World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Exploring and forecasting of solar radiation with machine learning methods

Sasmita Kumari Nayak *

Associate Professor, Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India.

World Journal of Advanced Research and Reviews, 2023, 20(03), 824–828

Publication history: Received on 31 October 2023; revised on 13 December 2023; accepted on 15 December 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.20.3.2513

Abstract

Machine learning has recently advanced to the point that a wide range of solar predicting works have been produced. Specifically, one of the most widely used methods at the moment for hourly solar forecasting is machine learning. However, it appears that there is a misconception regarding forecast accuracy—almost all study articles assert to be better than others. However, it appears that there is a misconception regarding forecast accuracy—almost all study articles assert to be better than others. It is illogical for solar forecasters to place their initial wager on a single model for any new forecasting project. Only commercially available versions of these algorithms are employed, and no hybrid models are taken into account to guarantee an equitable comparison. Additionally, the package's automatic tuning algorithm is used to train each model. Overall results show that tree-based methods consistently yield good results. Reliable generation forecasts are becoming more and more necessary for grid operation as distributed renewable power grows in penetration. In order to produce the most accurate day-ahead hourly irradiance forecasts, the current work combines cutting edge Weather Research and Forecasting (WRF) model implementations with machine learning techniques.

Keywords: Solar radiation forecasting; Machine Learning; Linear Regression; Decision Tree Regressor; Random Forest

1. Introduction

At all times, an electrical operator must make sure that the production and consumption of electricity are precisely balanced. This is frequently exceedingly challenging to maintain with a traditional, controllable energy production system, particularly in a small or disconnected electrical grid. These days, many nations are thinking about integrating renewable energy sources into their electrical systems. Because the resource (solar radiation, wind, etc.) is unstable, this leads to even more issues. As a result, it is crucial to have good solar radiation prediction skills, particularly when dealing with high energy integration. Science-based projections and machine learning models were the two main types of solar radiation prediction techniques. The aim of this work is to use machine learning algorithms to provide a summary of solar radiation prediction in this context [1].

The significant integration of renewable energy sources, especially those that are unpredictable like solar and wind, into the current or future energy supply structure will be one of the biggest challenges for the world's energy supply in the near future. At all times, an electrical operator must make sure that the production and consumption of electricity are precisely balanced. In actuality, the operator frequently faces challenges in maintaining this balance with traditional, controllable energy production systems, particularly in small or isolated electrical grids that are not connected. The electrical system's dependability then hinges on its capacity to handle disturbances and both anticipated and unforeseen changes (in production and consumption) while preserving the level of quality and uninterrupted customer service. Next, with different temporal horizons, the energy supplier has to manage the system (Figure 1).

* Corresponding author: Sasmita Kumari Nayak; Email: nayaksasmita484@gmail.com

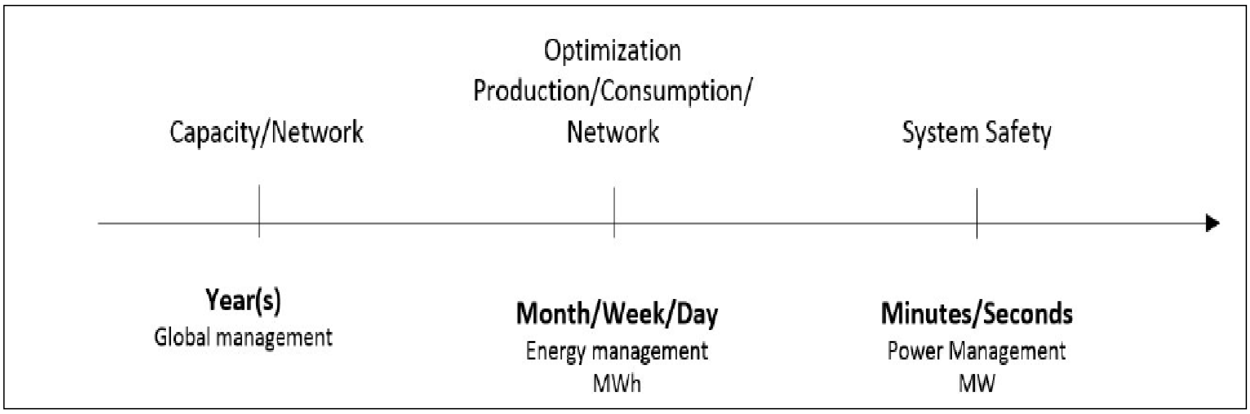

Figure 1 Energy Management Prediction scale of an electrical network

The following section will describe the motivation for the work followed by data, methods for deriving classifications, and constructing the models in section 3. Section 4 compares the performance of four ML models used for accuracy detection models (i.e., Decision Tree, Random Forest, Naïve Bayes and SVM). A summary of the study is given in the Section 5.

