

# Wind power prediction based on meteorological data visualization

Shengchi Liu<sup>1</sup>, Shuangyue Xiao<sup>2</sup>, Li Liu, Junqiao Liu✉  
{1830997169@qq.com<sup>1</sup>, [1546536983@qq.com](mailto:1546536983@qq.com)<sup>2</sup>, [dalianliujunqiao@126.com](mailto:dalianliujunqiao@126.com) ✉-};

Department of Information Science and Engineering Dalian Polytechnic University  
Dalian, P. R. China

**Abstract.** With the development of clean energy, wind power generation has become one of the most important power generation methods. However, the output power of wind power generation system is characterized by uncertainty, so the effective interval prediction of wind power is an effective method to reduce the uncertainty. In this article, through multi-channel multi-dimensional meteorological data, visual correlation analysis, and in-depth analysis of the main factors affecting wind power, put forward based on the extreme gradient promotion (XGB) improved LGB model to forecast. In addition, in order to improve the model calculating speed and accuracy, using principal component analysis was carried out on the original data dimension reduction analysis and visualization processing, then predicted the results compared with the actual situation, to verify the validity of the established model, it shows that this method can be applied to the era of big data of wind power prediction in the future.

**Keywords:** decision tree; wind power prediction; meteorological data visualization; correlation analysis

## 1 Introduction

In order to ensure the safe production, reliable operation and regulation of the wind power generation system [1], the micro-grid department must make reasonable plans of energy storage and output [2], so it is necessary to predict the wind power generation system [3]. For the power system with access to wind power generation, due to the randomness of the output power of wind power generation system [4], the prediction of the output power of wind power generation system has been a research hotspot at home and abroad.

The world's most developed countries have stronger ability to predict the wind power [5], Denmark is one of the first countries to use wind power in the enhancement power prediction [6], in the early 1990 s, Denmark began to use the wind power prediction system for wind power prediction, the Danish national grid by Danish meteorological administration, the European centre for the weather forecast and ConWX numerical weather prediction products [7] of the company, and then through the national statistical prediction model for wind power forecast, finally to realize the integration of the forecast[8]. The United States takes a leading position in

---

✉ Corresponding author: Junqiao Liu

Fund Project: Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Danlian Science and Technology Innovation Fund (2020JJ26GX029)

wind power prediction and microgrid research and practice[9], and is one of the countries with the largest number of microgrid demonstration projects in the world. Among them, the microgrid project of the University of Wisconsin-Madison is a strong representative.

In wind power prediction in China started late, the micro grid wind power prediction research is less, the current domestic have based on time series method and the neural network study of wind power [10], on the basis of time series short-term prediction, for a long time prediction effect is poorer, numerical weather forecast data, forecast the time scale of the general is limited to six hours, for micro power dispatching operation is of great significance, the research questions for emergency micro power grid scheduling needs to be improved.

At present, there are two commonly used methods at home and abroad. One is the direct method, that is, the wind power generation environment is not predicted [11], but the historical meteorological data is directly used to predict the power. The other is the physical method, which takes terrain, roughness and other information into consideration and uses physical equations to predict and establish a short-term prediction model.

In this paper, through multi-dimensional historical meteorological data, correlation analysis will be carried out on factors affecting air pressure, temperature, humidity, visibility, , wind direction, wind speed and other factors, so as to obtain further in-depth analysis of the main factors affecting wind power [12]. On the basis of the original data, further wind direction, wind speed are the main influencing factors of extraction of the data processing [13], and then respectively the average power and total power of the wind direction visualization, According to the data visualization effect [14], the characteristics of local wind can be observed more intuitively, and then through the direction of the wind power curve, clear to observe the influence of the wind to the wind power, and by comparing the wind direction, wind speed 3 d histogram [15], and the wind rose, intuitive to observe the wind direction, wind speed distribution characteristics. Through visualization, the optimal prediction model can be established more clearly, so as to improve the accuracy of prediction.

