IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AUTOMATED ARCHITECTURE DESIGN FOR SOLAR IRRADIANCE PREDICTION

Mrs. SUSHMA V, Ms. AMRUTHA, Ms. LIKITHA B, Mr. YASHWANTH M, Mr. AKSHAY GOWDA S G

ASSISTANT PROFESSOR , STUDENT, STUDENT, STUDENT, STUDENT
COMPUTER SCIENCE AND ENGINEERING,
ATME COLLEGE OF ENGINEERING, MYSURU, INDIA

Abstract: Precise estimation of solar radiation required for a continuous flow of electrical power has always been a challenging undertaking, in the domains of artificial intelligence and physical simulation. Power grid operators would find it simpler to take preventative measures ahead of time to guarantee failsafe outcomes during instances of abrupt or cascading drops in solar energy if solar irradiance could be predicted with extremely high precision. This paper used historical meteorological data from two consecutive years to forecast hourly solar irradiance of Johannesburg city using three machine learning techniques. The objective was to develop a model that could easily make use of historical meteorological data in order to forecast extremely precise hourly solar irradiance data. According to the results, Random Forest using normalized root mean square.

Index Terms - Solar Irradiance prediction, machine learning, Random Forest.

Introduction

Renewable energy integration into power systems has attracted great interest in recent years because of the growing demand for electricity, the aim to diversify the energy basket in most of countries, and the need to reduce CO2 25 emissions 26 by protecting the environment. Because energy consumers rely heavily on fossil fuels, the world's supply of these resources is running out. Because they release harmful gases into the atmosphere, fossil fuels, being non-renewable energy sources, have a negative impact on the environment by causing pollution and climate change. Only when dependable substitutes for fossil fuels are created will the overreliance on them begin to decrease.

Recently, solar energy has become more popular all over the world . Its accessibility and advancements in photovoltaic (PV) cell technology are to blame for this. Solar energy is an abundant and clean energy source that has minimal impact on the environment .Many investments have been made worldwide to increase solar power efficiency, which has led to the development of technologies that have produced new kinds of cell candidates to take the place of the more conventional and costly crystal silicon cells. Thin-Film and Multijunction Cell technologies are two examples of these new cell candidates. Solar power is now one of the most practical renewable energy sources thanks to this technological trend and the global effort to reduce greenhouse gas emissions. The key concept in the field of solar power production is solar irradiance It is the solar energy released, manifested as a broad range of light waves. In a nutshell, it's the amount of power (W/m^2) that is obtained from the sun's electromagnetic radiation. It is the solar radiation that is emitted and manifested as a broad range of light waves. To put it simply, it is the amount of power (P/A) that is received from the sun in the form of electromagnetic radiation. In a nutshell, electromagnetic fields cause vibrations that are represented by light waves.

The sun radiation that enters the atmosphere is filtered mostly by organic substances like nitrous oxide, ozone, water vapor, and carbon (IV) oxide. The atmospheric matter absorbs, emits, and scatters incident solar radiation as a result of these interactions. While scattering causes the radiative energy to be redistributed in all directions, absorption of incident radiation results in a reduction of radiative energy moving at the incident angle. Solar radiation is the energy that, in the end, makes its way to the photovoltaic (PV) cells through the interaction of radiation and matter. This paper will use machine learning techniques to predict the amount or intensity of solar radiation. solar power generation is an intermittent process, accurate forecasting of solar irradiance is crucial for grid operators and solar power supply companies alike. Grid operators need prediction data in order to create supply-demand plans that minimize power outages. This can be done by creating a backup plan for electricity generation based on the prediction data, which will provide fail safe scenarios in the event of cascading power outages caused by variations in solar radiation. In order to predict hourly solar irradiance in this work, historical meteorological data was acquired as input parameters via subscription from the Meteoblue website. The measured (actual) GHI, sunshine duration, temperature, and relative humidity are among the data. Two years' worth of data were used. Support vector regression (SVR), random forest (RF), and artificial neural network (ANN) are the three machine learning techniques that were applied.

