Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A CZEKANOWSKI'S QUANTITATIVE INDEXED FUZZY BROWN BOOST DATA CLUSTERING FOR WIND POWER PREDICTION IN SMART GRID

JENCY W.G [1], DR. JUDITH J.E[2]

[1]Research Scholar, Department of Computer Science and Engineering,
Noorul Islam Centre For Higher Education, Kumaracoil, Thuckalay, Tamil Nadu 629180, India.
[2]Associate Professor, Department of Computer Science and Engineering,
Noorul Islam Centre For Higher Education, Kumaracoil, Thuckalay, Tamil Nadu 629180, India.

E-mail: jencywg@gmail.com, judith@niuniv.com

ABSTRACT

Wind power prediction has received greater attention recently for generating energy continuously and sustainably and therefore new functions of electricity networks are necessitated. Smart grid is certainly in the evolution of renewable energy. This paper proposes a novel technique called Czekanowski's Quantitative Indexed Fuzzy Brown Boost Data Clustering (CQIFBBDC) is proposed with the objective of to increase the wind power prediction performance with higher accuracy and acceptable time consumption. The CQIFBBDC technique initially identifies essential energy feature for power prediction from the wind farm database using Czekanowski's quantitative index so that equilibrium between creation and electricity consumption is ensured therefore reducing the prediction time. After that, Fuzzy Brown Boosting Data Clustering of wind power is integrated into smart grids for predicting wind power according to the characteristics of wind turbines. Fuzzy Brown Boosting Data clustering process is carried out to group the input wind data in the smart grid using a FLAME clustering method. The FLAME clustering algorithm performs the clustering process based on neighborhood relationships between the data using a fuzzy triangular membership function. The Brown Boost method combines the results of weak learners to increase the clustering performance with reliability and efficiency of grid operation put forward exorbitant specifications for precise and accurate prediction.
Keywords: *Wind Power Prediction; Feature Selection; Czekanowski's Quantitative Index; Brown Boost Clustering; FLAME Clustering Algorithm; Fuzzy Triangular Membership Function*

1. INTRODUCTION

Wind power prediction and integration of smart grid is the most significant process in renewable energy production. Accurate prediction of wind energy is a major demanding issue for wind power grid problems due to the uncertain and complex nature of wind signals. Wind power is an essential resource for electrical power generation. Many researchers have introduced different techniques for performing wind power prediction by integrating smart grid. Data mining is the method of finding the patterns with the help of machine learning. Machine learning algorithms are utilized to describe the behavior of the dataset, model input features for the expected output, and predict the results for their historical records. This paper proposes a new prediction model based on the ensemble clustering concept for accurate wind power prediction under smart grid environment. The clustering refers to the process of partitioning a dataset into several groups along with the similarity and distance between the data.

A new DE clustering-GRNN model was introduced in [1] to enhance the prediction accuracy for wind power prediction with minimum computational complexity. However, the prediction errors were not minimized with innovative data analysis. A VMD-K means-LSTM model was developed in [2] to conduct short-term wind power forecasting on various scales for improving the precision and reliability. Though the model reduced the error involved, however, accurate prediction was not made with time efficient prediction.

1.1 Problem Statement

In renewable energy production, the wind power prediction is the most important process. But, the prediction accuracy was minimized using traditional methods. But, it failed to reduce the prediction error. Therefore, few research works have been implemented to perform feature extraction and clustering methods. However, the amount of time needed for efficient wind power prediction was not improved. Thus, it failed to minimize the prediction time by using feature selection models. Also, the prediction performance was not adequate. The above issues are overcome by introducing the proposed Czekanowski's Quantitative Indexed Fuzzy Brown Boost Data Clustering (CQIFBBDC) Technique.

1.2 Major Contribution

To overcome the above mentioned conventional issues in the wind power prediction, a novel CQIFBBDC technique is introduced. The contributing factors of CQIFBBDC technique are presented in below,

1) To solve wind power prediction issues as compared to state-of-the-art works, the proposed CQIFBBDC technique widely is used. The CQIFBBDC technique is designed by combining the Czekanowski's quantitative index, Fuzzy brown boost clustering algorithm.

2) To select similar features for smart grid environment with lesser the time complexity as compared to existing works, the Czekanowski's quantitative index used in proposed CQIFBBDC technique. The Czekanowski's similarity coefficient is used for evaluating and arranging the similarity among the attributes in the dataset and the objective function. The similarity coefficient result gets the concise and essential similarity value among 0 and 1 based on the threshold value.

3) To enhance the prediction accuracy as compared to conventional works, the Fuzzy Brown Boosting Data clustering process is introduced in the CQIFBBDC technique. The Fuzzy Brown Boosting Data clustering process is used to group wind farms data by constructing wind power predicting system using FLAME method with the training sample to form strong results on grid scheduling. The fuzzy triangular membership function is applied for assigning the data depending on neighborhood relationships between the data objects. The Manhattan distance measurement is used to calculate neighborhood relationships between data objects under smart grid circumstance.

4) To reduce the false positive rate and minimize the potential loss as compared to state-of-the-art works, brown boost algorithm is applied in CQIFBBDC technique. The weak learner results are combined into strong cluster that in turn significantly improve prediction accuracy and enhance grid potentiality of optimal resource allocation. To achieve this, the weight here is assigned to weak learner by evaluating the potential loss. Then, the margin between the clusters is adjusted to correctly group the data into the different cluster based on the error value. At last, the strong clustering outputs are obtained for limiting the evolution of wind power.

1.3 Paper Outline

The paper is ordered into the different sections as follows. Section 2 introduces the related reviews in the field of wind power prediction and smart grid. Section 3 describes the proposed CQIFBBDC technique with a neat architecture diagram for smart grid. Section 4 experimentally verifies its efficacy of the proposed technique and existing methods using the wind farm database. Section 5discusses the proposed CQIFBBDC technique for smart grid wind power prediction and compares it to other prediction models. Finally, the conclusion section is presented followed by which the references are cited.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. RELATED WORKS

A new hybrid intelligent technique was designed in [3] to enhance the forecasting performance using random forest and artificial neural networks. The designed technique was designed with the implementation of predictive deep belief network and optimized random forest to improve the accuracy of prediction for short-term wind speed prediction. But, the technique did not use any feature selection models for minimizing the wind power forecasting time. A direct quantile regression (DQR) method was designed in [4] for effectively performing the probabilistic prediction of wind power generation. The DQR method was integrating the extreme learning machine (ELM) and quantile regression to improve the high computational efficiency and handle different decision-making issues in power systems. The predictive quantiles of wind power with multiple proportions were not predicted simultaneously via one single linear optimization process.

