Potential Assessment of Solar Photovoltaic Generation using Machine Learning Algorithms for Northern Region of India

P.Upendra Kumar[1],K.Lakshmana Rao[2],Ramakrishna S S Nuvvula[3]

Abstract

Nowadays, with more and more large-scale linked solar power generation systems, the stable operation of the power grid depends heavily on the amount of solar electricity connected to the grid. This is not only necessary for stable operation but also for generation allocation and load planning. For this, an accurate method for estimating the potential is needed. In this paper, a modest attempt was made to estimate the potential for SPV production in northern India. The methodology presented is based on an efficient machine learning algorithm based on the regression method. A decision tree, linear and support vector regression model for predicting the number of units generated was presented. Key performance metrics such as mean absolute error, mean square error, root mean square error, and R2 score were considered to assess the efficiency of these algorithms. Linear regression models were observed to perform better than all the other methods considered in this study, and the same was summarized in the results.

Key Words: Grid interconnection, Assessment, Machine Learning, Regression.

DOI Number: 10.14704/nq.2022.20.8.NQ44020 **NeuroQuantology 2022; 20(8):175-181**

1. Introduction

Global energy demand in developing countries is projected to quadruple by 2030 and grow at a rate of nearly 4-5 per year. Furthermore, the share of global electricity demand in emerging markets is expected to rise from 27% in 2000 to 43% in 2030. To accommodate this growth, renewable energy sources in the country need to be improved. According to the latest statistics from the Central Electricity Authority (CEA), the installed capacity in India in May 2022 is 4,02,817 MW. In the total generation of, the share of renewable energy sources is around 39.7%, i.e. 1,59,949 MW, with the production of solar energy being around 56,951 MW, i.e. 35.60% of all renewable energy sources according to CEA statistics as of May 2022. Peak demand in India is around 2,15,888 MW with peak met of 2,07,231 MW at a deficit of 4.0%. With the associated environmental problems and depletion of fossil resources, the focus is now shifting to the use of renewable energy. This requires extending the grid to include renewable energy sources to take advantage of generation diversity and dispersed renewable resources [1].

It is necessary to plan and construct additional generating capacity simultaneously in order to meet the rising demand for energy. One of the renewable energy technologies that has received the most investigation in this area is solar photovoltaic (SPV) technology, which can meet rising energy demand while lowering carbon emissions, preserving fossil fuels, and using up less natural resources. It also has the advantage that the panels are industrially manufactured and easily assembled at the project site, facilitating their installation and operation.

It is evident that such a large amount of grid-tied solar power generation requires proper planning in execution and operation to maintain grid stability. In this regard, accurate estimation of solar power generation is required for effective planning. However, the output power generation of the solar array is random in nature.

Corresponding author: P.Upendra Kumar

Address: 1,3 Department of EEE,GMRIT, Andhra Pradesh, India,2 Department of CSE,GMRIT, Andhra Pradesh, India

It depends on weather conditions, ambient temperature, humidity, etc. One of the most reliable ways to meet the growing load demand by increasing SPV generation is to conduct a solar power generation assessment to get advance power generation information so that other activities related to generation planning and O&M can be planned without effects on network operation. To connect the solar system to the grid, an efficient prediction mechanism is required to avoid grid instability problems [2]. Predicting solar energy is not a simple process as it is highly dependent on weather and atmospheric conditions that change over time. Previously, the estimation process relied on metrological data from numerical weather forecasts and satellite imagery illustrating cloud movement to predict insolation and other dependent parameters. The fundamental problem with the older methods is that the required meteorological data for the SPV location is not always accessible and is not always available at the required resolution, which limits their applicability for extremely accurate forecasts. To conquer the issue it's miles critical to apply new and smart techniques to get legitimate and correct consequences. Presently, superior estimation algorithms and strategies for strength output estimation which mixes the blessings of synthetic intelligence and gadget mastering algorithms are gaining significance because, they are able to extract certain facts from photo voltaic strength information and bring greater dependable forecast consequences [2]. In latest past, Machine Learning introduced radical modifications in various domains. In this connection, recently, maximum of the researchers commenced integrating Machine Learning primarily based totally prediction techniques with inside the subject of electrical engineering like grid management, fault prediction, load balancing, output power prediction and load prediction etc. [3,4]. Regression models together with Linear Regression, Support Vector Regression and Decision tree models are famous methods of supervisory mastering with inside the gadget mastering domain. In this paper it is proposed to evaluate the capacity of SPV technology the usage of the formerly noted regression primarily based totally gadget mastering strategies for northern region of India and the consequences are summarized.

