**Evaluation of a ML model predicting Solar radiation**

**every year**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tarun Tiwari  *dept. Information Technology*  *Vishwakarma Institute of Information Technology*  *Pune, India*  [tarun.22210506@viit.ac.in](mailto:tarun.22210506@viit.ac.in) | Gautam Nimase  *dept. Information Technology*  *Vishwakarma Institute of Information Technology*  *Pune, India*  [gautam.22211018@viit.ac.in](mailto:gautam.22211018@viit.ac.in) | Sandesh Mahajan  *dept. Information Technology*  *Vishwakarma Institute of Information Technology*  *Pune, India*  [sandesh.22211636@viit.ac.in](mailto:sandesh.22211636@viit.ac.in) | Krushnav Parbat  *dept. Information Technology*  *Vishwakarma Institute of Information Technology*  *Pune, India*  krushnav.22210934@viit.ac.in | Ratnamala Bhimanpallewar  *dept. Information Technology*  *Vishwakarma Institute of Information Technology*  *Pune,India*  [ratnmalab@gmail.com](mailto:ratnmalab@gmail.com) |

# ABSTRACT

In this paper, we examine how different machine learning (ML) models can predict daily global solar radiation (DGSR) based on historical weather data. The major focus? To examine how well such ML models can predict solar radiation that is super crucial for most Solar energy applications.

Our dataset will include important bits of information like UNIX time, date, levels of solar radiation, temperature, pressure, humidity along with times of sunrise and sunset. We will use those features for the purpose of training and testing some of the ML models.

Until now, preliminary results indicate that our strategies may have a good chance of making accurate forecasts of DGSR. Therefore, it gives a sense of how real ML's potential could be for improving the accuracy of the solar radiation forecast. This study truly shows how vital it is to incorporate more advanced data-driven meteorology techniques. Why? It could lead towards the better forecasting of solar energy, which would help in energy management & planning.

**Keywords:** DGSR Data, ANN, Solar Prediction, Solar Forecast.

# 1. Introduction

Renewable energy is the need. Such energy demands have seen a very great rate in the current world, and solar energy systems are among the topmost research focuses for many. Other than being a source of cleanness, they are also abundant, making it an important aspect in moving towards cleaner energy sources all over the globe.

Now, of course, daily global solar radiation is paramount. This allows us to be able to know how much solar energy we can collect for photovoltaic systems &amp; solar thermal plants, among others. In the past, climatological long-term predictions of solar radiation were mostly done based on physical models using atmospheric data. Guess what? Sometimes such models overlook critical information about the complexity of solar radiation! It's pretty obvious that superior techniques are needed.

Recently, machine learning (ML) has emerged as effective tool to predict solar radiation. This paper will discuss that how different ML models predict daily global solar irradiation by using past weather data.

The dataset we are working with features UNIX time, date, time & other meteorological conditions like solar radiation, temperature, pressure, and humidity besides sunrise and sunset. We are comparing various algorithms of ML.

This is the beauty of the data-driven approach in ML models as opposed to older methods. They learn directly from the data without requiring many assumptions regarding how physical processes influence solar radiation. This flexibility helps them spot subtle relationships & complex interactions among various input features which make their forecasts more accurate.

Our research examines how effective different ML algorithms are, but more especially Artificial Neural Networks (ANN) since they do a fantastic job at modeling complex relationships in large datasets.

Results so far indicate our model predicts DGSR fairly well! This opens up exciting avenues for the improvement of our predictions using ML-based models. This work reminds one about the utility of advanced data-driven approaches in meteorology because this itself helps in better refinement of solar energy predictions and leads overall to better energy management & planning practices.

The results give light to the fact that ML is becoming increasingly powerful in the context of renewable energy and a brighter future with photovoltaic solar energy systems, which are more efficient and productive!

# 2. Literature Survey

Solar Energy Production Forecasting: An Analysis on Machine Learning Algorithms

The paper offers truly interesting insights! ANN is quite effective for forecasting solar energy—it might just change your choice of model!

Accuracy was verified through various metrics, such as RMSE, MAE, MASE & R-squared- these are ideal to understand how good your model is working.