An enhanced fuzzy clustering method was developed in [5] for forecasting wind power. But the error rate of power prediction was not minimized. The integration of deep learning network and k-means clustering algorithm was introduced in [6] to ensure the prediction accuracy of wind power and to correctly predict the wind power accuracy. But efficient and automatic selection method was not developed to search optimal models for wind power prediction. Two-step ultra-short-term forecast model was combined in [7] with aid of fuzzy clustering and RBF neural network. A fuzzy C-means clustering (FCM) method was designed to group data along with the characteristics of wind turbines for enhancing the power prediction accuracy. BF neural network prediction model was used for achieving the total wind power forecast power. The FCM method similar features were not evaluated, therefore compromising time factor. Probabilistic forecasting of wind power generation was performed in [8] for minimizing power prediction computational complexity by efficient feature extraction. It was used to consider the historical wind-speed and wind-power data. But the probabilistic method failed to provide accurate forecasting results.

An Extreme Learning Machine (ELM) method was introduced [9] to predict wind power in short-term time scale. An ELM method was applied to improve the efficiency and reduce the error rate by using error correction. The forecasting accuracy was not obtained at a significant level. A new hybrid method using Boosting algorithm was designed in [10] to enhance forecasting accuracy. The designed method failed to focus on performing accuracy validation with more existing approaches. An improved Long Short-Term Memory-enhanced forget-gate network method called as LSTM-EFG was developed in [11] to forecasting wind power. The LSTM-EFG method was used to improve the forget-gate effect and optimize the convergence speed. The method did not minimize prediction error at a required level. A hybrid model was designed in [12] to ensure short-term wind power prediction accuracy and reliability depended on the multi-objective optimization. However, effective wind power prediction as not ensured.

Different clustering methods were introduced in [13] to predict the output power of wind and minimize mean square error. Therefore, improve the prediction accuracy. But the prediction time was not evaluated while validating clustering algorithm. A Bayesian nonparametric framework was presented in [14] for short-term wind power probabilistic forecasts. It was used to provide accurate and reliable predictive distribution. However, the framework did not employ any forecasting model for dependent wind power data from various farms. The clustering analysis data processing method was designed in [15] to predict the wind power generation. Though prediction was ensured but at the cost of accuracy.

An integration of Bayesian model averaging and Ensemble learning (BMA-EL) was developed in [16] for forecasting the wind power with higher precision and minimum error rate. However the model involved higher amount of time for wind power prediction. Various machine learning algorithms were designed in [17] to accurately predict the wind power based on historical wind speed data. A data mining approach with K-means clustering and bagging neural network (NN) was introduced in [18] for short-term wind power forecasting with lesser computational complexity. But, effective meteorological forecasting was not performed to improve accuracy. The nonlinear partial least square based prediction was introduced in [19] for short-term wind power prediction to improve the precision. But the feature selection approaches were not used to further improves the training performance of the proposed algorithm.

A multi-objective salp swarm algorithm was designed in [20] for wind speed forecasting to improve the forecasting accuracy with minimum error. But the seasonal variation forecasting was not performed using unadjusted time series data. Smart grids are necessitated with the objective to counter swiftly and significantly to different types of electrical incidents. This is said to be arrived at upon application of sophisticated monitoring and decision support tools where the real time data are collected from the entire power system by the respective operators. In [21], an investigation was made on early detection of anomalies by means of smart meter data. Here, an anomaly detection algorithm was designed to analyze smart meter voltages prefixed at the residential loads. A novel mathematical model was introduced in [22] using Markov to analyze failure factors on the overall system reliability. During the apex period, power supplier on one hand cannot bear to counterbalance all the requirements from customers. In this case scenario, an Incentive-Based Demand Response (IDR) program comes into existence. But, it failed to allocate the optimum points of reliability and costs by using an optimization framework. A user transparent renewal scheme was proposed in [23] with the objective of reducing both the computation cost and delay incurred simultaneously. Yet another distributed strategy was introduced in [24] to minimize cost involved in generation and enhancement in voltage profiles.

The variational mode decomposition (VMD) model was presented in [25] for enhancing the prediction accuracy. However, the prediction performance was not sufficient. Operative strategies for coordinating battery energy storage with wind generation were introduced in [26] to minimize the variability of wind power generation. But, the prediction accuracy was not improved. A new ANN models were developed in [27] for wind power forecasting.

It is of particular interest that the assignment of failure to process involves a technique that is broadly utilized in several aspects of life, insurance companies, aviation and surveillance. Based on this concept, this paper proposes a novel method to evaluate the reliability of wind power prediction integrated with the smart (grid) monitoring systems. The proposed technique is selected based on our objective. In our objective is to increase the wind power prediction performance with higher accuracy and acceptable time consumption based on the feature selection and clustering process.

3. CZEKANOWSKI'S QUANTITATIVE INDEXED FUZZY BROWN BOOST DATA CLUSTERING

Being conscious of what is going to occur in upcoming days can always make people experience secure, therefore weather forecasting is so predominant in people's daily life. The same circumstance is said to persist for electric power systems, when speaking about robust wind power prediction for smart grid. In this section, a robust wind power prediction technique for smart grid using CQIFBBDC technique widely used for solving wind power prediction problems is presented. The proposed CQIFBBDC is based on data mining which consists of the czekanowski's quantitative index and fuzzy brown boost clustering. The proposed CQIFBBDC technique is introduced to solve the wind power prediction problems. With the novelty of Czekanowski's quantitative index is used to choose the similar features for wind power prediction. The czekanowski similarity coefficient is computed among the attributes in the dataset and the objective function. With the novelty of fuzzy brown boost clustering algorithm is applied for performing wind power prediction to improve the prediction accuracy. Manhattan distance measure is used to detect the nearest neighboring data. Local/Neighborhood fuzzy membership Approximation is carried to create the clusters. With the contribution of Brown boost algorithm is applied for minimizing the false positive rate and potential loss. The boosting technique is used to combine the weak learner to for improving the accuracy.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

