PREDICTING WIND ENERGY PRODUCTION IN THE SHORT TERM USING MACHINE LEARNING ALGORITHM

Hilmi KUŞÇU [1], Taşkın TEZ [2]

[1] Department of Mechanical Engineering, Trakya University, Edirne, 22030, TR Türkiye
[2] Edirne Provincial Health Department, Ministry of Health, Edirne, 22030, TR Türkiye

Abstract

The prediction of electricity generation from wind power plays a critical role in the formulation and management of future energy production plans. These predictions are highly important for wind energy facilities to achieve optimal performance, meet energy demands, and stabilize energy prices. Therefore, in this study, the Support Vector Machine (SVM) Regression Algorithm, a traditional machine learning algorithm, was preferred to forecast the weekly electricity production of wind power plants. Performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-Squared (R^2) were utilized to assess the accuracy of the predictions. The results of this study indicated that the SVM Regression Algorithm with the Radial function yielded the best outcomes. Consequently, it is recommended to employ the SVM Regression Algorithm with the Radial function for weekly electricity production predictions.

Keywords: Energy, Wind Energy, Renewable Energy, Wind Speed Prediction, Production, Machine Learning.

INTRODUCTION

Global energy demand is projected to increase by 46% by 2050, driven by factors such as population growth, economic expansion, and technological progress [1]. However, energy resources are finite. Consequently, wind-powered electricity generation has become one of the most preferred energy sources in recent years to meet global energy demand [2].

Due to factors such as night-day cycles, seasonal variability, meteorological events, and geographic location, wind energy production is a non-continuous process [3]. Therefore, data analysis and prediction methods are essential for wind energy production. A forward-looking accurate prediction of electricity generation holds significant importance for the effective planning, operation, and management of electrical grids and for balancing energy prices [4]. For optimization, wind energy production forecasts are classified into short-term, medium-term, and long-term intervals based on periods [5].

When reviewing the literature, numerous methods have been employed in studies aimed at predicting wind energy generation. These studies conducted by Singh et al. involved the prediction of short-term wind energy production in wind turbines located in western Turkey using Random Forest(RF), Gradient Boosting Machine(GBM), K-Nearest Neighbor(kNN), Decision-Tree(DT), and Extra Tree(ETR) regression algorithms. The research findings indicate that the GBM algorithm yielded the best results [6]. Yürek et al. proposed a methodology for comparing various machine learning algorithms to investigate the performance of different machine learning algorithms on real data from wind turbines and to select the most accurate final model for wind energy production forecasting [7]. Phan et al. applied a hybrid forecasting model that combines different numerical weather prediction (NWP) models and the XGBoost learning model for short-term wind energy forecasting [8]. Shirzadi et al. showed that non-linear auto-regressive exogenous (NARX) neural network and recurrent neural networks (Long Short-Term Memory-LSTM) methods outperformed SVM and RF machine learning algorithms in terms of accuracy for hourly resolution regional electricity demand forecasting [9]. Similarly, several scientific studies have been conducted using the support vector machine SVM algorithm and have yielded good performance results [10, 11-12].

While machine learning algorithms can be used to make efficient and reliable predictions from wind power plants, some factors limit these predictions. These include unexpected failures in the turbine or plant, personnel errors, changes in wind direction caused by urban sprawl, and sudden climate changes.

In this study, SVR algorithm, a machine learning algorithm, was used to predict wind power generation. The aim was to determine the best-performing kernel function by evaluating the performance of the Linear, Radial, Polynomial, and Sigmoid kernel functions of this algorithm.

EXPOSITION

Research Hypothesis

The purpose of this study is to investigate the performance results of sub-functions within the SVM regression algorithm, obtained through the utilization of traditional machine learning regression algorithms, to determine the most accurate prediction outcomes for electricity generation from wind energy.

Applied methodology and methods

The data used in this study were obtained from Kaggle[13], a data science competition platform owned by Google LLC and an online community of data scientists and machine learning practitioners. These records were extracted from the SCADA system, which collects and logs data from an operational wind turbine that generates electricity, capturing readings at 10-minute intervals. A weekly period, of 988 data points, three independent variables, and one dependent variable were used. All dependent and independent variables are of numerical data type. These variables are as follows:

- Active Power (kW): Turbine electricity production at that time.
- Theoretical PowerCurve (kWh): Turbine power rating at specified wind speed.
- Wind Speed (m/s): Wind speed at the height of the rotor hub.
- Wind Direction (°): Direction of the wind at the hub height of the turbine.

