Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FORECASTING SOLAR PHOTOVOLTAIC ENERGY PRODUCTION USING LINEAR REGRESSION-BASED TECHNIQUES

SHEREEN A. EL-AAL[1] , MAHA ABDULKARIM ALQABLI[2] , AMANY A. NAIM[1]

[1]Faculty of Science, Al-Azhar University, Cairo, Egypt
[2]University College of Duba, University of Tabuk,Tabuk, Saudi Arabia
E-mail: [1]shereen.a.elaal@azhar.edu.eg, [2]malgibly@ut.edu.sa, [1]amany.naim@azhar.edu.eg

ABSTRACT

Photovoltaic (PV) systems are now viewed as being crucial to the advancement of renewable energy. In order to maximize the performance of the power grid in accordance with market demands and prevent issues with solar generation due to instability, smooth solar energy generation requires precise and trustworthy forecasts. In order to estimate the solar system's output power, machine learning regression models were used to assess the solar system's performance. For the solar generation dataset, machine learning models based on linear regression, such as Multiple, Ridge, Lasso, Decision Tree, and Polynomial regression, have been applied. Performance measures were used to quantify how well the behavior of the suggested technique fit the dataset. The study's findings demonstrate a 93.7% accuracy rate for the polynomial regression model with a lower value of Mean Square Error (MSE).

Keywords: *Photovoltaic, Solar Energy, Machine Learning, Regression, Polynomial Regression*.

1. INTRODUCTION

Energy is one of the primary sources for economic development and industrialization. Therefore, rationalizing the use of energy and renewable energy resources is one of the most important research points. Renewable energy plays an increasing role in achieving the goals of sustainable development and protecting the environment [1].

At present, solar energy is recognized as one of the most promising clean sources of energy in the world. Solar energy consumption rises as technology advances, which leads to an increase in the number of applications that use it [2]. Furthermore, prediction of photovoltaic systems is a key aspect of science to support efforts to integrate solar energy with power grids and it improves the planning and operation of PV systems [3]. In addition, power prediction maintains a balance between power generation and consumption, which leads to the economic advantages of electrical utilities.

On the other hand, Machine Learning (ML) is primarily concerned with designing and developing algorithms that allow a system to learn from historical data in such a way that they can learn and improve their performance automatically. Supervised machine learning, including classification and regression, is used to predict a discrete class label and a continuous quantity, respectively [4]. Photovoltaic systems can be predicted using statistical methods such as regression models. Several models based on linear regression can be used to predict the solar energy covering a given area in a given time period, such as multiple linear regression [5].

In this research, the proposed model used different types of regression techniques, such as simple

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

linear regression, multiple linear regression, lasso regression, ridge regression, decision tree regression, and polynomial regression, for predicting the amount of power generated by a PV system.

The remainder of this study is organized as follows: In section 2, a literature review of the work developed with techniques based on regression is provided. Section 3 describes the methodologies used in the proposed model. In Section 4, the proposed model is introduced. The description of the dataset, dataset visualization, and the experimental results are introduced in Section 5. Finally, the conclusion and future work are discussed in the last section.

2. RELATED WORK

Authors in [1] studied the relationship between solar radiation and temperature and applied this relationship to different datasets for various countries. They used the linear regression technique to assess the solar radiation. In [3], the researchers applied regression models using only one year of data to better estimate solar power generation from the overall data usage. The researchers found that the saturation intensity variables are the most important feature of solar power generation in daylight.

Thombare et al. in [5] used Python to apply a linear regression model to predict solar energy. They used the root mean square error and score value to evaluate the model. Rishal et al. [6] proposed model by using a simple regression by Python open source. This model can be used anywhere as long as all the necessary conditions are met for the model.

The researchers in [7] proposed a mathematical model to forecast solar energy. The ambient temperature, cell temperature, and sun irradiation are the main variables considered when estimating solar power in their model. A multiple linear regression was used by Mohamed and Badrul in [8] to forecast the possibility of solar energy. The artificial neural network technique and regression are used in [9] to estimate the amount of solar energy. They used the weather conditions in India and Malaysia. They evaluated the proposed model by using root mean square error. In [10], Chia-Sheng T. et al. used grey wolf optimization and regression neural networks to predict the photovoltaic systems with more accuracy in less time.

The researchers in [11] proposed model by using the decision trees and the regression. Tests on different datasets show the effectiveness of the method compared to the output of traditional regression and decision tree methods. The polynomial regression model has been used to characterize the relationship between stresses and drilling depth in [12]. In [13], authors proposed a polynomial cube least squares regression to predict cost in the commercial sector.

3. METHODOLOGY

This section presents the methodologies used to predict the AC power generated by a solar system.

