Forecasting renewable energy for an integrated smart grid

Submitted in partial fulfilment of the requirements of the degree

BACHELOR OF ENGINEERING IN **COMPUTER ENGINEERING**

By

Osama Saleh 3119035

Qureshi Waqas Raza 3119034

Sharjil Ali Momin 3119030

Name of the Mentor

Prof. Mohammed Ahmed Shaikh

Department of Computer Engineering

M.H. SABOO SIDDIK COLLEGE OF ENGINEERING

M.H. SABOO SIDDIK COLLEGE OF ENGINEERING University of Mumbai

(AY 2021-22)

CERTIFICATE

This is to certify that the Mini Project entitled **Forecasting renewable energy for an integrated smart grid** is a bonafide work of **Osama Saleh (3119035) Qureshi Waqas Raza (3119034) Sharjil Ali Momin (3119030)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering” .**

(Prof. Mohammed Ahmed Shaikh)
Mentor

(Prof.Zainab Pirani)
Head of Department

(Prof.Ganesh Kame)
Principal

Mini Project Approval

This Mini Project entitled **Forecasting renewable energy for an integrated smart grid** by **Osama Saleh (3119035) Qureshi Waqas Raza (3119034) Sharjil Ali Momin (3119030)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1……………………………………………
(Internal Examiner Name & Sign)

2……………………………………………
(External Examiner name & Sign)

Date:

Place:

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 is 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: Energy management, forecasting models, photovoltaic system, smart grid, solar energy, Machine learning algorithms(linear regression algorithm).

Acknowledgement

We wish to express our sincere thanks to our director Dr. Mohiuddin Ahmed and our principal Dr. Ganesh Kame,M.H. Saboo Siddik College of Engineering for providing us all the facilities, support and wonderful environment to meet our project requirements.
We would also take the opportunity to express our humble gratitude to our Head of Department of Computer Engineering Dr. Zainab Pirani for supporting us in all aspects and for encouraging with her valuable suggestions to make our project success.

We are highly thankful to our internal project guide Mohammed Ahmed Shaikh Whose valuable guidance helped us understand the project better, her constant guidance and willingness to share her vast knowledge made us understand this project and its manifestations in great depths and helped us to complete the project successfully.

We would also like to acknowledge with much appreciation the role of the staff of Computer Department, especially the Laboratory staff, who gave the permission to use the labs when needed and the necessary material to complete the project.

We would like to express our gratitude and appreciate the guidance given by other supervisors and project guides, their comments and tips helped us in improving our presentation skills.

Although there may be many who remain unacknowledged in this humble note of appreciation but there are none who remain unappreciated.

Table of Contents

Sr. No	Topic	Page No.
	Abstract	4
1	**Introduction**	7
1.2	Motivation	9
1.3	Aim & Objectives	9
2	**Literature Survey**	10
3	**Proposed System**	12
3.1	Introduction	12
3.2	Algorithm & Process Design	12
3.3	Architecture	15
3.4	Details of Hardware & Software	16
3.5	Analysis and Results	17
3.5.1	Comparing with existing system	21
4	**Scopes**	22
5	**Conclusion and Future work**	23
	References	24

List of Figures

Fig. No.	Figures	Page No.
1	Regression graph for Independent VS dependent Variables	12
2	Graph for Cost function	13
3	Process flow of Linear Regression Algorithm	15
4	Architecture of the system	15
5	Graph: Actual solar generation in UK in MW	17
6	Graph: Diffused horizontal radiation	18
7	Graph: Direct horizontal radiation	18
8	Graph: Temperature (Celsius)	19
9	Corelational matrix	19
10	Graph: GB_radiation_direct_horizontal and GB_UKM_solar_generation_actual	20
11	Graph: GB_radiation_diffuse_horizontal and GB_UKM_solar_generation_actual	20

1. Introduction

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. ***"Renewable energy forecasting is the technique of gathering and analyzing data in order to predict Renewable-energy generation across various time horizons with the objective to reduce the impact of energy intermittency. This technique of predicting future renewable energy needs to achieve demand and supply equilibrium".***

Generally, the energy forecasting can be done in different fashions:

- Now-casting (Forecasting few hours ahead)
- Short-term forecasting (minimum up to week ahead) and
- Long-term forecasting (Weeks, months, years ahead).

