Solar Power Forecasting

Saurav Dharmeshbhai Patel, Vaibhav Dipesh Parikh, Dr. Jinan Fiaidhi

*Abstract*—The climate change is forcing government to shift there attention to the renewable resources like solar power. However, solar panels do not produce constant electricity throughout the day. Due to that accurate prediction of the amount of electricity generated in the future, it saves a lot of resources and reduces electricity waste. This paper provides a novel approach for short-term solar power forecasting to help solar power energy to integrate with the grid and to manage fluctuation. This will help in the smart integration of electricity. We have tried different algorithms, and later on we applied Long Short Term Memory(LSTM) for solar power forecasting. Long Short-Term Memory captures the temporal features and the non-dynamic features. This research paper will provide some interesting insights and into which way further research can help in the improvement of accuracy.

*Index Terms - Solar Power Forecasting,Long Short Term Memory (LSTM), Auto Regressive Integrated Moving Average(ARMIA),* *Artificial Neural Network (ANN),* Deep Belief Networks (DBN), *Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R2 Score*

# INTRODUCTION

G

lobal focus on solar power generation and increasing renewable energy has dramatically increased in the last few years to fulfill the climate change target. Another reason solar power forecasting has become a leading trend is the reduction in the cost of the basic materials required for solar power forecasting. Consumers are also willing to shift to solar power because of improved efficiency. Many countries have set goals to increase electricity production using renewable resources. India's solar capacity increased from 6,763 megawatts in 2016 to 6,6781 megawatts in 2023 [12]. The largest producer of solar power, which is China, has increased its solar capacity three times in just four years. It had 77,420 MWatt in 2016 and 253,430 in 2020 [11]. Both are the leading countries in solar power forecasting and doing projects on a vast scale. Many developed countries have set targets to achieve a particular. The percentage of energy generation from solar power forecasting as a primary energy source has increased.

As the importance of solar power has its adoption, it has increased, but this has introduced a new set of problems. This energy generated needs to be connected to the grid, and the grid should be ready to handle fluctuation from the solar grid. Short-term solar forecasting improves and makes the system ready for fluctuations and increases energy efficiency. Due to climate change, the weather is more chaotic and unpredictable, and on a large scale, the fluctuation of the current is very high. This paper has considered many features like Albedo, Global irradiance, Direct irradiance, cloud cover, wind speed, humidity, season, time, solar radiance, lower wind speed, upper wind speed, snowfall dept, sun position, etc. This paper accounts for past research papers and works on deep learning models to improve efficiency. This model will help increase the accuracy of solar power forecasting and suggest further works where improvement can take place.

# Related Work

Solar power forecasting is an ongoing research process. The time series algorithm has performed much better than any other model in accuracy. Some algorithms are Auto Regressive Integrated Moving Average ARIMA model (Support Vector Machine) SVM, Random Forest, etc. In deep learning, there are also research papers that use Multi-Layer Perceptron (MLP), Artificial Neural Network (ANN) and Deep Belief Networks (DBN) [15][16][17]. Meanwhile, the Multilayer perceptron and artificial neural network struggled with storing and using the seasonal data. MLP also struggles with temporal data. Arima Model is designed for the seasonal data storage, but it has limited memory and needs to handle other exogenous data well.

# Importance of SHORT-TERM solar power forecasting

There 4 types of solar power forecasting.

1)For the long term, there are already good algorithms present, and they are already being used for forecasting.

2) For the weekly Solar forecasting, the accuracy of the data decreases, which in turn affects the outcome. Many features are required, like wind speed, topology, solar irradiance, snow cover and, etc. Our algorithm will work for them, but accuracy will not be as good in the short term.

3) For the Short Term, which is usually of 3-4 days. Which will help predict very accurate results and help power companies, agriculture, and other companies make decisions much better.

# LSTM Model

Many models are available for forecasting, but choosing a model that captures the trends and sequences is difficult. The LSTM model is a perfect candidate for forecasting as it is able to capture long-term dependencies.

LSTM model stands for Long Short-Term Memory. LSTM is a widely used deep learning model. The LSTM model is able to capture long-term dependencies, which makes it suitable for the prediction of sequential tasks [1]. The LSTM model holds the information for a longer period of time [2].

