Prediction of Wind Speed using Machine Learning

Nabanita Mandal
Computer Engineering Department
Thadomal Shahani Engineering College
Affiliated to University of Mumbai
Bandra, Mumbai, India

Tanuja Sarode
Computer Engineering Department
Thadomal Shahani Engineering College
Affiliated to University of Mumbai
Bandra, Mumbai, India

ABSTRACT

Climate is the average state of the atmosphere and related components of earth system. Latitude, distance from sea, presence or absence of mountains or other geographic information determines the climate of a place. Climate models are used to study the behavior, components and interactions of the climate systems. Climate data includes historical as well as real-time data. Machine Learning (ML) is the subset of Artificial Intelligence (AI) which has the ability to learn from the training data. This learning experience improves which helps in predicting future values. Mean Wind Speed (MWS) is one important characteristic which influences the climate. Prediction of mean wind speed using ML algorithms is described in this paper. It is useful for determining abnormal weather events and also the future potential for wind energy.

General Terms

Linear Regression, Machine Learning, Deep Neural Network

Keywords

Climate Prediction, Wind Speed, Random Forest

1. INTRODUCTION

The world's climates is classified on the basis of Koppen Climate Classification System . There are mainly five types denoted by capital letters: A-Tropical Moist Climates, B - Dry Climates, C - Moist Mid-latitude Climates, D - Moist Mid-Latitude Climates, E - Polar Climates. The climate of a region is decided by a number of factors like latitude and its influence on solar radiation received, Air mass influences, location of global high and low pressure zones, heat exchange from ocean currents, distribution of mountain barriers, distribution of land and sea and altitude, pattern of prevailing winds [1]. Since the wind plays an important role in deciding the type of climate, so an attempt as been made to predict the mean wind speed. Wind Speed is defined as how rapidly the wind is moving at a certain point. It is calculated as an average over a given unit of time. Monitoring wind speed is important for determining climate patterns. Direction and speed of wind helps in deciding weather patterns also. Both these parameters influence the rate of evaporation which in turn influences the water level [2]. Direction of wind is measured in degrees whereas speed of wind is measured in miles per hour. Anemometer is used to measure wind speed and wind vane is another instrument which is used to measure the direction. Both the parameters are used to detect storms, dust storms, hurricanes.

There exists a correlation between these two parameters [3]. Prediction of wind speed not only helps to detect storms but also helps in determining the potential for wind energy at that place. This can be helpful for power generation through wind turbines [4]. Applying Artificial Intelligence and Machine Learning techniques help us to analyze the current situation,

predict the future events and helps in creating new solutions. Factors affecting wind speed and direction are: temperature, air pressure, centripetal acceleration and earth's rotation. Temperature of a place varies between day and night. Air masses also vary with temperature due to which warm and cold winds are created. Wind blows horizontally from high to low pressure regions. The rate of air pressure change between the pressure regions helps in determining the wind speed. The speed is more if this difference is more. Air speed increases due to the centripetal force. It also influences the direction of wind. Changes in direction of wind happen because of the rotation of earth on its axis [5].

This paper shows how effectively machine learning and deep learning techniques can be applied for mean wind speed. Section II gives a brief overview of the existing techniques. Section III and IV consists of the proposed system and results respectively.

