iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XI **Month of publication:** November 2022

DOI: https://doi.org/10.22214/ijraset.2022.47345

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

Forecasting Solar Radiation by the Machine Learning Algorithm & their Different Techniques

Yogesh Singh[1], Amarendra Singh[2]
[1]*M.Tech Scholar,* [2]*Assistant Professor, Department of Civil Engineering, Institute of Engineering and Technology, Lucknow-226021 UP, India*

Abstract: ***The objective of this study is to give a summary of machine learning-based techniques for solar irradiation forecasting in this context. Despite the fact that numerous research describe methods like neural networks or support vector regression. Ranking the performance of such methods is difficult because of the diversity of the data collection, time step, forecasting horizon, setup, and performance indicators. The prediction inaccuracy is quite comparable overall. Others write. Global solar radiation recommended utilising ensemble forecasting or hybrid models to improve prediction accuracy. Forecasting the output power of solar systems is required for the smooth operation of the power grid or for the optimal control of the energy flows into the solar system. Prior to projecting the output of the solar system, it is essential to focus on solar irradiance. The two primary categories of methods for predicting the global solar radiation are machine learning algorithms and cloud pictures combined with physical models.***

Keywords: solar radiation Prediction, Random Forest Method, and gradient boosting Method.

I. INTRODUCTION

For many uses, including meteorology, hydrology, and especially the development and use of renewable solar energy systems, the global solar radiation that reaches the Earth's surface is of essential importance. However, it is not common to get direct measures of global solar radiation, particularly in developing nations. This is most likely because installing and maintaining the measuring equipment is expensive and complicated. Different methods have been developed to estimate global solar radiation because observed data are not always available. These methods include empirical models that establish linear and nonlinear relationships between meteorological variables and global solar radiation and machine learning models that simulate the complex and nonlinear mapping from meteorological variables to global solar radiationas well as satellite-based techniques for tracking the spatiotemporal variations in solar radiation on both a global and regional scale, as well as radiative transfer models to simulate the scattering and absorption of solar radiation in the atmosphere. Additionally, there are worldwide databases like Meteonorm, SolarGIS, and NASA-SSE that provide data on large-scale global sun radiation (Surface meteorology and Solar Energy). Due to their low computing costs and high prediction accuracy, respectively, empirical and machine learning models are more frequently utilised in practise among the aforementioned methodologies. The isotropic models and anisotropic models can further predict global solar radiation on PV panel surfaces with specific tilt angles based on the horizontal global solar radiation. The prediction accuracy of various types of machine learning models, particularly their computational efficiency on large-scale datasets for predicting Global solar radiation, have rarely been compared in different parts of the world. In general, machine learning models provide more accurate predictions of Global solar radiation than empirical models do. For instance, Wang et al. only evaluated the prediction accuracy of three ANN models, including the MLP, RBF, and GRNN models, for daily Global Solar Radiation estimation with Sunshine Duration and other Meteorological Variables at 12 Stations in China.. The MLP and RBF models were found to perform better than the GRNN model. At three locations in China's Hunan Province, Zou et al. tested the effectiveness of the ANFIS model for forecasting daily Global solar radiation in comparison to two empirical models (such as the Bristow-Campbell Model and Yang's Hybrid Model). The ANFIS model was shown to provide estimates of global solar radiation that were more accurate than the two empirical models. Wang et al. also contrasted the ANFIS and M5Tree models for daily estimation of the global solar radiation at 21 locations around China. According to the findings, the ANFIS model outperformed the M5Tree and empirical models. Additionally, Fan et al. contrasted two machine learning models (such as the Gradient Boost Method) for the daily forecast of global solar radiation in humid subtropical China. They discovered that the Gradient boost method models outperformed the investigated empirical models, and due to greater model stability, efficiency, and comparable prediction accuracy, they proposed the Gradient boost model as a potential machine learning model for global solar radiation estimation.

