International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I7P208

Volume 69 Issue 7, 56-63, July, 2021

Development of Artificial Intelligent Techniques for Short-Term Wind Speed Forecasting

Amit Verma[1], K G Upadhyay[2], M M Tripathi[3]

[1]*Asst.Professor, Faculty of Technology, U.T.U.,Dehradun, India*
[2]*Professor, M.M.M. University of Technology, Gorakhpur, India*
[3]*Professor, Delhi Technological University, Delhi*

amit.v.ashu@gmail.com, kgueed@gmail.com, mmtripathi@dce.ac.in

Abstract – *The main aim of this research work is to present an assessment of machine learning algorithms for wind speed forecasting. Five machine learning algorithms considered for performance comparisons are such as Multiple Linear Regression, Polynomial Regression, Random Forest, Decision Tree, and Gradient Descent. The prediction is based on the data collected for Brussels in Belgium from open source for the period of 1 Jan 2020 to 30 Jun 2020. The overall best prediction results are obtained by the Gradient Descent algorithm to predict the wind speed for the next 300 minutes. The performance of Polynomial Regression is also found satisfactory.*

Keywords – *Data Preprocessing, Gradient Descent, Polynomial Regression, Mean Absolute Percentage Error (MAPE), Root mean Square Error (RMSE).*

I. INTRODUCTION

Renewable energy sources in general and wind energy is especially having grown significantly in the last decade, becoming an increasingly important component of the global supply of energy.

It is expected that global energy demand will continue to grow, even with the declared goal of increasing energy efficiency. Global warming continues to call for substantial carbon dioxide emissions reductions.

Different renewable sources of energy have been used for environmental purposes. Using different renewable resources has led to strong interest and growth. Wind, which can be turned into electricity, is safe, sustainable energy that has become very prevalent in recent times. New energy resources now take the market and deployment of wind power technology closer to conventional energy and have scope for competing with traditional power plants.

However, the biggest issue with wind is its erratic nature, which makes it tough to track wind generation production. This wind design contributes to an improvement in the complexity and efficiency of the current system control and reserve specifications (also known as auxiliary services on the electricity market). Because of wind fluctuations and instability, which make it difficult to monitor wind farm's output capacity, wind penetrators will increase their needs and reserves to ensure their stability and efficiency. A review of wind forecasting models is illustrated in [1]. A multistage wind forecasting model including artificial neural network (ANN) and support vector regression is developed and results compared with single-stage application [2]. Different ANN techniques have been developed for wind speed forecasting with minimum error [3]. A comparative analysis of wind speed forecasting techniques such as linear regression, autoregressive integrated moving average (ARIMA), ANN, and K- nearest neighbors (K-NN) is given in [4]. Polynomial Regression performs better than Ridge Regression, and ANN in terms of root mean square error (RMSE) and R-squared metrics for wind speed prediction [5]. Rainfall prediction for Sudan is made by utilizing fourteen algorithms in [6] with a different type of regression analysis. A comprehensive literature survey described the performance of Naïve Bayes, Random Forest, and SVM algorithms [7]. Forecasting of wind speed for altered heights by applying the multilayer perceptron(MLP) and adaptive neural fuzzy inference system (ANFIS) illustrated in [8]. Very short–term wind speed forecasting [9] is provided by the different machine learning models such as K-NN, ARIMA, and ANN. A short–term wind forecasting by data adjustment with the K-means clustering method included the data variability shown in [10].

The rest of the paper is structured as: Section II provides wind forecasting, which includes wind speed as well as wind power forecasting. Section III presents data preprocessing stages for data ready to use by the different algorithms. Section IV describes the evaluation criteria. Section V presents the regression analysis of wind speed with temperature and pressure. Section VI gives the performance of five applied machine learning algorithms. Results are discussed in section VII. Conclusions drawn from this work are highlighted in section VIII.

II. WIND FORECASTING

The forecast of wind speeds must be distinguished from the prediction of wind energy. Where, for weather forecasts and how people function in meteorology, the first is significant. Many methods for Numerical Weather Prediction(NWP) have now been developed and are used to estimate wind speeds at various locations, which is also the reason that anyone can access and use data from open source.

For wind speed forecasting, atmospheric temperature, humidity, and air pressure are the important factors contributing to wind speed creation. Of course, the main source of prediction is the atmospheric temperature, humidity, and air pressure, as the purpose of the system is to predict the feedback of the predictive tool, the comprehensive input we will get from our predictive tool.