Finally, XGB and LGB predictions [16] are selected to achieve better results. XGB model cost function adds regular term to control model complexity. The regular term contains the number of leaf nodes of the tree and the sum of squares of L2 module of score output on each leaf node. From the perspective of biase-variancetradeoff, the regular term reduces the variance of the model, makes the learned model more simple and prevents overfitting, which is also a feature of XGBoost over traditional GBDT. Pruning is available, and leaf node output L2 smoothing is new. It has good effect.

## 2 Data preprocessing

The data used in this study were from DarkSky's historical data, including cloud cover, dew point temperature, humidity, ozone, atmospheric pressure, temperature, vuindex, intensity, visibility, wind direction, gust, wind speed and power multidimensional characteristics. The data accuracy used in this study can reach the level of hours, and the corresponding model can be established with hours as the time sampling interval. Finally get the full year 2018 data, will obtain the multidimensional feature information of pretreatment, the unified data format, fill in the missing data, data visualization and convenient prediction model is established, in these data visualization using annual data, then by further processing data, the prediction model [17] is set up when the data from January 2018 to November as training set of forecasts [18], in December as the data set of predictionv [19]. Compare the results of prediction model analysis [20].

## 2.1 correlation analysis

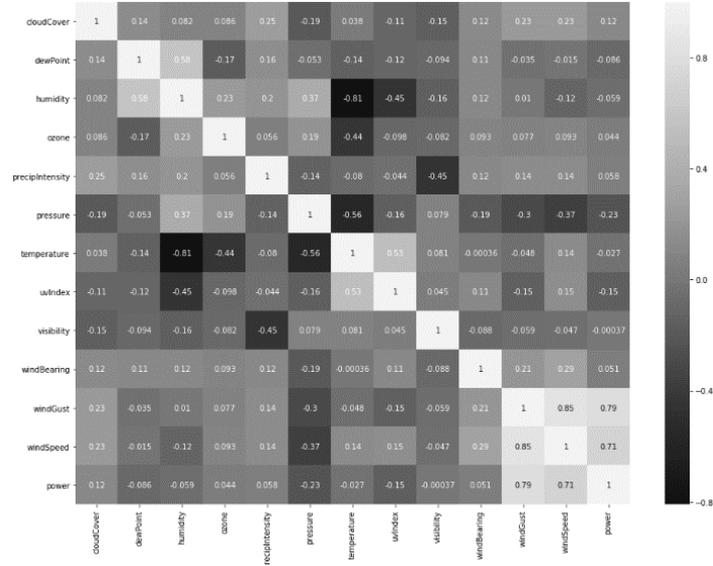


FIG. 1 Multi-dimensional correlation analysis diagram

Based on the obtained data, in python environment, the heatmap function in seaborn library was called to draw the correlation analysis diagram with multidimensional feature information. and the heatmap was obtained as shown in figure 1.

As shown in the multi-dimensional correlation analysis diagram, the power is taken as the dependent variable and the weight of the influencing factors is compared. At this time, the main factors affecting the power are wind direction and wind speed [21]. Other factors have very little effect on wind power. The main factors affecting wind power are wind direction, wind speed and gust wind. Further process the visual data, and extract the wind direction and speed data. At the same time, remove other factors affecting wind power, such as air pressure, temperature, humidity, visibility and uvindex , etc. to further visualize the new data.

## 3 Visual data analysis

Smaller portions of removal influence factors and the new data by the theory of wind direction, wind speed, power, actual power, and wind direction data further processing, converting wind 0-360 degrees 12 wind direction, respectively, using N, NNE, NEE, E, SEE, SSE, S, SSW, SWW, W, NWW and NNW said, according to the different wind data, draw the theory of the average power of the actual power and power comparing histograms, as shown in figure 2.

According to the data of different wind direction, draw the comparison diagram of the actual power and theoretical power of the total power, as shown in FIG. 3.

