IJDI-ERET

INTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON
ENGINEERING RESEARCH AND EMERGING TECHNOLOGIES
Vol. 11, No. 2, 2022
www.ijdieret.in

(Research Article)

Solar Power Generation Forecasting Using Adaptive Boost and Random Forest

Vidhi Rajeshkumar Patel[1*], Mihir Sanjaykumar Patel[2], Manish Upendrakumar Patel[3], Tejas C. Patel[4], Mihir R. Patel[5]

[1,2,3,4,5]M & V Patel Department of Electrical Engineering, Chandubhai S. Patel Institute of Technology, CHARUSAT- Charotar University of Science & Technology, Changa, Gujarat, INDIA*

Abstract

In the recent few years, the conscious approach towards green energy has gained significant momentum. These green sources, a noteworthy amount of power is generated through them and integrated into the grid. Hence, the need of forecasting these energy sources; especially solar energy, for the most efficient utilization is very crucial. Hence, measuring the accurate solar radiance, precise load management and efficient solar energy usage can be ensured.

This research utilizes the use of ML algorithms; through which hourly, daily and monthly prediction of solar generation is possible. Concurrently, shift towards smart grid concept and solar energy by Indian power system will be benefited by use ML algorithms, that will fulfill an important aspect; load forecasting. This in turn will assure an economic operation as well as planning of the power system. In addition, sort term planning along with future expansion of generation can be pre-determined with high accuracy.

Keywords: Energy, Economic, ML algorithms, Solar, Adaptive Boost, Random Forest.

1. Introduction

Globally, the shift towards clean energy in a sustainable manner is a challenging task. It is predicted that global energy need will double in the upcoming 25 years; and latter double again in the upcoming 50 years. Load forecasting plays a vital role in ensuring reliable and adequate power supply to the end users. Load forecasting is defined as a procedure used for predicting the future electricity demand using historical data to be able to manage the electric generation and electric demand of electric utilities [5]. In the present scenario the load forecasting is essential task in smart grid. The smart grid is an electrical grid which uses computers, digital technologies, or other advance technologies for real time monitoring, maintaining generation and demand, also to act on particular information (information such as behavior of electric utilities or consumers) for improving efficiency, reliability, sustainability, and economic [1].It is evident that power system will be surfaced/disrupt with instability/stability as solar energy is integrated into; as it is not a constant source of energy and its out varies accounting of many aspects. The output to solar energy varies due environmental conditions, installation site geographical coordinates. This research describes the use of Adaptive Boost and Random Forest model of machine learning to predict the solar generation with and without solar radiation for weekly data analysis.

Solar Energy based power generation; which is subject to be influenced by various factors. This research utilizes the use of Adaptive Boost and Random Forest model to predict the solar generation, allowing better energy demand planning. Solar generation data of past 4 years is stored in one sheet taking the base of which another sheet is generated, containing values of generation that are predicted in terms of week. The Adaptive Boost method calculation fabricates a model and gives equivalent loads to every one of the data of interest. It then allocates higher loads to focuses that are wrongly classified. Presently every one of the focuses which have higher loads are given more significance in the following model. It will continue to prepare models until and except if a lower mistake is derived [7]. Random forest is a classifier that uses a variety of choice trees on various subsets of the provided dataset and uses the normal to operate on the dataset's predictive accuracy. The raw data used in this paper is available through smart grid competition-forecast, Portugal [8].

*[*]Corresponding Author: e-mail: vidhipatel.vp3011@gmail.com
Tel-+91-9898207527*
ISSN 2320-7590 (Print) 2583-3863 (Online)

This research paper flow contains four sections. Introduction in first section, following which methods used in prediction are described in second section. Third section is dedicated to results and comparison between both the methods [6]. At last, final section elicits the conclusion, further prospects of the research and references.

