Predicting the energy output of Solar Stations and Wind Farms using Machine Learning Techniques



Supervisor: Dr. Saket Verma

Dept. of Mechanical Engineering

BITS Pilani, Pilani Campus

Co-Supervisor: Mr. Arun Kr. Choudhary

MNRE New Delhi

25th December 2022

BITS Pilani

Submitted By: Gaurav Masand

2019B1A40899P

Birla Institute of Technology and Science, Pilani, Pilani

(Rajasthan), INDIA

**Table Of Contents**

|  |  |
| --- | --- |
| Topic | Page Number |
| TABLE OF CONTENTS | 2 |
| INTRODUCTION | 3 |
| DATASET | 4 |
| MATERAILS AND METHODS | 6 |
| DATA ANALYSIS | 6 |
| LSTM | 10 |
| MATHMATICS | 10 |
| RESULTS | 12 |
| XGBoost | 13 |
| MATHMATICS | 13 |
| RESULTS | 15 |
| References | 17 |

**INTRODUCTION**

Predicting the power production of renewable energy sources, such as solar plants and wind farms, is essential for maintaining a consistent and predictable electricity supply. Traditional approaches for estimating power generation rely on physical models and statistical methodologies, which might lack precision and adaptability. In recent years, machine learning algorithms have demonstrated considerable potential for enhancing the forecast of renewable energy sources' power production. These algorithms can learn and adapt to complicated data patterns and relationships, resulting in more accurate and trustworthy predictions.

Using machine learning to properly anticipate the electricity production of solar plants and wind farms is crucial for a number of reasons.

1. Stable energy supply: Accurate power generation estimates are required to provide a stable electricity supply because they enable grid managers to balance electricity demand and supply. This is particularly crucial for renewable energy sources, which might be intermittent and weather-dependent.
2. Optimization of resources: Accurate power generation predictions can assist maximise the utilisation of resources, such as storage batteries and backup generators, by allowing operators to more accurately estimate when and how much electricity will be required.
3. Financial planning: Accurate projections of electricity generation may also aid in financial planning, since they can be used to estimate revenue and influence investment decisions.
4. Improved efficiency:  Machine learning algorithms may learn and adapt to complex patterns and correlations in the data, resulting in more accurate and trustworthy predictions. This can assist increase the efficiency of renewable energy systems by lowering the demand for extra capacity and the cost of energy generation.

Overall, the application of machine learning to estimate the power production of solar plants and wind farms may help assure a consistent and dependable electricity supply, optimise the use of resources, enhance financial planning, and boost the efficiency of renewable energy systems.

**DATASET**

For this report we had worked on Data from 6 Wind Farms and 8 Solar Stations. The data was for the beginning of 2019 (i.e., 2019-01-01 00:00:00) to the end of 2020 (i.e., 2020-12-31 23:45:00) with the time interval of 15 minutes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Wind Farm | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 |
| Nominal Capacity(MW) | 99 | 200 | 99 | 66 | 36 | 96 |
| Datapoints | 70176 | 70176 | 70176 | 70176 | 70176 | 70176 |

Table 1. Capacity of the wind farms and the Number of Datapoints

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Solar Station | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 | Site 7 | Site 8 |
| Nominal Capacity (MW) | 50 | 130 | 30 | 130 | 110 | 35 | 30 | 30 |
| Datapoints | 70176 | 70176 | 52608 | 70176 | 70176 | 70176 | 70176 | 69408 |

Table 2. Capacity of the Solar Stations and the Number of Datapoints.

|  |  |  |
| --- | --- | --- |
| Variable Name | Shortened Name | Description |
| Wind speed at height of 10 meters (m/s) | WS\_10 | The wind speed was recorded at 10 meters above the ground |
| Wind speed at height of 30 meters (m/s) | WS\_30 | The wind speed was recorded at 30 meters above the ground |
| Wind speed at height of 50 meters (m/s) | WS\_50 | The wind speed was recorded at 50 meters above the ground |
| Wind direction at height of 10 meters (°) | WD\_10 | The wind direction was recorded at 10 meters above the ground |
| Wind direction at height of 30 meters (°) | WD\_30 | The wind direction was recorded at 30 meters above the ground |
| Wind direction at height of 50 meters (°) | WD\_50 | The wind direction was recorded at 50 meters above the ground |
| Wind speed - at the height of wheel hub (m/s) | WS\_cen | The wind speed was recorded at centre of wheel hub |
| Wind direction at the height of wheel hub (˚) | WD\_cen | The wind direction was recorded at centre of wheel hub |
| Air temperature (°C) | Air\_T | Air dry-bulb temperature at 1.5  meters above the ground |
| Atmosphere(hpa) | Air\_P | Atmosphere at 1.5 meters above the ground |
| Relative humidity (%) | Air\_H | Air relative humidity at 1.5 meters above the ground |
| Power(MW) | Power(MW) | The total wind power generation |

