e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

# FORECASTING RENEWABLE ENERGY FOR AN INTEGRATED SMART GRID

**Waqas Qureshi*1, Osama Saleh*2, Sharjil Momin*3**

*1,2,3Mumbai University, Computer, M.H.S.S.C.E., Mumbai, Maharashtra, India.

## ABSTRACT

Nowadays, the use of renewable energy to mitigate the effects of climate change and global warming has become an increasing trend. Several countries present a high potential in Photovoltaics, which allows these two energy sources to be studied with considerable impact.

However, due to the challenge of climate and energy crisis, renewable energy generation are intermittent and variable, as the energy sources at the ground level is highly dependent on cloud cover variability, atmospheric aerosol levels, and other atmosphere parameters such as the air temperature, humidity, insulation and so on. This inherent variability of large-scale energy generation poses substantial challenges to smart grid energy management. Consequently, there is a need to mitigate those errors. Thus, accurate predictions of the amount of renewable energy that can be produced in future is an important task. Renewable energy prediction represents an important and active job in the renewable energy sector. In order to improve the prediction ability of renewable energy, Machine Learning Algorithms play an important role in this field. In this Proposed work, renewable energy sources are forecast by utilizing various data mining techniques, Numerical Weather Data (NWD), including pre-processing historical load data and the load time series' characteristics. After forecasting, with the objectives of minimizing overall cost and minimizing power loss, the optimal model is built.

**Keywords:** Solar Energy, Machine Learning Algorithms, Energy Management, Forecasting Models, Photovoltaic System.

## I. INTRODUCTION

The introduction should be typed in Times New with font size 10 With the rapid development of global industrialization, the world's energy demand is ever-increasing, it has been recognized that excessive consumption of fossil fuels to satisfy these demands will not only accelerate the depletion in fossil fuel reserves but also have an adverse impact on the environment. These influences will result in increasing health hazards and threats of global climate change. Further, these Non- renewable energy sources such as coal, oil, natural gas, fossil fuels, nuclear, minerals, etc., cannot be regenerated in a short period, and their consumption rate far exceeds their regeneration rate. For instance, fossil energy is not only finite and will eventually dry up, but also, it's becoming expensive day by day. Moreover, these energy resources are exhausted, and their future existence is questionable. Thus, there is an increasing need in the twenty-first century to decrease fossil fuels' consumption and boost their consequent replacement by other cleaner and more environment friendly energy sources. One of the most widely adopted action plans, towards obtaining a more environmentally sustainable planet, involves the integration of renewable energy as a primary source of energy production. Renewable energy refers to reusable energy that can be recovered in nature, such as solar energy, wind energy, biomass energy, hydropower, waves, tides, and geothermal energy etc. With characteristics of sustainability and low environmental pollution, the topic of renewable energy has attracted attention, and plenty of relevant studies have been performed recently. With deep learning become more accessible and mainstreamed it has brought new challenges and opportunities to forecasting renewable energy. At the same time, due to the fact that renewable energy is affordable, low-carbon, stable, and reliable, consumers in emerging markets and enterprises have continuously increased their demand for renewable energy. These driving factors and demand trends are particularly evident in both established and developing regions throughout the world and have forced the world to move to green energy. Many countries have demonstrated their interest in this topic by implementing new policies, norms, and laws in their respective communities. Some countries have already proposed annual goal to achieve certain percentages of their energy consumption by renewable energy sources. On average, 26.2% of the 2018 world's energy demand was supplied by renewable energies, and it is expected to increase the percentage up to 45% by the year 2040.

One of the most important challenges of renewable energy in the near future is to provide a safe and reliable power supply for the consumers. The renewable supply is the integration of renewable sources with non-renewable energy sources into electric grids to meet energy demand requirements. Hence, it is also necessary to know power demand accuracy for stable and efficient operation of power systems. Storage of electrical energy is necessary in the case when there is excess power production from the Renewable energy sources (RES) but less load demand. However, it cannot be massively stored as energy storage is costly, requires high maintenance and have limited life spans. Because of this, utilities have to maintain supply and demand equilibrium at every moment. However, due to the large volatility and the intermittent and random nature of renewable energy, this generation of numerous energy sources is intermittent and chaotic. Which may seriously affect the quality of electric energy and the operation of the power grid. If the output of power generation can be accurately forecasted, the negative impacts to the grid can be reduced by a large extent. Thus, Renewable energy forecasting technology plays a vital role for saving energy, reducing power generation costs, improving social and economic benefits, management and the policy making of energy system. Moreover, the results of these predictions will serve as the basis for future operational and strategic decisions with local and regional impact. Therefore, renewable energy forecasting is a highlighted topic in the twenty-first century.

