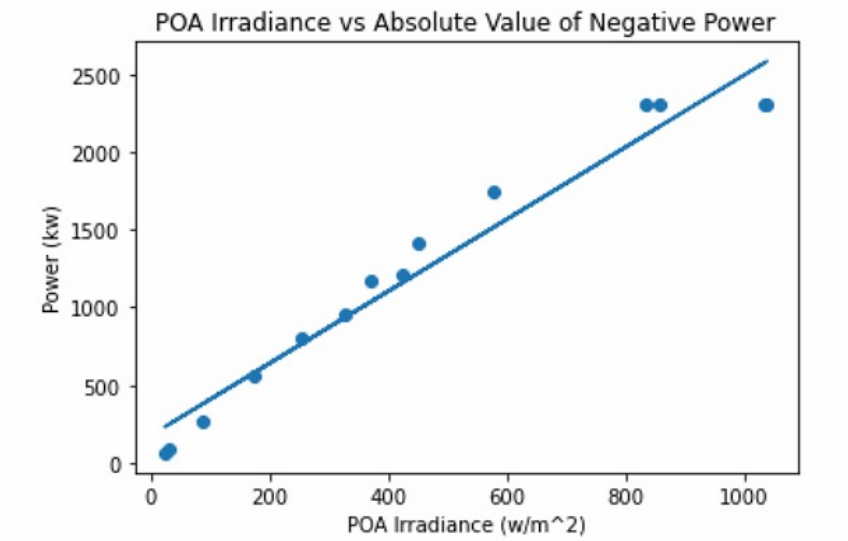


Filter Design II

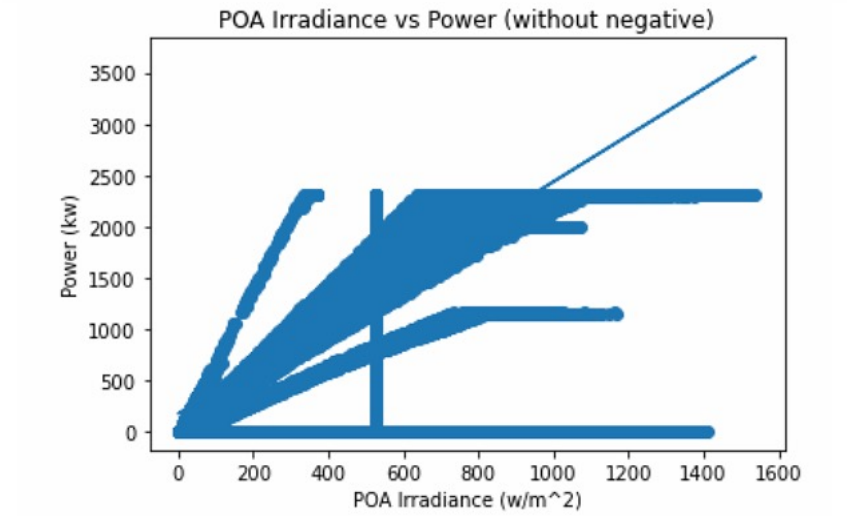




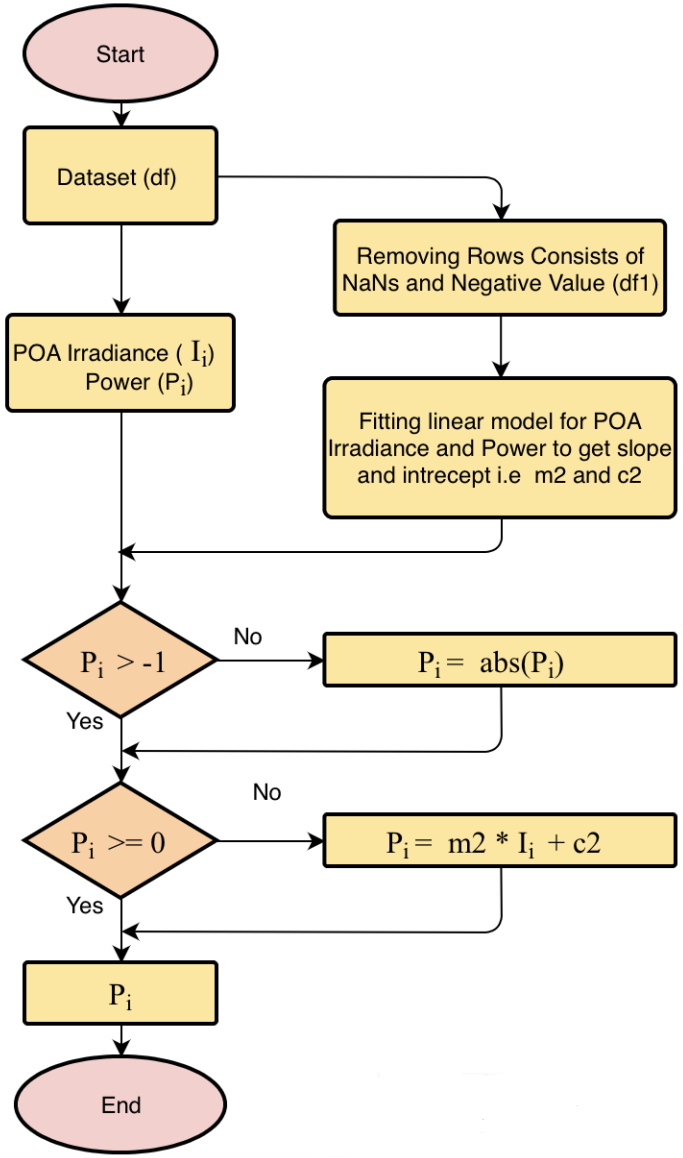
Filter Design III



R^2=0.95

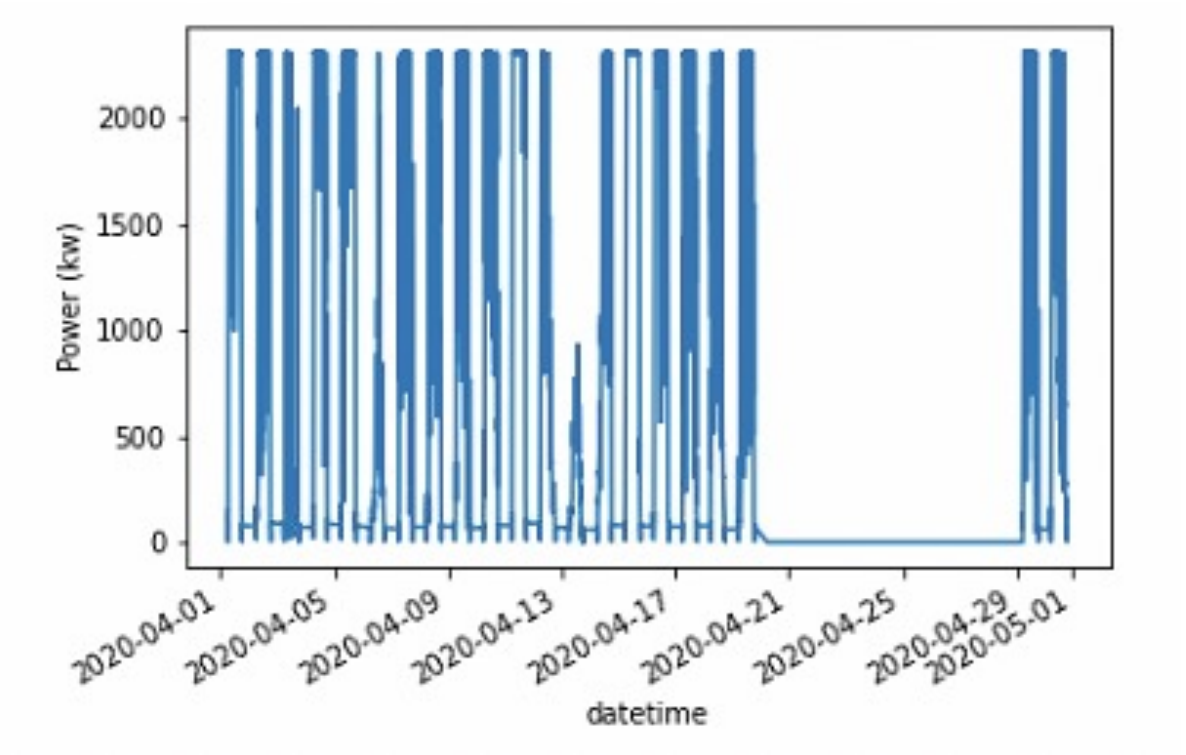


R^2=0.90



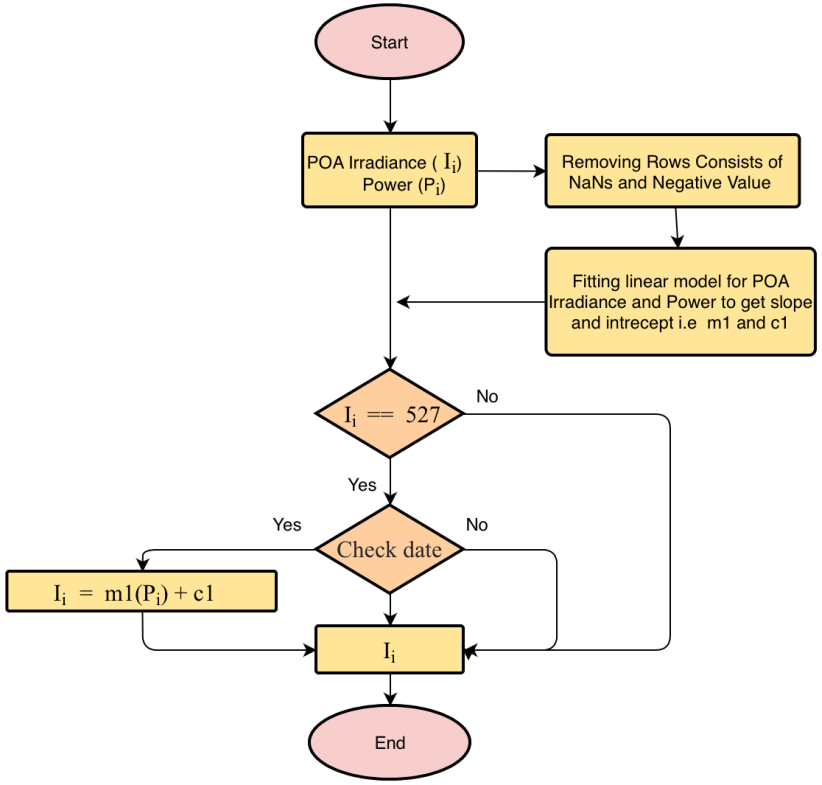
Other data issues & Filtering