©IJRASET: All Rights are Reserved

II. METHEDOLOGY

1) *Data Collection:* Twelve photos per day were utilised in the satellite computations. The satellite data and other data used in this study are the same as those used for images from Meteosat 5 and 7 (HRI-VIS channel).
2) *Data Cleaning:* The original data obtained has several gaps that must be filled for the subsequent analysis to be successful. In order to account for the missing units, we eliminate the columns from which a considerable quantity of data is missing as well as those particular samples whose integrity is compromised.
3) *Feature Selection:* The vast majority of useless elements in the acquired data should be eliminated. The majority of the characteristics won't help the prediction's outcome; some of them might even make things worse. The filtration is essential for those elements that indicate certain incredibly irrelevant aspects. A mathematical analysis would be more effective and compelling than just identifying each aspect and estimating their relativeness by examining their meanings.

A. Machine Learning Method

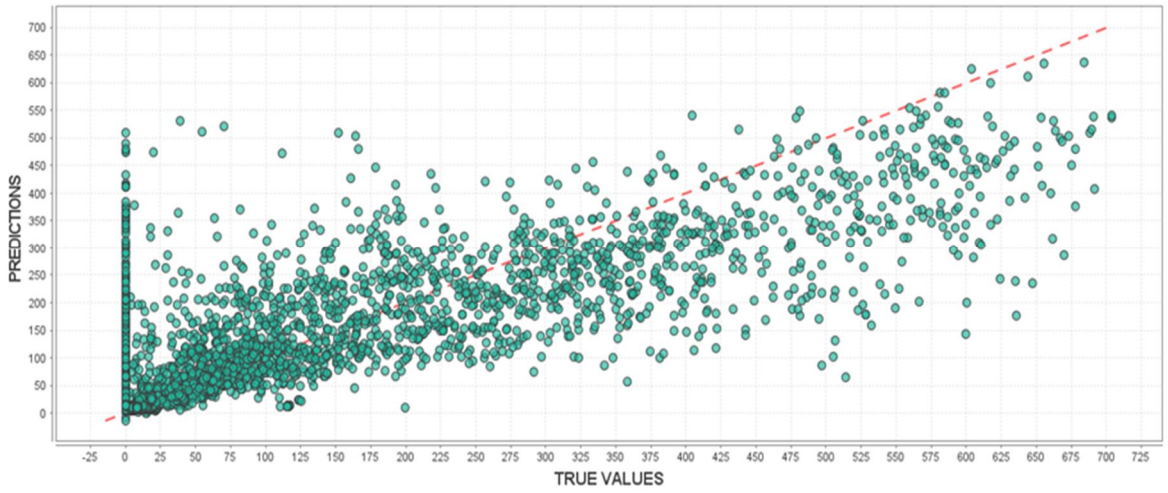
A branch of computer science known as machine learning is categorised as an artificial intelligence technique. The benefit of this approach is that a model can handle problems that are impossible to represent by explicit algorithms. The reader can get a full review of some deterministic and machine learning approaches to solar forecasting. In various situations, such as pattern recognition, classification issues, spam filtering, as well as data mining and forecasting issues, machine learning models can be used because they can identify relationships between inputs and outputs even when a representation is unavailable. Because one must work with large datasets and machine learning models are capable of pre-processing and data preparation, categorization and data mining are particularly intriguing in this field. The machine learning models can then be used to forecasting issues after this stage. Machine learning is a subset of artificial intelligence, as was already explained. It focuses on the development and research of systems that can learn from data sets, enabling computers to learn without explicit programming. The methods listed below are used to forecast solar radiation.