2. EXISTING TECHNIQUES

John V Neuman in the year proposed the Meteorology Project for developing numerical weather prediction (NWP). The team developed a Binary Electronic Sequence Calculator (BESK) for real time barotropic forecasts. But the duration of forecast was less because the computers didn't have the capacity to store longer forecast data. Later on capacity of computer was increased. Bert Bolin used this which resulted in extended forecast to cover a large area and also increase in the forecast duration. Analysis of numerical weather map was done after a few years [6]. Stanley L Barnes developed a method for weighted average interpolation scheme which is used to obtain desired amount of detail in the analysis of randomly spaced data [7]. A.C Lorenc has proposed a Better NWP which uses Bayesian Probabilistic Arguments [8]. Acosta, G., & Tosini, M. proposed a firmware digital neural network for climate prediction application. For predicting climate variables in a bounded environment, artificial neural network (ANN) was implemented in a Field Programmable Gate Array (FPGA) [9]. The variables like temperature, soil humidity were kept under control. ANN was helpful to predict the evolution of these variables in a temporal horizon. Wang Deji, Xu Bo proposed application of SVM to reconstruct climate variables from noisy environment [10]. Claire Monteleoni et al have identified goal of climate informatics. The aim is to inspire collaboration between climate scientists and data scientists [11]. Sonkusare, B., et al developed a Multi Model Ensemble approach based on ANN to correct the bias for the dynamical model generated forecast [12]. Huntingford, C et al proposed Machine Learning Techniques for identifying links between location, times and quantitites in the datasets[13]. Eleftheriou A. et all have proposed linear regression for forecasting wind speed which depicts the advantage of stochastic gradient descent algorithm over gradient descent algorithm [14]. Senthil Kumar P has

described different artificial neural network techniques for prediction of wind speed [15]. Gong Li and Jing Shi have compared three different artificial neural network techniques for wind speed forecast [16].

3. PROPOSED SYSTEM

Place The proposed system consists of the input data which is obtained from India Meteorological Department, Pune, India. It consists of parameters like sea level pressure, temperature, relative humidity, wind directions etc. Fig.1 shows the basic blocks of the proposed system
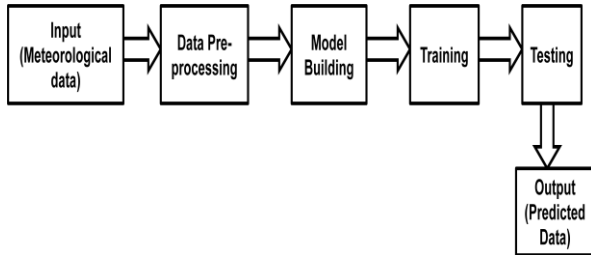

Figure 1: Block Diagram

The proposed system consists of six blocks. Data collection, pre-processing, model building are all different steps whereas splitting the data into training and testing is considered to be a single step. the next step is to do the prediction which is the output block. The first block is the Meteorological Data. This is raw data which needs to be pre-processed before using it. So, this step includes removal of missing values if any, scaling it to a standard form. The next step is to build a model. The three models that are used are: Linear Regression, Random Forest and Deep Neural Network. The data set has been split into training and testing in the next step. Predictions are obtained which are then matched against the actual values to calculate the error.

The problem of predicting is a regression problem. So, the first technique applied is Linear Regression.

A. Linear Regression: The requirement is to have a relationship between the dependent and independent variables. In this technique a straight line is fitted on the data. It assumes normal distribution of the dependent variable. The dependent variable is the Mean Wind Speed. The independent variables are temperature, pressure, wind direction, precipitation. All values of these independent variables are associated with the dependent variable. Since, the dependent variable is influenced by a number of independent variables so multivariate regression will be applicable here instead of univariate regression [17].

It is formulated as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... \beta_n x_n + \varepsilon \qquad (1)$$

y is the dependent variable, β_i is the parameter, x_i is the independent variable, ε is the error

The error is obtained by the difference between actual and predicted value [18]. The second technique is Random Forest.

B. Random Forest: In terms of training the model, it is simple. It is easier to find a good robust model using this technique. As the number of trees in the forest grows, the complexity also increases. Random forests are a mixture of tree predictors such that each tree depends on the values of a random vector sampled autonomously and with the same distribution for all trees in the forest [19].

The third technique is Deep Neural Network. A subset of Machine Learning is Deep learning, which involves creation of a neural network by using a number neurons and connections between them.