## II. METHODOLOGY

Method and analysis which is performed in your research work should be written in this section. A simple strategy to follow is to use keywords from your title in first few sentences.

### Introduction

A step-by-step framework to the current research on ‘forecasting solar energy’ is presented in this section. It is in-depth review to facilitate selection of the appropriate forecast method of ML, for designing and demonstration of the project that produce consistent and transparent results. The section will continue to discuss several important modules of the system.

For selection of appropriate method for our proposed work we have gone through several commonly used machine learning techniques. Linear Regression algorithm is proved the most basic and widely used technique for the forecasting purpose.

### Algorithm & Process Design

Linear Regression is supervised learning method of a machine learning algorithm. It performs a regression task, in which independent variables are used to model a target prediction value. A regression technique is applied when the output variable has real or continuous value. Different regression models differ; if there is only one input variable (x), such regression model is referred to as simple linear regression. When there are more than one input variables, this type of linear regression is referred to as multiple linear regression.

Linear regression performs the task of predicting a value of dependent variable (y) when given an independent variable (x). So, regression technique determines a linear relationship between x (input) and y (output). Therefore, it's known as Linear Regression.

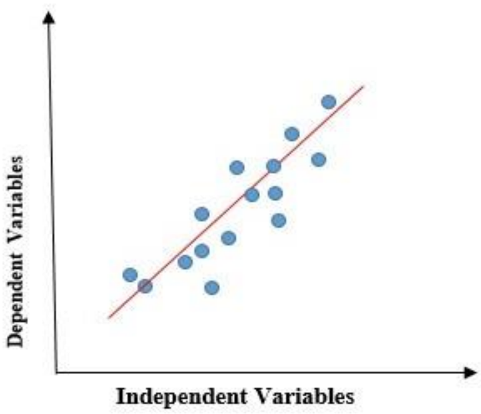


**Figure 1:** Regression graph for Independent VS dependent Variables.

In Regression, we plot a graph between the variables that best match the provided data points. In other words, Regression is defined as “a line or curve that goes through all of the data-points on a target-predictor graph

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

with the minimum vertical distance between data points and the regression line" It is most widely used for forecasting, time series modeling, and identifying the relationship between variables.

The above graph represents the linear relationship between the dependent variable and independent variables. The red line is referred as the best fit regression line based on the given data points shown as blue dots.

To construct the best-fit line linear regression uses a standard slope-intercept form.

**For Simple Linear Regression:**

$$y = c + mx$$

$$\boxed{y = \theta_1 + \theta_2.x}$$

**For Multiple Linear Regression:**

$$Y = \theta_1 + \theta_2.X + \theta_3.X_1 + \theta_4.X_3 + \ldots\ldots\ldots + \theta_n.X_n$$

While training the model we are given:

x: input training data

y: Independent variable.

When training the model, it fits the best regression line by least square method to predict the value of y for a given value of x. The model gets the regression line by finding the best θ1 and θ2 values.

θ1: intercept

θ2: coefficient of x

Once we find the best θ1 and θ2 values, we get the best fit regression line. By obtaining the best fit regression line, model aims to predict y value such that the error difference between estimated value and actual value is minimal.

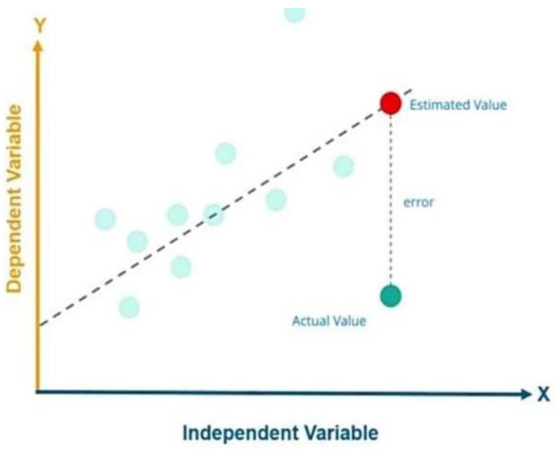


**Figure 2:** Graph for cost function.

So, it is very important to update the **θ1** and **θ2** values in order to find the ideal value that minimizes the error between predicted y value and true y value. Here the Cost Function is used.