2. Methods Used In the Research of Solar Radiance

2.1 Adaptive boost: There are numerous machine learning algorithms accessible right now. The use of ensemble learning in practical settings is widespread. In particular, almost all high-scoring teams use ensemble learning while playing. Ensemble learning can often do better than any other single learners, especially single "poor learners," because it involves several learners. A learner who performs poorly is referred to as a weak learner. [2]

We use a voting approach for ensemble learning, in which more than half of the base learners' results are taken as the classification results in the case of a binary classification problem with the parameters yyyy {-1, +1}, a real function f, and an individual learner (or base learner) hii whose m (odd number) error probabilities are independent.[2]

$$H(x) = sign(\sum_{i=1}^{M} h_i(x)) \cdots\cdots\cdots\cdots\cdots\cdots \quad (1)$$

$$P(H(x) \neq f(x)) \leq \exp(-\frac{1}{2} M (1 - 2 \in)^2) \cdots\cdots \quad (2)$$

After ensemble learning, the error probability for a large M of base learners with an independent error probability is near to 0, which is consistent with the following core belief: It's unlikely that the majority of people will commit mistakes at the same time. The premise behind each of the aforementioned assumptions is that base learners make mistakes on their own. In actual applications, these learners can't be independent of one another. In this regard, the fundamental problem to be solved for assemble learning is how to make base learners "relatively autonomous" (i.e., an increase in the diversity of base learners). [2]

Based on whether there is a dependency between base learners, there are two forms of assembly learning.

Boosting is the representative algorithm when there is a strong dependency between base learners and when there is no strong dependency, a series of base learners can be generated in parallel. Bagging and Random Forest are the representative algorithms when there is no strong dependency between base learners. [2]

When a series of base learners can be generated in parallel, an ensemble technique called "boosting" is utilised to transform a weak base learner into a strong learner. [2]

Following is a description of the steps:

- First, the training set's initial base learner is trained using it, with each sample's weight being equal.
- Using the learner obtained in the previous round of training, the weight of samples in the training set is adjusted based on the performance of the training set forecast (for example, an increase in the weight of misclassified samples causes the samples to receive an increasing amount of attention in the next round of training), and a new base learner is trained on this basis;
- Repeat Step 2 until you have m basis learners, and the ultimate ensemble result is the m base learners combined.

Boosting is a sequential procedure as a result. The Adaptive Boost boosting technique is the most widely used boosting clustering algorithm.

2.1.1 Adaptive boost principle: Using the linear weight sum approach, the weight of the samples that were incorrectly identified during the previous round of training is increased while the weight of the samples that were correctly classified is decreased. The weight of based learners with a low error rate is higher, whereas the weight of based learners with a high error rate is lower.

The Adaptive Boost structure is depicted in the following figure.

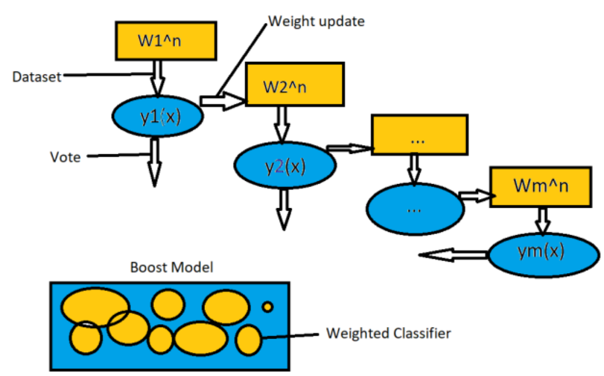

Figure 1. Adaptive boost structure

2.2 Random forest: Due to its simplicity, RF is one of the most used machine learning techniques. It can be applied to classification and regression problems. It belongs to the same group as the naive Bayes algorithm and other tree-based algorithms like Adaptive boost, which is the class of supervised learning algorithms and SVMs [3].

In general, the more trees in the ensemble, the more reliable the findings are. RF is made up of integrated decision trees that produce a more accurate and consistent forecast. Instead of splitting a node, this model seeks out the best feature from a random group of features, increasing the unpredictability of the decision tree model and growing the trees [3].

Out-of-bag error, or OOBE, also known as the generalisation error, is a type of cross-validation that is already included into the algorithm and refers to the average prediction error of the first observations. It is a key feature used in this project. The OOBE is used to calculate how generalizable the algorithm is [3].

$$OOBE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Y)^2 \text{ .... (3)}$$

The variable importance measure (VI), the last characteristic, is created by permuting a feature and summing the difference between the OOBE before and after the permutation over all the trees. [3]

$$VI(X^j) = \frac{1}{q}\sum_{i=1}^{N}(\overline{OOBE_l} - OOBE_l) \text{ .... (4)}$$

Where the estimated OOBE is the average of those estimates. The general architecture of the RF algorithm's functioning is shown in figure.