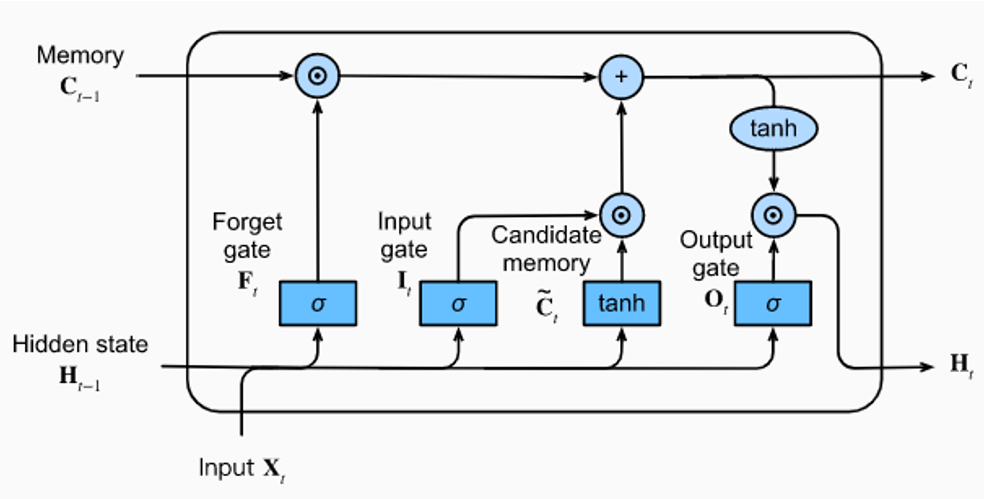


Fig 1. LSTM Architecture [3]

A memory cell in the LSTM retains the information and gates are responsible for performing operations related to memory [2]. During each time instance, the update, forgot, or retrieve operation of the memory cell Ct takes place according to the input gate, forgot gate and output gate. There are three gates in LSTM: 1) Forget Gate, 2) Input Gate, and 3) Output Gate.

### Forget Gate

The information that is not needed is removed by the forget gate. Input Xt at the current time and previous hidden state Ht-1 are added and then the sigmoid function is applied if the output of the sigmoid function is 1 then it retains the output. On the other hand, if the value of the sigmoid function is 0 then it forgets the value [3].

(1)

Where Wf is the weight matrix of the previous cell output, ht-1 is the previous hidden state Xt is the current input sigmoid function, and bf is the bias.

### Input Gate

The new information is added to the cell state in the input gate. The input of the input gate is Hi-1 and Xt [2]. The candidate memory also has the same input as the input gate. Candidate memory generates the candidate that is to be added to the cell state. Candidate memory gives the output using the tangent function. The output of the input gate is a selector vector and the output of the candidate memory is a candidate vector. The candidate vector and selector vector are then multiplied. The cell state vector is added with the output of the multiplication operation between the candidate vector and selector vector [3].

where, Wi is the weight matrix of the previous cell output, ht-1 is the previous hidden state Xt is the current input sigmoid function, and bf is the bias. is the candidate memory. Ct is the candidate memory output.

### Output Gate

The output gate is responsible for providing the hidden state of the network. The output gate determines which cell state is to be given as output as a new hidden state. The sigmoid function is used in the output to decide which cell state to be given as output [4].

(5)

where, Wo is the weight matrix of the previous cell output, ht-1 is the previous hidden state Xt is the current input sigmoid function, and bo is the bias.

Afterwards, the cell state is given as input to the tanh function. The tanh function's output and the sigmoid function's output are then multiplied to generate the hidden state, which can be the input to next layer or final layer [4].

# Implementation

## Dataset

In our implementation we have used German solar power plant data. The dataset is of 21 solar power plant facilities located around Germany. The capacity of these solar power plants is 100kW and 8500kW. The data is recorded at every 3 hours for 990 days. Using the min max normalization technique all the features are normalized between 0 and 1 except power\_normed. The target variable power\_normed is normalized by taking the nominal output capacity of each solar power plant [13].

## Feature Correlation

There are different features available in the dataset. Plotting the graph of each and every feature with a target variable is a very difficult task, so in order to know which features are positively correlated and negatively correlated, the feature correlation method comes into play.