Table 3. Description of the different variables for Wind Farms.

|  |  |  |
| --- | --- | --- |
| Variable Name | Shortened Name | Description |
| Total solar irradiance  (W/m2) | TSI | Solar power over all wavelengths per square meter |
| Direct normal irradiance  (W/m2) | DNI | The amount of solar radiation received per square meter by a surface that is always held perpendicular to the rays |
| Global horizontal  irradiance (W/m2) | GHI | The total amount of shortwave radiation received by a surface horizontal to the ground |
| Air temperature (°C) | Air\_T | Air dry-bulb temperature at 1.5  meters above the ground |
| Atmosphere(hpa) | Air\_P | Atmosphere at 1.5 meters above the ground |
| Relative humidity (%) | Air\_H | Air relative humidity at 1.5 meters above the ground |
| Power(MW) | Power(MW) | The total wind power generation |

Table 4. Description of the different variables for Solar Stations.

**MATERIAL AND METHODS**

**DATA ANALYSIS**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Wind farm name** | **Statistics** | **Power output (MW)** | **WS\_cen** | **WD\_cen** | **Air\_T** | **Air\_H** |
| Farm site 1 | Mean | 23.4 | 6.4 | 217.0 | 8.5 | 37.6 |
| Minimum | 0.0 | 0.0 | 0.0 | −24.1 | 0.0 |
| Maximum | 98.1 | 30.2 | 358.5 | 36.1 | 93.1 |
| Standard deviation | 24.1 | 3.9 | 85.4 | 13.4 | 18.9 |
| Farm site 2 | Mean | 72.7 | 7.5 | 206.8 | 8.7 | 33.4 |
| Minimum | 0.0 | 0.0 | 0.0 | −24.5 | 0.0 |
| Maximum | 201.2 | 28.8 | 359.8 | 37.6 | 97.6 |
| Standard deviation | 55.7 | 5.7 | 87.0 | 13.2 | 7.1 |
| Farm site 3 | Mean | 18.1 | 4.0 | 179.1 | 17.4 | 58.5 |
| Minimum | 0.0 | 0.0 | 0.0 | −14.3 | 0.0 |
| Maximum | 94.3 | 36.9 | 360.0 | 36.3 | 94.3 |
| Standard deviation | 22.6 | 3.3 | 110.5 | 9.9 | 23.8 |
| Farm site 4 | Mean | 17.4 | 5.5 | 147.3 | 13.8 | 80.7 |
| Minimum | 0.0 | 0.0 | 0.0 | −3.8 | 0.0 |
| Maximum | 64.6 | 31.1 | 356.8 | 35.3 | 100.0 |
| Standard deviation | 20.0 | 3.9 | 120.7 | 8.2 | 18.8 |
| Farm site 5 | Mean | 6.7 | 4.7 | 184.9 | 13.6 | 69.9 |
| Minimum | 0.0 | 0.0 | 0.0 | −9.9 | 0.0 |
| Maximum | 35.4 | 26.2 | 358.6 | 35.8 | 100.0 |
| Standard deviation | 10.1 | 3.1 | 113.2 | 8.9 | 32.2 |
| Farm site 6 | Mean | 28.8 | 8.1 | 94.0 | 21.2 | 78.6 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Maximum | 114.4 | 23.8 | 360.0 | 37.1 | 99.4 |
| Standard deviation | 28.0 | 3.8 | 91.2 | 6.4 | 10.9 |

Table 5. Statistics of the different variables for Wind Farms.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Solar station name** | **Statistics** | **Power output (MW)** | **TSI** | **DNI** | **GHI** | **Air\_T** |
| Solar station site 1 | Mean | 9.7 | 266.4 | 93.3 | 67.7 | 13.1 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | −18.2 |
| Maximum | 48.3 | 1359.0 | 980.0 | 989.0 | 41.2 |
| Standard deviation | 13.7 | 368.0 | 200.8 | 111.2 | 14.3 |
| Solar station site 2 | Mean | 19.6 | 169.6 | 122.4 | 78.3 | 13.7 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | −13.9 |
| Maximum | 109.4 | 1041.9 | 751.8 | 561.8 | 40.5 |
| Standard deviation | 28.0 | 248.4 | 179.2 | 117.6 | 12.1 |
| Solar station site 3 | Mean | 5.2 | 81.1 | 111.1 | 66.3 | — |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | — |
| Maximum | 29.9 | 1117.0 | 893.0 | 656.0 | — |
| Standard deviation | 8.1 | 205.8 | 199.1 | 98.9 | — |
| Solar station site 4 | Mean | 16.5 | 150.1 | 138.9 | 20.8 | 18.6 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | −5.3 |
| Maximum | 114.7 | 1237.4 | 1010.3 | 151.0 | 49.8 |
| Standard deviation | 27.5 | 253.5 | 210.6 | 31.5 | 10.3 |
| Solar station site 5 | Mean | 14.5 | 164.3 | 147.9 | 115.0 | 17.8 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | −6.6 |
| Maximum | 99.6 | 1467.0 | 1962.0 | 1208.0 | 39.5 |
| Standard deviation | 23.9 | 273.5 | 234.9 | 203.1 | 9.6 |
| Solar station site 6 | Mean | 6.4 | 244.1 | 216.0 | 54.1 | 20.6 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | 2.9 |
| Maximum | 31.2 | 1365.4 | 1179.8 | 296.2 | 36.7 |
| Standard deviation | 9.2 | 355.9 | 338.0 | 69.4 | 5.8 |
| Solar station site 7 | Mean | 5.4 | 206.8 | — | — | — |
| Minimum | 0.0 | 0.0 | — | — | — |
| Maximum | 29.8 | 3262.0 | — | — | — |
| Standard deviation | 8.0 | 300.5 | — | — | — |
| Solar station site 8 | Mean | 4.2 | 163.2 | 142.0 | 21.2 | 18.0 |
| Minimum | 0.0 | 0.0 | 0.0 | 0.0 | −8.0 |
| Maximum | 29.4 | 1214.5 | 1056.7 | 157.9 | 47.6 |
| Standard deviation | 6.5 | 245.4 | 213.5 | 31.9 | 8.6 |