Cost Function:

It's a function that determines our model's performance on the provided data. It calculates the difference between predicted values and actual values and displays it as a single real number. Whichever choice of **θ1** and **θ2** will provide the optimal value for total error or cost function will be best for the model. In this work, following approaches are used to measure performance:

**A. R-Square Or Coefficient Of Determination:**

$$R2\_score = 1 - \frac{\sum \left| y_{forecasted} - y_{observed} \right|}{\sum \left| y_{forecasted} - y_{mean} \right|}$$

**B. Mean Absolute Error:**

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_{forecasted} - y_{observed} \right|$$

**C. Root Mean Square Error:**

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( y_{forecasted} - y_{observed} \right)^2}$$

**Note:** The optimal value for performance measures for MAE, RMSE is 0 and 1 for $R^2$.

Gradient Descent:

The model utilizes Gradient Descent method to update **θ1** and **θ2** values in order to optimize the Cost function and achieving the best fit line. The idea is to begin with random values of **θ1** and **θ2** and then iteratively updating the values until minimum cost is achieved.

Process design for linear regression algorithm:

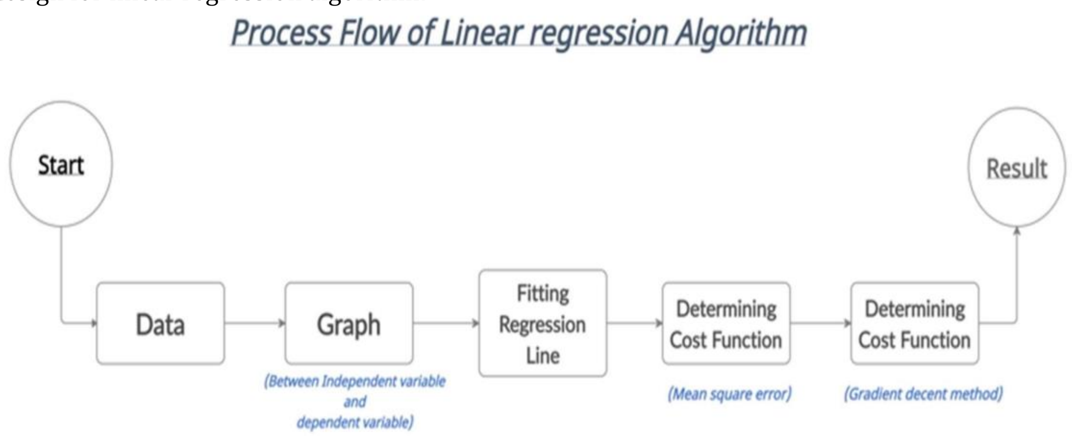


**Figure 3:** Flowchart of linear regression algorithm.

## III. MODELING AND ANALYSIS

To address the research question, we separated the proposed system into distinct functional areas which are listed below:

**1. Data Collection or Dataset:**

Reliable data availability and the choice of right attributes from the data are crucial used to train and test the forecasting model. In this work we relied on the dataset imported from Open Power System Data https://open-power- system-data.org/ . A free-of-charge platform which is dedicated to electricity system researchers and share data that are publicly available but currently inconvenient to use. The platform provides data for 37 European countries in a file, but in this project, we focused on data for UK in 2019 as an example. Further, we have utilized two datasets:

- Time series containing load, solar prices in hourly resolution
- Weather data which comprises values of radiation, temperature and other parameters

**2. Data pre-processing:**

The data pre-processing is important phase to make data smooth for machine learning algorithm. As the data we imported contains data for 37 European countries, we have to filter the dataset only with data of United Kingdom. We begin with a CSV file containing time series data for 37 European countries, but only read the data

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

for UK in 2019. Then we filter the CSV file with weather data for UK 2019. After this stage we are ready with the dataset containing the row of UK in 2019.

**3. Training and Validating data on approach model:**

For better understanding, evaluation of a model is necessary. Hence, after filtering, we train a processed dataset on our approached linear regression algorithm and validate it to establish forecasting model.

