Research Article.
Journal of Intelligent Systems: Theory and Applications 4(2) 2021: 142-154
DOI: 10.38016/jista.889991

Short-Term Wind Power Prediction Approach Based On Bayesian Optimization and Ensemble Learning

Kübra Yazıcı [1*], Semra Boran [2]

[1] Turkish-Deutsche University, Faculty of Engineering, Department of Industrial Engineering, İstanbul/Turkey

[2] Sakarya University, Faculty of Engineering, Department of Industrial Engineering, Sakarya/Turkey

kubra.yazici@tau.edu.tr, boran@sakarya.edu.tr

Abstract

In wind energy studies, predicting the short-term energy generation amount for wind power plants and determining the production offer to be placed on the market play an important role. In this study an hourly short-term wind power estimation of a wind turbine located in Turkey with an installed power of 3600 kW has been made. Estimation results were evaluated on a seasonal and annual basis. New hybrid models have been developed for short-term wind power prediction, consisting of Bayesian Optimization (BO), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Decision Tree (DT), stacking, and bagging algorithms. In the proposed prediction approach, it is aimed to reduce prediction errors by combining different regression algorithms with the BO method and ensemble algorithms. Unlike other wind prediction studies, BO was used for the first time in the hyperparameter selection of the regression algorithms selected as the basic learner in the study. Bayesian optimized decision tree (BO-DT) with the lowest error values among the base learners, and Bayesian optimized gaussian process regression (BO-GPR) combined with bagging and stacking. The efficiency of ensemble learning algorithms was measured by the statistical measurement methods Normalized Absolute Mean Error (NMAE), Normalized Root of Mean Squares Error (NRMSE), and determination coefficient (R^2). According to the results, the bagging method created with the BO-DT took the annual average NRMSE, NMAE, R^2 criteria of 11.045%, 4.880%, 0.899, respectively, and the model with the best performance was selected in terms of both annual and seasonal results.

Keywords: Renewable energy, Wind power prediction, Bayesian Optimization, Ensemble Learning

Bayes Optimizasyonu Ve Topluluk Öğrenmesine Dayalı Kısa Dönem Rüzgar Gücü Tahmin Yaklaşımı

Öz

Rüzgar enerjisi çalışmalarında, rüzgâr santralleri için kısa dönem enerji üretim miktarının tahmini ve piyasaya verilecek üretim teklifinin belirlenmesi önemli bir rol oynamaktadır. Çalışmada Türkiye'de bulunan ve kurulu gücü 3600 kW olan rüzgar türbinin saatlik kısa dönem rüzgar enerjisi tahmini yapılmıştır. Tahmin sonuçları mevsimsel ve yıllık olarak değerlendirilmiştir. Kısa dönem rüzgar gücü tahmini için bayes optimizasyonu, destek vektör regresyonu, gauss süreç regresyonu, karar ağacı, stacking ve bagging algoritmalarının birleşiminden oluşan yeni hibrit modeller geliştirilmiştir. Önerilen tahmin yaklaşımında farklı regresyon algoritmaları ile bayes optimizasyon yöntemi ve topluluk algoritmaları birleştirilerek tahmin hatalarının azaltılması amaçlanmıştır. Çalışmada temel öğrenen olarak seçilen regresyon algoritmalarının hiper parametre seçiminde diğer rüzgar tahmin çalışmalarından farklı olarak ilk defa bayes optimizasyonu kullanılmıştır. Temel öğreniciler içerisinde en düşük hata değerlerine sahip bayes algoritması ile optimize edilmiş karar ağacı ve gauss süreç regresyonu, torbalama ve istifleme ile birleştirilmiştir. Topluluk öğrenmesi algoritmalarının etkinliği istatistiksel ölçüm yöntemleri olan Normalize Mutlak Ortalama Hata (NMAE), Normalize Ortalama Hata Kareleri Kökü (NRMSE) ve determinasyon katsayısı ($\mathbf{R^2}$) ile ölçülmüştür. Sonuçlara göre bayes algoritması ile optimize edilmiş karar ağacı ile oluşturulan torbalama yöntemi yıllık ortalama NRMSE, NMAE, $\mathbf{R^2}$ kriterleri sırasıyla 11.045 %, 4.880 %, 0.899 değerlerini almış ve hem yıllık hem de mevsimlik sonuçlar açısından en iyi performansa sahip model seçilmiştir.