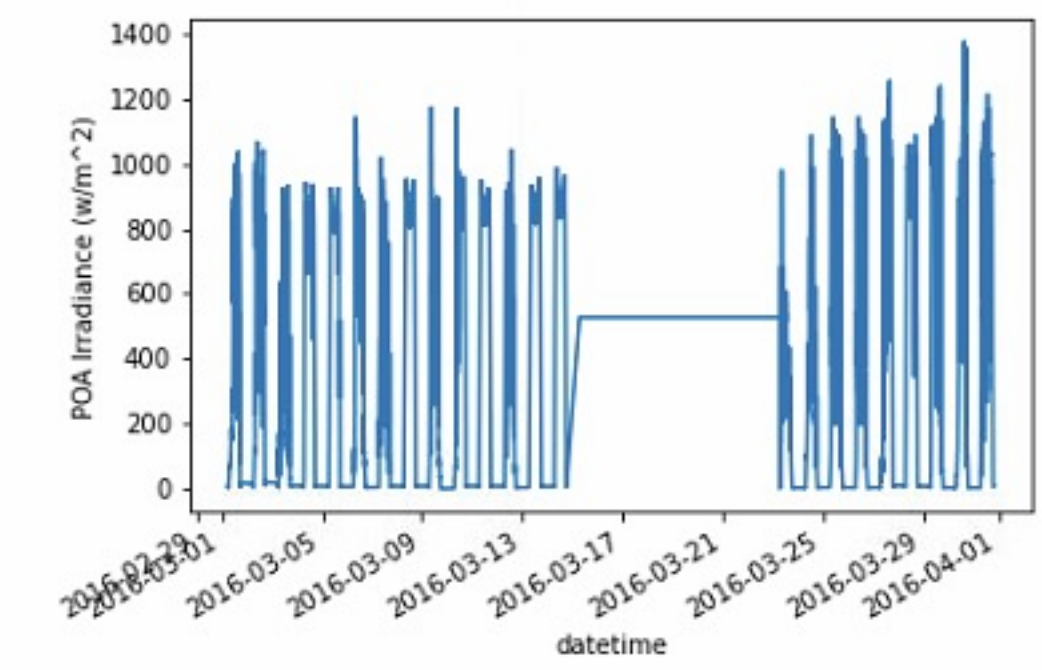
- It is clearly seen that the PV power plant is not producing energy for 9 days from April 20th to April 28 of 2020.
-
- So, we can replace it by using POA Irradiance data like Filter design III
- However, we should investigate in detail about the cause.



