



A New Hybrid Approach of Clustering Based Probabilistic Decision Tree to Forecast Wind Power on Large Scales

Mansoor Khan^{1,2} · Chuan He⁴ · Tianqi Liu² · Farhan Ullah³

Received: 5 March 2020 / Revised: 8 April 2020 / Accepted: 18 November 2020 / Published online: 4 January 2021
© The Korean Institute of Electrical Engineers 2020

Abstract

The wind power forecasting plays a vital role in renewable energy production. Due to the dynamic and uncertain behavior of wind, it is really hard to catch the actual features of wind for accurate forecasting measures. The patchy and instability of wind leading to the assortment of training samples have a main influence on the forecasting accuracy. For this purpose, an accurate forecasting method is needed. This paper proposed a new hybrid approach of clustering based probabilistic decision tree to forecast wind power efficiently. The collected data is screened for noisy information and selected those variables which mainly contribute in accurate predictions. Then, the wind data is normalized using mean and standard deviation to extract playing level fields for each feature. Based on the similarity of the data behavior, the K-means clustering algorithm is applied to classify the samples into different groups which contain the historical wind data. Further, the Naïve Bayes (NB) tree is proposed to extract probabilities for each feature in the clusters. The NB tree is a hybrid model of C4.5 and NB methods that successfully applied on three big real-world wind datasets (hourly, monthly, yearly) collected from National Renewable Energy Laboratory (NREL). The forecasting accuracy exposed that the proposed method could forecast an accurate wind power from hours to years' data. Comprehensive comparisons are made of the proposed method with the most popular state of the art techniques which show that this method provides more accurate prediction results.

Keywords Wind power forecast · K-means clustering · NB tree · Machine learning · Windfarm

1 Introduction

Currently, the creation of power energy is examined broadly because of energy disasters and global environmental change. The production of a sustainable power source plays out a fundamental job in the financial development of a nation. Wind power is viewed as an essential asset for electrical power generation. The introduced limit of wind power worldwide has expanded multiple times to a sum of 435 GW, with 17% aggregate development over the most recent couple of years. In 2020, wind energy is relied upon to supply roughly 12% of the all-out overall prerequisite [1, 2].

The unsure features and inferior controllability of wind power can grow the issue of its constancy and uncertainties of power systems. Moreover, wind speed might be effectively affected by stature and various sorts of obstacles. Thus, a reliable power forecasting system is required to enhance the accuracy of stable power generation and decline operational expenses. Numerous sorts of research have structured different kinds of calculations to forecast wind power [3]. Generally, the wind power forecasting

✉ Chuan He
he_chuan@scu.edu.cn
Mansoor Khan
khan007_bet@yahoo.com
Tianqi Liu
tqliu@scu.edu.cn
Farhan Ullah
farhankhan.cs@yahoo.com

¹ School of Electronics and Materials Engineering, Leshan Normal University, Leshan 614000, China
² School of Electrical Engineering, Sichuan University, Chengdu 610065, China
³ School of Software, Northwestern Polytechnical University, Taicang Campus, Xi'an 215400, Jiangsu Province, China
⁴ School of Electrical Engineering, Sichuan University, Chengdu 610065, China

techniques are partitioned into three principle classifications such as Numeric Weather Prediction (NWP), statistics-based and hybrid [4, 5]. Mathematical models are intended to predict valid NWP, which is progressively significant for longer horizons with better accuracy. But, it is difficult to build up an accurate mathematical model without the deep analysis of material science and the environment. These type of models uses various methodological factors that are difficult to measure. Statistics-based methods are utilized to predict the correlation among various features of historical-based wind data with the assistance of illustrative factors. These require just wind information for forecasting, and consequently, these methods are quite compelling for different kinds of engineering applications. The forecast accuracy of statistical methods reduces in long horizons. The Artificial Neural Network (ANN) [6], Convolution Neural Network (CNN) [7], Support Vector Regression (SVR) [8] and Back Propagation Neural Network (BPNN) [9] used mostly statistical techniques. The ANN algorithm contains different machine learning algorithms to process the complex type of data. Such type of algorithm needs data samples as input and then produce particular actions. CNN is a specific kind of artificial neural network which utilizes a perceptron to examine the data in supervised learning. The BPNN is used to compute the aggregate of gradient weights for generalization of feed-forward strategy in a multilayer neural network. It adopts the chain rule to calculate the gradient in a loop structure for each layer.

Due to real-time behaviour of wind speed, the Time series based models are used to forecast wind power mainly in minutes or hours. These are valuable for ultra-term low wind signals since they can mine the hidden stochastic features, such as Box–Jenkins models [10], Kalman filters [9] and ANN [11]. The Box–Jenkins model works at nonlinear estimation, so forthcoming perceptions may not be consumed to alter parameters straightforwardly; thus, it needs complete assessment for each observation based on complex computations. The Kalman filter, also called a direct quadratic estimator, utilizes a finite straight evaluated an incentive for determining future qualities so it needs less computations when contrasted with Box–Jenkins. In this way, Box–Jenkins provide better performance however needs an enormous degree of computations when contrasted with Kalman filters. The Kalman filter is best for fewer perceptions and forecasts. The ANN has gotten consideration for managing dynamic information in light of the fact that such information might be incredibly nonlinear. It utilizes neural system calculations, for example, a multilayer perceptron, yet doesn't completely utilize an efficient way due to its formal specification. The prediction capability of these methods falls in longer forecasting horizons [10, 12, 13].

1.1 Novelty and Contributions

Because of uncertain real-time behaviour of wind power, it is really a challenging task to extract meaningful features based on hourly or monthly wind data collection. The wind power features can be clustered based on their related features in terms of longitude, latitude, speed, the capacity factor of each turbine, etc. [14]. In this paper, we proposed a novel hybrid approach of clustering based probabilistic decision tree to forecast wind power on short as well as long term data. The wind data features are clustered using a K-means clustering approach and then the NB Tree is proposed to extract wind power forecasting probabilities. These probabilities are really helpful to predict the meaningful features from uncertain dynamic behaviour of wind speed. The main contributions of this paper are as follows:

- To capture the standard data, the wind data is normalized using mean and standard deviation.
- To categorize the related features, the K-means clustering method is used to group the real-time wind features.
- NB decision tree (NBT) is used to extract probabilities to predict MAE and RMSE for accurate wind power forecasting.
- Comprehensive comparisons are made with state of the art methods.

The remaining part of the paper is organized as follows. Section 2 explains the important literature work, and Sect. 3 describes the proposed methodology, including subsections, the comprehensive experiments including comparisons are shown in Sect. 4 and finally, Sect. 5 explains the conclusion.

