Techno-Science
Scientific Journal of Mehmet Akif Ersoy University
https://dergipark.org.tr/pub/sjmakeu

Original Research Article

hosted by Turkish JournalPark ACADEMIC

WIND SPEED PREDICTION USING DATA MINING APPROCHES: A CASE STUDY OF GOKCEADA, TURKEY

Fatma Kadriye DUDEN[1*]

[1] *Burdur Mehmet Akif Ersoy University, Department of Energy Systems Engineering, Burdur, Turkey*

ARTICLE INFO

Article History

Received	:	08/09/2023
Revised	:	03/11/2023
Accepted	:	26/11/2023
Available online	:	31/12/2023

Keywords

Data mining, Kstar, Wind speed, Prediction, Gokceada: Turkey

ABSTRACT

In this study, the meteorology data set covering the wind speed, humidity, pressure and temperature data between the years 2014-2021 obtained from the Turkish State Meteorological Service is utilized. With this data set, an estimation is made for the Gokceada district in Canakkale-Turkey, with the WEKA software as pressure and temperature inputs and wind speed output. Gaussian Processes, Linear Regression, Multilayer Perceptron, Simple Linear Regression, SMOreg, Kstar, Decision Table, M5P algorithms in WEKA software are used for estimation. It is made for 7 different groups as temperature-pressure-humidity, temperature-pressure, temperature-humidity, humidity-pressure, temperature, pressure and humidity. According to the results, the best estimation for the temperature-pressure-humidity group is found to be 0.999 for the CC (correlation coefficient) value and 0.2994 for the RMSE (root-mean-square error) with the Kstar algorithm. For the temperature-humidity group, the CC value is 0.9607 and the RMSE value is 0.2777. Estimates from the temperature-pressure and humidity-pressure groups is not give accurate results in comparison to the other groups. The CC and RMSE results are obtained from the humidity and pressure groups are found to be 0.9998 and 0.9985, 0.2679 and 0.0464, respectively.

1. INTRODUCTION

In recent years, renewable energy and clean energy sources have gained a lot of importance around the world. As it is known, wind energy has become a popular alternative energy source because it is sustainable and clean [1,2]. The analysis of the energy potential of the energy fields where wind energy is established is very important. Sufficient and continuous wind speeds are required for these systems to operate effectively [3-5]. However, since wind energy has a structure that changes over time, wind power plants must be planned reliably. Therefore, highly accurate forecasts are needed for the regions where wind turbines are planned to be installed [2,6]. There are many studies available in literature that analyzed wind speed and energy potential of specific regions using different types of machine learning and statistical approaches [7-12]. Zhang et al. [13] used variational mode decomposition for short term wind speed predictions. The stated model has been used to improve the accuracy of genetic algorithm-artificial neural networks combination. In another scientific research, Sareen et al. [14] analyzed India's wind energy potential using deep learning algorithms. In their work, decomposition algorithms are utilized to denoise the existing data. Moreover, Bi-directional long-short term memory technique has been utilized to predict each series. Five cities have been analyzed within the scope of the stated work. The obtained results of the study clearly showed the accuracy of using hybrid modeling techniques in wind speed prediction. Deep et al. [15] estimated wind energy potential of coastal areas of India using statistical techniques. They utilized Weibull distribution in their work. Their findings indicated that the availability factor could be calculated using the parent two-parameter Weibull model, whereas the amount of wind energy that is available between the cut-in and rated wind speeds should be calculated using a three-parameter Weibull model, where the location parameter could be equated to the cut-in wind speed. Serban et al. [16] analyzed the wind energy potential of Galati, Romania using Rayleigh and Weibull distributions. They used the wind speed data between 01/2017-12/2018. It is noted that the annual average wind speed is greater than the wind speed that occurred most frequently. Another key finding is that the highest energy-producing speed of the wind regime, not the average wind speed or the most common wind speed, has the biggest impact on the wind power potential. Nonlinear autoregressive exogenous neural networks and nonlinear autoregressive neural networks with various network

