International Journal of Sustainable Energy and Environmental Research
2024 Vol. 13, No. 2, pp. 83-94
ISSN(e): 2306-6253
ISSN(p): 2312-5764
DOI: 10.18488/13.v13i2.3777

A machine learning approach to forecast wind speed based on geographical location in Bangladesh

Elmeeh Hasan Shipra[1+]
Md Sujaur Rahaman[2]
Tasnim Ara[3]
Saeed Mahmud Ullah[4]

[1,2,3]*Department of ICT, Bangladesh University of Professionals, Dhaka, Bangladesh.*
[1]*Email: elmeehhasan111@gmail.com*
[2]*Email: suja312998@gmail.com*
[3]*Email: tasnimaramim96@gmail.com*
[4]*Department of Electrical and Electronic Engineering, University of Dhaka, Bangladesh.*
[4]*Email: ullahsm@du.ac.bd*

(+ Corresponding author)

Article History
Received: 26 March 2024
Revised: 30 April 2024
Accepted: 17 May 2024
Published: 20 June 2024

Keywords
Bangladesh
Geographical location
Machine learning
Water bodies
Wind speed forecast
XGBoost.

ABSTRACT

This paper examines a learning approach to forecasting wind speed based on geographical location in Bangladesh. Most people use wind energy, a rapidly expanding renewable energy source, to produce electricity, replacing traditional fossil fuel-based electricity generation. The variable nature of wind speed necessitates an accurate estimate for planning wind power generation and grid integration. Machine learning models, based on various methods, commonly make wind speed predictions. This study implements machine learning algorithms, namely Random Forest Regression, XGBoost Regression, Multi-Layer Perceptron Regression, Ridge Regression, and Lasso Regression, to determine wind speed prediction accuracy. Firstly, we divide the dataset on wind records of Bangladesh into seven categories, each encompassing different geographical locations in Bangladesh, to account for the change in wind characteristics based on location, and then apply the algorithms to each category. The performance metrics Mean Absolute Error, Mean Squared Error, and Root Mean Square Error are utilized to draw comparisons. The findings showed that the performance of XGBoost makes it the most reliable tool for predicting wind speed in Bangladesh. The near proximity to large water bodies causes considerable variation in accuracy, i.e., it showed higher accuracy compared to the other three algorithms after applying the ensemble learning approach, which is more effective and accurate. This research will help in identifying optimal power plant locations and efficient linking methods to Bangladesh's power grid, ensuring smooth electricity access and efficient utilization of the country's energy resources.

Contribution/Originality: The research evaluated the continuous value of wind speed by categorizing the dataset according to the distance from water bodies and identified the most effective method for wind speed prediction in Bangladesh.

1. INTRODUCTION

Wind power is quickly becoming one of the most popular and sustainable forms of renewable energy. Globally, induced air pollution is a problem since the current energy structure relies heavily on the burning of fossil fuels to produce electricity.

Solar, hydro, wind, biomass, tidal, geothermal, etc. are all examples of renewable energy that have collected significant attention from scientists and business people alike. Wind energy has been considered one of the most important renewable energy sources due to its low environmental impact, great efficiency, and widespread construction of wind farms around the world. As the motion of wind turbines and energy production depend on wind behavior, it is vital to accurately anticipate wind speed. Utilizing machine learning techniques, this prediction can be made (Zucatelli, Nascimento, Santos, Arce, & Moreira, 2021). The amount of wind that blows across the earth is more than enough to meet the world's electricity demand.

In 2019, fossil fuels continued to provide 63.3 percent of the world's power, mostly in the form of coal and gas, which was a significant contributor to the almost 33 gallons of carbon dioxide that were released into the atmosphere that year (Hess, 2021).

A significant reduction in carbon dioxide emissions may be achieved by moving the generation of electricity away from fossil fuels and toward renewable sources. In Bangladesh, the energy industry is the major contributor to CO2 emissions, with 93.09 metric tons, or 55.07 % of total emissions. The Cox's Bazar wind farm project in Bangladesh will play an active role in fostering economic development, energy savings, pollution reduction, and environmental protection in our nation. It has the potential to reduce emissions of 109,200 tons of Carbon Dioxide (CO_2), 25.15 tons of sulfur dioxide (SO_2), and 50.69 tons of Nitric Oxide (NO). When finished, the construction project will have a potential to generate more than 145,600 MWh of electricity annually (Bangladesh's First Large Wind Power Project Launched, 2021).

2. LITERATURE REVIEW

Predicting wind speed is possible using a variety of methods, including physical modeling, statistical analysis, machine learning, and deep learning.

Time series study of wind speeds, especially for short-term forecasting, makes use of statistical approaches such as the Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models, the Kalman filter and the Markov chain, the auto-regressive model, the autoregressive moving average, the autoregressive conditional, the Heteroscedasticity model, etc. we use physical approaches to make climate forecasts.

The development of Numeric Weather Prediction (NWP) models that are capable of predicting future data is made possible by the exploitation of existing physical data in conjunction with topographical information. The neural network models used by the authors, their accuracy, and their shortcomings are briefly given in Table 1.

Table 1. Neural network model approaches.