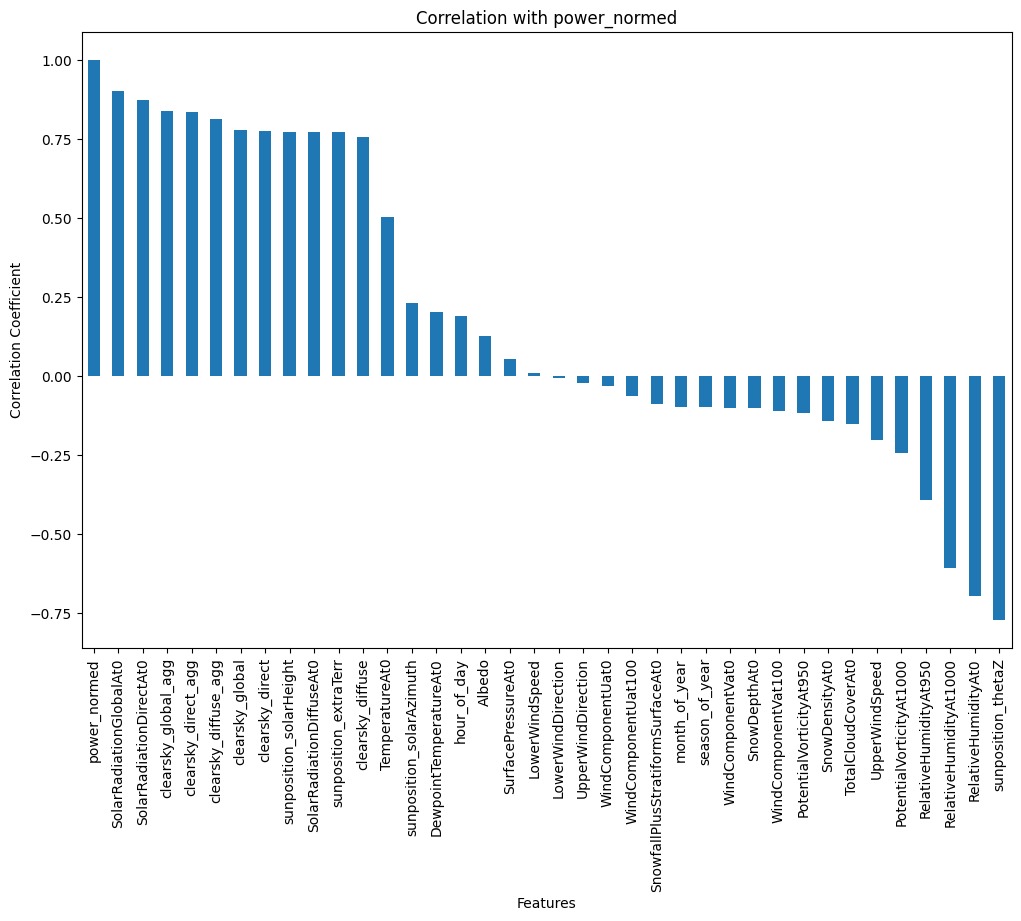


Fig 2. Feature Correlation

Fig 2. shows the correlation graph of the different features with power generation. From this, it can be inferred that Solar Radiation Global, Solar Radiation direct, clear sky global aggregate, clear sky direct aggregate, clear sky diffuse aggregate, sun position, Solar radiation diffuse, temperature and solar azimuth are positively correlated. It can also be inferred that the sun's position theta z, relative humidity, Snow density, and cloud cover are negatively correlated with each other. Furthermore, the remaining features do not affect much power generation. This gives an estimate of how the power will be generated and how the weight distribution will take place. From the correlation graph, it can be noticed that solar radiation global and solar radiation direct have the highest positive correlation with power generated, which is 0.86 and 0.84, respectively. In contrast, sun position theta z and relative humidity are negatively related to power generated, which is -0.75 and -0.68, respectively.

This can also be verified with the help of the Spearman correlation formula. The Spearman correlation can be inferred as how good the relationship is between two variable [10]. The correlation between solar radiation global can be reconfirmed by using the Spearman correlation. The formula for the calculation of the Spearman correlation is following.

The Spearmen correlation between power output and the sun position thetaZ can be shown in the fig 2.