2. Literature Review

Various devices can be used to measure and utilize Rs for different purposes; however, the cost and lack of measuring equipment present difficulties in precisely determining and keeping track of Rs levels. To tackle this issue, substitute techniques for estimating Rs have been devised, such as machine learning (ML) and artificial intelligence models [2]. In this work, five distinct machine learning-based methods were used to estimate monthly solar radiation (SR). Long short-term memory (LSTM), Support vector machine regression (SVMR), extreme learning machines (ELM), K-nearest neighbors (KNN), and Gaussian process regression (GPR) are the models that are employed [3]. Therefore, predicting solar system output power is necessary for either the best possible management of energy fluxes into the solar system or the efficient operation of the power grid. In this case, we employed a variety of machine learning models, including Random Forest, Decision Tree, and Linear Regression models [4–7].

There are several ways to forecast solar power, but two main approaches are machine learning models and cloud imagery paired with physical models. The prediction horizon is the primary determinant of the method to be employed; in actuality, the accuracy of each model varies depending on the horizon chosen. Depending on the desired forecasting time, several methods are available to forecast solar irradiance. These techniques are divided into two classes by the literature:

- Short-term forecasts up to six hours are typically suitable for extrapolation and statistical procedures utilizing satellite images or measurements on the ground level and sky images [8–11]. There are two sub-classes within this class: Now casting (0–3 h), where the extrapolation of real-time measurements must serve as the basis for the forecast [5]; Short-Term Forecasting (3–6 hours), in which real-time measurements or satellite data are combined with Numerical Weather Prediction (NWP) models and post-processing modules.
- NWP models are capable of predicting up to two days in advance or more (up to six days in advance). Post-processing modules and satellite data are commonly utilized in conjunction with these NWP models.

3. Methodology

Machine learning is an artificial intelligence approach and a subfield of computer science. This method has the benefit of allowing a model to solve problems that cannot be represented by explicit algorithms, and it can be applied in multiple domains. A thorough analysis of various deterministic and machine learning techniques for predicting food, crop yield, weather, and hepatitis is provided in [4- 6, 12-15]. Even in situations where representation is not feasible, machine learning models identify relationships between inputs and outputs; this characteristic allow the use of machine learning models in many cases, for example in data mining and forecasting problems, spam filtering, classification problems, and pattern recognition. Because one must work with large datasets and machine learning models can handle pre-processing and data preparation, the classification and data mining aspects of this field are especially intriguing.

Following this stage, forecasting issues can be solved using the machine learning models. Three approaches are available for using the models in global horizontal irradiance forecasting:

- Structural models that rely on additional geographic and meteorological factors;
- Time-series models that solely employ historical observations of solar irradiance as input features (endogenous forecasting);
- Hybrid models that treat exogenous variables (exogenous forecasting) such as solar irradiance and other variables.

As was previously mentioned, machine learning is a subfield of artificial intelligence. It focuses on building and researching systems that can learn from datasets, enabling computers to acquire knowledge without explicit programming. The main goal of machine learning, which is a collection of computer programs, is to create mathematical models by utilizing statistics to draw conclusions from samples. Learning is the process of running a computer program that optimizes a model's parameters based on training data or experiences, given a model that specifies certain parameters. The model can characterize the knowledge gleaned from the data, forecast the future state, or do both. A training data set is used to train an ML algorithm and produce a model. The machine learning algorithm uses the model to make predictions when it receives new input data. The prediction's accuracy is assessed, and the ML algorithm is used if it meets acceptable standards. The machine learning algorithm is repeatedly trained using an augmented training data set if the accuracy is deemed unacceptable. Fig. 2 depicts the machine learning model's process.

