*2021 International Conference on Computational Performance Evaluation (ComPE)*
*North-Eastern Hill University, Shillong, Meghalaya, India. Dec 1-3, 2021*

# The Prediction of Power in Solar Panel using Machine Learning

Umang Garg
*Dept. of CSE*
*Graphic Era Hill University*
Dehradun, India
umangarg@gmail.com

Deepak Kumar Chohan
*Dept. of SOC*
*Graphic Era Hill University*
Dehradun, India
dchauhan.geit@gmail.com

Dinesh C. Dobhal
*Dept. of SOC*
*Graphic Era Deemed to be University*
Dehradun, India
dineshdobhal@gmail.com

2021 International Conference on Computational Performance Evaluation (ComPE) | 978-1-6654-3656-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ComPE53109.2021.9751901

*Abstract*— **The Solar panels are depending on the various parameters like air pollution and environment. The air pollution and bad weather conditions are generated very critical condition for the generation of power from solar panels. The prediction of solar panel's power in advance improves the overall functionality of the solar panels and generated the best results for end-users. In this paper, a prediction of solar power using machine learning algorithms has been done and analysis the impact of air pollution and bad weather on it. A machine learning model is able to generate the best results with the extraction of features like bad environment, air and weather pollutions. The experimental results show the efficient results on the dataset collected from the open-source system.**

***Keywords— Machine Learning, Weather pollution, Bad environmental conditions, Solar Panel, Air Pollution.***

## I. Introduction

We claim in this paper that the current environmental condition is dire. Solar panels are used to generate the solar power which is harvesting from the sun and a cost-effective solution. The major installation of solar panel is conducted by the South Korea which is a great resource of topmost energy resources [1]. In 2019, South Korea contributed 3.1 GW (gigawatt) of electricity [1]. It is a difficulty to fully utilise solar power's potential, given the growing interest in the technology [2]. External factors such as weather and air pollution have a compelling impact for the production of power using solar panels. Solar panels have a hard time performing to their full potential when the weather is bad or the air is polluted. Knowing the power output of solar panels ahead of time can aid in the proper installation of solar panels and ensure that they perform to their full capacity.

The electricity can be generated with the help of solar panels is a critical process. Once the light shines and the panels are not partially obscured, solar panels can obviously operate at their full capability. However, some temporary reasons like weather and air pollution might result in partial shade (e.g., fine dust). Using machine learning techniques, this article proposes an approach for evaluating the output in terms of power which were generated from the different factors. We want to figure out what combination of environmental elements and machine learning techniques is most beneficial. The dataset can be downloaded for free from an open-source website [11].

To construct power output prediction models, this research employs cutting-edge algorithms of machine learning such as Support Vector Regression (SVR), Linear Regression, MultiLayer Perceptron (MLP), K-Nearest Neighbors (kNN), Random Forest Regressor (RF), and Gradient Boosting Regressor (GB). For each and every model, there are mainly three features. Humidity, sunlight, solar radiation, and clouds are among the initial group of meteorological variables. Air pollution includes ozone (O2), nitrogen dioxide (NO2), fine dust (PM10), sulphurous acid gas (SO2), carbon monoxide (CO), and fine particulate matter (FPM) (PM2.6). Further, the top correlated traits from the other two groups are chosen to construct the third set. We used prominent error rate measures including Root Mean Squared Error (RMSE), Coefficient of Determination (R2), Mean Absolute Error (MAE), and to analyze the models. The findings of the experiments suggest that weather and air pollution can be effective predictors of solar panel power output.

The following is how we ordered the paper. The second section contains brief descriptions of research that are similar to ours. Section 3 describe the methodology of the article. The analysis and evaluation using practical implementation is justify in section 4. A discussion of the result and recommendations for future work concludes in section 5.

## II. Related Work

When evaluating the several factors related to environment and pollution must be taken into account [2-9], as we described in Section 1. Several academics, for example, have proposed weather-based prediction models for estimating solar panel power generation [3-5]. There were several past researchers [3] predicts the outputs for solar panels and proposed some accurate models such as support vector regression and multi-input-based models. The output of solar panel which is connected with the grid can be predicted using different machine learning models. Humidity, temperature, pressure, and wind speed were all used by the writers. This study compared the analytical and SVR models, the SVR model is more accurate than other models which are analytical in nature. Artificial neural networks (ANNs) were utilised by Saberian et al. [4] to forecast the temperature and

**978-1-6654-3656-4/210/$31.00 ©2021 IEEE**

other parameters in terms of metrological data. From 2006 to 2010, the authors used a five-year dataset. Experiments have demonstrated that the proposed model outperforms the competition.

