Solar Power Prediction using Regression Models

Article *in* Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi · December 2022

DOI: 10.29137/umagd.1100957

Mustafa Yasin Erten
Kırıkkale University

Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2022) 14(3), s333-s342.

10.29137/umagd.1100957

Cilt/Volume:14 Sayı/Issue:3 Aralık/December 2022 Özel Sayı / Special Issue

Araştırma Makalesi / Research Article

Solar Power Prediction using Regression Models

Mustafa Yasin Erten[*,1], **Hüseyin Aydilek** [1]

[1]*Kırıkkale University,Faculty of Engineering & Architecture, Department of Electrical & Electronics Engineering, 71450 Kırıkkale, TURKEY*

***Başvuru/Received**: 09/04/2022* ***Kabul / Accepted**: 26/12/2022* ***Çevrimiçi Basım / Published Online:** 31/12/2022*

***Son Versiyon/Final Version**: 31/12/2022*

Abstract

Solar power prediction is an important problem that has gained significant attention recently due to the increasing demand for renewable energy sources. In this paper, we present the results of using four different regression models for solar power prediction: linear regression, logistic regression, Lasso regression, and elastic regression. Our results show that all four models are able to predict solar power accurately. However, the Elastic regression outperforms Linear, Ridge, and Lasso regressions in terms of predicting the maximum solar power output. In addition, PCA was applied to the dataset within the scope of the study, and better results were obtained in the elastic regression model. Within the scope of the study, the contribution of the feature extraction method to the study was examined. We also discuss the advantages and disadvantages of each model in the context of solar power prediction.

Key Words

"Solar power prediction, regression models, lasso regression, machine learning"

***Corresponding Author:** mustafaerten@kku.edu.tr*

1. Introduction

Solar power is a form of renewable energy generated by converting sunlight into electricity (Bagher et al., 2015). Solar power is a clean and renewable source of electricity. Unlike fossil fuels, which release greenhouse gases and other pollutants when burned, solar power is generated without causing environmental pollution or emissions (Shahsavari & Akbari, 2018). As a result, solar power is a sustainable and eco-friendly source of energy that can help reduce our environmental impact and combat climate change. Moreover, solar power is increasingly becoming more affordable and accessible. As the technology for generating solar power has improved, the cost of solar panels and other equipment has decreased, making it more affordable for individuals and businesses to install solar systems. In addition, solar power can be generated in various locations, from rooftops to open fields, making it accessible to a wide range of users. Also, solar power can help reduce our reliance on fossil fuels and improve energy security. By generating electricity from the sun, we can reduce our need for fossil fuels and finite resources subject to price fluctuations and other market forces. Thus, it can help increase energy security and reduce our vulnerability to energy market disruptions (Nwaigwe, 2019).

In recent years, there has been a significant expansion in the use of solar power worldwide. As the cost of Photovoltaic (PV) technology has decreased and its efficiency has increased, many individuals, businesses, and governments are turning to solar power to meet their energy needs. This trend is expected to continue, with considerable experts predicting that solar power will play a significant role in the transition to a clean energy future (Gielen, 2019).

Expanding solar power and developing PV generation technology is essential to achieving a sustainable and renewable energy future. By reducing our reliance on fossil fuels and minimizing our environmental impact, we can help create a cleaner, healthier, and more sustainable world for future generations.

For several reasons, estimating the amount of electricity produced by a PV system is essential. First, accurate PV power estimates ensure that a solar installation is correctly sized to meet users' energy needs. Therefore, it helps prevent over- or under-sizing of a PV system, which can be costly and inefficient. In addition, planning and scheduling electricity generation and use are done through PV power forecasts. By knowing how much power a PV system is expected to produce, power grid operators can better manage electricity supply and demand and ensure that enough power is available to meet users' needs. In this way, power outages and other malfunctions that may occur in the power grid are prevented. PV power estimates are also used to determine the financial viability of a solar project. Accurate estimates of the electricity production of a PV system help investors, lenders, and other financial stakeholders make informed decisions about investing in a particular solar project. As a result, this causes solar projects to be economically viable and provide a good return on investment (Sharadga, 2020).