The wind speed is a highly variable non-linear phenomenon, making it very difficult to forecast the use of a standard approach. The method of overcoming a non-linear problem is the way to use the intelligent engineering described by the networks of the neurons, a genetic algorithm, and a chaos fractal, etc.

It's difficult to estimate because the majority of positions on wind farms are not actually in the same place in the weather station (this is referred to as the horizontal interpolation "from of the grid points to the co-ordinate of the turbine") because the altitude of the wind turbine is not identical as the weather station (this is known as vertical interpolation). Estimation techniques are also available to transmit measured data to the interest location (or even the projected data from the NWP).

The Numerical Wind Prediction's output is translated into a wind turbine (farm) by two separate methods

1- Statistical Method

2- Physical Method

A. Statistical Method

The statistical method relies on the training for the available measuring dataset. It was intended that there be a statistical relationship between the given weather input and the calculated wind turbine power output (farms). Consequently, these systems rely entirely on data processing that ignores the weather information. Artificial neural networks (ANN) can be employed for these systems. The system operates on artificial neural networks(ANN) that train the past wind farm dataset available or calculate the transformer stations where a range of wind farms are interlinked to each other. An online calculation of the actual wind power input into the grid is also generated by the system, based on extrapolation measurements at representative wind power farms.

The benefit of the statistical method is that forecasts are tailored automatically to the wind farm's position to eliminate structural errors. The downside is that long-term measurement data are required, and additional training effort is needed. Also, these systems can't determine accurately when unusual atmospheric conditions become too unusual during the training process. Unfortunately, it is very important to correctly predict these unusual situations, and else it can lead to large forecasts.

B. Physical Method

The physical method utilizes parameters depending upon the precise physical descriptions of the lower atmosphere. This basic equation can be used to calculate a wind turbine's power yield:

$$P_{total}=\frac{1}{2} C_p \rho \pi R^2 v^{3}$$

For this purpose, we need to collect data on the air density and wind speed measured in the studied region, as well as know the swept segment of the selected wind turbine. While the energy output per month/year is the multiplication of power with time:

$$E=P \times t$$

And it is necessary to know that the wind speeds simultaneously in the future to measure the output power of wind turbines over a given period of the span. We need a prediction tool to provide us with the necessary details (like NWP). This will also help to control the energy market, where the energy supplier is supplied one day before the coordination center that handles the data and chooses energy prices for the following day. In this tool, we use the usual energy measurement methods, but note that we should take poor losses and other parameters into consideration when talking about a wind farm.

III. DATA PREPROCESSING STAGES

A. Data Collection

The data is collected from the open-source website timeanddate.com for the Brussels region of Belgium. The data covered in this work is in between the period of 01 Jan 2020 to 30 Jun 2020, which include 7 parameters such that time, temperature, weather wind speed, humidity, pressure, and visibility in column-wise and 8323 rows. It is required to clean the data because it contains a lot of unnecessary information

B. Data Cleaning

a) Removing Missing Data and Handle Nan Values

At this stage, a consistent for the whole data model is developed, which searches missing data (Nan values), finding out duplicate data, and weeding the bad data. The collected data are month wise which has been taken in a single file. After checking the N/A values, it is found that weather, humidity, and visibility have 43, 19, and 4283 N/A values. If we ignore all these N/A values, then definitely our model will be affected. A heat map is also plotted to clearly visualize N/A values, in which the visibility column contains approximately 52% N/A values, it is very difficult to fill the N/A values, so we drop this column. The heat map is shown in Fig 1.Weather column contains categorical values in which a lot of different climates. It is not possible to encode each type of climate, so we remove this column and keep on preprocessing the left data. Humidity column comprises only 19 N/A values, so we fill the N/A values by the mean. Some of the rows consist of NO WIND in the wind speed column, so we replace it with 0 Km/hr. After completing the data cleaning, a new csv (comma-separated value) file is created, and all the clean data is stored in it.