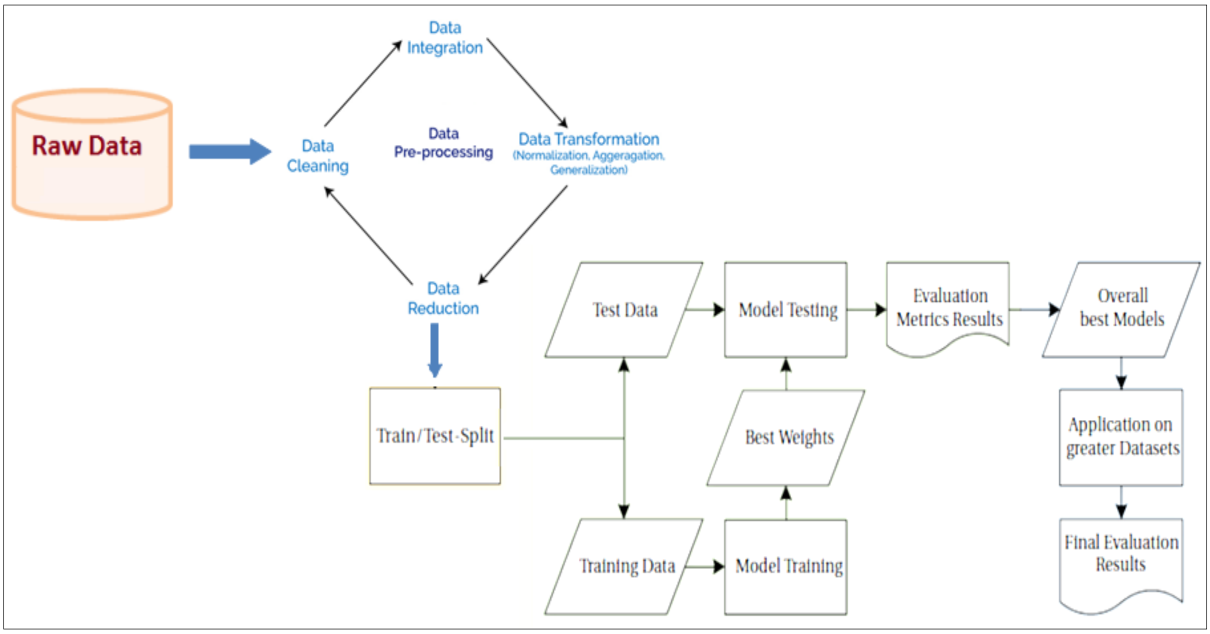

Figure 2 Experimental Process of Machine Learning [4]

3.1. Linear Regression

A supervised machine learning technique called linear regression [4–7, 12–15] yields continuous output predictions with a constant slope. Instead of attempting to categorize them, it is used to predict values within a continuous range (e.g., price, sales). The notion of a deterministic world pervaded early attempts to study time series, especially in the 19th century. The concept of stochasticity in time series was first introduced by Yule (1927), who made the important assumption that each time series can be thought of as the realization of a stochastic process. Since then, several time series techniques have been created based on this straightforward concept. The formulation and resolution of the linear forecasting problems were made possible by World's decomposition theorem. Since then, a sizable body of work on time series has been published, covering topics such as identification, forecasting, model checking, and parameter estimation.

3.2. Decision tree

This is a very basic idea. Predicting a response or class Y from inputs X1, X2,..., Xp is necessary. To accomplish this, a binary tree is grown. A test is applied to one of the inputs, let's say Xi, at every node in the tree. Either the left or the right sub-branch of the tree is chosen, depending on the test's result. When a leaf node is eventually reached, a prediction is made there. The training data points that arrive at that leaf are all averaged or combined in this prediction. Using each of the independent variables, a model is produced. The optimal split is found using mean squared error for each of the individual variables. The total number of features is the maximum number that must be taken into account at each split [4–7, 12–15].

3.3. Random Forest

A random forest is a meta estimator that employs averaging to increase prediction accuracy and manage over-fitting by fitting multiple classifying decision trees on different subsamples of the dataset [4-7, 12–15].

4. Results and discussion

The dataset indicates that the solar power system's radiation (W/m^2), sunrise, wind direction (Degrees), humidity, pressure, and temperature (F) were recorded at various times during the day between September 1, 2016, and December 31, 2016. We use the Anaconda application for the experiment, writing and running the code in a Jupyter notebook. The model score, r2 score, median absolute error, and explained variance are what we're looking for here.

We have performed Linear Regression, Decision Tree, and Random Forest model for our collected data to found out which model has better accuracy and the results we found were given in the below Table 1. From the table, we can easily conclude that Random forest ML model has the best accuracy in comparison with other models.