Anahtar Kelimeler: Yenilenebilir enerji, Rüzgar gücü tahmini, Bayes Optimizasyonu, Topluluk öğrenmesi

* Corresponding Author.
E-mail: kubra.yazici@tau.edu.tr

Recieved : 02 March 2021
Revision : 22 May 2021
Accepted : 12 July 2021

1. Introduction

The need for energy is increasing day by day due to the increase in population, as well as the developments in the technology and automation sector developing in recent years. However, providing energy from fossil fuels both harms human health and causes irreparable environmental problems. Increases in population, urbanization and energy demand require sustainable energy management, technological infrastructure and innovative applications. With the developments in the last 20 years, it is estimated that the 2% growth in energy need will increase by 1.3% in the next 20 years, which requires incentives to continuously support the sector with innovative systems, products and investments (Eroğlu, 2019). In the next 10 years, especially alternative energy use will be solar, wind, etc. It will depend on the resources, bringing along important developments in the energy production industry (Öz & Alyürük, 2020). Wind energy is the fastest-growing energy type among renewable energy sources globally and the most invested energy type in the last 6 years. In 2019 approximately 15% of the electricity demand in Europe, in Turkey about 7% is obtained from wind power plants (YEKDEM, 2020). Wind power plant installed power in 2023 in Turkey total power installed power (100,000 MW) 20%. It is aimed to reach a value of 20,000 MW, corresponding to the amount of energy (Şenol & Koç, 2015). Considering the wind potential and the 2023 targets, although the current installed capacity of wind power plant is low, significant developments have been achieved in the last decade. The electricity consumption in Turkey, which reached 290.4 kilowatt-hours in 2019, is expected to reach 375.8 TWh with an annual average increase of 4.8% (ETKB, 2019). This increase in energy consumption increases foreign dependency and causes a current account deficit for countries that cannot produce the energy they consume and meet this need with imports. When examining Turkey's current account deficit, this deficit constitutes a significant portion of energy imports(Bağcı,2019).

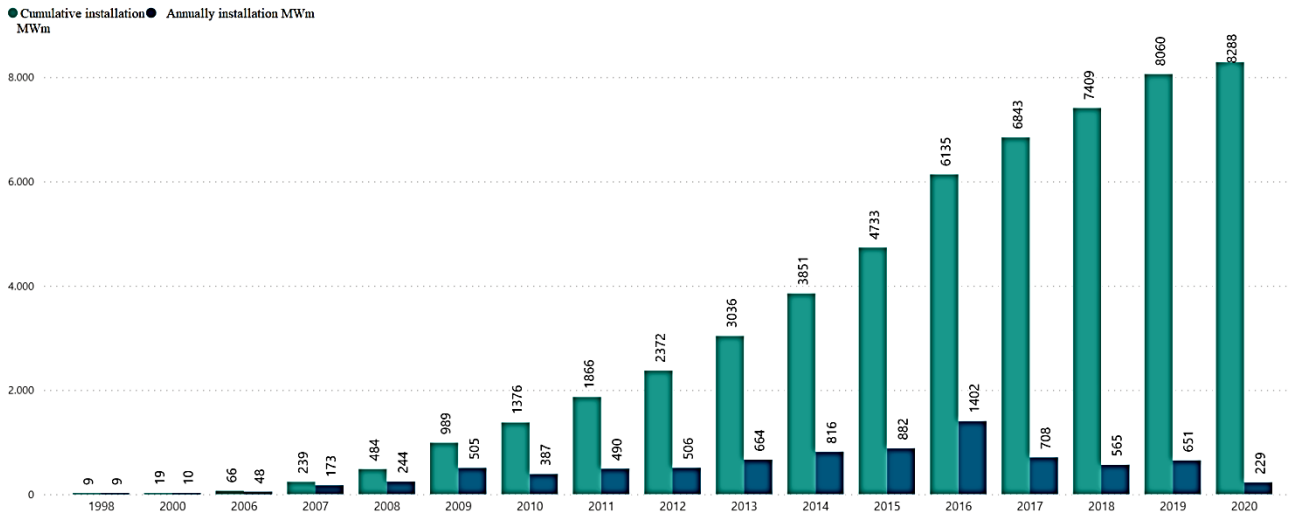

Figure 1. Installed wind power plants by years

In Figure 1, the last twelve years for Wind Power Plants in Turkey (1998-2020) is given to the development of total installed capacity (TÜREB, 2021). As shown in the graph, installed wind power has increased 10 times in the last 10 years and exceeded 8 GW. The ratio of electricity generated from wind energy to total electricity generation in Turkey in 2020 was 8.4%. A ratio of the 8.4% of electricity consumed in Turkey is produced from wind energy. It is aimed to increase this value to 20% in 2023.