No.	Author	NN model	Accuracy	Shortcomings
1	Zhu, Chen, Zhu, Duan, and Liu (2018)	PDCNN	99%	Model selection complexity, such as the number of layers, the dimension of each layer, etc.
2	Afrasiabi, Mohammadi, Rastegar, and Afrasiabi (2020)	CNN, GRU	75%	Time complexity
3	Azad, Mekhilef, and Ganapathy (2014)	FFBPNN	75%	High percentage of error in long term prediction of wind speed
4	Filik and Filik (2017)	ANN	69%	Not suitable for long term prediction
5	Kulkarni, Patil, Rama, and Sen (2008)	ARIMA, ANN	70%	No dynamical methods could be developed for the prediction

The machine learning models used by the authors, their accuracy and shortcomings are briefly given in Table 2.

Table 2. Machine learning model approaches.

No.	Author	ML model	Accuracy	Shortcomings
1	Salehin, Billah, Zahin, and Haque (2022)	RF classifier, DT, KNN, XGBoost	99%	Not suitable for short term prediction
2	Cai et al. (2020)	XGBoost	60%	Low accuracy
3	Khosravi, Machado, and Nunes (2018)	MLFFNN, SVR, FIS, ANFIS, ANFIS-PSO, ANFIS-GA, GMDH	99%	A number of hidden layers created complexity
4	Gupta, Natarajan, and Berlin (2022)	ANN	95%	Not suitable for long term prediction
5	Ghorbani, Khatibi, FazeliFard, Naghipour, and Makarynskyy (2016)	MLR	96%	Individual models have to be studies in greater detail to gain insight into their performance

In recent years, researchers have been formulating suitable mathematical models to accurately predict wind speed, as the predictability of wind data at a certain site is crucial for the evaluation of a wind energy utilization project.

To enhance the planning of wind energy generating facilities and minimize economic costs, predicting the wind speed accurately is necessary.

3. METHODOLOGY

3.1. Data Collection

In this work, we have collected our dataset from the website 'Kaggle,' which is a combination of different datasets, most of which were taken from various sources and merged from the Bangladesh Meteorological Department (BMD). It contains monthly average data for 65 years, e.g., maximum and minimum temperature, rainfall, relative humidity, wind speed, cloud coverage, bright sunshine, etc. of Bangladesh from 1948 to 2013 in 32 districts (Reza, 2019). This dataset is chosen to determine the wind speed in various parts of Bangladesh as its records provide insight into historical wind patterns.

3.2. Data Categorization

The distinct places in the dataset are categorized into six categories based on the geographical locations and wind state. Because the stations are located in different parts of Bangladesh, some of them have different weather conditions and landforms.

This categorization aids in the creation of a dependable and universal method for wind speed prediction, applicable across Bangladesh, irrespective of distinctive characteristics of individual locations. The goal of categorization is to identify the most suitable method that can provide precise and uniform forecasts of wind speed at any location in Bangladesh, regardless of the local topography, with no fluctuation.

The picture of the map of Bangladesh, taken from Google Maps is shown in Figure 1, where we indicated different stations, in each category. The small red circles are the different stations and the large oval-shaped markings are the six different categories.

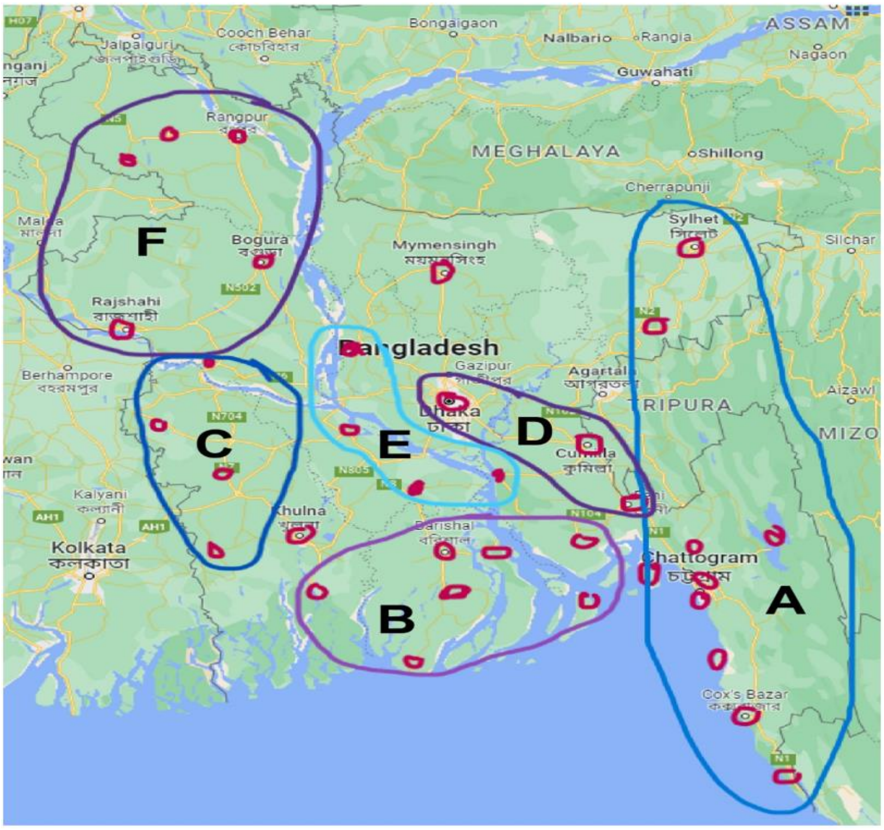

Figure 1. Categories marked on Bangladesh map.

