MATEC Web of Conferences **393**, 02003 (2024) https://doi.org/10.1051/matecconf/202439302003
STAAAR-2023

Machine Learning-Driven Wind Energy Forecasting for Sustainable Development

Magesh T[1*], *Samuel Franklin* F[1], *Santhi* P S[2] and *Thiyagesan* M[1]

[1]R.M.K. Engineering College, Department of Electrical and Electronics Engineering, Kavaraipettai, Tamil Nadu, India.
[2]Murugappa Polytechnic College, Department of Electrical and Electronics Engineering, Avadi, Tamil Nadu, India.

Abstract. The growing need for energy, in addition to the depletion of fossil fuel supply, has underlined the importance of renewable energy for long-term growth. Renewable energy stands out among these, but its broad usage is hampered by the inherent uncertainty of wind power generation. This study uses machine learning to predict wind energy yield. Several regression models were used, including decision tree regression, linear regression, and random forest regression. The results emphasize the random forest regression, which has a high R-squared score, suggesting strong predictive ability. The paper also contains wind power output projections, which provide insights for optimal wind energy planning and usage. Overall, this attempt gives vital insights to improving the effective use of renewable energy, advancing the cause of sustainable development.

1 Introduction

The growing global population, as well as the need for modernization and industrialization, highlight the urgent requirement for sustainable and secure power sources. Unfortunately, traditional nonrenewable Fuels like petroleum and natural gas are essential for meeting energy needs, but they can significantly worsen the environment through pollution and greenhouse gas emissions. As a result, several countries are investigating renewable, plentiful, and clean sources of energy such as wind, hydroelectricity, solar power, geothermal, and biomass [1]. This transition to renewable energy demonstrates a commitment to reducing the impact on the environment and ensuring an environmentally friendly energy future. Wind energy has grown in popularity as a dependable and environmentally acceptable energy source. Wind, being an informal energy source, is widely available and may be used in a variety of locales. As a result, wind power already occupies a significant part of the worldwide mix of energy sources and is anticipated to do so for the foreseeable future [2]. Wind energy is growing into a more important energy source for the power grid, which makes it increasingly difficult to predict its behavior. The power system control center may need to boost backup capacity and incur more expenditures as an outcome of the unexpected variations in wind output, which can lead to uncertainty and imbalance in

* Corresponding author: tmh.eee@rmkec.ac.in

the system. However, reliable forecasting models can lessen these difficulties by enhancing the system's stability, dependability, and power quality while lowering operating expenses. Developing models that reduce wind power forecast errors is essential to overcoming these obstacles and advancing the incorporation of wind energy into the electricity grid.

2 Literature Review

In the domain of forecasting wind power, there are two primary types: Short Term (ST) and Long Term (LT) forecasting. ST Forecasting focuses on predicting wind power for brief durations, spanning from one hour to several days ahead. On the other hand, long-term forecasting extends its predictions to periods ranging from several days to a year in advance. Statistical approaches have shown to be useful in the ST wind forecasting area. Lima et al. [3] suggested a climatic-statistical method that can anticipate wind power up to 72 hours ahead of time with Wang et al. [4] effective performance. Robles-Rodriguez and Dochain [5] introduced an ARMAX technique which incorporates a process including division of series and non-linearity management, resulting in accurate wind power forecasts up to 48 hours in advance. Pearre and Swan [6] presented statistical strategies for detecting and correcting prediction mistakes, demonstrating high performance in forecasts one to six hours in advance.

Yu Zhang, Yanting Li *, and Guangyao Zhang [7] developed a wind output forecast model using machine learning for ST forecast. This model beat other existing approaches as the Deep Belief System and Random Forest by combining a variable time series clustering system with a deep learning system. Sideratos and Hatziargyriou [8] proposed a technique for precisely estimating wind power output with a 48-hour horizon that incorporates neural networks and fuzzy logic. To anticipate electricity output, Lahouar and Slama [9] used the Random Forest method in conjunction with numerous meteorological data such as wind speed.[10] suggested an architecture for improving mathematical forecasts and data using deep belief networks and the k-means clustering technique.

After evaluating the research in the literature, it is clear that statistical approaches are not the most effective choice for predicting wind power, especially in the long run. They struggle with nonlinear wind data adaptation, massive dataset processing, and forecasting over long time horizons [10, 11]. In this paper, several machine learning approaches were applied to anticipate LT power. The results indicate that using machine algorithms to anticipate wind power shows promising outcomes. The investigation also looked at the effect of the standard deviation of the model. Finally, the best strategy for reliable wind power estimates was discovered.

3 Architecture

In order to minimize cost overruns and guarantee that demand is satisfied, forecasting for wind energy must be accurate. Instead of utilizing generalized global data to train the model, local wind data must be used to provide precise forecasts. In this study, we compare the model's accuracy in predicting wind power output and concentrate on the training process.

Depending on the type of learning input or response available, the research takes regression machine learning systems into consideration.