Table 6. Statistics of the different variables for Wind Farms.

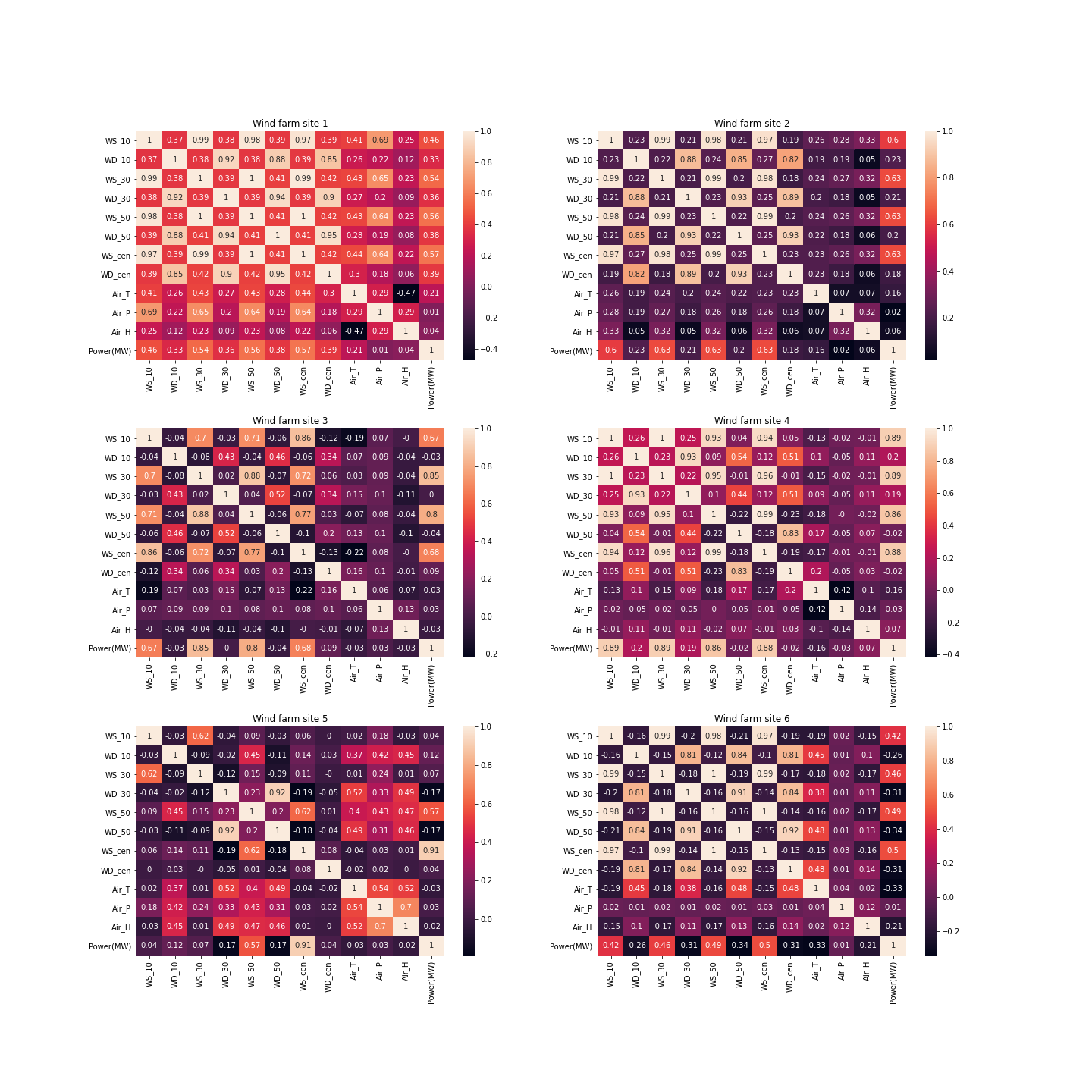


Fig 1. Heatmap of the variables of the Wind Farms.