2. Solar PV Technology

The world's most plentiful and sustainable energy source is solar energy. Solar power systems are gadgets that transform sunlight's heat or light into different types of energy. Figure 1 shows a typical layout of a solar power system. The main components of solar systems are the PV system, power processing systems and batteries for energy storage [5, 6]. 176

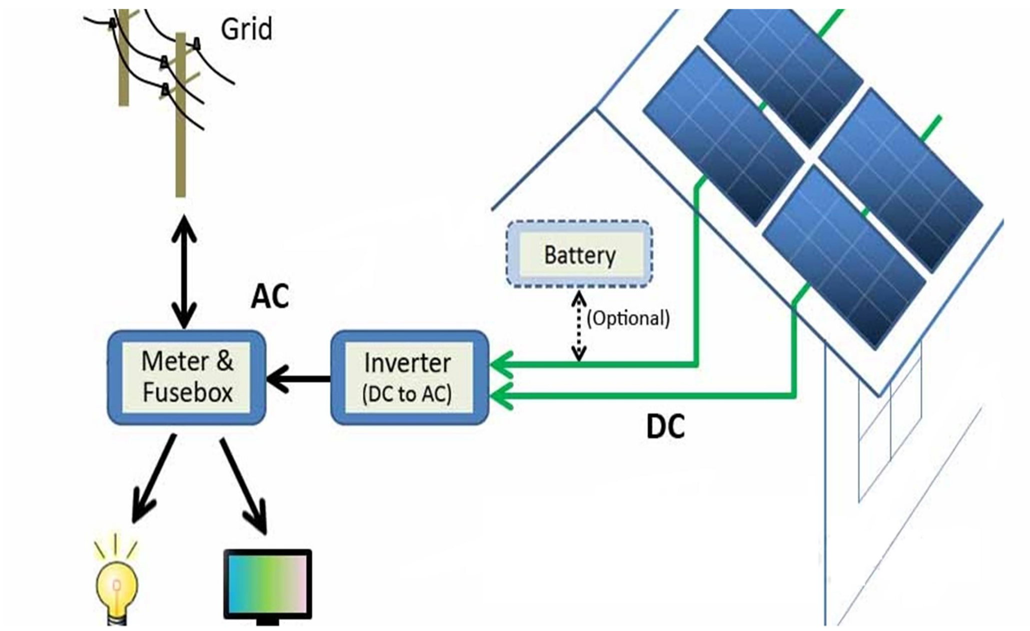

Figure 1: Layout diagram of Solar Power Plant

2.1 Photovoltaic (PV) Array

It is a series of photovoltaic panels that convert sunlight into direct current. Solar cells are the individual cells that combine to form modules, which in turn form photovoltaic modules, and a group of modules forms a photovoltaic array. Each solar cell generates a modest amount of electricity. A solar panel is constructed by joining several solar cells together and produces a sizable amount of electricity. Photovoltaic systems are available in different sizes depending on the application. There are three different types of solar panels that can be used in photovoltaic systems. Mono crystalline, polycrystalline and amorphous thin films are the three types. The cost and potency of electrical device varieties are the key differentiators.

2.2 Power Conditioning Equipment

As shown in Figure 1, a solar power conditioner device is an integrated system consisting of a solar charging controller, an inverter, and a main charger. The power conditioner unit continuously monitors battery voltage, solar power and load status. Due to the constant power consumption, the power controller automatically switches the load to the grid and at the same time charges when the battery voltage drops below the set value. The conditioner unit always prioritizes PV and uses grid power only if the PV or battery charger does not meet the load requirements.