3.1. Regression

Regression is a statistical tool for assessing the relationship between a dependent variable and one or more independent variables. So, the regression analysis is used to estimate the expectation and conditional expectation of the dependent variable considering independent variables, and its use interferes with fields of machine learning [14, 15]. There are many regression analysis techniques available to make predictions [14]:

3.1.1. Simple linear regression

The most widely used modeling technique is linear regression. Whereas linear regression analysis is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used when studying the relationship between one independent variable and a dependent variable. Mathematically, the relationship between a single independent variable x and a dependent variable y, it can be represented as [15], [16]:

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (1)$$

Where, β_0 is the intercept and indicates the estimated value of y when x = 0. β_1 is the regression coefficient and indicates that the estimated increase in the dependent variable per unit increase in the independent variable. ε is the error term. The estimator of the linear regression is known by β as follows:

$$\beta = (X'X)^{-1}X'Y \qquad (2)$$

Where, Y and X are dependent and independent variable, respectively. X denotes the transpose of the matrix.

The sum of the squares of the remaining values must be the least, so the cost function for the linear regression becomes:

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij} \ \beta_j)^2 \qquad (3)$$

Where, the number of independent variables is represented by *n*. The number of the regression coefficients is represented by *p*.

3.1.2. Multiple regression

Multiple linear regression is identical to simple linear regression except:

- There is more than one independent variable.
- There must be linear relationship between the independent variables.

The function can be represented as [16]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (4)$$

3.1.3. Ridge regression

Ridge regression is one of the ways dealing with multi collinearity. The ridge regression estimators are completely dependent on an external parameter, which is known as the bias constant or the ridge parameter (L). The estimator of the ridge regression is known by βr as follows [17]:

$$\beta_r = (X'X + \lambda I)^{-1}X'Y \quad , (0 < \lambda < 1) \qquad (5)$$

I identities matrix and shows the slope of the ridge depends on adding the ridge parameter (λ = alpha) to the diagonal of the X'X matrix. So that the estimation of the ridge coefficient is affected by the amount of the ridge parameter (λ). Then, the cost function of the ridge regression is changed by adding (λ) to the square of the magnitude of the coefficients, and it will be as follows:

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij} \ \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (6)$$

Where, $\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij} \ \beta_j)^2$ is the loss term and $\lambda \sum_{j=1}^{p} \beta_j^2$ is the penalty term. So, the ridge regression imposes constraints on the coefficients (β). The penalty term adjusts the operand, so if the operand takes a large value, the optimization function will be penalized. Therefore, the ridge regression reduces the coefficients and helps reduce the complexity of the model.

3.1.4. Lasso regression

Lasso is the operator with the least absolute shrinkage and selection. The cost function for the lasso regression is represented as [18]:

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij} \ \beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j^2| \qquad (7)$$

The only difference between lasso and ridge is that instead of taking the square of coefficients, the quantities are considered. Thus, lasso regression can aid in not only reducing over fitting but also in feature selection.

3.1.5. Decision tree regression

Decision tree is used for a regression task, which can be used if the dependent variable is continuous. The tree structure is designed in the form of different types of nodes, roots, internals, and leaves. Decision tree is top-down, because it starts at the top of the tree and divides the predictor space in two. The root node is the starting point of the decision tree, and the leaf nodes are the final classes. Most common traditional statistical models, such as least-squares, logistic, poisson, and proportional hazards models, as well as models for long-term and multi-response data, can be fitted by regression trees [11, 19].

3.1.6. Polynomial regression

Polynomial regression is one of the most recent techniques used in regression and avoids all the problems found in other systems. Polynomial regression is considered a special case of linear regression in which the relationship between the independent variables and the dependent variables is studied. The power of the independent variable is greater than 1, and it can be expressed as [12]:

$$y = \beta_0 + \beta_1 x_1^1 + \beta_2 x_2^2 + \ldots + \beta_n x_n^n + \varepsilon \quad (8)$$

In some datasets, linear regression cannot model the relationship between the independent variable and the target variable, which means the straight line is not the best fit line to model the dataset. So, the polynomial regression technique is used if the relation between the independent variable and the target variable is represented as curvilinear, which means it fits the data points. This implies that the value of the target variable changes in a non-uniform way regarding the independent variables in a curvilinear relationship. Additionally, n denotes the polynomial's degree. The equation gets more challenging as the polynomial's degree increases. A polynomial's degree is determined by the correlation between its parameters and its target variables; the degree value must be greater than one. A degree of one indicates a straightforward linear regression. If the n value is carefully selected, the polynomial model will accurately match the data model. Furthermore, polynomial regression involves two phases. First, it transforms the data into a polynomial form and then fits the parameters via linear regression [13, 12].

3.2. Evaluation Metrics

- **Mean Absolute Error (MAE)**

MAE represents the difference between the true values and the expected values in testing the data set. MAE is calculated by the following equation [20, 21]:

$$MAE = 1/T \sum_1^T |u^* - u| \quad (9)$$

Where, T is the total number of samples. u^* and u are the expected values and the true values respectively.