Identifying the importance of renewable energy forecasting, researchers have applied various technologies to forecast renewable energy. Many studies have revealed that various machine-learning models have aims to gain more reliable and accurate systems and has proven to be an excellent tool for solving various energy applications problems. The data-driven models do provide realistic ways of renewable energy predictions. In addition, hybrid machine-learning methods were created to improve renewable energy prediction accuracy. In this proposed work, we have used machine learning to predict the renewable energy generation by power sources. The data we used in this project are an ordered collection sampled and recorded at a particular time interval, so they are a time series. As a case study, we obtained the actual data from Open Power System Data [a free-of-charge platform with data on installed generation capacity by country/technology, individual power plants (conventional and renewable), and time series data]**.** https://open-power-system-data.org/ . This platform 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, but in this project, we focused on data for United Kingdom in 2019 as an example. In particular, we have utilised two datasets: **(1) Time series** with load, solar, prices in hourly resolution. **(2) Weather data** with radiation, temperature and other parameters.

The remainder of the paper is arranged as follows: Section 2 discusses publications, articles, and related materials. **Section 3** describes the algorithm, process design, architecture of propose system, Analysis and Results. **Section 4** describe about the scopes of the work. **Section 5** highlights the conclusion and future work of the presented work.

1.2 Motivation:

With the growing usage of renewable energy resources in the modern power system has made energy forecasting a popular theme. It is very essential for grid operators and decision-makers to know how much power renewable energy sources(RES) will produce in near future. Along with this, prediction of load demand and consumption is crucial in management and planning for the power system. 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. These constraints give rise to various interesting characteristics of energy forecasting, which include data collection and the need for precise accuracy. Forecasting errors lead to unbalanced supply-demand, which negatively affects the operational cost, reliability and efficiency. These important factors and demands of forecasting renewable motivate us to detailed study this topic by implementing a unified model.

1.3 Aim & Objectives:

- To develop a system to accurately forecast renewable energy for an integrated smart grid using machine learning.
- The main objective of the proposed work is to benchmark the different techniques to forecast renewable energy and to build a unified forecasting model to predict renewable energy generated by solar energy sources.

2. Literature Survey

As solar energy becomes more integrated into the energy system, forecasting solar power generation becomes more critical for controlling energy quality and improving system reliability. To estimate sunshine per hour, Asrari et al [7]. proposed utilising a hybrid prediction system. Many cost-effective and cost-improvement approaches and model structures have been developed, however, due to the competitiveness and weight of accounting. Gradient-descent optimization is employed in the first stage to meet the artificial neural network's basic criteria. In the second stage, an ensemble empirical mode decomposition (EEMD) meta-optimization model is constructed to determine the best artificial neural networks.

Hu et al. [8] employed a hybrid method that included ensemble empirical mode decomposition and support vector machine degradation to improve the quality of wind speed prediction. Using the proposed strategy, better prediction results were obtained. Ahmad et al. [9] gave an overview of how power forecasting outcomes have evolved over time utilising artificial intelligence approaches like support vector machines and artificial neural networks (ANN). They came to the conclusion that the hybrid strategy is better for forecasting energy usage. The enhanced cuckoo search-extreme learning machine (ELM) model is used as the basic model in Rui et al’s [10] technical document. It creates a hybrid model by combining the standard genetic algorithm with the auto-regressive and moving average models. Following the order of experimental analysis, it can be shown that the upgraded cuckoo search-extreme learning machine model solves the performance restrictions; that is, the initial weight growth is not easy to make; it is varied and steady. The distance between the test input vector and the distribution matrix was determined based on the results.
The upper and lower prediction boundaries of the anticipated values were provided, as well as a feasible predictive sample space. By enhancing the accuracy of the performance prediction range, the proposed strategy might avoid major single point prediction failures.

In order to integrate natural renewable energy into electrical utility systems, it is necessary to predict sun radiation in photovoltaic power generation. Paiva et al. [11] looked at two machine learning techniques for predicting sunlight throughout the day: multi-gene programming and multilayer perceptron. The findings revealed that the model's accuracy is influenced by site definitions, prediction ranges, and error computations. Ju et al. [12] looked at a degradation cost model and developed a two-tier discrete management system with a hybrid energy storage system. They posed the problem in such a way that it may be solved with low operational expenses while taking into consideration renewable energy's energy variations. They also devised a more cost-effective cost model for batteries and supercapacitors, allowing them to convert long-term capital expenses into short-term operational issues.
Microgrids with common points of connection to the public grid, hybrid energy storage systems, renewable energy systems, and total load employ the proposed energy management system. They ran a simulation and discovered that

different energy storage technologies, such as super batteries and capacitors, can be employed for a variety of decision-making tasks in different tiers. However, in their current work, they did not take into account merging random planning and the proposed energy management system with renewable energy forecast and modelling uncertainty in renewable energy production. Li et al. [13] used wavelet noise reduction and CatBoost-based models for short-term weather forecasting based on the Beijing Meteorological Administration's weather observation and forecast dataset.