Other data issues & Filtering

- POA Irradiance for 7 days cannot be constant
- It can be replaced by using power

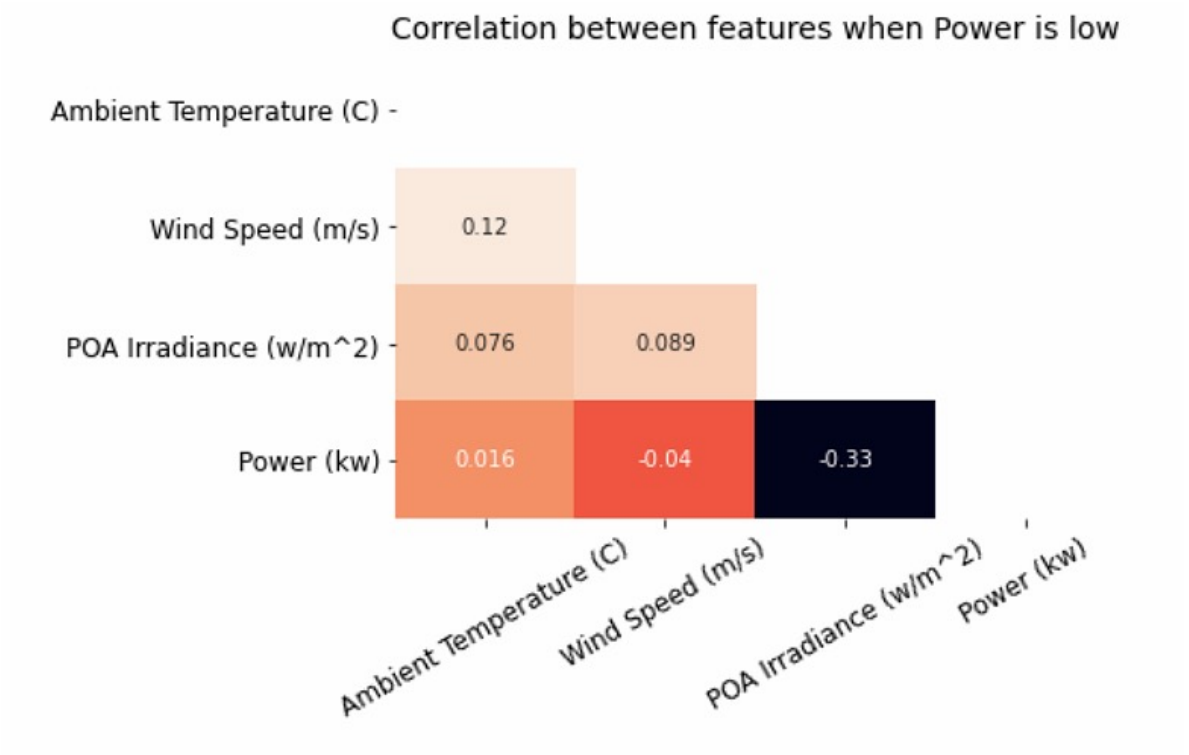




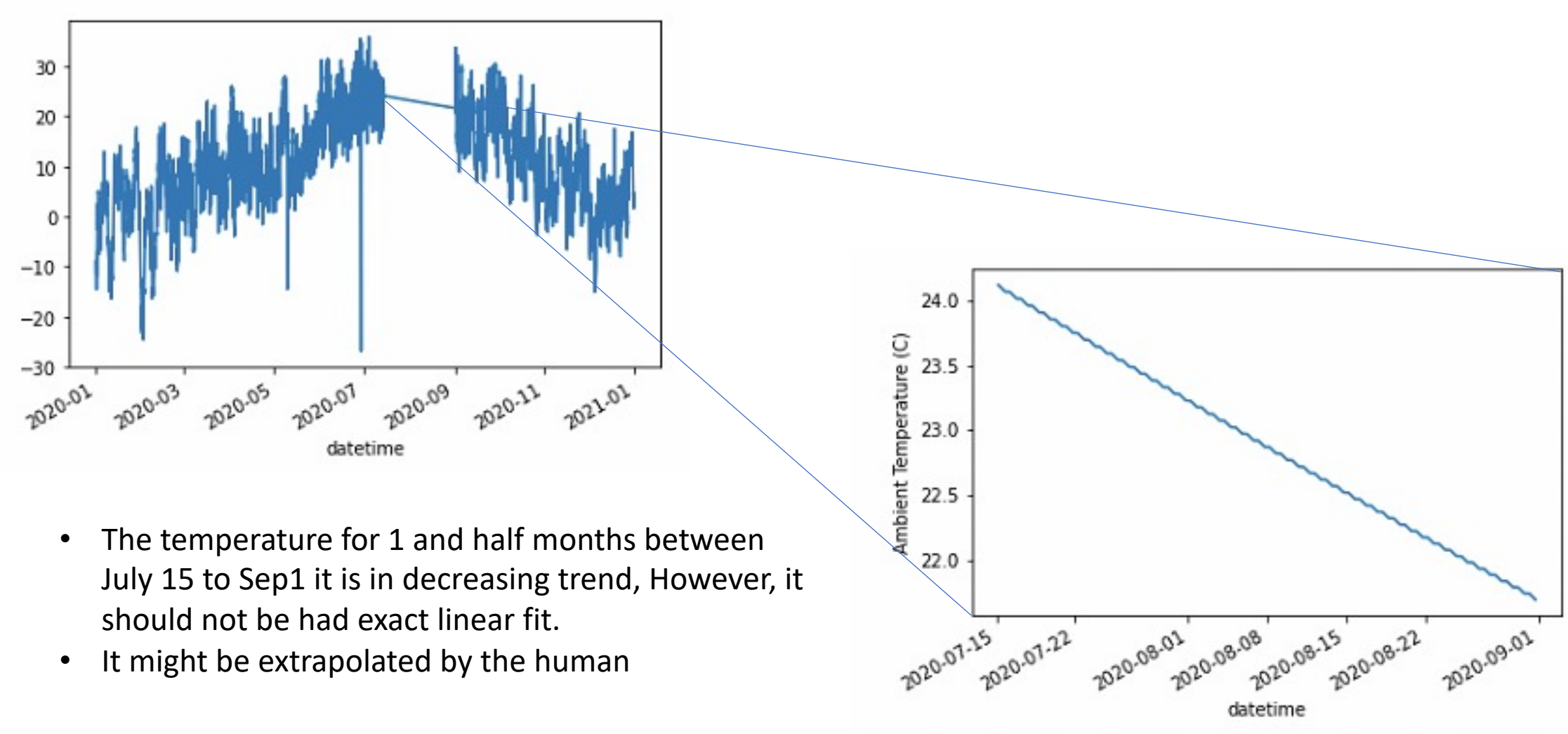
	count	mean	std	min	25%	50%	75%	max
Ambient Temperature (C)	5085.0	15.343889	4.482094	2.798603	12.412933	16.172625	18.912327	22.832796
Wind Speed (m/s)	5085.0	4.016026	2.872170	0.199602	1.797912	2.997993	6.190864	15.896758
POA Irradiance (w/m^2)	5085.0	527.000000	0.000000	527.000000	527.000000	527.000000	527.000000	527.000000
Power (kw)	5085.0	1469.695644	858.995473	1.542992	596.239082	1690.962458	2305.685610	2312.886007