Furthermore, several methods for predicting solar panel output power using bad content of air and water parameters exist [6-9]. The impact of particulate matter (PM) for the production of energy by using solar power was investigated by Son et al. [6]. The scientists used PM2.5 and PM10 readings from 2015 to 2017 to create the dataset. According to the authors, PMs frequently lower solar power generation by more than 10%. Their findings demonstrate that the harmful effects of PM on solar panels should be addressed when determining solar power output targets. Bergin et al. [9] calculated the loss of energy generated by the solar power due to air pollution and dust particles. The researchers analyzed the effect of dust and PM for solar electricity output by combining field measurements with global modelling. According to the data, the solar panel's PM level reduces production by 17 to 25%.

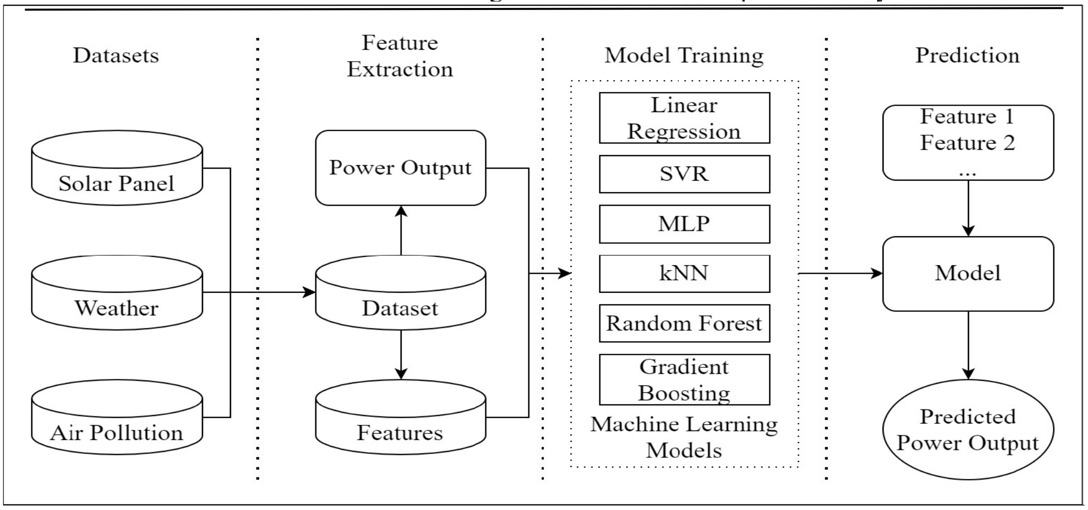


Figure1: Methology in one Frame

## III. Methodology

The overall methodology of the paper is divided into sub-parts:

### A. Overview

The general flow of our methodology is depicted in Figure 1. The collection of data, feature extraction, and train the module are different steps involved in our proposed methodology. The dataset can be obtained from the various online resources and experimental evaluation. After that, we select different sets of features to develop machine learning models. Here, we want to show you which features and machine learning technologies work well together. The subsections that follow explain each stage in detail.

### B. Datasets

We obtain our experimental datasets from three different data sources. Solar panel power output data is given by Daeyeon C&I, a South Korean renewable energy power output firm. All datasets are open-source and available online [11]. The feature description of the gathered data is shown in Table 1. The power output is our objective feature, as we can see from the table. Other features, such as the solar panel, are inputs to our models and are grouped into four areas. Features can be derived from the date and used for evaluation purpose. Based on the date, we integrated all datasets. From 2017 to 2019, there are 14191 observations in the dataset. We divided the dataset into train and test groups of 80 percent and 20 percent, respectively. Our models anticipate the power output over the next one hour.

### C. Feature Extraction

The initial stage after joining the separated datasets is to identify the traits that can be used to measure the output in terms of the power generated by the solar panels. First, we

use derived features (month, hour) to extend each set of features, which can be evaluated using the registered date of output and can be converted as an energy resources. The time is selected when there are less chances of rainfall and more chances of sunlight during the day time.