**4. Forecasting:**

Once the predictive model gets ready for accurate prediction, the process of forecasting has been performed for solar energy sources

**Analysis**

After filtering time series data for the rows for UK 2019, we end up with a Data Frame 8760 entries and 10 columns (each relative to a different quantity) having no error values. To have an idea about the data we make a couple of plots. The fig.5 below shows the actual solar generation in UK;

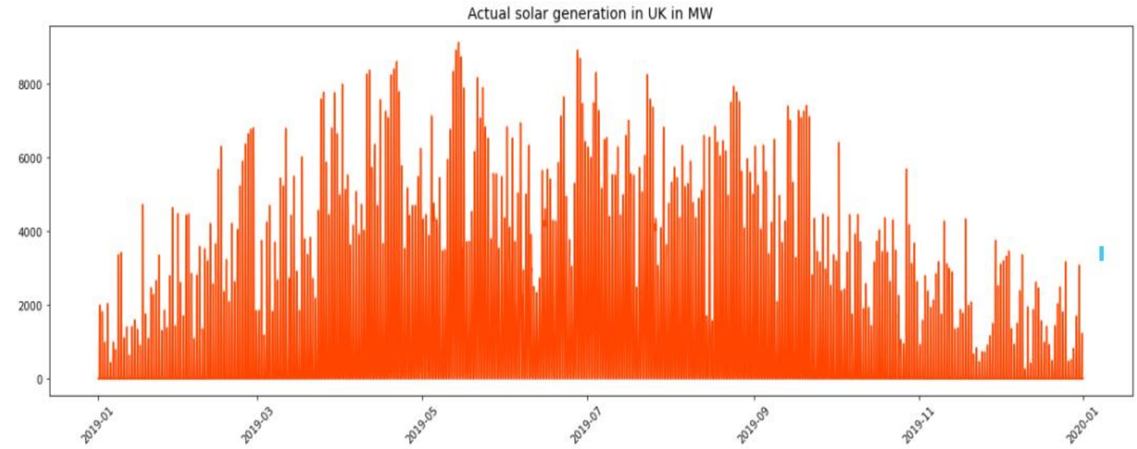


**Figure 4:** Actual solar generation in UK in MW.

It can be seen there is no clear pattern for the solar generation across the year, even though there is significantly larger production in the middle months of the year.

Now, we read the CSV file containing the weather data for UK 2019 and we obtained 350640 entries, each characterized by the different quantity as follows:

- GB_temperature
- GB_radiation_direct_horizontal
- GB_radiation_diffuse_horizontal

The behavior of these averaged weather quantities in UK 2019 are shown in fig.6 and fig.7;