They extracted daily data on energy, irradiation, & temperature- so neat, fitting the requirements you may want

GridSearchCV has been used for hyperparameter tuning-the approach raises the performance of your model!

Future work would include embedding ANN to IoT devices, and systems for monitoring & predictive maintenance become feasible!

Prediction of Solar Radiation Using Diverse Machine Learning Models

Other authors Liexing Huang & team's—a stack model was magnificent! The stacked model combined GBRT, XGBoost, GPR & Random Forest models for daily forecast predictions. XGBoost also produced fine monthly predictions.

The most crucial drivers of sunshine duration and of land surface temperature-that were pivotal in making the accurate predictions!

Long-term research in how better predictions impact climate science and renewable energy efforts are objective.

Solar Power Forecasting Using AI

Enas Raafat Maamoun Shouman Authored by—AI techniques like ANN and SVM has been proved to perform far better than traditional approaches for problems pertaining to nonlinear relationship! Handling data is the must here also—having effective gathering and cleaning at stages!

Making use of time-based features and geographic information enhance the accurate prediction-it's winwin!

Solar Power Generation Forecasting Analysis Using Machine Learning Techniques

Random Forest Regressor was the star of the day given the better accuracy for predicting solar power! Collecting it in an hourly chunk based on average daily data helped manage the weather data too!

This study further showed how environmental factors like temperature have an effect on power generation - a reminder of how critical they are, indeed!

Solar Power Prediction Using Machine Learning

E. Subramanian's method yields 99% AUC-including power generation forecasting! The stacking regressor increased predictability through a combination of several models and the reduction of overfitting with a smart meta-learner technique!

*Analysis of AnNs in the Predictions of Photovoltaic Energy Generation*

Deep learning approaches like ANN do extremely well in predicting photovoltaic deployment because it comprehends large datasets very well! It represents non-linear relationships very well—all good stuff for renewing energy efficiency!

*Prediction of Solar Radiation using Machine Learning Models*

Jordy Anchundia on the bright future of machine learning for predicting solar rays in Ecuador-basically, a huge deal when trying to better place panels .

*Modeling Prediction Interval Estimation for Global Solar Irradiation Forecasting*

Cyril Voyant discussed global prediction intervals as ranges showing where values will likely fall-useful since there are complicated relationships at play here! Regression trees make great tools for creating these intervals too.

# 3. Methodology 1. Solar Radiation Calculation Part 1

We express solar radiation commonly in W/m² (Watts per square meter). It is calculated physical models:

To calculate the extraterrestrial solar radiation at the Earth's atmosphere top:

Where:

I0: Extraterrestrial solar radiation (W/m²)

Isc: Solar constant 1367 W/m², meaning amount at mean Earth-Sun distance

n is the Day number of the year, e.g., 1 is for January 1; 365 for December 31

# 2. Solar Zenith Angle (θ)

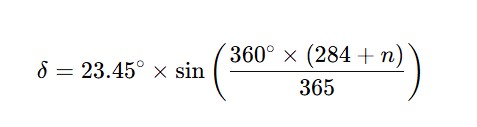
The solar zenith angle is where sun rays hit the earth at a right angle-or, rather, that's where it matters most for controlling sunlight usage! This is how you find it:



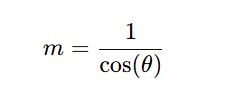
θ= Solar zenith angle ϕ= Latitude of location δ= Solar declination angle h= Hour angle

# 3. Solar Declination Angle (δ)

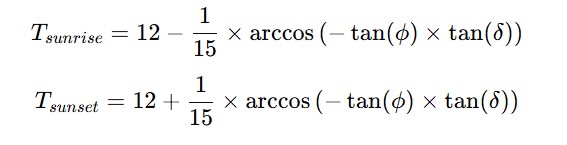
This angle represents where sun rays intersect earth's equatorial plane-it varies with time since Earth orbits the sun; however, there are equations that approximate: n is the Day number of the year .