* Corresponding Author: fkorgen@mehmetakif.edu.tr

To cite this article: DUDEN F. K., (2023). Wind Speed Prediction Using Data Mining Approches: A Case Study of Gokceada, Turkey. *Scientific Journal of Mehmet Akif Ersoy University*, vol. 6, no. 2-p.19-23

adjustments have been used by Sarkar et al. [17] to estimate the wind speed estimation in Malaysia. Since activation function is the fundamental component of every artificial neural network model, the main objective of their study is to compare the impact various activation functions, such as tansig and logsig, on the performance of time series forecasting. From various meteorological stations in Malaysia, located in Kuala Lumpur, Kuantan, and Melaka, a set of wind speed data is gathered. The findings of the comparison for the logsig transfer function results as well as the proposed activation functions tansig of nonlinear autoregressive and nonlinear autoregressive exogenous neural networks showed promising results in terms of very tiny error between actual and anticipated wind speed. In another research, Sarkar et al. [18] used adaptive neuro-fuzzy inference system (ANFIS) to estimate wind power in Malaysia. Field information is gathered from two locations in Malaysia over a three-year period in each location, from 12.2013 to 12.2015. ANFIS members developed the model. Additionally, the ANFIS's soft computing capabilities have been used to anticipate wind energy. Utilizing MATLAB, it has been done with four input variables—temperature, humidity, pressure, and wind speed—and one output variable—wind power. Wind speed data is gathered from Malaysia's Meteorological Department to produce wind power. Based on the error values of the anticipated wind power, the proposed predicted model of ANFIS clearly indicates excellent capability and reliability. Purohit et al. [19] used three different machine learning methods to estimate velocity and turbulence intensity of a wind turbine wake. Extreme gradient boosting, support vector regression and artificial neural networks are employed in their work. While operating at a similar speed to low-fidelity wake models, machine learning techniques estimate velocity and turbulence intensity in the wake in a manner comparable to computational fluid dynamics simulations. The findings show that machine learning-based methods are more accurate at predicting velocity and turbulence intensity than conventional analytical wake models. Demolli et al. [20] used five different algorithms to perform wind speed forecasting. The findings demonstrated that using past wind speed data, machine learning algorithms might be utilized to anticipate long-term wind power values. The outcomes also demonstrated that machine learning-based models might be used in regions other than those where they had been trained. However, as could be seen in the literature summary, studies using the WEKA method in wind speed estimation are quite limited. Also no study has been found in the literature on the estimation of the wind potential produced in Gokceada using data mining.

In this study, the meteorology data set covering 8 years of wind speed, humidity, pressure, and temperature data obtained from the Turkish State Meteorological Service is used. With this data set, wind speed estimation is made with Weka software for Gokceada district in Canakkale province. Eight different algorithms are used for the analyses.

2. MATERIALS AND METHODS

Data mining tools are divided into commercial and open-source software [21]. WEKA is an open-source program developed in Java language at Waikito University [22] Data mining is the information processing at researched data base. Weka presents new data by using old data base. In the resulting model must be determined dependent and independent variable. In this study, the Weka, by using the temperature, pressure, humidity and wind speed data in Table 1 obtained from the State Meteorology Station, wind speed estimations are made for the groupings in Table 2 with Gaussian Processes, Linear Regression, Multilayer Perceptron, Simple Linear Regression, SMOreg, Kstar, Decision Table, M5P algorithms [23,24].

Table 1. Average yearly environmental conditions of Gokceada (Coordinates of the meteorological observation station: 40°11'27.6"N 25°54'27.0"E)

Year	Ambient temperature	Pressure	Humidity	Wind speed
	(°C)	(atm)	(%)	(m/s)
2014	16.7	0.9936	83.6	3.7
2015	16.1	0.9954	70.4	4.1
2016	16.6	1.0103	66.3	4.2
2017	15.9	0.9933	68.0	3.8
2018	16.8	0.9921	76.2	4.0
2019	16.7	0.9923	85.7	4.0
2020	16.9	1.1159	67.5	4.1
2021	17.1	0.9926	64.2	4.1

Table 2. Developed group for wind speed prediction

Group 1	Temperature-Pressure-Humidity
Group 2	Temperature-Pressure
Group 3	Temperature-Humidity
Group 4	Humidity-Pressure
Group 5	Temperature
Group 6	Pressure
Group 7	Humidity

C and RMSE indicators are used to determine the wind speed prediction error rate. The correlation coefficient shows the relationship between the actual data and the estimated data. The closer it gets to the +1 value, the stronger the relationship between the two, and the closer it gets to -1, the lower the relationship. If it is zero, the relationship disappears. The equations in Table 3 are used for the CC and RMSE values.