Filtering Issues

- No linear relationship when there is less amount of power production i.e smaller than 100 kW



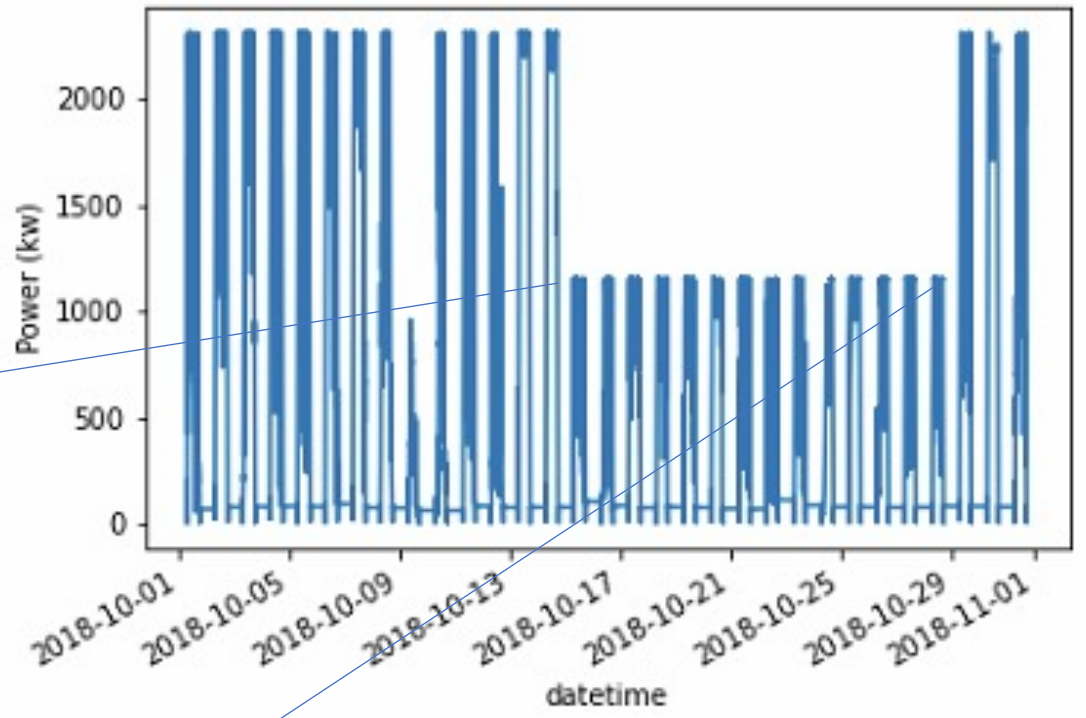
Other Data Issues



- The temperature for 1 and half months between July 15 to Sep1 it is in decreasing trend, However, it should not be had exact linear fit.
- It might be extrapolated by the human

Other Data Issues

- It is seen that the power produced for 15 days form 16th Oct to 29th Oct of the year 2018 is lower than other days.



	Ambient Temperature (C)	Wind Speed (m/s)	POA Irradiance (w/m^2)	Power (kw)
count	9132.000000	9132.000000	9132.000000	9132.000000
mean	17.221796	3.261355	522.410732	815.098878
std	5.235884	2.511559	298.351579	507.389199
min	4.191915	0.000000	2.801524	0.537126
25%	13.014186	1.597893	235.191832	356.310737
50%	18.100684	2.498122	670.574639	993.839502
75%	21.489092	4.005762	771.417087	1151.446204
max	28.296380	15.204024	1071.076406	2311.433237

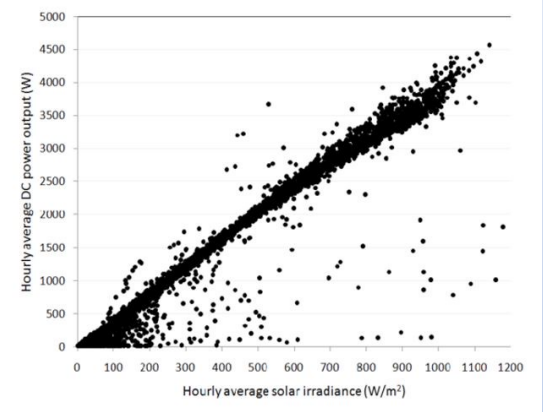
Possible Causes of Data Quality Issues

- Extrapolated by human
- Weather conditions like rain, cloudy days, snowfall on PV array
- Extreme weather condition like high wind speed than normal, much colder than other years for a day
- Issues on Pyranometer (About 8000 POA Irradiation values are zero)

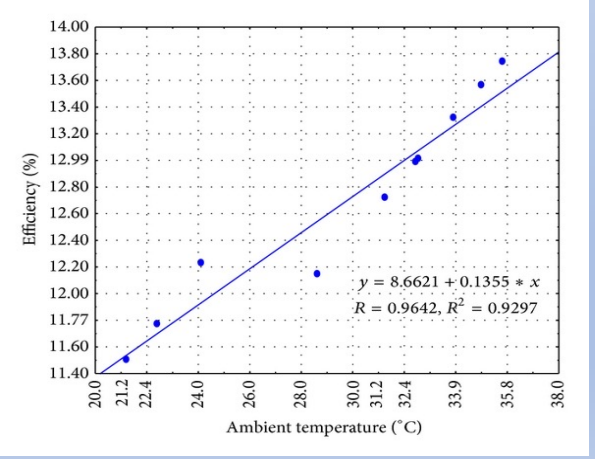
Meteorological Normalization

Effect of Meteorology

- Meteorological variables has a strong influence on the performance of PV power plant
  - Irradiance effect the power produced
  - Ambient temperature effect the efficiency of PV module
- Interdependencies of meteorological variables.
- Better to account all available meteorological variables