Table 3. Categorization of the places referenced in dataset.

Category A	Category B	Category C	Category D	Category E	Category F	No category
Chittagong (City-Ambagan)	Barisal	Ishurdi	Dhaka	Faridpur	Dinajpur	All 32 places
Chittagong (IAP-Patenga)	Bhola	Jessore	Feni	Madaripur	Syedpur	
Sandwip	Khepupara	Chuadanga	Comilla	Tangail	Rangpur	
Sitakunda	Mongla	Satkhira		Chandpur	Bora	
Kutubdia	Patuakhali	-	-	-	Rajshahi	
Cox's Bazar	Hatiya	-	-	-	-	
Teknaf	-	-	-	-	-	
Sylhet	-	-	-	-	-	
Srimangal	-	-	-	-	-	
Rangamati	-	-	-	-	-	

In Table 3, the stations falling under the six categories are listed. Category A includes the hilly regions of the south-east and north-east. Category B includes the districts of the Barisal division. The Khulna Division classifies most locations as Category C. Comilla, Feni, and Dhaka make up Category D. The areas close to large water bodies are included in Category E. we classify the districts in the north as Category F. Then, under the 'No category' column, all the different locations are listed.

3.3. Data Preprocessing

In order to use the data in the algorithm, we need to use clean data, i.e. converted into a more meaningful and usable form from raw state.

i) We have used the is null () function in Pandas to check the null or missing values of the dataset. We didn't have any missing values in our dataset, which we understood by observing the output of the sum() function.

ii) The dataset we used in our paper contained 17 features, and we have reduced the number of features to 7 while keeping the important features named MinTemp, Rainfall, Relative Humidity, Cloud Coverage, Latitude, Longitude, and Period to make the computation easier and faster, resulting in higher accuracy of the target output. We used correlation and Random Forest Classifier Technique to choose which features to keep.

iii) We divided the dataset into 80% and 20% for our model; the first percentage is used for training, and the remaining percentage is utilized for testing.

iv) We adopted feature scaling, which converts the data framework into a common format that brings consistency to the environment. In other words, it normalizes the range of the data features.

3.4. Machine Learning Algorithms

In this paper, we adopted five machine learning algorithms, which are Random Forest Regression (RF), Extreme Gradient Boosting Regression (XGBoost), Multi-Layer Perceptron Regression (MLP), Ridge Regression, and Lasso Regression.

3.4.1. Random Forest Regression

A reliable and accurate regression approach is the Random Forest model. The diagram in Figure 2 shows how a Random Forest is structured.

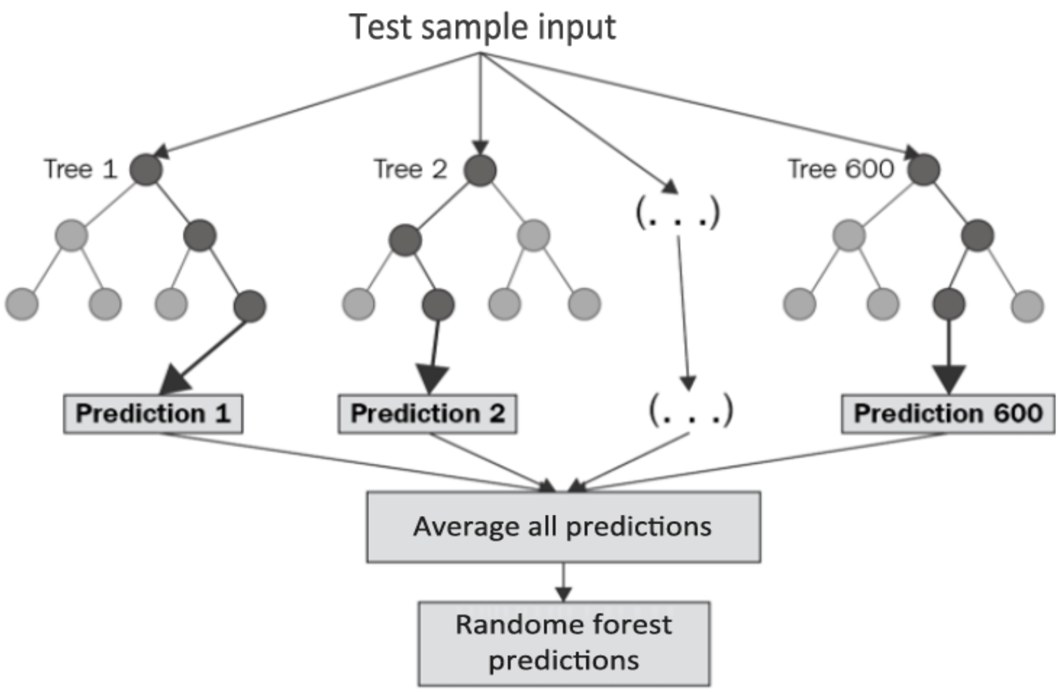

Figure 2. Random forest structure.

Source: Chaya (2020).

Since the trees are growing simultaneously, there is no touch between them. A Random Forest develops a number of decision trees during the training phase, and the class mean is generated as the forecast for each tree (Singh, 2019).