Catalao et al. [14] suggested a new hybrid technology for forecasting short-term wind power in Portugal in their research. A wavelet transforms, particle swarm optimization, and an adaptive network-based fuzzy inference system were combined in their suggested solution. The proposed method is a game-changer in terms of wind energy forecasting. The average results outperform the other approximations, allowing for a shorter calculation time. As a result, the findings back up their proposed short-term wind energy projection approach. Yu et al. [15] developed a new import selection algorithm that combines the group data handling approach with SVR to forecast short-term hourly load. They set the network multiple times after configuring the group technique of data handling networks multiple times under the identical testing conditions. Due to the random distribution of the training dataset, each entry was unique. Only one network configuration has the potential to cause skewed input selection outcomes. They used hourly Korean datasets to compare the estimated effects for one hour, one day, and one week to illustrate the execution of the strategy proposed in the experimental evaluation. Experiments reveal that the proposed strategy beats existing strategies in terms of expected performance.

Zhaojing et al. [16] developed a probabilistic low voltage load hybrid ensemble deep learning model with consistent and generally accurate predictions. Deep training models and deep belief networks were used to map nonlinear relation graphs. To improve network stability, five-way bagging and boosting algorithms were utilised. During the installation procedure, various transmission mechanisms were used to assure the site's safety during the loading and unloading operation. This was validated by the accuracy of the augmented Dickey–Fuller test. The results of the experiments demonstrate the great potential of real-world applications in the distribution network and have a substantial impact on decision-making and assumptions.

3. Proposed system:

3.1 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 propose work we have go through several commonly used machine learning techniques. Linear Regression algorithm is proved the most basic and widely used technique for the forecasting purpose.

3.2 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.

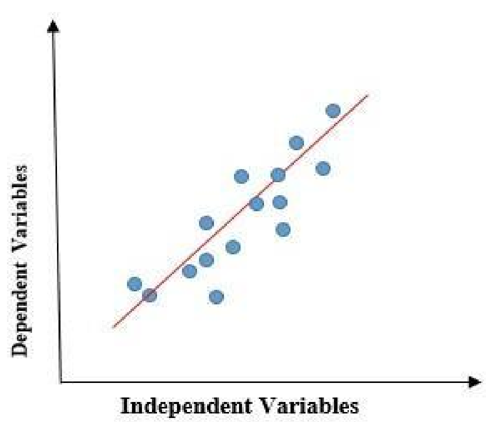

Fig.1 Regression graph for Independent VS dependent Variables

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 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.

<u>For Simple Linear Regression:</u>

$$\mathbf{y = c \quad + mx}$$

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

<u>For Multiple Linear Regression:</u>

$$\mathbf{Y = \theta_1 + \theta_2.X + \theta_3.X_1 + \theta_4.X_3 + ... + \theta_n.X_n}$$

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 θ_1 and θ_2 values.

θ_1: intercept
θ_2: coefficient of x
Once we find the best θ_1 and θ_2 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.

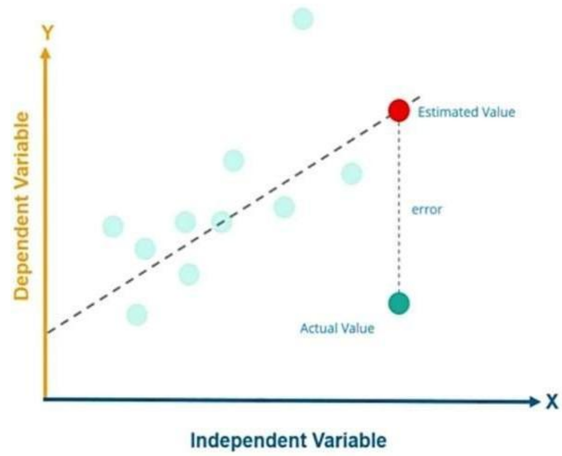

Fig.2 Graph for cost function

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.

<u>Cost Function:</u>

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 $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ 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 θ_1 and θ_2 values in order to optimize the Cost function and achieving the best fit line. The idea is to begin with random values of θ_1 and θ_2 and then iteratively updating the values until minimum cost is achieved.