Table 3. Equations of CC and RMSE metrics [25]

Performance metric	Equation
CC	$1-\left[\left(\sum_{m=1}^{n}(t_{m,m}-t_{m,m})\right)/\left(\sum_{m=1}^{n}(t_{m,m}-\overline{t_{m,m}{}^{2}})\right)\right]$
$RMSE$	$\left[\left(\sum_{m=1}^{n}(y_{p,m}-t_{m,m}{}^{2})\right)/\eta\right]^{1/2}$

3. RESULTS AND DISCUSSION

In this study, humidity, temperature, pressure and wind speed data of Gokceada between the years 2014-2021 obtained from the Turkish State Meteorological Service are used. The data in csv format is introduced to the Weka program. The percentage split method is used for estimation. Kstar algorithm give better results than other algorithms for all times. In the analysis make with this algorithm, the correlation coefficient is high and the RMSE value is low. Among the seven different groupings given in Table 2, the predicted success rate of the temperature-pressure-humidity, temperature-humidity group with the Kstar algorithm, and the pressure and humidity group with the multilayer perceptron algorithm is high. Since three factors are effective for wind speed, the K-star algorithm is used, which provides the best estimation for pressure, temperature, humidity. The wind speed estimation results make with 8 different algorithms using the data in Table 1 are given in Table 4. The best estimation values make with the Kstar algorithm using the wind speed, pressure and humidity data between 2014-2021 and the graph containing the actual values are shown in Figure 1 together with the error bars.

Table 4. Performance indicators for developed groups and use algorithms

Algorithm	Performance metric	T-P-H	T-H	T-P	H-P	T	H	P
Gaussian Processes	RMSE	0.1346	0.1291	0.0656	0.1503	0.0657	0.1453	0.0703
	CC	0.5675	0.4871	-0.9902	0.7978	-0.9912	0.7886	0.9985
Linear Regression	RMSE	0.2398	0.2398	0.0712	0.2398	0.0712	0.2398	0.0712
	CC	0.7886	0.7886	-	0.7886	-	0.7886	-
Multilayer Perceptron	RMSE	0.2708	0.2488	0.2123	0.1209	0.1209	0.2679	0.1150
	CC	0.0662	-0.0532	-0.9777	-0.9992	-0.9992	0.9998	0.9986
Simple Linear Regression	RMSE	0.2398	0.2398	0.1424	0.2398	0.1424	0.2398	0.1076
	CC	0.7886	0.7886	-0.9912	0.7886	-0.9912	0.7886	0.9985
SMOreg	RMSE	0.2750	0.2748	0.1442	0.2277	0.1442	0.2275	0.0464
	CC	0.3053	0.2912	-0.9912	0.7912	-0.9912	0.7886	0.9985
Kstar	RMSE	0.2994	0.2777	0.2654	0.2534	0.1701	0.2197	0.2076
	CC	0.9990	0.9607	0.0355	0.2734	-0.2789	0.9766	-
Decision Table	RMSE	0.1869	0.1869	0.1869	0.1869	0.1869	0.1732	0.1732
	CC	0.5000	0.5000	0.5000	0.5000	0.5000	0	0
M5P	RMSE	0.2409	0.2409	0.0712	0.2409	0.0712	-	0.0712
	CC	0.7886	0.7886	0	0.7886	0	0	0