Electricity generation with wind power has a highly variable profile. For this reason, energy prediction models based on wind prediction have an important place in the reliable, economical and quality operation of wind energy resources. Wind energy prediction models are used in power systems planning, reserve planning, maintenance and repair planning, and bidding in the electricity market. Thanks to prediction models, power plants can increase their revenues by reducing prediction errors in the day-ahead market, thus reducing energy imbalances in the electricity market and consequently, costs (Karık et al., 2017). Especially, short-term wind prediction play an important role in day-ahead electricity trade, planning the day-ahead electricity system, determining the required reserve amount, and making unit commitment decisions (Kerem, 2018). Thanks to short-term wind prediction, problems such as excessive production planning and allocation of excess reserves can be avoided, reducing operating costs and integrating more wind energy into the system.

Better predictive models for wind power significantly reduce the need for conventional power

plants to control energy. At this point, among prediction models, especially machine learning methods play a major role in the successful integration of wind power generation into electrical networks.

In this study, short-term wind power prediction four models based on the Bayesian optimization algorithm (BOA) and ensemble learning that can assist system operators and producers in more accurate production planning and accurate price proposal in electricity markets are proposed. Ensemble learning algorithms have been preferred in this study because it provides the opportunity to obtain higher performance models by combining more than one regression algorithm. In the parameter setting of regression algorithms, unlike other methods, the BOA, which uses the information obtained from previous experiments and works quickly, is used. Decision tree, Gaussian process regression and Support vector regression, which are frequently used in short-term wind power prediction in the literature, are used as regression algorithms. Among the optimized regression algorithms, Bayesian optimized Gaussian process regression (BO-GPR) and Bayesian optimized decision tree (BO-DT) have been chosen as the base learner for ensemble learning. In bagging and stacking methods, which are ensemble learning methods, different combinations of BO-GPR and BO-DT were tried and the results were evaluated seasonally and annually in terms of NRMSE, NMAE and R^2 performance criteria. The developed models were used for short term wind energy estimation of a wind turbine operating in Yalova. The model consisting of the combination of BO-DT and bagging algorithm, which has the best performance criteria for four seasons, has shown that it is the most effective model in short term wind power with annual average NRMSE 11.045%, NMAE 4.880% and R^2 0.899.

The remainder of this work is organized as follows. Section 2 summarizes the literature on short-term wind power prediction studies. Section 3 and Section 4 introduces the methods used in the study and proposed approach, respectively. Section 5 shows the implementation stages of the proposed models in the study and the analysis results on the determined wind turbine. And finally, Section 6 summarizes the results of the proposed short-term wind prediction models and their contribution to the literature.

2. Literature Survey

There are many studies of machine learning and ensemble learning applications in short-term wind power prediction in the literature. Lee and Baldick proposed 52 artificial neural networks (ANN) models and 5 Gaussian process regressions (GPR) models based on the GPR model and ANN for prediction at 48 hours time horizon. Artificial neural network (ANN) sub-models predicted future wind power over a 48-hour period based on past and predicted wind power data. Parallel to the ANN, GPR predict only from historical wind power data. The most appropriate prediction from more than one prediction value formed for the same hour is determined by the decision process (Lee & Baldick, 2014).

Chen et al. (2014), combined NWP (Numerical weather prediction) model and GPR models to predict short term wind power. Wind power was forecasted based on the relationship between the corrected data after the wind speed data received in the proposed model were corrected with Gaussian process. The data set on which the model has applied consists of wind speed, wind direction, temperature, pressure, humidity and wind power and is in a time period of 10 minutes. The proposed model was compared with classical wind prediction methods such as Multilayer perceptron (MLP) and ARIMA, and an improvement between 9% and 14% was achieved in error values compared to the artificial neural network.

Li et al. (2018), proposed a hybrid model based on Support vector machine (SVM) using Wavelet transform (WT) and Cuckoo search (CS) methods.

In the proposed model, Fourier, Gaussian and polynomial fitting was used to deal with the missing and erroneous data, and then the original signal was eliminated with WT. With the CS optimization, core function of the SVM and the penalty factor are optimized, and the prediction accuracy is increased. The model created has been applied to the data set consisting of hourly wind speed, wind direction, and wind power belonging to China's wind power plant. The past 60 days (1440 hours) data was used for the training, and the next 3 days (72 hours) data was used for the test. According to the results, the proposed model has fewer error values than ARIMA, Support Vector Regression (SVR) and Back Propagation Neural Network (BPNN).