I. EXPERIMENTAL SET UP

For an accurate prediction, the training data's composition and quality, as well as the data's source, are essential. Meteoblue provided meteorological weather data for Johannesburg, which was used in this paper. We used hourly data from 2017 and 2018, two years in a row. The data were cleaned before being used, with all nighttime data that included zero measured solar irradiance (SI/d) removed. Only daily data from 7 a.m. to 6 p.m. was used. The three models were trained independently using data from each of these years, and the outcomes were then compared. 30% of the data were designated as testing data, and the remaining 70% as training data. The following parameters are present in the dataset and were used as input:

a. Temperature at two meters above sea level
b. Relative humidity at an elevation of two meters
c. Duration of sunshine and
d. Solar radiation measurements

The general process input data were split, pre-processed, and fed into the three machine learning algorithms in order to produce a prediction, as the chart in figure 1 below shows.

II. Flow chart

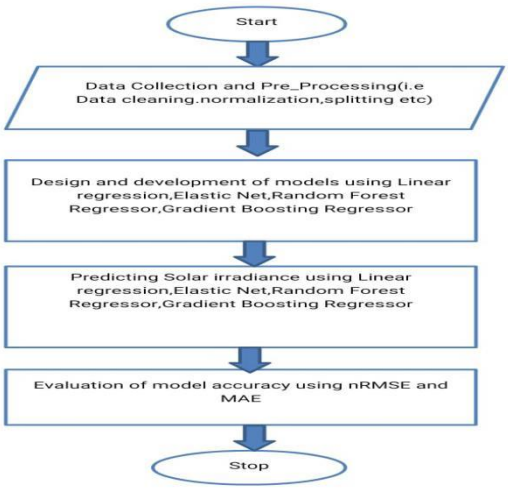

Fig 1 : Flow chart explaining process of predictions using ML models

A chart showing the general process used to generate predictions using the three ML models
Below is a brief description of the three machine learning models used to predict hourly solar irradiance of Johannesburg city:

Linear regression:

By fitting a linear equation to observed data, the statistical technique known as linear regression is used to model the relationship between a dependent variable and one or more independent variables. The objective is to identify the hyperplane or straight line that "best fits" the data to depict the relationship between the variables.

Simple linear regression is the most basic type of linear regression; multiple linear regression is the most complex type of regression involving multiple independent variables.

The equation of a simple linear regression model can be represented as:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Where:

- y is the dependent variable (the variable we are trying to predict).
- x is the independent variable (the variable used to make predictions).
- β_0 is the intercept (the value of y when x is 0).
- β_1 is the slope (the change in y corresponding to a one-unit change in x).
- ϵ represents the error term (the difference between the observed and predicted values of y).

The goal of linear regression is to estimate the values of β_0 and β_1 that minimize the sum of squared differences between the observed and predicted values of y, often referred to as the "least squares" criterion.

Random forest regression:

Random forest regression is a supervised learning algorithm that performs regression using the ensemble learning technique. In order to produce a prediction that is more accurate than one from a single model, the ensemble learning method combines predictions from several machine learning algorithms.

During the training phase, a Random Forest builds multiple decision trees, with the mean of the classes serving as each tree's prediction. Let's go through the steps of the Random Forest algorithm to better understand it:

Choose k data points at random from the training set. Construct a decision tree for these k data points.

For regression tasks, the Random Forest Regression is a well-liked ensemble learning technique. It is predicated on the idea of decision trees, in which various decision trees are trained on various data subsets and the average of their predictions is used to generate the final result.

A Random Forest Regression operates as follows:

Random Subset Selection: A random subset of the training data is sampled with replacement before each tree in the forest is constructed. We call this procedure "bootstrapping."

Random Feature Selection: A random subset of features is taken into consideration for splitting at each decision tree node. This strengthens the model's resistance to over fitting and aids in decor relation of the trees.

Tree Building: Every decision tree is expanded to its greatest depth or until a predetermined threshold is reached. The feature that maximizes the mean squared error or variance reduction is used to split the nodes.

Ensemble Prediction: Following the construction of each tree, predictions are generated by adding together the predictions of each **individual tree.** Predictions for regression tasks are usually averaged over all trees.

Gradient Boosting for Classification and Regression:

Loss Function (L): The difference between the actual and predicted values is measured by this function. The deviance (negative log-likelihood) or another suitable loss function, such as exponential loss for AdaBoost, is frequently used for classification tasks, whereas the mean squared error (MSE) is the typical loss function for regression tasks.

Weak Learner (h): A decision tree is usually used as a weak learner, and it is trained to predict the residuals, or the discrepancies between the ensemble's previous predictions and the actual values. The predictions of the weak learner are represented by *h*h(x).

