



Wind power plant forecasting and power prediction methods using Machine Learning Algorithms

E.Bijolin Edwin¹, Darren Bert Lyngdoh², M.Roshni Thanka³

¹Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Tamil Nadu, INDIA, bijolin@gmail.com

²Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Tamil Nadu, INDIA, darrenbert73@gmail.com

³Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Tamil Nadu, INDIA, mroshni.thanka@gmail.com

ABSTRACT

A significant and eligible source such as wind energy has the potential for producing energy in a continuous and sustainable manner among renewable energy sources. However, wind energy has several challenges, such as initial investment costs, the stationary property of wind plants, and the difficulty in finding wind-efficient energy areas. In this study, wind power forecasting was performed based on daily wind speed data using machine learning algorithms. The proposed method is based on machine learning algorithms to forecast wind power values efficiently. Tests were conducted on data sets to reveal performances of machine learning algorithms. The results showed that machine learning algorithms could be used for forecasting long-term wind power values with respect to historical wind speed data. Furthermore, several machine learning models were built for analysis on the accuracy level of the respective models, i.e, the accuracy levels of the machine.

Key words : Forecasting, machine learning, neural network, power prediction, regression, wind power, wind speed.

1.INTRODUCTION

A major part of the world's energy need is fulfilled by utilizing depleting sources of energy like fossil fuels. However utilizing fuels like coal, oil and gas to produce energy releases a large number of greenhouse gasses, which causes a major problem for the environment. A way of producing clean and carbon-free energy is to use renewable sources of energy like wind, tidal and solar power to meet the increasing demand for energy production. Wind power as it stands is one of the fastest growing sources of power generation, with countries like China, United States, India

and Germany moving towards wind energy for their energy needs. The intermittency of the power output characteristics of wind power resources became one of the biggest challenges for the power system operators[1]. Wind power or wind energy is the use of wind to provide the mechanical power through wind turbines to turn electric generators and traditionally to do other work, like milling or pumping. Wind power is a sustainable and renewable energy, and has a much smaller impact on the environment compared to burning fossil fuels[3]. In 2018, global wind power capacity grew 9.6% to 591 GW and yearly wind energy production grew 10%, reaching 4.8% of worldwide electric power usage, and providing 14% of the electricity in the European Union. Denmark is the country with the highest penetration of wind power, with 43.4% of its consumed electricity from wind in 2017 [3]. At least 83 other countries are using wind power to supply their electric power grids. India as it stands also has an exponential potential for developing power plants generated through wind power.

A significant and eligible source such as wind energy has the potential for producing energy in a continuous and sustainable manner among renewable energy sources[2]. However, wind energy has several challenges, such as initial investment costs, the stationary property of wind plants, and the difficulty in finding wind-efficient energy areas. The energy produced by wind depends heavily on weather variables such as wind speed, wind direction, temperature, etc. Since the weather cannot be controlled, the amount of energy produced by wind cannot be controlled as well. This makes the integration of wind power to the grid a major challenge. To overcome this issue, it is important to make accurate wind energy forecasts in order to reduce the uncertainty in the amount of energy that is going to be produced in future.

2.LITERATURE REVIEW

Wind power forecasting based on certain research studies could be divided into two categories, such as,

short-term wind power forecasting and long-term wind power forecasting. Short-term wind power forecasting techniques forecast wind power with respect to short time periods, i.e. one hour to consecutive days ahead, while long-term wind power forecasting methods try to forecast wind power with respect to longer time periods, i.e. a few days to one year ahead [2]. Each category is divided into two separate sub-categories, such as statistical and machine learning methods.

Wind Power forecasting has its applications in generation and transmission maintenance planning, energy optimization as well as in energy trading [4]. Wind power forecasting models have been designed at different technical background and scales and they can be used to predict the production for a single wind turbine to a whole wind energy farm. In recent years, machine learning algorithms, such as support vector machine, artificial neural networks (ANN) and convolutional neural network (CNN), have been studied and implemented to understand and solve the complex challenges related to these studies on wind power forecasting. Given the powerful learning capacity, neural networks can be trained and regress a complicated relationship with appreciable accuracy[5]. Successful applications of neural networks dealing with fluid mechanics can be found in many recent studies.