T: Temperature, **H**: Humidity, **P**: Pressure

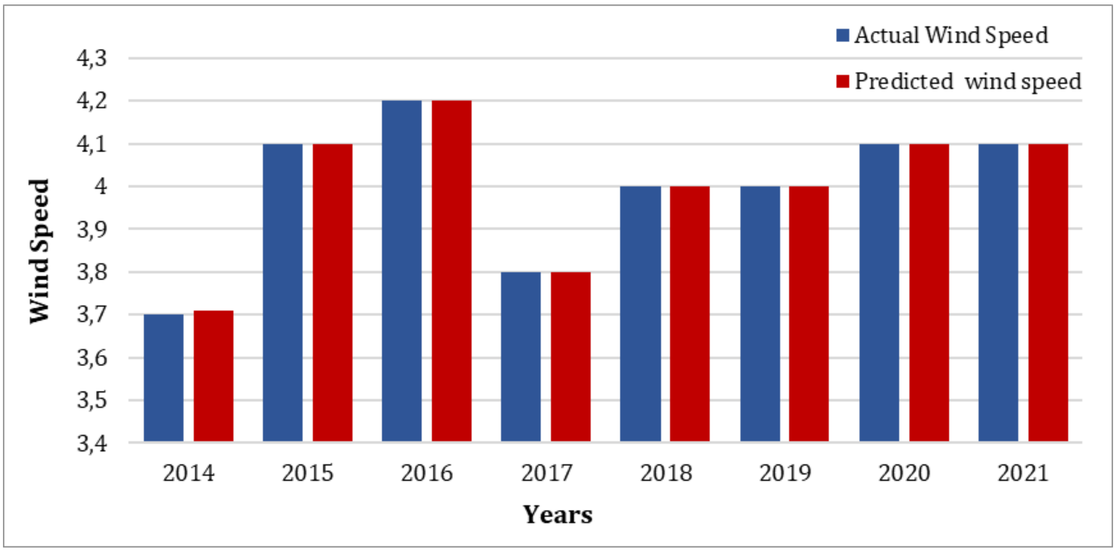

Fig. 1. Actual and predicted wind speed values

4. CONCLUSIONS

In wind speed estimation between 2014-2021, the best results are obtained with the Kstar algorithm, with humidity-temperature-pressure and temperature-humidity inputs. The highest CC values obtained are 0.9994 and 0.9607, respectively, and the lowest RMSE values are 0.2994 and 0.2777, respectively. Accurate predictive values for humidity-pressure, temperature-pressure and temperature inputs could not be obtained. It is obtained by Multilayer perceptron algorithm for moisture input. While the best CC value is 0.9998, the lowest RMSE value is calculated as 0.2679. For the pressure input, the best estimates are made with the SMOreg algorithm. The resulting CC and RMSE values are 0.9985 and 0.0464, respectively. The fact that the RMSE value is close to zero indicates that the predictions are very close to the actual data.

5. ACKNOWLEDGEMENT

The author is thankful to the General Directorate of State Meteorological Affairs for providing Gokceada datas.

REFERENCES

[1] Shabbir, N., Ahmadi Ahangar, R., Kutt, L., Iqbal, M. N., Rosin, A. (2019). Forecasting short term wind energy generation using machine learning. In 2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON) (pp. 1-4). IEEE. DOI: 10.1109/RTUCON48111.2019.8982365

[2] Isık, A. H., Duden Orgen, F. K., Sirin, C., Tuncer, A. D., Gungor, A. (2019). Prediction of wind blowing durations of Eastern Turkey with machine learning for integration of renewable energy and organic farming-stock raising. Tecnho-Science, 2, 47-53.

[3] Rushdi, M. A., Rushdi, A. A., Dief, T. N., Halawa, A. M., Yoshida, S., Schmehl, R. (2020). Power prediction of airborne wind energy systems using multivariate machine learning. Energies, 13(9), 2367. https://doi.org/10.3390/en13092367.

[4] Fugon, L., Juban, J., Kariniotakis, G. (2008). Data mining for wind power forecasting. In European Wind Energy Conference & Exhibition EWEC 2008 (pp. 6-pages). EWEC.

[5] Colak, I., Sagiroglu, S., Yesilbudak, M. (2012). Data mining and wind power prediction: A literature review. Renewable Energy, 46, 241-247. https://doi.org/10.1016/j.renene.2012.02.015

[6] Marugán, A. P., Márquez, F. P. G., Perez, J. M. P., Ruiz-Hernández, D. (2018). A survey of artificial neural network in wind energy systems. Applied energy, 228, 1822-1836. https://doi.org/10.1016/j.apenergy.2018.07.084.

[7] Mora, E., Cifuentes, J., Marulanda, G. (2021). Short-term forecasting of wind energy: A comparison of deep learning frameworks. Energies, 14(23), 7943. https://doi.org/10.3390/en14237943.