2 Literature Review

Recently, mostly researchers worked on hybrid approaches to take the benefits of combined techniques. Normally, individual methods have lower performance than hybrid based techniques. The hybrid methods are classified into main sections [15]. First, the hybrid method computes aggregating factors of each technique and afterwards, estimate the cumulative value for weighing based predictions. A hybrid method based on distributed and rational grey features of wind speed used to estimate the average weighting values [16]. It merged two techniques such as Least Square Support Vector Machine (LSSVM) with the Radial Basis Function Neural Network (RBFNN). The outcomes exposed that the hybrid method retrieves better forecasting predictions as compared to a single technique

for short-term wind features. Artificial intelligence and negative constraint theory based hybrid method are proposed to forecast wind power [17]. The chaos optimization and genetic algorithm are applied to extract weighted features. It is given that, the hybrid method enhanced the forecasting accuracy by merging the meaningful features from each procedure. The preprocessing steps are helpful to get the features from nonlinear wind data. Further, these steps are used to transform the highly correlated data into linear and normalized form. Recently, the decomposition methods merged to form algorithm are widely used. A hybrid approach contains four different methods is designed to predict power in multi-step wind features, such as Wavelet Decomposition, Wavelet Packet Decomposition, Empirical Mode Decomposition and False Ensemble Empirical Mode Decomposition [18]. The Extreme Learning Machine (ELM) method is utilized for wind power forecasting and classification. The combination of three methods such as Beveridge-Nelson decomposition (BND), Relevance Vector Machine (RVM) and Anti Lion Optimizer (ALO) are used for efficient power predictions [19]. The time series method is applied to convert the nonlinear information into deterministic linear features. After that, the BND method is applied to excerpt the standardized stochastic features. Finally, the RVM is utilized to forecast the wind power from the stochastic features. To examine the efficiency, the suggested approach is used to forecast wind power from hourly based wind data using Xinjiang territory in China.

The Fast Ensemble Empirical Mode Decomposition (FEEMD) and MLP models are merged enhance the forecasting accuracy [20]. First, the FEEMD model is utilized to transform the historical wind data various sub-layers. After that, MLP is used to forecast the wind power for these layers efficiently. The BPNN and SVM are useful machine learning algorithms which are merged to investigate the wind power statistically based on probability values. The hard step is to estimate the uncertain features which can help in predictions. The probabilistic features are used to analyze the uncertain behaviour of wind data. The proposed method is investigated on seven wind turbines data. The accuracy shows that the designed approach provides better outcomes for short term wind [21]. The wind data is collected with 24 h' time and further analyzed using State Estimation Based Neural network (SENN). This method utilized the Weighted Least Square State Estimation (WLSSE) for prediction on input and output layers. The resultant score shows that forecasted accuracy is healthier than BPNN [22]. A hybrid approach of Back Propagation (BP) and Stacked auto-encoders (SAE) is designed for predicting wind power. The neural network with SAE is used to extract the effective features from different wind structures, and loss method is applied to mine the best-linked weights. The BP method is utilized to fine-tune

the model for better predictions. The swarm optimized particles based algorithm is proposed to choose the best number of neurons in each neural network layer. The predictions show that the designed approach offers improved results as compared to SVM and ordinary neural network [23]. The deep learning-based algorithms are effective to forest wind power on large scales. The Principal Component Analysis (PCA) is a statistical technique used to extract independent variables in reduced dimensional space without losing the actual information. After that, the Long Short-Term Memory (LSTM) is utilized to forecast wind power using NWP. The performance shows that the proposed approach provides better results as compare to SVM and BP models [24].

The stationary and non-stationary models can be used to investigate the behavior of time series data [25–27]. A hybrid proposed a method to detect and analyze the time series and correlated data. The Simple harmonizable processes (SHP) structure is studied to test the behaviour of time series data. The periodically correlated processes are analyzed and predicted using extensive Monte Carlo tool. The MAE and RMSE showed the competence of the proposed approach [25]. In [28], presented goodness of fit test for nearly cyclo-stationary discrete-time models. The principal technique is focused on estimating the spectral support and applying multiple research. The results of employing the method presented on simulated and real datasets indicate that the applied method works healthy in light of the analysis of power energy. In [29], the asymptotic distribution of the discreet Fourier transform periodically correlated time series is introduced to the deriving hypothesis testing for both the equal treatment of two periodically correlated time series. The analysis of the simulation in Monte Carlo is then presented to examine the efficiency of the proposed approach. The stationary and non-stationary methods are really effective to analyze the dynamic behaviour of the short term, time series and periodical data [30, 31].

The clustering-based methodology can assist to group the uncertain dynamic features of data receiving from a large number of wind turbines. A hybrid method of K-means clustering and bagging neural network is proposed to forecast wind power in short-term data [14]. The hours' based historical data is grouped using K-means clustering. Then the BPNN is configured to design neural network layers and to tackle overfitting problems. The comparisons show that the proposed approach has better accuracy for short-term wind data features. In [32] proposed a hybrid approach of K-mean clustering and deep belief neural network for wind power forecasting based on NWP. To improve the efficiency of the model, the K-means clustering approach is used to group a large number of samples based on NWP data. After that, the deep belief network model is proposed to predict wind power accurately. The probability-based approach is really supportive of predicting the actual probabilities of the

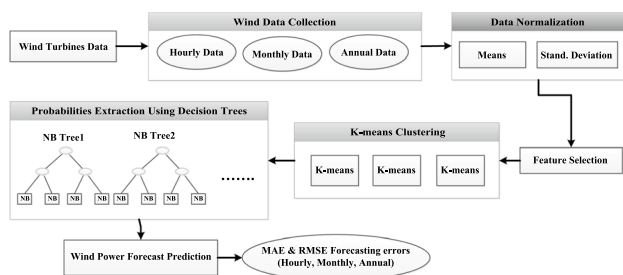


Fig. 1 System Architecture using Probabilistic Decision Tree Approach with K-means clustering

coming dynamic features based on the collected historical data. The probabilistic wind power forecasting is made using gradient boosting decision trees (GBDT) [33]. The GBT approach is designed to develop the wind power quantile regression method. The designed approach is used to deal with the spatial cross-correlation properties of wind power based on transfer learning. The negative transfers are tackled by assigning weights to wind data. It provides improved results using probabilities estimation of each attribute.

3 Proposed Scheme: Probabilistic Decision Tree using K-means Clustering

Wind power predictions of different sorts of interests including the beginning from traditional points forecasts, then progressing to univariate probabilistic forecasts representing wind power production at a fixed location for defined lag time. After that, the univariate probabilities explore to multivariate space–time patterns. The data mining methods are broadly utilized for wind classification and forecasting. We proposed the hybrid approach of K-means clustering and probabilistic decision tree for efficient wind power forecasting from uncertain behaviour of wind data. The complete system architecture is shown in Fig. 1.

3.1 Feature Selection and Normalization

First, the dynamic historical data is collected from a large collection of wind turbines based on hourly, monthly and yearly. Each turbine has different characteristics, such as capacity, capacity factor, wind speed, etc. Similarly, the collected wind has different type features in terms of longitude, latitude, wind speed, direction etc. The dynamic features of wind turbines data are mainly uncertain, which may contain noisy information which is not useful for prediction. For instance, more input variables may transfer more distinguishing information, but practically, unnecessary variables are disposed to many ambiguities. Hence, choosing the appropriate variables can help accurate power forecasting.

After the selection of suitable variables, the data normalization is the second essential step. There may be some attributes containing higher values than the others. Due to this, it can dominate the lower values, which may affect prediction accuracy. For efficient clustering making, all the values should be uniform so that one cluster may not reflect the whole information. We used two main statistical methods, such as mean and standard deviation. We normalized all the values so that the mean average and standard deviation of approximately zero and 1, respectively [34–36]. It creates a level playing field through which we can handle the higher ranges of the dynamic wind features. The mean is the sum of all samples variables ($x_1 + x_2 + \dots x_n$) divided by the total number of the value n calculated using Eq. 1.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n (x_i) = \frac{x_1 + x_2 + \dots x_n}{n}, \quad (1)$$

where \bar{x} is the mean value, n is the total number of observation and x_i is the subsequent samples for each feature in Eq. 1. The standard deviation is the measurement of the dispersion of wind features. It explains how different values in a group are spread out from the mean value. For instance, the low and high standard deviation means that the distribution of values is close or far from the mean, respectively. It can assist us in estimating the clusters more efficiently and accurately. The standard deviation is calculated using Eq. 2.