Ensemble Prediction (F): This is the total of all the weak learners' predictions, weighted by a shrinkage parameter (learning rate). Every time a new weak learner is added to the ensemble, its prediction is first multiplied by the learning rate. This process is repeated for each iteration. The group forecast is represented by the symbol F(x).

Update Rule: The weak learner is trained to predict the loss function's negative gradient in relation to the ensemble prediction for each iteration. The pseudo-residual is the name given to this negative gradient. The predictions of the weak learner are then added to the ensemble prediction after being scaled by a factor known as the learning rate. Usually, the update rule is provided by:

$$F_{m+1}(x) = F_m(x) + \nu \cdot h_{m+1}(x)$$

where:

$F_{m+1}(x)$ is the ensemble prediction after adding the $(m+1)$-th weak learner,

$F_m(x)$ is the current ensemble prediction,

ν is the learning rate (shrinkage parameter),

$h_{m+1}(x)$ is the prediction of the $(m+1)$-th weak learner.

III. Result and discussion

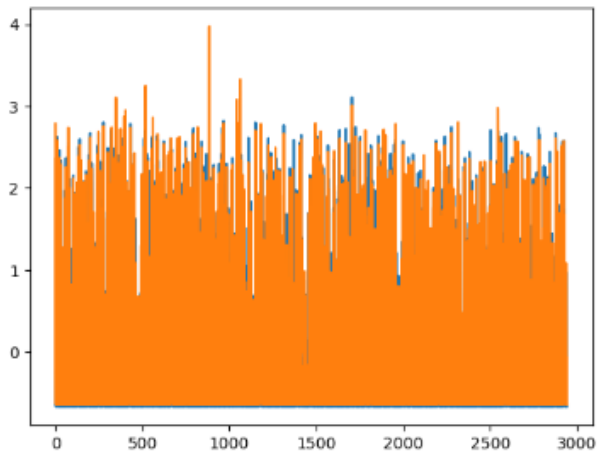

Fig 2 : Accuracy Graph

Classifier	Accuracy
Linear Regression	62.97%
Elastic Net	62.56%
Random Forest Regressor	93.44%
Gradient Boosting Regressor	88.50%

The comparison of regression models for predicting solar irradiance is crucial for ensuring the accuracy and reliability of solar energy forecasting systems. In this evaluation, four regression models were assessed: Linear Regression, Elastic Net, Random Forest Regression, and Gradient Boosting Regression.

Linear Regression and Elastic Net are traditional linear models commonly used for regression tasks. However, they may struggle to capture complex non-linear relationships present in solar irradiance data, leading to lower accuracy. As observed in this analysis, both Linear Regression and Elastic Net achieved accuracies of around 62-63%, indicating their limited performance in this context.

On the other hand, Random Forest Regressor and Gradient Boosting Regressor are ensemble learning techniques capable of capturing non-linear relationships and interactions within the data. Random Forest Regressor constructs multiple decision trees and combines their predictions to produce more robust results. Gradient Boosting Regressor sequentially builds an ensemble of weak learners, with each new model focusing on the mistakes of the previous ones, thereby improving accuracy.

The results of the evaluation demonstrate the superior performance of Random Forest Regressor and Gradient Boosting Regressor compared to linear models. Random Forest Regression achieved an impressive accuracy of 93.44%, indicating its ability to effectively model the complex patterns present in solar irradiance data. Gradient Boosting Regression also performed well with an accuracy of 88.50%, albeit slightly lower than Random Forest Regression.

In conclusion, for accurate prediction of solar irradiance, the Random Forest Regression emerges as the most suitable model among those evaluated. Its superior accuracy suggests that it can effectively capture the intricate relationships and variability inherent in solar irradiance data, making it a valuable tool for solar energy forecasting applications. However, the choice between Random Forest Regression and Gradient Boosting Regression may depend on factors such as computational resources, interpretability, and specific requirements of the forecasting task.