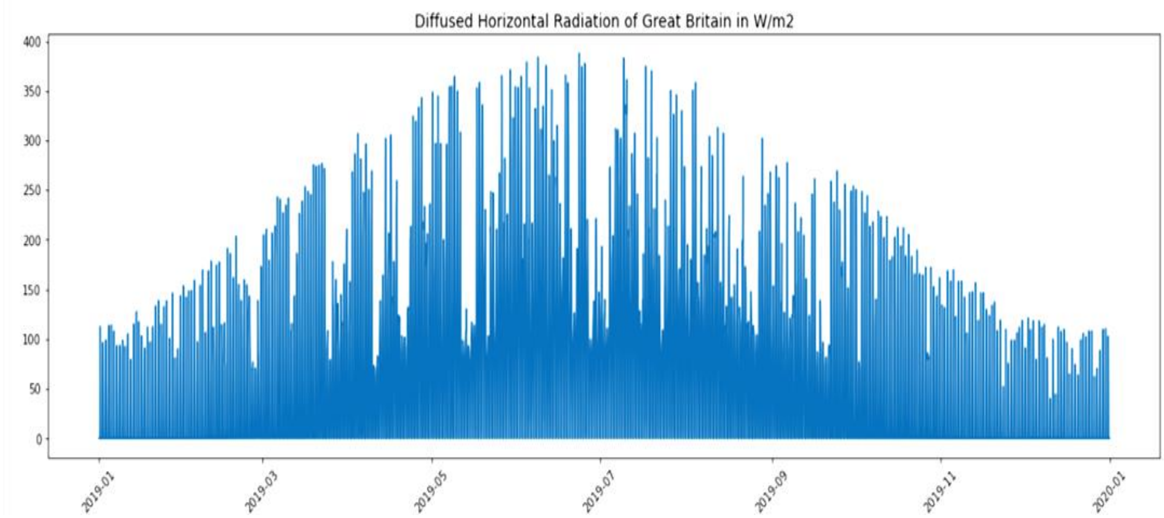


**Figure 5:** Diffused horizontal radiation in UK in W/m2.

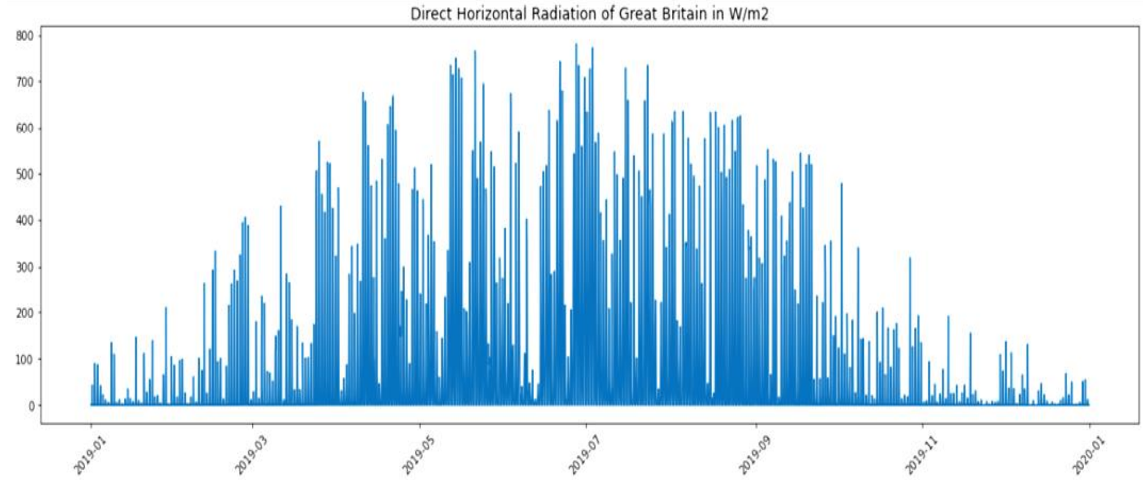


**Figure 6:** Direct horizontal radiation in UK in W/m2.

As expected, the horizontal radiation at the ground level was larger during the summer months, likewise with the temperature, as plotted below.

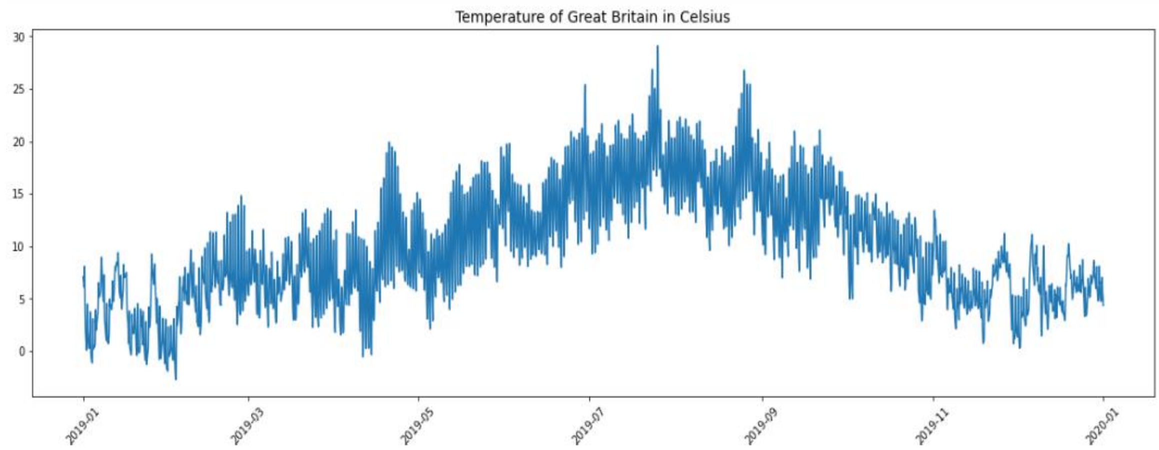


**Figure 7:** Temperature in UK in celsius.

Now, we simply merged both the data frames and a matrix of pair correlation coefficients is generated for a set of features under investigation in order to find collinear factors as shown below in fig.7;