Source: Choi et al 2014



Source: Bhattacharya et al 2014

Model Development

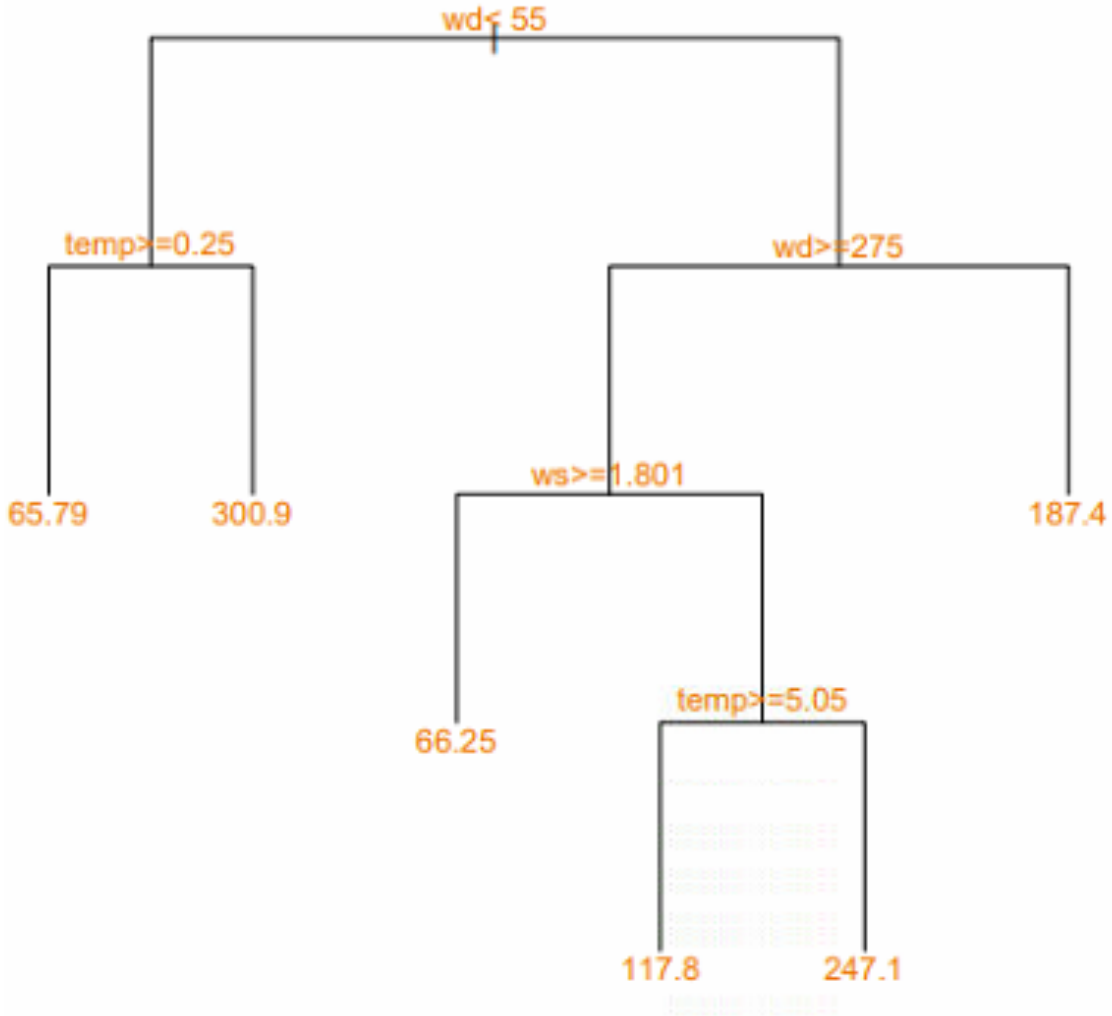
- Model is developed using the 5 years historical solar power and meteorological parameters
- Different types of model could be used including multiple linear regression, neural networks etc.
- However, regression tree approach is used here

Why Decision Trees – Modelling Benefits

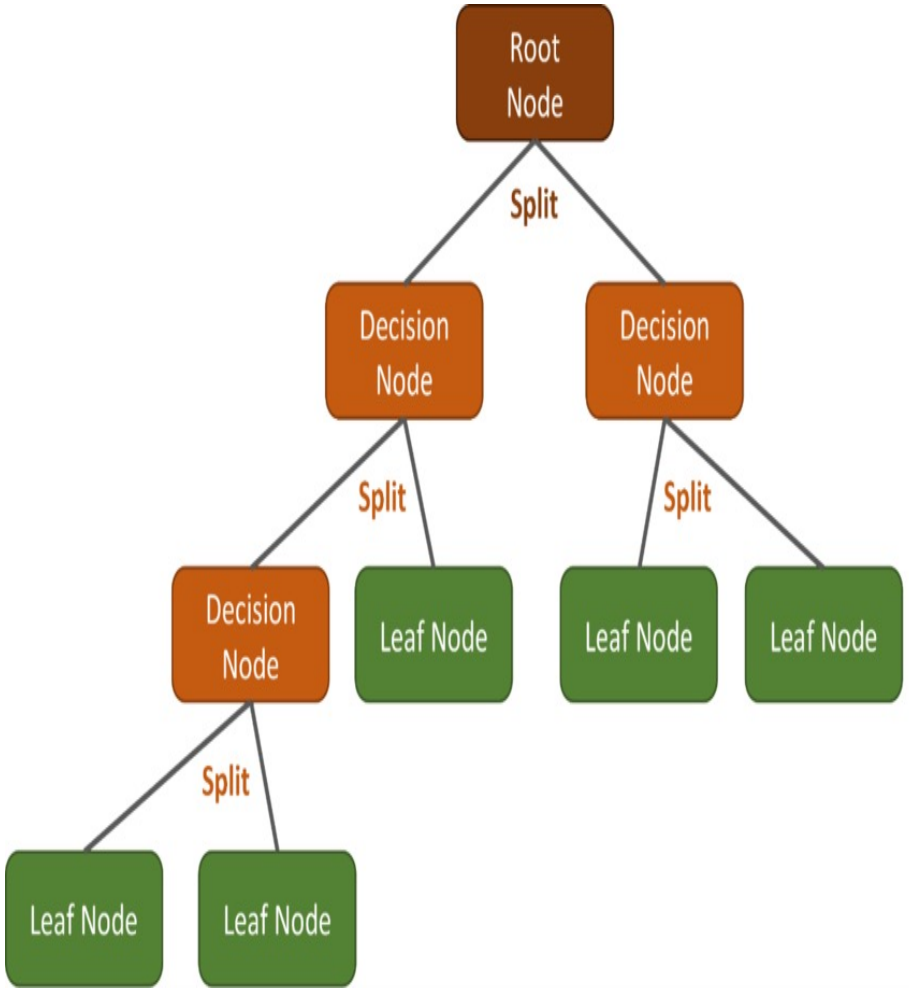
- Can model non-linear relationships
- Can take account of complex interactions
- Can model abrupt changes
- Good treatment of missing data