Process design for linear regression algorithm:

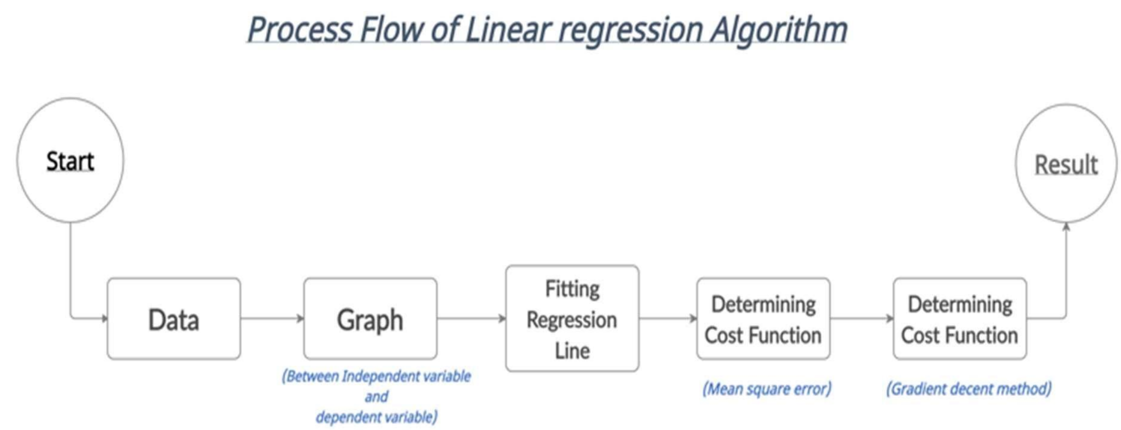

Fig. 3 Flowchart of linear regression algorithm

3.3 Architecture

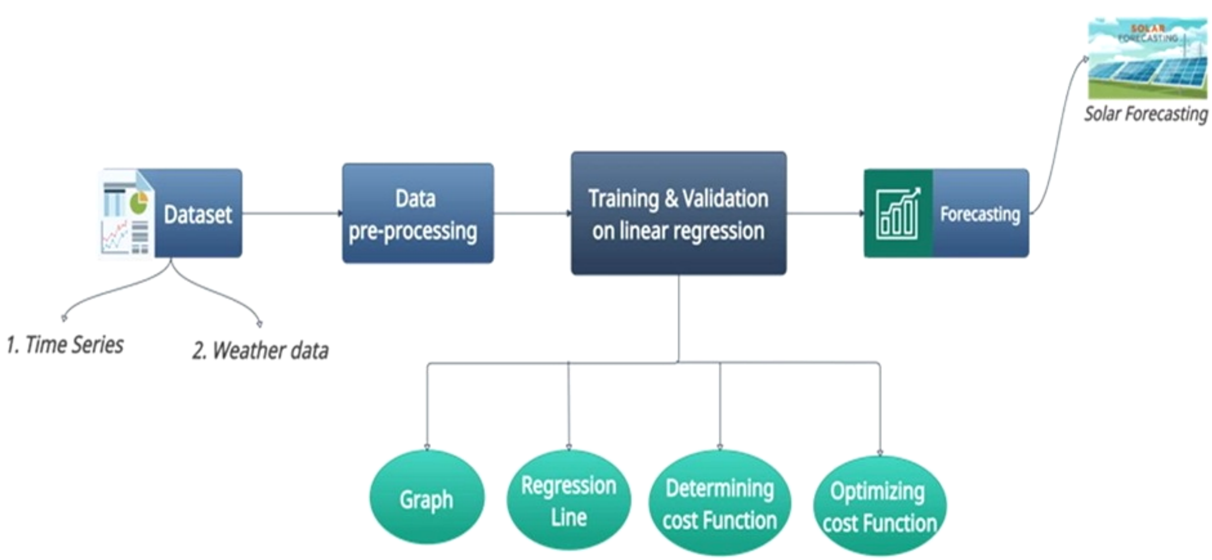

Fig.4 Architecture of the system

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 utilised 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 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.

3.4 Details of Hardware & Software

The most common set of hardware and software requirements for the product to operate efficiently without any lag or issue is given following.