1) *Gradient Boosting Method:* It is a powerful machine learning technique which produces complex model using ensembles of decision trees. Gradient boosting consists of three components they are loss function which is to be optimized, a weak learner and an additive model to make the weak leaner to minimize the loss function. Gradient Boosting Decision Tree is a kind of ensemble algorithms which is based on multiple weak classifiers. The result was then assayed by weighting method. Weak classifiers are commonly adopting Regression Decision Tree employing several times' iteration. Every weak classifier is trained based on previous weak classifier's residual error in which way GBDT can reach the classification target by decreasing the residual error in the training process. Gradient Boosting Decision Tree is a sort of boosting machine learning method. Therefore, we adopt the addition model and forward stagewise algorithm. The following formula represents the mth weak classifier, which the M represents the number of the classifiers and indicates the parameter of specific classifiers.
2) *Random Forest Learning:* Random Forest Tree is a classification algorithm that was introduced by Tin Kam in 1995 and improved later by Leo Breiman in 2001. It uses the bagging technique and operates by constructing simultaneously many independent decision trees during the training phase using feature randomness to build an uncorrelated forest of trees. For the classification of a new instance, the random forest classifier collects the classification from each individual decision tree then combine them using either the majority vote or the confidence vote strategy to have a more accurate prediction than that of any individual tree, based on the idea that the combination of learning models improves prediction accuracy. It is also called random decision forests which are method for ensemble learning for classification and regression it includes multiple randomized trees.
3) *Support Vector Method:* SVM is powerful machine learning algorithm for supervised learning it is relatively fast model, especially for non-linear relationships. Support vector machine is another kernel based machine learning technique used in classification tasks and regression problems introduced by Vapnik in 1986. Support vector regression (SVR) is based on the application of support vector machines to regression problems. This method has been successfully applied to time series forecasting tasks.

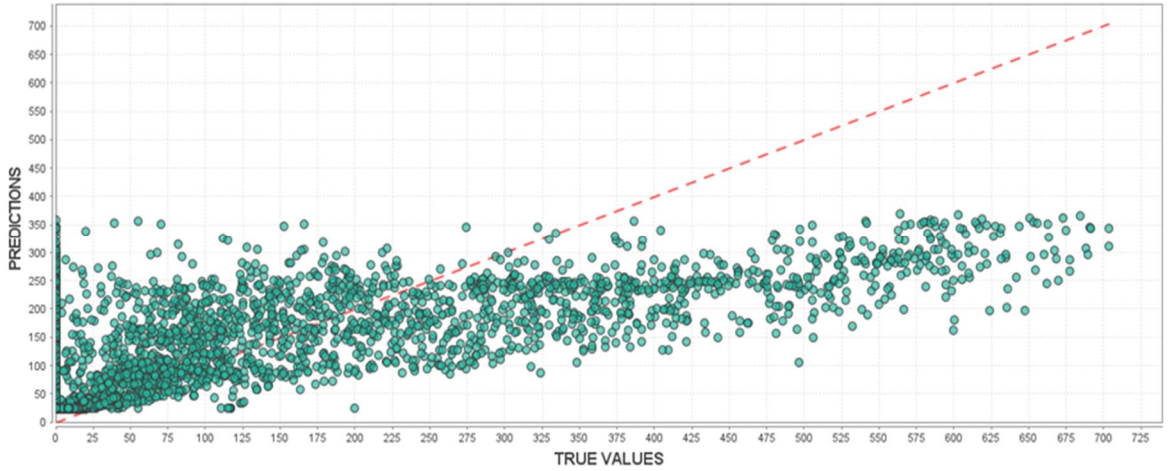
III. RESULT AND DISCUSSION

We claim that machine learning algorithms are more effective and efficient than other traditional methods as evidenced by the results obtained. The following are some intriguing outcomes we obtained after utilising various machine learning algorithms, including Deep Learning, Decision Tree Learning, and the Generalized Learning Method.