In addition, accurate and reliable estimates of PV electricity generation support the growth and development of solar energy as a clean and renewable energy source.

There are various studies on the power generation prediction of PV. Considering machine learning, especially with processors that have achieved high computing power due to developments in silicon technology, is one issue that attracts attention when estimating the generated power. Several machine-learning algorithms can be used for solar photovoltaic (PV) power prediction. These algorithms use data and mathematical models to predict the amount of electricity a PV system is expected to generate (Jebli, 2021, Munawar, 2020). Some common types of machine learning algorithms that are used for PV power prediction include:

Regression algorithms use training data to build a mathematical model that can predict the output of a PV system based on input variables such as solar radiation and temperature. Regression algorithms can make short-term and long-term predictions of PV electricity generation (He, 2019). Neural network algorithms are composed of multiple interconnected nodes, or "neurons," that can process and analyze data to make predictions. Neural networks are often used for PV power prediction because they can learn and adapt over time, allowing them to make increasingly accurate predictions as more data is collected (Pazikadin, 2020). Support vector machine algorithms use training data to build a model that can classify PV electricity generation data into different categories. In addition, support vector machines are often used for short-term PV power prediction, as they can quickly and accurately predict the output of a PV system based on the current weather conditions (Buwei, 2018). Finally, decision tree algorithms use a set of rules to predict PV electricity generation. Decision trees can be used to make both short-term and long-term predictions of PV power and are often used in conjunction with other machine-learning algorithms to improve the accuracy of PV power predictions (Wang, 2018).
Different machine-learning algorithms can be used for PV power prediction. The specific algorithm will depend on the type of data available, the time frame of the predictions, and the desired level of accuracy. Literature review of solar power prediction with machine learning algorithms are given in Table 1.

Table 1. Literature Review of Solar Power Prediction with Machine Learning Algorithms

Paper	Prediction Method	Compared Method	Database	Inputs	Forecasting Horizon	Metric
Didavi, 2021	XGBoost	DT, RF, XGBoost	PVGIS database for the city of Natitingou (Benin) for 12 years	Wind speed, sun position, temperature, direct irradiation, diffuse irradiation and reflected irradiation	3-days	MSE, R
Khandakar, 2019	ANN	LR, M5P DT, GPR	Own acquired data (Rooftop) in Qatar	Irradiance, relative humidity, ambient temperature, wind speed, PV surface temperature and accumulated dust	Day-ahead	RMSE, MSE, MAE, R
Massaoudi, 2020	Hybrid (BRR, CWT, Catboost)	-	The Australian weather data	temperature, relative humidity, horizontal irradiation, previous PV power, wind direction, and diffuse horizontal radiation.	24h	RMSE, MSE, MdAE, MAE, R
Li, 2022	XGBoost	ELM, RF, SVR	The NRELhourly weather and solar irradiance data for ten years	Dew point temp, Total Cloud Cover, Wind Speed, Sea-level pressure, solar irradiance	Day-ahead	MAE, RMSE
Carneiro, 2022	Ensemble with Ridge Regression	CFBP, SOM, RBF and MLP	Algeciras, Spain, obtained by European Commission for Energy and Transport (IET) PV Geographical Information System			RMSE, MAE, MAPE, R
Kumar, 2020	Hybrid (ANN with GWO)	PSO, LM, ANF	Own acquired data - 5 kWp grid-connected rooftop PV	solar irradiance (W/m2) incident on the PV panel, cell temperature (°C), Linke turbidity, and wind speed (m/s).		NE, NSE, NRMSE, NMBE, NMAE, MSE
Munawar, 2020	XGBoost + PCA	Random forest, ANN and XGBoost	Kaggle database, Hawaii, collected by NASA	UNIX time, date, time, radiation, temperature, pressure, humidity, wind speed, wind direction, sun rise time and sun set time	Day-ahead	RMSE, R
Cervone, 2017	ANN + AnEn	ANN and AnEn combination	Threesolar power plant in Italy	Global horizontal irradiance, percent cloud cover and air temperature, solar azimuth and elevation	72h	RMSE, MRE, CORR, BIAS
Yang, 2020	complete-history persistence ensemble, OLS, AnEn, quantile regression	Markov-chain mixture	PLC Dataset	clear sky index, image	Intra-hour, Day-ahead	PICP, CRPS,PIAW, Pinball, Skill
Mohana, 2021	LASSO, Rain Forest,Linear Regression, XGBoost, SVM, DL	Polynomial Regression,	Own Data, Saudi Arabia, Abha City, King Khalid Univ.	Ambient Temp Sensor, Relative Humidity Sensor, Wind Speed Sensor, Wind Direction Sensor, Sollar Irradiation Sensor, Precipitation Sensor, Pyronometer, PV Sensor,	5 min	MSE
Perveen, 2018	Fuzzy logic	Empirical models	Own Data, photovoltaic module of 210 W power output, Delhi India	Global solar radiation, Sunshine hours, Ambient temperature, Relative humidity, Wind speed, Dewpoint	Day-ahead	MPE, MBE