3.4.2. XGBoost Regression

The Extreme Gradient Boosting technique, or XGBoost, enables direct regression predictive modelling. People often refer to Boosting as an additive model because it maintains the consistency of the existing trees by gradually incorporating weak learners into the model. The fully evolved model gets better at forecasting as we merge more and more basic models.

Regression gradients are improved by computing the difference between the current forecast and the known, accurate goal value. Gradient boosting regression is then used to transfer features to the residual and train a subpar model. By adding the predicted residual from a weak model to the current model's input, this strategy propels the model towards its intended goal. The model forecast improves overall when this step is repeated (Dhiraj, 2019). The method of XGBoost Regression is shown in Figure 3.

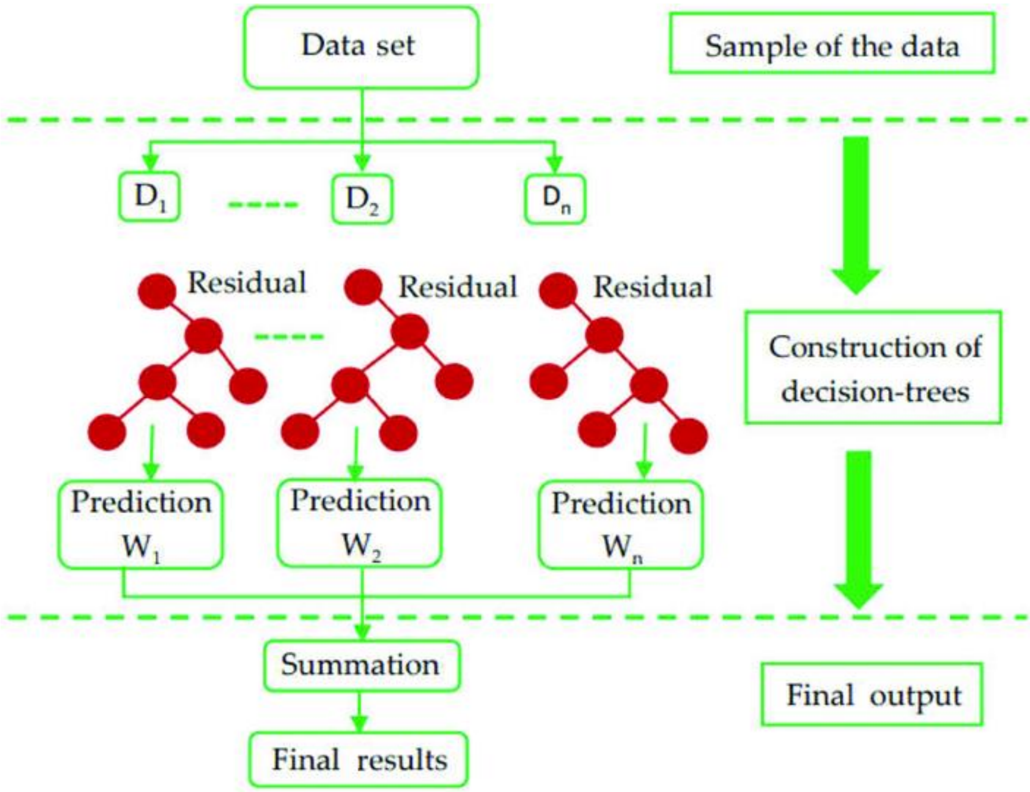

Figure 3. XGBoost regression method.

Source: Amjad et al. (2022).

3.4.3. MLP Regression

MLP, referred to as a multilayer perceptron, is made up of a single layer of neurons or several layers. The output layer produces predictions. For classification prediction issues when the inputs are labeled or classified, MLPs perform well (Brownlee, 2022).

3.4.4. Ridge and Lasso Regression

Ridge regression reduces linear model overfitting by adding a penalty term to the error function, which reduces the magnitude of the coefficients. When the data is noisy, this might protect the model from being unduly sensitive to certain data points (Kumar, 2023). Lasso Regression is a shrinkage technique that also adds a penalty term, causing the coefficients to decrease as well as the complexity and multi-collinearity of the model (Lasso and Ridge Regression in Python Tutorial, 2022).

3.5. Performance Measures

A regression model is assessed using three metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These measures show us the degree of divergence from the real data and the accuracy of our predictions.

Typically, we use MAE to measure performance using continuous variable data. The model performs better when its value is lower. When a dataset has outliers—unexpected values that are either too high or too low—the mean square error (MSE) is most helpful. When there are significant mistakes that significantly affect the model's performance, RMSE is far more helpful. Its ability to avoid considering the error's absolute value is helpful in many mathematical computations.

Four commonly used statistical measures are used to evaluate the model's overall forecasting performance, i.e., R-squared (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). R^2 shows how much of the variance in a dependent variable in a regression model can be accounted for by independent variables.

If its value is 0.5, the inputs to the model explain about half of the observed variation. We may plot fitted values versus observed values to visually illustrate how R^2 values show the dispersion around the regression line (Frost, 2022). The equation of R^2, is given in Equation 1.

$$R^2 = \frac{Variance\ explained\ by\ the\ model}{Total\ variance} \quad (1)$$

The variance between the instance's actual value and its expected value represents each prediction inaccuracy in MAE.

$$\text{MAE} = \frac{\sum_{n}^{i=1} abs(y_i - \lambda(x_i))}{n} \quad (2)$$

In Equation 2 y_i is the actual target value for test instance x_i , $\lambda(x_i)$ is the anticipated target value, and n is the total number of test instances. By calculating the mean of the squared difference between the data set's original values and its predicted values, the mean squared error is determined.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_l)^2 \quad (3)$$

In Equation 3 y_i is the actual output value, (y_i) is the predicted output and n is the sample size. Squaring the MSE allows one to find the Root Mean Square Deviation, or RMSE. When the RMSE number is zero, a model fits the data perfectly.