A. Prediction Charts



Gradient Boosting Method



Random Forest Learning

B. Some Predictive Values

1) Gradient Boosting Method

Solar Radiation (SR)	3.77	3.58	4.45	156.49	151.06	4.05	4.02	4.19
prediction(SR)	9.506987	6.90012	9.419813	200.341	90.16443	10.89493	9.196775	6.787829

2) Random Forest Learning

Solar Radiation (SR)	3.77	3.58	4.45	156.49	151.06	4.05	4.02	4.19
prediction(SR)	25.20451	25.09222	25.01659	147.7936	140.4441	23.95128	25.09281	25.06323

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

C. Validation

In order to determine the correctness of prediction models, we use a process called validation in which we discover many validation points, such as r square, squared error, relative error value, root mean square error value, and absolute error value. In order to choose the optimal prediction model for our task, validation gives us a clearer picture of our models. Validation only determines the error values' accuracy in the outcomes.

MODEL	R SQUARE VALUE	SQUARED ERROR VALUE	RELATIVE ERROR VALUE	ROOT MEAN SQUARE ERROR VALUE	ABSOLUTE ERROR VALUE
GRADIENT BOOST METHOD	0.692710915	164.6313863	0.001602392	1.002406655	0.542154811
RANDOM FOREST METHOD	0.606190927	146.5791903	0.001793453	0.770415784	0.305207387
SUPPORT VECTOR MACHINE	0.555571952	9070.79001	0.00419110147	0.226723317	0.038680899

The mean absolute error (MAE) is suitable for applications with linear cost functions, i.e., when the costs associated with inaccurate forecasting are inversely correlated with forecast error:

$$MAE = \frac{1}{N} \times \sum_{i=1}^{N} |\hat{y}(i) - y(i)|$$

The square of the difference between the observed and anticipated values is used to calculate the mean square error (MSE). The highest gaps are penalised by this index:

$$MSE = \frac{1}{N} \times \sum_{i=1}^{N} (\hat{y}(i) - y(i))^2$$

MSE is often the parameter that the training method strives to minimise..

The root mean square error (RMSE) is more sensitive to big forecast errors, and hence is suitable for applications where small errors are more tolerable and larger errors cause disproportionately high costs, as for example in the case of utility applications. It is probably the reliability factor that is most appreciated and used:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (\hat{y}(i) - y(i))^2}$$

©IJRASET: All Rights are Reserved

The mean absolute percentage error (MAPE) and the mean absolute error (MAE) are nearly identical; however, the relative gap is taken into account by dividing each gap between observed and forecasted data by observed data.

$$MAPE = \frac{1}{N} \times \sum_{i=1}^{N} \left| \frac{\hat{y}(i) - y(i)}{y(i)} \right|$$

D. Correlation & Heat Map

Correlation refers to a process for establishing the relationships between two variables. You learned a way to get a general idea about whether or not two variables are related, is to plot them on a "scatter plot". While there are many measures of association for variables which are measured at the ordinal or higher level of measurement, correlation is the most commonly used approach.

Correlation is a statistical term describing the degree to which two variables move in coordination with one another. If the two variables move in the same direction, then those variables are said to have a positive correlation. If they move in opposite directions, then they have a negative correlation. Correlation means association - more precisely it is a measure of the extent to which two variables are related. There are three possible results of a correlational study: a positive correlation, a negative correlation, and no correlation. A heatmap (or heat map) is a graphical representation of data where values are depicted by color. They are essential in detecting what does or doesn't work on a website or product page. By experimenting with how certain buttons and elements are positioned on your website, heatmaps allow you to evaluate your product's performance and increase user engagement and retention as you prioritize the jobs to be done that boost customer value. Heatmaps make it easy to visualize complex data and understand it at a glance.

