Exploring Machine Learning Models for Solar Energy Output Forecasting

Idamakanti Kasireddy(✉), V Mamatha Reddy, P. Naveen, and G Harsha Vardhan

EEE Department, Vishnu Institute of Technology, Bhimavaram, India
kaasireddy.i@vishnu.edu.in

Abstract. Engineering, science, health, and other fields have all used machine learning algorithms. The idea of machine learning is used in this study to forecast solar energy output. Predicting solar energy, a well-known renewable source with a number of advantages, can help with energy consumption planning. Inconsistent weather makes it difficult for grid operators to manage solar energy output, which makes it harder to satisfy customer demand. Utilizing various algorithms including Lasso, Ridge, Linear and Support Vector Regression (SVR) algorithms, our suggested strategy entails developing prediction models. These algorithms produce forecasts based on past weather information such as temperature, dew point, wind, cloud cover, and visibility. SVR algorithm outperformed the other algorithms, according to the Jupyter Notebook examination.

Keywords: Solar production · Algorithms · Jupyter Note book

1 Introduction

Solar energy offers numerous merits; however, it is plagued by the issue of irregular production, which creates difficulties for grid operators trying to meet consumer demand. By anticipating the output of solar energy, grid operators can address this problem and ensure that they are adequately prepared to reach the needs of loads. For prediction of solar energy, one must acquire data on various factors that impact production, including wind speed, humidity, temperature, and dew point. This information can be obtained from meteorological departments or online sources. The effective way to examine this data is through the application of machine learning, which can be segregated into three categories: supervised, unsupervised, and reinforcement learning [1–14].

Unsupervised learning just includes input data, letting the model to independently find patterns. Supervised learning necessitates providing the model both the input data and the desired output. Contrarily, in reinforcement learning algorithm, a software deals with an environment to accomplish a certain goal and receives feedback in the form of incentives. Regression and classification are two subcategories of supervised learning, which is used in this work. Regression techniques including linear, ridge, lasso, and support vector regression are the most often used ones. Regression involves using independent variables to predict the dependent variable values.

P. Pareek et al. (Eds.): IC4S 2023, LNICST 536, pp. 210–217, 2024.
https://doi.org/10.1007/978-3-031-48888-7_18

The implementation of algorithms involves a set of steps, including data collection, analysis, wrangling, training and testing, and accuracy evaluation. The first step involves collecting data, which is then analyzed for duplicates, missing values, and incorrect formatting. Data cleaning, which involves removing duplicates, imputing missing values, and correcting formatting errors, is performed as necessary. The model is then trained using 80% of the data, with the remaining 20% used for testing. The final step is to examine the accuracy of the model using error method. The efficiency of the model is proportional to its accuracy.

Python's numerous libraries make it easier to build machine learning methods. These libraries facilitate and improve the implementation of algorithms. Due to its simplicity of use and abundance of libraries, Python is frequently used for machine learning. Machine learning algorithms are run on platforms like Anaconda, Google Colab, and Jupyter. Jupyter stands out among them for its ease of use and ability to share files via Jupyter Notebook. Users can conduct mathematical operations like trigonometry and Fourier transforms using Jupyter, which is open source.

2 Methodology

2.1 Linear Regression

Several regression approaches, including linear regression, were used in this study [15]. A fundamental regression technique that aids in determining the relationship between two variables (independent and dependent) is linear regression. It is based on minimizing the error between the line and the data points by applying several strategies to determine the line that best fits the data points. The least-squares approach is used to find the regression line, also referred to as the line of best fit. Accuracy is measured using the R-squared value, which measures how closely the data adhere to the regression line. A good model is one that has an R-squared value better than 0.5. The evaluation of trends and projections of sales and trend forecasting are two common uses of linear regression. The slope for the predicted regression equation is represented by formulas (1).

$$b_1 = \frac{\sum(x_1 - \overline{x})(y_1 - y)}{(x_1 - \overline{x})^2} \quad (1)$$

2.2 Lasso Regression

This algorithm is applied as a regularization method to prevent over fitting in machine learning models. Over fitting occurs when a model performs exceptionally well on the training dataset, but poorly on the testing dataset, due to a high cost function. To mitigate this issue, Lasso Regression incorporates a regularization term represented by lambda, which multiplies the weights of the model.