C. Deep Neural Network: Deep Neural Network are very much useful for prediction. It gives a good result because it consists of a number of hidden units and it learns by continuous training [20]. Information is passed on from each layer to the other. Weights are assigned and the final result is the sum of the weights.
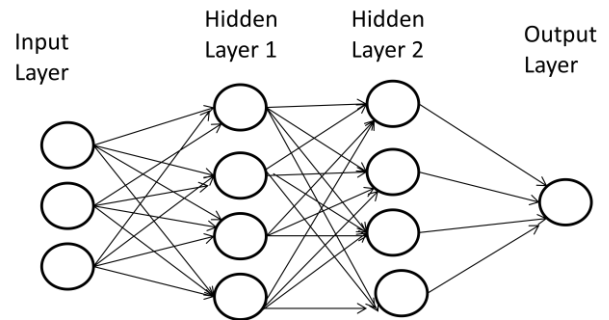

Figure 2: Deep Neural Network

Fig.2 gives a representation of a deep neural network which is fully connected.

4. RESULTS

The number of training samples is 216. The input data consists of features like temperature, direction of wind, precipitation, sea level pressure. The results are obtained by considering 90% training and 10% testing data split, 80% training and 20% testing data split and 70% training and 30% testing data split. The number of epochs are 500.

The difference between actual and predicted are values are calculated using: Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2 \qquad (2)$$

$$RMSE = \sqrt{\frac{i}{n}\sum_{i=1}^{n}(y_i - y'_i)^2} \qquad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y'_i| \qquad (4)$$

The Formulae for MSE,RMSE and MAE is given by equation (2), (3) and (4) respectively
For Training-Testing Split of 70:30, the results are shown in Table 1.

Table 1. Performance evaluation at 70:30

Algorithms	MSE	RMSE	MAE
Linear Regression	3.032244	1.741334	1.429564
Random Forest	0.137142	0.370327	0.283060
Deep Neural Network	2.069286	1.438501	1.052631

Here, only 70 percent data is given for training and rest 30 per cent is kept for testing. The error obtained is more in case of Linear Regression.

For Training-Testing Split of 80:20, the results are shown in Table 2

Table 2. Performance evaluation at 80:20

Algorithms	MSE	RMSE	MAE
Linear Regression	2.246709	1.498902	1.228233
Random Forest	0.105940	0.325484	0.251742
Deep Neural Network	1.339082	1.157187	0.901564

In this, 80 percent data is given for training and 20 per cent is given for testing. It is observed that the error reduces as compared to the previous training testing split.

In Machine Learning and Deep Learning Techniques, the model learns by itself using the data that is fed into it [10-19]. So, if the training data is more, the model performance is better. When the training data is increased to 90 percent, there is no significant I mprovement in model performance. The results are shown in Table 3.

This is because the model starts learning too well. The model suffers from overfitting problem in this case.

For Training-Testing Split of 90:10, the results are shown in Table 3.

Table 3. Performance evaluation at 90:10

Algorithms	MSE	RMSE	MAE
Linear Regression	2.706346	1.645097	1.328296
Random Forest	0.116405	0.341182	0.260059
Deep Neural Network	1.458042	1.207494	0.947262