- **Mean Squared Error (MSE)**

MSE is the difference between the true values and the expected values in testing the data set and squaring them. MSE is given by the following equation [20]:

$$MSE = 1/T \sum_1^T (u^* - u)^2 \quad (10)$$

- **Root Mean Squared Error (RMSE)**

RMSE is defined as the square root of the mean squared error. RMSE is given as follows [20]:

$$RMSE = \sqrt{1/T \sum_1^T (u^* - u)^2} \quad (11)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- **Coefficient of Determination (R2_score)**

R2_score is the coefficient of relevance of the expected values compared to the true values. R2 score ranges from 0 to 1 and is it is given by [21]:

$$R2\_score = 1 - \frac{\sum(u - u^{*})^{2}}{\sum(u - u^{-})^{2}} \qquad (12)$$

Where u^{-} is the mean value of u.

4. PROPOSED METHOD

This section provides an overview of the proposed system model.

The amount of energy produced is determined by the amount of sunlight that reaches the solar panels. As noticed in Figure 1, solar panels are made up of arrays of solar modules, which are made up of smaller solar cells. These cells are also known as photovoltaic cells (PV). PV systems produce energy in the form of direct current (DC) and the DC power can then be stored in a battery. Then this DC energy is passed through an inverter to convert it to Altering Current (AC), which is better suited for transmission through the centralized electrical panel. This electrical panel separates that home solar power into individual circuits that run household appliances. Solar energy production is exposed to instability in the grids, and this happens because the intensity of sunlight varies from time to time. In addition, another problem for solar energy is the imbalance between supply and demand. Therefore, tracking solar system performance and measuring the amount of AC power is essential to meet market demand as well as detecting when the system needs maintenance.

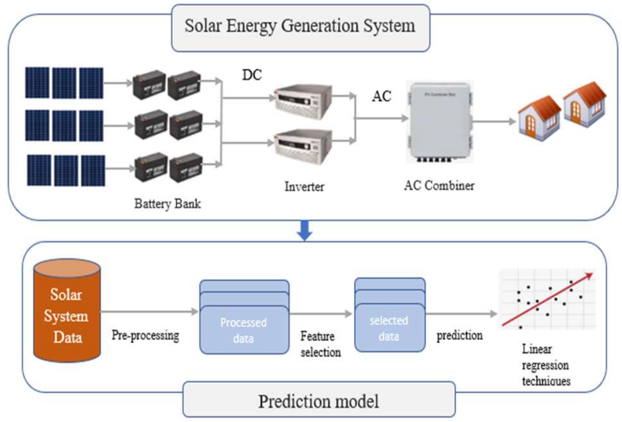

Figure 1: Proposed model for AC power Prediction

As shown in the Figure 1, in the proposed model, solar system data passes through three steps, including processing and cleaning, feature selection and prediction based on regression techniques. The techniques based on linear regression used to predict the amount of generated AC power, including multiple, Ridge, Lasso, Decision tree and Polynomial regression. Furthermore, the performance of the proposed model is evaluated according to the performance metrics.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the PV solar power dataset is used for testing the proposed learning model, and the used dataset is analyzed and visualized to gain insights from the data. In the first subsection, the structure of the used dataset is explained. In the following subsection, a visualization of the dataset is presented. In addition, the results of the proposed model are presented and compared against performance metrics.

5.1. Dataset Description

Solar energy is one of the renewable green energy sources that receive a lot of sunlight. A solar power system converts sunlight into electrical energy through photovoltaic (PV) panels.

In this research, the PV solar power dataset used was obtained from different two solar planets in India [22]. Each plant produces two files of data: solar power generation data and weather data. In this research, data visualization and experimental results were made on plant-2 data. Data was collected every 15 minutes, and this was done over a period of 34 days. In addition, the planet-2 data file shows there are 22 inverters with only one weather monitoring unit.

5.2. Dataset Visualization and Analysis

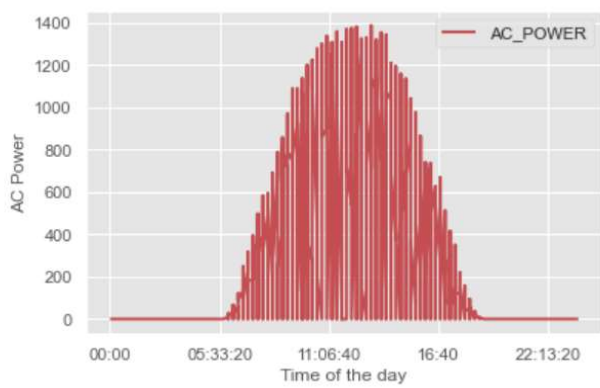

Figure 2: AC power generation against the time of day

The daily yield power rises from sunrise to sunset, remains constant until midnight, and then falls back to zero after midnight. The intensity of sunlight and the wavelength of sunlight striking the photovoltaic cells indicate the presence of sunlight. Figure 2, 3 show the generated power of AC and DC against the time of day. As shown in Figures, the AC power varies in the range [0, 1358.4] *W* and DC power varies in range [0, 1420.93] *W* with maximum power being achieved around 13: 00 PM.