3.1 Regression

Since the objective of this research is to anticipate continuous outputs, regression, a supervised learning problem, is frequently used. In order to properly estimate wind power, regression techniques are used. Data preprocessing is a crucial step in turning raw data into a format that can be used. There are several preprocessing processes involved in the research.

3.1.1 Data cleaning

Data cleaning deals with irrelevant or missing data parts. Handling missing data, noisy data, and other data quality concerns are part of this process.

3.1.2 Missing and Noisy Data

Missing data may be handled in a variety of methods, including disregarding tuples with missing values, manually adding the missing values, utilizing the attribute mean, or choosing the value that is most likely. Noisy data is unintelligible and lacks useful information. The data can be smoothed using regression by fitting it to a regression function.

Data transformation is performed to convert information into a suitable format for the mining process. Two common techniques are feature scaling and dividing sets.

In machine learning, this stage is essential. Algorithms that determine the distances between features may be biased toward numerically bigger values if data is not scaled. In order to scale features to a similar scale, common feature scaling techniques include normalization (or Min-Max) and standardization (R-Score).

In dividing train and test sets the dataset is divided into two sets to confirm the accuracy of the machine learning model. Train set is used for training, while a Test set is used for assessment. This ensures a reliable assessment of the model's performance and helps prevent biased findings.

4 Methodology

Various machine learning models that are employed for Wind forecasting in this research.

4.1 Random Forest

A common decision tree approach that builds several decision trees based on the input information is the Random Forest algorithm. It generates decision trees for every smaller subdivision of the input parameters rather than using all of them. The ultimate choice is then reached by combining the output from each tree. Using this method makes the issue of juggling several features easier to grasp.

In RF, a random vector of size k is created, representing a portion of the dataset's feature space. The training dataset and k are used to construct each tree.

$$PE^{*}=PX,Y(mg(X,Y) <0)$$

4.2 Linear Regression

The statistical method known as linear regression is employed to examine the relationship that exists between a scalar result, often referred to as the variable of interest, and one or more independent variables that serve as explanatory factors. These correlations are represented by the method's use of linear predictor functions, and the model's unknown variables are estimated using the given dataset.

4.3 Decision Tree Regression

Forecasting continuous numerical values is an area in which decision tree regression, a supervised machine learning method, shines. With the expected numerical output represented by the leaf nodes and judgments made based on the input data by the inside nodes, it constructs a structure like a tree. The interpretability of this strategy is valued since it makes it easier for customers to comprehend how decisions are made.

5 Results and Discussion

The dataset used in this study, consisting of 5 years of every-hour measurements of wind speed from Coimbatore, India, was utilized to train and assess learning methods for wind power forecasts. The training and evaluation of the models were conducted using the Google Colab platform, for running Python code, along with the scikit-learn (sklearn) library for machine learning. The wind output power for every hour is calculated on the standard power formula. The generated power values are then used as the target variable for training and evaluating the learning algorithms for wind forecasts.

Table 1. Different statistical measure

	Square R	RMSE	MAE	MSE
Linear	0.9251	0.4487	0.3917	0.2013
Decision Tree	0.8722	0.5861	0.4677	0.3435
Random Forest	0.9264	0.4447	0.3839	0.19776

The Table contains the various scores of the differential statistical method obtained by training the models and testing. The effectiveness of different learning algorithms, such as linear, Random Forest and decision tree was evaluated using this dataset. The accuracy of these models was assessed by comparing their predictions against the actual wind power values.