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2)$$

where $(x_1 + x_2 + \dots x_N)$ are perceived values computed from samples, \bar{x} is the mean average calculations of these observations and, N is the total number of observations in the sample. The mean and standard deviation extract normalized features from dispersed dynamic data. The standard deviation represents the dispersion of the distribution, which may indicate the level of uncertainty of the prediction. It can show the overall distribution of complex wind data characteristics which support us to analyze the uncertain features of hourly, monthly and annual wind turbine data.

3.2 K-means Clustering

The wind speed data is quite dynamic and uncertain, which mainly relate to the grid covered area, direction and time. It may be possible that wind speed can be high in some specific time while low in some other intervals. Due to this, wind characteristics may produce inconsistent values. The clustering method can categorize the collected values in the related group, which can solve the problem of uncertainty. As it makes groups from different intervals of wind data so, it can overcome the missing

values, and a cluster can reflect the actual impact of the related wind data features. By doing this, we can easily identify the effectiveness of each cluster for wind power forecasting. Clustering assumes to be unsupervised learning, which doesn't need tagging in the training set. In this paper, we used clustering method to extract groups of similar patterns to handle large collection of dynamic data efficiently. Many researchers [14, 37, 38] have suggested the K-means cluster algorithm for dynamic and uncertain wind data. This method is based on splitting, which commonly evaluates the similarity by computing distance. The main idea of the K-means method is to select k center points randomly and divided the data using distance value. The Euclidean distance method allocates each data to its closest center value P_k , as shown in Eq. 3 [14, 32].

$$P_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i^k \quad (3)$$

where x_i^k is the i -th data point of the cluster k and, N_k is the number of points in the corresponding cluster. The center value of each cluster requires to be updated until it does not variate further. The normalized wind data is distributed into k the number of clusters such that data points in those clusters have high similarity. The clustering algorithm has the following main steps:

- (1) The k objects are chosen as the early clustering centers from the data which achieve N objects.
- (2) The closest cluster is classified by calculating the distance value among objects and centers of clusters.
- (3) The cluster's mean value is calculated continuously and update the center of the cluster in the cycle.
- (4) Step 2 and 3 are repeatedly executed until the center value of each cluster updates no more.

The clustering process is carried out on the basis of steps 1 and 2 on the large scale of wind datasets (hourly, weekly, yearly). The training sets are prepared for further evaluation by using these clusters. We used the optimum number of k , which achieved better accuracy. The wind data is grouped into the different clusters gathered from the historical samples according to the above steps. Such clusters are used for the creation of training sets to implement probabilistic modelling. The K-means clustering is simple to implement as it just needs K value to select the centroid and distance value to create each cluster of similar patterns. We used real-time wind turbines data at different time intervals such as hourly, monthly and annual. So, each type of wind data may have a specific type of features which needs to be categorized in the same K-means cluster.

3.3 Probabilistic Decision Tree

The NB tree composed of C4.5 and NB algorithms which are used to mine the most significant variables from wind turbines data. As the wind turbines data are uncertain and dynamic in nature, so a probability-based approach can be really helpful to estimate the effect of the coming wind speed on the basis of particular wind characteristics. The NB tree is further described into following main sections.

3.3.1 C4.5

C4.5 is a decision tree that categorizes data into clusters and derives a dataset rule. The decision tree is represented in the form of a binary tree which is useful for classification purposes. It is made up of the root, the split, and the nodes of the leaf.

The root node signifies the classification start point. The working of the algorithm starts from the root node. The node on which a split into two clusters can play an important role which follows if-then procedure. The leaf node offers the final wind data classification. The if-then rules are used to trace the link path from the root to the leaf node. The C4.5 algorithm follows two main steps, such as the growth of tree and pruning, to develop a decision tree. The decision tree grows to reduce the data spreading by dividing it two clusters in each cycle. Given that a set of input variables, we construct subsequent probabilities for each class of cluster among a set of output variables. C4.5 reduces the contamination index that involves the variance of data at the node. When the specific index approaches zero, then all the mined data becomes the same as it applies the information entropy procedure [39–41]. Mathematically, the information entropy is given in Eq. 4.

$$info(t) = - \sum_{j=1}^n \{p(j|t) \log_2 p(j|t)\} \quad (4)$$

where $info(t)$ represents the entropy at a node t , $p(j|t)$ denotes the ratio of j -th class of the compiled samples at the note t . The impurity reduction is computed by the subtraction of entropy between the parent and the child nodes. Mathematically, it is given in Eq. 5.

$$Gain(t) = info(t) - info(t_L) - info(t_R) \quad (5)$$

where, $info(t_L)$ is the left of the child node in the corresponding tree. Further, the gain and splitting ratios are given in Eqs. 6 and 7, respectively.

$$Gainratio(t) = \frac{Gain(t)}{Split Info(t)}. \quad (6)$$

$$\text{Split Info}(t) = \sum_j^c \frac{N(t_j)}{N(t)} \times \log_2 \frac{1}{N(t_j)/N(t)}. \quad (7)$$

where, $N(t)$ is the cumulative number of data at t , $N(t_j)$ is the number of j -th class at the node t and, C is the number of the specific class. The C4.5 algorithm contains the following steps:

- (1) For each attribute a , search the normalized information gain ratio from splitting on a .
 - (2) If a is the best attribute with peak normalized information gain, then mark the a as decision node on which split occurs
 - (3) Extract the sub-nodes of a splitting node then make them as children nodes and so on.
- (1) Select the start conditions.
 - (2) Calculate the clustering data and splitting node value where the split occurs.
 - (3) Prune the tree to estimate the optimum point and the cross-validation error.
 - (4) Input test variables to a tree and pinpoint the leaf nodes.
 - (5) Predict the one step ahead wind power with NB algorithm at each leaf node.

3.3.2 Naïve Bayes

After developing the decision tree using C4.5 then, the NB algorithm is applied on terminal nodes to extract probabilities for each wind data feature in a specific cluster. The NB is a probabilistic algorithm that uses Bayes theorem with naïve independencies. The NB is the most suitable algorithm for uncertain dynamic-wind data as it works on estimated probabilities. So, it can efficiently calculate the prediction probability of the coming wind data signals. It has the power to tackle the arbitrary number of independent variables based on predicted probabilities from historical data. We applied the NB model on each leaf node in a decision tree to extract probabilities for wind data features. Mathematically, the probability for each independent variable is given in Eq. 8 [42, 43].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (8)$$

where $P(C_i|X)$ represents the conditional probability of C_i in X , $P(X)$ is the probability for the independent variable X , $P(X|C_i)$ denotes the conditional probability for X in C_i . The C and X already defined in C4.5 algorithm. The Bayes theorem is used to label the new variable X as the class level C_i to obtain the maximum subsequent probabilities using Eq. 9.