Modelling Approach

- Simplified model with wind speed, wind direction and ambient temperature
  - Aim to predict power production
- Interpretation
  - Output looks like a ‘tree’
  - Actual model are considerably more complex



Modelling Parameters



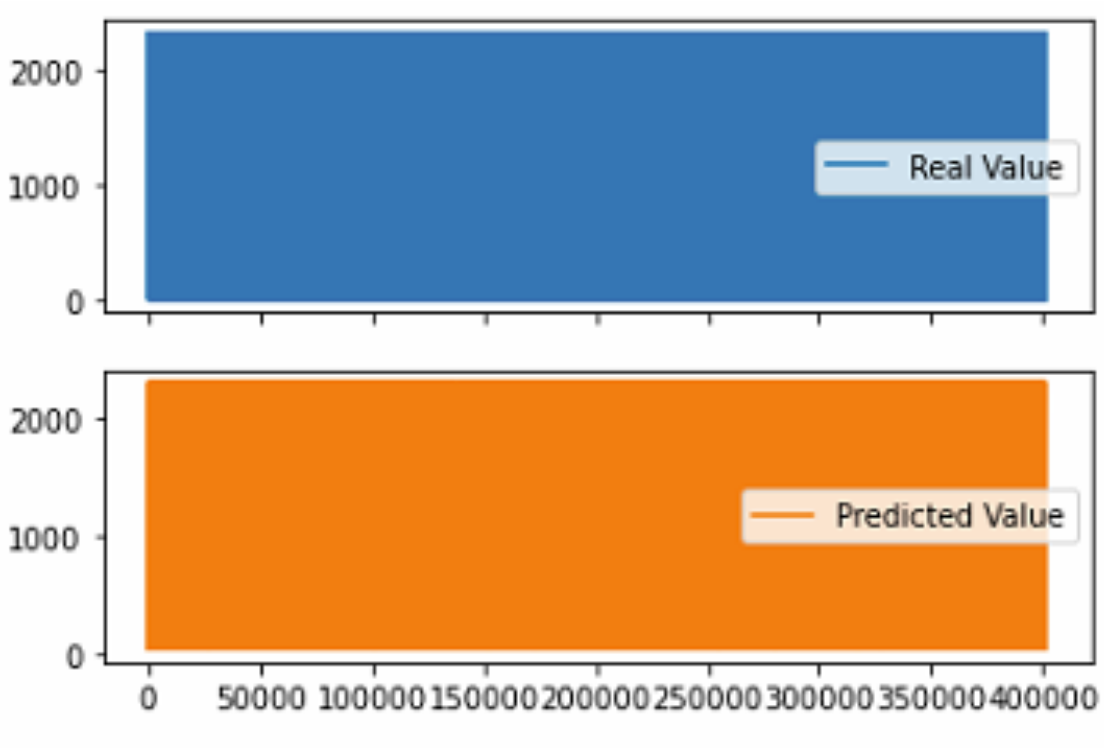
Normalizing the predictor in range of 0 to 1