2.3 Batteries for Energy Storage

The energy generated by photovoltaic panels can be stored in batteries for future use or at times when SPV generation decreases. Photovoltaic solar and storage technology is the optimal and powerful combination to achieve independent and autonomous energy production and consumption goals during day, night and bad weather. The battery resources DC present day for a sure duration of time. The existence of battery relies upon at the present day provided via way of means of the battery and the most prices the battery can hold. The maximum charge a battery can charge is usually stated on the battery in ampere-hours (A-h). This unit represents the current the battery can supply and how long the battery can supply current.

2.4 SPV energy generation

The energy production of a solar power plant is given by

$$E = A \times r \times H \times PR \qquad (1)$$

Where E is the energy in Kwh, A is the total area of the panel (m2), r is the efficiency of the solar panel (%), H is the annual average solar radiation and PR is the ratio efficiency. Performance reports estimate the quality of the SPV system to deliver system performance regardless of module orientation or tilt. This includes all losses depending on system size, technology used and location [7]. Solar power generation depends on various parameters, the characteristics of which are described in Table 1. The specifications and parameters to be monitored for the grid connected SPV generation system are shown in Figure 2.

Table 1: Parameters that affect SPV generation

Name of the Feature	Description of the Feature	Units
TT	Module Temperature	^oC
HH	Humidity	g/m^3
WD	Wind Direction	Degrees
WS	Wind Speed	Km/hr
PR	Atmospheric Pressure	mbar
SR	Irradiance	W/m^2
ST	Standard or Ambient Temperature	^oC
Net Reading	Generated billable units at Inverter	Kwh

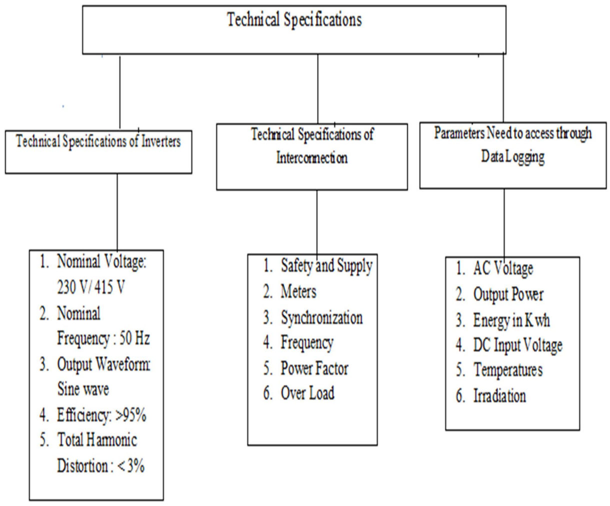

Figure 2: Specifications and typical parameters for grid connection

3. Methodology

The intensity of solar radiation in an area depends on latitude, time of year, atmospheric conditions, air quality and pollution index. The problems stem from solar radiation, heat, photo voltaics, expensive energy storage, grid stability, and constant fluctuations due to seasonal influences. Furthermore, integrating the SPV system into the grid as a backup power source to meet growing demand is not technically feasible directly. Such a structure affects the stability of the network. In short, fluctuations in weather conditions cause uncertainty in the operation of the SPV. In this respect, precise and accurate models are needed to estimate SPV emissions, especially for large-scale grid-connected systems. Previously, the estimation process depended on measurement data from digital weather predictions and satellite images illustrating cloud motion to predict solar irradiance and other dependent parameters [8]. To overcome all the shortcomings of previous prediction methods, this study focuses on various machine learning-based regression methods for predicting solar PV system performance. The proposed ML algorithms are used to forecast number of units generated by a solar power plant depending on various independent variables such as module temperature, humidity, wind direction, wind speed, barometric pressure, radiation, standard or ambient temperature [9-13].