AOD	1	0.1156582	-0.2163769	-0.0931418	-0.078386	-0.0673811	-0.0789726	-0.1753665	0.05362555	-0.00508322	0.06494505	-0.06981108	-0.0616833	-0.12487762	0.05044609	-0.14965603
Clearsky DHI	0.1156582	1	0.8195505	0.9087418	-0.022349	0.0254967	0.0859577	0.0620782	0.00846146	-0.3501923	-0.85520894	0.65133031	0.09597008	0.52224456	0.09727953	0.38491072
Clearsky DN	-0.2163769	0.8195505	1	0.9513638	-0.0247418	0.0106793	0.0514294	0.078879	0.01085776	-0.32392968	-0.86293966	0.68826883	0.05163374	0.48642135	0.06770632	0.47134359
Clearsky GHI	-0.0931418	0.9087418	0.9513638	1	-0.0045977	0.0424119	0.0888235	0.1086804	-0.0010634	-0.33621198	-0.88342734	0.73058984	0.09323225	0.54562195	0.06743585	0.45906151
Cloud Type	-0.078386	-0.022349	-0.0247418	-0.0045977	1	0.5514785	0.0388929	0.671473	-0.06823717	0.41593515	-0.02715384	-0.09268988	0.02125617	0.18175607	-0.24860391	0.06781352
Dew Point	-0.0673811	0.0254967	0.0106793	0.0424119	0.5514785	1	0.0790115	0.8932685	-0.0821864	0.72392548	-0.0974824	-0.02191343	-0.02581817	0.32035309	-0.32426189	0.04808943
Ozone	-0.0789726	0.0859577	0.0514294	0.0888235	0.0388929	0.0790115	1	0.1382874	0.40635329	-0.19160073	-0.06980094	0.06089908	0.56583333	0.44915015	-0.05214507	0.1813284
Precipitable	-0.1753665	0.0620782	0.078879	0.1086804	0.671473	0.8932685	0.1382874	1	-0.10881272	0.57069439	-0.15719059	-0.0010428	0.06634872	0.43152971	-0.35933153	0.16174286
Pressure	0.0536256	0.0084615	0.0108578	-0.0010634	-0.0682372	-0.0821864	0.4063533	-0.1088127	1	0.03779487	0.10734316	-0.00140703	0.33189857	-0.04640991	0.12615097	-0.01147712
Relative Hun	-0.0050832	-0.3501923	-0.3239297	-0.336212	0.4159352	0.7239255	-0.1916007	0.5706944	0.03779487	1	0.30958211	-0.28878504	-0.32041459	-0.36026699	-0.20580168	-0.21182194
SR	-0.0698111	0.6513303	0.6882688	0.7305898	-0.0926899	-0.0219134	0.0608991	-0.0010428	-0.00140703	-0.28878504	-0.63680419	1	0.06618687	0.37610662	0.04669985	0.29789212
Solar Zenith	0.0649451	-0.8552089	-0.8629397	-0.8834273	-0.0271538	-0.0974824	-0.0698009	-0.1571906	0.10734316	0.30958211	1	-0.63680419	-0.08561266	-0.55798425	-0.04398838	-0.43739583
Surface Albe	-0.0616833	0.0959701	0.0516337	0.0932322	0.0212562	-0.0258182	0.5658333	0.0663487	0.33189857	-0.32041459	-0.08561266	0.06618687	1	0.46521356	-0.05439022	0.23074947
Temperature	-0.1248776	0.5222446	0.4864214	0.5456219	0.1817561	0.3203531	0.4491501	0.4315297	-0.04640991	-0.36026699	-0.55798425	0.37610662	0.46521356	1	-0.11145644	0.40748901
Wind Directi	0.0504461	0.0972795	0.0677063	0.0674358	-0.2486039	-0.3242619	-0.0521451	-0.3593315	0.12615097	-0.20580168	-0.04398838	0.04669985	-0.05439022	-0.11145644	1	0.0041914
Wind Speed	-0.149656	0.3849107	0.4713436	0.4590615	0.0678135	0.0480894	0.1813284	0.1617429	-0.01147712	-0.21182194	-0.43739583	0.29789212	0.23074947	0.40748901	0.0041914	1