Table 1 *(cont)***.** Literature Review of Solar Power Prediction with Machine Learning Algorithms

Paper	Prediction Method	Compared Method	Database	Inputs	Forecasting Horizon	Metric
Yang, 2015	Lasso	OLS, ARIMA, ETS	Own Data, Hawaii Oahu Island,	Horizontal irradiance, direct normal irradiance, diffuse horizontal irradiance, global tilt, air temperature, relative humidity, barometric pressure, wind speed, wind direction	5 min	MAE, RMSE
Zazoum, 2021	SVM and GPR	SVM and GPR	PV modules in Port Harcourt	PV panel temperature, ambient temperature, solar flux, time of the day and relative humidity	Day-ahead	RMSE, MAE, R2
Chiteka, 2020	ANN and MLR	ANN and MLR	Harare Institute of Technology, Harare, Zimbabwe, Three 100Wp PV	PM10, relative humidity, precipitation, wind speed, wind direction, ambient temperature, air pressure, maximum and minimum temperature, dew point, and clearness index	Day-ahead	RMSE, R2
Alfadda, 2017	SVR	Ploynomial Regression and Lasso	Rooftop of Virginia Tech Research Center,	Temperature, Dew Point, Relative Humidity, Visibility, Wind Speed, Wind Direction, Cloud Cover	Hourly	RMSE
Jebli, 2021	ANN and RF	LR and SVR	Errachiddia, Morroco, semi desert climate	Solar radiation, temperature, wind direction, wind speed, humidity, and pressure	Day-ahead	MAE, RMSE, MSE, R2, NRMSE, ME
VanDeventer, 2019	GASVM	SVM	Own acquired data-Local weather station on 3kW PV in Deaken University	Air temperature and solar irradiance	Hourly	RMSE, MAPE

2. Material and Methods

In a prediction model, the properties of the feature vectors used as input directly affect the forecasting performance. While some features reduce the estimation achievement, overused features increase computational costs. Before the data is used in the prediction model, the PCA approach reduces the dimensionality of the data, thus providing a more accurate and faster result of the prediction model.

The main purpose of PCA is to describe features with fewer data without excessive loss of information. For this purpose, PCA extracts reduced features using the tendencies of the original data, the covariance matrix, the eigenvalue of the covariance matrix, and the eigenvectors. The eigenvalues describe the total amount of variance, and the eigenvectors represent the direction of the new feature space. As a result of this operation, its components are converted into a small set of new uncorrelated variables as the length of a large set of arguments. The newly formed uncorrelated new variables are called principal components.

Linear regression is a statistical method used to model the linear relationship between a dependent variable and one or more independent variables. It is called simple regression if the dependent variable is modeled with only one independent variable. The relationship between the dependent variable Y and the independent variable X is calculated by Equation 1.

$$Y = \beta_0 + \beta_1 X + \in \tag{1}$$

Here, β_0 is the constant (the point where it intersects the y-axis), β_1 is the regression coefficient (the slope of the regression line), and $\in$ is the error term.