3.1 Decision Tree Model

The call tree generates regression models with a tree structure. It divides information set into smaller associate in smaller subsets whereas additionally bit by bit developing an associated decision tree. The tip result's a tree with leaf nodes and decision nodes.

Coefficient of variation = $\frac{S}{x}\times 100\%$ (2)

Where S represents the standard deviation and x represents the average value of leaf nodes.

3.2 Linear Regression

This model effectively fits the given data and determines the optimal straight line with few errors. Find a linear / straight line between the predictor variables (module temperature, humidity, wind direction, wind speed, atmospheric pressure, solar radiation, standard or ambient temperature) and the response variables (number of billable units generated). This is known as linear regression. If Y is the structured variable and X is the unbiased variable, the populace regression line is given by

$$Y = B_o + B_1 X \qquad (3)$$

Where B_0 is a constant and B_1 is a regression coefficient.

3.3 Support Vector Regression

It finds a better hyper plane with fewer errors and overcomes the positive and negative limits of this hyper plane during its learning phase. Then in the test phase, it verifies which side of the hyper plane the new point lies to predict its value. The decision surface that separates the classes in the hyper line has the form

$$W^T X + b = 0 \quad (4)$$

Where W is the weight vector, X is the input vector, and b is the offset.

4. Performance Indices

The performance of the proposed model is evaluated using a specific key performance indicator [9]. The MAE, MSE, and RMSE validation values are close to zero, which means that the actual values are similar to the predicted values. The significance of the distinction among the genuine price of the estimate and the located price is known as absolutely the error. MAE calculates the average of the absolute errors in a series of predictions and observations to determine the size of the error for the entire set. MSE measures the root mean square difference between the estimated and actual values. MSE is the standard deviation of the residuals. Residual is a measure of the distance from a data point on the regression line. The RMSE is a measure of the distribution of these deposits. That is, how much the data is concentrated around the optimal line. The R^2 score shows how well the regression model fits the observations. In general, the closer the score is to 1, the better the model.

4.1 Mean Absolute Error (MAE)

It calculates absolutely the distinction among the real information point (A_i) and the expected (P_i). This difference produces an absolute error (E_i) committed by the model. It finds the sum of all absolute errors, such as $\sum_i^n E_i$ divided by the total number of statistics points, is called the MAE. Its mathematical representation is shown in the Eq. 5.

$$\text{Mean Absolute Error} = \tfrac{1}{n}\sum_{i=1}^{n}\left|A_i - P_i\right| \quad (5)$$

4.2 Mean Square Error (MSE)

It finds square distance between actual and expected information points. The sq. operation is helpful to avoid the cancellation of negative terms. Its mathematical illustration is as shown in Eq. 6.

$$\text{Mean Square Error} = \tfrac{1}{n}\sum_{i=1}^{n}(A_i - P_i)^2 \quad (6)$$

4.3 Root Mean Square Error (RMSE)

It tells however closely the information scattered round the line. It's measured as root of MSE as shown in Eq. 7.

$$\text{Root Mean square Error} = \tfrac{1}{n}\sqrt{\sum_{i=1}^{n}(A_i - P_i)^2} \quad (7)$$

4.4 R2-Score

This metric describes concerning the performance of the regression method. It's the important thing output of a regression analysis. It' mentioned due to the fact the fraction of the based variable' version so as to be foreseen with the aid of using the freelance variable. The determination constant ranges from zero to one. The upper worth of R^2-score indicates that the model higher fits the ascertained information points. The calculation of R^2-score divides the total of square Regression (SSR) with the total of Squares Total (SST) and subtracts its result from 1 as shown in Eq. 8. The SSR could be a sum of the distinction between every jth predicted worth ($\hat{P}_j$) and also the actual variable (A_j) adore $\sum_{i=1}^{n}(A_j - \hat{P}_j)^2$. The SST is square total of the distinction between every jth actual dependent value (A_j) and the mean of actual dependent variable ($\bar{A}$) such as $\sum_{j=1}^{n}(A_j - \bar{A})^2$.

