Predicting the irradiation effect on solar power plants and optimum performance evaluation model using Machine learning

B Kalaiselvi[1], Dr. Karthik B [2]Dr.A.Kumaravel [3]

[1]Research Scholar, Dept of ECE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India
[2]Associate Professor, Dept of ECE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India
[3]Dean, School of Computing, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India

[1]kalaigopal1973@gmail.com, [2]karthikguru33@gmail.com, [3]drkumaravel@gmail.com

Abstract.
Power harvesting using solar power is the recent trend and innovations happening in deploying many types of equipment working with solar power. This is harmless and greatly reduces pollution and is eco-friendly. The government also provides more concessions for establishing these solar power harvesting methods. There are two subsystems in solar power generation like sensor management systems. The subsystems have to be managed by predicting the power generation and identifying the right time for panel cleaning, and maintenance. In solar power generation systems, it is necessary to identify the faulty equipment and replace it for robust power generation. In the proposed article we are predicting the effect of ambient temperature, and module temperature on irradiation of the solar power generation system using the Weka machine learning tool using algorithms like SMOreg,Linear regression, KNN, and Multilayer Perceptron. The prediction model predicts the solar power irradiation with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of 0.0294 and 0.0558 of the Ambient and module temperature respectively. The prediction of irradiation in the solar power plantwill be helpful in grid maintenance, efficient use of accessories, identifying and servicing the sub-optimally performing unit to increase the daily yield, and reducing the operational cost.
Keywords. Solar power plant, Sensors, Ambient temperature, Power generation, Weka, ML, SMOreg, Machine Learning(ML), Renewable Energy,Regression, Power Generation.

1. INTRODUCTION

The advancement of human civilization has resulted in fast industrialization, which has been accompanied by urbanization and globalization, resulting in increased global energy demand. Conventional energy sources, which face rapid depletion, appear to be unable to meet soaring energy demand on their own. Solar, as a ubiquitous and environmentally beneficial source of renewable energy, has become one of the most widely used power production technologies in a variety of applications. To get the most out of a solar photovoltaic (SPV) system, it needs on be set to maximum power generation. The existence of life with sustainable energy is almost very difficult and the harvest of energy with an alternative source is very important in the context of developing countries, the conventional method is predicted to be extinct concerning time and globalization.The socio-Economic of a country greatly depends on the consumption of energy andalso has a direct impact on the economic growth of a country. Energy harvested has different forms la like thermal power plants, nuclear power plants, windmills, and so on. One of the trending methodologies of harvesting power is solar power plants. A country like India is best suited for establishing a solar power plant. Sun energy is available in abundance and can be transformed from solar energy to electrical energy. Since energy can neither be created nor destroyed. Methane, biogas is one of the most commonly used power from fossil fuels to generate energy which is very costly for establishing, comprehensive, unreplenishable, and decreases day by day. As a part of sustainable resources,India has stepped into solar power harvesting which is one of the renewable energy resources replacing fossil fuels and other types of power generation to meet the thirst for energy. This solar photovoltaicmethod of energy is relatively very cost-effective, and eco-friendly compared to other chemical or oil-based electricity generation. Hence in our near future, Solar power plant is believed to be the most effective, efficient, and sustainable type of energy generation and null effect on environmental degradation. Solar electricity, in general, is becoming more economical every year. Solar panels have cheap operational expenses compared to other power generation technologies since they can create a huge amount of electricitycompared to other power generation technologies since they can create a huge amount of electricity without compared to other power generation technologies since they can create a huge amount of electricity without the use of fuel. Another benefit of solar energy is that it is scalable, meaning it can be used to power a single residence or an entire factory. Overall, the sun emits undeniably more energy than we require; yet, the ability to convert it to electrical energy cost-effectively and in a cost-effective manner and to store electricity for nighttime and rainy days is a limitation. Solar power is a CO2-free, renewable energy source that has a substantially lower environmental impact than other power generation technologies. The key elements that affect the environment are the unique materials necessary for solar panel production, location, and the water required for solar panel cleaning[3-4]. The efficient use of solar panels, inverters, and grids in large-scale solar plants will reduce the cost and space necessary to generate the required power, reducing the environmental impact. Machine Learning (ML) methods will be increasingly beneficial in evaluating and generating models from solar plant data tomaximize hardware use.