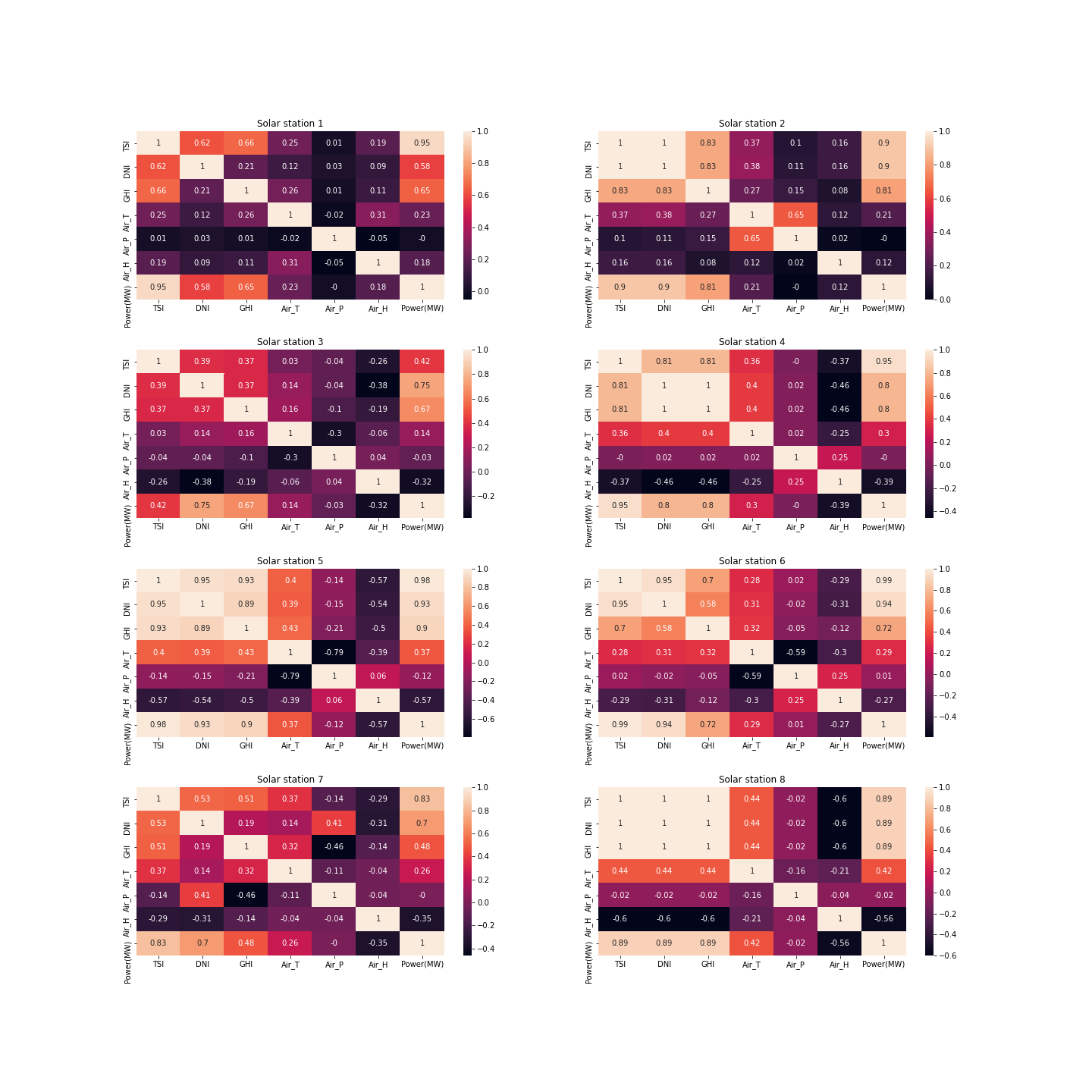


Fig 2. Heatmap of the variables of the Solar Stations.

Picking the appropriate input feature variables can increase the performance of a data-driven forecasting model; consequently, correlation analysis is essential for selecting the variables. Wind velocity and solar radiation are, respectively, the most significant elements for creating wind and solar energy. The Pearson correlation coefficient (PCC) quantifies the linear relationship between two data sets.

It can be seen that the PCC between wind speed and power production in the wind dataset is significantly greater than that of other factors, such as temperature and pressure (Figure 1). As seen in Figure 2, total sun irradiance has the greatest PCC with power production in the solar dataset.