Fu et al. (2019), put forward a SVM model optimized with the improved CS method on the data set consisting of hourly wind speed, wind direction, temperature and wind power. The classic CS method has been optimized to prevent the local optimum's easy capture in large data sets. The proposed model was applied in two data sets, 500 and the other 400 training data, and tested with 40 data. According to the results, the proposed model gave better results than the SVM optimized with classical CS, and it was shown that the number of training data was effective on the prediction.

Ma & Zhai (2019), established a 2-stage hybrid model consisting of WT, Feed-forward artificial neural network (FFANN) and Ant colony optimization (ACO) methods. In the first stage of the model, meteorological parameters such as wind direction, temperature, pressure and wind speed were predicted. In the second stage, wind speed predictions from the first stage were predicted with an ANN optimized with the ant colony and wind power sub-series. Wind power prediction results are obtained by applying the predicted sub-series WT. The proposed model has been more successful than ANN optimized with genetic algorithms, ANN

optimized with Particle swarm optimization in the next 24-hour wind power prediction.

Li et al. (2020), proposed a hybrid model consisting of a SVM and improved Dragonfly algorithm (IDA) has been proposed for short term wind power forecasting. The IDA is used to select the optimum parameters of the SVM. The proposed model was tested on a data set consisting of hourly wind speed, wind direction and wind power in 2017 of a wind farm located in France. In the hybrid model proposed for short-term wind forecasting, the past 6 days (144 hours) data are determined as training data, and the next 2 days (48 hours) data are determined as test data. The proposed model gave better short-term wind power forecasting results than ANN and GPR methods.

Recently, ensemble learning applications have increased in wind power forecasting by combining multiple forecasting methods, offering more generalized performance and reliability.

Heinermann & Kramer (2016), applied the k nearest neighbor (k-NN), Decision tree (DT) and SVM methods to short-term wind power. DT and SVM methods that predict with less error are combined with heterogeneous ensemble learning. With the created heterogeneous ensemble learning models, both an increase in prediction accuracy and a decrease in runtime were achieved.

Ahmad et al. (2018), applied Random Forest (RF) and Extra Trees (ET) method to investigate the effect of ensemble learning methods on improving calculation costs and prediction accuracy in hourly wind power prediction. The results were compared with the SVM. According to the prediction performance results of the methods, RF and ET, which are tree-based ensemble learning methods, gave better results than the SVM. In addition to the prediction accuracy, the ET method has less training and testing time than the other two methods.

Banik et al. (2020), used Boosting, Gradient Boosting and Extreme gradient boosting (XGBoost) methods as predictors, outliers were removed in the data set consisting of hourly data from 2014, and the relationship between wind power and meteorological parameters was determined with the Pearson correlation heat map. According to the results, a strong correlation was found between wind power and wind speed, wind direction, temperature and humidity, and other parameters were not included in the model. Then, 5-fold validation was applied to the models and predictions were made in two different time horizons, 1 hour and 1 week. According to R^2 and RMSE criteria, XGBoost ensemble gave the best results compared to other methods.

Tahir et al (2018), proposed a two-layer stacked ensemble-based model consisting of Random Forest, Support vector machine and Radial Basis Function Neural Network methods for very short-term wind power forecasting. The output values obtained with the support vector machine, random forest and radial basis function neural network models in the first layer of the model are combined with the support vector machine in the second layer and wind power are forecasted. According to the results, the proposed stacked ensemble-based model has been more successful than the classical support vector machine, linear regression, regression tree, random forest in very short-term wind power forecasting.

In this study, SVR, GPR and DT were applied in order to predict short-term wind power on a data set consisting of hourly wind speed, wind direction, temperature, pressure, relative humidity and wind power. In order to optimize the regression algorithms, unlike the optimization algorithms such as CS, ACO, PS, which are used in wind prediction studies in the literature, the BO was used. In order to reduce the prediction error of optimized regression algorithms and provide an improved prediction performance, these algorithms are combined with the ensemble learning methods, bagging and stacking. In the literature, DT and GPR algorithms, which are frequently used as single models in wind power prediction, are combined with the stacking algorithm in the study and different models are proposed.