$$R^2 \text{ Score} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{j=1}^{n}(A_j - \hat{P}_j)^2}{\sum_{j=1}^{n}(A_j - \bar{A})^2} \quad (8)$$

5. Results and Discussions

For preforming the estimation, the day sensible hourly statistics is taken into consideration from sun electricity plant life running in South India with diverse realistic parameters like voltage, current, electricity in phrases of AC and DC portions together with frequency. The dataset includes seven functions and 1116 instances. Of the entire statistics accrued from the dataset, 80% of statistics is used for schooling and 20% of statistics is used for checking out purpose. The based variable Net Reading i.e. quantity of devices generated is based on diverse unbiased variables consisting of module temperature, humidity, wind direction, wind speed, atmospheric pressure, irradiance and ambient temperature.

Table 2 and Figure 3 illustrate the results of the MAE, MSE, RMSE, and R^2 score metrics of the three proposed regression methods. From the performance indices, it can be seen that the R^2 score of linear regression is higher than that of all other regression methods. Of the three methods, it is clear that the linear regression model performs better than any of the methods performed in this study.

Table 2: Values of performance metrics for regression models

S. No	Regression Name	Mean Absolute Error	Mean Square Error	Root Mean Square Error	R^2 Score
1	Decision Tree Model	1.6637547456	10.0626437697	3.1721670463	0.9981856
2	Linear Regression	2.8697874256	1.3196600713	3.6327125833	1
3	Support Vector Regression	0.9078537099	3.5099638366	1.8734897482	0.9993671