**LSTM**

Long Short-Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), They work tremendously well on a large variety of problems, and are now widely used for time series forecasting

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behaviour, not something they struggle to learn.

This is very useful in case for time series forecasting as the model can predict the future power output based on the previous data.

**MATHMATICS**

The heart of a LSTM network is its cell or say cell state which provides a bit of memory to the LSTM so it can remember the past.

In LSTM we will have 3 gates:

* Input Gate
* Forget Gate
* Output Gate

Gates in LSTM are the sigmoid activation functions i.e they output a value between 0 or 1 and in most of the cases it is either 0 or 1.

“0” means the gates are blocking everything.

“1” means gates are allowing everything to pass through it.

The equations for the gates in LSTM are:

 [Input Gate]

 [Forget Gate]

[Output Gate]

*it→ represents input gate*

*ft→ represents forget gate*

*ot→ represents output gate*

*σ→ represents sigmoid function*

*wx→ weight for the respective gate(x) neurons*

*ht-1→ output of the previous lstm block(at timestamp t-1)*

*xt→ input at the current timestamp*

*bx→ biases for the respective gates(x)*

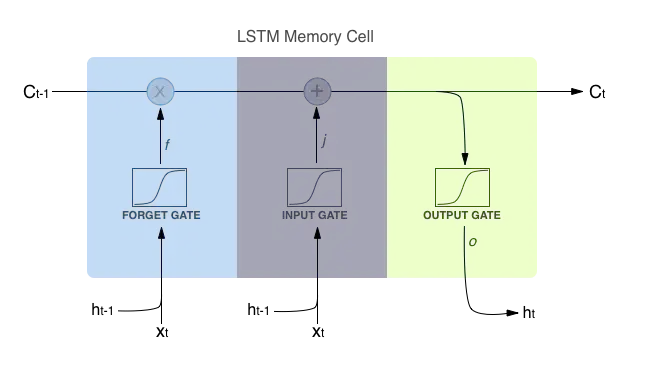


Fig 3. LSTM Memory Cell

The equations for the cell state, candidate cell state and the final output:

**

**

**

*ct1→represents candidate for the cell state at timestamp(t)*

*ct→ cell state(memory) at timestamp(t).*

Diagram

Description automatically generated

Fig 4. Block of LSTM at any timestamp

**Results of LSTM**

The Power at time t , t+15, t+30 is given as input to the Neural Network and Power at time t+45 is the corresponding output.