Throughout the years several tools and methodologies have been developed to forecast wind power. Physical methods that make use of atmospheric descriptions to forecast wind power have been used [6]. A general method to forecasting wind power is to use numerical weather prediction data in statistical forecasting models[7]. Methodologies that are statistical such as polynomial regression, ARIMA[9] have been extensively documented in various literature work for wind power forecasting as well as ARMA[10]. However, with the development of intelligent machine learning algorithms like neural networks and support vector machines (SVM) for time series forecasting, these were used to forecast wind power as well. Machine learning techniques such as artificial neural network (ANN) and SVM are capable of performing complex nonlinear modelling for a large amount of data making it suitable for time series forecasting [8].

The method mentioned above, Autoregressive Moving Average (ARMA) has also been compared with several machine learning algorithms like artificial neural networks (ANNs) as well as adaptive neurofuzzy inference systems (ANFIS) which resulted with the conclusion that both the ARMA method as well as the artificial neural neural networks (ANNs, ANFIS) were suitable for short term wind power forecasting[11].

3.METHODOLOGY

With regards to the main objective, four machine learning algorithms were used to build the models for the selected data set obtained. The first of the algorithms is Multiple Linear Regression (MLR) which is a supervised

learning model with associated learning algorithms that analyze data used for classification and regression analysis. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. The second algorithm is Decision Tree Regression which is a Decision tree which builds regression or classification models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. The third model will be similar to decision trees known as Random Forest Regression. The fourth model will be built using Recurrent Neural Network which is a type of Neural Network where the output from previous step are fed as input to the current step.

These machine learning models were built around a similar data set to form an unbiased evaluation of the most efficient and accurate model that would produce the more accurate predictions. Proper analysis of the models will be done by calculating the accuracy levels of the predicted output using the data set which would be divided into training set and testing set.

3.1 Multiple Linear Regression Model

Multiple Linear Regression (MLR) method helps in establishing correlation between the independent and dependent variables. Here, the dependent variables are the biological activity or physiochemical property of the system that is being studied and the independent variables are molecular descriptors obtained from different representations. In linear regression models, the dependent variable is predicted using only one descriptor or feature. Multiple linear regression models consider more than one descriptor for the prediction. The formula is given below:

$$Y = b_0 + b_1*x_1 + b_2*x_2 + \dots + b_n*x_n$$

where Y is the dependent variable needed to be predicted, x is the independent variable and b is considered as the rate of change per unit time.

The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables.
- The independent variables are not too highly correlated with each other.
- Y observations are selected independently and randomly from the population.
- Residuals should be normally distributed with a mean of 0 and variance σ .

3.2 Decision Tree Regression Model

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an

associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node (e.g., Hours Played) represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs and utility. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables.

3.3 Recurrent Neural Network model

The RNN used in this study is Long Short Term Memory. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn.

All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

3.4 Random Forest Regression Model

The random forest is a model made up of many decision trees. Rather than just simply averaging the prediction of trees (which we could call a “forest”), this model uses two key concepts that gives it the name *random*:

- Random sampling of training data points when building trees
- Random subsets of features considered when splitting nodes

4.RESULTS AND ANALYSIS

The following machine learning models were constructed using python as a programming language. They were trained in a open source software application called Spyder using Python 3 on the same data set that was collected shown in the following table below:

Table 1: Data set

AirTemperature	Pressure(atm)	Wind speed(m/s)	Wind direction(deg)	Power generated(kW)
10.926	0.979103	9.014	229	33688.1
9.919	0.979566	9.428	232	37261.9
8.567	0.979937	8.7	236	30502.9
7.877	0.980053	8.481	247	28419.2
7.259	0.979867	8.383	256	27370.3
6.57	0.979884	8.256	261	25805.9
5.897	0.980318	6.476	265	11546.8

This particular data set in Table 1 showsthe consistence of 50,000 rows. This was imported into each of the following machine learning models. It consisted of daily wind speed data (Wind Speed) as well as its contributing factors like air temperature, pressure, wind direction and power generated.

4.1 MLR Testing and Results

The data set was divided into two parts: Training set (70%) and Testing set (30%). Using the predict function of the regressor class, the y value for the testing set is predicted using the model and the accuracy is calculated using the variance metric that can be derived from the sk learn metrics library.