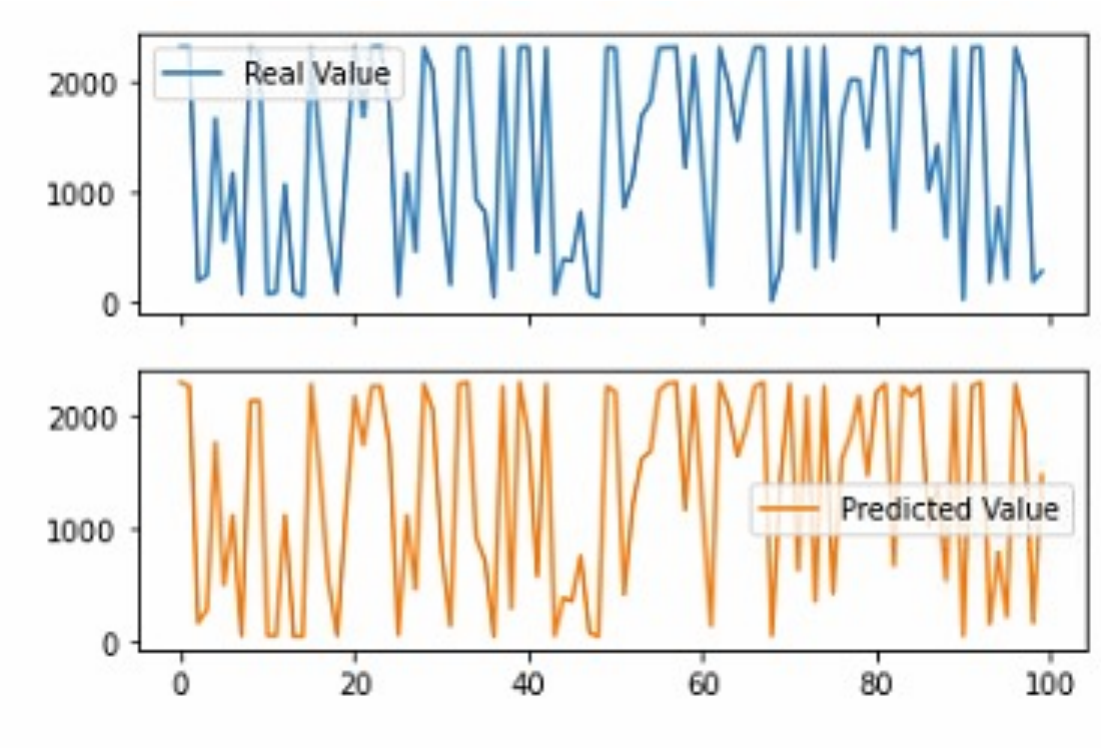
Train-Test split : 70/30

Hyperparameter Tuning

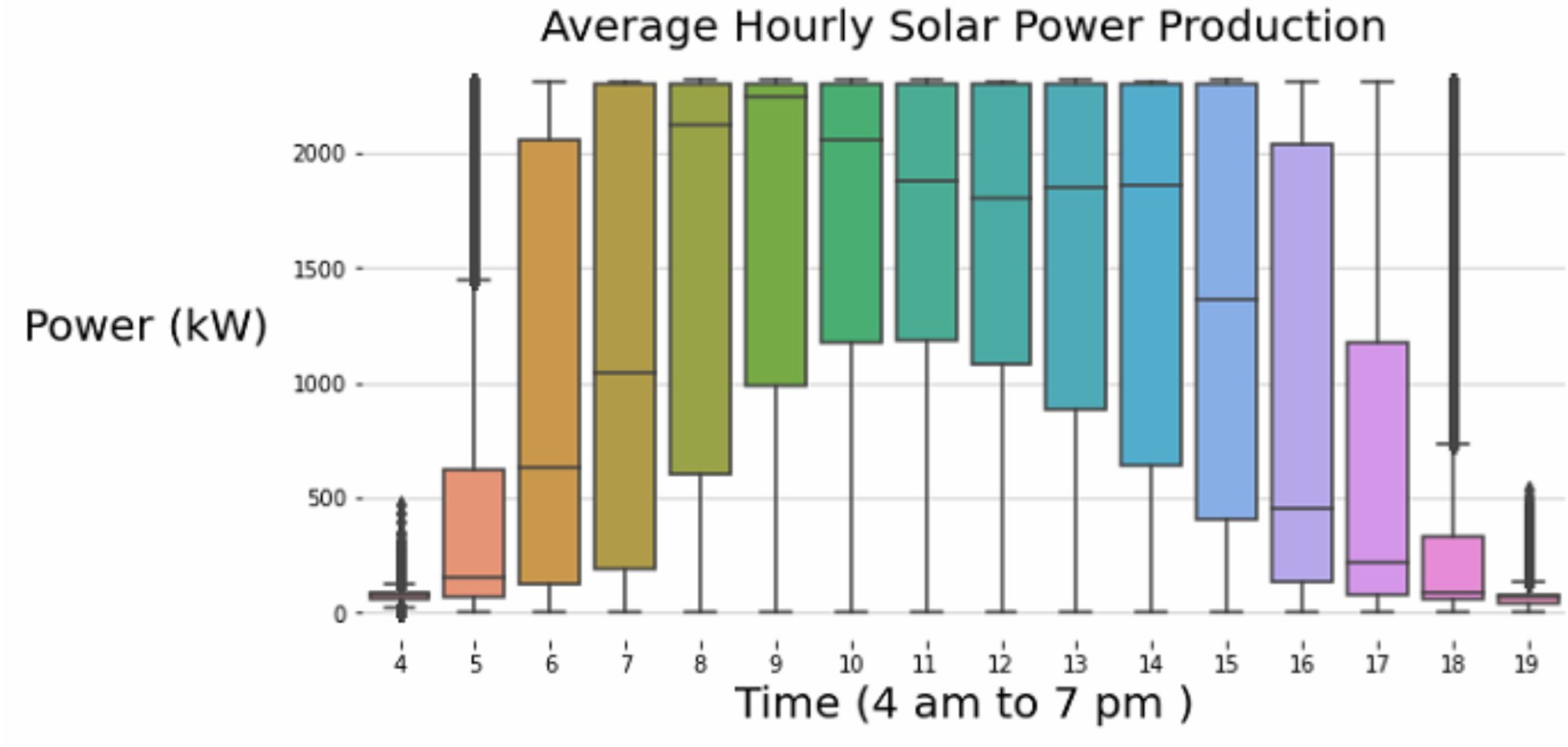
Maximum depth : 2, 5, 7 and 8 (optimum result at 7)