- **Recommended Operating System**

  Windows: 7 or newer

  Mac: OS X v10.7 or higher

  Linux: Ubuntu

- **Hardware Requirements**

1. **Processor:** Minimum 1 GHz i.e. Intel or AMD processor with 32-bit support; Recommended 2.8 GHz or faster i.e. 2GHz or more
2. **Ethernet Connection (LAN) OR a (Wi-Fi) is required for software activation**
3. **Hard Drive:** Minimum 32 GB; Recommended 64 GB or more
4. **Memory (RAM):** Minimum 2 GB; 4 GB or above is recommended
5. **GPU:** Integrated Graphics

- **Software Requirements**

1. **Scikit-Learn :** https://scikit-learn.org/stable/
2. **Matplotlib:** https://matplotlib.org/
3. **Seaborn:** https://seaborn.pydata.org/
4. **Pandas:** https://pandas.pydata.org/

3.5 Analysis and Results:

After filtering time series data for the rows for Uk 2019, we end up with a DataFrame 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;

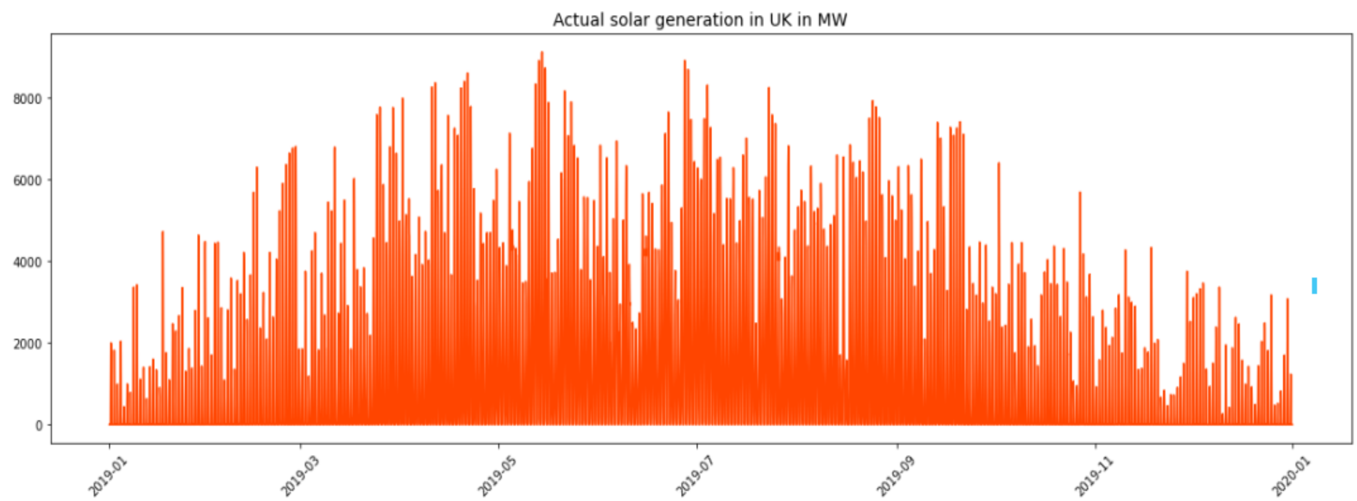

Fig.5 Actual solar generation in UK in MW

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 characterised by the different quantity as follows:

- GB_temperature
- GB_radiation_direct_horizontal
- GB_radiation_diffuse_horizontal