Our context of studies in this paper goes along the following steps: Firstly, the data collection from Kaggle is taken as such as plant sensor data of power plant 1. Next, the data is preprocessed to nine different features as mean,maximum,minimum,Variance, Standard deviation, kurtosis, Skewness, RMS, and Trimean is evaluated over a wide range ofdata, and the best possible acceptable performance within the set threshold is sought for getting the final optimal model. Secondly, the data set acquired is preprocessed using the Weka tool usingthe election module []. Training is performed by building a model and training the model with 66% of data with 10 cross-validationprocesses. The third step is the classification of the data set using some of the pre-selectedclassifiers and the result is being tabulated for its performance accuracy and mean square error in classifying the output class with the remaining44% of the data. Finally, the performance evaluation criteria for checking for better performance and attainment of an optimum model. Moreover, this model has the capability of giving irradiation level that affects the solar power generationat different irradiation level as predicted by the dataset information.

2. RELATEDWORKS

Singh, Sarabpreet, et al.(2012)applied the ambient air temperature that will degrade the entire efficiency of the power plant. In this article, it is finalized with the result that 5 – 40 degree rise in ambient temperature caused an 11% decrease in mass flow rate, net output flow by 24%, Plant efficiency by 11and %, and an 18% decline in net output power. Finally the reduction in plant capacity by 8MW in summer due to an average rise of 40 degrees rise in ambient temperature [1]. GünnurŞenet al. (2018) proposed seasonal temperature variations havea great influence on GT,ST, and CCB. The range of 8-23 degrees causes 15.4 MW energy loss, on enhancing the CCPP to give maximum power output by the conventional controllers the output power decreased with rising in temperature by 30.4MW for a raise in temperature to 23 ° from 15° [2]. Najjar, et al.(2020) proposed the performance is evaluated using statistical analysis, The Thermal power plant's four parameters like availability, reliability, capacity factor, and thermal efficiency are considered using mathematical modeling compared with the international best practices and the target values [3]. M. Ferri et al. (2010) applied energy harvesting using light energy by using integrated micro solar cells and the energy is stored in an external capacitor. The capacitor relates to the load for a predefined period slot. This proposed systsystem'serimentalreresulthows that this is best suited for discrete-time regime applications.[4]. S. Ghosh et al. (2014) proposed the solar energy harvesting technique is employed to generate high voltages from integrated photodiodes. This paper presents two switch transistors and solar energy harvesting using ICs. Switched inductor DC to DC [5]. Pathak et al. (2020) proposed in this articlethe efficacy of a solar photovoltaic system using the MPPT algorithm, this is also an automated control systemusing the solar intensity and temperature. The experiment involved uniform solar sunstroke concerning time and Global MPPT[6]. MO Lukyanov et al.(2021),applythe legislative rules and regulations for harvesting renewable energy and introduce the penalties and imbalances faced by the producers. This article focuses on the forecast and implementation latest technology and energy storage systems. Using MATLAB Simulink, the theoretical relations with input dataand a model predict tests its operability[7]. Md. Bengir Ahmed Shuvho et al.(2019) proposed Life with sustainable energy is becoming unimaginable and we must switch over to an alternative source of energy such as the solar power plant and the performance evaluation using Fuzzy logic controller for solar irradiation, ANN for the grid-connected to 80KWp investigations and sun simulator test [8], many literatures discussed on solar power plant using Machine learning algorithms [13-16].