	Linear regression	Random forest	Deep Neural Network
70:30			
80:20			
90:10			
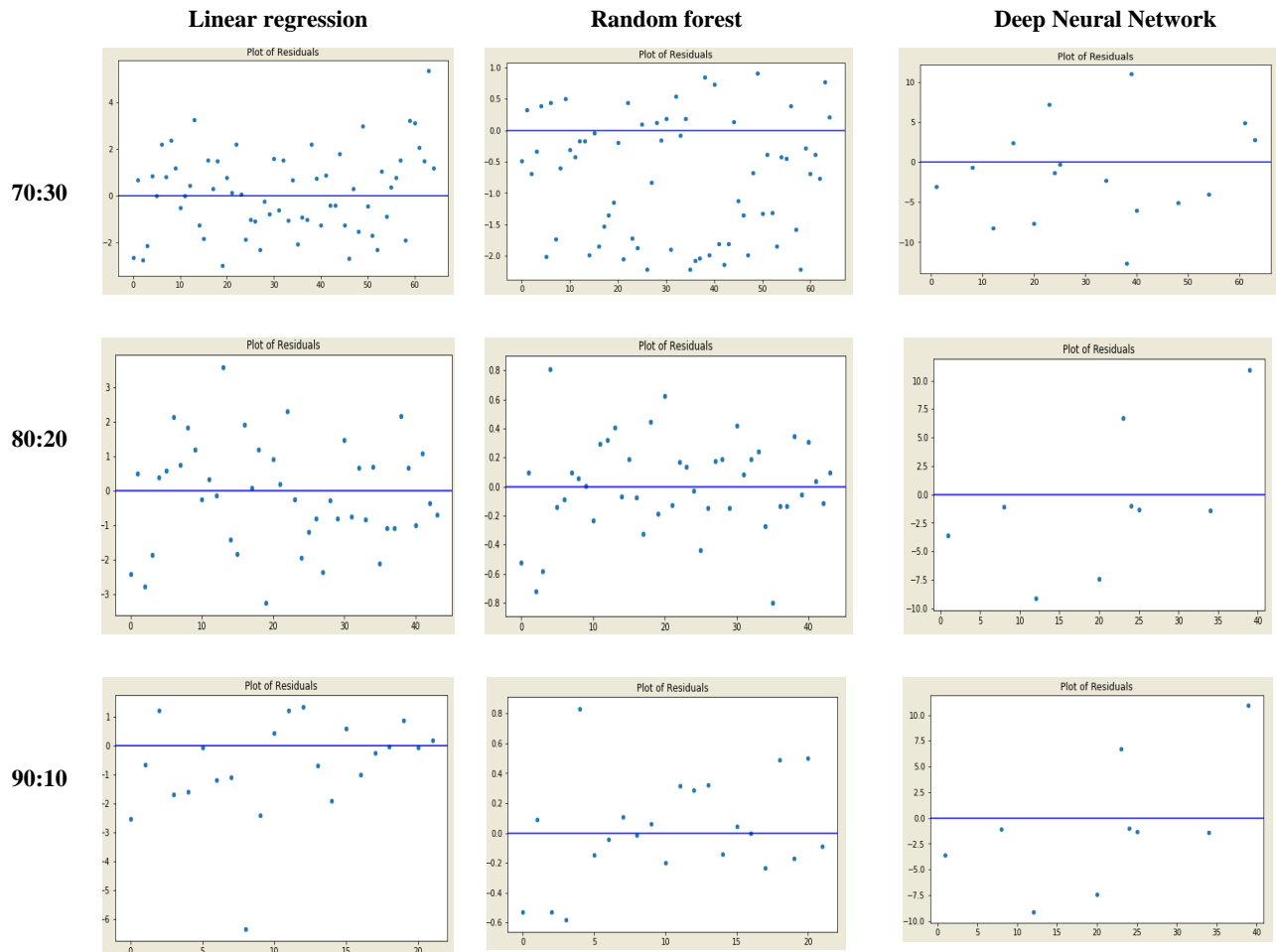
Figure 3: Scatter Plot

Overfitting problem occurs when the model learns the training data too well. Since, the training data is less here so the model learns each and every small detail. To prevent this, a number of techniques like K-fold cross validation, increasing the size of the training set, using regularization, dropout.

Another way of evaluating the model is to plot the residuals.Residuals are the difference between the actual and predicted values. Fig.3. shows the scatter plots of the residuals. The distance between a dot and the regression line in the scatter plot denotes the residuals.

When residual plots are randomly placed above and below the x-axis at y=0 then it is an appropriate model for the data. If there exists any kind of pattern in the residual plot then the model is not appropriate for the data. It is observed that the residual plot does not follow any pattern. It is scattered across the regression line

5. CONCLUSION

The goal of Machine Learning is to develop algorithms, automated techniques, to detect patterns in data. The models developed using these techniques can learn from past data and use its learning to predict the future. Climate Models are data driven models built based on solving classification and regression problems. Machine Learning and Deep Learning techniques can be used to solve these problems by using the real time and historical climate data. Here the MSE, RMSE and MAE for Random Forest is 0.105940, 0.325484 and 0.251742 respectively when the data set is split into 80 percent training and 20 per cent testing. These values give an indication that Machine Learning Techniques can efficiently predict wind speed. the model accuracy can be judged by the low error values obtained in this paper. Since the world is moving towards green energy, in the upcoming years prediction of wind speed will play a vital role in determining the potential for wind energy as well as climate for many geographical locations.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Pidwirny, M, Climate Classification and Climatic Regions of the World, Fundamentals of Physical Geography, 2nd Edition, 2006. Accessed on: April 10, 2020. [Online]. Available:http://www.physicalgeography .net/fundamentals/7v.html

