WIND POWER FORECASTING USING MACHINE LEARNING IN BHUTAN

Nabindra Sharma[1], Namgay Tenzin[2], Manoj Sharma[3]
Department of Electrical Engineering, College of Science and Technology [1,2]
Rinchending, Phuentsholing [3]
nabin15sharma@gmail.com[1], namgaytenzin.cst@rub.edu.bt[2], manojsharma.cst@rub.edu.bt[3]

DOI: 10.54417/jaetm.v3i1.110

Abstract— *In this research, an approach for predicting wind energy using machine learning has been explored. An indirect method has been adopted. Predicting wind speed at first using the hourly weather data and combining that predicted wind speed with the power curve of considered wind turbine prepared by the companies. This research aims to develop a generalized machine learning based wind power forecasting model for Bhutan. Thus, hourly weather data for the year 2018 and 2019 of 300kW On-grid Wind Farm at Rubesa was used to train the base model. Meanwhile, the trained base model was tested against the weather data sets for the selected sites namely Gaselo and Dagana. A Random Forest Regression machine learning algorithm was used in this research. The developed base model has five input variables which are time, temperature, global horizontal irradiance, relative humidity, and pressure, while the target is wind speed. The R- squared values, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) for the developed base model were found to be 0.88, 0.40 and 0.30 respectively. Energy output in the wind turbine was calculated via the predicted wind speed and power curve prepared by the wind turbine companies. The calculated energy output could shape the considered theoretical power curve. The power curve considered in the present research is 300kW On-grid Wind Farm at Rubesa, Wangdiphodrang.*

Keywords—machine learning, forecasting, random forest regression, variables, train and test.

1. INTRODUCTION

Despite its hilly geography, Bhutan is blessed with abundant wind and hydro resources in numerous places. Bhutan's overall wind regime is significantly influenced by the monsoon season, which means that Bhutan has more wind from November to April and less wind during the rest of the year. Coincidentally, Bhutan is short on hydro resources during this time, necessitating importing electricity from India. Additionally, 99 percent of the country's installed capacity comes from hydropower plants, with highly varying seasonality-based river discharge and the need for enhanced energy security allows Bhutan to diversify its power supply by utilizing wind energy resources, which is an increasing source of interest in Bhutan [1].

The amount of electricity a wind farm can produce is directly influenced by components of the wind farm, wind speed, and weather conditions [2] [3]. Forecasting wind power is quite challenging since wind speed is highly variables. Different wind speeds could make a wind turbine generate zero power to its maximum capacity. Besides, wind power is not storable and must be transported once is generated. Because of the high dependency of wind power generation on varying nature of wind speed, a reliable wind power forecasting method is required [4]. A power forecasting system like this can be quite useful in anticipating how much wind energy a wind farm can generate. It can also assist in predicting the proportion of required power that a wind farm can supply to the power grid [5] [6].

Many researchers have focused on forecasting wind power using various analysis methods. Each of these researchers attempted to forecast on a different time scale. Initially, wind power forecasting was carried out using a persistent and statistical approach. Recent research, on the other hand, has favored forecasting using a machine learning system [7] [8] [9]. Classification algorithms of random forest, support vector machines and deep learning have all acquired prominence in the literature [10] [11].

The regression algorithms of random forest are examined in this study as a novel forecasting technique. Solar energy was predicted using random forest regression with greater forecasting accuracy. The approach in this research was to predict wind energy using machine learning algorithm (Random Forest Regressor). An indirect method has been adopted. Predicting wind speed at first using the hourly weather data namely temperature, pressure, global horizontal irradiance, relative humidity, with time series, and combining that predicted wind speed with power curve of considered wind turbine prepared by companies. The power curve considered in the present research is 300kW On-grid Rubesa Wind Farm. The importance of this research is to have one base machine learning model that can predict wind energy at any site in Bhutan. Bhutan has a wind energy potential of 760 MW, according to the Renewable Energy Management Master Plan [1]. The current installed capacity of the country is only 600 kW. Wind energy in Bhutan is still in the early stages of development on a commercial scale thus a wind power forecasting model like this will be critical for any investment in wind energy in the country.