3. MATERIALS AND DATA DESCRIPTION

The proposed article is working on the Kaggle data set [9] which contains two sets of information one give the power generation data and the second one gives the sensor panel data from a solar power plant. The data is collected from the inverter wherein it is connected to an array of solar panels, the sensor data is collected from the optimally placed sensor at the plant every 15 minutes both the power generation and the sensor data or tabulated and given in this public data set like Kaggle.[9]

Variant Modes Features Extraction: In this study, for each mode, we have extracted the statistical features of variability quantifying the variables. The Table shows the statistical significance of the mode. For each mode, the computational statistical feature results in 9 different features with the resulting data set are 425 by 19

Feature number	Descriptive Statistics	Description
1	Mean Value MEDIAN(Array)	The central value of the distribution
2	Maximum value MAX(Array)	The maximum value in the data selected range
3	Minimum Value MIN(Array)	The minimum value in the data selected range
4	Variance VAR.S(Array)	Variance quantifies the deviation of data from the mean value in the data selected range
5	Standard deviation STDEV.S(Array)	Quantifies the amount of distribution each value in data has from the mean.
6	Kurtosis KURT(Array)	Quantifies the peakedness in the data[15]
7	Skewness SKEW.P(Array)	the symmetry in the dataset is quantified[16]
8	Root mean square SQRTPI((SUMSQ(Array)/COUNTA(Array)))	Quantifies the strength of variations in the data.
9	Trimean TRIMMEAN(Array,Percentage)	The mean of data excluding the highest and lowest 50% of the data points

Table 1: Variant mode feature extraction on weather sensor data

The power generation and the sensor data are collected for some time and date. To obtain a relationship between the sensor and the power generation of the Solar Power Plant at a particular time with varying intensity of light from the sun which has a direct relationship with the power generation. The table shows the distribution of Power Plant data for 34 days and every 15 minutes.

Variants	Ambient Temp °C	Module Temp in °C	Irradiation
Mean	28.0549	32.5802	0.2328
Variance	0.9887	12.0260	

Standard Deviation	0.8117	2.5227	
Kurtosis[15]	-0.5834	-0.4273	
Skewness[16]	0.044	0.1782	
Trimean	28.0568	32.588	
Maximum	29.2704	36.7395	
Minimum	26.8854	29.2872	

Table 2: Data validation in variant mode using Weather sensor data

3.1 Data Preprocessing

The data pre-processing is done using the weka tool by choosing the preprocessing we can discard all the irrelevant, outliers, and incomplete data transformation. The data is stored in comma-separated value file format (CSV) which is the explorable format in weka. The description of the data present in the data set of power generation is as follows:

i) Date and time of the reading recorded, Plant ID for the plant, Source key of the plant, Temperature at the plant, Solar Panel Module temperature, Solar panel Ambient temperature Irradiation for every 15 minutes duration,

ii) The data is normalized using the pre-processingin weka tool which Normalizes all numeric values in the given dataset (apart from the class attribute, if set) before analyzing and classification of the built model using machine learning algorithms.

iii) The Variant mode features are extracted for the available data set and analyzed in the weka tool for building a model and predicting the correlation coefficient and error evaluations.

4. EXPERIMENTAL ANALYSIS AND PREDICTION MODEL:

The main experimental setupcomes from the Kaggle dataset, which contains a system of 2 solar power plant datasets in a comma-delimited format, and the output class chosen is irradiation soambient temperature and module temperature are the two categories that affect the irradiation in solar power plants.

4.1 Experimental setup

The collected data is pre-processed in the weka tool for the implementation of the machine learning algorithm with the Normalized weather sensor dataset. Data is classified using selected classifiers such as Linear Regression, SMOreg,IBk, and Random Forest. The next step in the advancement is to check for a better correlation coefficient, very less Mean square error, root mean squared error, relative absolute error, and Root relative squared error. We go to the preprocessing step called attribute evaluator, Kernel selection, and Neighbour selection for increasing the performance and decreasing the error percentage are performed with the same base classifiers and the level of precision is checked. If the threshold is reached, the model attained is an optimal model.