There is no association between the features if the result is near to zero. The link between power output and other variables is depicted in Figure 3. We constructed five sets of characteristics based on the results shown in Fig. 2. The first group includes solar panel features, the second group includes weather elements, and the third group includes air pollution features. Furthermore, the fourth set consists of the most highly associated features from the previous three sets, whereas the fifth set consists of all features. Several factor have been combined for the extraction of features such as temperature, power factor, solar radiation, and different levels of gas (power factor, slope, horizontal irradiation, module).

## IV. Experimental Analysis

To test model efficacy, a comparison have been take place between solar panel output and the estimated power of solar energy. We have measured several factors such as $R^2$, MAE, and RMSE to evaluate the estimated error and predict different accuracies.

Table 1: Description of Features in Dataset

| Variant | Energy Source | Factor | Details |
|---|---|---|---|
| Dependent (y) | Solar Panel | Power output | Solar panel output in terms of power (kWh) |
| Independent (xi) | | Power factor | It indicates the ration in consume and the originated power. |
| | | Horizontal irradiation | The entire amount of solar radiation that strikes a horizontal surface. |
| | | Module temperature | It indicates the temperature measured by solar panels |
| | Weather | Humidity | the amount of water vapour existing in the atmosphere. |
| | | Sunshine | Clouds do not block sunlight from reaching the earth. |
| | | Solar radiation | The radiation energy reaches to the ground. |
| | | Cloud | The amount of Cloud |
| | Air Pollution | $SO_2$ | Sulfurous Acid Gas |
| | | $PM_{2.5}$ | Fine Particulate Matter |
| | | $O_2$ | Ozone |
| | | $NO_2$ | Nitrogen Dioxide |
| | | $PM_{10}$ | Fine Dust |
| | | CO | Carbon Monoxide |
| | Derived Features | Month | Month |
| | | Hour | Hour |

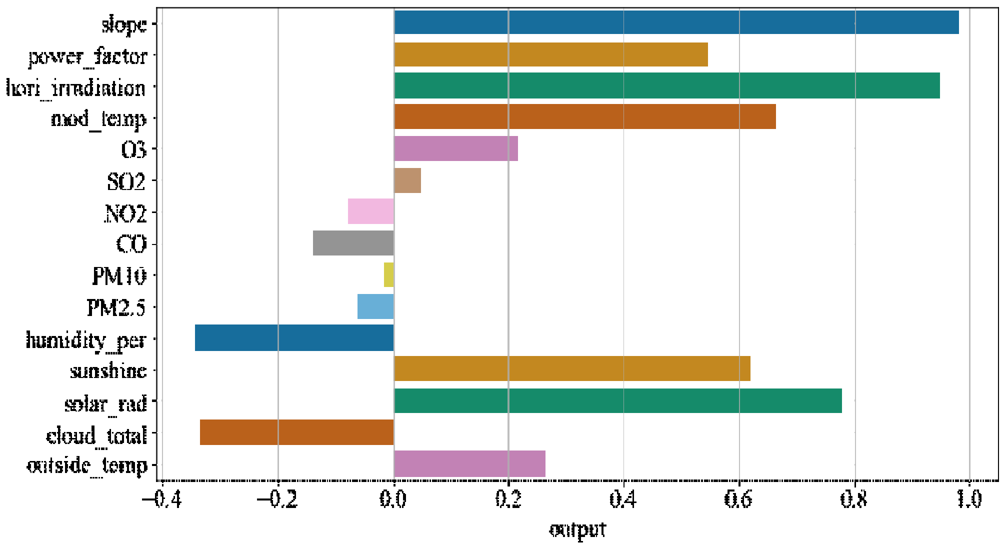

Figure 2: Correlation between Output Values and Features

The average of variations between the actual and anticipated values is calculated using MAE. Lower RMSE and MAE values indicate that the MAE and RMSE are more precise. A regression method is also to evaluate the R2 score. The value of $R^2$ is 1 indicates the best score, however it could be negative, implying inferior accuracy.

*A. Results*

When evaluating model performance, it evaluate the ideal scores for the hyperparameters in a very first step that is used to build an effective model, specifically related to SVR which is sensitive. To identify the optimum option, we ran each model with various different combinations of hyperparameter values. Table 2 shows the values of each model's hyperparameters.