2. MACHINE LEARNING

2.1. Machine Learning

Artificial Intelligence includes machine learning as a subset (AI). It is the study of making machines behave and make decisions in a more human-like manner [12]. In general, supervised, unsupervised, and reinforcement learning are three different machine learning algorithms [13]. In supervised learning, both the dependent variable sets(y) and independent variables (x) are provided. During the learning phase the machine finds the underlying pattern in these two variable sets. The learning is supervised here. Unsupervised learning is a model where only independent input data are provided without labeled dependent data. During the learning phase, the model discovers the underlying pattern and determines the output independently. There is no supervisor here. In reinforcement learning, machine learning model is trained through trial and error. In the learning phase, machine learning models are trained to make a series of decisions. The desired decision gets a reward, and the undesired gets punished.

Each of these models has its own algorithm that is specialized up to a specific problem. Classification algorithms to predict category and regression algorithms to estimate the quantity, both with supervised methods are commonly used for forecasting. In this research, the regression analysis algorithm is introduced, namely Random Forest Regression.

2.2. Random Forest Regression

Random Forest (RF) algorithm is a supervised learning technique. The algorithm employs the popular technique of bootstrap aggregation or bagging to train tree learners. Bagging takes a random sample each time (B times) the training set is replaced, and then fits trees to these samples, given a training set $X = x_1, \ldots\ldots\ldots x_n$ with responses $Y = y_1, \ldots\ldots\ldots y_n$:

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b\,(x')$$

Where f_b: is regression tree and x' is an unseen sample. After the learning phase, trained random forest regression after feeding the unseen samples x' gives its prediction. Prediction made for this unseen sample x' is the average prediction made by each of the regression trees [14].

3. METHODOLOGY

Wind turbine energy production can be determined using wind speed data and power curves prepared by wind turbine manufacturers [15] [16]. The technique of calculation is based on integrating the turbine's power curve with wind speed data collected as a time series. An indirect method has been adopted. Part 1: Predicting wind speed at first using the hourly weather data and Part 2: combining that predicted wind speed with power curve of considered wind turbine prepared by companies. Part 3: Testing the developed base machine learning model against the datasets of other candidate sites. The experimental setup is presented in figure 1.

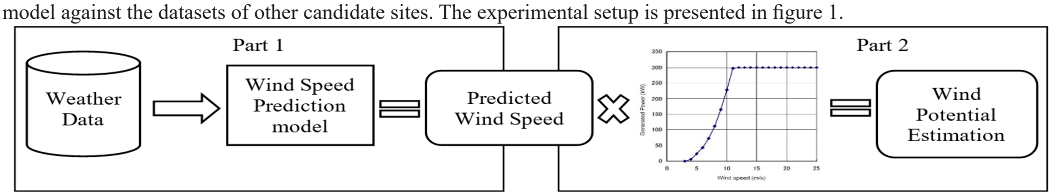

Fig. 1. Experimental Setup

Anaconda's Python 3.9 environment was used to evaluate and preprocess the collected weather datasets. NumPy, Scipy, Matplotib, Seaborn, SciKit, and Pandas were among the libraries used. The same Python 3.9 environment was used for both training and testing. Imported libraries included Sklearn.

3.1. Source and Description of Case Study

Wind speed, temperature, global horizontal irradiance (GHI), relative humidity, and pressure are among the independent variables included in the dataset. The dependent variable is the wind speed measured at the wind turbine's (WT) hub height. The predicted wind speed is then further combined with a 300kW On-grid wind turbine to obtain the wind energy output.

The dataset for the base location was obtained from the Department of Renewable Energy (DRE) and the dataset for the target locations was obtained from the National Centre for Hydrology and Meteorology (NCHM). On-grid 300kW Rubesa Wind Farm in the WangduePhodrang being the only wind farm in Bhutan were used as base location for this research.

The base machine learning model was trained using the data collected from this base location. The data from Dagana and Gaselo were then used to test the trained base machine learning. Dagana and Gaselo were chosen because they represent two different regions, the southern foothills and the inner central valley, respectively. Both the selected locations have good wind potential capacity, as well as good infrastructure and electrical grid connections [1]. Another reason for choosing was that a small number of null values were discovered during the data pre-processing step and were imputed and cleaned.