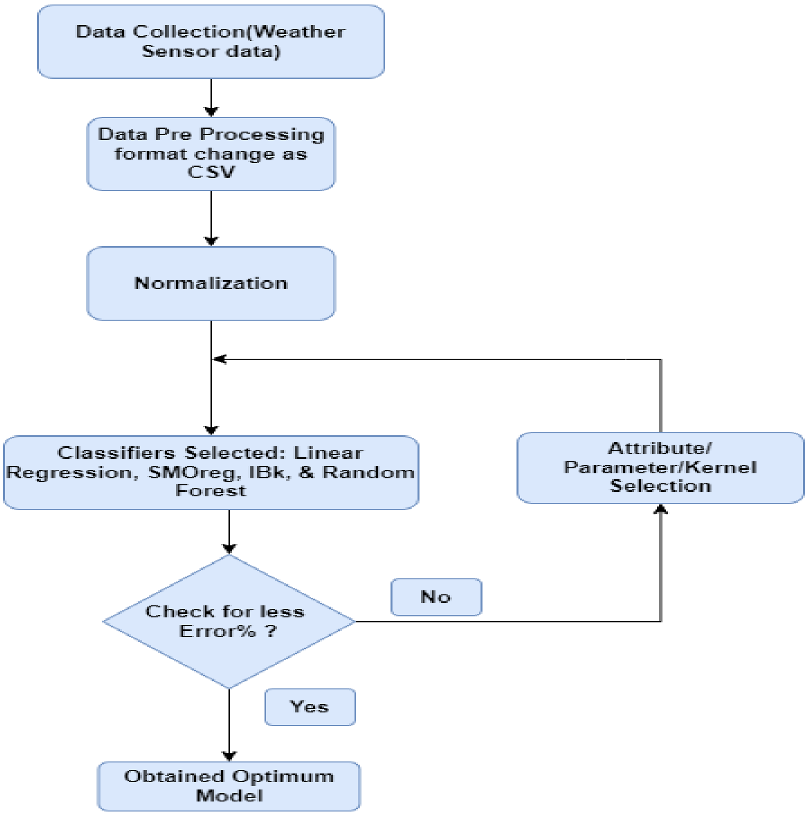

Figure 1. Experiment flow diagram

Figure 1 illustrates an experimental flow diagram starting from the data collection, format changing for exploring in the Weka tool with nearly 425 instances 19 attributes. First, the data is normalized in preprocessing step in the wekatool and the classification begins by building a model for the selected base classifiers. The next process is decision making by adjusting the attribute selection in linear regression as the M5 method and Greedy method for getting less percentage of error and enhanced correlation coefficient. In SMOreg the kernel like Polykernel and PUK kernel is selected out of these two PUK worked out to be best, IBk is instance base classifier the

Weka 3.8.5 (Waikato Environment for knowledge analysis) is one of the popular machine learning software written in Java that was developed by the University of Waikato, New Zealand, open-sourcesoftware available under the general public license.[12] Weka is a graphical user interface way for learning machine learning. It is a collection of machine learning algorithms for performing datasets directly or it can be called from Java code. It also contains tools for data preprocessing, classification, regression, associated rules, clustering, and visualization.[12]