If the dependent variable is modeled with more than one independent variable, it is called multivariate linear regression. The linear relationship between the dependent variable Y and more than one independent variable is calculated by Equation 2.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \in \tag{2}$$

Here, X_n denotes independent variables β_0 is the constant, β_n are the regression coefficients, and ∈ is the error term.

In the context of solar power prediction, the dependent variable is the solar power output, and the independent variables are factors such as weather conditions, time of day, and solar panel characteristics. In linear regression, the relationship between the predictor variables and the response variable is modeled as a linear equation. The goal is to find the values of the coefficients in the equation that best fit the data.

Linear regression has some advantages that is listed as: linear regression is relatively simple to understand and implement, it can be used to model the relationship between a continuous outcome variable and one or more predictor variables. Moreover, it is widely used and well-understood, so there is a wealth of resources available for learning about it and using it effectively. Although linear regression has some advantages, it also has disadvantages. Linear regression assumes that the relationship between the predictor variables and the response variable is linear, which may not always be the case. It can be sensitive to outliers, which can affect the estimated coefficients and the predictions made by the model.

Logistic regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables when the dependent variable is binary (i.e., it can take on only two values, such as "yes" or "no", "1" or "0", etc.). It is used to predict the probability that an event will occur, based on the values of the predictor variables. The predicted probability is transformed into a binary outcome using a threshold value. Logistic regression is commonly used in classification problems, where the goal is to predict the class label of an observation based on the values of the predictor variables. In the context of solar power prediction, logistic regression can be used to predict the probability that the solar power output will exceed a certain threshold value. The predicted probability is transformed into a binary outcome using a threshold value. The logistic function is given in Equation 3.

$$p(x) = \frac{1}{1 + e^{-\frac{(x-\mu)}{s}}} \tag{3}$$

where μ is a location parameter and s is a scale parameter. This expression is rewritten as in Equation 4:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{4}$$

Logistic regression has some advantages that is listed as: Logistic can be used to predict the probability of an event occurring, which can be useful in many real-world applications. It is relatively simple to understand and implement. Furthermore, it is widely used and well-understood, so there is a wealth of resources available for learning about it and using it effectively. While linear regression has some advantages, Logistic regression also has disadvantages. It assumes that the relationship between the predictor variables and the response variable is linear, which may not always be the case. It can be sensitive to outliers, which can affect the predicted probabilities and the class labels predicted by the model.

Lasso regression is a type of linear regression that uses a regularization term in the optimization process. The regularization term is a penalty applied to the coefficients of the predictor variables in the model, which helps to prevent overfitting by reducing the complexity of the model. Lasso regression is particularly useful for selecting important features in a dataset since it tends to drive the coefficients of unimportant features to zero. In the context of solar power prediction, Lasso regression can be used to select the most important features in the dataset, which can improve the accuracy of the predictions. Lasso regression has some powerful side. Lasso regression can be used to select important features in a dataset, since it drives the coefficients of unimportant features to zero. Also, it can help to prevent overfitting by reducing the complexity of the model. Lasso regression formula is given in Equation 5.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j x_{ij} \right)^2 \right\}, \qquad s.t. \sum_{j=1}^{P} |\beta_j| \le t \tag{5}$$

where $t \ge 0$ is a tuning parameter which controls the amount of shrinkage. Equation 6 is equivalent to the ℓ_1-penalized regression problem of finding:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{P} |\beta_j| \leq t \right\} \tag{6}$$

Elastic regression is a type of linear regression that combines the strengths of both Lasso and Ridge regression. Like Lasso regression, it uses a regularization term in the optimization process to prevent overfitting. However, unlike Lasso regression, which uses the L1 norm as the regularization term, elastic regression uses a combination of the L1 and L2 norms. This allows elastic regression to balance the trade-off between model complexity and goodness of fit, which can be beneficial in some situations. Elastic regression formula is given in Equation 7.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{P} |\beta_j|^2 \quad \lambda \sum_{j=1}^{P} |\beta_j| \right\} \tag{7}$$

3. Experiments

3.1. Dataset

The solar energy power generation dataset from Kaggle was used to compare the performance of the regression models in power generation from solar panels. The data set consists of 4213 data in 21 different dimensions. The dimensions included in the dataset are given in Table 2 with mean and standard deviation values.