Wind power generation forecasting was calculated using the SVM regression algorithm, a machine learning algorithm employed for kernel-based classification and regression problems, known for its effective prediction capabilities using kernel methods. SVM regression algorithm find a line or curve to classify or predict data points using a plane or hyperplane. SVR can also be applied to nonlinear data, achieved through a technique known as the kernel trick. The kernel trick transforms the data into a higher-dimensional feature space so that the data can become linearly separable [14]. In the predictions, the following subfunctions of the SVM regression algorithm were utilized:

- Linear
- Radial
- Polynomial
- Sigmoid

Once machine learning regression models are developed, performance metrics are used to assess the performance of these models objectively. The performance metrics used in this paper are as follows:

- Mean squared error (MSE): MSE is the mean squared error, which is a measure of how close the predicted values are to the true values.
- Root mean squared error (RMSE): RMSE is a metric used to measure how well a model fits, and the square root of MSE shows how far the model is from the actual values.
- Mean absolute error (MAE): MAE is the average of the absolute errors between the model's predictions and the actual values.
- Mean absolute percentage error (MAPE): MAPE measures the average percentage error between predicted and actual values.
- R-squared (R^2): R^2 is a statistical measure of how well a regression model fits the data. It is calculated as the percentage of the variance in the dependent variable that is explained by the independent variables. A higher R^2 value indicates that the model is a better fit for the data [15].

Results and Discussion

The analysis of the independent variables Theoretical PowerCurve (KWh), Wind Speed (m/s), and Wind Direction (°) with the dependent variable Active Power (kW) in this study was performed using the JASP 0.17.1.0 statistical computing program, and descriptive statistics are presented in Table 1.

Table. 1. *Descriptive statistics.*

	Active Power (kW)	Theoretical PowerCurve (KWh)	Wind Speed (m/s)	Wind Direction (°)
Valid	988	988	988	988
Mode	0.000	0.000	0.291	0.000
Median	1043.951	1045.875	7.066	200.663
Mean	1523.195	1589.501	7.510	185.359
Std. Deviation	1401.628	1444.221	4.023	65.286
Minimum	0	0.000	0.291	0.000
Maximum	3604.561	3600.000	16.294	358.190

In the SVM regression algorithm, 791 out of 988 data points were used for training, while 197 were used for testing, with 562 support vectors shown in Table 2 and Fig. 1.

Table. 2. *SVM regression algorithm data.*

Support Vectors	n(Train)	n(Test)	Test MSE
562	791	197	0.003

Fig. 1. *Data split*

The performance metrics obtained using the subfunctions of the SVM regression algorithm are shown in Table 3. As can be seen from the performance metrics in Table 3, the best prediction of wind power generation is provided by the Radial subfunction of the SVM regression algorithm. As shown in Table 3, the MSE, RMSE, MAE, and MAPE values are the closest to zero. The R^2 = 0.997 value also belongs to the Radial function, which is closest to 1.

Table. 3. *Evaluation Metrics.*

Sub-Functions of SVM Algorithm	Evaluation Metrics				
	MSE	RMSE	MAE	MAPE%	R^2
Linear	0.008	0.089	0.055	279.63	0.993
Radial	0.003	0.055	0.035	13.75	0.997
Polynomial	0.046	0.214	0.123	37.25	0.96
Sigmoid	2094.823	45.769	35.701	18744.02	0.05

The regression line obtained by using the test data in the SVM regression algorithm and the predicted values of the SVM regression algorithm is shown in Fig. 2. The fact that all data points in this regression line lie on the same line indicates how well the regression line fits the data.