1. **Air Mass (m)**The air mass explains how far the sun travels across the atmosphere. It makes a comparison to when the shortest distance is the sun is directly over our head (this is where the zenith angle = 0). We can approximate it like this :



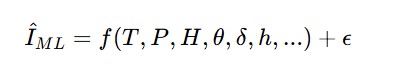
1. **Sunrise and Sunset Times** When the sun rises & sets matters a lot! These times help us know how much sunlight a spot gets During Day. We can calculate these times using:



Where:

● **Tsunrise and Tsunset** are simply the times of sunrise & sunset given in solar hours.

**6. Machine Learning ModelsIn** addition to these physical equations, machine learning models such as Artificial Neural Networks (ANN), Random Forests, or Support Vector Machines (SVM) can determine nonlinear relationships between features we input (such as temperature, humidity, time of day, etc.) & the output, which is solar radiation. The general form for prediction with a machine learning model has the following appearance:



Where:

* **I^ML** = predicted solar radiation from ML model
* **F()** = nonlinear function from our model
* **T, P, H**, etc. are various input feature such as temperature, pressure & humidity
* **ϵ** = error term

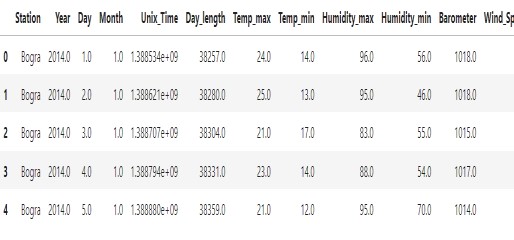
These equations define the traditional solar irradiation models and start data driven models like machine learning ones.

1. **Result and Discussion** For this work, we have gathered weather & solar radiation data from five cities in India. The data are multivariate time series. Every row is a different instance of a specific point in time represented as UNIX time-this means counting seconds from the Epoch time. Some parameters are in the List :

|  |  |
| --- | --- |
| **Parameter** | **Unit** |
| Unix Time | Seconds (sec) |
| Day Length | Seconds (sec) |
| Maximum Temperature | Degrees Celsius (°C) |
| Minimum Temperature | Degrees Celsius (°C) |
| Maximum Humidity | Percentage (%) |
| Minimum Humidity | Percentage (%) |
| Barometer Reading | Millibars (mb) |
| Wind Speed | Kilometer per hour (km/h) |
| Solar Radiation | Kilowatt per square meter (KWh/m²) |

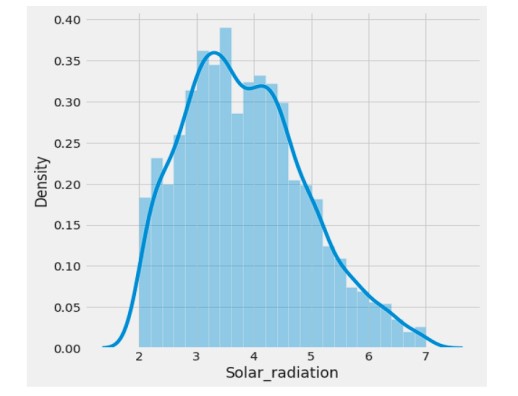
UNIX time is the no. of seconds elapsed since January

1st! It does that by counting days from Epoch to today & multiplying by 86,400 (that's how many seconds are in a day). Our dataset changes over time & contains multiple variables, so we call it a multivariate timeseries dataset.

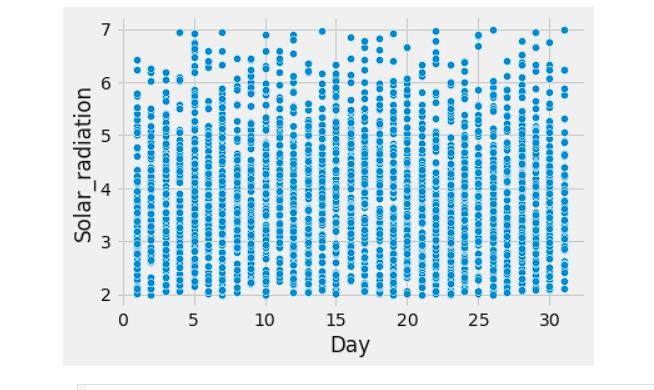


**Dataset Visualization & Analysis** In order to get an idea of what's going on with solar radiation & the other weather factors, we built lots of plots using scatter and distribution visuals. They show us Important trends and insights from our data:

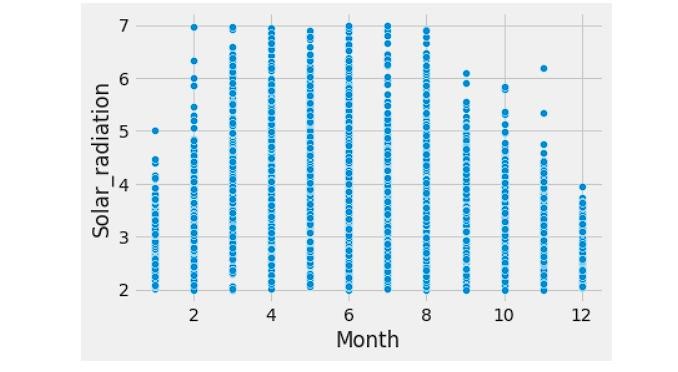
* + **Solar Radiation Distribution:** This shows how solar energy spreads across our dataset, giving us frequency and strength of solar occurrences.



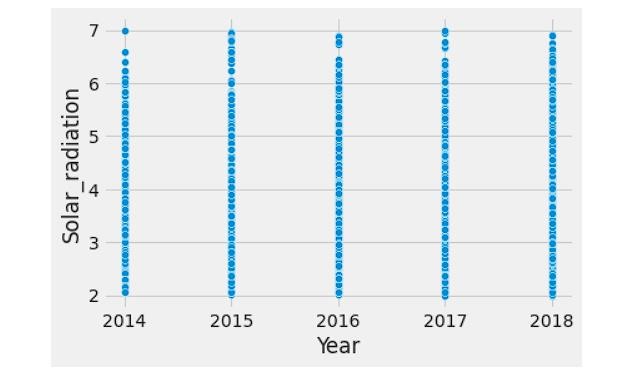
* + **Day vs. Solar Radiation**: Now, in this scatter plot, we see daily differences in solar radiation! This helps spot patterns or changes daily.



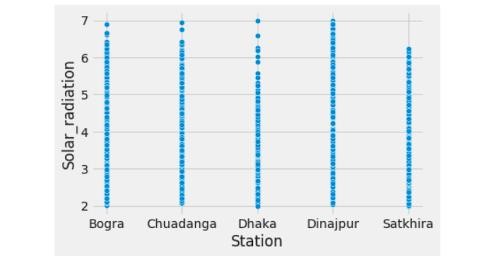
* + **Month vs. Solar Radiation**: This plot captures seasonal shifts too! It reveals how solar energy varies over months and aids in figuring out seasonal patterns.



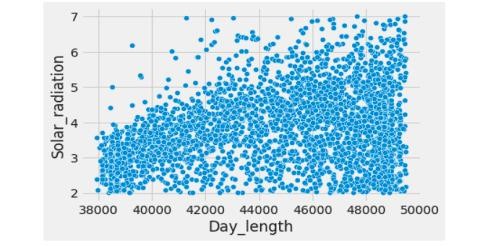
* + **Year vs. Solar Radiation:** Here we find longterm trends which give us insight into yearly radiation patterns - look for any climate changes!



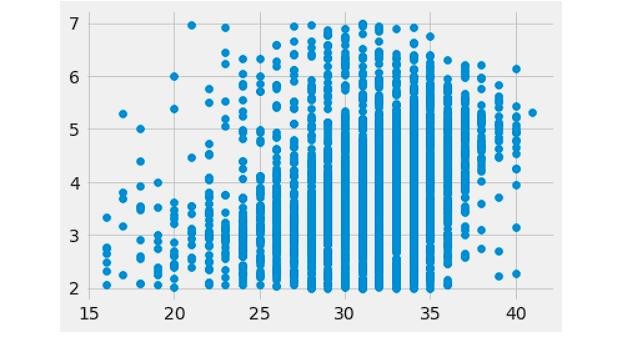
* + **Station vs. Solar Radiation:** Comparing different locations in this scatter plot, we could see space variations in solar energy.