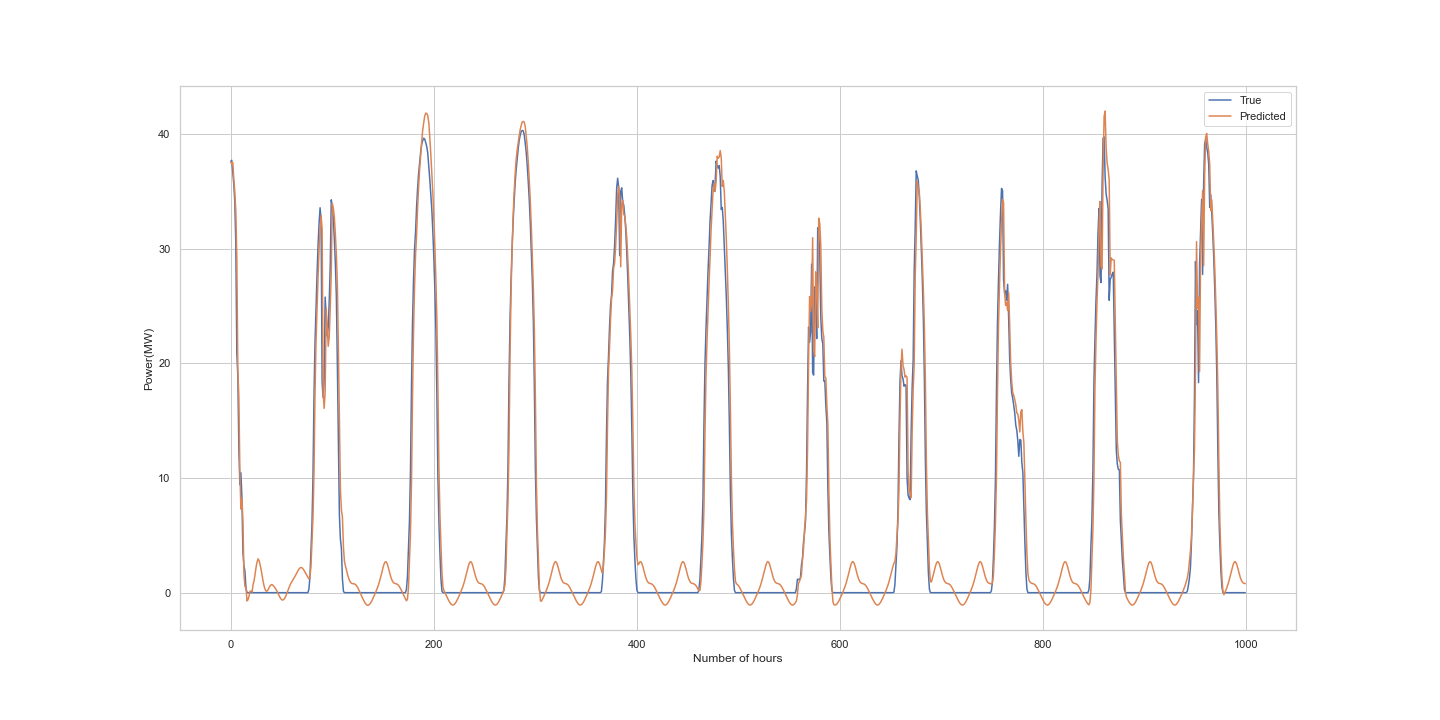


Fig 5. Expected vs Predicted values of Power by using LSTM for Solar Station 1

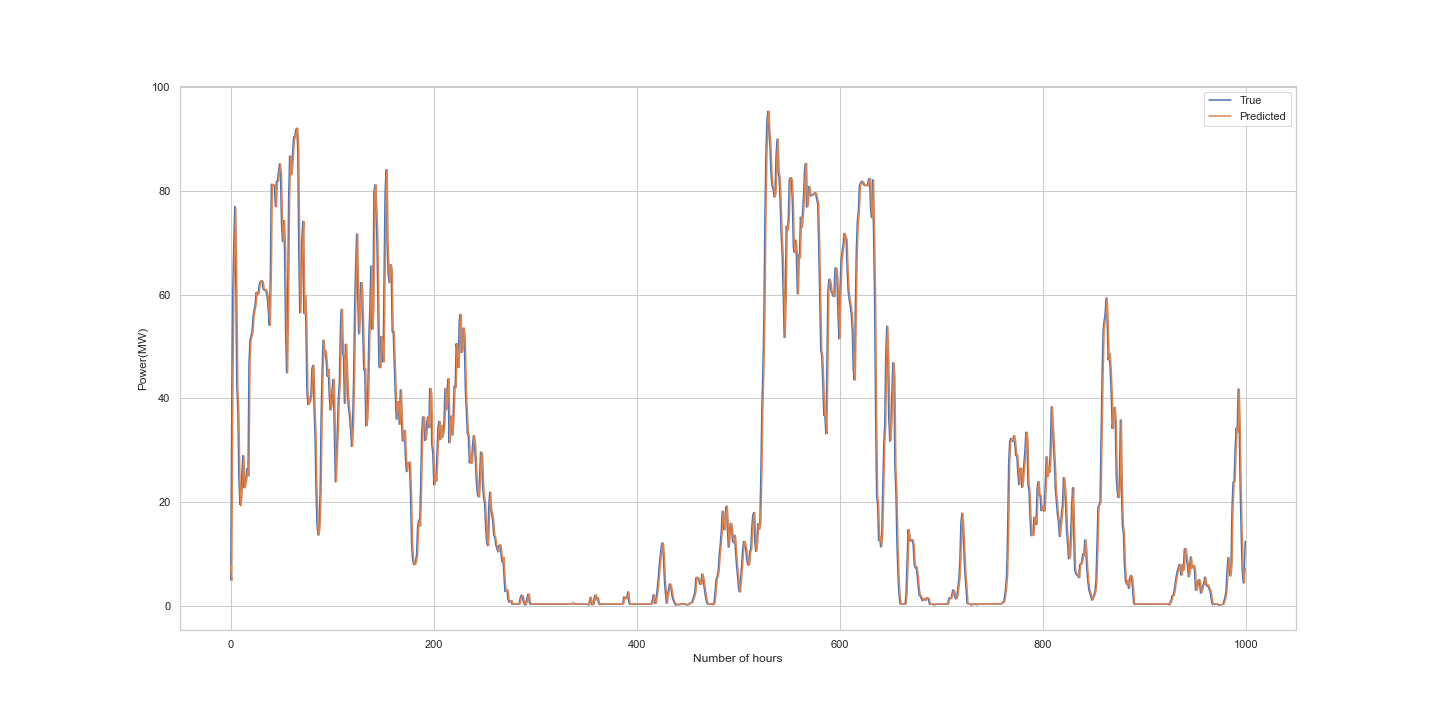


Fig 6. Expected vs Predicted values of Power by using LSTM for Wind Farm 1

**XGBoost**

XGBoost is an implementation of Gradient Boosted decision trees. In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined prediction problems.