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i) \quad (9)$$

The x_k is further calculated using Eq. 10.

$$P(x_k|C_i) = \frac{1}{\sqrt{2\pi\sigma_{C_i}}} \exp\left(-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}}\right) \quad (10)$$

where, $P(x_k|C_i)$ is the conditional densities for each variable, μ_{C_i} and σ_{C_i} denote the mean and standard deviation for each conditional probability C_i . The NB model efficiently

classifies the wind data due to the conditional probability for each class which is also known as class densities computed distinctly for each independent variable. By doing this, the NB diminishes the high dimensional density jobs to one-dimensional kernel density approximation. The NB tree algorithm is represented in the following steps:

The NB model is operated on the leaf nodes of each decision tree. The wind data features are first processed by decision trees, and then the NB model is applied to extract the probabilities for each leaf node. These probabilities values are used to forecast wind power for each clustered feature.

3.4 Evaluation Measures

The proposed method is evaluated by two most popular matrices such as Mean Absolute Error(MAE) and Root Mean Square Error(RMSE). These matrices are used to analyze the power forecasting accuracy for each dataset. The MAE is the average value of the estimated errors in a set of forecasting values. This value is the absolute difference between predicted and actual observations with equal weight. Mathematically the MAE defined in Eq. 11 [44, 45].

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (11)$$

where y and x are two coordinates and n is the number of data points occurs between two coordinates. The RMSE is the quadratic of the average of the squared subtraction between predicted and actual values. The RMSE is given in Eq. 12.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (12)$$

These two measures are used to analyze the performance of the proposed approach.

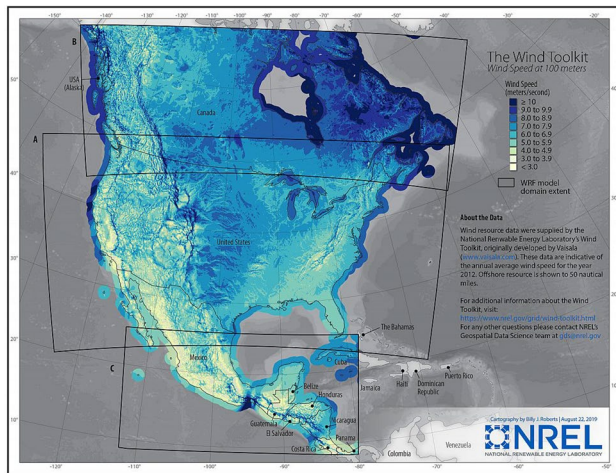


Fig. 2 The wind turbines architecture compiled by NREL

4 Experiments

The proposed hybrid approach is analyzed on different datasets comprehensively for efficient wind power forecasting. First, the useful variables are selected from wind data and then apply the K-means clustering algorithm to mine the related group of wind data signals. After that, we applied the NBT model to classify each feature based on the decision trees and their mined probabilities. The detailed experimentations are provided in the subsequent sections.

4.1 Datasets

Three real-world datasets are collected from NREL¹ (US) database, which comprises the hourly, monthly and yearly wind turbines data. Each dataset has particular characteristics as follows: The hourly dataset is compiled in a range of 20–160 m distance on the ground with a period of 1, 4 and 6 h, respectively. It includes the wind data of 126,000 wind farms sites with various meteorological factors. The monthly wind dataset is gathered from Hawaii area in the US with a mean of 2 km each for each grid in the month of January. This monthly data includes the cumulative average of 17 years Modern-Era Retrospective analysis for Research and Applications (MERRA) based on time series compilation from various wind turbines. The annual dataset is gathered from offshore wind statistics geodatabase that taken distinctive wind speed factors for the Hawaii territory. The real-world historical data is examined by the MERRA for 17 years distinctive meteorological parameters from nearly 2 km distance for each grid. The wind turbines located in different regions with speed ranges taken from NREL is

¹ <https://www.nrel.gov/analysis/>.

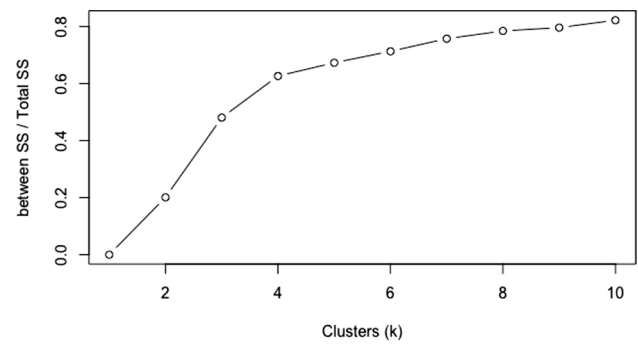


Fig. 3 K value selection for analysis of 10 clusters in hour wind dataset

shown in Fig. 2. The wind speed is uncertain and dynamic in nature, and it mainly depends on some specific variables. We collected the most effective variables from actual wind turbines data, such as grid identification (id), wind speed, latitude, longitude, wind direction, covered areas as used area, capability as a capacity factor, etc. Each grid has a unique id which provides the specific grid statistic for the given variables. These variables are gathered from hourly, monthly and yearly datasets for wind power forecasting analysis.

4.2 Result Analysis

The raw data is collected from different wind turbines and organized based on hourly, monthly and annual. We selected the useful variable which can carry worthwhile information from the vast collection of wind data. As the data variables are in different ranges, which means that high magnitude variables may dominate the lower variables. Thus, it may affect the overall prediction accuracy. We used mean and standard deviation techniques to transform all the variables in level playing fields. The normalized features are then used for extracting clusters. We applied the K-means clustering algorithm to group the related normalized wind data features. Here, the difficult task is that, what is the optimum value for K in different mining sizes of clusters. For this purpose, we conducted an experiment to convert all data to the same number of clusters in order to evaluate the difference between them. Figure 3 shows the 10 cluster ranges for the hour wind dataset to select the K value. The K value for a number of clusters up to 10 is shown horizontally while the division of between Sum of Squares (SS) and total SS is shown vertically. The curve denotes the difference for K between a group of clusters. It can be seen that up to 3rd cluster the K covers the large areas and behave differently. After 3rd cluster, the range is very low and behave almost similar. So, according to this experiment, the best value is K = 3 for hour wind dataset. The same experiment is applied

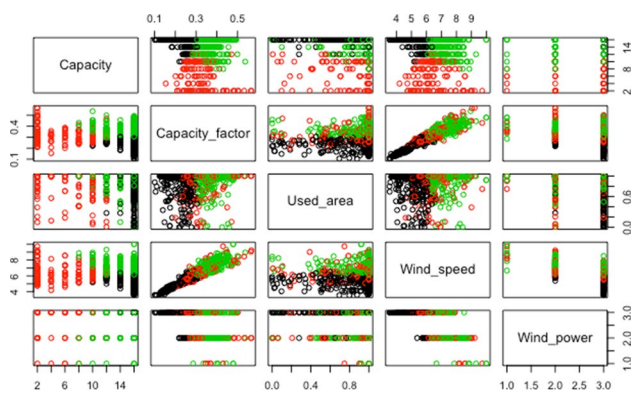


Fig. 4 K-means clustering with K=3 for hour wind dataset

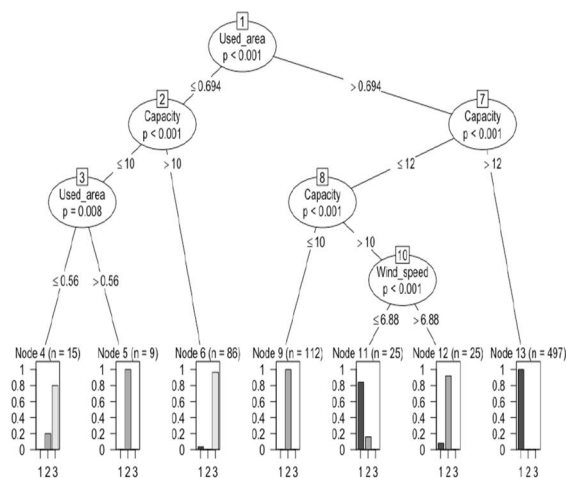
Table 1 K-means clustering details for each variable in hour wind dataset