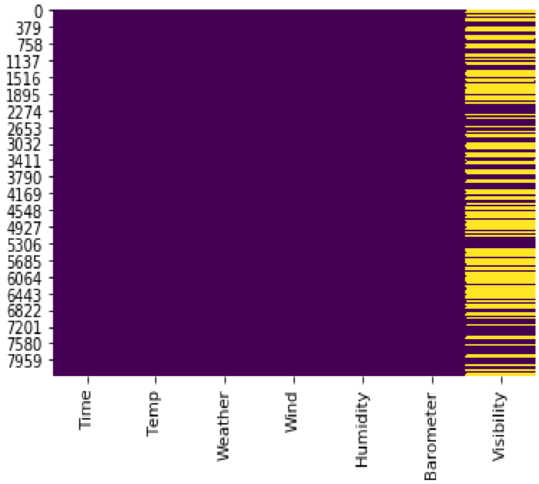

Fig 1: Heat map showing Nan values of the seven attributes

b) Outlier Treatment

Outliers are the observations that diverge from the overall pattern in a data sample. In other words, they are not useful in the data set for statistical analysis. The five most popular methods to detect outliers within the data set are Standard Deviation, Box Plot, DBScan(Density-based spatial clustering of applications with noise), Isolation Forest, and Robust Cut Forest. In this paper, box plots have been plotted with the help of Python library Scipy along with Pandas and Numpyto remove outliers. Box plots are a graphical representation of any data set through their quintiles. Lower and upper whiskers are the boundary of the whole data set. Any data points outside these whiskers are the outliers. The Figure shows a box plot of temperature, pressure, and wind data distributions in which it has been clearly seen that some data points are out of the boundaries called whiskers. These are outliers here, and these data are removed. The left data are stored in a new csv file.