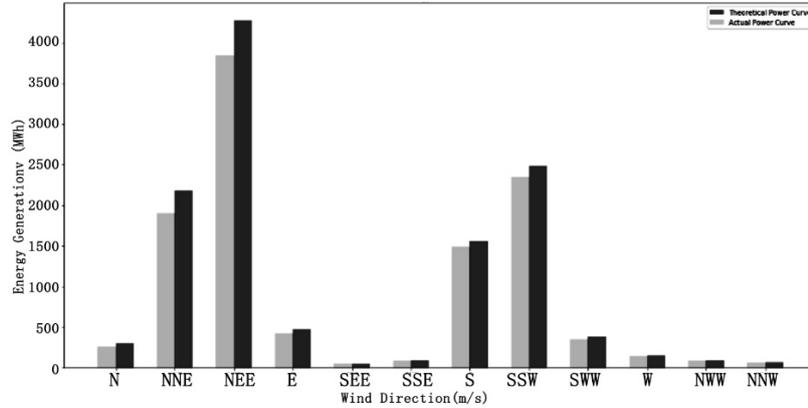


FIG. 2 Bar chart of average power of each wind direction

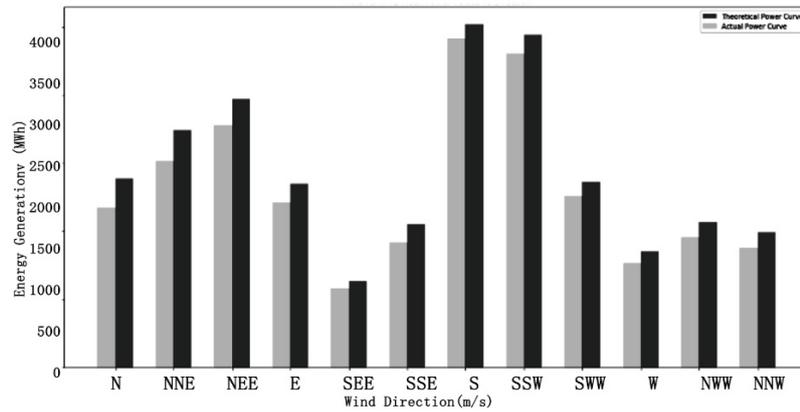


FIG. 3 Histogram of total power at each wind direction

According to Figure 1, the average power distribution of each wind direction can be seen. By comparison, the total power of northeast wind and southwest wind is relatively large, while that of southeast wind and southwest wind is relatively small. Moreover, the theoretical power value of EAST-northeast wind differs greatly from the actual power value, and the characteristics of the average power and total power of each wind direction can be seen by comparing Figure 2 and Figure 3. According to the comparison of total power and average power, it can be seen that there are more south and southwest winds in this region, and the wind speed and generating power are relatively high; the average power of northeast wind is particularly high, and the total power is medium; the northeast wind is relatively less, but the wind is very strong. The characteristics of local wind direction can be preliminarily obtained, and the spatial layout of wind turbines can be optimized. Improve the efficiency of electricity generation [22].

### 3.1 Draw the relationship curve between wind speed and wind power

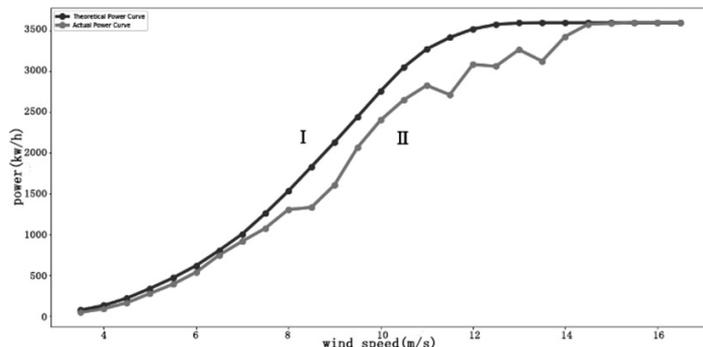


FIG. 4 Wind speed total power curve

In order to further analyze the characteristics of wind data in this region, the actual and theoretical power curves of each wind direction can be drawn to compare the actual and theoretical power curves of different wind directions and observe the power changes under different wind directions. First draw the wind speed and power curve, as shown in the following figure 4. Which I theoretical value, says II said the actual value.