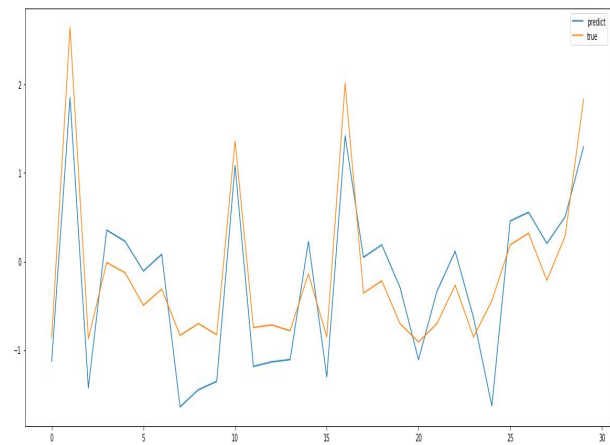


Fig. 1 Prediction graph for MLR model

In the above Fig 1 shows, the orange curve is the actual value of power generated taken from the data set (y value from the testing set). The blue curve is the predicted value of y obtained from the machine learning model. The data collected from the model is shown below:

Analysis:

Variance = 0.841

Accuracy = 84%

4.2 Decision Tree Testing and Results

The dataset was splitted into two parts: Training set (70%) and Testing set (30%).The data set was divided into two parts: Training set (70%) and Testing set (30%). Using the predict function of the regressor class, the y value for the testing set is predicted using the model and the accuracy is calculated using the variance metric that can be derived from the sk learn metrics library.

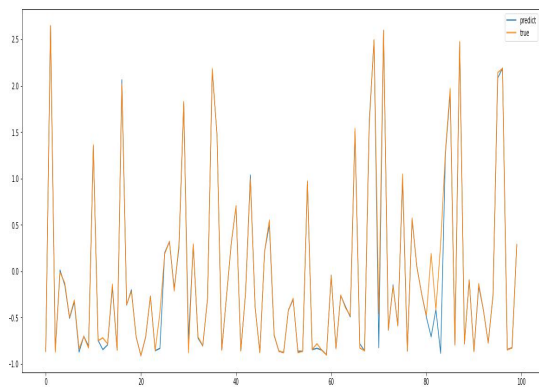


Fig. 2. Prediction graph for Decision Tree Model

In the above Fig 2 graph shows, the orange curve is the actual value of power generated taken from the data set (y value from the testing set). The blue curve is the predicted value of y obtained from the machine learning model. The data collected from the model is shown below:

Analysis:

Variance = 0.979

Accuracy = 97%

4.3 RNN Testing and Results

The data set was divided into two parts: Training set (70%) and Testing set (30%). Using the predict function of the regressor class, the y value for the testing set is predicted using the model and the accuracy is calculated using the variance metric that can be derived from the sklearn metrics library.

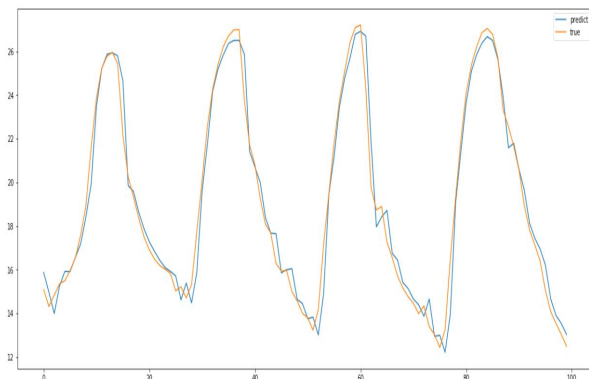


Fig. 3. Prediction graph for RNN model

In the above Fig 3 graph shows, the orange curve is the actual value of power generated taken from the data set (y value from the testing set). The blue curve is the predicted value of y obtained from the machine learning model. The data collected from the model is shown below:

Analysis:

Variance = 0.981 (R2 value)

Accuracy = 98%

4.4 Random Forest Testing and Results

The dataset was splitted into two parts: Training set (70%) and Testing set (30%). The data set was divided into two parts: Training set (70%) and Testing set (30%). Using the predict function of the regressor class, the y value for the testing set is predicted using the model and the accuracy is calculated using the variance metric that can be derived from the sklearn metrics library.

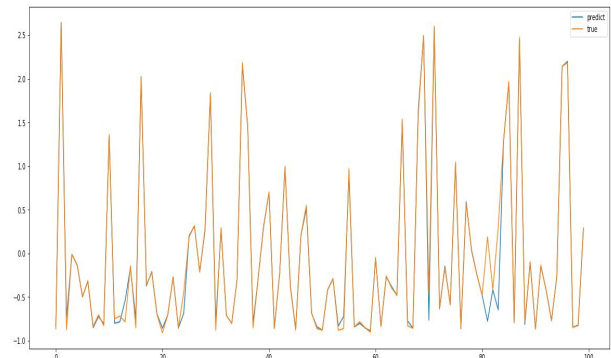


Fig. 4. Prediction graph for Random Forest Model

In the above Fig 4 graph shows, the orange curve is the actual value of power generated taken from the data set (y value from the testing set). The blue curve is the predicted value of y obtained from the machine learning model. The data collected from the model is shown below:

Analysis:

Variance = 0.986

Accuracy = 98%

5.CONCLUSION AND FUTURE SCOPE

Wind energy sources is one of the main renewable sources of energy because of environment friendly nature, availability, natural and less complex than other sources. It is much more relevant and possible to produce electrical energy from wind turbines at every hour of the day which is suitable for working systems that require a lot of energy continuously. One of the main problems however is that wind energy is challenging due its initial investments in construction of wind farms, careful analysis before establishing such farms, distance of these areas to commercial cities, and identification of high potential wind energy areas.