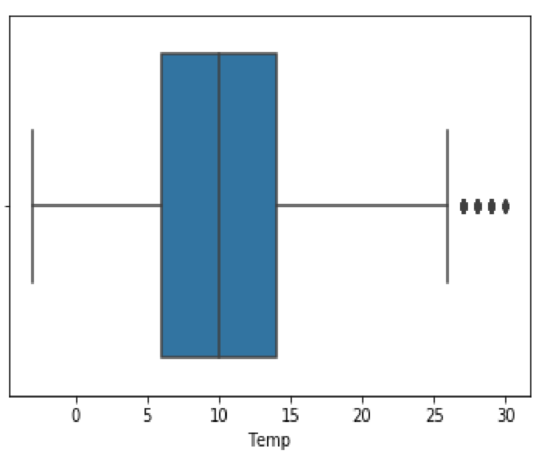

Fig 2: Box-plot removing outlier of temperature

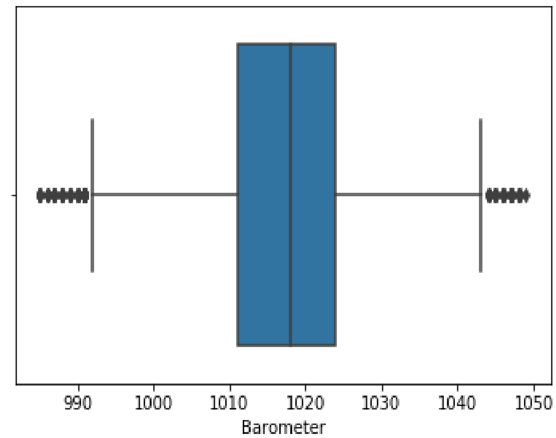

Fig 3: Box-plot removing outlier of pressure

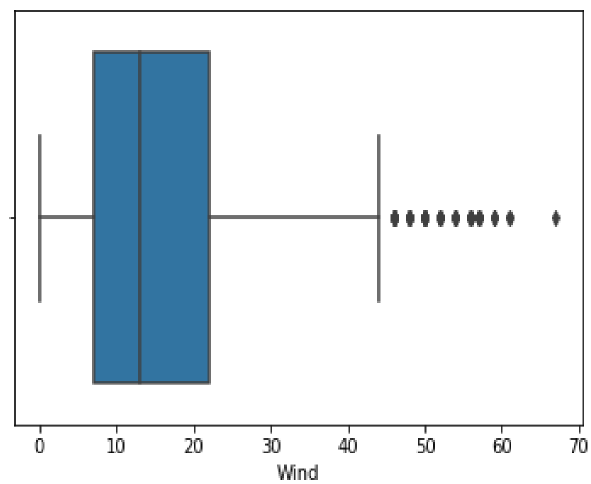

Fig 4Box-plot removing outlier of wind speed

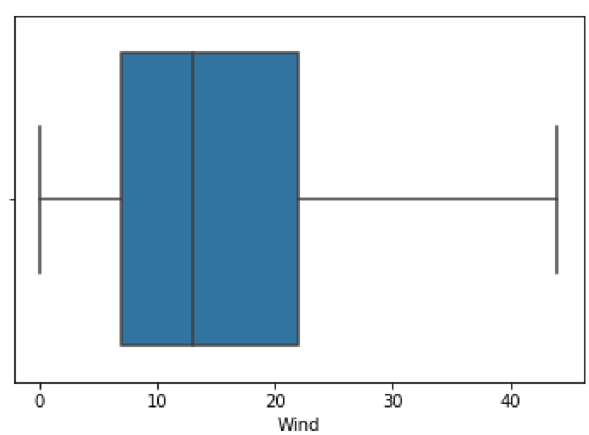

Fig 5: Box-plot with a removed outlier in wind speed

C. Data Selection

At this stage, the data for the analysis of linear regression is decided. Wind speed depends on three major attributes temperature, pressure, and humidity. At this stage, some regression plots and heat maps are drawn for a clear understanding of the data set.

The heat map shown in the Fig indicates the correlation between the attributes in which temperature and pressure have a negative correlation with the wind speed.

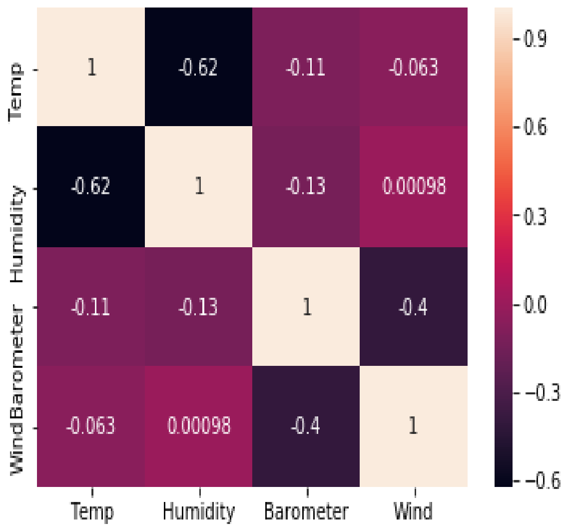

Fig 6: Correlation Heat map

D. Data Normalization

It is the process in which the data obtained from the preceding section has to be transformed into the appropriate format for simple computing. Min Max Scalar method is applied for the data normalization, and this task is performed by the Python sklearn library. The mathematical formula used for data scaling is given in the equation below, where N_x is the normalized value of the dataX_d, X_{min} and X_{max} Are the minimum and maximum values in the data set?

$$N_x = \frac{X_d - X_{min}}{X_{max} - X_{min}}$$

After normalization, data is stored in a new csv file and kept for the next stage.

E. Data Mining

This is the last stage of data preprocessing in which the data set is sectionalized into two parts as training and testing set in accordance to 80% and 20% ratio. Now the data set is ready to use by five supervised machine learning algorithms such as Multiple Linear Regression, Polynomial Regression, Decision Tree, Random Forest, and Gradient Decent.

IV. EVALUATION FACTORS

The Root-Mean-Square Error (RMSE) or Root-Mean-Square Deviation (RMSD) is also used by the model or the estimator to denote the differences between the forecasted values and observed values. When the measurements are carried out over the data set used for estimation, these individual variations are called residuals and are called prediction errors when measured out-of-sample. The RMSE serves to integrate into a single metric of predictive power the magnitudes of the errors in forecasts for different periods. The RMSE is the square root of the variance for an impartial estimator, referred to as the standard error. The RMSE of the predicted value Ft for times t of the actual dependent variable At of regression is determined as the square root of the average of deviation square for n distinct predictions:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(At - Ft)^2}{n}}$$

The Mean Absolute Percentage Error (MAPE) is a measure of precision in the value of the time sequence equipped. Absolute values of all the errors are measured, and then its mean is calculated. It reveals accuracy as a percentage and has the same units as the initial data, which will take value from zero to infinity. This is why several researchers have used it in large part in wind speed forecasts. This study also uses MAPE to estimate wind speeds because of its precision, where At is the actual or real value and Ft is the forecasted value.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{At - Ft}{At}\right|$$

V. REGRESSION ANALYSIS

A Scatter plot between temperature and wind speed is plotted, and the best-fitted regression line is drawn with the Python programming as shown in Fig. A regression line is also drawn on the data plot of pressure and wind speed, and each time RMSE is calculated. These regressions Ines are in such a way that error from each point should be minimum. Temperature shows a good correlation to the wind speed than the pressure. Pressure shows a negative correlation with wind speed as the slope of its regression line is negative.