2.3 Ridge Regression

Regularization further makes use of Ridge Regression [16]. It is a method to stop a model from fitting too tightly. Sometimes a machine learning model performs well when tested on training data, but when tested on testing data, it performs poorly compared to training data, leading to over fitting. When using this, the objective function is modified by the addition of a penalty factor. Ridge Regression Penalty is the measurement of the penalty that was applied to the model.

2.4 Support Vector Regression

The Support Vector Regression (SVR) [17–20] algorithm is a machine learning method that is used to predict continuous values. The SVR algorithm is based on the same principle as Support Vector Machines (SVMs), where the aim is to find the best fit hyperplane. The algorithm tries to minimize the distance between the predicted values and the actual values, while also maintaining a margin of error. This margin of error is controlled by the threshold value, which determines the trade-off between the accuracy of the model and the complexity of the hyperplane. Overall, SVR is a powerful algorithm for predicting continuous values and can be used in a variety of applications. When compared to other algorithms, SVR offers the benefits of being extremely simple to use and having a high degree of prediction ability.

2.5 Purpose

The study compares and analyzes the performance of four different machine learning algorithms used for solar energy prediction. The main objective of the study is to identify the most efficient method for forecasting energy output of solar system.

The algorithms used in the study are carefully selected and include the most commonly used methods for time-series forecasting. These algorithms are trained on historical solar energy data, and their performance is evaluated based on their ability to predict future solar energy output accurately.

The study utilizes various performance metrics to evaluate the accuracy of each algorithm, including the following,

- mean absolute error,
- root mean square error,
- Coefficient of determination.

The results of the study are then analyzed and compared to determine which algorithm performs best for solar energy prediction.

By conducting a comparative analysis of these four machine learning algorithms, the study aims to provide valuable insights into the most effective method for solar energy forecasting. This information can help inform decision-making processes and support the development of more accurate and reliable solar energy prediction models.

3 Results

Understanding the link between the dependent and independent variables is crucial before conducting an accurate analysis of solar energy forecast using machine learning algorithms. In this scenario, solar output is the dependent category, which implies that numerous things impacting it, which are independent variables, are the output or consequence of solar energy.

The study moves further with the implementation of machine learning models after identifying the variables. In order to forecast solar energy output, the analysis employs four alternative regression algorithms: support vector, lasso, linear, and ridge.

To facilitate the analysis, various Python libraries are imported into Jupyter Notebook, including pandas, numpy, matplotlib.pyplot, and seaborn. The weather data is imported into the platform i.e. Jupyter book using the various commands, which varies depending on the file type (.xls or.csv).

The imported data is then analyzed using several methods, such as info(), describe(), and isnull().sum(), to identify any empty, duplicate, or incorrectly formatted values. The goal of this step is to minimize differences between values and improve the accuracy of the analysis. If null values are present, they are replaced with suitable values through mean or median methods. Rows with duplicate values are dropped, while rows with incorrect formats are corrected to ensure the accuracy of the data.

Overall, this rigorous analysis of the solar energy data using different machine learning algorithms and data analysis techniques helps to provide accurate and reliable predictions of solar energy output, which can be valuable for various applications in the renewable energy industry.

Finally, evaluating the effectiveness of different algorithms are obtained as follows (Table 1).

Table 1. Evaluating the effectiveness of different algorithms

Algorithm Accuracy
In linear regression, 51.3%
In Lasso regression,51.2%
In Ridge regression, 51%
In Support vector regression, 88.4%

The results of various algorithms' predictions for solar energy are depicted in Figs. 1, 2, 3 and 4. Upon analyzing the figure, it becomes evident that the SVR algorithm performs better than the other algorithms tested. The SVR algorithm's superior performance indicates that it is the most accurate model among the algorithms compared. Therefore, the results suggest that SVR is the optimal choice for predicting energy of solar system.