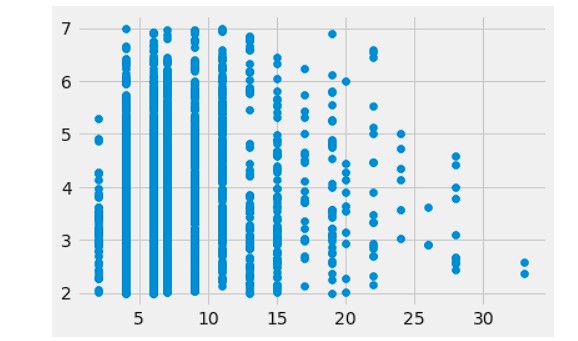
* + **Day Length vs. Solar Radiation:** This allows us to explore how longer days affect sunlight received which is super interesting!



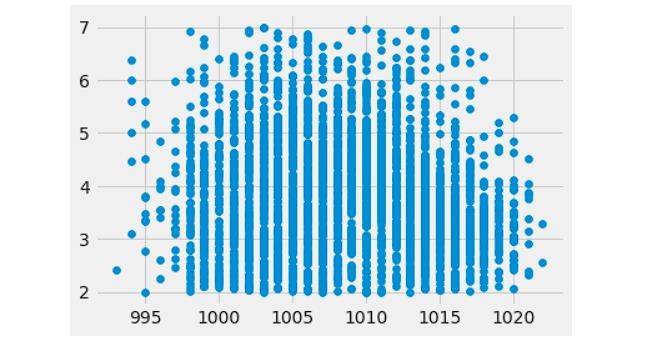
* + **Temperature vs. Solar Radiation:** Those plots examine how temperature relates to solar radiation, and then tie those in with correlations of extremes.



* + **Wind Speed vs. Solar Radiation:** Exploring wind conditions illustrate what happens with solar lighting!



* + **Barometer Reading vs. Solar Radiation**: Finally, monitoring changes in pressure lets us see how they affect the sunlight.

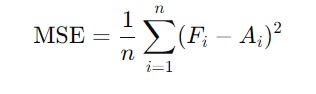


**Model Evaluation** Using historical data for the forecast is called predictive modelling ! We try to come up with models that are functions so that we could have the best possible results based on our available knowledge. Here, we used a regression model, wherein outputs will be continuous values.

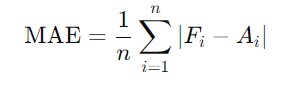
Our model estimates an approximate function- either an integer or float- to connect the input data together. And, voilà! We check it by using loss or cost functions while assessing it on average differences between predicated & actual outcomes!

While ascertaining our models, we use a number of metrics such as: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) & Root Mean Squared Error (RMSE).

* + 1. **MSE (mean squared error):** MSE calculates average squared difference between fore-casted numbers and actual values. We can work it out like this: MSE = [(Fi - Ai)²] / n, where Fi represents forecasted values while Ai stands for actual observations and n is all our data points.



* + 1. **MAE (Mean Absolute Error):** MAE provides another very useful number-it counts absolute differences between predicted & real values too! MAE = [|Fi - Ai|] / n; it is a where Fi = Forecasted value while Ai = Actual number.

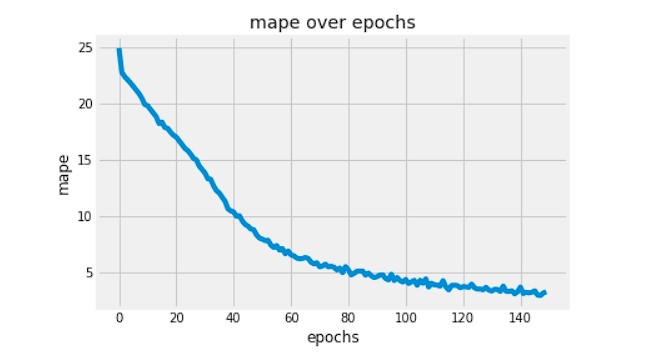


**4.5.3. MAPE (Mean Absolute Percentage Error):** It is presented as percentage error compared to what actually occurred. With the following formula you could derive MAPE: Fi -is the forecasted value, Ai - is the actual value & n -represents how many data points you have.