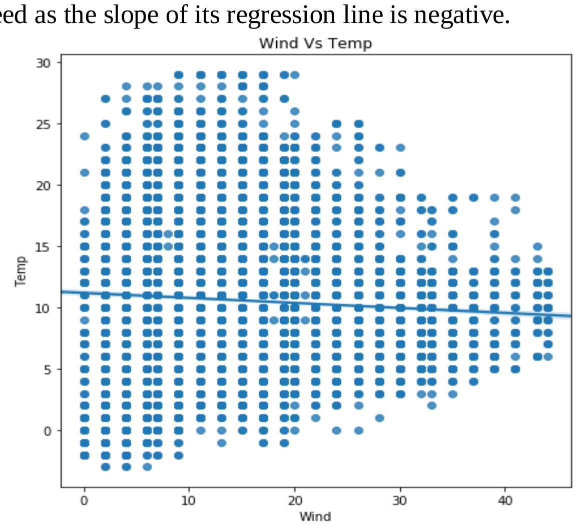

Fig 7: Regression plot of temperature vs. wind speed

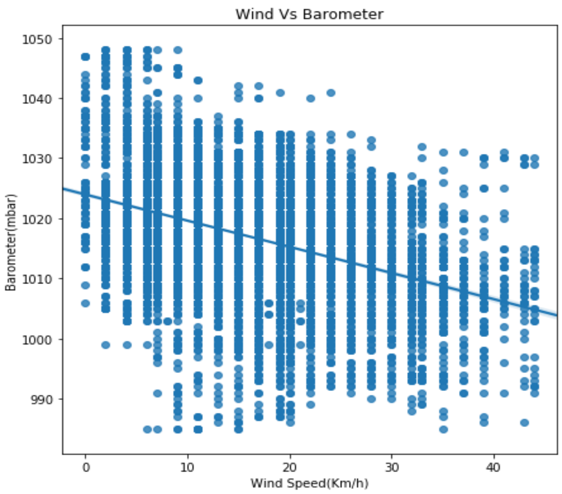

Fig 8:Regression plot of pressure vs. wind speed

VI. PERFORMANCE OF SOFTCOMPUTING ALGORITHM

A. Multiple Linear Regressions

Regression analysis is mostly used for the prediction of the dependent variable. Linear regression is frequently applied among many statistical models. When there is more than one independent variable is considered, then a multiple linear regression algorithm is applied. It comes under the category of supervised learning technique in which a single line is generated beside numerous lines with deferent y-intercept values and slopes. Each time error is calculated from all the data points and improved as close to the target, and finally, a regression line is developed. This process is performed with the help of Python library scikit-learn, and the "linear. fit(x,y)" function has been applied. Multiple linear regression given in the form as

$$y = m_0 + m_1x_1 + m_2x_2 + \ldots\ldots.. m_nx_n + \varepsilon$$

Where $m_0, m_1, m_2, \ldots.. m_n$ are the regression coefficients, $x_1, x_2, x_3 \ldots. x_n$ are dependent variables and ε as the error.

B. Polynomial Regression

When the data sets are not in the form to be fitted by a single straight line, then for getting good results, polynomial regression should be applied. It is a special case of multiple linear regressions. In this paper, second-order polynomial regression is applied.

The mathematical form of this algorithm with three variable is given as

$$y_p = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3^2 + \varepsilon$$

Where interpretation of parameter α_0 is at $x_1, x_2, x_3 = 0$ with zero error, coefficients α_1 and α_2 are the linear effect parameter, α_3 is the quadratic effect parameter, and ε is the error.

C. Decision Tree

A decision tree algorithm is a supervised learning technique in which the node is the root that utilizes to test any decision. The outcomes represent branches, and the leaf denotes the target. The source data set is divided into subsets according to the different attributes. This is the recursive process continues until we get no variation in results. Decision tree is also designated as categorization of data sets by various soft computing methods. A most popular term in this context is known as classification, and regression tree (CART), in which classification denotes the category and regression comprises numerical data. Some specific algorithms used as decision trees are known as Iterative dichotomiser 3 (ID3), CART, multivariate adaptive regression splines (MARS), C4.5, etc. In this paper, a decision tree algorithm is utilized to predict wind speed, and the error is computed in the form of RMSE and MAPE.

D. Random Forest

Random forest is an ensemble supervised learning method which combines various decision trees to enhance performance. The main point is to decide the number of decision trees with the maximum depth. Target has to be fixed by representing the leaf node. We have to prepare the model in such a way that it should not be over-fitted or under–fitted. Random forest algorithm is able to control the over-fitting problem. The function "RandomForestRegressor()" is applied in Python. Estimator range is selected between 10 to 200 with an interval of 10, and a plot score vs. n estimator plotted as shown in Fig. The R-squarer score is computed as 0.1795.