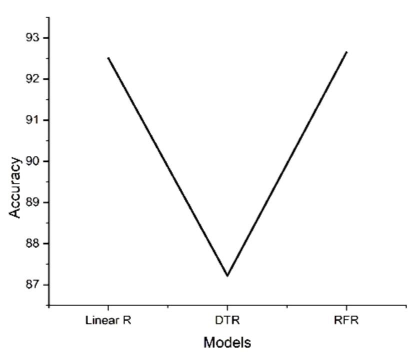

Fig 1. Models and Scores

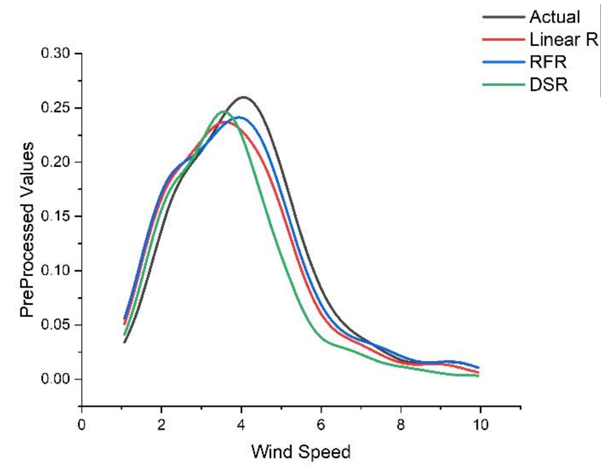

Fig 2. Accuracy Graph Comparison

The study's findings show that a significant number of models based on learning methods, like linear regression and random forest, perform well in forecasting wind power. The high R-scores approach one indicates that the models explain a big part of the difference in wind energy data. This closeness to 1 indicates a high level of predictability in wind power, highlighting the efficiency of modern machine learning algorithms. However, decision tree regression appears to have lower accuracy compared to the other models, as indicated by its line deviating significantly from the black line and its lower R-score. Among the different machine learning models evaluated, the results indicate that the random forest method performed the top in predicting wind power having score as 0.9264. From fig 2 The model shows the line in the graph closely aligning with the black line representing the actual values, indicating the ability to capture the underlying patterns in the wind data and provide accurate predictions. These findings suggest that learning algorithms, particularly linear and random forests, have the potential to accurately forecast wind energy using the provided dataset.

6 Conclusion

In conclusion, the necessity for sustainable development and the expanding global energy demand have highlighted the significance of renewable energy sources. A lot of attention has been paid to wind energy in particular because of how abundant it is and how much electricity it can provide. The difficulty, however, is that wind power generation is unpredictable, which makes it difficult to use it effectively.

This study attempted to address the difficulty of wind power prediction using several models of machine learning methods, like linear regression, decision tree regression, and random forest networks. The results revealed significant potential for prediction in both the linear and random forest regression models, as evidenced by their high R-squared values. Compared to both the models the random forest regression is more efficient. Accurate wind power forecasting is critical for successful planning and maximum exploitation of renewable energy resources.

The objective of the study was to handle the problem of power prediction by using several machine-learning methods. The outcomes, which were characterized by the differential statistical methods, demonstrated the linear regression and random forest regression models' strong predictive power. Precise wind power prediction is important for successful wind resource planning. It can be used for grid planning for balancing the demand and supply and in cooperating into the existing system.

Furthermore, accurate wind power generation projections allow for more accurate power balancing and economic operations calculations, which enhance system stability, lower operating expenses, and improve power quality. An energy infrastructure that is more robust and sustainable is promoted by the usage of power in power networks.

In conclusion, the study highlights the need for precise wind power forecasts and offers a machine learning-based solution to the problems brought on by the unpredictable nature of wind power output. We can open the door to a more sustainable future by using the potential of renewable sources.

References

1. Global Wind Report 2019; Global Wind Energy Council: Brussels, Belgium, 2020. https://gwec.net/global-wind-report-2019/ .
2. Tsikalakis, Antonis & Hatziargyriou, Nikos & Katsigiannis, Yiannis & Georgilakis, Pavlos. (2009). Impact of Wind Power Forecasting Error Bias on the Economic Operation of Autonomous Power Systems. Wind Energy. 12. 315 - 331. 10.1002/we.294.
3. Lima JM, Guetter AK, Freitas SR, Panetta J, de Mattos JGZ. A meteorological-statistic model for ST wind power forecasting. J. Control Autom. Electr. Syst. (2017);28(5):679–91. doi: 10.1007/s40313-017-0329-8.
4. Wang J, Zhou Q, Zhang X. IOP Conf. Ser.: Earth Environ.Sci. (2018);199:022015.https://doi.org/10.1088/1755-1315/199/2/022015
5. Robles-Rodriguez C, Dochain D. Decomposed threshold armax models for short- to medium-term wind power forecasting. IFAC-PapersOnLine (2018);51(13):49–54, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2018.07.253.
6. Pearre NS, Swan LG. Statistical approach for improved wind speed forecasting for wind power production. Sustainable Energy Technol. Assess. (2018);27:180–91. https://doi.org/10.1016/j.seta.2018.04.010.
7. Yu Zhang, Yanting Li, Guangyao Zhang, Short-term wind power forecasting approach based on Seq2Seq model using NWP data, Energy, Volume 213, (2020), 118371, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2020.118371.
8. Sideratos G, Hatziargyriou ND. An advanced statistical method for wind power forecasting. IEEE Trans. Power Syst. (2007);22(1):258–65. https://doi.org/10.1109/TPWRS.2006.889078.
9. Lahouar A, Slama JBH. Hour-ahead wind power forecast based on random forests. Renewable Energy (2017);109:529–41. https://doi.org/10.1016/j.renene.2017.03. 064.
10. Sun G, Jiang C, Cheng P, Liu Y, Wang X, Fu Y, He Y. STwind power forecasts by a synthetical similar time series data mining method. Renewable Energy (2018);115:575–84. https://doi.org/10.1016/j.renene.2017.08.071.
11. Yu R, Gao J, Yu M, Lu W, Xu T, Zhao M, Zhang J, Zhang R, Zhang Z. LSTM-EFG for wind power forecasting based on sequential correlation features. Future Gener. Comput. Syst. (2019);93:33–42. https://doi.org/10.1016/j.future.2018.09.054.