Then, for 12 different wind directions, draw main influence wind power generation of the main direction, such as NNE, NEE, SSW, SWW, draw the curve relation diagram of actual power and theoretical power of different wind directions, as shown in the figure below.

According to the relationship between wind direction and wind power, the influence of different wind directions on wind power output can be seen. Take the wind direction of SSE and SEE for example, when the wind speed is relatively high, the power generation attenuation is relatively large.

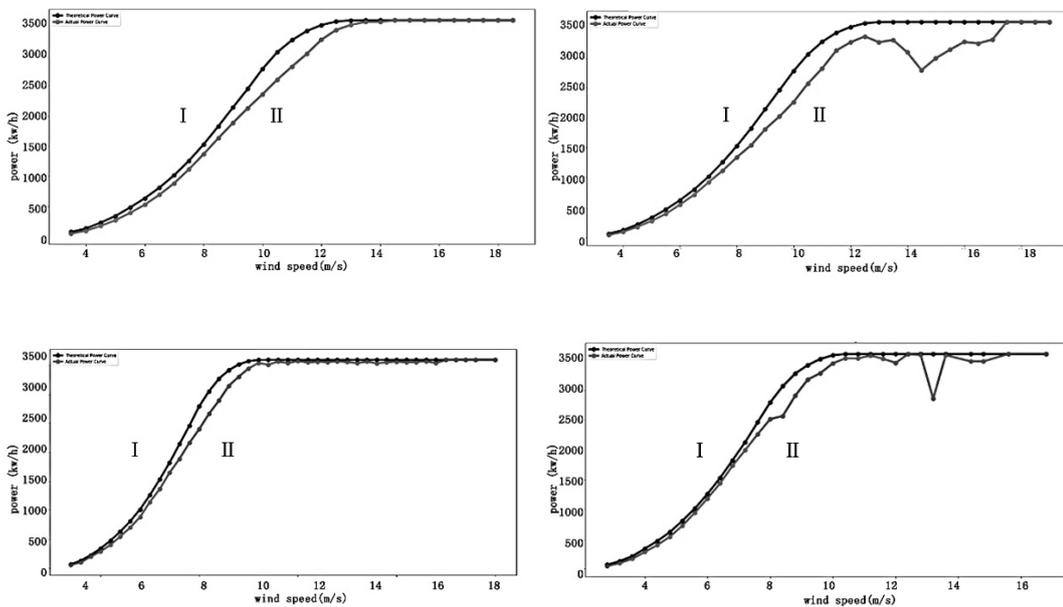
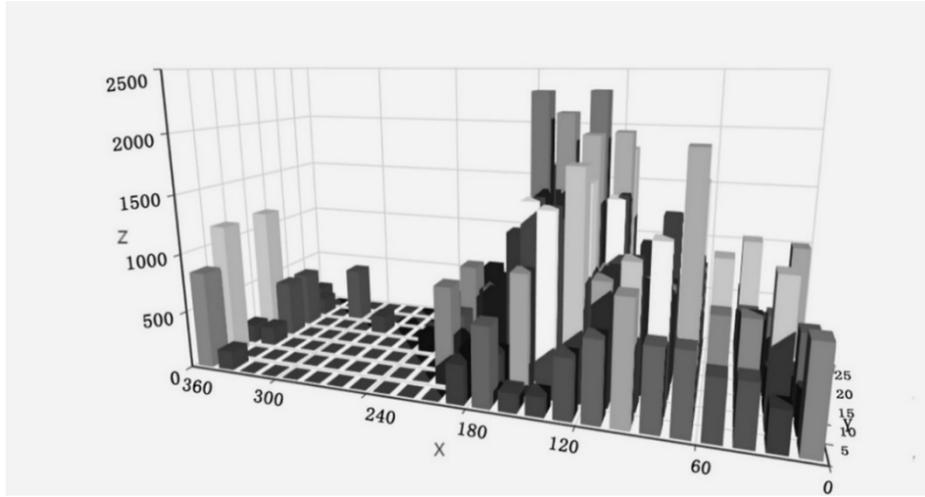