Table 1 Results of different ML Models

Result	Linear Regression	Decision Tree	Random Forest
Model Score (%)	62.94	86.29	92.94
R Square (%)	62.94	86.29	92.94
Median Absolute Err (%)	115.354	64	84.78
Explained Variance	0.6295	0.8629	0.9294

5. Conclusion

As demonstrated, there are a wide variety of methods available. Many techniques exist for estimating solar radiation; some are commonly used (ANN, naive methods), while others are used less frequently (boosting, regression tree, random forest, etc.). Still others are used more frequently (SVM, SVR, k-mean). One is the greatest in certain circumstances, but not always, and vice versa. These techniques' accuracy is typically influenced by the training data quality. SVM, regression trees, and random forests are the three approaches that ought to be applied consistently in the upcoming years due to the very encouraging results that have been shown thus far, and the intriguing research that will undoubtedly be conducted in the coming years. In actuality, these techniques produce comparable error statistics. The mistakes noted in the literature might have less to do with the methods' execution and more to do with how they were used. In comparison to other ML models, the Random Forest model has the best accuracy of 92.94% for our data, it can be concluded. In addition to handling missing values, the Random Forest model can be applied to regression and classification tasks. Finding the most significant features from the training dataset is its primary benefit.

References

[1] K. Venu, K. V. Indhu Prakash, S. Jayaram, N. S. Karan, M. M. Raja and K. Renu, "Solar Radiation Prediction using Machine Learning Model," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 52-57, doi: 10.1109/ICSCDS56580.2023.10104904.

[2] Hissou, H., Benkirane, S., Guezzaz, A., Azrour, M., & Beni-Hssane, A. (2023). A novel machine learning approach for solar radiation estimation. Sustainability, 15(13), 10609.

[3] Demir, V., Citakoglu, H. Forecasting of solar radiation using different machine learning approaches. Neural Comput & Applic 35, 887–906 (2023). https://doi.org/10.1007/s00521-022-07841-x

[4] Sasmita Kumari Nayak, Swati Sucharita Barik, Mamata Beura," Weather Forecasts Based on Rainfall Prediction Using Machine Learning Methodologies," Adalya Journal 9 (6), Page No: 72 – 80, ISSN NO: 1301-2746.

[5] Nayak, S. K. (2020). Analysis and high accuracy prediction of coconut crop yield production based on principle component analysis with machine learning models. International Journal of Modern Agriculture, 9(4), 359-369.

[6] Nayak, S. K., Barik, S. S., & Beura, M. (2020). Analysis of Infectious Hepatitis Disease with High Accuracy Using Machine Learning Techniques. TEST Engineering & Management, 83, 83.

[7] Mishra, S. P., Siddique, M., Beura, M., & Nayak, S. K. (2021). Analysis of Indian Food Based on Machine learning Classification Models. Journal of Scientific Research and Reports, 27(7), 1-7.

[8] Barik, Swati & Nayak, Sasmita & Panda, Stitiprajna & Tripathy, Aeisuriya & Mohapatra, Gourav. (2020). Human Face Recognition using LBPH. International Journal of Recent Technology and Engineering. 8(2), 3208-3212. 10.35940/ijrte.F8117.038620.

[9] Nayak, S. K., Panda, C. S., & Padhy, S. K. (2019, February). Dynamic Task Scheduling Problem Based on Grey Wolf Optimization Algorithm. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-5). IEEE.

[10] N. Dash, S. K. Nayak and J. Majumdhar, "Detection of Cut Transition in Videos Using Optical Flow and Clustering," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-7, doi: 10.1109/ASIANCON51346.2021.9544553.

[11] Majumdhar, J., & Nayak, S. K. (2021, August). A Novel Method on Summarization of Video Using Local Ternary Pattern and Local Phase Quantization. In 2021 2nd International Conference on Range Technology (ICORT) (pp. 1-6). IEEE.

[12] Nayak, S. K. (2023). Classification of cyclones using machine learning techniques. World Journal of Advanced Research and Reviews, 20(2), 433-440.

[13] Jagadeesan, S., Barman, B., Agarwal, R. K., Srivastava, Y., Singh, B., Nayak, S. K., & Venu, N. A Perishable Food Monitoring Model Based On Iot And Deep Learning To Improve Food Hygiene And Safety Management. interventions, 8, 9.

[14] Sasmita Kumari Nayak. (2020). CONSTRICTING THE BIVARIATE ANALYSISOFHIGHEST CROP YIELD PRODUCTION BASED ON DIFFERENT ZONES OF INDIA. International Journal of Modern Agriculture, 9(4), 216 - 226. Retrieved from https://modern-journals.com/index.php/ijma/article/view/205

[15] Sasmita Kumari Nayak, Mohammed Siddique. (2020). EFFECT OF STOCK INDEX PARAMETERS ON FORECASTING THE HIGH STOCK VALUE OF VISA STEEL USING DEEP LEARNING NEURAL NETWORK MODEL. International Journal of Modern Agriculture, 9(4), 227 - 236. Retrieved from https://modern-journals.com/index.php/ijma/article/view/207