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

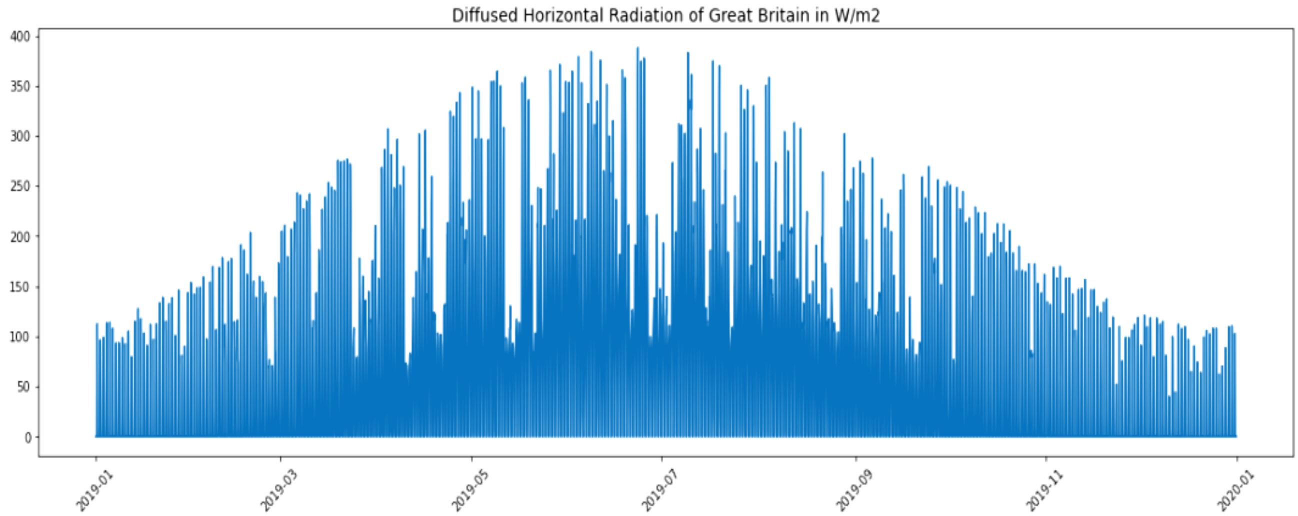

Fig.6 Diffused horizontal radiation

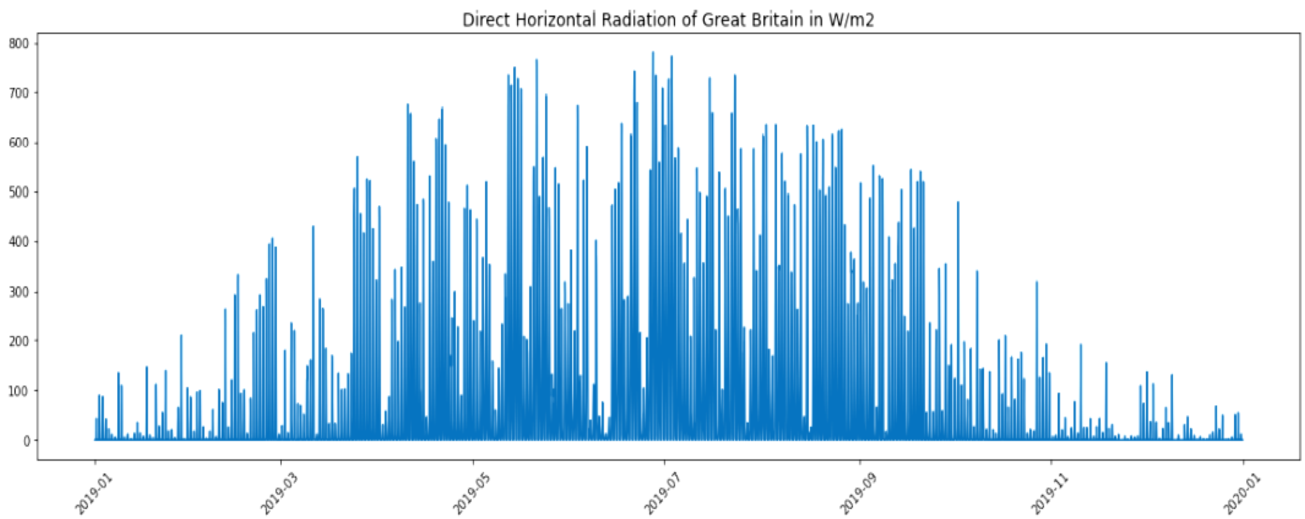

Fig.7 Direct horizontal radiation

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

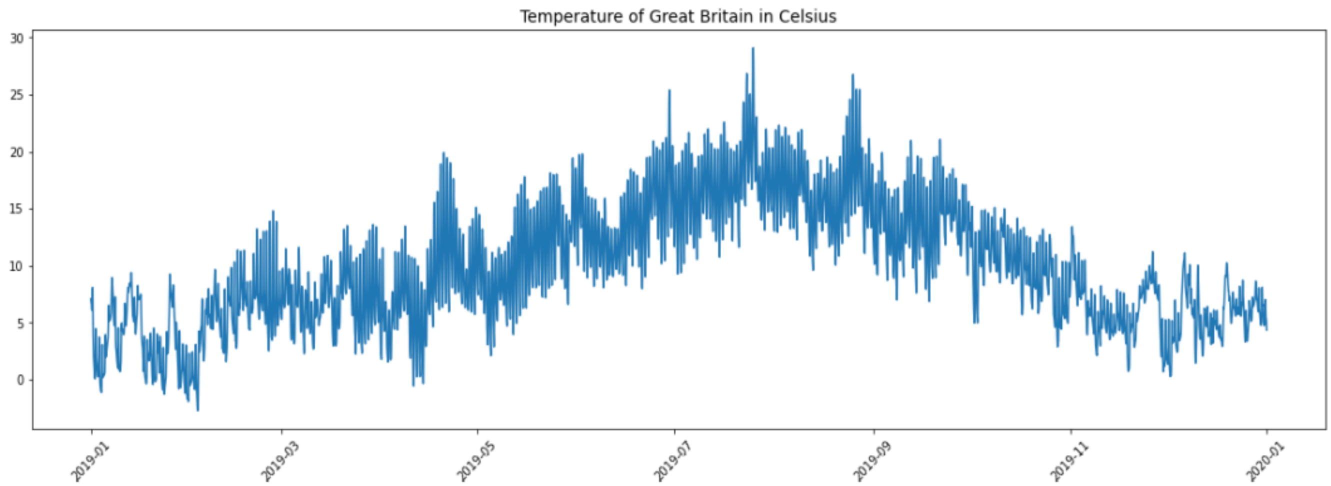

Fig.8 Temperature (Celsius)