FIG. 5 Different wind direction curves



**FIG. 6** 3D histogram of wind direction and speed distribution

### 3.2 Visualization of wind direction and speed distribution

By comparing the wind direction of NEE and SSW, the theoretical value of wind power curve is close to the actual value, showing good power generation. And according to the wind direction total power diagram, we can preliminarily judge the rationality of wind turbine installation in this area.

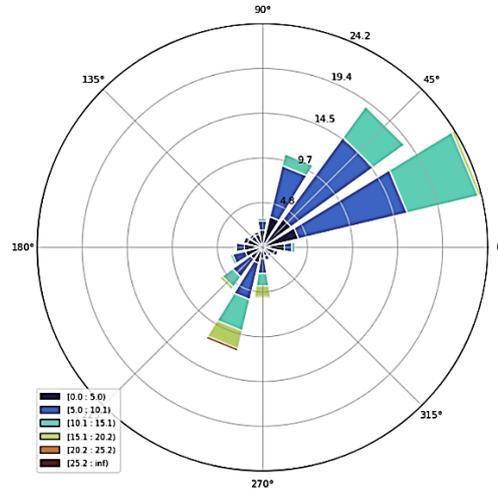
Wind direction and speed have special properties. According to the change of wind direction and speed, the distribution of detected data is converted into a visual 3D rendering using the drawing tool, as shown in Figure 6.

In the 3D figure, the direction of X axis represents the wind direction, the direction of Y axis represents the size of the wind speed, and the Z axis represents The Times of data appearing in the corresponding X and Y axes. Raw data further processing in the visual process, first of all for the sampling, will have a uniform wind direction, wind speed and height of histogram said corresponding to the number of occurrences of wind direction, wind speed, through the intuitive see above the wind distribution characteristics of concentration distribution between 30 degrees - 90 degrees, which are mainly concentrated in NEE, NNE direction, the distribution of wind speed are mainly concentrated in the range of 5 to 15 m/s, according to the characteristics of the wind can be optimized in the region of the wind power equipment installation program.

The data is divided into 16 directions according to the wind direction, as shown in the corresponding direction table 1.

**Table 1.** Corresponding azimuth chart of wind direction rosette.

<b>Number</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
<b>Direction</b>	N	NNE	NE	ENE	E	ESE	SE	SSE
<b>Number</b>	9	10	11	12	13	14	15	16
<b>Direction</b>	S	SSW	SW	SWS	W	WNW	NW	NNW



**FIG. 7** Wind direction roses map

Then correspond the data to the wind direction rose chart, as shown in **Figure 7**. Draw wind direction roses according to the processed data; Can see clearly the main distribution of wind direction, wind speed, and the percentage of each wind direction and wind speed. As shown in the figure, the concentric circles represent the frequency of different wind directions. The farther away from the center, the greater the frequency. The different colors of the chart indicate the difference of wind speed. The wind speed range is 0 to 25.2m/s. According to the data statistics, the frequency of various wind directions during this period, Calculation formula:

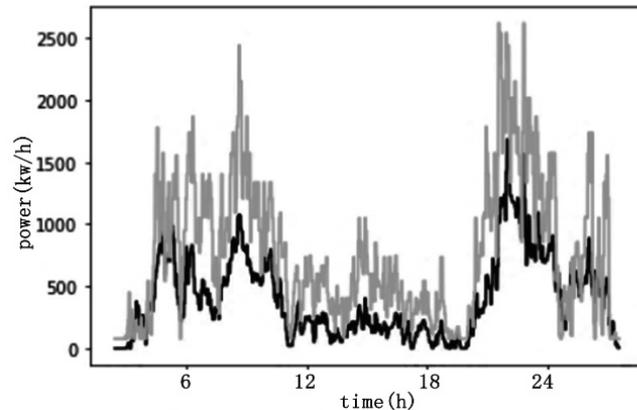
$$g_n = \frac{f_n}{\sum_{n=1}^{16} f_n} \quad (1)$$