Table 2. Dimensions of Dataset

Name of the Value	Mean	Standard Deviation
Temperature 2m above ground	15.1	8.85
Relative humidity 2m above ground	51.4	23.5
Mean sea level pressure	1.02k	7.02
Total precipitation	0.03	0.17
Snowfall amount	0	0.04
Total cloud cover	34.1	42.8
High cloud cover	14.5	30.7
Medium cloud cover	20	36.4
Low cloud cover	21.4	38
Shortwave radiation backwards	388	278
Wind speed 10m above ground	16.2	9.88
Wind direction 10m above ground	195	107
Wind speed 80m above ground	19	12
Wind direction 80m above ground	191	109
Wind speed 900 m	16.4	9.88
Wind direction 900 m	192	107
Wind gust 10m above ground	20.6	12.6
Angle of incidence	50.8	26.6
Zenith	60	19.9
Azimuth	169	64.6
Generated power kW	1.13k	938

3.2. Performance evaluation criteria

To quantitatively evaluate the energy estimation results produced in solar panels, some metrics were used within the scope of the study. Metrics calculate performance by comparing forecast data with actual values to determine accuracy. It is also necessary to use these metrics to decide which model should be the most accurate.

Root mean square error (RMSE) and R-squared metrics were used to constrain the estimation of these variable models. The RMSE value is calculated by Equation 8 and the R squared value by Equation 9.

$$RMSE = \sqrt{\frac{\sum_{\tau=1}^{m}(y_\tau - \widehat{\lambda y}_\tau)^2}{n}} \tag{8}$$

$$R^2 = 1 - \frac{\sum(y_\tau - \hat{y}_\tau)^2}{\sum(y_\tau - \bar{y}_\tau)^2} \tag{9}$$

Here, y_τ denotes actual values, $\hat{y}_\tau$ denotes predicted values, $\bar{y}_\tau$ denotes mean values, and m is the number of data points.

3.3. Experimental results

Within the scope of this study, four different models were developed with Linear, Ridge, Lasso, and Elastic Regression algorithms and applied to the dataset. In the framework of the developed model Ridge Regression Model and Lasso Regression Model coefficients are given in Table 3 and Table 4, respectively, in matrix form.

Table 3. Coeffient matrix of Ridge Regression Model

$$\begin{bmatrix} -9{,}25E{+}00 & -4{,}27E{+}00 & 1{,}62E{+}01 & -3{,}58E{+}00 \\ 5{,}55E{+}02 & -1{,}47E{+}00 & -1{,}24E{+}00 & -1{,}15E{+}00 \\ -1{,}37E{+}00 & 1{,}24E{+}00 & 1{,}67E{+}01 & 3{,}02E{-}01 \\ 9{,}13E{+}00 & 1{,}59E{-}01 & -3{,}05E{+}01 & -2{,}63E{-}01 \\ -3{,}07E{+}00 & -1{,}51E{+}01 & -7{,}73E{+}00 & -6{,}18E{+}00 \end{bmatrix}$$

Table 4. Coeffient matrix of Lasso Regression Model

$$\begin{bmatrix} -9{,}32E{+}00 & -4{,}28E{+}00 & 1{,}61E{+}01 & 0{,}00E{+}00 \\ 3{,}54E{+}02 & -1{,}45E{+}00 & -1{,}24E{+}00 & -1{,}17E{+}00 \\ -1{,}37E{+}00 & 1{,}24E{+}00 & 1{,}67E{+}01 & 3{,}03E{-}01 \\ 9{,}04E{+}00 & 1{,}60E{-}01 & -3{,}04E{+}01 & -2{,}62E{-}01 \\ -3{,}08E{+}00 & -1{,}51E{+}01 & -7{,}71E{+}00 & -6{,}19E{+}00 \end{bmatrix}$$