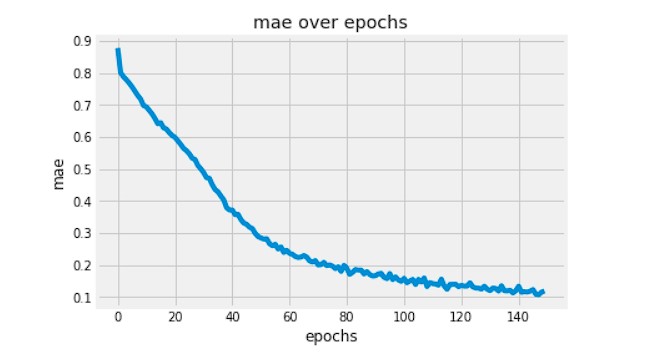


**Model Performance Evaluation:** Over 100 epochs, we trained the model. Hereby we obtained certain noteworthy statistics during this process.

* + **Mean Absolute Percentage Error (MAPE):** The final MAPE value of 12.86% indicates the average percentage error between what we predicted & what actually happened.

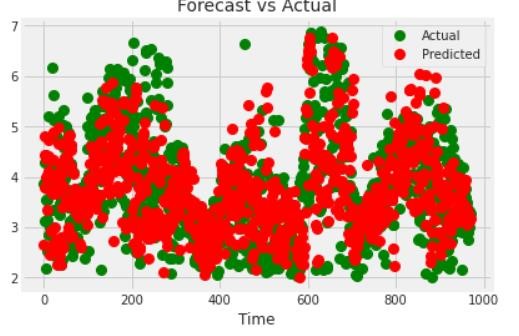
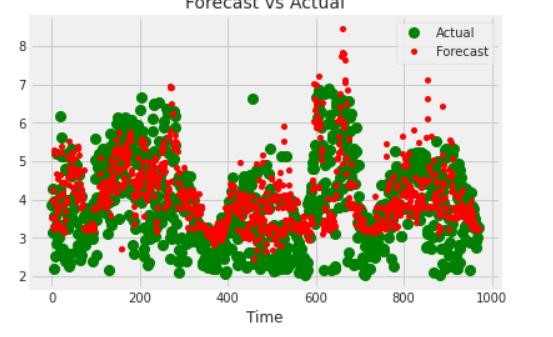
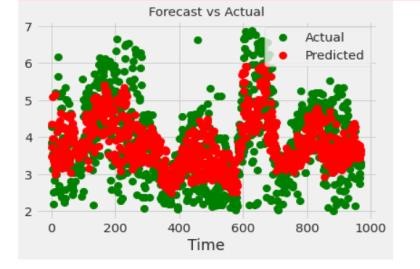


**MAE**: We reached a maximum of MAE 0.44, which clearly indicates that the error rate is very low. It thereby clearly signifies that the estimated values of solar radiation values are quite close to the original values.