Table: Correlation & Heat Map of Different Parameters

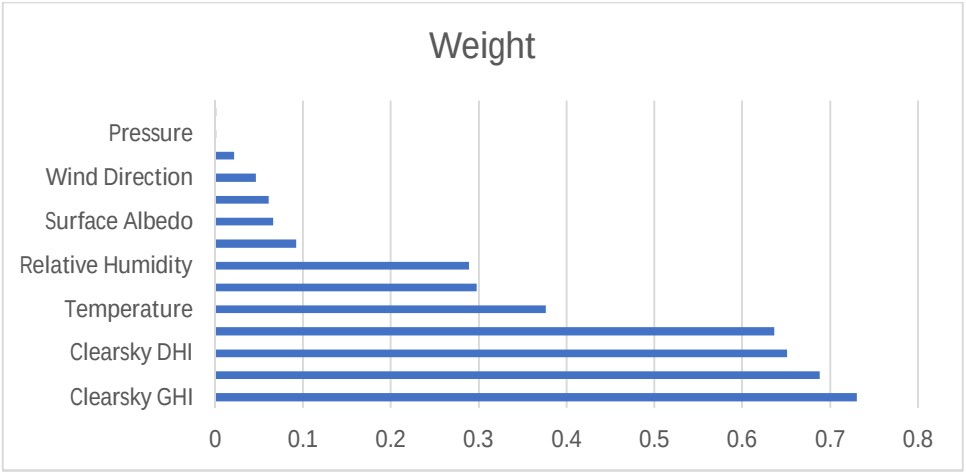


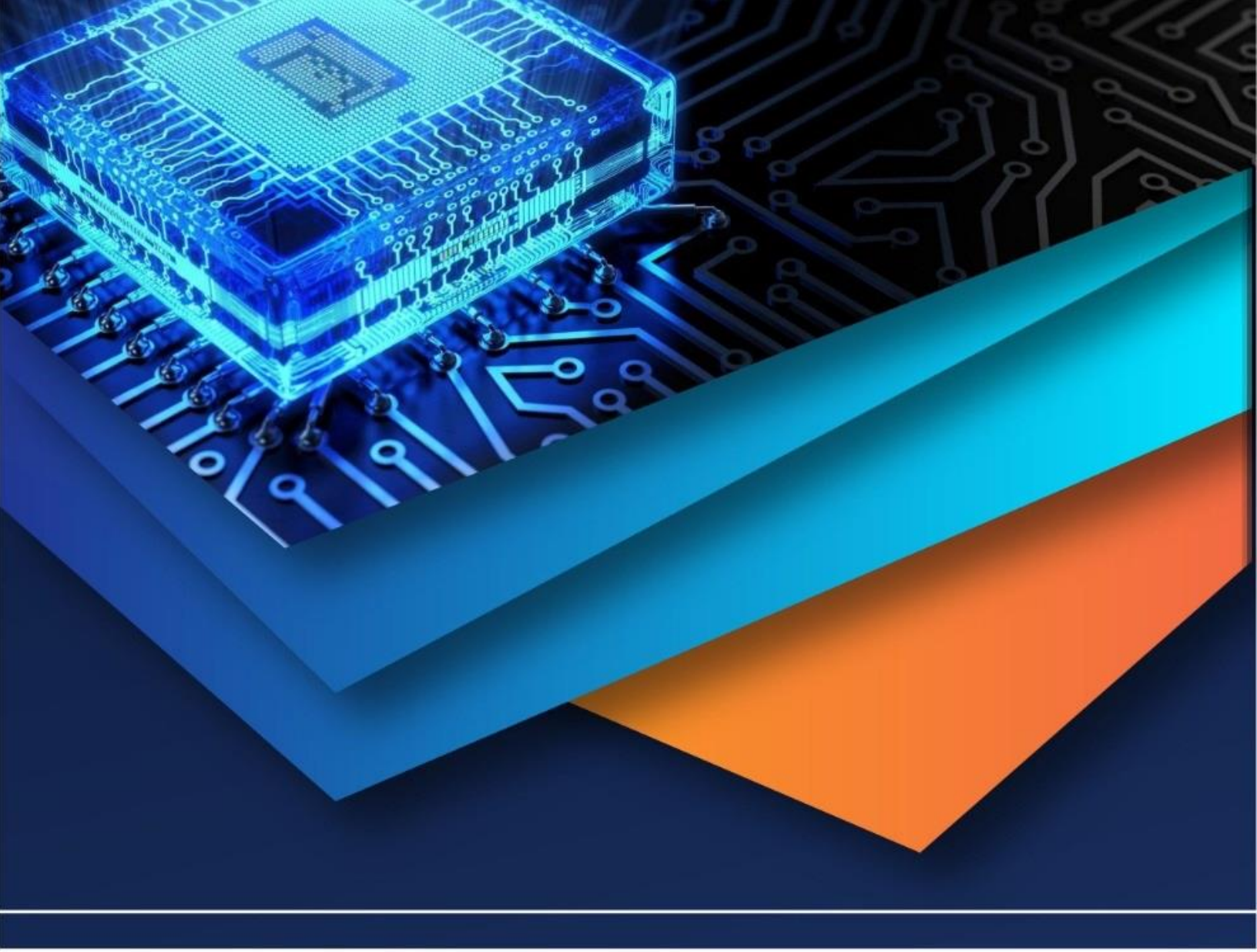
Fig: Weightage of different parameters in prediction