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.9;

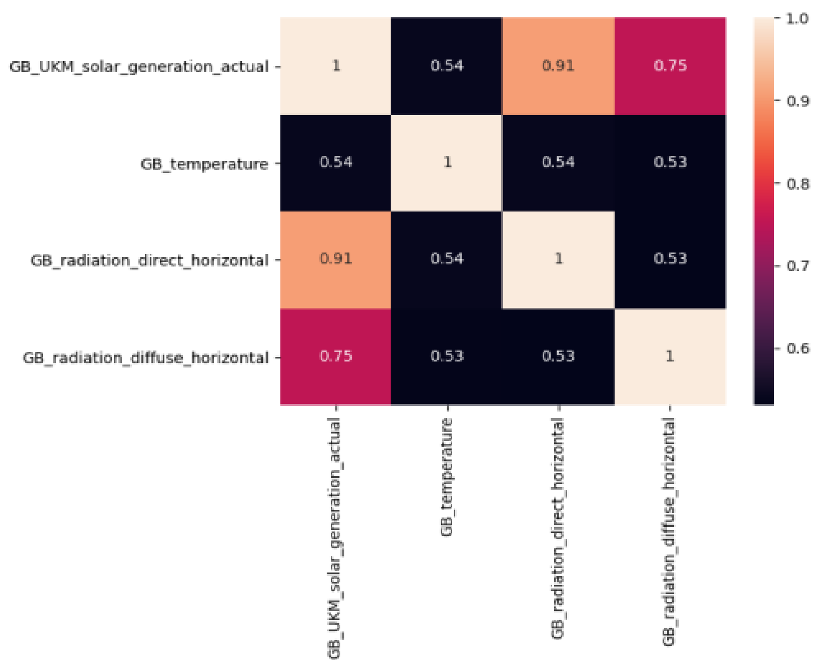

Fig.9 Corelational matrix

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

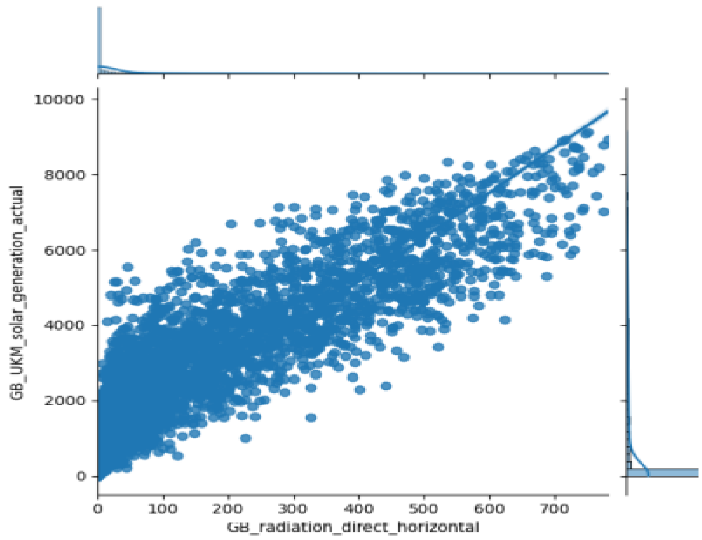

Fig.10 GB_radiation_direct_horizontal and GB_UKM_solar_generation_actual

The Fig.10 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;