Result

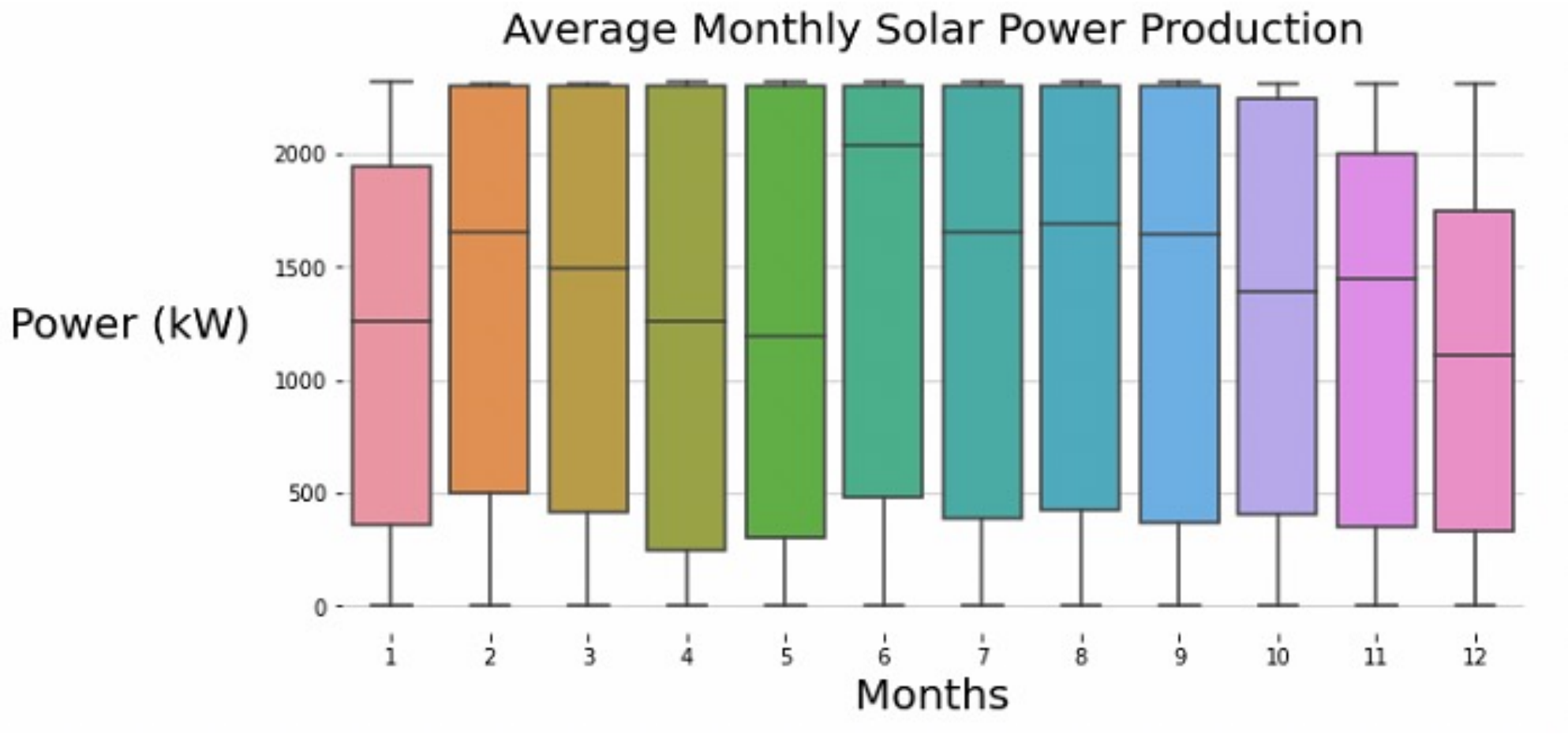

2000
1000
0
Real Value
2000
1000
0
Predicted Value
0
50000
100000
150000
200000
250000
300000
350000
400000


Real Value
2000
1000
0
2000
1000
0
Predicted Value
0
20
40
60
80
100

Insight on PV Plant



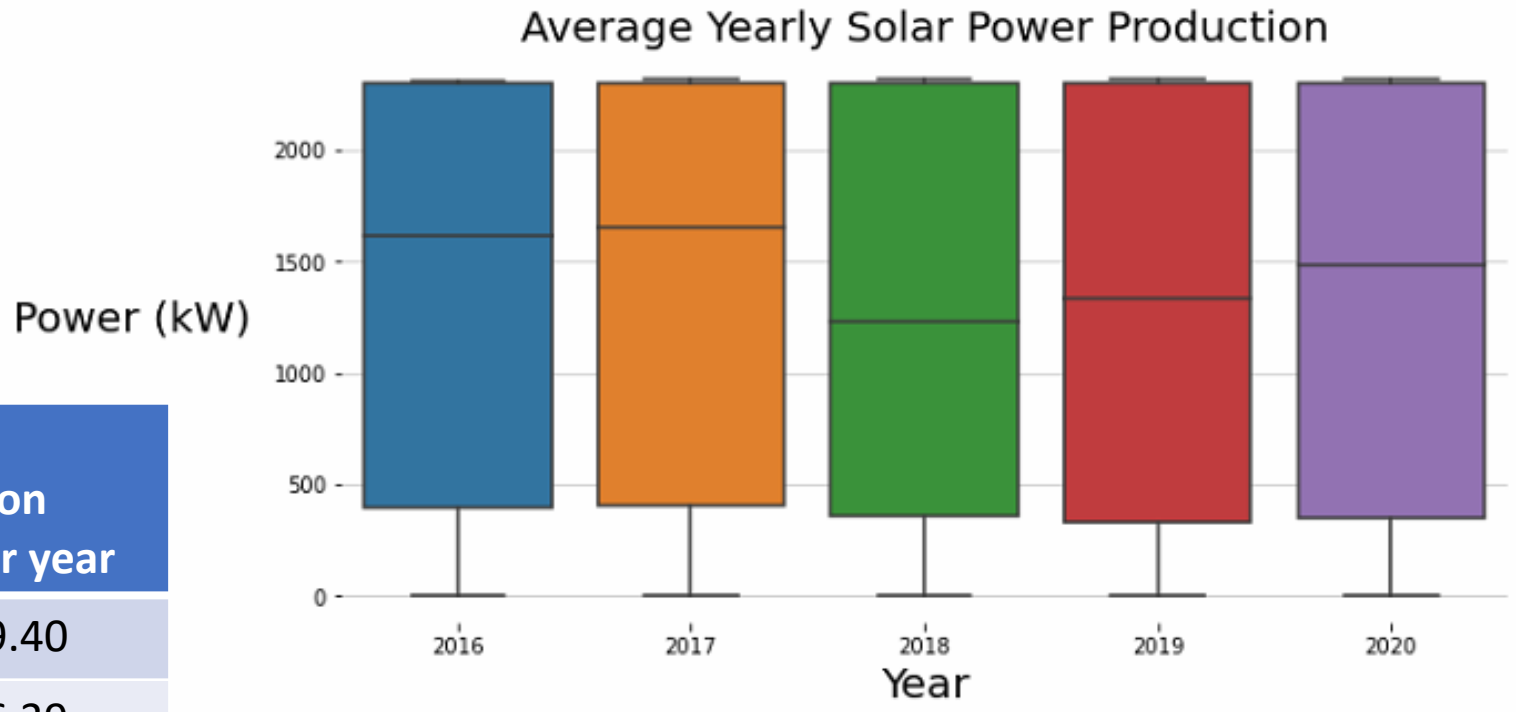
Insight on PV Plant



Insight on PV Power Plant

Assuming 24.5 % as Capacity Factor
1 MW Plant produce 2146 MWh

Plant Capacity 2.75 MW



Year	Mean Power Production (kW)	Energy Production MWh per year
2016	1379.74	6159.40
2017	1382.07	6156.39
2018	1279.23	5698.23
2019	1264.77	5633.82
2020	1316.36	5876.25

Future Works

- Optimization of the Data Filtration Techniques
- Comparative analysis with at least three other plants at different geo-location
- Normalizing performance using other techniques and compare performance.

References

1. Lovrić, Mario, et al. "Machine Learning and Meteorological Normalization for Assessment of Particulate Matter Changes during the COVID-19 Lockdown in Zagreb, Croatia." *International Journal of Environmental Research and Public Health* 19.11 (2022): 6937.
2. David Carslaw, "Meteorological Normalisation Accounting for meteorology in trends ", LAQN Seminar, King's College London 24th April 2009.

Thank You

?