The Lasso regression model gives better results than the Ridge regression model because, in the Lasso regression model, one of the coefficients is penalized and given a value of 0. In addition, the most suitable alpha values were selected in the Ridge and Lasso regression models, and the most optimum results were obtained. In this study, the alpha value for Ridge regression was 616,423 and the alpha value for Lasso regression was found to be 143,434.

Firstly, four different regression models were applied to the data set to evaluate the performance. The performance outputs of the regression models performed with the raw dataset are given in Table 5.

Table 5. Performance Evaluation Results of Four Different Regression Models to Raw Data

	RMSE	R- Squared
Linear Regression	507,532322	0,718008
Ridge Regression	507,379842	0,718178
Lasso Regression	506,864894	0,718750
Elastic Regression	506,796370	0,718826

When the results in Table 5 are examined, it is seen that the four regression models give similar results, and the Elastic model produces better results than the others.

Principal component analysis was applied to our 20-component dataset to examine the effect of reducing some dimensions on data analysis; it was found that 99,5% of the variance was explained with 16 components. The variance result matrix is given in Table 6, and the number of components / expected variance graph is given in Figure 1.

Table 6. Explained Variance Ratio Matrix

22,1379	41,2584	55,7907	66,9964	73,6667
79,8258	84,3156	88,4635	92,5043	94,3980
96,1544	97,0091	97,7361	98,3392	98,8200
99,2069	99,5242	99,8251	99,9758	100,0000

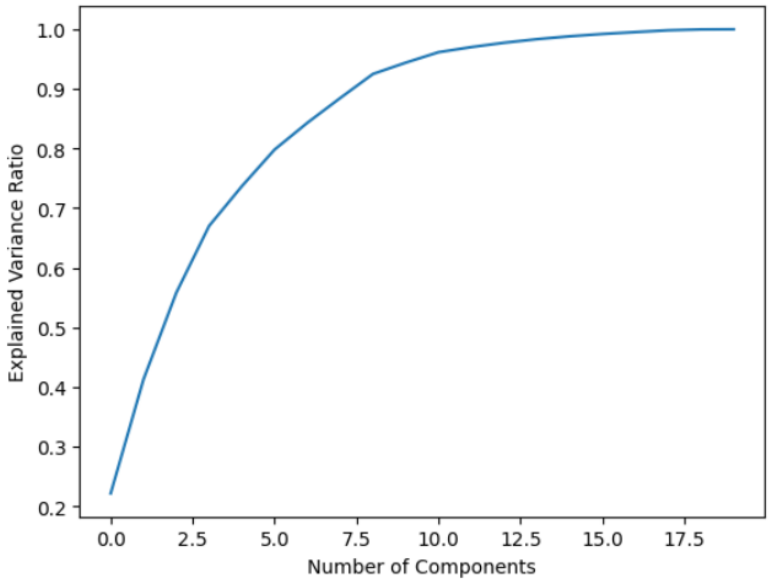

Figure 1. Explained Variance Ratio per Number of Components

On the reduced data set, the model developed with four different regression algorithms was applied again, and the results were obtained and given in Table 7.

Table 7. Performance Evaluation Results of Four Different Regression Models with PCA

	RMSE	R- Squared
Linear Regression	507,532321	0,718008
Ridge Regression	507,523041	0,718019
Lasso Regression	507,368822	0,71819
Elastic Regression	506,718803	0,718912

When the results in Table 7 are examined, it is seen that the four regression models give similar results, while the Elastic model produces better results than the others. While it is seen that the elastic regression model with PCA produces better results than the raw data, the Lasso regression model produces worst results than the raw data. The RMSE values per number of components in the Elastic regression model are given in Figure 2.