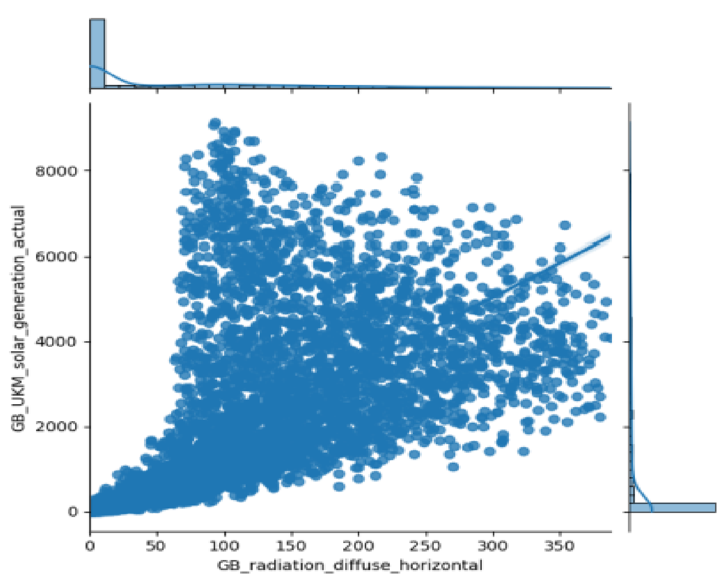

Fig.11 GB_radiation_diffuse _horizontal and GB_UKM_solar_generation_actual

Given these observations, 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**.

3.5.1 Comparing with Existing System

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(a): Performance analysis of solar dataset A and B at two different solar stations.

	Solar dataset A			Solar dataset B		
Models	MAE	RMSE	R2_score	MAE	RMSE	R2_score
Linear regression-I	**4880032.24**	**6798604.39**	**0.2638796**	**4820050.11**	**6616022.41**	**0.3025184**
Linear regression-II	5486982.49	7252964.41	0.1621314	4820050.12	6616022.42	0.3025184

Table 1(b): Performance analysis of solar dataset A and B at one solar station.

Models	MAE	RMSE	R2_score
Linear regression-I	4360158.37	5958222.87	0.432
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.

4. Scopes:

After reviewing the related research work on forecasting renewable energy, and determining possible real-world situations where such systems would be of use significantly. We have identified the several important potentials of this system.

As Renewable energy forecasting provides valuable information about the expected changes in the energy to be generated in the near future it has many scopes in real-life problems, some of them are listed below:

1. The major scope of renewable energy is to maintain supply demand equilibrium.
2. It plays an important role for proper functioning and management of power system.
3. Due to uncertainty in energy by renewable energy sources, forecasting future energy generation proved significant for operators and decision-making authorities for future planning Operational problems in energy sector.
4. It Reduced Energy Losses, and conserve energy for future use.
5. Energy forecasting using renewable energy sources not only helps to reduce operational cost but also improve reliability of energy and helps to mitigate the greenhouse effect.
6. Now-casting (Forecasting few hours ahead) can help renewable energy companies and grid operators to improve risk management practices, and to increase the stability of the grid.
7. Seasonal forecasts can assist renewable energy producers in planning their production and maintenance schedule by determining available electric power in the near future with high precision and over longer time periods, resulting in increased performance for renewable energy businesses.
8. Forecasts for the weeks and months ahead can also alert power providers to the possibility of disruption and damage from storms and flooding at an earlier stage, allowing them to better prepare for such situations.

5. Conclusion and future work

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 algorithms. 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 operation 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.

Future Works:

- The accuracy of system can be increased and made the system accurate completely. i.e. 100 percent accurate.
- We can make system more advance so it can forecast solar energy, wind energy hydroelectric energy as well as geothermal energy based on the data given.
- Two cleanest sources of renewable energy are solar and wind which have gained a lot of popularity in production for residence as well as for the national grid. These are promising sources which can not only produce clean energy but also they can add a lot to local economies.
- Clean energy in future will is very important as it can contribute to many things like in improving global health, providing jobs and promoting economic growth. It is far better for environment and will help in reducing the risks of flood and droughts.
- As the power functioning and forecasting will be done accurately and completely more field related to power supply and consumption will be rising in near future and ever increasing energy demand will increase even more which will give rise to the world completely based on the renewable energy power and using fossil fuels would be seeing as an old-fashioned way of doing things.
- Solar energy will become 35 percent cheaper by 2025 and wind energy capacity

will be increased by 75 percent by the year of 2025.

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, 10 pages, 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.

[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-term wind 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 of Data 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.