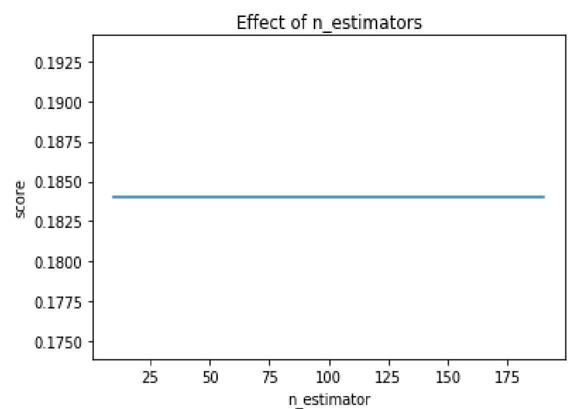

Fig 9: Plot of score versus estimator

E. Gradient Descent

Gradient descent is a machine learning algorithm used for the optimization of the coefficient of a function. In this algorithm, we have to find out global minima with a learning rate of steps not too small or not too big. The slope is minimum wherever the partial derivative of the function with respect to the slope has to attain zero value. A curve between cost versus iterations is plotted in Fig.

Fig 10: Plot of cost versus iterations

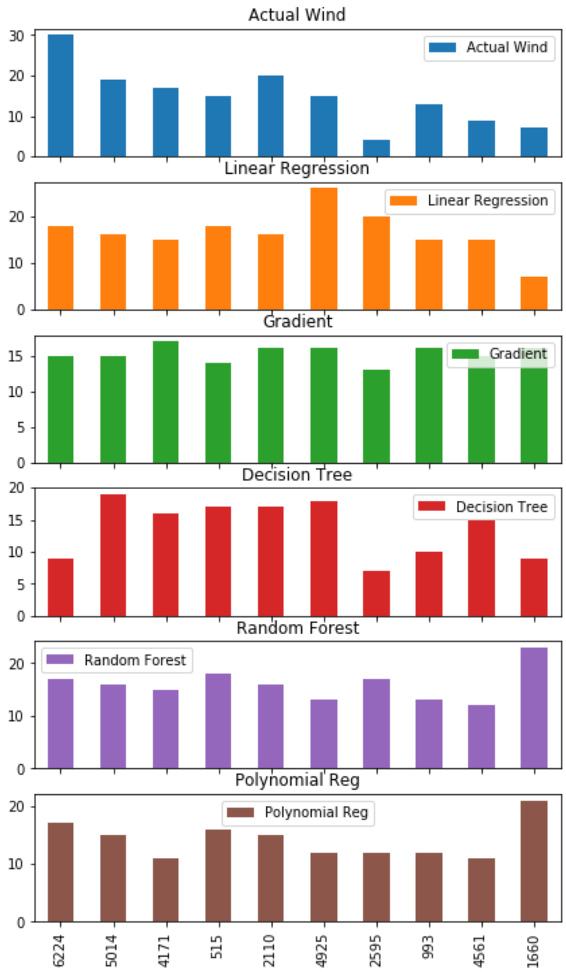

Fig 11: Actual wind speed and the predicted wind speed while applying five algorithms

VII. RESULTS AND DISCUSSION

The above discussed five algorithms are applied to predict the wind speed of the Brussels region of Belgium. These algorithms also compute the RMSE and MAPE. Fig. shows the actual wind speed and the predicted wind speed by the algorithms such as Linear Regression, Gradient Descent, Decision Tree, Random Forest, and Polynomial Regression at different random values of the data set.

These results give the empirical comparison of performance between the five applied algorithms. The wind speed predicted by Gradient Descent and Polynomial regression is very near to the actual wind speed, while wind speed obtained by Decision Tree, Linear Regression, and Random forest is somehow far to the actual wind speed. Fig. 12 shows the comparison of actual to forecasted wind speed while applying all five algorithms with respect to the time in minutes. A comparison of RMSE and MAPE computed by these algorithms is presented in the table1 and table2 on the priority basis of RMSE and MAPE, respectively.