[2] Fondriest Staff , Wind Speed and Direction, August 12, 2010. Accessed on: March,11 , 2020. [Online]. Available: https://www.fondriest.com/news/wind-speed-and-direction.htm

[3] A.R. Dehghanpour , A.H. Halabian, M. Fallahpour, "Impact of wind direction and speed on dusty days", International Journal of Advanced Biological and Biomedical Research, Vol. 2, No. 5, 2014,pp. 1742-1749.

[4] Annemarie D., Nicole L., and Matthias D., "Should future wind speed changes be taken into account in wind farm development?", Environmental Research Letters, Vol. 13, No. 6, May 2018.

[5] Barber, David. "The Four Forces That Influence Wind Speed & Wind Direction", Sciencing.com, Accessed on: April 25, 2020[online]. Available: https://sciencing.com/ list-7651707-four-wind-speed-wind-direction.html

[6] Bergthórsson, P., & Döös, B. R.,"Numerical Weather Map Analysis", Tellus,1955, pp.329–340.

[7] Barnes, S. L, "A Technique for Maximizing Details in

[8] Lorenc, A. C., "Analysis methods for numerical weather prediction", Quarterly Journal of the Royal Meteorological Society,1986, pp.1177–1194.

[9] Acosta, G., & Tosini, M., "A firmware digital neural network for climate prediction applications", Proceedings of the 2001 IEEE International Symposium on Intelligent Control.

[10] W. Deji, X. Bo, Z. Faquan, L. Jianting, L. Guangcai and S. Bingyu, "Climate Prediction by SVM Based on Initial Conditions", 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, Tianjin, 2009, pp. 578-581

[11] C. Monteleoni,G. Schmidt, S. McQuade,"Climate Informatics: Accelerating Discovery in Climate Science with Machine Learning", IEEE Computing in Science and Engineering (CISE) Magazine, Special Issue on Machine Learning. Vol. 15, no. 5, Sept.Oct.2013, pp. 32–40.

[12] B. Sonkusare, et al., "Improved Performance of Multi-Model Ensemble through bias correction on ANN technique", ECIAC International Conference, 2016.

[13] Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., & Yang, H.. "Machine learning and artificial intelligence to aid climate change research and preparedness", Environmental Research Letters, Vol.14, No.12, 2019.

[14] Eleftheriou, A., Kouvaris, K., Karvelis, P., & Stylios, C., "Micro Climate Prediction Utilising Machine Learning Approaches", IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea). 2018.

[15] Paramasivan, Senthil, "Improved Prediction of Wind Speed using Machine Learning", EAI Endorsed Transactions on Energy Web, Vol. 6, 2019.

[16] Li, G., & Shi, J. ,"On comparing three artificial neural networks for wind speed forecasting", Applied Energy, 2010, Vol. 87, No. 7, pp. 2313–2320.

[17] MAI Navid, NH Niloy. "Multiple Linear Regressions for Predicting Rainfall for Bangladesh",Communications. Vol. 6, No. 1, 2018, pp. 1-4.

[18] Sreehari, E., & Srivastava, S. ,"Prediction of Climate Variable using Multiple Linear Regression", 4th International Conference on Computing Communication and Automation (ICCCA), 2018

[19] S. Karthick, D. Malathi, C.Arun, "Weather Prediction Using Random Forest Algorithm", International Journal of Pure and Applied Mathematics, Vol. 118, No. 20, 2018, pp.255-262

[20] Hu, Q., Zhang, R., & Zhou, Y., "Transfer learning for short-term wind speed prediction with deep neural networks", Renewable Energy, Vol. 85, 2016, pp. 83–95.