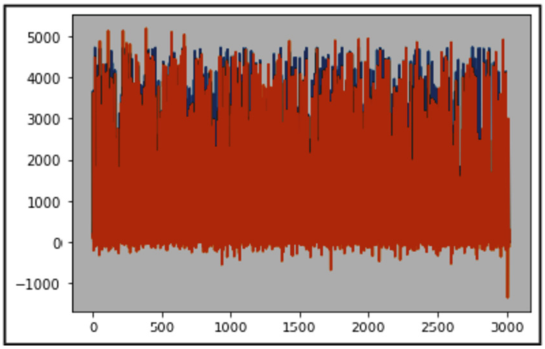

Fig. 1. The output from the SVR test is shown in blue, while the expected output is shown in red. (Colour figure online)

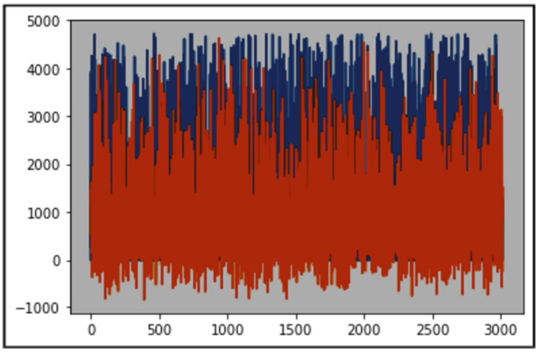

Fig. 2. The output from the linear regression test is shown in blue, while the projected output is shown in red. (Colour figure online)

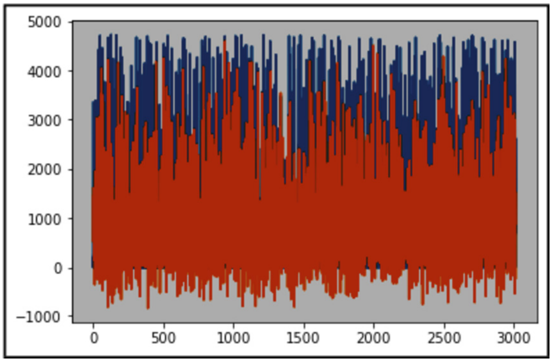

Fig. 3. The output from the Lasso Regression test is shown in blue, while the expected output is shown in red. (Colour figure online)

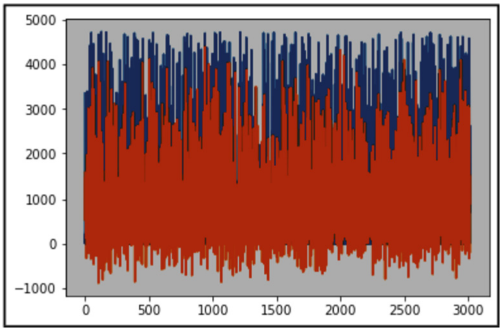

Fig. 4. The output from the Ridge Regression test is shown in blue, while the expected output is shown in red. (Colour figure online)

4 Conclusion

This study's objective was to forecast solar energy using different models. Jupyter Notebook was used to implement and analyse the models. After investigation, it was discovered that the Ridge, Lasso, and linear regressions had accuracy levels between 51% and 52%. The Support Vector Regression model, however, has a substantially higher accuracy rate of 88.4%. The results clearly indicate that SVR outperformed the other regression models in predicting solar energy. The higher accuracy of SVR suggests that it is a more appropriate and reliable model for predicting solar energy as compared to other models. Therefore, this study recommends the use of Support Vector Regression for predicting solar energy.

References

1. Javed, A., et al.: Predicting solar irradiance using machine learning techniques. In: 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1458–1462 (2019)
2. Ak, R., et al.: Two machine learning approaches for short-term wind speed time-series prediction. IEEE Trans. Neural Networks Learn. Syst **27**(8), 1734–1747 (2016)
3. Nasir, A.W., Kasireddy, I., Rahul Tiwari, B.K., Ahmed, I., Furquan, A.: Data-Based Tuning of PI Controller for First-order System. In: Bhaumik, S., Chattopadhyay, S., Chattopadhyay, T., Bhattacharya, S. (eds.) Proceedings of International Conference on Industrial Instrumentation and Control: ICI2C 2021, pp. 547–555. Springer Nature Singapore, Singapore (2022). https://doi.org/10.1007/978-981-16-7011-4_52
4. Pragaspathy, S., Aravindh, G., Kannan, R., Dhivya, K., Karthikkumar, S., Karthikeyan, V.: Advanced control strategies for the grid integration of wind energy system employed with battery units. In: 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), pp. 1–5. Chennai, India (2022). https://doi.org/10.1109/ICPECTS56089.2022.10046771
5. Phani Kumar, Ch., Elanchezhian, E.B., Pragaspathy, S.: An adaptive regulatory approach to improve the power quality in solar PV-integrated low-voltage utility grid. J. Circ. Syst. Comput. **31**(17), 2250301 (2022). https://doi.org/10.1142/S0218126622503017