From the study conducted, it is observed that machine learning models can be used for the prediction of wind power from daily wind speed data and its contributing factors. The machine learning models were built and trained using the selected data set so as to evaluate their accuracy with no partiality. As far as existing systems go, the majority of the models being used were time series models and neural networks which are complex and takes up very high computing power. Regression models like multiple linear regression and decision tree regression models were constructed and both displayed positive accuracy results with 88% and 97% respectively. A Neural network using recurrent network algorithm like LSTM was also modeled using the same data set giving an accuracy level of more than 98% resulting in a prediction potential that is greater than multiple linear regression but much similar to decision trees. A final model was constructed using Random Forest Regression model which achieved a slightly higher accuracy level than the RNN model. This concludes that for the selected data set, regression models are enough or more than required for the prediction of wind power.

Experimenting with these models constructed in real-time is now crucial to determine the applicability of such machine learning models. An area with high potential for wind should be monitored in real-time using IOT or a mini wind turbine. This can be used to monitor the wind speed and several other contributing factors to power generation. The selected machine learning models then can be evaluated using this real-time data of a specific location to determine the most accurate model.

REFERENCES

1. Umit Cali , Vinayak Sharma ,**Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection.** University of North Carolina at Charlotte, Charlotte, NC, USA. Pg (2-30) 2019.
2. Demolli, Ahmet Sakir Dokuz, Alper Ecemis, Murat Gokcek ,**Wind power forecasting based on daily wind speed data using machine learning algorithms Halil**, Pg (3), March 2019.
3. Jaume Manero, Javier Bejar , Ulises Cortes,**Wind Energy Forecasting with Neural Networks: A Literature Review**,**Barcelona**, Volume 22 (Pg 4), 2018.
4. Morgan Svenson ,**Short term wind power forecasting using artificial neural networks**, Master's Thesis at School of Computer Science and Communication at Royal Institute of Technology KTH. Page 4-5, 2016.
5. Zilong Ti, Xiao Wei Deng, Hongxing Yang ,**Wake modeling of wind turbines using machine learning**, Applied Energy 257 (2020).
6. Carlini, EM., et al. **Physical and statistical downscaling for wind power forecasting**,*Proc. of 2016 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*(2016).
7. Tewari, S, Geyer CJ, Mohan N. **A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets**, *Transactions on Power Systems (IEEE)*, 2011.
8. Mao, Y. and W. Shaoshuai. **A Review of wind power forecasting & prediction**, *Proc. of 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. 2016.
9. Nayak, AK, et al. **ARIMA based statistical approach to predict wind power ramps**, *Power & Energy Society General Meeting (IEEE)*, 2015.
- 10 Cao, Y, et al. **Wind power ultra-short-term forecasting method combined with pattern matching and ARMA model**. IEEE Grenoble Conference, 2013.
11. Q.Chen, K.A.Folly. Wind Power Forecasting. IFAC PapersOnLine Volume 51, Issue 28, 2018, Pages 414 -419
12. Yingying Zhao, Dongsheng Li, Ao Dong, Dahai Kang, Qin Lv, Li Shang, **Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA** ,*Data Energies* 2017 2017
13. Emrah Ozahi, Aysegul Abusoglu, Alperen Tozlu, **A comparative thermoeconomic analysis and optimization of two different combined cycles by utilizing waste heat source of an MSWPP**, *Energy Conversion and Management* 228 (2021), 113583, Year 2021
14. Marcos Tostado-Vélez, Paul Arévalo, Francisco Jurado **comprehensive electrical-gas-hydrogen Microgrid model for energy management applications**, *Energy Conversion and Management* 228 (2021) 113726, Year 2021
15. Ceyhun Yildiz , Hakan Acikgoz, Deniz Korkmaz, Umit Budak ,**An improved residual-based convolutional neural network for very short-term wind power forecasting**, *Energy Conversion and Management* 228 (2021) 113731