Table 2: Setting of Parameters in Model

| Algorithms | Factors | Range |
|---|---|---|
| SVR | C | 199 |
| | epsilon | 6 |
| MLP | max_iter | 470 |
| | epsilon | 1e-8 |

| | | |
|---|---|---|
| KNN | n_estimators | 16 |
| RF | K value | 16 |
| | max_depth | 12 |
| GB | n_estimators | 100 |
| | learning_rate | 0.11 |

The full results of all models and feature sets are summarised in Table 3. The table shows that GB, MLP, and RF are more precise than Linear Regression, KNN, and SVR. Air pollution controls are also less effective than solar panels and weather patterns. Because slope and horizontal irradiations are significantly connected with solar panel power production, solar panel features perform best with all models. The model used for some correlated features and weather feature are having significant difference between them. It indicates that weather characteristics can be used to anticipate production in terms of power factor. However, air pollution features with RF and GB demonstrate good accuracy, with roughly 67 percent of R2. The most suited model is RF, which has the best accuracy across all feature sets. The RMSE factor is 0.9%, while MAE is 98.25%, and RF score is 0.3% in accordance. Solar panels have a power output range in between from 0 to 25.

The RMSE of all models is shown in Figure 3, given by the h three arrangements of highlights which show the shifting of accuracy of each model. The after effects of RF, GB, and MLP, just as KNN, SVR, and Linear Regression, are tantamount. The most dependable are RF and GB, and no extreme qualifications exist between them. The GB model's is 0.8 percent greater than the RF model's and RF model's is 42.6 percent lower than Linear Regression, SVR is 27.9% greater than RF, 6.6 percent lesser than MLP, and 11.9 percent lesser than KNN models. The predictions of meteorological features, as shown in Table 3 and Fig. 3, are more similar to true values than the predictions of air pollution features.

Table 3: Result Analysis of the Models generated by different algorithms

| **Algorithms** | **Factors** | **$R^2$** | **RMSE** | **MAE** |
|---|---|---|---|---|
| Linear Regression | Solar Panel | 94.17% | 1.50 | 0.59 |
| | Weather | 72.99% | 2.9 | 2.90 |
| | Air Pollution | 30.93% | 5.61 | 4.57 |
| | All | 94.97% | 1.51 | 0.61 |
| MLP | Solar Panel | 97.42% | 0.90 | 0.40 |
| | Weather | 90.1% | 3.01 | 2.01 |
| | Air Pollution | 59.53% | 5.00 | 2.21 |
| | All | 97.89% | 0.98 | 0.52 |
| SVR | Solar Panel | 90.52% | 3.02 | 1.99 |
| | Weather | 81.2% | 3.01 | 2.29 |
| | Air Pollution | 50.67% | 5.87 | 3.99 |
| | All | 92.50% | 2.1 | 1.8 |
| RF | Solar Panel | 97.90% | 1.00 | 0.30 |
| | Weather | 88.02% | 2.40 | 1.40 |
| | Air Pollution | 67.46% | 3.49 | 2.70 |
| | All | 98.2% | 0.9 | 0.29 |
| kNN | Solar Panel | 96.9% | 1.20 | 0.40 |
| | Weather | 82.16% | 3.1 | 1.9 |
| | Air Pollution | 50.20% | 5.20 | 4.1 |
| | All | 97.3% | 2.0 | 0.39 |
| GB | Solar Panel | 97.30% | 1.01 | 0.39 |
| | Weather | 88.30% | 3.20 | 1.5 |
| | Air Pollution | 66.28% | 4.20 | 2.85 |
| | All | 99.10% | 0.90 | 0.4 |