Table 1. Station Details

Station Name	Latitude	Longitude	Elevation (m)	Authority
Rubesa	27°28′6.4″	89°54′9.1″	1678	DRE
Gaselo	27°42′4.4″	89°88′6.9″	1820	NCHM
Dagana	27°07′7.3″	89°88′6.4″	1504	NCHM
Rubesa	27°28′6.4″	89°54′9.1″	1678	DRE

4. RESULT AND DISCUSSION

4.1. Exploratory Data Analysis

For the years 2018 and 2019, 1 hour time series data of meteorological variables such as wind speed, temperature, global horizontal irradiance, relative humidity, and pressure were collected. Each of the variable has 17,520 observations in the dataset. The Data frame for the Rubesa is shown in Table 2. The top five and lowest five observation points are shown in the data frame.

Table 2. Data frame for Rubesa

Year	Month	Day	Hour	Temperature	GHI	Relative Humidity	Pressure	Wind Speed
2018	1	1	0	13.2	0	81.72	1017	1.3
2018	1	1	1	12.7	0	83.96	1017	1.3
2018	1	1	2	12.2	0	86.83	1017	1.3
2018	1	1	3	11.8	0	89.52	1017	1.3
2018	1	1	4	11.3	0	93.28	1017	1.3
-	-	-	-	-	-	-	-	-
2019	12	31	19	16.9	0	82.42	1018	1.2
2019	12	31	20	16.2	0	84.51	1018	1.2
2019	12	31	21	15.6	0	85.97	1018	1.2
2019	12	31	22	15.1	0	86.82	1018	1.2
2019	12	31	23	14.7	0	87.31	1018	1.2

Tables 3 present the summary statistics for the response variables for each of the three sites. The distribution of wind speed measured at hub height is not normally distributed, according to the summary statistics shown in the table. The distribution of the response variable, wind speed recorded at hub height, is not normally distributed, as shown by the time series plot, density plot, and box plots in Figures 2. There is a variation of wind speed *(m/s)* between 2018 and 2019. With different lengths of time and in different areas, wind speed has shown to change in a pattern.

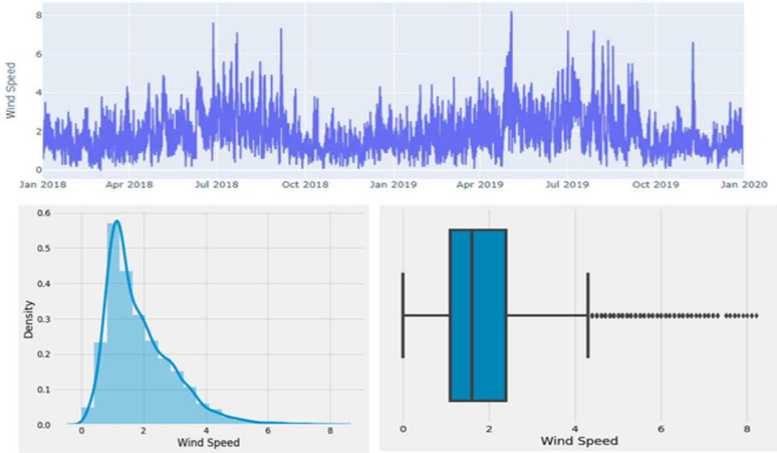

Fig. 2. Diagnostic plots for the response variables for Rubesa

Table 3. Summary statistics of the response variables for all the sites

Mean	Std	Min	25%	50%	75%	Max
			Rubesa			
1.869075	1.075071	0.000000	1.100000	1.600000	2.400000	8.200000
			Gaselo			
2.219970	1.048089	0.100000	1.400000	2.100000	3.00000	7.400000
			Dagana			
2.930375	1.750904	0.10000	1.40000	2.70000	4.20000	11.30000

Variation of other variables was also visualized. Variations of the temperature (a), pressure (b), relative humidity (c), and global horizontal irradiation (d) for all months of the years are shown in this Figure 3. All of the variables have shown variation with time for a year and the pattern is repeating for the corresponding year. Each month, or even each day, relative humidity and global horizontal irradiance change. The temperature pattern from September to January is growing and then declining, whereas the pressure trend is the reverse.