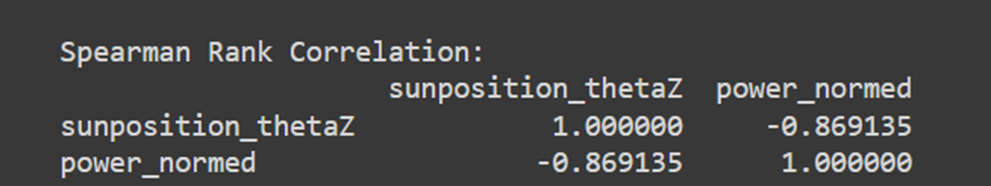


Fig 3. Spearman Rank Correlation

In the fig 3. it is seen that sunposition\_theta Z has negative correlation with the power\_normed.

The Kendall Rank correlation coefficient was used to find the strength and direction of relation [4]. The formula for it is as follows.

The Kendall Rank Correlation Coefficient of solar radiation global and power normalized is displayed at fig 4.

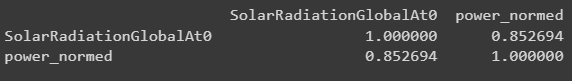


Fig 4. Kendall Rank Correlation Coefficient of solar radiation global and power normalized

It is seen that solar radiation global and power normalized are highly related with each other from fig 4.

The Kendall Rank Correlation Coefficient of solar radiation direct and power generated is displayed at fig 5.

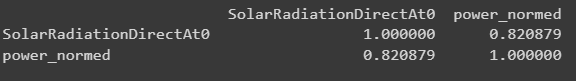


Fig 5. Kendall Rank Coefficient Correlation of solar radiation direct and power normalized

It is seen that solar radiation direct and power normalized are highly related with each other from fig 5.

The Kendall Rank Correlation Coefficient of sun position theta Z and power generated is displayed at fig 6.

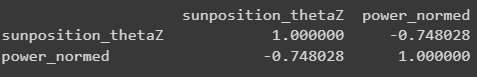


Fig 6. Kendall Rank Correlation Coefficient of sun position theta Z and power generated

It is seen that sun position theta Z and power normalized are inversely related with each other from fig 6.

The Kendall Rank Correlation Coefficient of Relative Humidity and power generated is displayed at fig 7.

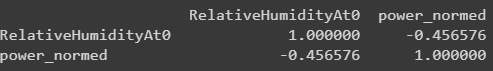


Fig 7. Kendall Rank Correlation Coefficient of relative humidity and power generated

It is seen that Relative Humidity and power normalized are inversely related with each other from fig 7.

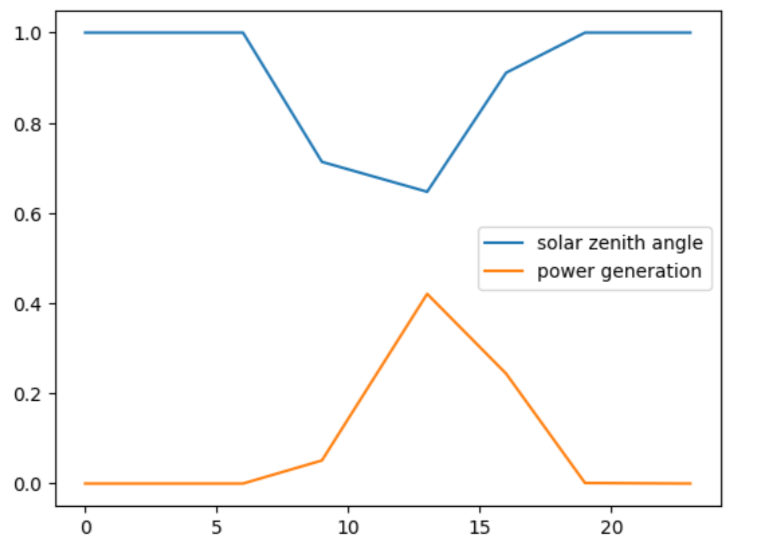


Fig 8. Solar Zenith Angle vs Power Generation

The fig 8 is the graph of solar zenith angle vs power generation during 24-hour period. From the fig 4. it can be seen that during the morning time solar zenith angle value increases and during the after-noon time when the sun is at overhead position solar zenith angle value decreases after that when the sun position gets lower in the sky during the evening time sun position zenith angle value increase. The power generation is less during morning hours because sun position is lower in the sky and as a result the solar zenith angle value increase. During the afternoon hours, power generation is at its peak because the sun is at the overhead position as a result solar zenith angle value decrease. At last, during the evening time when sun position gets lower in the sky power generation also decreases but on the other hand solar zenith angle increase. Solar zenith angle and power generation are negatively correlated.