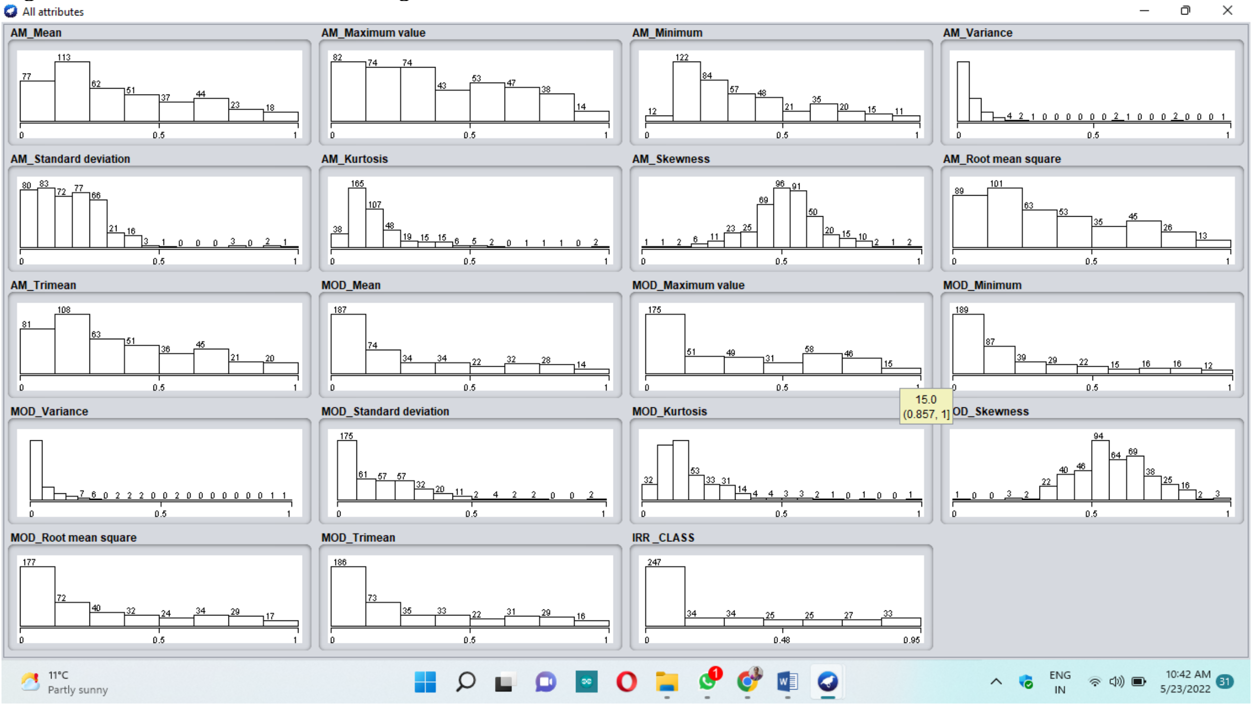

Figure 2. Normalized data distribution in weka3.8.5

Figure 2 shows the visual distribution of Normalized data which is 19 attributes and 425 instances [12].The proposed experimental setup is for the structure of building a model to estimate or predict the irradiation level.The selected classified are linear regression model function category, SMOreg model from function category IBk in the lazy category, and random forest from Tree category using weka tool.

4.2 Designated Base Machine Learning Algorithms

Selected two classifiers are from the set of supervised learning algorithms to fit our modeling procedure and their knowledge and modest implementation foremost to easier clarifications.

4.2.1 Linear Regression:

Linear regression is one of the popular machine learning algorithms usually used for the statistical method of predictive analysis decision-making predictions for continuous data sets, real or numeric variables such as salaries, age, product values, and sensor's timely varying values and so on. The linear regression algorithm follows a linear relationship between the dependent variable and independent variables and is hence called linear regression.

$$Y = a_0 + a_1 \times \qquad (1)$$

The formula for calculating the mean square error is as follows:

$$MSE = 1\frac{1}{N} + \sum_{i=1}^{n}\big(Y_i - (a_0 + a_1 \times i)\big)^2 \qquad (2)$$

Where N is the total number of observations Yi is the actual value and (a1xi+a0)=the Predicted value.

4.2.2 SMOreg:

SVM (Support Vector Machine) to carry out the estimate of one-variable persistent and restricted genuine capacities characterized in the scope of reals through PUK portion (Pearson VII capacity based Universal Kernel); the goal is to show the way that exemplary AI relapse can accomplish intriguing degrees of exactness with incredibly short learning times.