The value of RMSE is lower as the model and its predictions are better. The formula of RMSE is given in Equation 4 (Padhma, 2021).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_l)^2} \quad (4)$$

3.6. Working Flowchart

We have shown how the complete model operates in the workflow diagram given in Figure 4 from dataset input through model estimation.

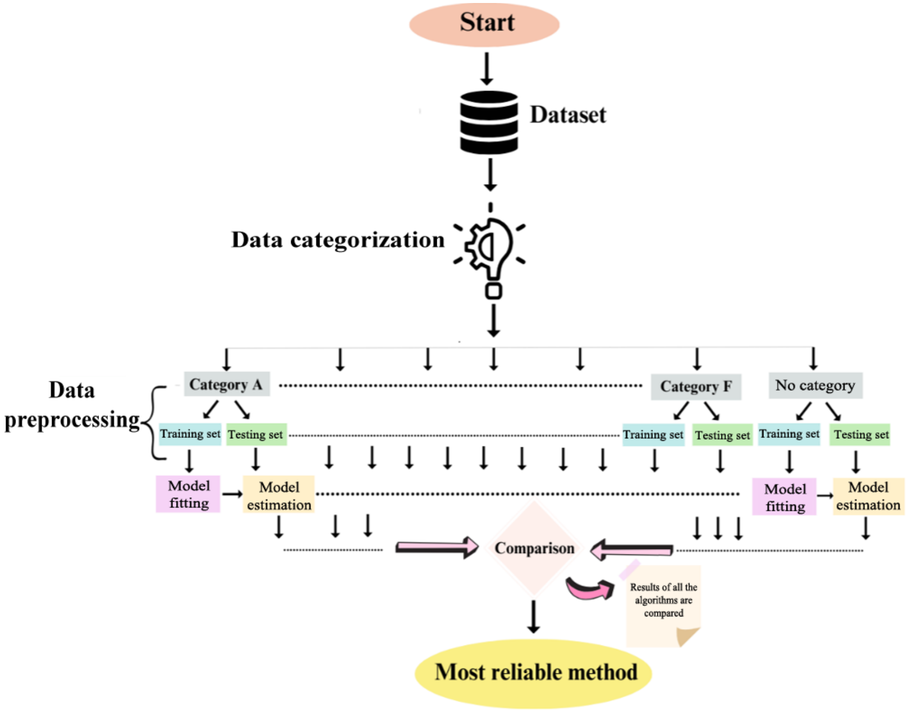

Figure 4. Working flowchart of our work.

We created six categories from the dataset, including one with all kinds of data. Each category is subdivided into an 80% training set and 20% testing set. Each training set uses the previously mentioned five algorithms. After training, the model is tested with new data samples. In the final step, we compare the data using the four performance measures and determine the best method.

4. SIMULATION AND RESULTS

We have programmed all the algorithms in Python.

4.1. Graphical Representation of Accuracy and Error of Different Algorithms

The algorithms used in this paper are compared with the help of line chart.

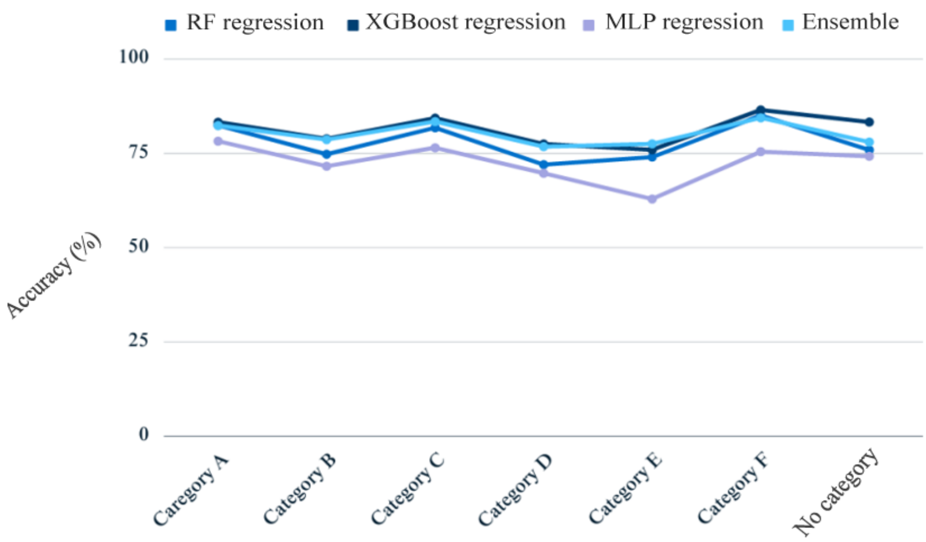

Figure 5. Accuracy of various algorithms.