6. Rama Rao, R.V.D., Pragaspathy, S.: Enhancement of electric power quality using UPQC with adaptive neural network model predictive control. In: 2022 International Conference on Electronics and Renewable Systems (ICEARS), pp. 233–238. Tuticorin, India (2022). https://doi.org/10.1109/ICEARS53579.2022.9751866
7. Kannan, R., Karthikkumar, S., Suseendhar, P., Pragaspathy, S., Chakravarthi, B.N.C.V., Swamy, B.: Hybrid renewable energy fed battery electric vehicle charging station. In: 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 151–156. Coimbatore, India (2021). https://doi.org/10.1109/ICESC51422.2021.9532995
8. Kumar, C.P., Pragaspathy, S., Karthikeyan, V., Durga Prakash, K.N.S.: Power quality improvement for a hybrid renewable farm using UPQC. In: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 1483–1488. Coimbatore, India (2021). https://doi.org/10.1109/ICAIS50930.2021.9396048
9. Saravanan, S., Karunanithi, K., Pragaspathy, S.: A novel topology for bidirectional converter with high buck boost gain. J. Circ. Syst. Comput. **29**(14), 2050222 (2020)
10. Pragaspathy, S., Rao, R.V.D.R., Karthikeyan, V., Bhukya, R., Nalli, P.K., Korlepara, K.N.S.D.P.: Analysis and appropriate choice of power converters for electric vehicle charging infrastructure. In: 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 1554–1558. Coimbatore, India (2022). https://doi.org/10.1109/ICAIS53314.2022.9742853
11. NarasimhaRaju, V.S.N., Premalatha, M., Pragaspathy, S., Rao K, D.V.S.K., Korlepara, N.S.D.P., Kumar, M.M.: Implementation of instantaneous symmetrical component theory based hysteresis controller for DSTATCOM. In: 2021 International Conference on Advancements in Electrical, pp. 1–8. Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India (2021). https://doi.org/10.1109/ICAECA52838.2021.9675526
12. Sharma, A., Kakkar, A.: Forecasting daily global solar irradiance generation using machine learning. Renew. Sustain. Energy Rev. **82**(3), 2254–2269 (2018)
13. Kasireddy, I., et al.: Application of FOPID-FOF controller based on IMC theory for automatic generation control of power system. IETE J. Res. **68**(3), 2204–2219 (2022). https://doi.org/10.1080/03772063.2019.1694452
14. Kasireddy, I., et al.: Determination of stable zones of LFC for a power system considering communication delay. AIP Conf. Proc. **2418**, 040014 (2022). https://doi.org/10.1063/5.0081986
15. Maulud, D., Abdulazeez, A.M.: A Review on Linear Regression Comprehensive in Machine Learning. JASTT **1**(4), 140–147 (2020)
16. Yang, X., et al.: Lasso regression models for cross-version defect prediction. IEEE Trans. Reliab. **67**(3), 885–896 (2018)
17. Crone, S.F., Guajardo, J., Weber, R.: A study on the ability of support vector regression and neural networks to forecast basic time series patterns. In: Bramer, M. (ed.) IFIP AI 2006. IIFIP, vol. 217, pp. 149–158. Springer, Boston, MA (2006). https://doi.org/10.1007/978-0-387-34747-9_16
18. Le, D.N., Parvathy, V.S., Gupta, D., et al.: IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. Int. J. Mach. Learn. & Cyber. **12**, 3235–3248 (2021). https://doi.org/10.1007/s13042-020-01248-7
19. Kalaiyarasi, M., Saravanan, S., Narukullapati, B.K., Kasireddy, I., Naga Malleswara Rao, D.S., Nagineni Venkata Sireesha, D.: Analysis of SAR ImagesDe-speckling using a bilateral filter and feed forward neural networks. In: 2023 Second International Conference on Electrical, pp. 1–6. Electronics, Information and Communication Technologies (ICEEICT), Trichirappalli, India (2023). https://doi.org/10.1109/ICEEICT56924.2023.10156987

20. Venkata Subbarao, M., Padavala, A.K., Harika, K.D.: Performance analysis of speech command recognition using support vector machine classifiers. In: Jason, G., Dey, R., Adhikary, N. (eds.) Communication and Control for Robotic Systems, pp. 313–325. Springer Singapore, Singapore (2022). https://doi.org/10.1007/978-981-16-1777-5_19