IV. CONCLUSION

In order to solve one of the oldest issues in the solar power sector, this study demonstrates for the first time that the prediction model (Deep Learning Method) is capable of accurately forecasting Global Solar Radiation based on hydrological, geographical, etc. characteristics. However, because different machine learning techniques have different mechanisms for making predictions, other prediction models perform less accurately with future prediction data. Nevertheless, machine learning is superior to other traditional methods in terms of ease of future data discovery and time required for prediction work. As automated data gathering becomes routine, it is possible to identify solar radiation and minimise losses by constructing, training, and testing such predictive models. As a result, we can conclude that machine learning is the ideal approach given how simple it is to use and how accurate the results are.As automated data gathering becomes routine, it is possible to identify solar radiation and minimise losses by constructing, training, and testing such predictive models. As a result, we can conclude that machine learning is the ideal approach given how simple it is to use and how accurate the results are.

REFERENCES

[1] Ahmed F, Ulfat I. Empirical models for the correlation of monthly average daily global solar radiation with hours of sunshine on a horizontal surface at Karachi, Pakistan. Turk J Phys 2004;28:301–7.
[2] Ajayi OO, Ohijeagbon OD, Nwadialo CE, Olasope O. New model to estimate daily global solar radiation over Nigeria. Sustain Energy Technol Assess 2014;5:28–36.
[3] Aladenola OO, Madramootoo CA. Evaluation of solar radiation estimation methods for reference evapotranspiration estimation in Canada. Theor Appl Climatol 2014;118:377–85.
[4] Allen RG, Pereira LS, Raes D, Smith M. others. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome 300, D05109; 1998.
[5] Almorox JY, Hontoria C. Global solar radiation estimation using sunshine duration in Spain. Energy Convers Manag 2004;45:1529–35.
[6] Al-Mostafa ZA, Maghrabi AH, Al-Shehri SM. Sunshine-based global radiation models: a review and case study. Energy Convers Manag 2014;84:209–16.
[7] Ampratwum DB, Dorvlo ASS. Estimation of solar radiation from the number of sunshine hours. Appl Energy 1999;63:161–7.
[8] Angstrom A. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. Q. J. R. Meteorol. Soc. 50; 1924. p. 121–6.
[9] Badescu V, Dumitrescu A. Simple solar radiation modelling for different cloud types and climatologies. Theor Appl Climatol 2016;124:141–60.
[10] Bahel V, Bakhsh H, Srinivasan R. A correlation for estimation of global solar radiation. Energy 1987;12:131–5.
[11] Bakirci K. Models of solar radiation with hours of bright sunshine: a review. Renew Sustain Energy Rev 2009;13:2580–8.
[12] Bakirci K. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. Energy 2009;34:485–501.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)