K-means cluster	Capacity	Capacity_factor	Used_area	Wind_speed
1	-0.0400	-0.2606	-0.3046	-0.2543
2	0.5416	-1.3590	0.1604	-1.3684
3	-0.2342	0.9765	0.2460	0.9746

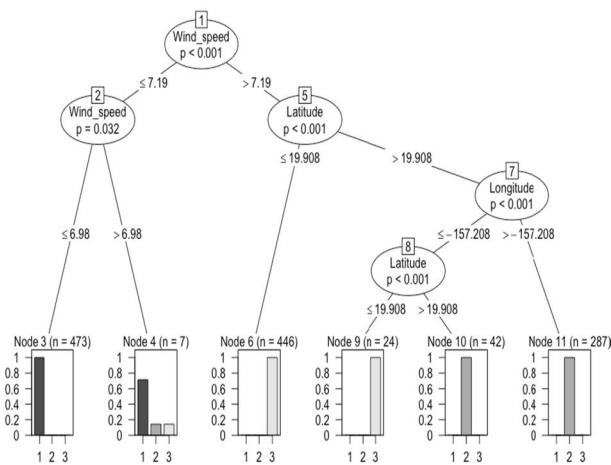
for monthly and annual wind dataset and got the same value for these datasets such as K=3. The K-mean clustering algorithm is applied on all three datasets and extracted the clustered features from uncertain wind turbines data.

Figure 4 shows the clustering visualization for hour dataset. The black, red and green color shows the three clusters captured from hourly wind data. The capacity, capacity_factor, used_area and, wind_speed are considered as input variables, and wind_power is the response variable. The wind power values are categorical, so it seems in straight lines with respect to each variable. The correlation between wind speed and capacity factor shows in linear form, which means these values are directly proportional to each other. The black and green data points are quite high as compared to red data points which means that these two clusters cover the maximum data. The same clustering experiment is applied to all three datasets.

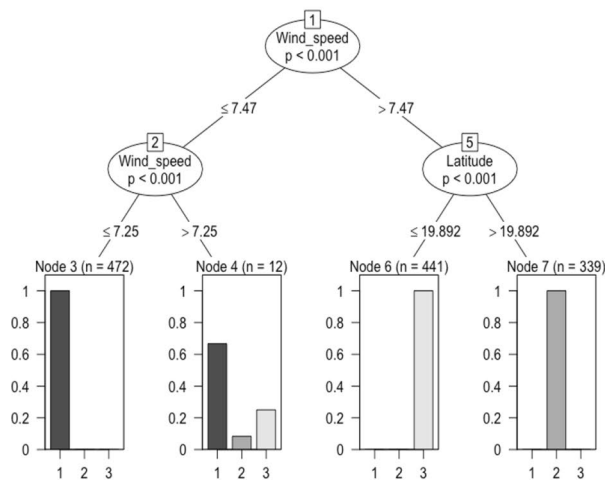
Table 1 shows the number of clusters for each variable for hour wind dataset. For efficient clustering, the difference among clusters should be high for each variable. On the other hand, the difference should be lower among different variables in each cluster. The wind_speed has -0.2543, -1.3684, 0.9746 for cluster 1, 2 and 3, respectively. It means that the wind_speed variable can play a significant role in wind power forecasting as the difference among these values is high for each cluster. Similarly, the capacity_factor can also contribute well, but the capacity variable is the lowest effective variable because it has minimum the minimum



(a) Hourly Wind data



(b) Monthly Wind data



(c) Annual Wind data

Fig. 5 Probabilistic Decision Trees for wind datasets using C4.5 and NB model, i.e. (a) Hourly Wind data (b) Monthly Wind data (c) Annual Wind data

Table 2 Power forecasting probabilities using C4.5 and NB model against each cluster in hour wind dataset

Prob. 1	Prob. 2	Prob. 3	Capacity	Capacity_factor	Used_area	Wind_speed	NBT with K-means Prediction
0.93	0.07	0.00	16.00	0.29	1.00	6.67	1.00
0.98	0.02	0.00	16.00	0.28	1.00	7.03	1.00
0.00	0.00	1.00	2.00	0.31	0.10	6.03	3.00
0.00	0.00	1.00	4.00	0.32	0.13	6.16	3.00
1.00	0.00	0.00	16.00	0.25	1.00	5.51	1.00
0.92	0.08	0.00	16.00	0.42	1.00	7.64	1.00

difference. In each cluster, the values are very close to each other for all four variable which means that they reflect the related features. For instance, cluster 1 can contribute more in power forecasting as compared to cluster 2 and 3 as all the values in this cluster are very close to each other. The same values are calculated for monthly and annual wind datasets. The C4.5 algorithm is used to extract decision trees for all three datasets, and the NB model is then used on terminal nodes to extract the probabilities for efficient power forecasting. The decision tree for each dataset is shown in Fig. 5. The root node has a higher impact than the other nodes. On each node, the if–then algorithm is used to traverse the path of the desired terminal node. On each terminal node, the NB model is applied to capture the probabilities for each cluster with respect to the traversed terminal node.

For instance, cluster 3 has the highest, and cluster 1 has the lowest probabilities for used_area variable if the condition is true on the left side as compared to other clusters in hour wind dataset. But, if the condition is true for the right side, then cluster 2 has the maximum probability contribution using the NB model. The probabilistic decision tree is quite different, with fewer number nodes as compared to hourly and monthly datasets. It means that the hybrid approach of C4.5 and NB model captured the most influenced nodes which can efficiently contribute to power forecasting. Predicted power forecasting probabilities using C4.5 and NB Model for each cluster are shown in Table 2. The prob. means the probability for each cluster, capacity, capacity_factor, used_area, wind_speed are the hour dataset variables, and the last column shows the corresponding K-means clusters for each record. It shows the probability contribution of for each cluster using the NBT approach. All three predicted probabilities are compared with each other for the corresponding clusters, and then, NBT placed the corresponding cluster according to the maximum contribution of each cluster. For instance, the maximum probability is 0.933 captures from cluster 1; thus, cluster 1 efficiently contributes for the first record. The lowest contribution of probability 3, which has 0.0001 value. Similarly, cluster 1, 3, 3, 1, 1 have the maximum probabilities, and thus, they have more impact on forecasting power for the corresponding records.