[8] Wu, T., Snaiki, R. (2022). Applications of machine learning to wind engineering. Frontiers in Built Environment, 8, 811460. https://doi.org/10.3389/fbuil.2022.811460

[9] Keyhani, A., Ghasemi-Varnamkhasti, M., Khanali, M., Abbaszadeh, R. (2010). An assessment of wind energy potential as a power generation source in the capital of Iran, Tehran. Energy, 35(1), 188-201. https://doi.org/10.1016/j.energy.2009.09.009

[10] Nezhad, M. M., Heydari, A., Neshat, M., Keynia, F., Piras, G., Garcia, D. A. (2022). A Mediterranean Sea Offshore Wind classification using MERRA-2 and machine learning models. Renewable Energy, 190, 156-166. https://doi.org/10.1016/j.renene.2022.03.110

[11] Mortezazadeh, M., Zou, J., Hosseini, M., Yang, S., Wang, L. (2022). Estimating urban wind speeds and wind power potentials based on machine learning with city fast fluid dynamics training data. Atmosphere, 13(2), 214. https://doi.org/10.3390/atmos13020214

[12] Akpinar, E. K. (2006). A statistical investigation of wind energy potential. Energy Sources, Part A, 28(9), 807-820. https://doi.org/10.1080/009083190928038

[13] Zhang, Y., Pan, G., Chen, B., Han, J., Zhao, Y., & Zhang, C. (2020). Short-term wind speed prediction model based on GA-ANN improved by VMD. Renewable Energy, 156, 1373-1388. https://doi.org/10.1016/j.renene.2019.12.047

[14] Sareen, K., Panigrahi, B. K., Shikhola, T., Sharma, R. (2023). An imputation and decomposition algorithms based integrated approach with bidirectional LSTM neural network for wind speed prediction. Energy, 278, 127799. https://doi.org/10.1016/j.energy.2023.127799

[15] Deep, S., Sarkar, A., Ghawat, M., Rajak, M. K. (2020). Estimation of the wind energy potential for coastal locations in India using the Weibull model. Renewable energy, 161, 319-339.

[16] Serban, A., Paraschiv, L. S., Paraschiv, S. (2020). Assessment of wind energy potential based on Weibull and Rayleigh distribution models. Energy Reports, 6, 250-267. https://doi.org/10.1016/j.egyr.2020.08.048

[17] Sarkar, R., Julai, S., Hossain, S., Chong, W. T., Rahman, M. (2019). A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia. Mathematical Problems in Engineering. https://doi.org/10.1155/2019/6403081

[18] Sarkar, M. R., Nahar, M. J., Nadia, A., Halim, M. A., Rafin, S. S. H., Rahman, M. M. (2019). Proficiency assessment of adaptive neuro-fuzzy inference system to predict wind power: A case study of Malaysia. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-5). IEEE.

[19] Purohit, S., Ng, E. Y. K., Kabir, I. F. S. A. (2022). Evaluation of three potential machine learning algorithms for predicting the velocity and turbulence intensity of a wind turbine wake. Renewable Energy, 184, 405-420. https://doi.org/10.1016/j.renene.2021.11.097

[20] Demolli, H., Dokuz, A. S., Ecemis, A., Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. Energy Conversion and Management, 198, 111823. https://doi.org/10.1016/j.enconman.2019.111823

[21] Berti, A., Van Zelst, S. J., van der Aalst, W. (2019). Process mining for python (PM4Py): bridging the gap between process-and data science. arXiv preprint arXiv:1905.06169.

[22] Hammoodi, M. S., Al Essa, H. A., Hanon, W. A. (2021, February). The Waikato Open Source Frameworks (WEKA and MOA) for Machine Learning Techniques. In Journal of Physics: Conference Series (Vol. 1804, No. 1, p. 012133). IOP Publishing. 10.1088/1742-6596/1804/1/012133

[23] Gao, W., Alsarraf, J., Moayedi, H., Shahsavar, A., Nguyen, H. (2019). Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. Applied Soft Computing, 84, 105748. https://doi.org/10.1016/j.biombioe.2009.12.016

[24] Al-Rousan, N., Al-Najjar, H., Alomari, O. (2021). Assessment of predicting hourly global solar radiation in Jordan based on Rules, Trees, Meta, Lazy and Function prediction methods. Sustainable Energy Technologies and Assessments, 44, 100923. https://doi.org/10.1016/j.seta.2020.100923

[25] Reiss, M. A., Temmer, M., Veronig, A. M., Nikolic, L., Vennerstrom, S., Schongassner, F., Hofmeister, S. J. (2016). Verification of high-speed solar wind stream forecasts using operational solar wind models. Space Weather, 14(7), 495-510. https://doi.org/10.1002/2016SW001390

Techno-Science *Paper ID: 1357259*