4.2.3 4.2.3IBk Instance base classifier:

One of the conventional calculations in the AI idea is the closest neighborhood calculation, where the information is dispersed in the hyperdimensional space, and the calculation occurring relying upon the "k" closest neighbors by basically duplicating the names of the load with next to no particular impacts on calculation for the forecast. Even though it is computationally concentrated, it is exceptionally simple for execution.

4.2.4 4.2.4 Random Forest:

This is additionally a prestigious sort of tree classification AI calculation. The group AI calculation in weka. make numerous choice trees and union them to get a more precise and stable forecast. These choice tree-based indicators are most popular for their computational power and adaptability. Nonetheless, on account of profoundly uneven preparation information, as is many times found in information from clinical examinations with enormous benchmark groups, the preparation calculation or inspecting strategy ought to be changed to further develop the forecast quality for the classes. minority. In this work, a balanced random forest approach is proposed for Weka 3.8.5.

5. Results and Discussions

5.1 Evaluating the Performance Criteria

The performance metrics are obtained from the prediction patterns generated by the applied classifiers, which are reflected in the available inputs in the confusion matrices. Evaluating the performance by building a model using machine learning algorithms

Selected Classifiers	Time to build the model	Correlation coefficient	MAE	RMSE	RRSE	Root relative squared error	No of Instances
Linear Regression Model	0.03	0.9881	0.0329	0.0468	12.62%	15.35%	425
SMOreg(PUK)	0.32	0.9903	0.0241	0.0424	9.23%	13.89%	425
Ibk @k=3	0	0.9831	0.0306	0.0561	11.72%	18.39%	425
Random Forest	0.26	0.9874	0.0274	0.0483	10.50%	15.82%	425

Table 3:Experimental results of the built model

Table 3 is the consolidated results of all classifiers with changes in attribute and kernel changes for obtaining minimum error.

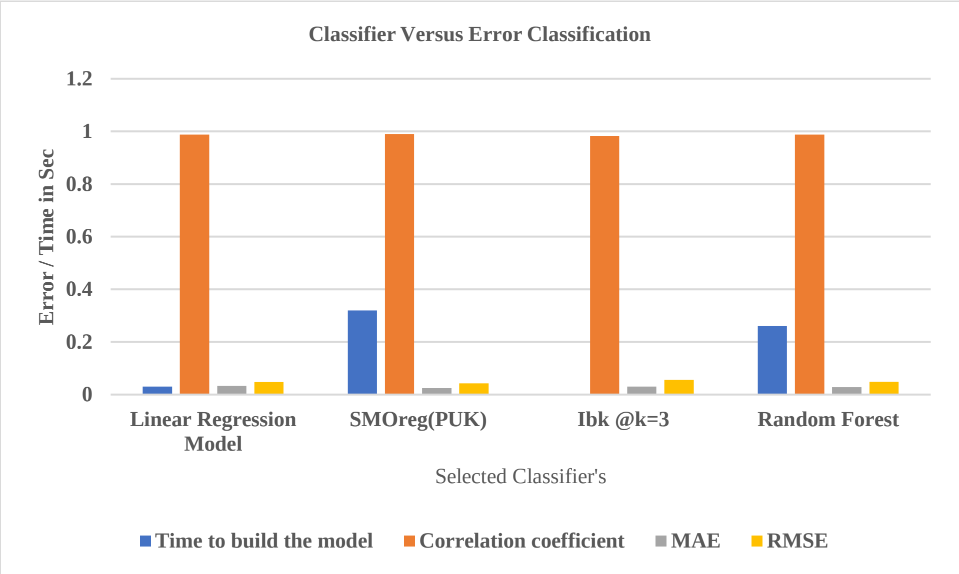

Figure 3. selected classifiersfor the error and time to build the model

Figure 3 illustratesa similar correlation coefficient obtained for all the four different ML algorithms that are in the range of 0.9831 to 0.9903 for Instance-based classifier and SMOreg respectively. The Instance-based classifier has taken negligible time to build the model, SMOreg ML algorithm has taken more time of 0.36 seconds compared to all other classifiers. The mean absolute error is 0.0241 to 0.0329 for the function-based classifier's SMOreg and Linear regression ML algorithms. The root mean square error value is also in the range of 0.0424 to 0.0561 for SMOreg and Instance base Classifier.