Table 3 Cluster-type based comparisons with nbt probabilistic decision tree

Dataset	Errors	Hierarchical	Density-based	K-means
Hourly-based	MAE	0.0608	0.0442	0.0200
	RMSE	0.1730	0.1115	0.0858
Monthly based	MAE	0.0573	0.0133	0.0111
	RMSE	0.0900	0.1100	0.0899
Annual-based	MAE	0.1399	0.0489	0.0443
	RMSE	0.2343	0.2088	0.1594

Table 4 Comparisons of the proposed approach for MAE and RMSE at different training ratios

Datasets	Errors	50%	60%	70%	80%
Hourly-based	MAE	0.0773	0.0644	0.0527	0.0200
	RMSE	0.1730	0.1517	0.1244	0.0858
Monthly based	MAE	0.0413	0.0295	0.0242	0.0111
	RMSE	0.1561	0.1186	0.1047	0.0899
Annual-based	MAE	0.0609	0.0461	0.0380	0.0443
	RMSE	0.2111	0.1842	0.1667	0.1594

To get the optimum wind power forecasting results, we analyzed the NBT approach on different clustering algorithms. We select the most popular clustering algorithms such as hierarchal, density-based and K-means methods, as shown in Table 3. The power forecasting errors such as MAE and RMSE are evaluated for each clustering algorithm on each wind dataset (hourly, monthly, annual). It can be seen that, the K-means provider better prediction results with NBT approach. The MAE and RMSE of the proposed approach with K-means for hourly, monthly, annual datasets have 0.2, 0.0858, 0.0111, 0.0899, 0.0443, 0.1594, respectively. Next, the density-based clustering algorithm performs better with NBT approach as compared to hierarchal clustering. Hence, it is proved that the K-means clustering algorithm outperforms for the respective evaluation matrices, such as MAE and RMSE.

Table 5 Comparison of the mean absolute error (MAE) and root mean square error (RMSE) of the proposed approach with other methods for hourly, monthly and yearly wind turbine data

Datasets	Errors	Random forest	J48	REP Tree	SVM	Ensemble selection	BPNN	Proposed approach
Hourly-based	MAE	0.0499	0.0426	0.0248	0.2279	0.0376	0.0319	0.0200
	RMSE	0.1340	0.1897	0.1362	0.2825	0.1259	0.1229	0.0858
Monthly based	MAE	0.0154	0.0276	0.0172	0.2267	0.0362	0.0325	0.0111
	RMSE	0.1043	0.1662	0.1175	0.2802	0.1356	0.1140	0.0899
Annual-based	MAE	0.1017	0.0495	0.0501	0.2345	0.0560	0.0797	0.0443
	RMSE	0.1861	0.2024	0.1893	0.2938	0.1768	0.1964	0.1594

Table 6 Comparison of the mean absolute error (NMAE) and root mean square error (NRMSE) of the proposed approach with other methods for hourly, monthly and yearly wind turbine data

Datasets	Errors	Random forest	J48	REP tree	SVM	Ensemble selection	BPNN	Proposed approach
Hourly-based	NMAE	0.0080	0.0068	0.0040	0.0366	0.0060	0.0051	0.0032
	NRMSE	0.0215	0.0305	0.0219	0.0454	0.0202	0.0197	0.0138
Monthly-based	NMAE	0.0021	0.0038	0.0023	0.0308	0.0049	0.0044	0.0015
	NRMSE	0.0142	0.0226	0.0160	0.0381	0.0184	0.0155	0.0122
Yearly-based	NMAE	0.0133	0.0065	0.0065	0.0307	0.0073	0.0104	0.0058
	NRMSE	0.0243	0.0265	0.0247	0.0384	0.0231	0.0257	0.0208

On the basis of this experiment, we selected the K-means clustering algorithm, which is more effective for efficient power forecasting accuracy. To analyze the running smooth and effectiveness of the proposed approach, we conducted an experiment to compare the power forecasting values such as MAE and RMSE on different training data ratios. Table 4 shows the MAE and RMSE comparisons on different training data ratios from 50 to 80%. In each cycle, the remaining ratio is the testing ratio to complete the 100% of the total.

For example, for 50%, 60%, 70%, 80% the testing ratios are 50%, 40%, 30%, 20%, respectively. The minimum training ratio provides the lower prediction results while higher training ratio gives better prediction scores. For instance, all three wind datasets provide the lowest MAE and RMSE scores on 50% training ratio, but when we increase the training samples, these scores are growing higher. It can be seen that the highest training ratio such as 80%, the MAE and RMSE scores are much higher as compare to the lower training ratios. This means that, for applying the predictive model, the training ratio should be optimum standard which is 80%.

To evaluate the effectiveness and performance, we conducted an experiment to compare the proposed approach with the popular state of the art wind forecasting algorithms, as shown in Table 5. We selected random forest, J48, Rep tree, SVM, ensemble selection, BPNN as the state of the art approaches. The comprehensive comparison is made for

all three datasets such as hourly, monthly and annual based wind turbines data. The same process is followed, such as extracting significant features, normalized them and then capture the K-means clusters. After that, we apply the NBT hybrid model to extract the MAE and RMSE wind forecasting scores for each method on three datasets. It can be seen that the proposed approach outperforms as compared to other given methods with MAE and RMSE scores for all three datasets (hourly, monthly, annual) as, 0.2, 0.0858, 0.0111, 0.0899, 0.0443, 0.1594, respectively. The random forest, J48, and Rep tree work on decision-based trees to predict the wind power forecast. After our proposed approach, the Rep tree and BPNN provide better MAE and RMSE scores (0.0248, 0.1229) for hourly based wind dataset. Next, the random forest performs better with MAE and RMSE forecasting scores (0.0154, 0.1043) for monthly based wind dataset as compared to the remaining methods. Furthermore, the J48 and ensemble selection provide good MAE and RMSE scores (0.0495, 0.1768) using annual wind data turbine dataset after the proposed approach. This evaluation proved that the proposed approach provides significant results for all three wind datasets.

We investigated the proposed approach for Normalized Root Mean Square Error (NRMSE), and Normalized Mean Absolute Error (NMAE) values among the state of the art methods. It facilitates us to compare the performance of the proposed approach with different scales. The NRMSE value

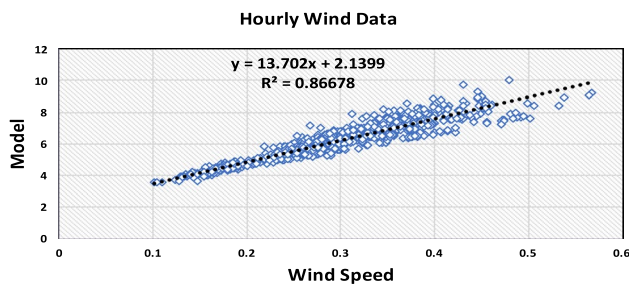


Fig. 6 R-Squared Curves for Hourly wind turbines dataset

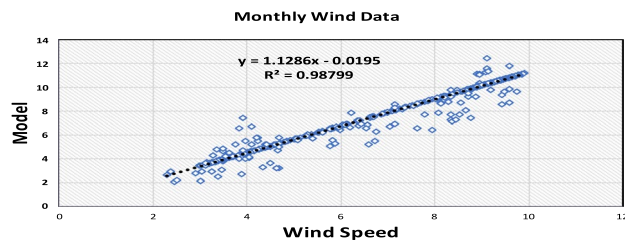


Fig. 7 R-Squared curves for Monthly wind turbines dataset

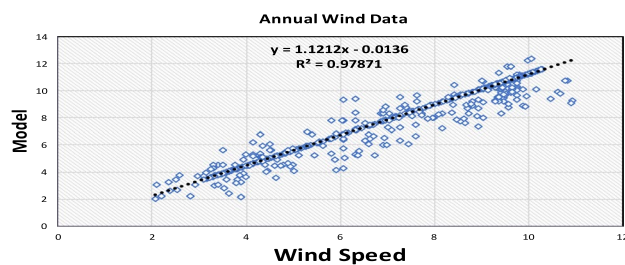


Fig. 8 R-Squared Curves for Annual wind turbines dataset

indicates the effectiveness of the proposed approach. For instance, lower the NRMSE value, better will be the performance of the power forecasting method. Table 6 shows the comparisons of the proposed approach with state of the art methods on the basis of NRMSE and NMAE. We choose Random Forest, J48, REP Tree, SVM, Ensemble Selection, BPNN as the popular state of the art power forecasting methods. It can be seen that our approach outperforms with NMAE and NRMSE values for hourly, monthly, annual based wind turbines data, such as (0.0032, 0.0138), (0.0015, 0.0122), (0.0058, 0.0208), respectively. The R-Squared can be used to show the correlation relationship between the independent and dependent variables of wind data features. R-squared describes the degree to which the variance of one variable determines the variance of the second variable. It is a statistical measure of how related wind data is for the regression line [12]. The R-Squared curves for the hourly, monthly and annual wind turbines are calculated as shown in Figs. 6, 7 and 8, respectively.