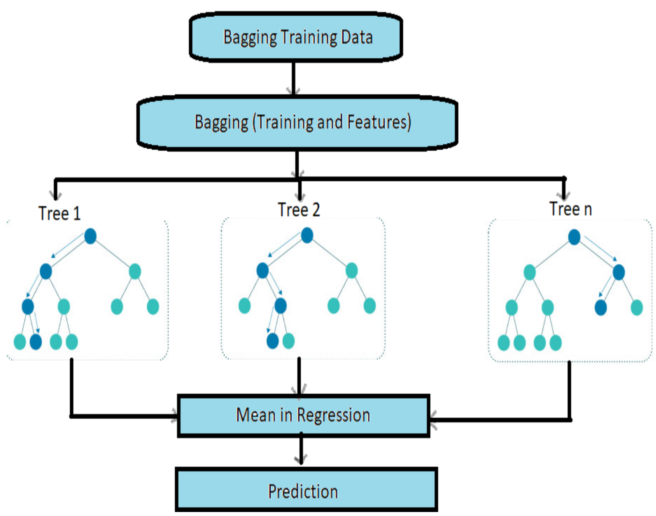

Figure 2. Random forest structure

3. Results and Comparison between Both Methods

This research paper will describe the accuracy of the results through:

- The Root Mean Square Error (RMSE). RMSE is nothing but the square root of the calculated Mean Square Error (MSE). It means the average magnitude of the error. Here, the reason for taking root is to error same as the actual value unit.

$$MSE= \frac{\sum(\text{Actual - Forecast})^2}{n} \text{ .... (5)}$$

$$RMSE= \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \text{ .... (6)}$$

Where Pi is the predicted values and Oi is the observed or actual values. n is the length of the forecasted values.

- The Mean Absolute Percentage Error (MAPE). It is a statistical measure of how accurate a forecast model/method is. The MAPE is the simple average of absolute percentage errors and the measure is in relative percentage. The MAPE is not affected by the unit of actual values. Also, the closer the value of MAPE to zero, the better the prediction results. This measure is given by

$$MAPE: \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| *100 \text{ .... (7)}$$

Where, at is the actual values and Ft is the forecasted values, and n is the length of the forecasted values.

3.1 Implementation and results: The results in both these methods have common implementation. In the very first stage the raw data of the solar generation is collected and distinguished through various factor affecting generation. This raw data is fed to the algorithm through two different sets. First, testing data set which trains the algorithm and second one being training data set; in which algorithm acts through results achieved by performing calculations on testing data. As both the methods may encounter error; the algorithm used by both the methods tries to minimise the error to the best possible extent.

3.1.1 Results of both the methods with RMSE and MAPE error when solar generation is zero (Counting in night hours and unpredictable weather condition):

Table 1. Results of accuracy of both the methods when solar generation is zero (counting in night hours and unpredictable weather condition)

Methods	RMSE	MAPE
Adaptive Boost	1171.4293	6.8456
Random Forest	1389.8614	9.8585

3.1.2 Results of both the methods with RMSE and MAPE errors when there is energy generate through solar only:

Table 2. Results of accuracy of both the methods when there is energy generate through solar

Methods	RMSE	MAPE
Adaptive Boost	2331.766	12.6518
Random Forest	2671.307	15.4923

3.2 Graphs:

3.2.1 Accuracy of both the methods when solar generation is zero (Counting in night hours and unpredictable weather condition):

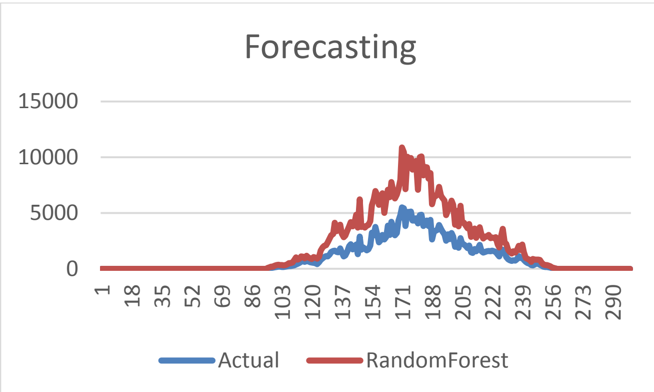

Figure 3. Predicted vs actual solar power generation through the day with random forest