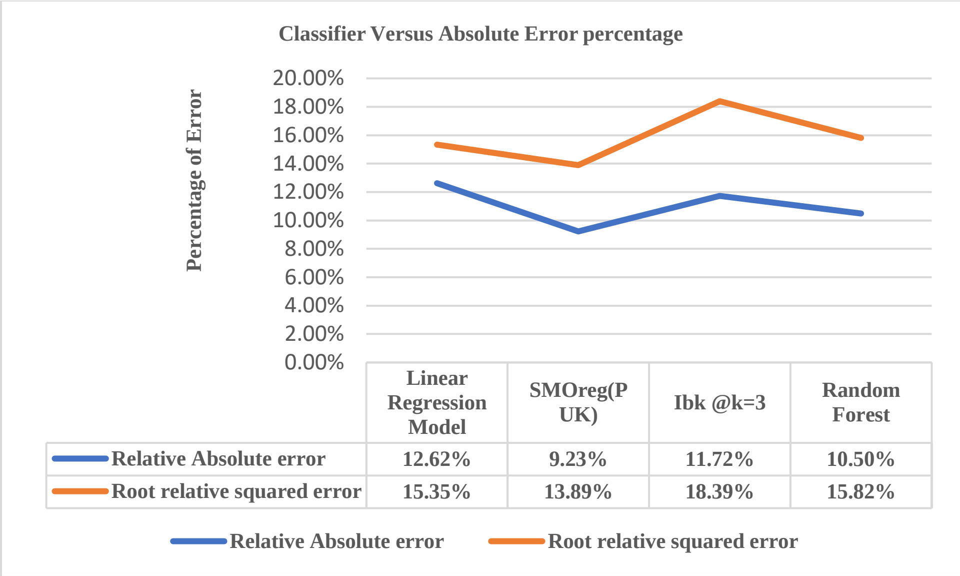

Figure4. Classifierconcerning the percentage of error characteristics

Figure 4 illustrates the Selected ML algorithms versus the relative absolute errorpercentage(RAE) and the Root relative squared error percentage (RRSE). The range of RAE is 9.23% to 12.62% for SMOreg and Linear Regression ML algorithm. The Root relative squared error lies in the range of 13.39% to 18.39% for SMOreg and the Instance-based classifier IBk ML algorithm. Overall the optimum model that is considered with minimum error is the function-based ML algorithm SMOreg.

6. Conclusion:

Solar Energy harvesting is one of the favorable alternate energy resources in the declination of ecological energy resources. This prediction of irradiation in photovoltaic energy resources using machine learning algorithms gives a model that predicts the irradiation with very less error and a good correlation coefficient of 0.9903 and RMSE of 0.0424. Future research examinations can be conducted on mapping solar potential in the fuzzy logic ANN (Artificial Neural network) model or go into deep learning techniques.

7. Future recommendations:

Stability analysis of the fuzzy logic and ANN models will be carried out in the future for verifying the meeting stability criteria of these results. Fuzzy logic and ANN models can be utilized to break down a mixture of sustainable power for future energy emergency alleviation. Notwithstanding that these techniques can be used to plan the executive arrangement of an Earth-wide temperature boost and environmental change difficulties.

8. References:

[1] Singh, Sarabpreet& Kumar, Rajesh. (2012). Ambient air temperature affects power plant performance. International Journal of Engineering Science and Technology. 4. 3916-3923.