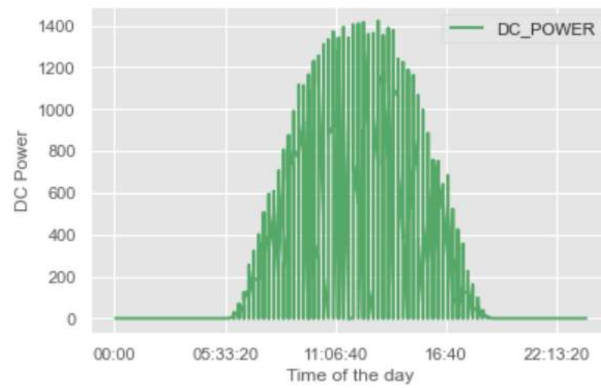

Figure 3: DC power generation against the time of day

Figure 4 depicts the average AC and DC power over 34 days, with 241.2778 W and 246.70196 W, respectively. As noticed in the figure, the maximum amount of AC and DC power is generated on 2020/05/25 and the minimum amount of generated power is achieved on 2020/06/11.

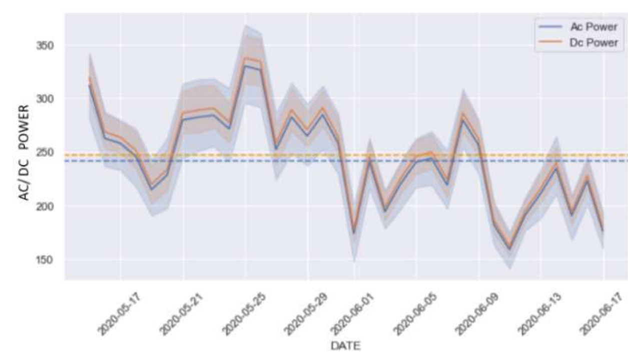

Figure 4: AC and DC power over 34 days

Figure 5 demonstrates the values of irradiation against time of day which mean the density of the sun. As noticed in the figure, the values of irradiation range from 0 to 1.099KW/m2, increasing around 5:40 am, peaking around 13: 00 PM and then gradually decrease until they reach 0 around 17:30 pm. Furthermore, Figure 6 shows the solar irradiation over 34 days, where the average of the irradiation over the recorded days equals 0.233. As shown in the figure, the maximum and minimum values of irradiation are obtained on 2020 /05 /17 and 2020/ 06/11, respectively. Figure 7 shows the range of ambient temperature

Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

throughout the day.

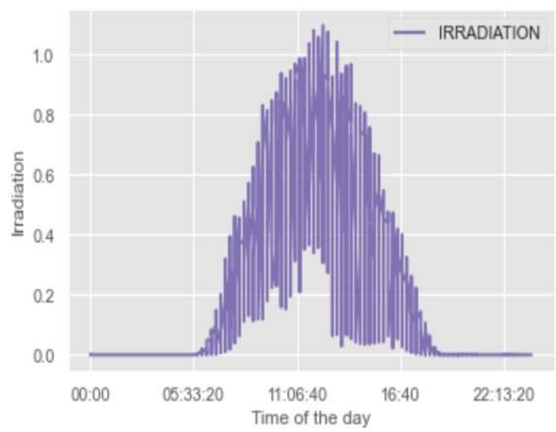

Figure 5: Irradiation against time of the day

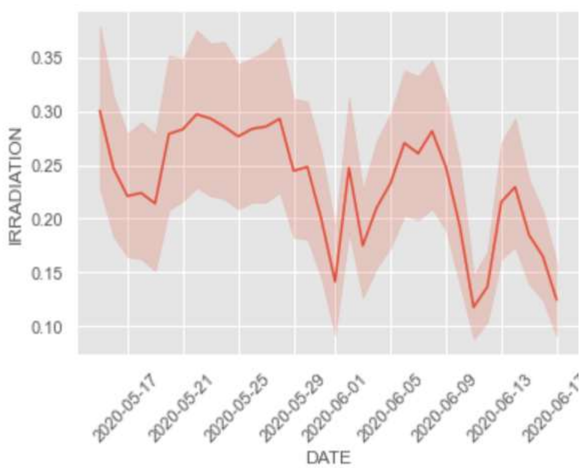

Figure 6: Solar irradiation over 34 days

As noticed in the figure, the maximum range of ambient temperature is achieved at 4.00 PM. The ambient temperature ranges from 20.94 to 39.18 degree C. In addition, Figure 8 shows the ambient temperature range over 34 days.

The yield DC power is passed through 22 inverters, whose output ranges from 520758 *W* for *"q49J1IKaHRwDQnt"* to 2247916000 *W* for *"9kRcWv60rDACzjR"* with an average output equal to 964540730.197 *W*. Figure 9 shows the inverter efficiency where the efficiency of an inverter is defined as the amount of *AC* output power it provides for a given DC input, and it is given by *Eff = Pac/Pdc* where *Pac, Pdc* represent *AC* output power and *DC* input power in watts, respectively. As shown in the figure, the inverter efficiency ranges from 97.76% to 97.92% with an average efficiency of 97.8%. The output efficiency of inverters is uneven, it may be due to various reasons, such as faults in the inverters or because the units connected to these inverters are in cloudy areas; this means that the system needs to be checked or maintained.