Where  $g_n$  is the frequency of wind direction in  $n$  direction and  $f_n$  is the frequency of wind direction in  $n$  direction in this period. According to the pie chart can clearly see that wind speed are mainly concentrated in 5 to 15 m/s, and the larger wind speed generally appeared between 180 degrees to 270 degrees, namely SSW, SWW direction, combining the characteristics of figure 1, figure 2 power, in order to establish the prediction model in this paper, to provide better modeling scheme, after the visualization of data can be applied to meteorological data statistics, industrial layout and other applications.

#### 4. Building prediction model

Decision tree [23] is a very common classification method. It is based on the known probability of occurrence of various situations, The probability that the expected value of NPV is greater than or equal to zero is obtained by constructing a decision tree. Compared with the XGBoost method, XGBoost uses the method of pre sorting [24], the calculation process is sorted according to the value, one by one data samples to calculate the division of income, this algorithm can accurately find the best division value, but the cost is similar, there is no good generalization.

LGB has stronger adaptability and higher efficiency. LGB does not use the traditional idea of pre sorting, but divides each value into a series of discrete fields, that is, silos. Take floating-point data as an example, the value of an interval will be taken as a cylinder, and then the histogram of these cylinders as the precision unit will be used.



**Fig.8.** Prediction effect diagram based on LGB model

In this way, the data expression becomes more simplified, the use of memory is reduced, and the histogram brings a certain regularization effect, which can make our model avoid over fitting and has better generalization. When processing feature dimension is high and data volume is large, LGB has faster training speed on the basis of ensuring accuracy and precision.

In this paper, the data from January to November 2018 is used as the training data of the model [25], and the data of December is used as the prediction data for prediction. First, the training data is input into the prediction model for training [26], and then the test data set is used to compare the prediction results, which are shown in **Figure. 8** below.

As shown in figure 8, compared to forecast power and real power, according to the test data for predicting power value is generally greater than the actual power value, but the trend of wind power prediction and the actual value, which have kept broadly in step with good analysis of the effect, the prediction results can optimize energy storage and wind power system dispatching plan, realize the adjustment of wind power generation system in advance to maximize the use of energy, reduce more economic expenditure.

## 5. Conclusion

In this paper, multi-dimensional correlation analysis is provided based on the multi-dimensional historical meteorological data. Visual tools are used to sort out the statistical historical data to intuitively display the characteristics of wind power data in this region. LGB model, a learning algorithm based on decision tree, is proposed to predict wind power. First by pretreatment of data, the multidimensional characteristics data correlation analysis, analysis of the main factors influencing the wind power, wind direction, wind speed is put forward the main factors affecting the data further processing, the processed data visualization analysis, respectively, analyze the impact of wind power and the influence of wind speed of power, so as to provide more choice model and optimization model of optimal solution. The main contribution of this paper is to use visualization tools, to establish a prediction model and analyze the characteristics of wind power in the region with the intuitive data support, to comprehensively consider the effect of different influence factors on wind power, the selection of the optimal data of wind power prediction model, finally compare prediction model, the LGB model based on decision tree has better prediction accuracy and faster calculation speed. The advanced prediction of wind power can greatly help wind power generation system to save more

energy and realize the maximization of economic benefits.

### **Acknowledgment**

The authors acknowledge financial support from the Natural Science Foundation of Educational Department of Liaoning Province (Grant: J2020053), and Technology Innovation Fund (Grant: 2020JJ26GX029) and would like to express many thanks to the support of Dalian Key Laboratory of Smart Micro-grid and Green Recycling Industry.