**MATHMATICS**

The objective function (loss function and regularization) at iteration t that we need to minimize is the following:

**

where

Simplest linear approximation of a function f(x) using Taylor’s theorem:

**

If we decide to take the second-order Taylor approximation, we have:

**

**

Where:

** and **

Finally, if we remove the constant parts, we have the following simplified objective to minimize at step t:

**

The above is a sum of simple quadratic functions of one variable and can be minimized by using known techniques, so our next goal is to find a learner that minimizes the loss function at iteration t.

* *

The tree learner structure q scoring function

**

Log loss objective function:

**

where **

**Results of XGBoost**

The Data was split into Training (80%) and Testing Data (20%).

The variables TSI, DNI, GHI, Air\_T, Air\_P, Air\_H were used to predict the target variable: Power(MW)

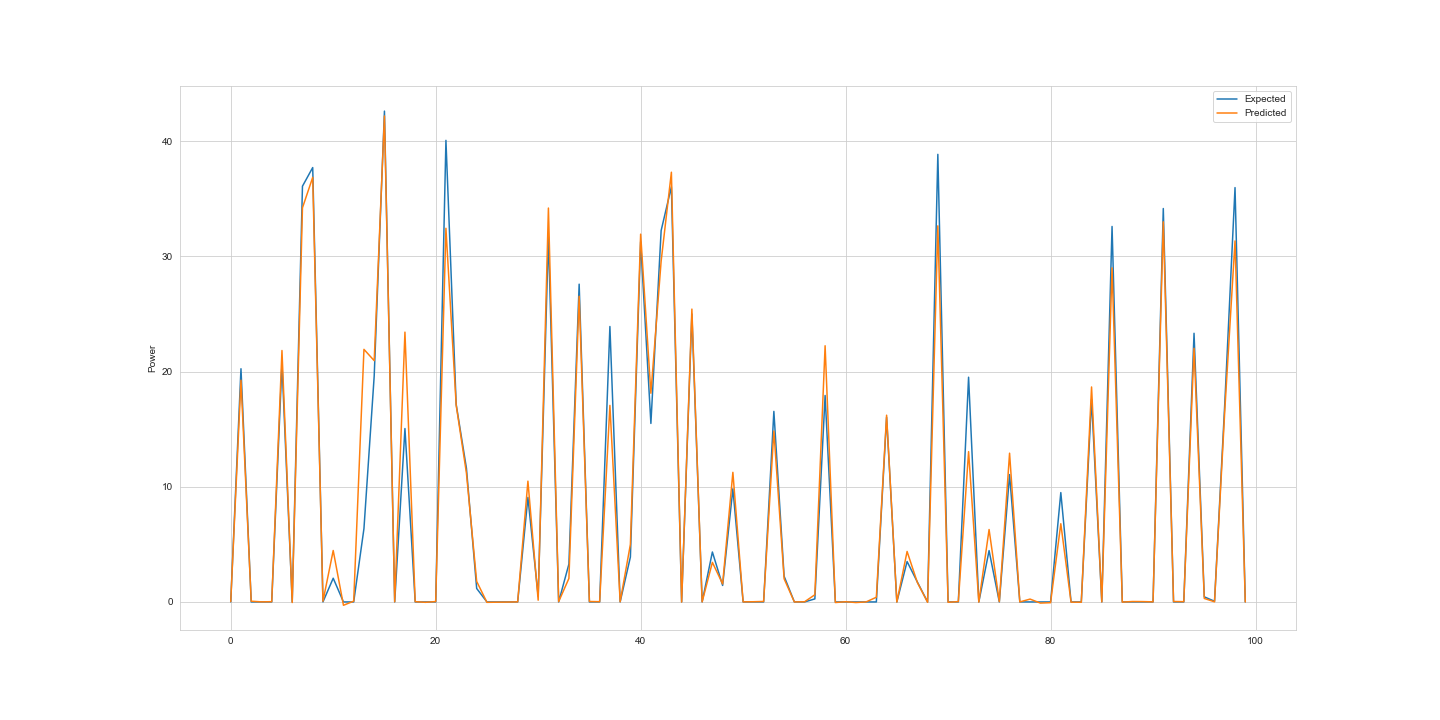


Fig 7. Expected vs Predicted values of Power by using XGBoost for Solar Station 1