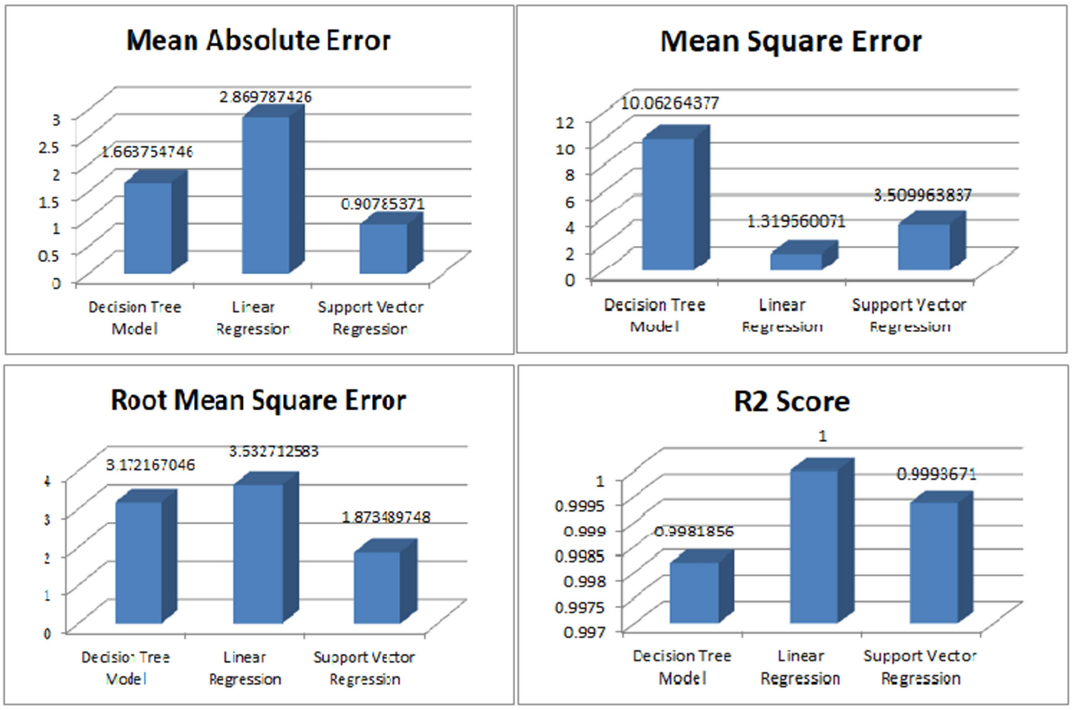

Figure 3: Illustration of regression model performance indices

6. Conclusions

In this paper, a modest try has been created to estimate the potential of SPV generation for northern region of India. Since massive scale grid interconnected SPV generation systems are increasing day by day, the stable operation of grid extremely depends on the quantity of SPV energy penetrating into the grid. This is often not solely essential for stable operation however additionally necessary for generation allocation and load scheduling. So as to attain this, an explicit methodology for estimating the potential is necessary. The methodology employed in this study, associate machine learning algorithmic based regression strategies namely decision tree, linear and support vector regression models for prediction of range of units' generated has been conferred. The analysis for potency of those algorithms was tested supported key performance indicators adore MAE, MSE, RMSE and R^2 score. The linear regression model was determined to outperform all adversarial strategies considered in this study. The on top of techniques used and methodology proved to be economical and effective for potential assessment for SPV systems. It's to be noted that the methodology presented is extended to all or any renewable power generation sources for addressing the issues relating to grid operation.

Further, the projected methodology is extraordinarily helpful throughout starting stage for capability fixation, generation and load scheduling.

References

CEA, Load generation balance report, New Delhi, India, 2022.

Mariam AlKandari and Imtiaz Ahmad, Solar power generation forecasting using ensemble approach based on deep learning and statistical methods, Applied Computing and Informatics, 2019, PP. 01-20.

Nils Jakob Johannesen, Mohan Kolhe, Morten Goodwin, Relative evaluation of regression tools for urban area electrical energy demand forecasting, Journal of Cleaner Production, 2019, PP. 01-20.

Aroosa Tahir, Zahoor Ali Khan, Nadeem Javaid, Zeeshan Hussain, Aimen Rasool, and Syeda Aimal, Load and Price Forecasting Based on Enhanced Logistic Regression in Smart Grid, Springer Nature Switzerland AG 2019, PP. 221-233.

S Venkateswarlu, T S Kishore, "Enhancement of Static Voltage Stability Margin using STATCOM in Grid Connected Solar Farms", TEQIP-III Sponsored Electric Power and Renewable Energy Conference-2020 (EPREC-2020), 29th - 30th May 2020, NIT Jamshedpur, India.

Kishore T S, Koushik S D, Venu Madhavi Y, "Modelling, Simulation and Analysis of PI and FL Controlled Microgrid System", 3rd IEEE International Conference on Electrical, Computer and Communications Technologies (ICECCT 2019), 20-22 Feb, 2019, SVS College of Engineering, Coimbatore, India.

Jose R. Andrade and Ricardo J. Bessa, Improving renewable energy forecasting with a grid of numerical weather predictions, IEEE Transactions on Sustainable Energy, 2017, 8(4), PP. 1571–1580.

Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., Lughi, V. Advanced methods for photovoltaic output power forecasting: A review. Appl. Sci., 2020, 10, 487.

Dinh Van Tai, Solar photovoltaic power output forecasting using machine learning technique, IOP Conf. Series: Journal of Physics: Conf. Series 1327, 2019, PP. 01-05.

Tserenpurev Chuluunsaikhan, Aziz Nasridinov, Woo Seok Choi, Da Bin Choi, Sang Hyun Choi, Young Myoung Kim, Predicting the Power Output of Solar Panels based on Weather and Air Pollution Features using Machine Learning, Journal of Korea Multimedia Society Vol. 24, No. 2, 2021, PP. 222-232.

Jung-Pin Lai , Yu-Ming Chang, Chieh-Huang Chen and Ping-Feng Pai, A Survey of Machine Learning Models in Renewable Energy Predictions, Appl. Sci., 10, 5975, 2020.

S. Sobri, S. Koohi-Kamali, N.A. Rahim, Solar photovoltaic generation forecasting methods: a review, Energy Convers. Manage, 2018, 156, PP. 459–497.

R. Ahmed, V. Sreeram, Y. Mishra, M.D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, Renewable and Sustainable Energy Reviews 124 , 2020.