For each curve, the vertical line shows the predicted values for the model and the horizontal line indicates the wind data features as observed values. The blue data points show the model and wind data features and the linear regression curve show the fitness of the model based on R-Squared analysis. We extracted R-Squared curves to analyze the fitness of the proposed model. For instance, the closer wind data points to the linear regression line, and more will be the variance and performance of the model. The hourly, monthly and annual wind data have R-Squared values, 86.678%, 98.799%, 97.971%, respectively. It can be seen that the data points in monthly dataset are closer to the linear line as compared to hourly and annual data. Therefore, it gives maximum R-Squared values for the wind speed and model data points. Similarly, the data points in hourly wind dataset are dispersed to the linear line, so it provides less R-Squared value as compared to the other two wind datasets.

4.3 Discussion

The power forecasting for different wind turbines is really dynamic and uncertain. The proposed approach targets two main problems, such as grouping the same properties of wind speed and probability-based wind power forecasting. In order to get the optimal choice for clustering approach, we analyzed three popular methods, i.e., hierarchal, density and K-means. We selected the K-means clustering approach as it gives maximum performance with NBT. The wind turbines data is dynamic and uncertain, and because of this, we used NBT model to forecast the wind power based on their probabilities. The proposed approach is investigated with different level of comparisons which show the effectiveness of the proposed approach, i.e., cluster type based, different data rates and state of the art methods. The MAE and RMSE values show that the proposed approach has a significant gain. Although the proposed approach is not particularly designed for distinctive wind turbines, the K-means clustering approach is still significant and vital in case of large-scale wind farms. Particularly, in one wind farm, the location of wind turbines may be placed in one direction, then the wind speed from such turbines can be classified into one category. By doing this, we can extend the proposed approach into the large scale of wind farms. This approach can boost power forecasting accuracy and diminish the computational cost. The NBT model can be trained for the given features. After that, the trained model can be further used to forecast wind power for the same behavior of wind farms.

5 Conclusion

A new hybrid approach of K-means clustering and probabilistic decision tree is proposed to forecast wind power on real-world wind datasets. Due to the uncertain behavior of wind turbines data, the data variables may have diverse ranges. To get the efficient wind power forecasting scores, all collected variables should be normalized. Due to this, these important variables equally contribute to the proposed approach. We used two statistical methods, such as mean and standard deviation to normalize the features. Next, the K-means clustering algorithm is proposed to extract the group of features having related information. It makes clusters of related normalized features based on the number of K. Then, the NB tree hybrid model is applied to extract the forecasting probabilities for each feature in a cluster. The C4.5 algorithm is used to extract if–then decision trees for each wind dataset, and then the NB model is applied to each terminal node to capture the individual probabilities for each wind data features. The NB tree uses the advantages of both decision tree and the NB model for accurate wind power forecasting. The decision tree is used to pick the best feature at each cycle for the next successive element and then the NB model is used to rank their probabilities as wind data features have the dynamic behavior. To get the optimal wind power forecasting accuracy, we conducted an experiment to compare the working of most popular clustering algorithms with NB tree model. It is proved that the K-means clustering algorithms perform better as compared to other state of the art cluster-based methods. To examine the effectiveness of the proposed approach, we designed an experiment to compare them with the popular state of the art methods in terms of power forecasting scores such as MAE, RMSE. It is proved that; our proposed approach outperforms based on different types of comparisons. The proposed approach can assist in the following:

- The NB tree produces an extremely accurate hybrid model in practice which can significantly improve the wind power forecasting accuracy for three real-world datasets on large scales.
- Each terminate node uses the NB algorithm, which provides highly accurate results on the basis of predicted probability.
- It can deal with the uncertain behavior of real-world wind data with better accuracy.
- The idea of the k-means clustering is significant and effective when a huge wind farm is being developed. In particular, in a wind farm, the position of turbines that fall in one direction, then the wind speed of such wind turbines may be categorized into one group. In such a

way, the approach can be expanded and broadly used in any actual wind farm, that not only improves the power forecast accuracy, but it also decreases computational complexity.

- The decision tree-based classification algorithm is easy to use and explain to others
- Naïve Bayes is a probabilistic model which can handle the real-world wind data with sure fast and accurate.
- The proposed approach handles the continuous and discrete wind data both and even needs less training samples for wind power predictions.

Though the proposed hybrid approach provides promising results for real-world wind turbines data, still it has some problems which need to be tackled in future work. The decision tree is easy to implement, but it needs more time to train the classifier, which may increase the complexity of the model. In future, we will work to handle these types of problems. In addition, the intensity of wind speed may vary on each wind farm regarding the location and wind direction. We will try to address the multi-dimensional clustering problem for each wind farm.

Acknowledgements This research was financially supported by the Education Department of Sichuan Province Foundation (No. 18ZB0273) Bamboo Diseases and Pests control and Resources Development Key Laboratory of Sichuan Province (No. ZL2019004) and Leshan science and technology bureau foundation (No.15NZD100).

References

1. Wan C, Lin J, Wang J, Song Y, Dong ZY (2016) Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. *IEEE Trans Power Syst* 32(4):2767–2778
2. Zhao Y, Ye L, Li Z, Song X, Lang Y, Su J (2016) A novel bidirectional mechanism based on time series model for wind power forecasting. *Appl Energy* 177:793–803
3. Quan H, Khosravi A, Yang D, Srinivasan D (2019) A survey of computational intelligence techniques for wind power uncertainty quantification in smart grids. *IEEE Trans Neural Netw Learn Syst*
4. Kehler J, Hu M, McMullen M, Blatchford J (2010) ISO perspective and experience with integrating wind power forecasts into operations. *IEEE PES General Meeting*, pp 1–5
5. Jiang Y, Chen X, Yu K, Liao Y (2017) Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm. *J Mod Power Syst Clean Energy* 5(1):126–133
6. Demirdelen T, Aksu IO, Esenboga B, Aygul K, Ekinçi F, Bilgili M (2019) A new method for generating short-term power forecasting based on artificial neural networks and optimization methods for Solar photovoltaic power plants. *Solar photovoltaic power plants*. Springer, Berlin, pp 165–189
7. Liu H, Mi X-W, Li Y-F (2018) Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. *Energy Convers Manag* 156:498–514
8. Chen Y et al (2017) Short-term electrical load forecasting using the support vector regression (SVR) model to calculate