Figure 5 indicates that the accuracy given by the XGBoost Regression is higher in all categories compared to other algorithms. On the other hand, Ridge and Lasso Regression is continuously showing the lowest accuracy, as these two algorithms have the tendency to keep only one variable and set other correlated variables to zero, which leads to some loss of information and lower accuracy in the model.

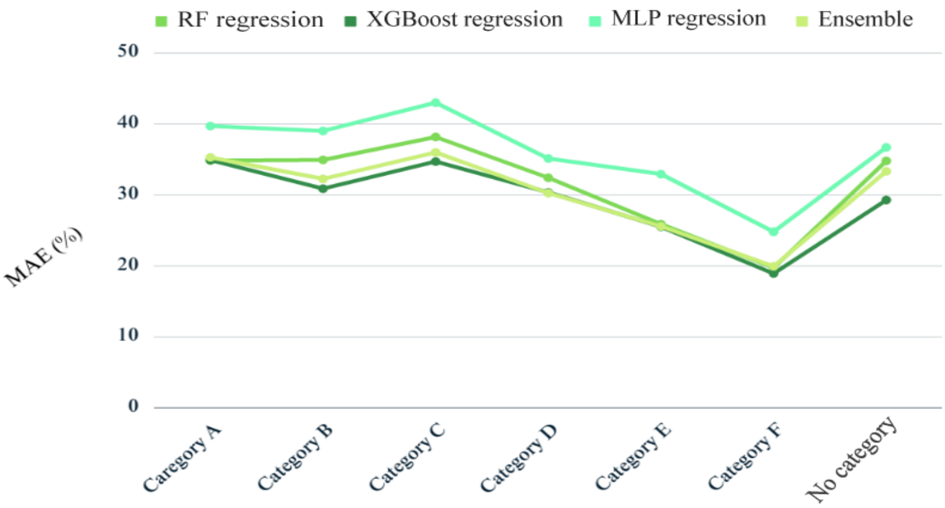

Figure 6. Graphical comparison of MAE of algorithms.

Figure 6 indicates the MAE of the five algorithms. This value is the highest for ridge and lasso regression in case of all categories. On the other hand, lowest error is found from XGBoost Algorithm.

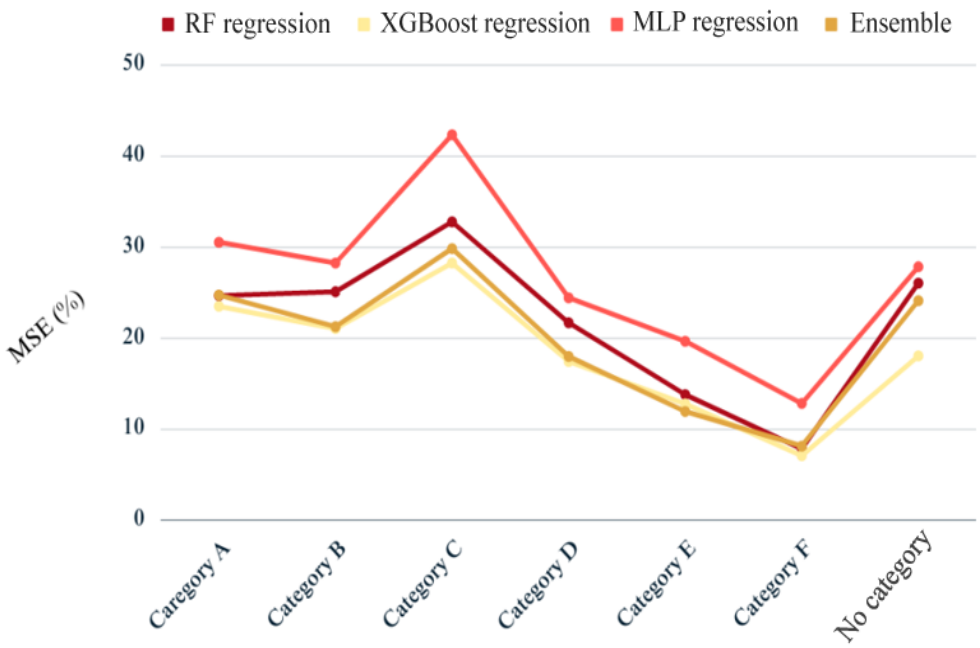

Figure 7. Graphical comparison of MSE of algorithms.

Figure 7 shows the MSE, whose value is significantly large for ridge and lasso regression in all categories. The XGBoost Algorithm, on the other hand, finds the lowest error, and category F has the considerably lowest value.

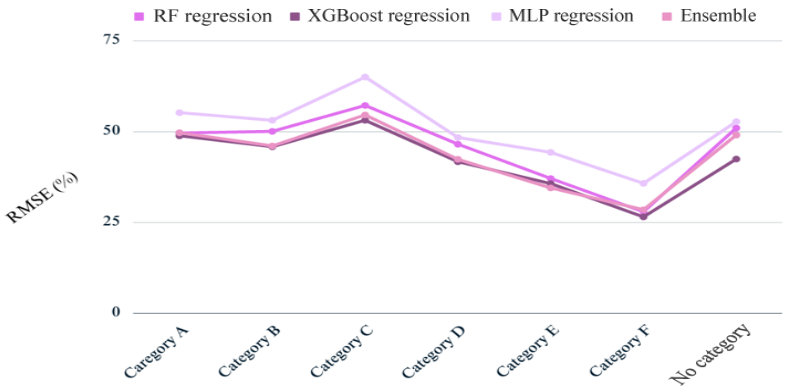

Figure 8. Graphical comparison of RMSE of algorithms.