VIII. Conclusion

This work estimates the wind speed in the Brussels region of Belgium by applying five machine learning techniques as Gradient Descent, Random Forecast, Multiple Linear Regression, Polynomial Regression and Decision Tree. Performance comparison is made on the basis of RMSE and MAPE. It depicted that prediction by Linear Regression is best when compared on the basiss of RMSE. While Gradient Descent shows best performance on the basis of MAPE. Overall we can conclude that Gradient Descent prediction shows better among all the five applied algorithms. Wind speed forecasting for next 300minutes is successfully achieved by applying the five machine learning algorithms.

Table1:Comparison Chart on the basis of RMSE

Rank	Applied Algorithm	RMSE	MAPE
1	Linear Regression	8.3855	17.437
2	Polynomial Regression	8.5825	21.21
3	Random Forest	8.75	16.025
4	Gradient Decent	9.8428	0.783
5	Decision Tree	55.2727	49.363

Table 2: Comparison Chart on the basis of MAPE

Rank	Applied Algorithm	MAPE	RMSE
1	Gradient Decent	0.783	9.8428
2	Random Forest	16.025	8.75
3	Linear Regression	17.437	8.3855
4	Polynomial Regression	21.21	8.5825
5	Decision Tree	49.363	55.2727

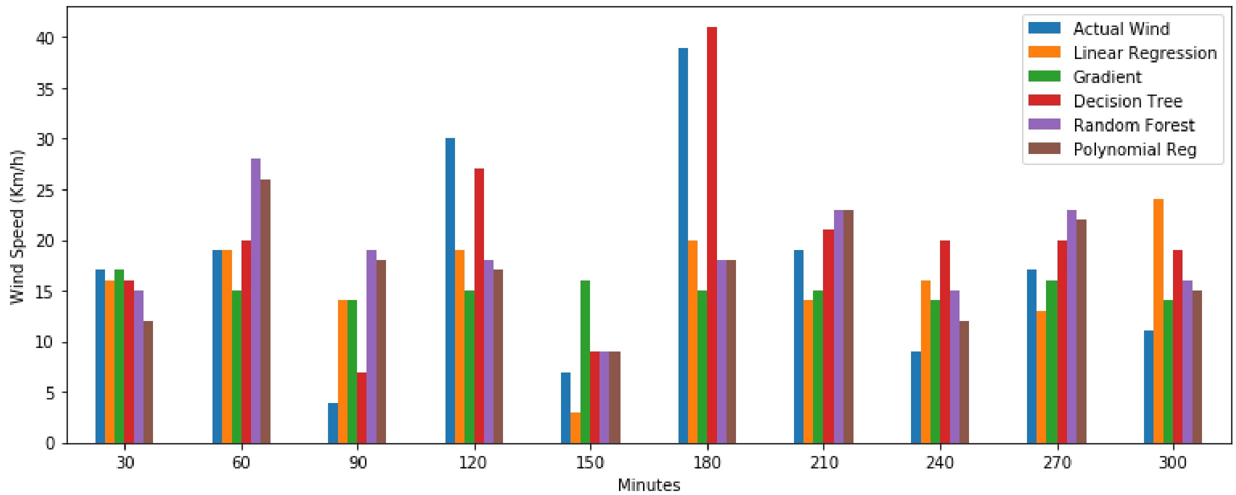

Fig 12: Wind speed forecasting for 300 minutes by applying the five machine learning algorithms

REFERENCES

[1] Xiaochen Wang, PengGuo and XiaobinHaung,., A review of wind power forecasting model, Energy Proceedia, 12 (2011) 770-778

[2] .SerkanBuhan and IsikCadirici., Multistage wind - electric power forecast by using a combination of advanced ststistical methods, IEEE Transactions on Industrial Informatics, 11(5) (2015) 1231-1242.

[3] TarlochanKaur, Sanjay Kumar and Ravi Segal., Application of artificial neural network for short-term wind speed forecasting, Biennial International conference on Power and Energy Systems: Towards Sustainable Energy,(2016).

[4] ArchitaVijavargia, Kailash Chand Sharma and RohitBhakar., A comparative study of short term wind speed forecasting models, 2nd International Conference on Large Scale Grid Integration of Renewable Energy in India, (2019).

[5] Ahmad FerdousAntor and Ebisa D. Wollega., Comparision of machine learning algorithms for wind speed prediction, Proceedings of 5th NA International Conference on Industrial Engineering and operations management, (2020) 857-866.

[6] Nazeem Osman Bushara and Ajith Abraham., Weather forecasting in Sudan using machine learning schemes, J. of Network & Innovative Comput., 2 (2014) 309-317.