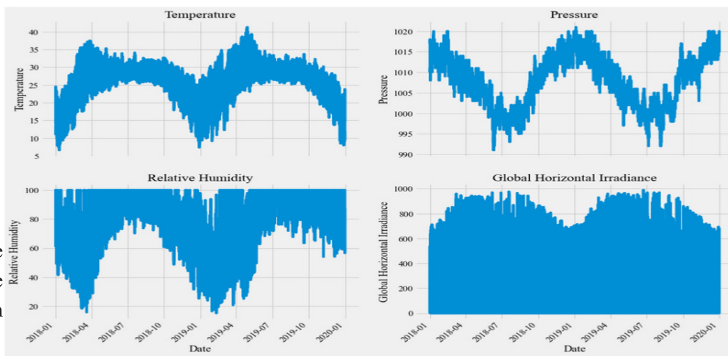

Fig. 3. temperature (a), pressure (b), relative humidity (c), and global horizontal irradiance (d) for On-grid 300kW Rubesa Wind Farm in 2018 and 2019

In machine learning, a feature is individual measurable property carrying useful information. The method adopted in building the random forest regression machine learning model is supervised learning. The supervised learning method consists of both dependent and independent features while training the random forest regression machine learning model. Choosing informative independent features plays a vital role in forecasting [17] [18] [19]. It helps to understand the vital role in the predictive problem. rf.feature_ importances_ was implemented in the Python 3.9 programming tool to award a score to input features based on how relevant they are at predicting the target value. Following figure 4 shows the correlation of independent features with dependent ones. Temperature variable showing its most importance with target variable (Wind Speed) and minute variable showing the least.

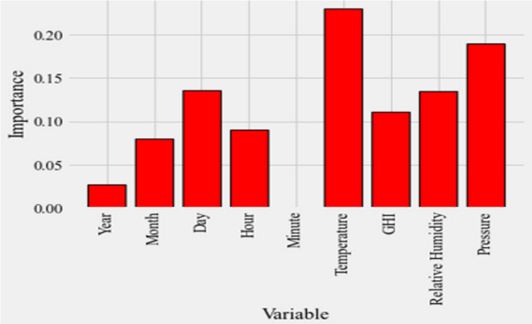

Figure 4. Variable Importance

4.2. Wind Spped Forecasting Result

The developed model comprises five inputs: time, temperature, global horizontal radiation, relative humidity, and pressure, and one output; wind speed. By default, the number of decision trees (n_estimators) was set to 100, a maximum feature was set to auto, and the random state was set to 42. Later, using RandomSearch Cross Validation (CV) from the Sklearn model selection library, hyper parameter tuning on the RF model was done for the best parameter and the least MAE. The proposed RF model had 800 trees, each with one leaf. Table 4 shows the parameters before and after hyper parameter tuning.

Table 4. Parameters before and after hyper parameter tuning.

	Default Parameter	**RandomSearchCV**
Bootstrap	True	False
Maximum depth	None	90
Maximum features	Auto	Sqrt
Minimum sample leaf	1	1
Minimum sample split	2	2
Number of decision trees	100	800
MAE	0.34	0.30

Random Search Cross Validation has narrowed the range of values for each hyper-parameter from the range of parameter grid created to sample from during fitting and has yielded a better model. With default parameter, the model performed with 0.34 MAE. After hyper-parameter tuning, model has improved its performance.

Table 5 shows the performance indices for the baseline model and proposed machine learning model. The model was evaluated using three accuracy metrics, namely RMSE, MAE, and R2.

Table 5. Performance Indices

	Baseline	**Rubesa**	**Gaselo**	**Dagana**
Number of used variables	5	5	5	5
Mean Absolute Error (MAE)	0.35	0.30	0.33	0.59
Root Mean Square Error (RMSE)	0.52	0.40	0.43	0.77
Coefficient of Determination (R^2)	0.00	0.88	0.83	0.81