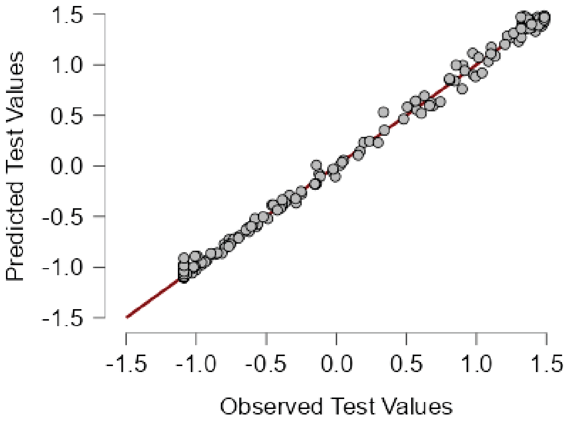

Fig. 2. Regression line

The subject of predictive analysis of renewable energy sources has been extensively explored by scientists in recent times, particularly in the context of energy transformation. This approach necessitates validation with real data obtained from a functioning wind turbine over an extended period. This is due to the fact that wind turbine energy production exhibits seasonality and cyclic patterns in various parts of the world. Furthermore, the density of air, which varies with temperature, directly affects electricity generation from wind. In other words, as the air density increases, the momentum of the air passing through the sweeping area increases and high torque is obtained. Additionally, In real operating conditions, many parameters such as turbulence, wind profile on the rotor, angle of wind incidence on the turbine blades, and the response time of the turbine in flow conditions affect the output power of turbines. However, in this study, a variable related to air temperature was not included in the prediction model due to its absence in the Kaggle[13] dataset.

CONCLUSION

This research presents an analysis conducted using the Support Vector Machine (SVM) Regression Algorithm, one of the traditional machine learning regression algorithms, for electricity generation from wind energy power plants. According to the research findings, the Support Vector Machine (SVM) Regression Algorithm exhibits the best performance, particularly when the Radial function is employed. The SVM Regression Algorithm has been evaluated based on performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-Squared (R^2). When the Radial function is used, the SVM algorithm yields lower MSE, RMSE, MAE, and MAPE values compared to its other kernel functions. Additionally, the highest R^2 value of 0.997 is achieved with the Radial function. These study results demonstrate the successful use of the SVM regression algorithm for predicting electricity generation from wind energy power plants. These prediction models can play a significant role in the energy sector and in the better utilization of renewable energy sources. In the near future, the aim is to enhance prediction accuracy by comparing it with other regression algorithms using datasets from wind energy power plants of different scales.

REFERENCE

[1] https://iea.blob.core.windows.net/assets/830fe099-5530-48f2-a7c1-11f35d510983/WorldEnergyOutlook2022.pdf(date of access)

[2]https://gwec.net/wp-content/uploads/2022/04/Annual-Wind-Report-2022_screen_final_April.pdf(date of access)

[3] Albadi, M. H., & El-Saadany, E. F. (2010). Overview of wind power intermittency impacts on power systems. Electric power systems research, 80(6), 627-632.

[4] A. Tuğba, Y. Cemal and S. Yusuf, "Elektrik Yükü Tahmin Yöntemlerinin Analizi," Gazi Mühendislik Bilimleri Dergisi, pp. 168-175, 2018.

[5] Hanifi, S., Liu, X., Lin, Z., & Lotfian, S. (2020). A critical review of wind power forecasting methods—past, present and future. Energies, 13(15), 3764.

[6] Singh, U., Rizwan, M., Alaraj, M., & Alsaidan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A

step towards smart grid environments. Energies, 14(16), 5196.
[7] Yürek, Ö., Birant, D., & Yürek, I. (2021). Wind power generation prediction using machine learning algorithms. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 23(67), 107-119.
[8] Phan, Q. T., Wu, Y. K., & Phan, Q. D. (2021). A hybrid wind power forecasting model with XGBoost, data preprocessing considering different NWPs. Applied Sciences, 11(3), 1100.
[9] Shirzadi, N., Nizami, A., Khazen, M., & Nik-Bakht, M. (2021). Medium-term regional electricity load forecasting through and deep learning. Designs, 5(2), 27.
[10] Zeng, J.; Qiao,W. Support vector machine-based short-term wind power forecasting. In Proceedings of the IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 20–23 March 2011.
[11] Li, L. L., Zhao, X., Tseng, M. L., & Tan, R. R. (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. Journal of Cleaner Production, 242, 118447.
[12] Narasimhan, A. (2022). Support Vector Machine Based Forecasting for Renewable Energy Systems. In Renewable Energy Optimization, Planning and Control: Proceedings of ICRTE 2021, Volume 1 (pp. 149-157). Springer Singapore.
[13]https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset
[14] Alkesaiberi, A., Harrou, F., & Sun, Y. (2022). Efficient wind power prediction using machine learning methods: A comparative study. Energies, 15(7), 2327.
[15] Chicco, D., Warrens, M.J., and Jurman, G., "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, p. e623, 2021.