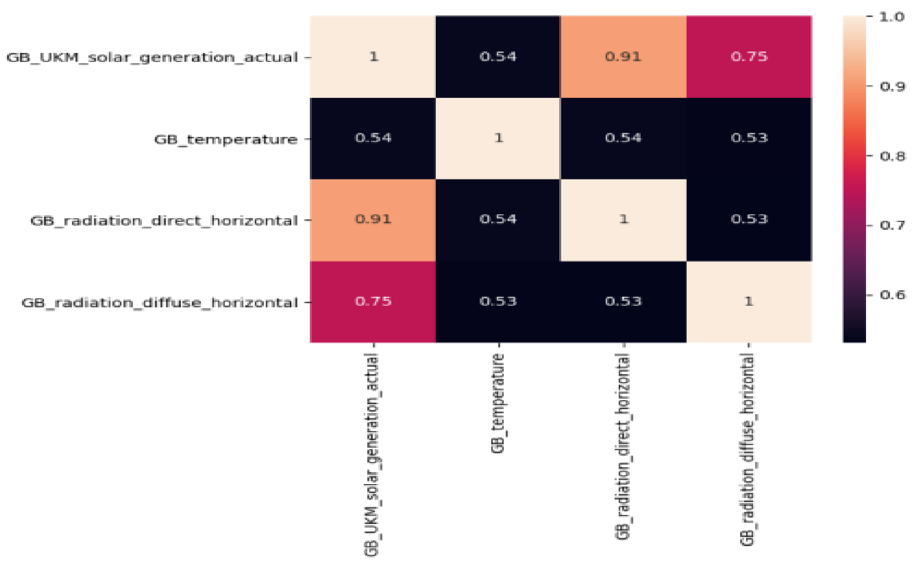


**Figure 8:** Correlation matrix.

The linear correlation coefficient value in this case is 0.91 and 0.75.

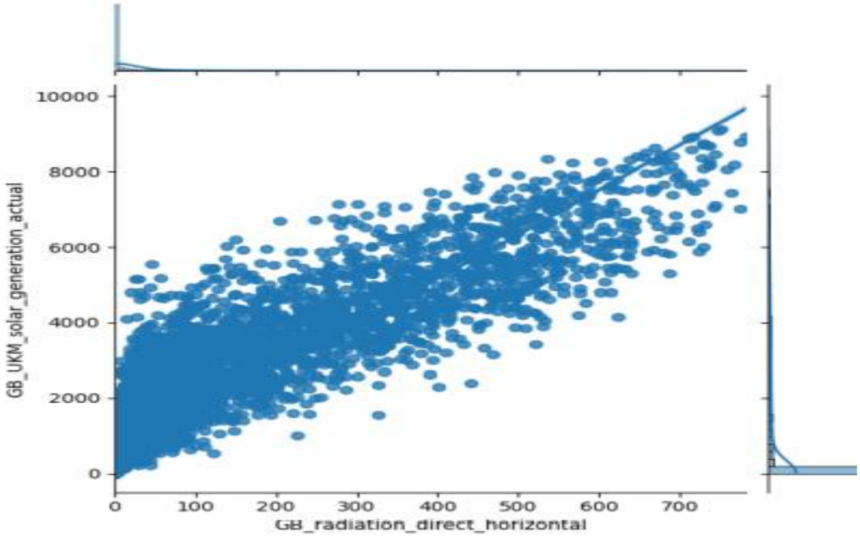


**Figure 9:** GB_radiation_direct_horizontal and GB_UKM_solar_generation_actual.

The Fig.8 shows that the linear correlation coefficient value between "GB_radiation_direct_horizontal and GB_UKM_solar_generation_actual" is 0.91. It is shows that a strong linear relationship between GB_radiation_direct_horizontal and GB_UKM_solar_generation_actual.