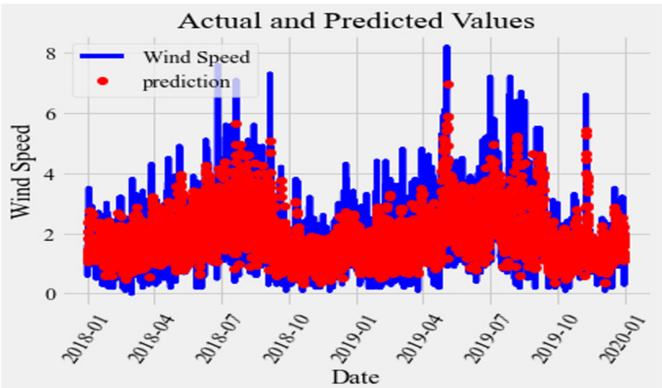

Fig.5. Point forecast plot for Rubesa

The base model was trained using data from the 300kW On-grid Rubesa Wind Farm, which has average wind energy as an option. According to the results, the developed base random forest regression machine learning model predicted response data that is wind speed throughout the year, with 0.88 r^2, 0.40 RMSE and 0.30 MAE.The graph shows that results with new wind speed datafor 300kW On-grid Rubesa Wind Farm have nearly a similar pattern with the measured wind speed dataset 300kW On-grid Rubesa Wind Farm.

Figure 6. Point forecast plot for Gaselo

Testing the base machine learning model was carried out using the weather data collected from Gaselo (DRE selected sites). Hourly weather data namely temperature, pressure, relative humidity and global horizontal irradiance with a time series for 2018 and 2019 was collected. These collected data are new datasets for the developed base machine learning model. According to the results, the developed base random forest regression machine learning model predicted response data that is wind speed throughout the year, with 0.83 r^2, 0.43 RMSE and 0.33 MAEfor Gaselo.This graph also shows that results with new wind speed data of Gaselo have nearly a similar pattern with the measured wind speed dataset of Gaselo.

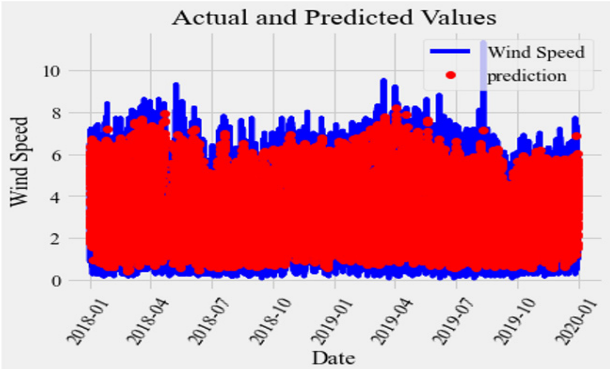

Fig. 7. Point forecast plot for Dagana

Dagana lies at an elevation of 1504 m above sea level is another DRE selected site whose weather data was used to test against developed base machine learning model. Hourly weather data namely temperature, pressure, relative humidity, and global horizontal irradiance with time series for 2018 and 2019 was collected. These collected data are new datasets for the developed base machine learning model. According to the results, the developed base random forest regression machine learning model predicted response data that is wind speed throughout the year, with 0.81 r^2, 0.77 RMSE and 0.59 MAE for Dagana. This graph also shows that results with new wind speed data of Dagana have similar trend with the measured wind speed data of Dagana though not as approximate as the above two graphs.

4.3. Power Curve Result

The hourly wind speed predicted by the proposed random forest regression machine learning model and On-grid 300kW wind turbine power curve prepared by the wind turbine companies were used to compute the energy output of the base site. The energy output calculated for the base site was plotted along with the theoretical power curve as shown in Figure 8.