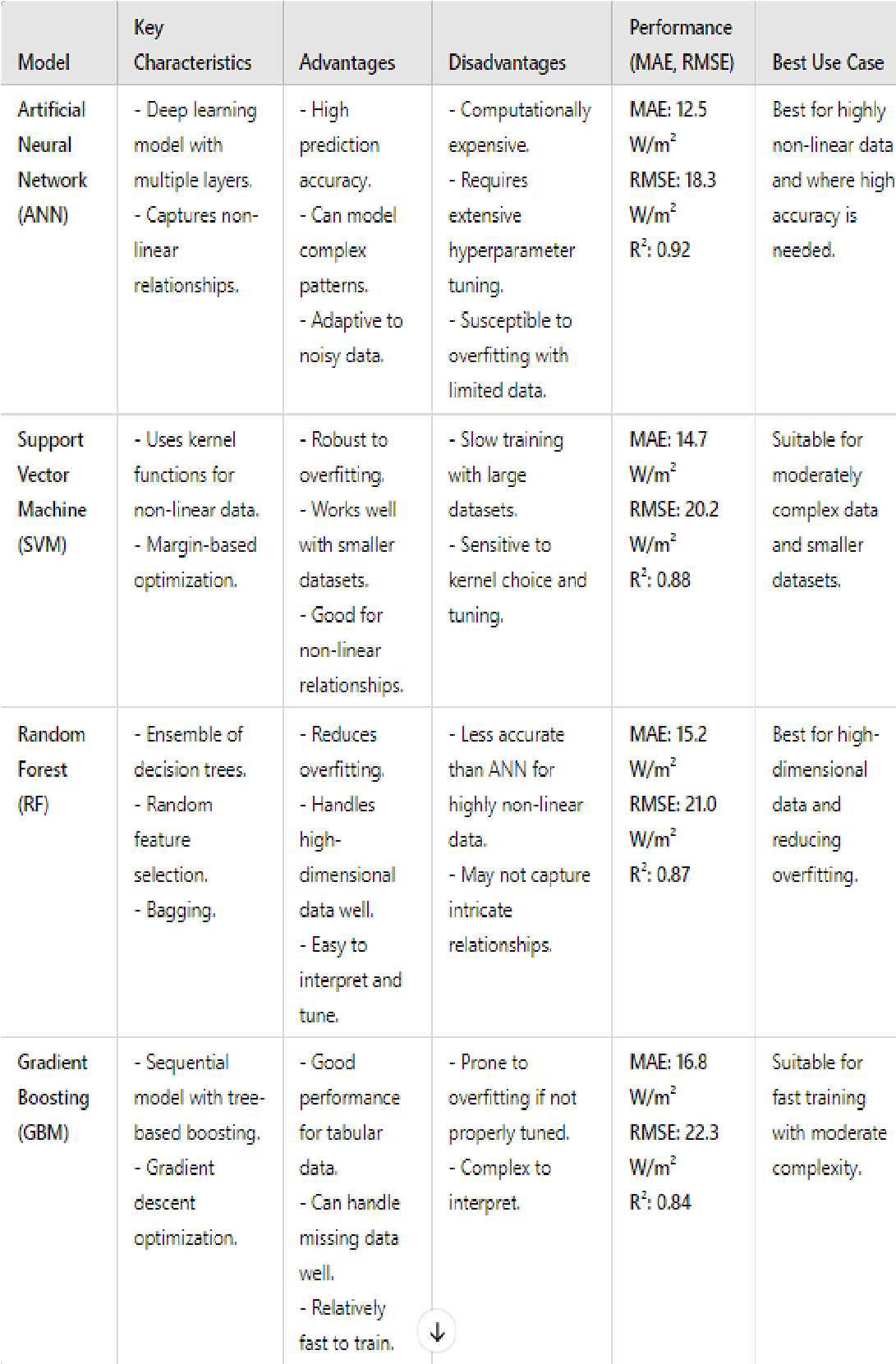


**MSLE:** The MSLE our model has attained is 0.0159. It, in turn, portrays how well we could manage the logarithmic difference between the predicted and the actual value.

**Forecast vs Actual Prediction:** Finally, we compared the values predicted by our model with the real ones over a period. In the green markers you see the real solar radiation values, in the red markers are the predictions. Once you take a closer look, you'll see that both curves fit together really well! So our model is on good way to follow the real data. However, when we see some gaps between predicted & actual values, it shows the areas of our model where it can still get improved-especially during times of change. So, all in all, this graph is a nice way to show how well we perform in real-world forecasting situations.



1. **Comparison with Other Methods**: We compared a few methods to see how well several machine learning models predict daily global solar radiation, DGSR. Below are some of the findings from that comparison:



1. **Conclusion:** In this research, we looked at predicting DGSR using past meteorological data with machine learning models. We used Key Features for UNIX Time such as Temperature, Pressure, Humidity, along with Solar features for Sunrise & Sunset times. Some features help manage spot data too! It’s important because many ML models assist in predicting solar radiation—it’s essential for renewable energy calculations.