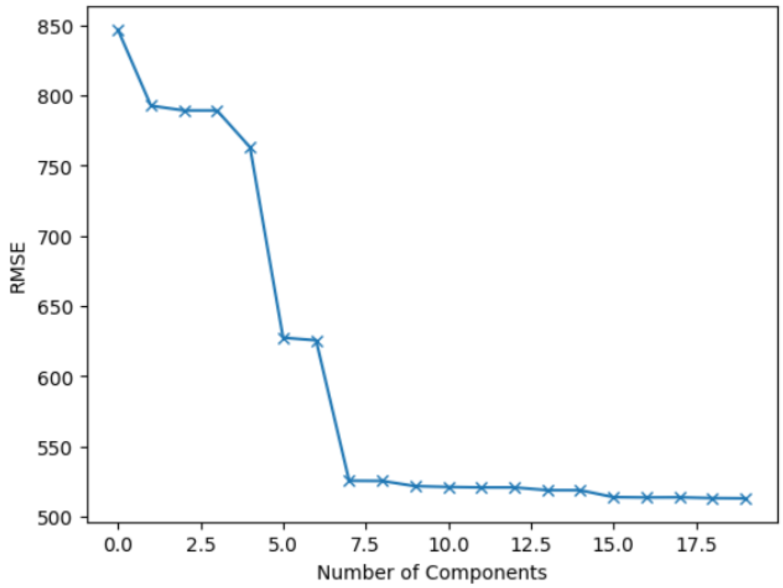

Figure 2. RMSE values per Number of Components

4. Conclusion

Fossil energy sources are rapidly depleting, renewable energy sources are an essential energy production alternative as an environmentally friendly and economical solution despite the increasing energy demand. However, electricity generation with renewable energy sources has its limitations. For example, production efficiency depends on environmental and meteorological data in generating electrical energy from solar energy with photovoltaic panels. For this reason, the electrical energy produced during the day constantly changes. Enabling healthy energy planning with the most realistically estimated production data; will ensure both uninterrupted meetings of energy demand and grid efficiency. With the PCA applied, 4 dimensions were reduced in the dataset consisting of 20 dimensions and a new data set with 0.5% sensitivity was created. After extracting the features with PCA, the RMSE values with the elastic model is increased. Due to the data set's low R^2 values, the estimations' accuracy was limited. However, it is considered to show higher accuracy prediction performance when applied to better datasets. In addition, the differences between the regression models will be better examined.

This study aims to investigate the performance effect of four different regression models on solar power prediction. In addition, the performance of these four different regression models was measured after reducing the dimensions that may affect the result in large data sets negligibly from the data set and reducing the size. When the results were evaluated with the general evaluation metrics RMSE and R square, it was seen that the Elastic model produced better results than the other models. In addition, principal component analysis and dimension reduction affect the results positively.

Acknowledgment

This study has been evaluated within the scope of the special issue of engineering sciences research and development applications.

References

Alfadda, A., Adhikari, R., Kuzlu, M., & Rahman, S. (2017, April). Hour-ahead solar PV power forecasting using SVR based approach. In 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) (pp. 1-5). IEEE.

Bagher, A. M., Vahid, M. M. A., & Mohsen, M. (2015). Types of solar cells and application. *American Journal of optics and Photonics*, *3*(5), 94-113.

Buwei, W., Jianfeng, C., Bo, W., & Shuanglei, F. (2018, November). A solar power prediction using support vector machines based on multi-source data fusion. In 2018 International Conference on Power System Technology (POWERCON) (pp. 4573-4577). IEEE.

Carneiro, T. C., Rocha, P. A., Carvalho, P. C., & Fernández-Ramírez, L. M. (2022). Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. Applied Energy, 314, 118936.

Cervone, G., Clemente-Harding, L., Alessandrini, S., & Delle Monache, L. (2017). Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy, 108, 274-286.

Chiteka, K., Arora, R., & Sridhara, S. N. (2020). A method to predict solar photovoltaic soiling using artificial neural networks and multiple linear regression models. Energy Systems, 11(4), 981-1002.