## Training

For training the model dataset is divided into training and testing. Training dataset includes the 80% of the data and the testing dataset includes 20% of the data.

The LSTM model requires sequential data so converting the dataset into sequential format is a crucial part of the training. Our focus is on the short-term solar power forecasting so in order to do that it is important to choose the number of past hours to predict the next hour solar power generation. Choosing the number of past hours is a difficult task so trial and error method is used to select the past hours. For training 24 past data is chosen means 24\*3= 72 hours is considered to predict the next hour power generation.

## Result

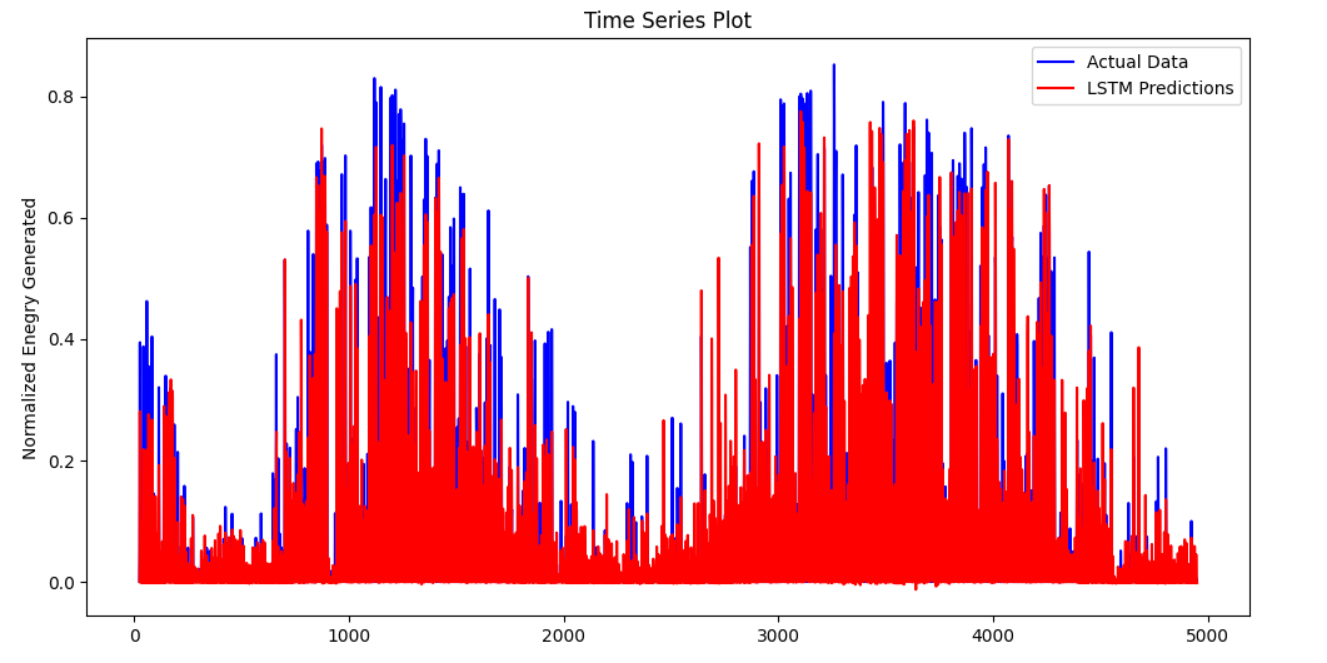


Fig 9. Actual vs Predicted Solar Power Forecasting

From the fig 9. it can be seen that LSTM model was successful to predict.

# Performance Evaluation

Many kind of evaluation metrics are available to evaluate the performance of solar power forecasting model it necessary to select the standard evaluation metrics to measure the accuracy of the model.

### Mean Square Error

Mean Square Error is an evaluation metric that is used to asses the effectiveness of the regression model. “In Statistics, Mean Squared Error (MSE) is defined as Mean or Average of the square of the difference between actual and estimated values.” [5].

The MSE of the trained model is 0.004905.