In Figure 8, XGBoost once again demonstrates the lowest RMSE value. Other algorithms, however, show comparatively large errors.

4.2. Analysis of Performance

In this study, predicted wind speed estimates fluctuated. For the final prediction, we used the ensemble approach. Averaging is one of the ensemble approach's methods. Regression problems and classification problems may both utilize it to make predictions and calculate probabilities. The idea of ensemble learning is to integrate different models' outputs to get a prediction that is more accurate. This method offers robustness against data uncertainties in addition to improving accuracy. Ensemble learning has proven to be a strong technique in a variety of fields, providing more robust and trustworthy estimates by efficiently combining predictions from numerous models. We used the XGBoost Regression algorithm, which returned the highest accuracy in both with and without-category cases among all the algorithms. Here, we ensembled the output values of all algorithms to produce a better predictive performance. In Table 4, we have mentioned the results of XGBoost, RF, MLP, and Ensemble method.

Table 4. Accuracy and error value of different algorithms in each category.

Category	Algorithm	Accuracy	MAE	MSE	RMSE
Category A	RF regression	82.44%	34.86%	24.64%	49.64%
	XGBoost regression	83.27%	34.94%	23.47%	48.45%
	MLP regression	78.24%	39.72%	30.53%	55.25%
	Ensemble	82.39%	35.30%	24.70%	49.70%
Category B	RF regression	74.75%	34.95%	25.07%	50.07%
	XGBoost regression	78.85%	30.91%	21.01%	45.83%
	MLP regression	71.57%	39.03%	28.23%	53.13%
	Ensemble	78.63%	32.29%	21.22%	46.06%
Category C	RF regression	81.80%	38.18%	32.77%	57.24%
	XGBoost regression	84.33%	34.73%	28.23%	53.13%
	MLP regression	76.49%	43.01%	42.35%	65.07%
	Ensemble	83.44%	36.01%	29.83%	54.61%
Category D	RF regression	71.99%	32.44%	21.66%	46.54%
	XGBoost regression	77.48%	30.38%	17.41%	41.73%
	MLP regression	69.72%	35.15%	24.42%	48.39%
	Ensemble	76.77%	30.29%	17.96%	42.38%
Category E	RF regression	74.00%	25.91%	13.75%	37.08%
	XGBoost regression	75.93%	25.55%	12.72%	35.67%
	MLP regression	62.89%	32.96%	19.63%	44.30%
	Ensemble	77.51%	25.62%	11.89%	34.48%
Category F	RF regression	85.05%	19.78%	7.78%	27.89%
	XGBoost regression	86.50%	18.97%	7.02%	26.50%
	MLP regression	75.43%	24.84%	12.77%	35.74%
	Ensemble	84.43%	19.96%	8.09%	28.46%
No category	RF regression	75.88%	34.81%	26.02%	51.01%
	XGBoost regression	83.29%	29.30%	18.01%	42.44%
	MLP regression	74.21%	36.72%	27.82%	52.74%
	Ensemble	77.99%	33.35%	24.09%	49.08%

The table also includes information about the output of the three different errors. In the case of XGBoost Regression, the categories had the lowest error rate and highest accuracy. Since this result remains constant throughout the analysis irrespective of the stations, i.e., none of the other two algorithms surpassed the XGBoost, we can say that XGBoost Regression is the most suitable algorithm for forecasting wind speed at any location in Bangladesh without any variation. Category E showed different result compared to other categories. Although its highest accuracy is given by the XGBoost, the ensemble method gave a value greater than its highest accuracy level. The one reason behind this effect could be the closeness to the large river banks. The area near the bodies of water is always windy. Both land and water bodies heat up and cool down at different rates. Large bodies of water heat up considerably more slowly and are substantially cooler than nearby land masses, creating a high-pressure zone. As a result, air moves from the water to the land each morning, but this reverses in the afternoon, since the land cools faster than the water. Wind is caused by differences in air temperature; therefore, there is always a temperature differential over ocean and adjacent land. Category E consists of the stations that are near the big rivers, i.e., Padma and Jamuna, which might affect the wind behavior of those areas and deviate from the wind speed prediction. RF and MLP regression also have an impact on the ensemble value, which increases accuracy and outperforms XGBoost.

5. CONCLUSION

Wind energy has come a long way in the past few decades due to the advancements in modern turbine and generator technology, making it feasible to use it to help power large cities and even our own houses. While harnessing wind energy, it is crucial to discover a trustworthy approach to estimating wind speed. Using a dataset from Bangladesh that has 65 years' worth of meteorological data, we have worked on predicting wind speed. The Random Forest, XGBoost, and MLP Regression were used to forecast the wind speed. Our research has demonstrated that XGBoost is more accurate at producing predictions than the others, irrespective of the landforms mentioned in categories A, B, C, D, and F, whereas category E showed different average results as it's near the notable water bodies. The significant limitation of our work was that we tested our data sets using monthly average values, whereas if we had worked with daily data, we could have achieved better results. Hybrid methods boost prediction accuracy by combining methodologies and optimization strategies. Future wind speed prediction research may focus on increasing generality through data preprocessing, as well as deploying and validating forecasting models using performance assessment criteria. Most importantly, we could conduct more research near water bodies in Bangladesh.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Designed the research and prepared the manuscript, E.H.S.; conducted simulation and interpreted data and result, M.S.R.; prepared manuscript, T.A.; developed the research concept and provided overall guidance and supervision, S.M.U. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Afrasiabi, M., Mohammadi, M., Rastegar, M., & Afrasiabi, S. (2020). Advanced deep learning approach for probabilistic wind speed forecasting. *IEEE Transactions on Industrial Informatics*, *17*(1), 720-727. https://doi.org/10.1109/tii.2020.3004436