Similarly, the linear correlation coefficient value between "GB_radiation_diffuse_horizontal and GB_UKM_solar_generation_actual" is 0.75 and there seems to be a linear relation between "GB_radiation_diffuse _horizontal and GB_UKM_solar_generation_actual" as can be seen in fig.11;

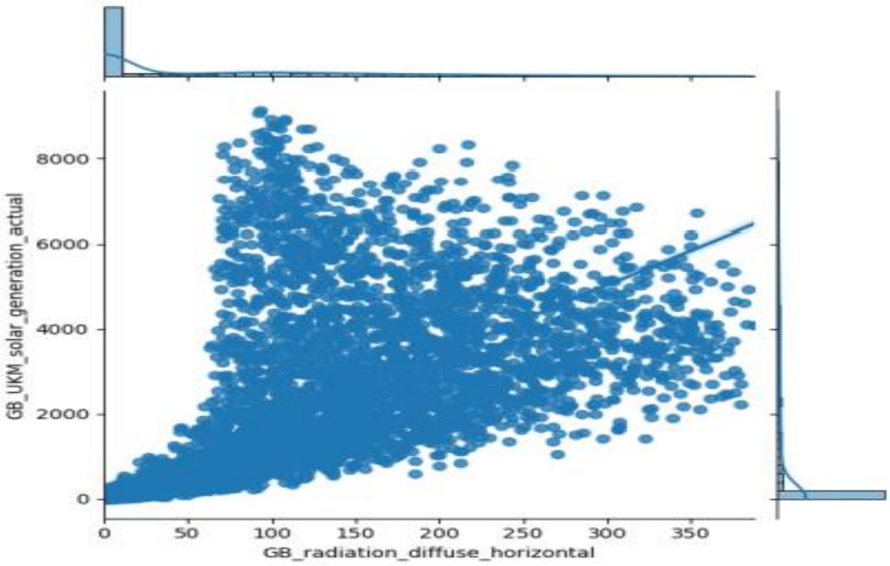


**Figure 10:** GB_radiation_diffuse _horizontal and GB_UKM_solar_generation_actual.

## IV. RESULTS AND DISCUSSION

We implemented linear regression algorithm in order to predict the solar energy generation from the above weather quantities. To evaluate the models' performance on data set, R squared, Mean Absolute Error (MAE) and Root Mean Square Error values have been computed. (RMSE) values and obtained the result as $R^{2\,s}$= 0.925,

MAE = 290.903 and RMSE = 529.114.

In order to check the ability and efficiency of proposed model, we compare our regression-based prediction models with existing systems on the basis of accuracy. This comparison is performed with respect to statistical error measures such as MAE, RMSE and $R^2$_score. First, the Results of proposed work are compared with "Short term solar energy prediction system" [18]. The work validated two datasets A and B on different approaches.

Moreover, it determined performance on two different solar stations simultaneously as well as on single station. However, the work gives performance on several approaches; we compared the results obtained from regression-based technique of existing system. Table 1[(a) and (b)] below shows the MAE, RMSE, R2 values obtained from existing system:

**Table 1:** Performance Analysis of solar dataset A and B at two different solar stations

| SN. | Models | Solar Dataset A | | | Solar Dataset B | | |
|---|---|---|---|---|---|---|---|
| | | MAE | RMSE | R2_score | MAE | RMSE | R2_score |
| 1 | Linear Regression – I | 4880032.24 | 6798604.39 | 0.2638796 | 4820050.11 | 6616022.41 | 0.3025184 |
| 2 | Linear Regression – II | 5486982.49 | 7252964.41 | 0.1621314 | 4820050.12 | 6616022.42 | 0.3025184 |

**Table 2:** Performance Analysis of solar dataset A and B at one solar station

| SN. | Models | MAE | RMSE | R2_score |
|---|---|---|---|---|
| 1 | Linear Regression- I | 4360158.37 | 5958222.87 | 0.432 |
| 2 | Linear Regression- II | 5031607.95 | 6563079.72 | 0.310 |

Improvements in our proposed work can be observed with $R^2$ value as 0.9247, RMSE value as 529.1138 and MAE as 290.9028.

Secondly, we compared with another demonstrated “linear regression-based model” [19]. This comparison is done on the bases R2 score. The result shows that proposed system provides a much more accurate model. The R2 Score for each model highlight this result: The R2 value for proposed model and existing model obtained is 0.9247 and 0.6054 respectively. Thus, Percentage improvement in R2 value is about 31% in a proposed work. These comparisons show presented work demonstrates enhanced performance.