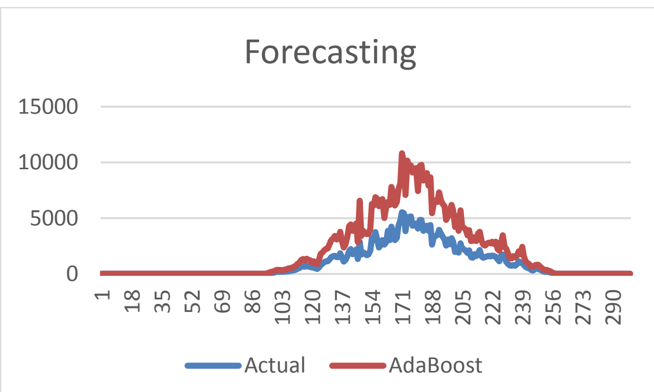

Figure 4. Predicted vs actual solar power generation through the day with adaptive boost

3.2.2 Results of accuracy of both the methods when there is energy generate through solar:

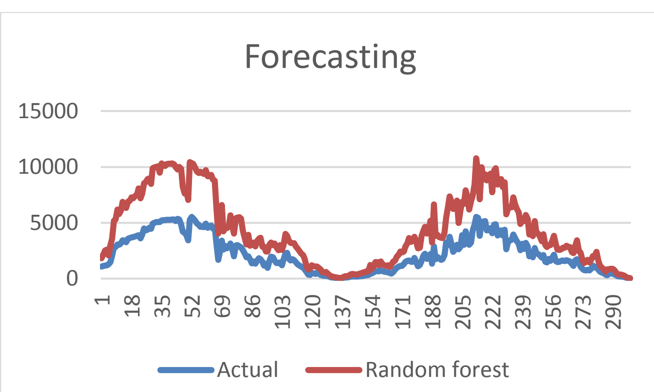

Figure 5. Predicted vs actual solar power generation during peak hours with random forest

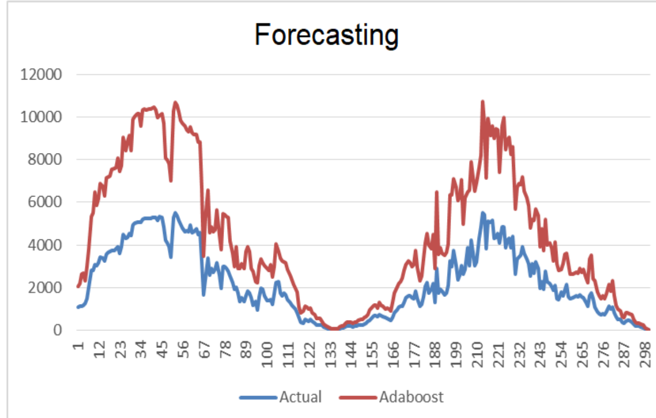

Figure 6. Predicted vs actual solar power generation during peak hours with adaptive boost

4. Conclusions

From this paper, it was clear that random forest has better prediction in comparison with Adaptive Boost. The error correction in both the methods is commendable but random forest still has an edge over Adaptive Boost since the level of error rectification it offers is extensively accurate. Hence this algorithm can be of great help for the expansion of solar system and for load forecasting.

References

1. Corentin Kuster, Yacine Rezgui, Monjur Mourshed, "Electrical load forecasting models: A critical systematic review" Sustainable Cities and Society, August-2017
2. Adaptive Boost-Based Power System Load Forecasting WuNeng Ling *et al, J. Phys.: Conf. Ser.* **2005** 012190 2021
3. Optimized Random Forest for Solar Radiation Prediction UsingSunshine Hours Cesar G. Villegas-Mier 1,† , Juvenal Rodriguez-Resendiz 2,*,† , José Manuel Álvarez-Alvarado 2,† ,Hugo Jiménez-Hernández 1,† and Ákos Odry 3,†
4. Rikin Patel, Mihir R Patel, Ravi V. Patel, "A Review: Introduction and Understanding of Load Forecasting", Journal of Applied Science and Computations (JASC), pp. 1449-1457, 2019.
5. Mihir R. Patel, Ravi Patel, Dharmesh Dabhi, Jignesh Patel, "Long Term Electrical Load Forecasting considering temperature effect using MultiLayer Perceptron Neural Network and k-Nearest Neighbor algorithms" International Journal of Research in Electronics and Computer Engineering (IJRECE), Volume 7, Issue 2, pp. 823-827, April- June 2019.
6. Rikinkumar B. Patel, Mihir R. Patel, and Dr. Nilaykumar A. Patel, "Electrical load forecasting using machine learning methods, RNN and LSTM," in Journal of Xidian University, Volume 14, Issue 4, pp. 1376-1386, April 2020.