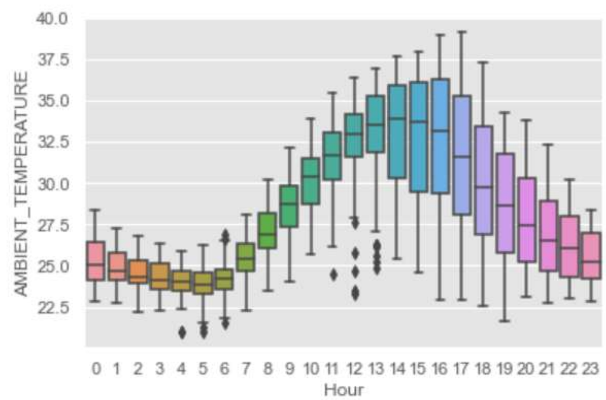

Figure 7: Ambient temperature throughout the day

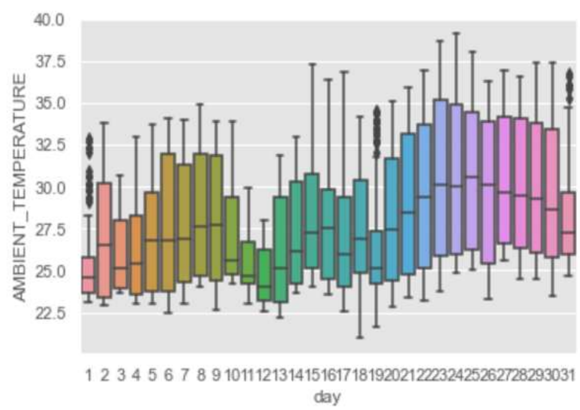

Figure 8: Ambient temperature over 34 days

5.3. Results and Discussion

After data visualization, machine learning methods are implemented on data to predict the *AC* power. Considering the data files, generation data and weather data files were merged to train and test the proposed models for predicting *AC* power. According to the demonstrated dataset, *DC* power attribute is excluded which *AC* power values are calculated depending on *AC* power. For the proposed model, forecasting is based on the attributes that have at least a 40% correlation with each other, which are irradiation, module temperature, and ambient temperature. Several regression methods are used for forecasting, including multiple linear regression, ridge regression, lasso regression, decision tree regression, and polynomial regression.

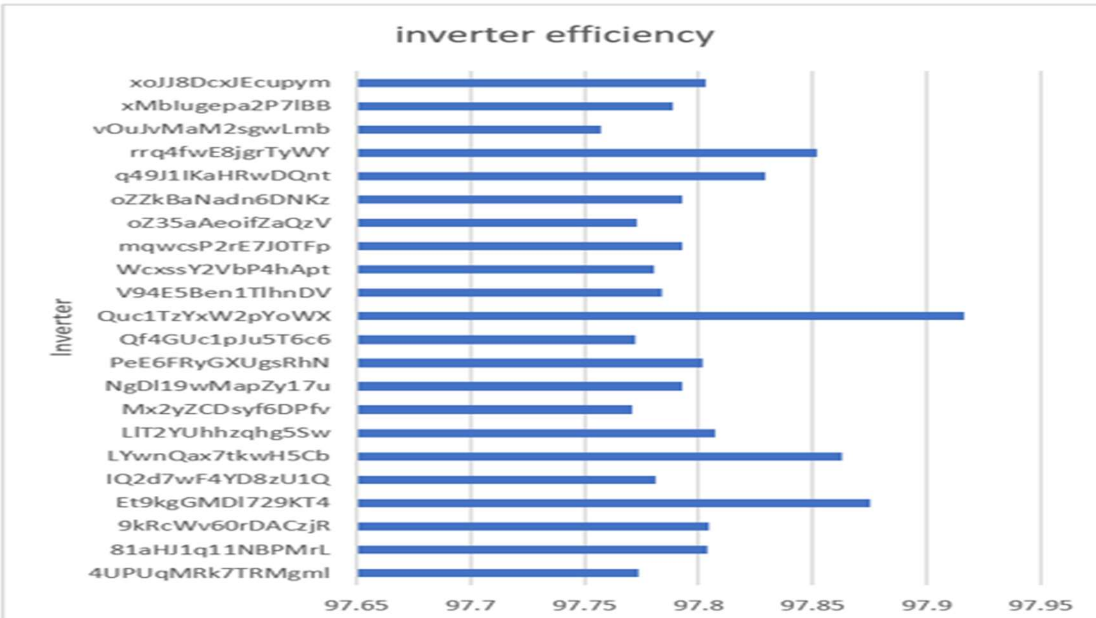

Figure 9: Inverter Efficiency

Firstly, multiple regression, which depends on dependent variables, is used without any regularization, which means the value of alpha is equal to 0, and the values of performance metrics as shown in table 1 are: MAE = 1585.39, MSE = 5688979.95, RMSE= 2385.16, and R2_score = 0.8624.

Subsequently, ridge and lasso regression were used to reduce the complexity of the model and prevent overfitting, which means different values of alpha. As a result, a cross validation using RidgeCV and LassoCV was performed to determine the optimal value of alpha for multiple regression in order to minimize MSE and optimize the models.