## V. CONCLUSION

We have successfully implemented a system that can accurately predict or forecast a renewable energy for an integrated smart grid using machine learning algorithms and techniques. So, main objective of the proposed work that is to benchmark the different techniques to forecast the renewable energy and to build a unified forecasting model to predict renewable energy generated by solar energy sources is completely done under a machine learning algorithm. The accuracy of prediction and forecasting in the existing system was low and it was necessarily required to being enhanced. So, we made system that can accurately do the forecasting of renewable energy system. Now using the raw data and processing the database completely and accurately can give us the results that can be used in many operations and working field. As the demand of renewable energy sources are extremely high and increasing day by day our system can make quite a difference and be very helpful for many industries.

## ACKNOWLEDGEMENTS

The authors can acknowledge professor, friend or family member who help in research work in this section.

## VI. REFERENCES

[1] Buturache, A. and Stancu, S. (2021) Wind Energy Prediction Using Machine Learning. Low Carbon Economy, 12, 1-21.

[2] Amila T. Peiris, Jeevani Jayasinghe, Upaka Rathnayake, "Forecasting Wind Power Generation Using Artificial Neural Network: “Pawan Danawi”—A Case Study from Sri Lanka", Journal of Electrical and Computer Engineering, vol. 2021, Article ID 5577547, 10pages, 2021.

[3] Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. Energy Convers. Manag. 2019, 198, 111799.

[4] F. Carmo, J. F. Martins and M. Sănduleac, "A methodology to assess home PV capacity to mitigate wind power forecasting errors," 2017 International Young Engineers Forum (YEF-ECE), 2017, pp. 47-52, doi: 10.1109/YEF-ECE.2017.7935639.

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

[5] What is Linear Regression

[6] Detailed Explanation of Simple Linear Regression, Assessment and, Inference with ANOVA

[7] Asrari, A.; Wu, T.X.; Ramos, B. A hybrid algorithm for short-term solar power prediction—Sunshine state case study. IEEE Trans. Sustain. Energy 2016, 8, 582–591.

[8] Hu, J.; Wang, J.; Zeng, G. A hybrid forecasting approach applied to wind speed time series. Renew. Energy 2013, 60, 185–194.

[9] Ahmad, A.; Hassan, M.; Abdullah, M.; Rahman, H.; Hussin, F.; Abdullah, H.; Saidur, R. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustain. Energy Rev. 2014, 33, 102–109.

[10] Wang, R.; Li, J.; Wang, J.; Gao, C. Research and application of a hybrid wind energy forecasting system based on data processing and an optimized extreme learning machine. Energies 2018, 11, 1712.

[11] Mendonça de Paiva, G.; Pires Pimentel, S.; Pinheiro Alvarenga, B.; Gonçalves Marra, E.; Mussetta, M.; Leva, S. Multiple Site Intraday Solar Irradiance Forecasting by Machine Learning Algorithms: MGGP and MLP Neural Networks. Energies 2020, 13, 3005.

[12] Ju, C.; Wang, P.; Goel, L.; Xu, Y. A two-layer energy management system for microgrids with hybrid energy storage considering degradation costs. IEEE Trans. Smart Grid 2017, 9,6047–6057.

[13] Diao, L.; Niu, D.; Zang, Z.; Chen, C. Short-term Weather Forecast Based on Wavelet Denoising and Catboost. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 3760–3764.

[14] Catalao, J.; Pousinho, H.; Mendes, V. Hybrid wavelet-PSO-ANFIS approach for short-termwind power forecasting in Portugal. IEEE Trans. Sustain. Energy 2010, 2, 50–59.

[15] Yu, J.; Park, J.H.; Kim, S. A New Input Selection Algorithm Using the Group Method ofData Handling and Bootstrap Method for Support Vector Regression Based Hourly Load Forecasting. Energies 2018, 11, 2870.

[16] Cao, Z.; Wan, C.; Zhang, Z.; Li, F.; Song, Y. Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting. IEEE Trans. Power Syst. 2019, 35, 1881– 1897.

[17] Linear Regression in Machine Learning

[18] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, 'Predicting solar generation from weather forecasts using machine learning', 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE, Oct-2011.

[19] S. Ibrahim, I. Daut, Y. M. Irwan, M. Irwanto, N. Gomesh, and Z. Farhana, 'Linear Regression Model in Estimating Solar Radiation in Perlis', Energy Procedia, vol. 18. Elsevier BV, pp. 1402–1412, 2012.