- the demand response baseline for office buildings. *Appl Energy* 195:659–670
9. Abedinia O, Amjady N, Ghadimi N (2018) Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm. *Comput Intell* 34(1):241–260
 10. Jafarian-Namin S, Goli A, Qolipour M, Mostafaiepour A, Golmohammadi AM (2019) Forecasting the wind power generation using Box–Jenkins and hybrid artificial intelligence. *Int J Energy Sect Manag*
 11. Chen Q, Folly K (2019) Effect of input features on the performance of the ANN-based wind power forecasting. In: *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*. IEEE, pp. 673–678.
 12. Kumar N, Singh A, Rai N, Chauhan N (2019) Investigation on short-term wind power forecasting using ANN and ANN-PSO. *Applications of computing, automation and wireless systems in electrical engineering*. Springer, Berlin, pp 1103–1116
 13. Chen Q, Folly K (2018) Wind power forecasting. *IFAC-Pap OnLine* 51(28):414–419
 14. Wu W, Peng M (2017) A data mining approach combining \$ k \$-means clustering with bagging neural network for short-term wind power forecasting. *IEEE Internet Things J* 4(4):979–986
 15. Tascikaraoglu A, Uzunoglu M (2014) A review of combined approaches for prediction of short-term wind speed and power. *Renew Sustain Energy Rev* 34:243–254
 16. Shi J, Ding Z, Lee W-J, Yang Y, Liu Y, Zhang M (2014) Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features. *IEEE Trans Smart Grid* 5(1):521–526
 17. Xiao L, Wang J, Dong Y, Wu J (2015) Combined forecasting models for wind energy forecasting: a case study in China. *Renew Sustain Energy Rev* 44:271–288
 18. Liu H, Tian H-Q, Li Y-F (2015) Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Convers Manag* 100:16–22
 19. Guo S, Zhao H, Zhao H (2017) A new hybrid wind power forecaster using the beveridge-nelson decomposition method and a relevance vector machine optimized by the ant lion optimizer. *Energies* 10(7):922
 20. Liu H, Tian H, Liang X, Li Y (2015) New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks. *Renew Energy* 83:1066–1075
 21. Lu H, Chang G (2018) Wind power forecast by using improved radial basis function neural network. In: *2018 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, pp 1–5
 22. Chandra DR, Kumari MS, Sydulu M, Ramaiah V (2018) State estimation based neural network in wind speed forecasting: a non iterative approach. *J Green Eng* 8(3):262–282
 23. Jiao R, Huang X, Ma X, Han L, Tian W (2018) A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. *IEEE Access* 6:17851–17858
 24. Xiaoyun Q, Xiaoning K, Chao Z, Shuai J, Xiuda M (2016) Short-term prediction of wind power based on deep long short-term memory. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, IEEE, pp 1148–1152
 25. Mahmoudi M, Nematollahi A, Soltani A (2015) On the detection and estimation of the simple harmonizable processes. *Iran J Sci Technol (Sci)* 39(2):239–242
 26. Mahmoudi MR, Maleki M (2017) A new method to detect periodically correlated structure. *Comput Stat* 32(4):1569–1581
 27. Nematollahi A, Soltani A, Mahmoudi M (2017) Periodically correlated modeling by means of the periodograms asymptotic distributions. *Stat Pap* 58(4):1267–1278
 28. Mahmoudi MR, Heydari MH, Avazzadeh Z, Pho K-H (2020) Goodness of fit test for almost cyclostationary processes. *Digit Signal Process* 96:102597
 29. Mahmoudi MR, Heydari MH, Roohi R (2019) A new method to compare the spectral densities of two independent periodically correlated time series. *Math Comput Simul* 160:103–110
 30. Mahmoudi MR, Heydari MH, Avazzadeh Z (2019) Testing the difference between spectral densities of two independent periodically correlated (cyclostationary) time series models. *Commun Stat-Theory Methods* 48(9):2320–2328
 31. Mahmoudi MR, Maleki M, Pak A (2017) Testing the difference between two independent time series models. *Iran J Sci Technol Trans A Sci* 41(3):665–669
 32. Wang K, Qi X, Liu H, Song J (2018) Deep belief network based k-means cluster approach for short-term wind power forecasting. *Energy* 165:840–852
 33. Cai L, Gu J, Ma J, Jin Z (2019) Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees. *Energies* 12(1):159
 34. Ackermann G (1983) Means and standard deviations of horizontal wind components. *J Clim Appl Meteorol* 22(5):959–961
 35. Johansson J, Christensen SS (2018) Wind direction variations in the natural wind—a new length scale. *J Wind Eng Ind Aerodyn* 2:2
 36. Demolli H, Dokuz AS, Ecemis A, Gokcek M (2019) Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manag* 198:111823
 37. Ghofrani M, de Rezende M, Azimi R, Ghayekhloo M (2016) K-means clustering with a new initialization approach for wind power forecasting. In: *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, 2016: IEEE, pp. 1–5.
 38. Bhavani M, Vasani SM, Kumar SM, Gokul NP (2020) Wind power forecasting using K-means clustering and convolutional neural network. *EasyChair* 2:2516–2314
 39. Quinlan R (1993) 4.5: Programs for machine learning morgan kaufmann publishers inc. San Francisco
 40. Quinlan JR (2014) C4.5: programs for machine learning. Elsevier, Amsterdam
 41. Mori H, Umezawa Y (2009) Application of NBTree to selection of meteorological variables in wind speed prediction. In: *2009 Transmission & Distribution Conference & Exposition: Asia and Pacific*, 2009: IEEE, pp. 1–4.
 42. Colak I, Sagiroglu S, Yesilbudak M (2012) Data mining and wind power prediction: a literature review. *Renew Energy* 46:241–247
 43. Nam S, Hur J (2018) Probabilistic forecasting model of solar power outputs based on the naive Bayes classifier and kriging models. *Energies* 11(11):2982
 44. Franses PH (2016) A note on the mean absolute scaled error. *Int J Forecast* 32(1):20–22
 45. Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev VA (2013) A survey of forecast error measures. *World Appl Sci J* 24(24):171–176

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mansoor Khan received the B.S. degree in Electrical (Electronics) Engineering from Bahria University, Pakistan, in 2012, the M.S. degree in Power Electronics and Power Drives from Beijing University of Aeronautics and Astronautics, China, in 2017, and the Ph.D. degree in Electrical Engineering and its Automation from Sichuan University, China, in 2020. He is currently serving as an Associate Professor with the Department of Electronics and Materials, Leshan Normal University, China. His research interests include power forecasting, renewable energy and power prediction using artificial intelligence.

Chuan He (S'13-M'18) received the B.S. degree, M.S. degree and Ph.D. degree in electrical engineering from Sichuan University, Chengdu, China in 2011, 2014 and 2018, respectively. He was a visiting Ph.D. student at Clarkson University, Potsdam, NY, USA from 2015 to 2017. He is currently an Associate Research Professor with the College of Electrical Engineering, Sichuan University, Chengdu, China. His research interests include robust optimization on power system operation and planning with renewable energy.

Tianqi Liu (SM'16) received the B.S. and the M.S. degrees from Sichuan University, Chengdu, China, in 1982 and 1986, respectively, and the Ph.D. degree from Chongqing University, Chongqing, China, in 1996, all are in Electrical Engineering. Currently, she is a professor

in college of electrical engineering at Sichuan University. Her main research interests are power system analysis and stability control, HVDC, optimal operation, dynamic security analysis, dynamic State Estimation and load forecast.

Farhan Ullah received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2008, and the M.S. degree in computer science from CECOS University Peshawar, Pakistan, in 2012, and the Ph.D. degree in computer science, Sichuan University, Chengdu, China in 2020. He is currently serving as an Associate Professor with the Department of Software, Northwestern Polytechnical University, Xi'an, China.