According to the cross-validation methods, the optimal values based on the RidgeCV and LassoCV were 0.001 and 1, respectively. As a result of the new alpha values, results for performance metrics remain virtually unchanged, as demonstrated in table 1. On the other hand, decision tree is used with hyper-parameter random state of 42, implying that the train and test sets are the same but have different executions. For the decision tree, although the performance metrics were significantly decreased and their values are: MAE = 970.84, MSE = 5213822.94, RMSE = 2283.38, and R2 score = 0.874, as shown in Table 1, it achieved underfitting to the model as shown in Figure 10.

Figure 10 shows the learning curve for different techniques based on linear regression. Multiple regression was run without regularization, ridge with an alpha value of 0.001, lasso with an alpha value of 1.0, and the learning curve for the decision tree is also shown in the figure. The difference in learning curves for multiple, ridge, and lasso is so convergent that it's almost

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

unnoticeable. For the decision tree, training error remains flat regardless of training set size, which means an underfitting model. In addition, Figure 11 shows the scatter diagram for actual power versus prediction for multiple, ridge, lasso and decision tree, which they are in roughly a good pattern.

Figure 12 shows the performance of the polynomial regression model with respect to the learning curve using three different values of the hyperparameter degree, which are 2, 3, and 4. As shown in the figure, the optimal learning curve is achieved with degree = 4. Furthermore, Figure 13 demonstrates the scatter diagram for actual irradiation versus prediction based on the polynomial with hyperparameter degree values 2, 3, and 4. As shown in the figure, polynomial regression is optimal at fitting the data comparing to the other used methods. Additionally, the RMSE of polynomial regression is also significantly lower than that of linear regression due to the better fit.

Table 1: Performance Metrics For The Regression Models

Regression Model	MAE	MSE	RMSE	R2-score
Multiple	1585.39	5688979.95	2385.16	0.8624
Ridge (alpha= 0.001)	1585.4	5688959.79	2385.15	0.8624
Lasso (alpha= 1.0)	1586.48	56886631.6	2384.66	0.8625
Decision tree	970.84	5213822.94	2283.38	0.874
Polynomial (degree= 4)	**774.48**	**2711517.5**	**1646.67**	**0.937**
Polynomial (degree= 3)	799.67	3136051.54	1770.9	0.924
Polynomial (degree= 2)	781.36	2725447.13	1650.89	0.934

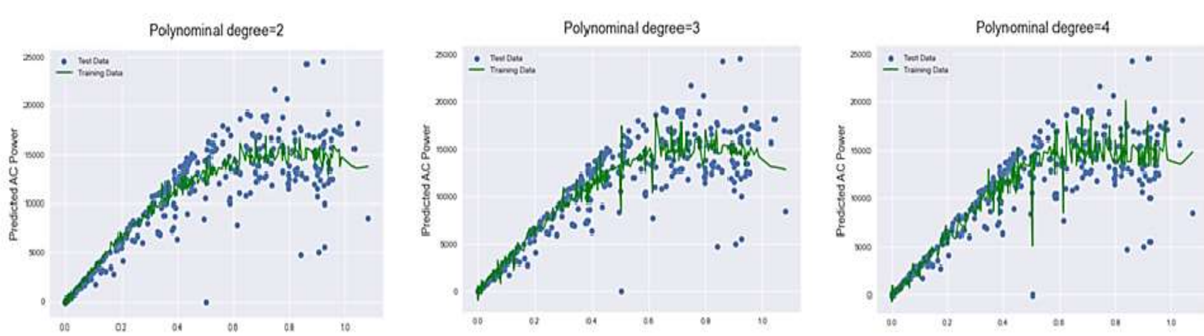

Figure10: Learning curve for Multiple, Ridge, Lasso and Decision tree base-line regressor

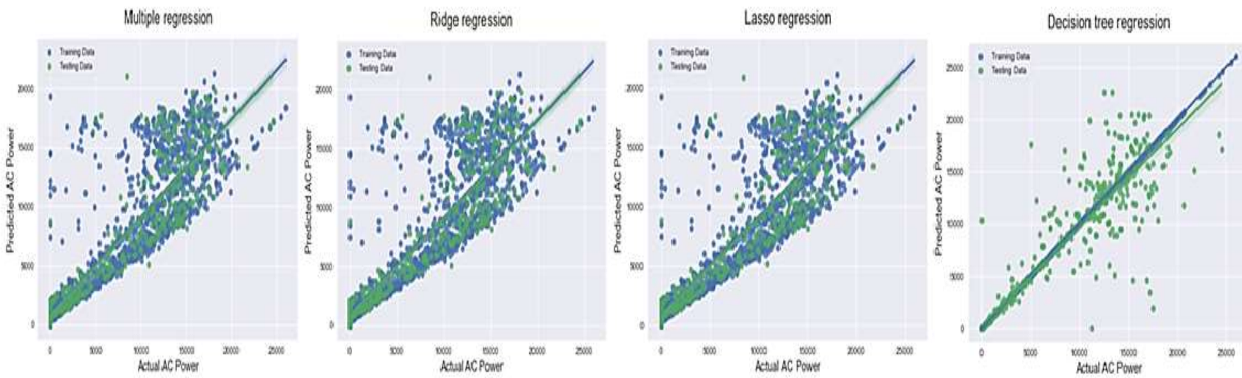

Figure 11: Actual versus prediction for Multiple, Ridge, Lasso and Decision tree base-line regressor