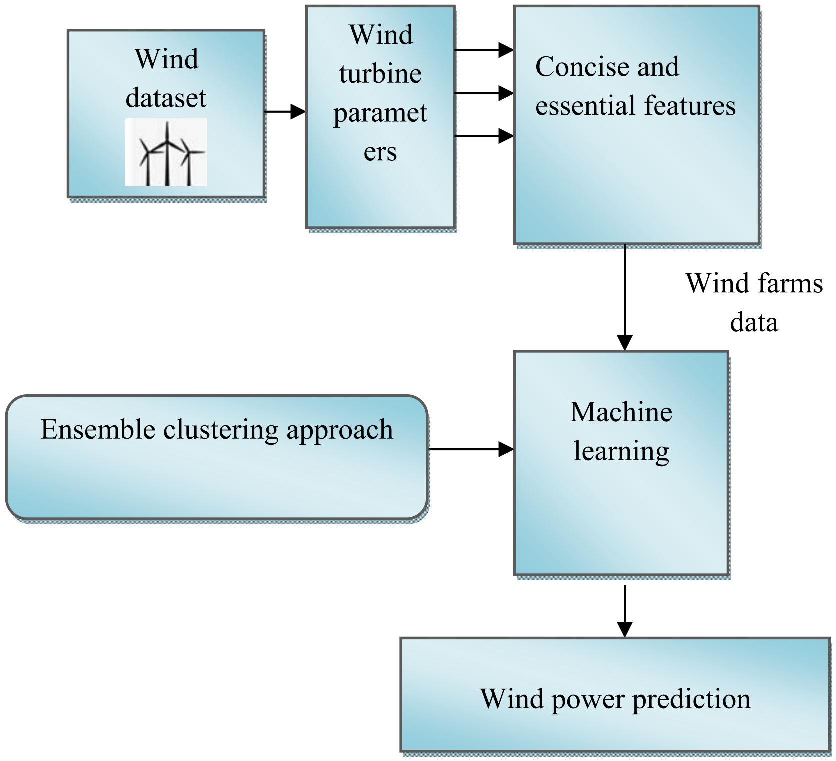


Figure 1: System Architecture Of Proposed Wind Power Prediction In Smart Grid Environment

Figure 1 shows the wind power forecasting model for smart grid using the CQIFBBDC technique. First, wind turbine parameters are collected and organized for constructing the wind data by selecting the most related attributes (i.e. features) that serves as the inputs for the brown boost clustering. Clustering is performed to obtain concise and essential information with aid of Fuzzy Brown Boost clustering technique via wind farms data. Fuzzy Brown Boost clustering technique is used for enhancing both effectiveness and accuracy of wind power prediction.

Figure 1 shows the architecture of the proposed CQIFBBDC technique in smart grid environment. The massive data collected from the wind farms dataset forms as the input to wind power prediction. The training dataset consists of massive amount of data pertaining to wind turbine and not all servers towards robust prediction. The proposed CQIFBBDC technique employs a similarity coefficient measure to find concise and essential features forming as the input towards the boosting ensemble machine learning clustering algorithm. To correctly predict the wind power by using ensemble machine learning clustering algorithm.

3.1 Czekanowski's Quantitative Index

The CQIFBBDC technique initially performs the attributes selection process for accurate wind power prediction for smart grid with minimum time. Czekanowski's quantitative index s used to identify concise and essential features from the dataset. Here, concise and essential features refer to the process of reducing the input to be given to the ensemble technique for wind power prediction. Therefore, Czekanowski's quantitative index employed with the aim of enhancing the prediction accuracy and reducing the prediction time for wind power prediction in smart grid. Let us assume the number of attributes in the wind dataset given as below.

$$a_1, a_2, a_3, a_4, \dots, a_n \in D_w \quad (1)$$

Where, $a_1, a_2, a_3, a_4, \dots, a_n$ denotes an 'n' number of attributes in the wind dataset 'D_w'. Then the similarity is measured for

Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

identifying concise and essential similar features for wind power prediction in smart grid. With the aid of Czekanowski's quantitative index, essential similar features or attributes are evaluated as given below.

$$\beta = 2 * \left[\frac{a_i \cap \vartheta}{\sum a_i + \sum \vartheta - a_i \cap \vartheta}\right] \quad (2)$$

Where, β indicates a czekanowski similarity coefficient, a_i denotes an attribute in the dataset, ϑ denotes an objective function (i.e. wind power prediction for smart grid). The intersection symbol '$\cap$' indicates a mutual dependence representing the statistical dependency of attributes for wind power prediction into smart grids. The similarity coefficient (β) provides the concise and essential similarity value between 0 and 1. The threshold is set to the concise and essential similarity value to find more similar features. The flow chart of the similar feature identification is shown in figure 2.

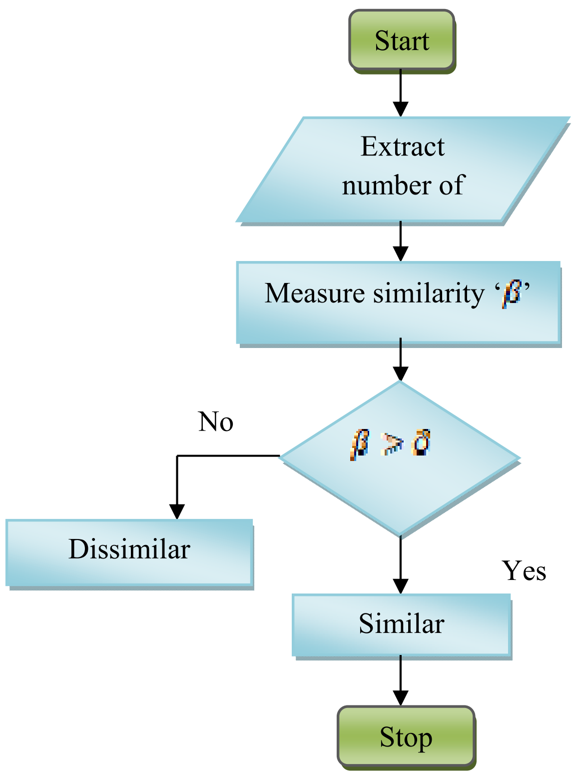


Figure 2: Flow Chart Of Proposed Feature Selection

Figure 2 illustrates the flow process of the Czekanowski's quantitative index-based essential similar feature selection for clustering the input training data instead of processing the large volume of data. This is because of the reason that effectiveness and flexibility of grid process put forward efficient prediction. The step by step process of the Czekanowski's quantitative index is clearly described in algorithm 1. The quantitative index is used to measure and organize the similarity between the attributes in the dataset and the objective function, therefore forming various patterns from data. The feature which is more similar and enhances the prediction paradigms is selected in smart grid based on the threshold value. Finally, the algorithm returns similar features removing the redundant information and obtaining essential information for the next process (i.e., wind power prediction resulting in it minimizes the time consumption).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Algorithm 1 Czekanowski's quantitative index-based similar feature selection

Input: wind dataset, number of features $a_1, a_2, a_3, a_4, \ldots, a_n$

Output: Select similar features for wind power prediction

Begin

  For each attribute $\in D_w$

  Measure the similarity 'β'

  if $(\beta > \delta)$ **then**

   Features are said to be a similar

   else

  Features are said to be a dissimilar

  end if

   Return (similar features)

  End for

End

3.2 Step 2: Fuzzy Brown Boost Data Clustering (Wind Power Prediction)

Another foremost locus is that as a smart grid needs the wind power prediction. Therefore in this work a prediction model is designed in such a way so as to change according to the changes occurring in the environment. Hence, combined prediction is said to more efficient than an individual prediction. The second process of the CQIFBBDC technique is to perform wind power prediction by using the fuzzy brown boost clustering technique. Fuzzy brown boost clustering technique is the machine learning ensemble algorithm that combines numerous base models to provide one optimal predictive model. The ensemble predictive model boosting method yields better accuracy than the normal machine learning technique. Therefore, the CQIFBBDC technique introduces the boosting ensemble technique for accurately predicting the power using wind farms data. The basic structure of the boosting ensemble algorithm is shown in figure 3.