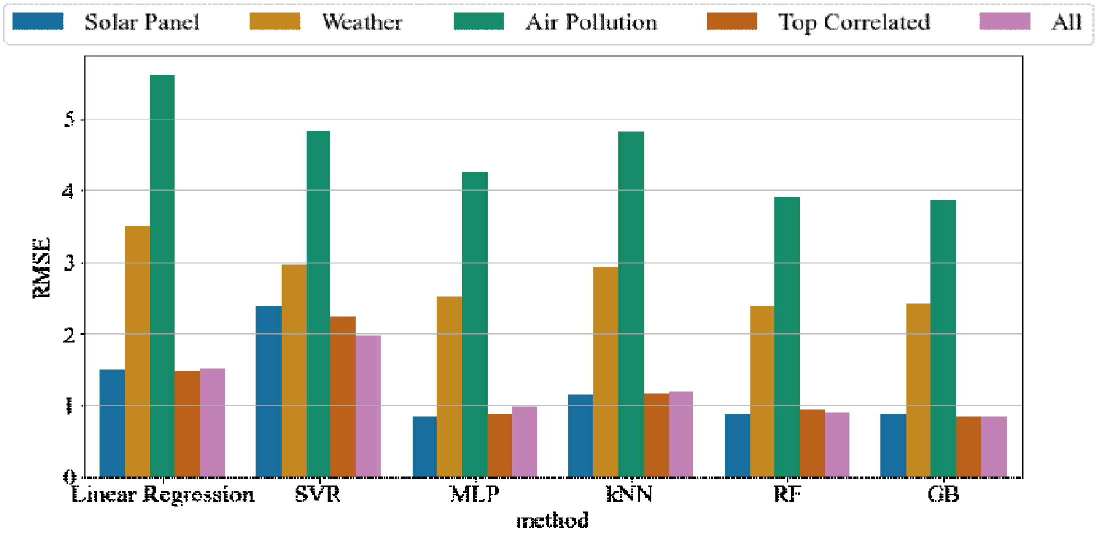

Figure 3: Algorithm RMSE

## V. Conclusion

The purpose of this research is to forecast solar panel output based on weather and pollution levels. Six machine learning algorithms were employed to test five sets of features, including LR, RF, GB, and SVR. We found that most of the algorithms provides more than 95% of the accuracy. Air, weather and other characteristics have also shown favourable results. However, unlike two earlier research, we found that there are some negative outcomes appears from the air attribute. Finally, the best model, the RF model, with a 98 percent accuracy, was established.

When the sun shines directly on the solar panels, they are able to collect a lot of electricity. However, due to environmental factors such as weather or air pollution, the sun does not always reflect straight on the solar panels. As a result, these factors related to the environment perform a key role in the proper location of solar panels. Users of solar panels can also determine when to clean their panels depending on predicted electricity output. Predicting the power production of solar panels ahead of time can have a number of advantages, including altering installation, monitoring functioning, and planning future business plans.

More environmental features, such as wind, temperature, and others, can also be added to improve model accuracy. Outliers in the power output data points may also occur as a result of irregular solar panel functioning. As a result, we can think about reducing outliers to improve forecast accuracy even more. In the future work, some deep learning model can be implemented such as LSTM, XGBoost, and other improve power output forecast accuracy even more.

## References

[1] International Energy Agency, Photovoltaic Power Systems Program:Snapshot of Global PV Markets 2020, Report IEA-PVPS T1- 37, 2020.

[2] J. Ko and H. Kim, “Implementing the Smart System for the Safety of the Elderly and the Weak - Focus on Energy Harvest -,” Journal ofKoreaMultimediaSociety, Vol. 23, No. 12, pp. 1507-1518, 2020.

[3] M. Vakili, S.R. Sabbagh-Yazdi, K. Kalhor, and C. Khosrojerdi, “Using Artificial Neural Networks for Prediction of Global Solar Radiation in Tehran Considering Particulate Matter Air Pollution,” EnergyProcedia, Vol. 74, pp. 1205-1212, 2015.

[4] R. Nageem and R. Jayabarathi, “Predicting the Power Output of a Grid-Connected Solar Panel Using Multi-Input Support Vector Regression,” ProcediaComputerScience, Vol. 115, pp. 723-730, 2017.

[5] S.G. Kim, J.Y. Jung, and M.K. Sim, “A TwoStep Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning,” Sustainability, Vol. 11, article. 1501, 2019.

[6] J. Son, S. Jeong, H. Park, and C. Park, “The Effect of Particulate Matter on Solar Photovoltaic Power Generation over the Republic of Korea,” EnvironmentalResearchLetters, Vol. 15, No. 8, 2020.

[7] Dataset Available at https://www.kaggle.com/anikannal/solar-power-generation-data.

[8] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, “Predicting Solar Generation from Weather Forecasts Using Machine Learning,” IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 2011.

[9] M.H. Bergin, C. Ghoroi, D. Dixit, J.J. Schauer and D.T. Shindell, “Large Reductions in Solar Energy Production Due to Dust and Particulate Air Pollution,” Environmental Science & Technology Letters, Vol. 4, No. 8, pp. 339-344, 2017.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, M. Perrot and E. Duchesnay, “Scikit-learn: Machine Learning in Python,” JournalofMachine LearningResearch, Vol. 12, pp. 2825-2830, 2011.

[11] A. Saberian, H. Hizam, M.A.M. Radzi, M.Z.A.A. Kadir and M. Mirzaei, “Modelling and Prediction of Photovoltaic Power Output Using Artificial Neural Networks,” International JournalofPhotoenergy, Vol. 2014, article. 469701, 2014.