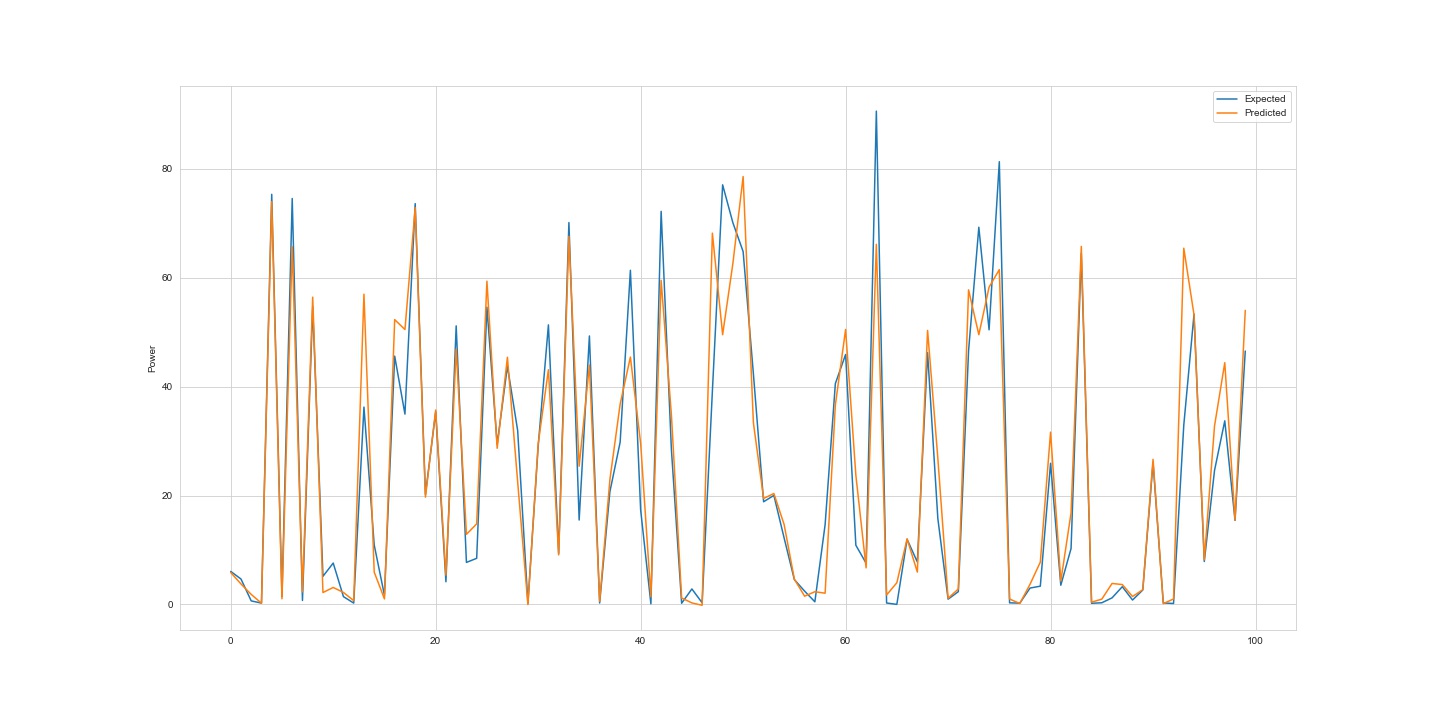


Fig 8. Expected vs Predicted values of Power by using XGBoost for Wind Farm 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Error metrics | Solar Station 1 XGBoost | Solar Station 1 LSTM | Wind Farm 1 XGBoost | Wind Farm 1 LSTM |
| MAE | 1.39 | 2.753 | 5.287 | 2.723 |
| MAPE | 1.74\*1014 | 6.387 | 1.78\*1014 | 0.251 |
| MSE | 9.80 | 12.754 | 77.933 | 22.10 |
| SMAPE | 111.19 | 32.597 | 46.927 | 20.70 |

Table 7. Error Metric for the Results obtained.

All the code and processing scripts used to produce the results of this paper were written in Python, Jupyterlab. Links to scripts and data for analysis can be found in the GitHub repository(<https://github.com/retrophozeac/BITSProject>)

**REFRENCES**

* https://www.youtube.com/watch?v=bDBE2A\_6VWw
* https://www.youtube.com/watch?v=WNZymwInLcY
* <https://medium.com/@divyanshu132/lstm-and-its-equations-5ee9246d04af>
* <https://dimleve.medium.com/xgboost-mathematics-explained-58262530904a>
* Chen, Yongbao, and Junjie Xu. "Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition." Scientific Data 9.1 (2022): 1-12.
* Sanjari, Mohammad Javad, Hoay Beng Gooi, and Nirmal-Kumar C. Nair. "Power generation forecast of hybrid PV–wind system." IEEE Transactions on Sustainable Energy 11.2 (2019): 703-712.
* Quan, Duong Minh, et al. "Hybrid model for hourly forecast of photovoltaic and wind power." 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, 2013.
* Alkhayat, Ghadah, and Rashid Mehmood. "A review and taxonomy of wind and solar energy forecasting methods based on deep learning." Energy and AI 4 (2021): 100060.
* Qadir, Zakria, et al. "Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids." Energy Reports 7 (2021): 8465-8475.
* Yadav, Harendra Kumar, Yash Pal, and Madan Mohan Tripathi. "24-Hour ahead PV power forecasting based on the univariate hybrid machine learning model." International Journal of Ambient Energy (2022): 1-11.
* Larson, David P., Lukas Nonnenmacher, and Carlos FM Coimbra. "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest." Renewable Energy 91 (2016): 11-20.