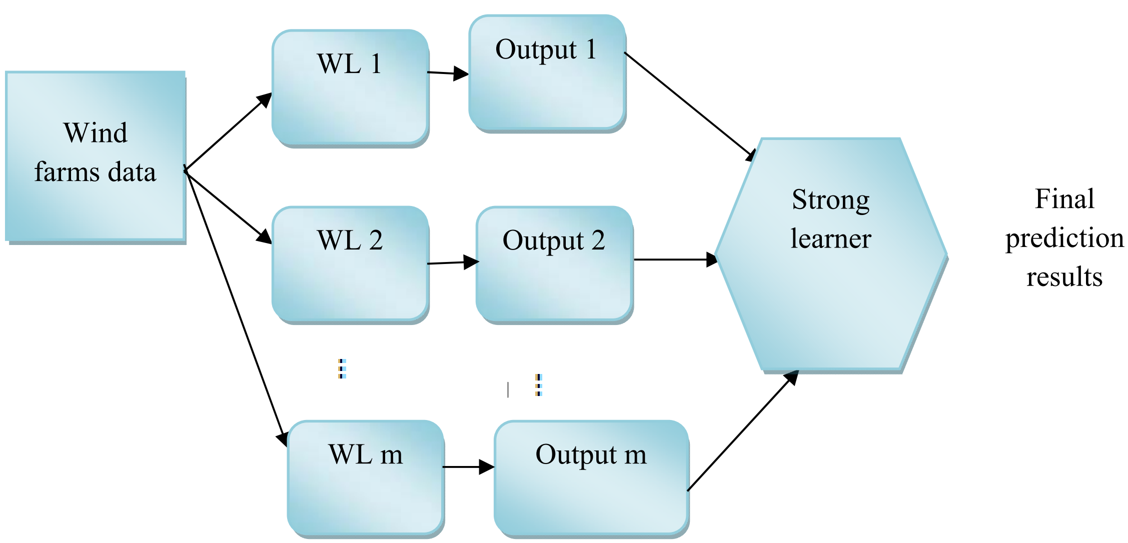


Figure 3: Basic Structure Of Fuzzy Brown Boost Clustering In The Proposed Technique

Figure 3 illustrates the process of the Fuzzy brown boost clustering algorithm. Here, the original training data set is applied as the input of the weak learners (WL). The boost clustering algorithm uses the set as $\{wd_i, Z_i\}$ where wd_i indicates the input wind farms data and Z_i indicates the output of the ensemble technique. The brown boost ensemble technique uses the weak learner as a Fuzzy clustering by Local Approximation of MEmberships

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(FLAME). A data clustering algorithm describes the clusters in the dense part and performs the clustering tasks based on neighborhood relationships between the data. Initially, the number of wind farms data is taken as input $wd_1, wd_2, wd_3, \ldots, wd_m$. The flow process of FLAME clustering is shown in figure 4.

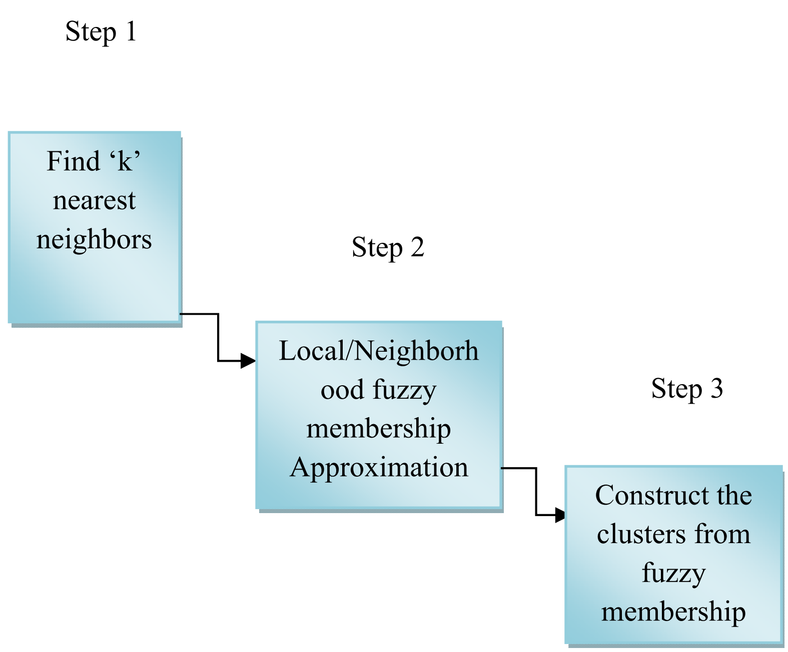


Fig 4: Step By Step Process Of The Flame Clustering

As shown in fig 4, three processing steps are involved in the FLAME clustering. Initially, the 'k' nearest neighbors is identified for each data. Followed by, the Local/Neighborhood fuzzy membership Approximation process is carried out. Finally, the Cluster is formed from fuzzy memberships.

Step 1: For each data, find their 'k' neighborhood using the Manhattan distance measure. The distance is calculated as a given blow,

$$d_t = \sum_{j=1}^{k} |wd - wd_j| \quad (3)$$

From the above mathematical equation (3), 'd_t' indicates a Manhattan distance between the wind farms data 'wd' and the other data 'wd_j'. Based on the distance measure, the 'k' neighboring objects are identified. These data are called as cluster supporting objects, around which to construct the clusters. In this way, the clusters are constructed based on the changing environment. Also, several individual clusters are evolved and accordingly combined prediction is said to be better than individual prediction.

Step 2: In this step, the Local/Neighborhood fuzzy membership Approximation is performed to construct the clusters. By applying the fuzzy concept, each data 'wd' is associated with a membership vector '$\mu(wd)$', in which each data indicates the membership degree of clusters 'm'. For each data in the cluster, the membership vector is consigned through Neighborhood approximation. The membership vector returns the value in the range from 0 to 1, representing how much a data fit into a cluster, or not. In this manner, the prediction model with the aid of approximation keeps adjusting itself therefore ensuring self adaptation a paramount factor

Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

required in smart grid. The proposed technique uses the triangular fuzzy membership function.