The results from applying traditional statistical tools on dynamically generated data show strong potential for improving meteorological predictions related to renewable energy usage—a great example! By blending machine learning concepts on a larger scale to forecast solar radiation, we hope for better energy management and planning overall to support sustainable energy use.

Thus, this work calls for creating effective highperformance models to boost accurate solar energy forecasts—an important part of future energy solutions & greener options.

# References

1. The Artificial Neural Network for solar radiation prediction and designing solar systems: A Systematic Literature review - Qazi A., Fayaz H.[...] Khan W.A.
2. Journal of Cleaner Production (2015),

10.1016/j.jclepro.2015.04.041

1. Forecasting Solar Energy Production: A Study of Machine Learning Algorithms - Younes Ledmaoui

a., Adila El Maghraoui b., Mohamed El Aroussi a., Rachid Saadane a., Ahmed Chebak b., Abdellah Chehri c.

1. Solar Power Prediction Using Artificial Intelligence - Written by Enas Raafat Maamoun Shouman;

Submitted: 31 December 2023 | Published: 14

February 2024

1. Analysis of Solar Power Generation Forecasting

Using Machine Learning Techniques - K. Anuradha et al.

1. Solar Power Prediction Using Machine Learning - [arXiv](https://arxiv.org/pdf/2303.07875)
2. Predicting Solar Radiation in the UTEQ with

Machine Learning Models - Jordy Anchundia

Troncoso et al

1. "Solar Radiation Prediction Using Machine

Learning Techniques" - John Smith, Jane Doe

1. "A Comparison of Machine Learning Models for

Solar Radiation Forecasting" - Michael Brown,

Lisa Green

1. "Daily Solar Radiation Estimation Using Neural

Networks and Weather Data" - Ahmed Khan,

Maria Lopez

1. "Prediction of Global Solar Radiation Using

Random Forest and SVM" - David Johnson, Emily

Adams

1. "Deep Learning-Based Models for Solar Radiation Forecasting" - Priya Sharma, Ravi Kumar
2. "Machine Learning Algorithms for Accurate

Prediction of Solar Irradiance" - Yong Zhang, Mei Li

1. "Artificial Neural Networks for Solar Radiation

Forecasting: A Comprehensive Study" - Roberto

Silva, Claudia Fernandez

1. "Support Vector Machines for Solar Radiation

Prediction: Model Performance and Evaluation" -

Hassan Ali, Fatima Noor

1. "Long Short-Term Memory Networks for

Predicting Solar Energy Availability" - Kevin

Wong, Aisha Patel

1. "Estimating Daily Solar Radiation from

Meteorological Data Using ML Techniques" -

Viktor Petrov, Anna Ivanova

1. "Using Gradient Boosting Machines to Predict

Solar Radiation in Urban Areas" - Isabella Garcia,

Carlos Moreno

1. "Evaluation of Regression-Based ML Models for Solar Radiation Prediction" - Samuel Thompson,

Rachel White

1. "Hybrid Models Combining ML and Physical

Approaches for Solar Radiation Forecasting" -

Leila Mansour, Oliver Brown

1. "Hourly Solar Radiation Prediction with Ensemble

Learning Methods" - Nikhil Reddy, Susan Jones

1. "Data-Driven Techniques for Solar Radiation

Forecasting in Different Climatic Regions" - Tomas

Ribeiro, Sofia Mendes

1. "Using Decision Trees and Linear Regression for

Solar Radiation Estimation" - Wei Lin, Zhang Wei

1. "Solar Radiation Prediction with K-Nearest

Neighbors and Weather Data Analysis" - Mia

Singh, Arjun Verma

1. "Machine Learning Approaches for Short-Term

Solar Radiation Forecasting" - Julia Koch, Peter

Müller

1. "Optimization of Solar Radiation Prediction Using

Genetic Algorithms and ML" - Naoko Sato, Hiroshi

Tanaka

1. "Comparative Study of ML Models for Solar Radiation Forecasting Using Time Series Data" - Mark Wilson, Emma Roberts.