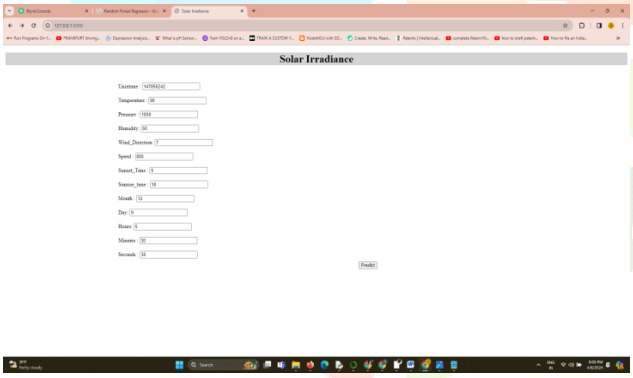

Fig 3 :User interface to provide inputs

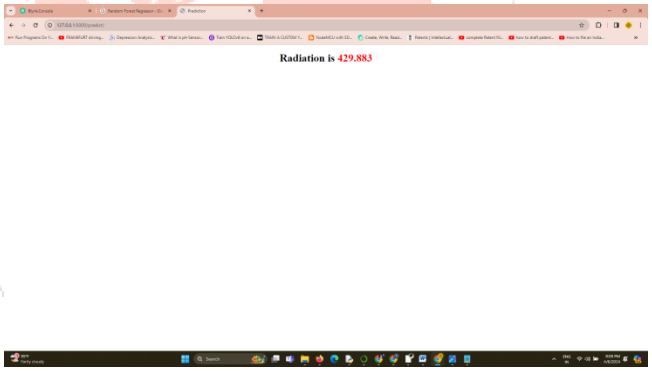

Fig 4 :output interface which predicts radiation

IV. CONCLUSION

The use of solar power has become more and more accepted worldwide in recent years as a clean, sustainable, and dependable replacement for fossil fuels. This momentum, though, can only be maintained if PV cell power generation efficiency is constantly increased. In addition to technological advancements such as the maximum power point tracking (MPPT) mechanism, which enables the system to maximize power output from each incoming radiative energy, there is a need for improved accuracy in solar irradiance prediction. The grid operators will be able to get over the uncertainties brought on by variations in the daily climate and weather patterns, which have a significant impact on solar power generation, by developing a trustworthy forecasting model.

V. REFERENCES

[1] A. M. Farayola, A. N. Hasan, and A. Ali, "Efficient photovoltaic MPPT system using Coarse Gaussian support vector machine and artificial neural network techniques," International Journal of Innovative Computing Information and Control (IJICIC), vol. 14, no. 1, 2018.
[2] R. Zhang, M. Feng, W. Zhang, S. Lu, and F. Wang, "Forecast of solar energy production-A deep learning approach," presented at the 2018 IEEE International Conference on Big Knowledge (ICBK), 2018, pp. 73–82.
[3] A. M. Farayola, A. N. Hasan, and A. Ali, "Optimization of PV systems using data mining and regression learner MPPT techniques," Indonesian Journal of Electrical Engineering and Computer Science, vol. 10, no. 3, pp. 1080–1089, 2018.
[4] J. D. M. do Rego and R. M. G. de Castro, "Solar irradiance forecast using artificial intelligence techniques".
[5]A. Moosa, H. Shabir, H. Ali, R. Darwade, and B. Gite, "Predicting solar radiation using machine learning techniques," presented at the 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1693–1699.
[6] R. Caballero, Physics of the Atmosphere, vol. 150. IOP Publishing, 2014.
[7] T. Shimazaki and L. C. Helmle, "A simplified method for calculating the atmospheric heating rate by absorption of solar radiation in the stratosphere and mesosphere," 1979.
[8] Meteoblue, "Weather history download South Africa", can be accessed in: https://www.meteoblue.com/en/weather/forecast/week /johannesburg_south-africa_993800.
[9] C. Yang and L. Xie, "A novel ARX-based multi-scale spatio-temporal solar power forecast model," presented at the 2012 North American Power Symposium (NAPS), 2012, pp. 1–6.
[10]B. Goswami and G. Bhandari, "Automatically adjusting cloud movement prediction model from 11.5 28.1 7.2 15.8 19.8 7.8 0 5 10 15 20 25 30 ANN Random Forest SVR nRMSE values (%) Models Comparison of nRMSE values 1st year error values 2nd year error values satellite infrared images," presented at the 2011 Annual IEEE India Conference, 2011, pp. 1–4.
[11]A. Radovan and Ž. Ban, "Predictions of cloud movements and the sun cover duration," presented at the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014, pp. 1210–1215.