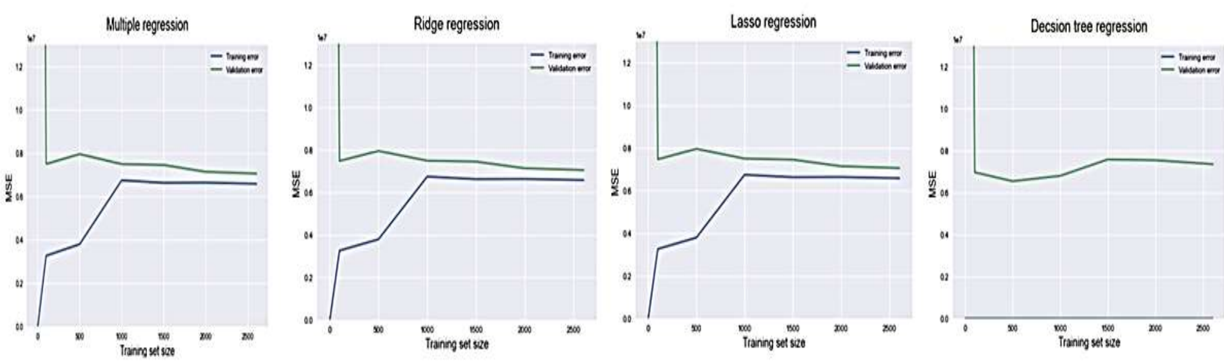

Figure 12: Learning curve for or polynomial regression with degrees =2, 3 and 4

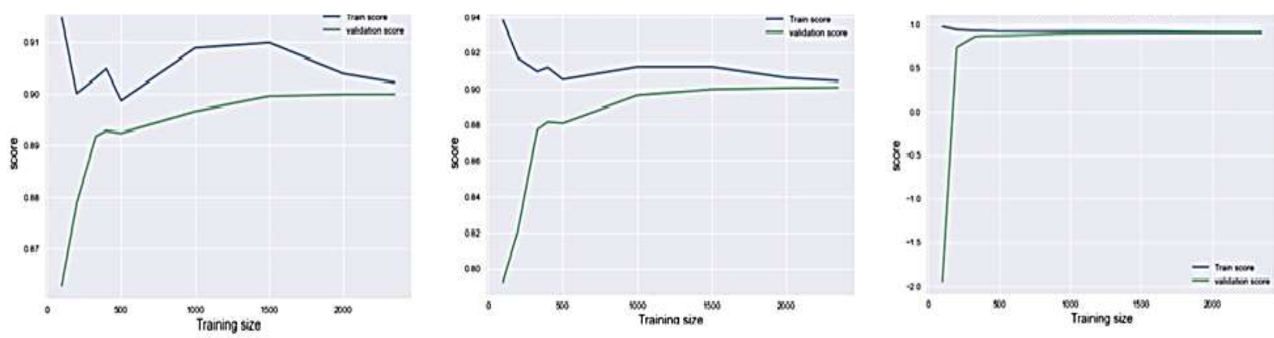

Figure 13: Actual versus prediction for the polynomial base-line regressor with respect to irradiation

Table 1 shows the performance matrices for multiple, ridge and lasso, decision tree and polynomial regression models. These matrices were used to measure the performance of the regression models. As demonstrated in the Table 1 the lowest error between actual and prediction, which is defined by MSE was achieved by a polynomial with a degree of 4. Although, the value of the R2_score for polynomial with different degrees is convergent, the optimal value is achieved with a degree of 4, which is 0.937.

6. CONCLUSION AND FUTURE WORK

Solar photovoltaic (PV) production in large-scale energy sectors is rising quickly as a result of the increased use of renewable energy sources. Predicting the output power of solar systems can be used to evaluate the operation of the solar system in order to fulfil market demands and prevent solar instability. In the current work, techniques based on linear regression, including multiple, ridge, lasso, decision tree, and polynomial regression were used to predict the generated power of the solar system. Based on cross validation approaches for ridge and lasso techniques to address underfitting, two optimal values of alpha were obtained. Three different degrees are used to accomplish the outcome of polynomial regression. Performance indicators such as MAE, MSE, RMSE, and R2 score for the five techniques were used to evaluate and compare the suggested model. As compared to other techniques based on linear regression, the findings of the polynomial of order 4 are notable. Polynomial regression with an ideal R2_score of 93.7% has been shown to produce the lowest value of MSE. The suggested polynomial curve fitting approach offers accurate and trustworthy predictions and estimations. In future work, applying this data on another model and can apply this model on another data.