[2] GünnurŞen, Mustafa Nil, HayatiMamur, HalitDoğan, Mustafa Karamolla, MevlütKaraçor, FadılKuyucuoğlu, NuranYörükeren, Mohammad Ruhul Amin Bhuiyan,"The effect of ambient temperature on electric power generation in natural gas combined cycle power plant—A case study",Energy Reports,Volume 4,2018, Pages 682-690,ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2018.10.009.

[3] Najjar, Y.S.H., Abu-Shamleh, A. Performance evaluation of a large-scale thermal power plant based on the best industrial practices. Sci Rep 10, 20661 (2020). https://doi.org/10.1038/s41598-020-77802-8

[4] M. Ferri, D. Pinna, E. Dallago, P. Malcovati, "Photovoltaic Energy Harvester with Power Management System", Journal of Sensors, vol. 2010, Article ID 349389, 6 pages, 2010. https://doi.org/10.1155/2010/349389

[5] S. Ghosh, H. Wang, and W. D. Leon-Salas, "A Circuit for Energy Harvesting Using On-Chip Solar Cells," in IEEE Transactions on Power Electronics, vol. 29, no. 9, pp. 4658-4671, Sept. 2014,DOI: 10.1109/TPEL.2013.2273674

[6] Pathak, Pawan &Yadav, Anil &Alvi, Dr. Parvez. (2020). Advanced Solar MPPT Techniques Under Uniform and Non-Uniform Irradiance: A Comprehensive Review. Journal of Solar Energy Engineering. 142. 1-31. 10.1115/1.4046090.

[7] MO Lukyanov, EV Verbytsky, and A. Blinov, "Solar power plant storage system", Microsyst., Electron. andAcoust. , issue 26, issue 2, p. 236748–1, Aug 2021.

[8] Md. Bengir Ahmed Shuvho, Mohammad AsaduzzamanChowdhury, Shameem Ahmed, MohammadAbulKashem,Prediction of solar irradiation and performance evaluation of grid-connected solar 80KWp PV plant in Bangladesh,Energy Reports,Volume 5,2019, Pages 714-722,ISSN 2352-4847,https://doi.org/10.1016/j.egyr.2019.06.011.

[9] https://www.kaggle.com/datasets/anikannal/solar-power-generation data?select=Plant_1_Weather_Sensor_Data.csv

[10] Vijayan T, Sangeetha M, A. Kumaravel, Karthik B, Gabor Filter and Machine learning Based Diabetic Retinopathy Analysis and Detection, Microprocessors and Microsystem, 2020; DOI: https://doi.org/10.1016/j.micpro.2020.103353

[11] Bakthavatchalam, K.; Karthik, B.; Thiruvengadam, V.; Muthal, S.; Jose, D.; Kotecha, K.; Varadarajan, V. IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms. Technologies 2022, 10, 13. https://doi.org/10.3390/technologies10010013

[12] WEKA: http://www.cs.waikato.ac.nz.

[13] Kamal, S., Ramapraba, P. S., Kumar, A., Saha, B. C., Lakshminarayana, M., Sanal Kumar, S., ...&Erko, K. G. (2022). Optimization of Solar Panel Deployment Using Machine Learning. International Journal of Photoenergy, 2022.

[14] Dhanraj, J. A., Mostafaeipour, A., Velmurugan, K., Techato, K., Chaurasiya, P. K., Solomon, J. M., ...&Phoungthong, K. (2021). An Effective Evaluation on Fault Detection in Solar Panels. Energies, 14(22), 7770.

[15] Maheswaran, U., Kallam, R. B., Arathi, B., Prawan, K., & Anitha, G. (2021). Efficient plant leaf disease identification Material Fabrication using lightweight device. Materials Today: Proceedings, 47, 381-386.
[16] Ramesh, S., Tamilselvi, M., Ramkumar, G., Anitha, G., &Nirmala, P. (2022, January). Comparison and analysis of Rice Blast disease identification in Greenhouse Controlled Environment and Field Environment using ML Algorithms. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-5). IEEE.