### Root Mean Square Error

“The sum of the squared differences between the predicted and observed values is divided by the number of observations, and the square root of the result is taken to yield the RMSE” [6].

The RMSE score of trained models is 0.0052

### Mean Absolute Error

Mean Absolute Error is the error between actual value and the predicted value. The advantage of using the MAE is that it is robust to outliers [7].

The MAE of the trained model is 0.0342.

### R2 score

R2 score means coefficient of determination. It determines the percentage of the variation in the dependent variable which can be predicted from the independent variable(s). It indicates how well model fits the dataset [9]. ”Best possible score is 1.0, lower values are worse” [8].

The R2 score of the trained model is 0.764.

Where Yi and Yi’ is the prediction value and observation value and n is the number of datapoints.

# Conclusion

The LSTM has captured the trend and predicted it according to the provided conditions while considering past conditions as well. However, many times, the predicted output is less than the outcome. Changing of the length of the sequential data and changing the model size won't affect the much in the accuracy. From this research use of hybrid model might improve the prediction. Deep belief network for the feature extraction and LSTM for the processing of the network might not have good outcome because deep belief network affects the temporal data. Moreover, Transformers and in that specifically BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) has showcased extraordinary results. The three main benefits of this algorithms that it can have sequential data, understand context and lastly understanding of the complex weather pattern.

References

1. Saxena, S. (2023, October 25). What is LSTM? Introduction to Long Short-Term Memory. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/#LSTM_Architecture>
2. GeeksforGeeks. (2023b, May 31). Deep Learning introduction to long short term memory. <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>
3. Calzone, O. (2022, April 10). An Intuitive explanation of LSTM - Ottavio Calzone - medium. Medium. <https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c>
4. Understanding LSTM Networks -- colah’s blog. (n.d.). <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
5. Great Learning Team. (2022, November 18). Mean squared Error: Definition, applications and examples. Great Learning Blog: Free Resources What Matters to Shape Your Career! <https://www.mygreatlearning.com/blog/mean-square-error-explained/>
6. What is Root Mean Square Error (RMSE). (2023, August 14). Deepchecks. <https://deepchecks.com/glossary/root-mean-square-error/>
7. Agrawal, R. (2023, October 9). Know the best evaluation metrics for your regression model ! Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/#h-mean-absolute-error-mae>
8. sklearn.metrics.explained\_variance\_score. (n.d.). Scikit-learn. <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.explained_variance_score.html#sklearn.metrics.explained_variance_score>
9. Numeracy, Maths and Statistics - Academic Skills kit. (n.d.). <https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html#:~:text=%C2%AFy)2.-,R%202%20%3D%201%20%E2%88%92%20sum%20squared%20regression%20(SSR)%20total,from%20the%20mean%20all%20squared>.
10. Wikipedia contributors. (2023, November 21). Spearman’s rank correlation coefficient. Wikipedia. <https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient>
11. Wikipedia contributors. (2023b, November 28). Solar power in China. Wikipedia. <https://en.wikipedia.org/wiki/Solar_power_in_China>
12. Wikipedia contributors. (2023b, November 22). Solar power in India. Wikipedia. <https://en.wikipedia.org/wiki/Solar_power_in_India#:~:text=India's%20solar%20power%20installed%20capacity,as%20of%2031%20August%202023>.
13. Gensler, A., Henze, J., Sick, B., & Raabe, N. (2016, October). Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In 2016 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 002858-002865). IEEE.
14. Wikipedia contributors. (2023d, November 29). Kendall rank correlation coefficient. Wikipedia. <https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient>
15. S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne and J. Macfie, "Forecasting of total daily solar energy generation using ARIMA: A case study," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2019, pp. 0114-0119, doi: 10.1109/CCWC.2019.8666481.
16. Çolak, M., Yeşilbudak, M., & Bayındır, R. (2020). Daily photovoltaic power prediction enhanced by hybrid GWO-MLP, ALO-MLP and WOA-MLP models using meteorological information. Energies, 13(4), 901. <https://doi.org/10.3390/en13040901>
17. Hu, W., Zhang, X., Zhu, L., & Li, Z. (2021). Short-Term photovoltaic power prediction based on similar days and improved SOA-DBN model. IEEE Access, 9, 1958–1971. https://doi.org/10.1109/access.2020.3046754