Didavi, A. B., Agbokpanzo, R. G., & Agbomahena, M. (2021, December). Comparative study of Decision Tree, Random Forest and XGBoost performance in forecasting the power output of a photovoltaic system. In 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART) (pp. 1-5). IEEE.

Gielen, D., Boshell, F., Saygin, D., Bazilian, M. D., Wagner, N., & Gorini, R. (2019). The role of renewable energy in the global energy transformation. Energy Strategy Reviews, 24, 38-50.

He, Y., Yan, Y., & Xu, Q. (2019). Wind and solar power probability density prediction via fuzzy information granulation and support vector quantile regression. International Journal of Electrical Power & Energy Systems, 113, 515-527.

Jebli, I., Belouadha, F. Z., Kabbaj, M. I., & Tilioua, A. (2021). Prediction of solar energy guided by pearson correlation using machine learning. Energy, 224, 120109.

Khandakar, A., EH Chowdhury, M., Khoda Kazi, M., Benhmed, K., Touati, F., Al-Hitmi, M., & Jr SP Gonzales, A. (2019). Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar. Energies, 12(14), 2782.

Kumar, A., Rizwan, M., & Nangia, U. (2020). A hybrid intelligent approach for solar photovoltaic power forecasting: impact of aerosol data. Arabian Journal for Science and Engineering, 45(3), 1715-1732.

Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., ... & Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. Energy Reports, 8, 1087-1095.

Massaoudi, M., Refaat, S. S., Abu-Rub, H., Chihi, I., & Wesleti, F. S. (2020, July). A hybrid Bayesian ridge regression-CWT-catboost model for PV power forecasting. In 2020 IEEE Kansas Power and Energy Conference (KPEC) (pp. 1-5). IEEE.

Mohana, M., Saidi, A. S., Alelyani, S., Alshayeb, M. J., Basha, S., & Anqi, A. E. (2021). Small-scale solar photovoltaic power prediction for residential load in Saudi Arabia using machine learning. Energies, 14(20), 6759.

Munawar, U., & Wang, Z. (2020). A framework of using machine learning approaches for short-term solar power forecasting. Journal of Electrical Engineering & Technology, 15(2), 561-569.

Nwaigwe, K. N., Mutabilwa, P., & Dintwa, E. (2019). An overview of solar power (PV systems) integration into electricity grids. Materials Science for Energy Technologies, 2(3), 629-633.

Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., & Faraj, M. A. (2020). Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend. Science of The Total Environment, 715, 136848.

Perveen, G., Rizwan, M., & Goel, N. (2018). Intelligent model for solar energy forecasting and its implementation for solar photovoltaic applications. Journal of Renewable and Sustainable Energy, 10(6), 063702.

Shahsavari, A., & Akbari, M. (2018). Potential of solar energy in developing countries for reducing energy-related emissions. *Renewable and Sustainable Energy Reviews*, *90*, 275-291.

Sharadga, H., Hajimirza, S., & Balog, R. S. (2020). Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy, 150, 797-807.

VanDeventer, W., Jamei, E., Thirunavukkarasu, G. S., Seyedmahmoudian, M., Soon, T. K., Horan, B., & Stojcevski, A. (2019). Short-term PV power forecasting using hybrid GASVM technique. Renewable energy, 140, 367-379.

Wang, J., Li, P., Ran, R., Che, Y., & Zhou, Y. (2018). A short-term photovoltaic power prediction model based on the gradient boost decision tree. Applied Sciences, 8(5), 689.

Yang, D., van der Meer, D., & Munkhammar, J. (2020). Probabilistic solar forecasting benchmarks on a standardized dataset at Folsom, California. Solar Energy, 206, 628-639.

Yang, D., Ye, Z., Lim, L. H. I., & Dong, Z. (2015). Very short term irradiance forecasting using the lasso. Solar Energy, 114, 314-326.
Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. Energy Reports, 8, 19-25.