### **References**

- [1] González-Sopeña, J, M.: An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renewable and Sustainable Energy Reviews*. pp. 245-252 (2020)
- [2] Zheng, X.: Deep Neural Network for Short-Term Offshore Wind Power Forecasting. *IEEE Transactions on Power Systems*. pp. 101-106 (2018)
- [3] Electronics.: Combining K-Means and XGBoost Models for Anomaly Detection Using Log Datasets. *IEEE Transactions on Industrial Informatics*. pp. 221-225 (2020)
- [4] Shi, Z.: Direct interval forecast of uncertain wind power based on recurrent neural networks. *IEEE Transactions on Sustainable Energy*. pp. 1177-1187 (2017)
- [5] Fu, X.: Spatiotemporal Attention Networks for Wind Power Forecasting. *International Conference on Data Mining Workshops*. pp. 149-154 (2019)
- [6] Kong, W.: Short-term residential load forecasting based on resident behaviour learning. *IEEE Transactions on Power Systems*. pp. 1087-1088 (2017)
- [7] Zhao, Y.: Research on Photovoltaic Power Generation Output Forecasting Model Based on Wavelet SVM and Micrometeorological Influencing Factor Analysis. *Innovation in Artificial Intelligence*. pp. 863-868 (2020)
- [8] Chen, G.: An ADMM-based distributed algorithm for economic dispatch in islanded microgrids. *IEEE Transactions on Industrial Informatics*. pp. 3892-3903 (2017)
- [9] Meng, W.: Distributed energy management in smart grid with wind power and temporally coupled constraints. *IEEE Transactions on Industrial Electronics*. pp. 6052-6062 (2017)
- [10] K oksoy.: Improved wind power forecasting using combination methods. *IEEE 14th International Conference on Machine Learning and Applications*. pp. 26-32 (2015)
- [11] Peng.: A very short term wind power forecasting approach based on numerical weather prediction and error correction method. *IEEE Transactions on Power Systems*. pp. 303-308 (2016)
- [12] Gupta, S.: Wind ramp event prediction with parallelized gradient boosted regression trees. *IEEE*. pp. 5296-5301 (2016)
- [13] Afrasiabi, M.: Advanced deep learning approach for probabilistic wind speed forecasting. *IEEE Transactions on Industrial Informatics*. pp. 720-727 (2020)
- [14] Khodayar, M.: Rough deep neural architecture for short-term wind speed forecasting. *IEEE Transactions on Industrial Informatics*. pp. 2770-2779 (2017)
- [15] Zhou, Q, Y.: A modern library for 3D data processing. *IEEE Transactions on Industrial Informatics*. pp. 236-242 (2018)
- [16] Jun, M.: Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. *IEEE Transactions on Industrial Informatics*. pp. 463-469 (2020)
- [17] Ozkan, M, B.: Statistical hybrid wind power forecast technique. *IEEE Transactions on Industrial Informatics*. pp. 375-387 (2015)

- [18] Lemenkova, P.: Processing oceanographic data by Python libraries NumPy, SciPy and Pandas. *IEEE Transactions on Knowledge and Data Engineering*. pp. 121-125 (2019)
- [19] Jiang, Z.: Focal-test-based spatial decision tree learning. *IEEE Transactions on Knowledge and Data Engineering*. pp. 1547-1559 (2014)
- [20] Wang, L.: Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications*. pp. 855-863 (2015)
- [21] Yang, W.: Load frequency control strategy for wind power grid-connected power systems considering wind power forecast. *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration*. pp.763-768 (2019)
- [22] Wang, Y.: Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. *Applied Energy*. pp. 1097-1112 (2017)
- [23] Wang, Y.: Optimal wind power uncertainty intervals for electricity market operation. *IEEE Transactions on Sustainable Energy*. pp. 199-210 (2017)
- [24] Alessandrini, S.: A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*. pp. 768-781 (2015)
- [25] Zjavka, L.: Direct wind power forecasting using a polynomial decomposition of the general differential equation. *IEEE Transactions on Sustainable Energy*. pp. 1529-1539 (2018)
- [26] Perry, R.: Multiview Machine Learning in Python. *IEEE Transactions on Knowledge and Data Engineering*. pp. 356-360 (2019)