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: Modeling and performance evaluation. *Applied Sciences Basel, Switzerland*, *12*(4), 2126. https://doi.org/10.3390/app12042126

Azad, H. B., Mekhilef, S., & Ganapathy, V. G. (2014). Long-term wind speed forecasting and general pattern recognition using neural networks. *IEEE Transactions on Sustainable Energy*, *5*(2), 546-553. https://doi.org/10.1109/tste.2014.2300150

Bangladesh's First Large Wind Power Project Launched. (2021). *Gov.cn.* Retrieved from http://bd.china-embassy.gov.cn/eng/zmjw/202110/t20211019_9594482.htm. [Accessed December 12, 2022]

Brownlee, J. (2022). *When to use MLP, CNN, and RNN neural networks, machine learning mastery.* Retrieved from https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/

Cai, R., Xie, S., Wang, B., Yang, R., Xu, D., & He, Y. (2020). Wind speed forecasting based on extreme gradient boosting. *IEEE Access, 8*, 175063-175069. https://doi.org/10.1109/access.2020.3025967

Chaya. (2020). *Random forest regression, level up coding.* Retrieved from https://levelup.gitconnected.com/random-forest-regression-209c0f354c84. [Accessed December 12, 2022]

Dhiraj, K. (2019). *Implementing gradient boosting regression in Python. Paperspace Blog.* Retrieved from https://blog.paperspace.com/implementing-gradient-boosting-regression-python/. [Accessed December 12, 2022]

Filik, Ü. B., & Filik, T. (2017). Wind speed prediction using artificial neural networks based on multiple local measurements in Eskisehir. *Energy Procedia, 107*, 264-269. https://doi.org/10.1016/j.egypro.2016.12.147

Frost, J. (2022). *How to interpret R-squared in regression analysis, statistics by Jim* Retrieved from https://statisticsbyjim.com/regression/interpret-r-squared-regression/

Ghorbani, M., Khatibi, R., FazeliFard, M., Naghipour, L., & Makarynskyy, O. (2016). Short-term wind speed predictions with machine learning techniques. *Meteorology and Atmospheric Physics, 128*, 57-72. https://doi.org/10.1007/s00703-015-0398-9

Gupta, D., Natarajan, N., & Berlin, M. (2022). Short-term wind speed prediction using hybrid machine learning techniques. *Environmental Science and Pollution Research, 29*(34), 50909-50927. https://doi.org/10.1007/s11356-021-15221-6

Hess, L. (2021). *Wind energy could cut global warming by up to 0.8 degrees C, says report, #ThinkLandscape global landscapes forum*. Retrieved from https://thinklandscape.globallandscapesforum.org/55376/wind-energy-could-cut-global-warming-by-up-to-0-8-degrees-c-says-report/

Khosravi, A., Machado, L., & Nunes, R. (2018). Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. *Applied Energy, 224*, 550-566. https://doi.org/10.1016/j.apenergy.2018.05.043

Kulkarni, M. A., Patil, S., Rama, G., & Sen, P. (2008). Wind speed prediction using statistical regression and neural network. *Journal of Earth System Science, 117*, 457-463. https://doi.org/10.1007/s12040-008-0045-7

Kumar, A. (2023). *Ridge regression concepts & python example.* Retrieved from https://vitalflux.com/ridge-regression-concepts-python-example/

Lasso and Ridge Regression in Python Tutorial. (2022). *Datacamp.* Retrieved from https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression. [Accessed December 12, 2022]

Padhma, M. (2021). *A comprehensive introduction to evaluating regression models. Analytics Vidhya.* Retrieved from https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/. [Accessed December 12, 2022]

Reza, R. B. (2019). *65 years of weather data Bangladesh preprocessed.* Retrieved from https://www.kaggle.com/datasets/emonreza/65-years-of-weather-data-bangladesh-preprocessed/code

Salehin, A. S., Billah, M., Zahin, M. A., & Haque, M. S. (2022). *A machine learning approach for wind speed prediction*: Bangladesh University of Professionals.

Singh, G. (2019). *Random forest regression - data science bridge - medium. Data Science Bridge.* Retrieved from https://medium.com/data-science-bridge/random-forest-regression-ddfc88c92689. [Accessed December 12, 2022]

Zhu, Q., Chen, J., Zhu, L., Duan, X., & Liu, Y. (2018). Wind speed prediction with spatio–temporal correlation: A deep learning approach. *Energies, 11*(4), 705. https://doi.org/10.3390/en11040705

Zucatelli, P. J., Nascimento, E. G. S., Santos, A. Á. B., Arce, A. M. G., & Moreira, D. M. (2021). Study of the wind speed forecasting applying computational intelligence. *In Aerodynamics. IntechOpen.*