REFERENCES

[1]. S.Ibrahimb, I.Dauta, Y.M.Irwan, M.Irwantoa, N.Gomesha, and Z.Farhanaa,"Linear Regression Model in Estimating Solar Radiation in Perlis", *Energy Procedia*, Elsevier, vol. 18, 2012, pp. 1402-1412.

[2]. Suruchi D., Vishal P., Niraj R., Gajanan H., Jyotiba B., "Solar Energy Prediction using Least Square Linear Regression Method", *International Journal of Current Engineering and Technology*, vol.6(5) , 2016, pp. 1-5.

[3]. Young S., Han Y. , Jae W, So Yun J., and Joo H, "Use of a Big Data Analysis in Regression of Solar Power Generation on Meteorological Variables for a Korean Solar Power Plant", *Applied Science*, vol. 11(4) , 2021, pp. 1-10.

[4]. Shereen A. El-aal, Ahmed A. Gad-ELrab, Afaf A. Zaghrout, Neveen I. Ghali, "Context aware reasoning model using deep Learning and fog computing for waste Management in IoTs environments", *Journal of Theoretical and Applied Information Technology*, vol. 98(18) , 2020, pp. 3767-3777.

[5]. Thombare, S., Pande, V.N., Kulkarni, R.S., Kakade, S.M., "Prediction of Solar Power Using Linear Regression", In: Panda, G., Naayagi, R.T., Mishra, S., (eds) Sustainable Energy and Technological Advancements, *Advances in Sustainability Science and Technology*, Springer, 2022, pp. 709–718.

[6]. Rishal A., Koko F., Sefrani S.,"Predicting Solar Irradiance Using Regression Model (Case Study: ITERA Solar Power Plant)", *IOP Conference Series: Earth and Environmental Science*, vol. 830(1) , 2021, pp.1-7.

[7]. Viswanath, A., Krishna, K., Chandrika, T., Purushotham, V., Harikumar, P., "Development of a Mathematical Model for Solar Power Estimation Using Regression Analysis", In: Singh, V., Asari, V., Kumar, S., Patel, R.,(eds) Computational Methods and Data Engineering, *Advances in Intelligent Systems and Computing,* Springer, vol. 1227, 2021, pp. 589–597.

[8]. Mohamed A. and Badrul C.,"Solar Power Probabilistic Forecasting by Using Multiple Linear Regression Analysis", *SoutheastCon 2015*, IEEE, 2015, pp. 1-6.

[9]. Ojaswa Y., Ramani K, Sheikh T., Ammar M.,"Machine Learning Based Prediction of Output PV Power in India and Malaysia with the Use of Statistical Regression", *Mathematical Problems in Engineering*,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Hindawi, 2022, pp. 1-10.

[10]. Chia-Sheng T., Wen-Chang T., Chih-Ming H., Whei-Min L., ”Short-Term Solar Power Forecasting via General Regression Neural Network with Grey Wolf Optimization”, *Energies*, vol. 15, 2022, pp. 1-10.

[11]. Tamas F.,”Using Regression Trees in Predictive Modelling”, *Production Systems and Information Engineering*, vol. 4, 2002, pp. 115-124.

[12]. Eva O.,”Modelling using Polynomial Regression”, *Procedia Engineering,* Elsevier, vol. 48, 2012, pp. 500-506.

[13]. I. Ajao, I. Raji, A. Abdullahi, ”Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, *Mathematical Theory and Modeling*, vol. 2(1), 2012, pp. 1-14.

[14]. Parveen A. and Ahtisham Y.,”Understanding and interpreting regression analysis”, *Evidence-Based Nursing 2021*, vol. 24, 2021, pp. 116-118.

[15]. John F., ”Applied regression analysis and generalized linear models”, 2nd ed, Sage, *Los Angeles*, 2008.

[16]. M. Rodr´ıguez del and N.Ben´ıtez-Parejo,”Simple linear and multivariate regression models”, *Allergologia et Immunopathologia*, Elsevier, vol. 39(3), 2011, pp. 159-173.

[17]. Khoirin N. and Netti H.,”Ridge Regression For Handling Different Levels Of Multicollinearity”, *Sci.Int.(Lahore),* vol. 30(4) , 2018, pp. 597-600.

[18]. Dastan H. and Adnan M. , ”A Review on Linear Regression Comprehensive in Machine Learning”, *Journal of Applied Science and Technology Trends*, vol. 1(4) , 2020, pp. 140-147.

[19]. Burcu K., Ipek D., ”Regression Analyses or Decision Trees?”, *Celal Bayar Universitesi ¨ Sosyal Bilimler Dergisi*, vol. 18, 2020, pp. 251-260.

[20]. Weijie W. and Yanmin L., ”Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model”, *IOP Conf. Series: Materials Science and Engineering*, vol. 324, 2018, pp. 1-11.

[21]. Davide C., Matthijs J. Warrens, Giuseppe J., ”The coefficient of determination Rsquared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”, *Peer J Computer Science*, vol. 7, 2021, pp. 1-24.

[22]. Raman Kumar, https://www.kaggle.com/code/pertetotale/solarpowergeneration, 2020.