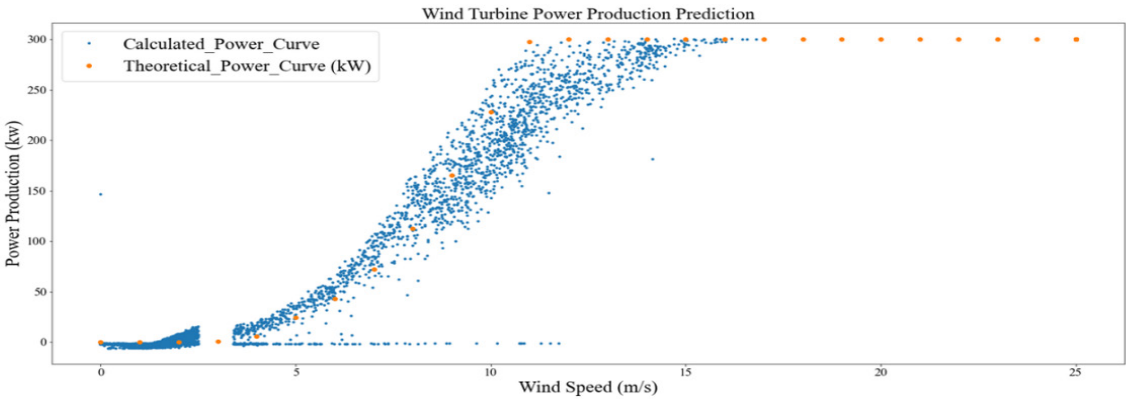

Fig. 8. Forecasted Power Curve for Rubesa

The electrical power output ratings of the machine for various wind speeds are present on the wind turbine power curve. In the above figure, the orange represents the theoretical power curve and the blue color represents the calculated power curve for the base site. The turbine starts generating power when the wind speed reaches the cut-in value and settles at rated wind speed. Calculated energy output takes the shape of the considered theoretical power curve. Both the theoretical and computed power curve rises when wind speed is at 3.0 (*m/s*) cut-in speed and settles between 11 and 12 (*m/s*)rated speed generating 300kW rated power. There is no power generation when wind speed is beyond 25 (*m/s)*cut-out speed.

4.4. Discussion

In this research, wind energy has been predicted using a random forest regression machine learning model. An indirect method has been adopted. Part 1: Predicting wind speeds at first using weather data. Part 2: Combining that predicted wind speed with power curve of considered wind turbine prepared by companies. Part 3: Testing the prepared base machine learning model against the datasets of other candidate sites (DRE selected sites).

Anaconda's python 3.9 programming language was used to do all the necessary exploratory data analysis for all the collected datasets. The random forest regression machine learning model was trained and tested using the same python statistics and machine learning toolkit.

Rubesa wind farm hourly weather data were used to train the base machine learning model then this model was tested against the datasets of other DRE selected sites. Gaselo and Dagana were two DRE selected sites. Each of these two DRE selected sites has different geographical regions and different wind characteristics with that of base location Rubesa.

In general, the developed model has five inputs which are temperature, global horizontal radiation, relative humidity, pressure, with time series and one output which is wind speed. Initially the number of decision trees (n-estimators) was made to be 100 by default, max_features as auto and the random state was given as 42. Later for the best parameter and the least MAE, hyper parameter tuning was done on the RF model using RandomSearchCross Validation from the Sklearn model selector library.

For 2018 and 2019, 1 hour time series data of meteorological variables were collected. There were 17,520 observation points in each of these input values. The datasets were split into training and testing datasets after thorough exploratory data analysis on the collected data. 75% of the data from the training datasets were used for training the model, which included both the independent and dependent variables. The remaining data were for testing the model that includes only the independent variables. Meanwhile, the proposed RF model had 800 trees, each with one leaf.

The proposed RF model was evaluated based on RMSE, MAE, and R2 which is said to be commonly used performance indices [20]. Before that, the baseline was set using the dummy regressor to understand how better the proposed model performed. The baseline sets were 0.35 of MAE and 0.52 of RMSE for this specific hourly weather data. Any values lower than this baseline set indicate model's better performance.

According to the results, the random forest predicted energy throughout the year for the base location. It's RMSE, MAE and R^2 were 0.40, 0.30, and 0.88, respectively. The developed base machine learning model random forest regression was then tested against the weather data sets for DRE selected sites namely Gaselo and Dagana. For Gaselo RMSE, MAE, and R^2 were respectively 0.43, 0.33, and 0.83 and for Dagana RMSE, MAE, and R^2 were respectively 0.77, 0.59, and 0.81. The proposed random forest regression machine learning model had least RMSE, MAE, and R^2 for the base train location data and maximum RMSE, MAE and R2 for Dagana.

Energy output in the wind turbine was then calculated via the predicted wind speed and power curve prepared by the wind turbine companies. The calculated energy output could shape the considered theoretical power curve. The power curve considered in the present research is 300kW On-grid Wind Farm at Rubesa, Wangdiphodrang.