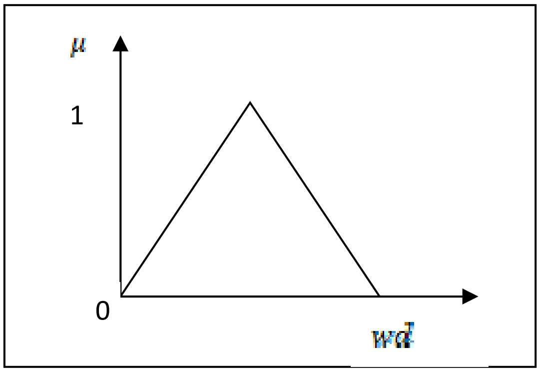


Figure 5: Triangular Fuzzy Membership Functions

Figure 5 illustrates the triangular fuzzy membership function where the horizontal line indicates the data and vertical line indicates the membership function. The membership vector of one data is approximately equated by a combination of its nearest neighbors' memberships which is calculated as given below,

$$\mu_{wd} \approx \sum_{j=1}^{k} \mu_{wd_j} * \varphi \quad (4)$$

Where, μ_{wd} denotes a membership vector of one data object, $\approx$ indicates the approximately equated symbol, $\sum_{j=1}^{k} \mu_{wd_j}$ denotes a combination of its nearest neighbors' memberships, φ denotes a weight ($\varphi = 1$) defining how much each neighbor contributes to the approximation of the fuzzy membership of that neighbor. Then the Local (Neighborhood) Approximation Error is computed as follows,

$$E_A = \left\| \mu_{wd} - \sum_{j=1}^{k} \mu_{wd_j} \varphi \right\|^2 \quad (5)$$

Where, E_A denotes a Local (Neighborhood) Approximation Error is the difference between the membership vector μ_{wd}, and the linear approximation of $\sum_{j=1}^{k} \mu_{wd_j} \varphi$ by its neighbors.

Step 3: Finally, the cluster is formed from fuzzy memberships to assign each data object to the cluster. It has the highest membership and minimum Approximation Error. In this way, the data objects are clustered and used for wind power prediction. The Brown Boost is a boosting algorithm used to improve the clustering accuracy. It is used to convert a weak learner into a strong by combining the weak classifier generated results with minimum potential loss. To further improving the accuracy, the output of each weak learner is summed as follows,

$$Z = \sum_{i=1}^{n} R_i\,(wd) \quad (6)$$

Where, Z denotes an output of the strong cluster, $R_i(wd)$ indicates an output of the weak cluster. After that, the weight is initialized to all the weak clusters as expressed as follows,

$$\vartheta_i = e^{\left(-\frac{(v_i\,(wd)+t)^2}{b}\right)} \quad (7)$$

Where, ϑ_i indicates the weight that the boosting algorithm assigns to the weak learner at iteration 'i', $v_i\,(wd)$, denote the margin of the data, t denotes an amount of remaining time of the weak cluster ($t = b$). After setting the weight, the potential loss for each data with a margin v_i is expressed as follows,

$$l_p = 1 - E\,(g)^{\frac{1}{2}} \quad (8)$$

Where l_p indicates the potential loss, E indicates the error, g is the positive real-valued parameter. Based on the error value, the margin of each weak learner is updated as follows,

$$v_{i+1}\,(wd) = v_i\,(wd) + \sum_{i=1}^{n} \vartheta_i\, R_i\,(wd)\; R_a \quad (9)$$

Where $v_{i+1}\,(wd)$ indicates the updated margins of the data sample, ϑ_i denotes the weight of the weak learner to ensure the final strong clustering results, $R_i\,(wd)$ denotes a weak learner result, R_a represents the actual output of the weak learner. The updated

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

marginal results show that the boosting technique efficiently groups similar data objects into the cluster to improve prediction patterns in smart grids and with false positive data prediction. Therefore, the final strong clustering results are obtained as the linear combination of the weak learner with a lesser error. Then the strong clustering results are obtained as follows,

$$Z = \sum_{i=1}^{n} \vartheta_i \, R_t \, (wd) \tag{10}$$

Where, Z denotes a strong clustering result, ϑ_i indicates the weight, $R_t\,(wd)$ denotes an output of the weak learner. In this way, the ensemble clustering techniques accurately groups similar data into the different clusters. Based on the clustering results, the wind power is correctly predicted with a minimum false positive rate. Algorithm 2 show the description of the Fuzzy brown boost clustering-based wind power prediction with the wind farms data. The boosting technique initially constructs the weak learner with the number of data. For each data, the nearest neighboring data are identified through the Manhattan distance measure. As a result, the neighboring data objects are identified to form the cluster. Then the fuzzy membership is assigned to group the data into clusters. The weak learner results are combined into strong to obtain higher prediction accuracy. The weight is set to the weak learner and calculates the potential loss. Based on error value, the margin between the clusters is adjusted to accurately group the data into the particular cluster. Finally, the strong clustering results are obtained to minimize the false positive rate. The clustering results are used for predicting the wind power at an earlier stage with minimum false positive rate.

Algorithm 2Fuzzy brownboost clustering-based wind power prediction

Input: selected similar features $a_1, a_2, a_3, a_4, \ldots a_n$, wind data $wd_1, wd_2, wd_3, \ldots wd_m$

Output: Improve wind power prediction accuracy

Begin

Construct the number of weak learners

For each data

Find 'k' nearest neighbors using d_t

Assign fuzzy triangular membership vector 'μ_{wd}'

Measure Neighborhood Approximation Error 'E_A'

Assign each data to the cluster

Obtain the clustering results

End for

Combine all weak learners results $Z = \sum_{i=1}^{n} R_t\,(wd)$

Initialize the margin $v_t = 0$

For each $R_t\,(wd)$

Set weight ϑ_i

Calculate the potential loss 'l_p'

for each data

Update margin $v_{t+1}\,(wd)$

Obtain strong clustering results $Z = \sum_{i=1}^{n} \vartheta_i\, R_i\,(wd)$

end for

end for

End

4. EXPERIMENTAL SETUP

In this section, the proposed forecasting technique CQIFBBDC and the three existing methods namely DE clustering-GRNN model [1], VMD-K means-LSTM model [2] and VMD model [25] has been experimented using Java programming language with the aid of world wind farms database obtained from https://www.thewindpower.net/store_windfarms_view_all_en.php. The database comprises of different wind farm data acquired from different countries. The data obtained from the database of a wind turbine are collected for wind power prediction. The database consists of 27 attributes to form wind data for each country. The attributes are listed in Table I.