7. http://www.gecad.isep.ipp.pt/smartgridcompetition-forecast/
8. Y. Wang, G. Cao, S. Mao and R. M. Nelms, "Analysis of solar generation and weather data in smart grid with simultaneous inference of nonlinear time series," (IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 600, 2015).
9. P. A. G. M. Amarasinghe and S. K. Abeygunawardane, "Application of Machine Learning Algorithms for Solar Power Forecasting in Sri Lanka" (2nd International Conference on Electrical Engineering (EECon), Colombo, Sri Lanka, 87, 2018).
10. Abuella, M., and Chowdhury, B.(2015) Solar power forecasting using artificial neural networks. In North American Power Symposium (NAPS) (pp. 1-5).IEEE, 2015.
11. W. Sun, T. Zhang, R. Tao and A. Wang, "Short-Term Photovoltaic Power Prediction Modeling Based on AdaBoost Algorithm and Elman," *2020 10th International Conference on Power and Energy Systems (ICPES)*, 2020, pp. 184-188, doi: 10.1109/ICPES51309.2020.9349730
12. Chen S, Gooi H, Wang M. “Sizing of energy storage for microgrids”. IEEE Trans Smart Grid, 3(1):142e51, 2012.
13. Lee, J.; Wang, W.; Harrou, F.; Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. Energy Convers.Manag.208, 112582, 2020.
14. Gupta, S.; Katta, A.R.; Bald Aniya, Y.; Kumar, R. Hybrid Random Forest, and Particle Swarm Optimization Algorithm for Solar Radiation Prediction. In Proceedings of the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 30–31; pp. 302–307, October 2020.
15. Sun, W.; Zhang, T.; Tao, R.; Wang, A. Short-Term Photovoltaic Power Prediction Modeling Based on AdaBoost Algorithm and Elman. In Proceedings of the 2020 10th International Conference on Power and Energy Systems (ICPES), Chengdu, China; pp. 184–188. 25–27 December 2020.
16. Sreekumar, S.; Sharma, K.C.; Bhakra, R. Optimized Support Vector Regression models for short term solar radiation forecasting in smart environment. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore; pp. 1929–1932. 22–25 November 2016.

Biographical notes

Mihir R. Patel is working as an Assistant Professor in M & V Patel Department of Electrical Engineering, Chandubhai S. Patel Institute of Technology, CHARUSAT University, Gujarat, India since January, 2016. He has completed his B.E. Electrical from C. S. Patel Institute of Engineering and Technology in 2012 and M.Tech. in Electrical Power System from C. S. Patel Institute of Engineering and Technology in 2014. He was lecturer in Electrical Department at IIET-Dharmaj from June, 2013 to July, 2014 and Assistant professor & GTU coordinator at IIET-Dharmaj from August, 2014 to January, 2016. His research interest area includes Power System Stability, Renewable Energy, and Forecasting. He has published Eight research papers in international journal and one conference research paper. He is Member of Indian Smart Grid Forum working group. His other roles are Department coordinator of EOC (Equal Opportunity Cell), Central council member and Hub Member of HRDC (Human Resources Development Center) at CHARUSAT University.

Manish U. Patel has received Diploma in Electrical Engineering from Vidhya Bharti Trust Polytechnic, Gujarat in 2010. He has worked as Junior Engineer in maintenance department from 2010 to 2019 at Kapil Engineering Works, Mehsana. His area of interest includes Industrial Automation and internet of things. Currently he is pursuing B. TECH in Electrical Engineering in Chandubhai S. Patel Institute of Technology, CHARUSAT UNIVERSITY, Changa.

Mihir S. Patel has received Diploma in Electrical Engineering from GUJARAT TECHNOLOGICAL UNIVERSITY, Gujarat in 2019. Currently he is pursuing B. TECH in Electrical Engineering in Chandubhai S. Patel Institute of Technology, CHARUSAT UNIVERSITY, Changa.

Vidhi R. Patel Currently She is pursuing B. TECH in Electrical Engineering in Chandubhai S. Patel Institute of Technology, CHARUSAT UNIVERSITY, Changa. Her area of interest is robotics and internet of things.

Tejas C. Patel has received Diploma in Electrical Engineering from B&B Institute of Technology, GUJARAT TECHNOLOGICAL UNIVERSITY, Gujarat in 2020. Currently he is pursuing B. TECH in Electrical Engineering in Chandubhai S. Patel Institute of Technology, CHARUSAT UNIVERSITY, Changa.