5. CONCLUSION

Wind energy is one of the primary renewable energy sources because it is natural, inexpensive, and clean. Every hour of the day may be used to generate energy using wind turbines, making them ideal for use in systems that need constant power. However, employing wind energy is difficult because of the high upfront costs, the need for thorough analysis before creating a wind plant, the distance between wind-efficient regions and the national networks, and its disruptive impacts on the environment.

In this work, the wind energy in Bhutan was predicted using a proper understanding of machine learning. The most intelligent model was introduced using various meteorological data where the number of meteorological data and the time interval of meteorological data used to train the model has a direct impact on how much power is produced by wind turbines. The key finding is that, machine learning algorithms in conjunction with the wind power model of a base location can be employed before the construction of wind farms in uncharted territory. However, weather is a continuous, multi-dimensions, data-intensive, dynamic and chaotic process. Owing to these characteristics, highly accurate forecasting remains a big challenge. Still, the proposed solution can be further enhanced by:

- Training the model with a greater number of observation points with a lesser time horizon.
- Increasing the number of independent variables that strongly correlate with the dependent variables.
- Adopting different hyper parameter tuning techniques.
- Handling the missing data with different imputation techniques.
- Perform the cross validation when splitting the dataset into train and test data.

REFERENCES

[1] DRE-MOEA, "Renewable Readiness Assessment," 2016.

[2] Z. O. O. Folly and K. A., "Wind energy analysis based on turbine and developed site power curves: a case-study of Darling City," Renewable Energy, vol. 53, pp. 306-318, 2013.

[3] C. Diyoke, "A new approximate capacity factor method for matching wind turbines to a site: case study of Humber region, UK," International Journal of Energy and Environmental Engineering, vol. 10, pp. 451-462, 2019.

[4] A. Khosravi, L. Machado and R. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," Applied Energy 224, 2018.

[5] Fattal.J, Ezzine.L, Aman.Z, Moussami.E.H and lachhab.A, "Forecasting of Demand Using ARIMA Model," International Journal of Engineering Business Management, 2018.

[6] C.Monteiro, R.Bessa, V.Miranda, A.Botterud, J.Wang and G.Conzelmann, "Wind Power Forecasting: State-of-the-Art," Argonne

National Laboratory ANL/DIS-10-1, 2009.

[7] H. Demolli, A. S. Dokuz, A. Ecemis and M. Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms," Energy Conversion and Management, 2019.

[8] M. Yousuf, I.AI-Bahandly and E.Avci, "Current perspective on the accuracy of deterministics wind speed and power forecasting," IEEE Access, 2019.

[9] A. Khosravi, R. Koury, L. Machado and J. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," Sustainable Energy Technologies and Assessments 25, 2018.

[10] M. Yousuf, I.AI-Bahandly and E.Avci, "Current perspective on the accuracy of deterministics wind speed and power forecasting," IEEE Access, 2019.

[11] T. Khatib, R. Deria and A. Isead, "Assessment of Three Learning Machines for Long-Term Prediction of Wind Energy in Palestine," Mathematical Problems in Engineering, 2020.

[12] A. Geron, Hands-On Machine Learning with Sci-kit Learn and TensorFlow, United State of America, 2017.

[13] Marsland and Stephen, MACHINE LEARNING An Algorithm Prespective, 2015.

[14] A. Chakure, "Random Forest and its Implementation," 29 6 2019. [Online]. Available: https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f.

[15] M. Gokcek and M. Genc, "Evaluation of electricity genration and energy cost of wind energy conversion system (WECSs) in central turkey," Appl. Energy 86, 2009.

[16] C. Carrillo, A. F. Obando Montano, J. Cidras and E. Diaz-Dorado, "Review of power curve modelling for wind turbines," Renewable and Sustainable Energy Reviews, vol. 21, pp. 572-581, 2013.

[17] L. Jung-Pin, C. Yu-ming, C. Chieh_Huang and P. Ping-Feng, "A Survey of Machine Learning Models in Renewable Energy Predicitions," applied sciences MDPI, 2020.

[18] Ahmed, S. K. Mukhiya and Usman, Hands-On Exploratory Data Analysis with Python, India, 2020.

[19] Geron and Aurelien, Hands-On Machine Learning with Sci-kit Learn and TensorFlow, United State of America, 2017.

[20] Sevensson and Morgan, "Short-Term Wind Power Forecasting Using Artifical Neural Network," Sweden, 2015.