Table 1: Attribute Description

S.No	Attribute	Description
1	ID	Internal ID
2	Continent	Country continent
3	ISO code	Country code according to ISO 3166.1
4	Country	Country name
5	State code	Area or state code (not available for all countries)
6	Area	Area or state
7	City	City name
8	Name	Wind farm name
9	2nd name	Wind farm 2nd name or wind farm part (if appropriate)
10	Latitude	Wind farm latitude (WGS84)
11	Longitude	Wind farm longitude (WGS84)
12	Altitude/Depth	Wind farm altitude (onshore) or sea depth (offshore)
13	Location accuracy	Yes = accurate location, No = estimated location or city location
14	Offshore - Shore distance	Yes = offshore wind farm (+ shore distance if available)
15	Manufacturer	Turbine manufacturer
16	Turbine	Turbine model
17	Hub height	Hub height
18	Number of turbines	Number of turbines
19	Total power	Wind farm total power
20	Developer	Developer(s) name(s)
21	Operator	Operator(s) name(s)

22	Owner	Owners(s) name(s)
23	Commissioning date	Wind farm commissioning date (format: yyyy or yyyy/mm)
24	Status	Wind farm status
25	Decommissioning date	Wind farm decommissioning date (format: yyyy or yyyy/mm)
26	Link	Direct link to The Wind Power datasheet
27	Update	Update date (format: dd-mm-yyyy)

Based on the information about the wind farm data collected from the database, the wind power prediction for smart grid gets predicted accurately.

5. CASE STUDY AND RESULTS

In order to simulate the proposed CQIFBBDC technique, three different cases, prediction accuracy, false-positive rate and processing time are analyzed and comparison made with two different models, DE clustering-GRNN model [1], VMD-K means-LSTM model [2], and VMD model [25].

5.1 Case 1: Prediction Accuracy

One of the most significant parameters for wind power prediction is smart grid is prediction accuracy. The prediction accuracy (PA) is defined as the ratio of number of wind data objects correctly predicted by means of clustering process with respect to the total number of data taken as input. The power prediction accuracy is mathematically expressed as follows,

$$PA = \left(\frac{\text{Number of wind data correctly predicted}}{\text{Total number of wind data}}\right) * 100 \quad (11)$$

Where, PA denotes a prediction accuracy. The unit accuracy is percentage (%). Initially, the prediction accuracy is discussed with the number of data taken in the counts from 20, 40, 60… 200. For fair computation, a total of ten different runs are carried out with different inputs.

As shown in Table 2, the performance of prediction accuracy is described for the number of wind data taken from the database. The tabulated results indicate that the prediction accuracy of the CQIFBBDC technique is found to be increased compared to the existing methods.

Table 2: Prediction Accuracy

Number of data	Prediction accuracy (%)			
	DE clustering-GRNN model	VMD-K means-LSTM model	VMD model	CQIFBBDC
20	75	70	68	80
40	80	75	69	85
60	82	78	72	87
80	86	84	76	90
100	83	80	75	88
120	85	82	78	93
140	86	81	80	92
160	84	78	76	91
180	85	81	79	93
200	82	79	77	91

The number of wind data correctly predicted is increased by the CQIFBBDC technique using Fuzzy Brown Boost Data Clustering. The Brown Boost ensemble method uses the FLAME clustering to find the neighboring wind data to form the cluster. The FLAME clustering uses the wind forms data like latitude, longitude, and the number of turbines used and their location, Altitude/Depth to be of paramount importance in smart grid. Based on these wind forms information, the output power generation results are accurately predicted in terms of kilowatts (KW). The prediction results of FLAME methods are combined to make a strong learner perform the data clustering process. As a result, the strong learner results are used for accurate prediction of output wind powers. The table values indicate the average prediction accuracy of the CQIFBBDC technique is increased by 7% than the DE clustering-GRNN model [1], 14% than the VMD-K means-LSTM model [2], and 19% than the VMD model [25]. The comparison results of different methods as shown in figure 6.

Figure 6 shows the prediction accuracy result of four different clustering methods. As shown in the above plots, four columns mean that the prediction accuracy of the four methods. In this process, the green color columns are represented by the prediction accuracy of the CQIFBBDC technique, the blue, red, and violet color columns indicate the correctly predicted rate of DE clustering-GRNN model [1], VMD-K means-LSTM model [2], and VMD model [25]. As shown in fig 6, the input data are given as the horizontal axis of the graph and the evaluation results are obtained at the vertical axis.

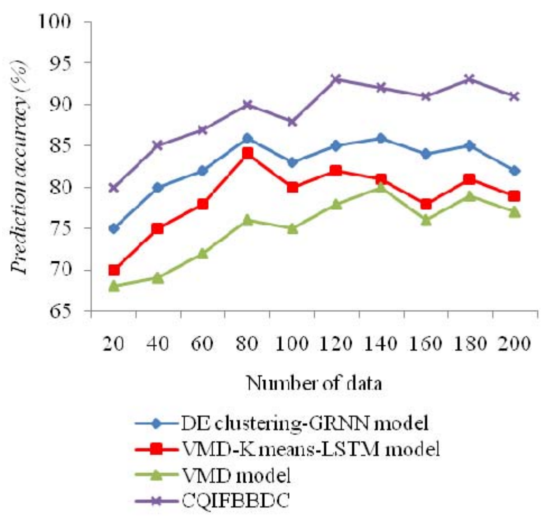


Figure 6: Impact Of Prediction Accuracy For Different Clustering Algorithm

For each iteration, similar counts of data are considered as input for all the methods. The prediction accuracy results from the different algorithms are illustrated in fig 6. The graphical results confirm that the proposed technique has a better performance compared with other baselines [1] [2].

5.2 Case 2: False-Positive Rate

A significant amount of wind data objects are wrongly prediction during wind power prediction in smart grid. Therefore, the second parameter used in our work is the false-positive rate. The false-positive rate (FPR) is measured as the ratio of the number of wind data objects that are wrongly predicted using clustering algorithm and the total wind data. This is mathematically calculated as follows,

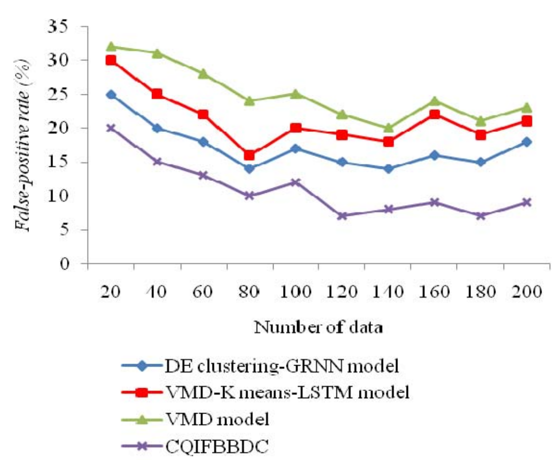
$$FPR = \left(\frac{\text{Number of wind data wrongly predicted}}{\text{Total number of wind data}}\right) * 100 \quad (12)$$

Where, FPR indicates a false positive rate. It is measured in the unit of percentage (%).

Table 3: False Positive Rate

Number of data	False-positive rate (%)			
	DE clustering-GRNN model	VMD-K means-LSTM model	VMD model	CQIFBBDC
20	25	30	32	20
40	20	25	31	15
60	18	22	28	13
80	14	16	24	10
100	17	20	25	12
120	15	19	22	7
140	14	18	20	8
160	16	22	24	9
180	15	19	21	7
200	18	21	23	9

Table 3 shows wrong wind power prediction of smart grid using CQIFBBDC, DE clustering-GRNN model [1], VMD-K means-LSTM model [2], and VMD [25]. The table values prove that, compared with two existing algorithms, the proposed CQIFBBDC exhibits better prediction capability and minimizes the percentage of the wrong prediction.



Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 7: Impact Of False-Positive Rates For Different Clustering Algorithms

Figure 7 illustrates the experimental results of a false positive rate for four different clustering algorithms with respect to number of data collected from the wind farm database. With the wind farms data, the incorrect predictions of four methods are calculated. The CQIFBBDC uses the ensemble method as the forecasting approach and it has better performances than the other clustering algorithms. The reason is that the ensemble method maximizes the margin of the clusters for accurately clustering data by recognizing various patterns of data required for smart grid construction and also enhancing the forecasting accuracy of the wind power. Besides, the FLAME method measures the neighbor approximation based on the fuzzy triangular membership function to minimize the square error resulting in it reduces the incorrect power prediction. The results of CQIFBBDC has compared with two existing base clustering algorithms. It shows that the proposed approach has the best forecasting accuracy among these approaches. Moreover, the comparative analysis indicates that the false positive rate is found to be minimized by 37%, 49%, and 57% when compared to two existing base clustering algorithms [1], [2], and [25] respectively.

5.3 Case 3: Processing time

Finally, the processing time indicates the amount of time taken by the clustering algorithm to predict output power based on wind farm data. The overall processing time of the algorithm is calculated using the given formula,

$$PT = \text{Total number of wind data} * t(\text{predict one data}) \quad (13)$$

Where, PT indicates a processing time of wind power prediction, t indicates the time taken to predicting one data. As a result, the processing time is measured in milliseconds (ms). Table 4 and figure 8 show the processing time for different clustering algorithms versus different data ranging from 20 to 200. The number of data is directly relative to the processing time. In other words, while increasing the number of wind forms data, the processing time gets increased.

Table 4: Processing Time

Number of data	Processing time (ms)			
	DE clustering-GRNN model	VMD-Kmeans-LSTM model	VMD model	CQIFBBDC
20	19	21	23	17
40	22	24	26	20
60	24	26	30	22
80	29	32	35	26
100	30	33	38	28
120	35	37	40	32
140	36	39	41	34
160	40	42	44	37
180	42	44	46	40
200	44	46	48	41

The obtained result validates that the CQIFBBDC exhibits better wind power prediction results for smart grid with minimum time consumption. The Czekanowski's quantitative index is employed in the CQIFBBDC to find similar attributes for wind power prediction from the database since it comprises of dissimilar attributes.

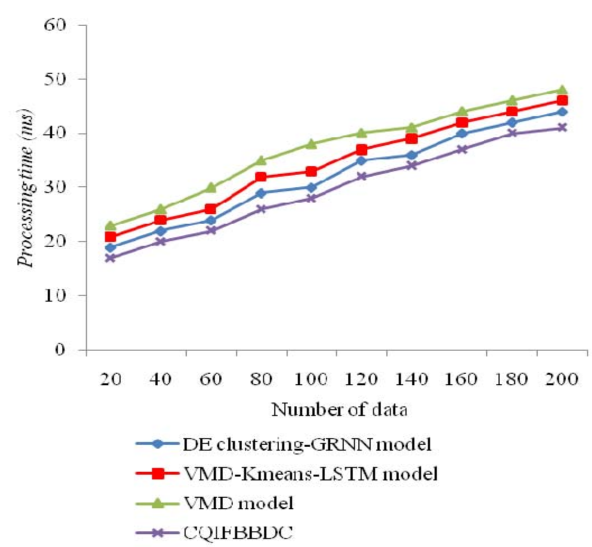


Figure 8: Impact Of Processing Time For Different Clustering Algorithms

While considering multiple attributes for forecasting process, the time gets increased, and also accurate prediction is a challenging one. Czekanowski's quantitative index finds more similar attributes and is used for clustering process. The ensemble clustering methods perform accurate wind power forecasting by considering more similar set of attributes. Therefore, the processing time of CQIFBBDC is said to minimize. In addition to these performances of processing time is compared to the other base clustering models DE clustering-GRNN model [1], VMD-K means-LSTM model [2], and VMD model [25].The average of ten results indicates that the CQIFBBDC minimizes the wind power forecasting time by 8%, 14%, and 21% upon comparison with two methods [1], [2], and [25] respectively.

6. CONCLUSION

An effective data mining technique CQIFBBDC for wind power prediction has been proposed, which comprises the FLAME clustering method and brown boots ensemble method under smart grid circumstance. The proposed CQIFBBDC technique with the construction of smart grid first identifies the related attributes and meteorological information for wind power prediction and therefore ensuring tremendous advantage in integration of large scale wind power. The Czekanowski's quantitative index is a similarity coefficient used to find the significant attributes, therefore ensuring optimal resource allocation from the wind farm database. Fuzzy Brown Boosting Data Clustering of wind power is combined into smart grids to predict the wind power with higher accuracy. Brown Boosting ensemble method applies FLAME clustering to cluster the input wind data in the smart grid. The FLAME clustering algorithm computes the neighborhood relationships between the data depended on the fuzzy triangular membership function. The input variables of the brown boost ensemble are selected to reduce the processing time of prediction. Besides, the wind farms data are clustered to predict the historical power and the ensemble algorithm is applied to optimize wind power forecasting accuracy of the CQIFBBDC technique. A brown boost algorithm is used to combine the weak learner results into strong for achieving higher prediction accuracy. To demonstrate the effectiveness, the proposed CQIFBBDC technique has been tested with the wind farm database. According to the comparison results, it observed that in the clustering stage, the proposed CQIFBBDC algorithm effectively cluster the wind data hence it has a better prediction accuracy. Similarly, the CQIFBBDC algorithm also minimizes the amount of prediction time and forecasting error when compared to traditional clustering algorithms. In future work, the proposed work further extended to improve the wind power prediction by using classification methods.

REFERENCES:

[1] Ying Hao, Lei Dong, Xiaozhong Liao, Jun Liang, Lijie Wang and Bo Wang, “A novel clustering algorithm based on mathematical morphology for wind power generation prediction”, *Renewable Energy,* Vol. 136, Jun 2019, pp. 572-585.

[2] Zexian Sun, Shenshen Zhao, Jingxuan Zhang, “Short-Term Wind Power Forecasting on Multiple Scales Using VMD Decomposition, K-Means Clustering, and LSTM Principal Computing”, *IEEE Access,* Vol. 7, Sep 2019,pp.166917 – 166929.

[3] Zexian Sun, Hexu Sun and Jingxuan Zhang, “Multistep Wind Speed and Wind Power Prediction Based on a Predictive Deep Belief Network and an Optimized Random Forest”, *Mathematical Problems in Engineering*, Hindawi, Vol. 2018, July 2018, pp. 1-15.