[7] Jay Kumar Sadhasivam, ArpitRathore, Indrajeet Bose, SaumyaBhattacherjee and SenthilJayavel., A survey of machine learning algorithms, International J. Engg. Trend &Technology,68(4) (2020) 64-71.

[8] EmrahKorkamaZ, ErcanIzgi and SalihTutun., Forecasting of short term wind speed at different heights using a comparative forecasting approach, Turkish J. of Elect. Engg& comp. Sc., (2018) 2553-2569.

[9] M. C. Alexiadis, P. S. Dokapoulas, H. S. Sahasamanoglou and I. M. Manausaridis, Short term forecasting of wind speed and related electrical power, Solar Energy, 63(1) (1998) 61-68.

[10] QianyaoXu, Dawei He, Ning Zhang, Chongqing Kang, Qing Xia, JianhuaBai, and Junhui Huang., A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining, IEEE Trans. on Sust. Energy, 6(4) (2015) 1283-1291.

[11] Mingjian Cui, DepingKe, Yuanzhang Sun,Di Gan,Jie Zhang, and Bri-Mathias Hodge., Wind power ramp event forecasting using a stochastic scenario generation method,IEEE Trans. on Sust. Energy, 6(2) (2015) 422-433.

[12] S. S. Chandel, P. Ramasamy and K. S. R. Murthy., Wind power potential assessment of 12 locations in western Himalayan region of India, Renw. andSust. Energy Rev., 39 (2014) 530-545.

[13] NicoleneBhotha and Christiaan M. van der Walt., A comparision of regression algorithms for wind speed forecasting at Alexander Bay, Pattern recognition association of South Africa and robotics and mechatronics international conference (PRASA-RobMech) (2016).

[14] MahmetBerisOzkan and Pinar karagoz., A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP), IEEE Trans. on industrial Informatics,11(2) (2015) 375-387.

[15] Jethro Dowell and Pierre Pinson., Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression, IEEE Trans. on Smart Grid, 7(2) (2016).

[16] RonayAk, Valeria Vitelli, and Enrico Zio., An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction, IEEE Trans. on Neural Networks and Learning Sys., 26(11) (2015) 2787-2800.

[17] Kristof De Vos and Johan Driesen, Active participation of wind power in operating reserves, IET Renewable power generation, 9(6) (2015) 566–575.

[18] Anna Hellmers, Marco Zugno, Anders Skajaa, and Juan M. Morales., Operational Strategies for a Portfolio of Wind Farms and CHP Plants in a Two-Price Balancing Market IEEE Trans. Power Syst, (2016).

[19] Yimu Fu, Mingbo Liu and Licheng Li., Multiobjective Stochastic Economic Dispatch With Variable WindGeneration Using Scenario-Based Decomposition and Asynchronous Block Iteration, IEEE Trans. on Sust. Energy, (2015) 1949-3029.

[20] Yaowen Yu, Peter B. Luh, Eugene Litvinov, TongxinZheng, Jinye Zhao, Feng Zhao., Grid Integration of Distributed wind generation: Hybrid Markovian and Interval Unit Committmen, IEEE Trans. on Smart Grid, 6(6) (2015) 3061-3072.

[21] KacemGairaa, Farouk Chellali, Said Benkaciali, YoucefMesslem and KhellafAbdallah., Daily global solar radiation over a desert area using NAR neural networks comparision with conventional methods, 4th International Conference on Renewable Energy Research and Applications, (2015) 567-571.
[22] RosnaliniMansoor, BahtiarJamiliZaini Catherine Chan May May, Medium term electricity load demand forecasting using Holt-Winter exponential smoothing and SARIMA in university campus, Int. J. of Engg Trend & Tech., 69(3) (2021) 165-171.
[23] A. K. Chaudhary, K. G. Upadhyay and M. M. Tripathi, Estimation of wind power using different soft computing methods, Int. J. of Electr. Sys., 1(1) (2011) 1-7.
[24] Doan Duc Tung and Tuan-Ho Le, S statistical analysis of short-term wind power forecasting error distribution, Int. J. Aplld. Engg. Research, 12(10) (2017) 2306-2311.
[25] AnishTalwar and Yogesh Kumar., Machine learning: An artificial intelligence methodology, Int. J. Engg& Comp. Sc., 2(12) (2013) 3400-3404.
[26] Nathaniel S. Pearre and Lukas G. Swan., Statistical approach for improved wind speed forecasting for wind power production, Sust. Energy Tech. & Asses., 27 (2018) 180-191.