[4] Can Wan, Jin Lin, Jianhui Wang, Yonghua Song, Zhao Yang Dong, “Direct Quantile Regression for Nonparametric Probabilistic Forecasting of Wind Power Generation”, *IEEE Transactions on Power Systems*, Vol. 32, No. 4, pp. 2767 – 2778, July 2017.

[5] Lei Li, Xiao-Li Yin, Xin-Chun Jia, Behrooz Sobhani, “Day ahead Powerful Probabilistic Wind Power Forecast Using Combined Intelligent Structure and Fuzzy Clustering algorithm m”, *Energy,* Elsevier, Vol. 192, Feb 2020, pp. 1-46.

[6] Kejun Wang, Xiaoxia Qi, Hongda Liu, Jiakang Song, “Deep belief network based k-means cluster approach for short-term wind power forecasting”, *Energy,* Elsevier, Vol. 165, Dec 2018, pp. 840-852.

[7] Huang Hui, Jia Rong, and Wang Songkai, “Ultra-Short-Term Prediction of Wind Power Based on Fuzzy Clustering and RBF Neural Network”, *Advances in Fuzzy Systems,* Hindawi, Vol. 2018, Sep 2018, pp.1-7.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] Yuan-Kang Wu, Po-En Su, Ting-Yi Wu, Jing-Shan Hong, Mohammad Yusri Hassan, "Probabilistic Wind-Power Forecasting Using Weather Ensemble Models", *IEEE Transactions on Industry Applications,* Vol. 54, No. 6, Dec 2018, pp. 5609 – 5620.

[9] Zhi Li, Lin Ye, Yongning Zhao, Xuri Song, Jingzhu Teng & Jingxin Jin, "Short-term wind power prediction based on extreme learning machine with error correction", *Protection and Control of Modern Power Systems,* Springer, Vol. 1, Jun 2016, pp. 1-8.

[10] Yu Jiang, Xingying Chen, Kun Yu & Yingchen Liao, "Short-term wind power forecasting using hybrid method based on enhanced boosting algorithm", *Journal of Modern Power Systems and Clean Energy,* Springer, Vol. 5, Jan 2017, pp.126-133.

[11] Ruiguo Yu, Jie Gao, Mei Yu, Wenhuan Lu, Tianyi Xu, Mankun Zhao, Jie Zhang, Ruixuan Zhang, Zhuo Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features", *Future Generation Computer Systems*, Elsevier, vol. 93,Apr 2019, pp. 33-42.

[12] Pei Du, Jianzhou Wang, Wendong Yang, Tong Niu "A novel hybrid model for short-term wind power forecasting", *Applied Soft Computing Journal,* Elsevier, Vol. 80, Jul 2019, pp. 93-106.

[13] Yimei Wang, Yongqian Liu, Li Li, David Infield and Shuang Han, "Short-Term Wind Power Forecasting Based on Clustering Pre-Calculated CFD Method", *Energies,* Vol. 11, No. 4, Apr 2018, pp. 1-19.

[14] Wei Xie, Pu Zhang, Rong Chen, Zhi Zhou, "A Nonparametric Bayesian Framework for Short-Term Wind Power Probabilistic Forecast", *IEEE Transactions on Power Systems,* Vol. 34, No.1, Jan 2019, pp. 371 – 379.

[15] Aoran Xu, Tao Yang, Jianwei Ji, Yang Gao, Cailian Gu, "Application of cluster analysis in short-term wind power forecasting model", *The Journal of Engineering*, Vol. 2019, No. 9, Apr 2019,pp. 5423 – 5426.

[16] Gang Wang, Ru Jia, Jinhai Liu, Huaguang Zhang "A hybrid wind power forecasting approach based on Bayesian modelaveraging and ensemble learning", *Renewable Energy*, Elsevier, Vol. 145, Jan 2020, pp. 2426-2434.

[17] Halil Demolli, Ahmet Sakir Dokuz, Alper Ecemis, Murat Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms", *Energy Conversion and Management,* Elsevier, Vol. 198, Oct 2019, pp.1-12.

[18] Wenbin Wu and Mugen Peng,, "A Data Mining Approach Combining K -Means Clustering With Bagging Neural Network for Short-Term Wind Power Forecasting", *IEEE Internet of Things Journal,* Vol. 4, No. 4, 2017,pp. 979 – 986.

[19] Qian Wang, Yang Lei, and Hui Cao "Wind Power Prediction Based on Nonlinear Partial Least Square", *Mathematical Problems in Engineering,* Hindawi, Vol. 2018, June 2018, pp. 1-9.

[20] Ping Jiang, Ranran Li, Hongmin Li, "Multi-objective algorithm for the design of prediction intervals for wind power forecasting model", *Applied Mathematical Modelling*, Elsevier, Vol. 67, Mar 2019, pp. 101-122.

[21] S.W. Yen, S. Morris, M.A.G. Ezra, T.J. Huat, "Effect of smart meter data collection frequency in an early detection of shorter-duration voltage anomalies in smart grids", *Electrical Power and Energy Systems*, Elsevier, Jul 2019.

[22] Mohammad Esmail Honarmand, Mohammad Sadegh Ghazizadeh, Vahid Hosseinnezhad, Pierluigi Siano. "Reliability modeling of process-oriented smart monitoring in the distribution systems", *Electrical Power and Energy Systems,* Elsevier, Jul 2019, pp. 20-38.

[23] Shaomin Zhang, Kai Wang, Baoy iWang, "A user-transparent pseudonym renewal scheme for smart meters in incentive-based demand response programs", *Electrical Power and Energy Systems,* Elsevier, Feb 2019, pp. 49-56.

[24] Nohora España, John Barco-Jiménez, Andrés Pantoja, Nicanor Quijano, "Distributed population dynamics for active and reactive power dispatch in islanded microgrids", *Electrical Power and Energy Systems,* Elsevier, Jul 2020, pp.1-11.

[25] Gang Zhang, Benben Xu, Hongchi Liu, Jinwang Hou, Jiangbin Zhang, "Wind power prediction based on variational mode decomposition multi-frequency combinations", *Journal of Modern Power Systems and Clean Energy*, Vol. 7, Dec 2019, pp. 281-288.

[26] Thais P. Teixeira; Carmen L. T. Borges, “Operation Strategies for Coordinating Battery Energy Storage with Wind Generation and their Effects on System Reliability”, *Journal of Modern Power Systems and Clean Energy,* Vol. 9, No. 1, Jan 2021, pp. 190-198.

[27] Sergio Velázquez Medina, Ulises Portero Ajenjo, “Performance Improvement of Artificial Neural Network Model in Short-term Forecasting of Wind Farm Power Output”, *Journal of Modern Power Systems and Clean Energy,* Vol. 8, No. 3, May 2020, pp. 484-490.