3. Materials and Methods

3.1. Gaussian Process Regression

Gaussian process regression (GPR) is a kernel-based nonparametric probabilistic model. The Gaussian process estimates the final probability distribution based on a previous probability distribution and updates the previous probability distribution based on training data. A Gaussian process model predicts response variables with the new input vector and training data by defining the relationship between input variables and target variables (Rasmussen & Williams, 2006). Regression model function:

$$y= f(x) + \varepsilon \qquad \varepsilon \approx N(0, \sigma_n^2 \ ) \tag{1}$$

where y is the target variable, and x is the input variable. The average of ε, which is additive noise, is 0 and it is assumed to be normally distributed (Wan & Sapsis, 2017). A Gaussian process mean function is expressed by $m(x)$ and covariance function $k(x,x')$:

$$m(x) = E[f(x)] \tag{2}$$

$$k(x,x') = E[(f(x) - m(x))\ (f(x') - m(x'))] \tag{3}$$

The model is defined as follows:

$$f(x) \approx GP(m(x), k(x,x')) \tag{4}$$

Assuming that the training dataset of the Gaussian model is represented by $D = \{(x_i, y_i) = |i =$

$1,2,\ldots\ldots N|\}\, x \in R^d$, $y \epsilon R$ where x is the input y is output, R^d, d-dimensional R is the one-dimensional vector space. The Gaussian distribution with D, a linear combination of Gaussian parameters, is given in Equation 5.

$$y \approx GP(m(x), k(x, x') + \sigma_n^2 I_n) \quad (5)$$

Where I denote the identity matrix. The covariance function is the central component in the GPR model, so function selection is critical. It is the most commonly used quadratic exponential in the literature from different functions such as linear, exponential, matern, rational quadratic, squared exponential. This can be explained by the fact that the function is infinitely differentiable and therefore, uniformly distributed. Because of these properties, the squared exponential kernel function is frequently used in studies on energy estimation (Heo & Zavala, 2012). In this study, the squared exponential function used as the covariance function is shown in Equation 6.

$$k(x\, i, x\, j \mid \theta) = \sigma_f^2 exp\left[-\frac{1}{2}\ \frac{(x\, i - x\, j)^t\ (x\, i - x\, j)}{\sigma_l^2}\right] \quad (6)$$

In the equation, σ_f refers to the signal standard deviation, and σ_l refers to the feature data's length.

3.2. Support Vector Regression

The support vector machine (SVM) is one of the most frequently used methods for predicting renewable energy generation (Zendehboudi et al., 2018). The method provides flexible control over the model complexity. Thus it gives successful results in solving nonlinear problems even in a small training set (Deng et al., 2018). The SVM aims at minimizing the structure risk in contrast to conventional experimental risk reduction. It tries to minimize the upper limit of education error and generalization error consisting of the sum of a confidence interval (Dong et al., 2005).

x_i represents the vector of input parameters, and y_i represents the output value (*i* represents data point *ith* in the data set). In this case, the sample set can be defined as $\{(x_i, y_i)\}_{i=1}^n$ where N represents the total number of samples.

$$y = f(x) = W.\emptyset(x) + b \quad (7)$$

In Equation 7, *W* denotes the feature vector, *b* is the intersection vector, and ∅ (x) is the high dimensional space. In Equation 8, the regulated risk function is given to estimate the *W* and *b* coefficients (Li et al., 2009).

$$Minimise = \frac{1}{2}|W|^2 + C\frac{1}{N}\sum_{i=1}^N L_\varepsilon(y_i, f(x_i)) \quad (8)$$

$$L_\varepsilon(Y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i) \le \varepsilon| \\ |y_i - f(x_i) \le \varepsilon| & others \end{cases} \quad (9)$$

$|W|^2$ is the regulated term, while C is the penalty factor determining the balance between pattern smoothness and training. The second term of Equation 8 is empirical error measured by the density loss function ε given in Equation 9. To estimate *W* and *b*, the equation given above is converted into the objective function in Equation 10.

$$Minimise\ \zeta_1\zeta_1^*\ W\ b: \frac{1}{2}|W|^2 + C\ \frac{1}{N}\sum_{i=1}^N(\zeta_1 + \zeta_1^*) \quad (10)$$

$$Subject\ to: \begin{cases} y_i - W.\emptyset x_i - b \le \varepsilon + \zeta_1 \\ W.\emptyset x_i + b \le \varepsilon + \zeta_1^*,\ i = 1,2,\ldots.N \\ \zeta_1 \ge 0\ \ \zeta_1^* \ge 0 \end{cases}$$

Here ζ_1^* refer to artificial variables. Equation 10 is written as when the kernel function is $K(x_i, x_j)$:

$$Minimise\ \{a_i\}\{a_i^*\} = \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N(a_i - a_i^*).(a_j - a_j^*).K(x_i, x_j) - \varepsilon\sum_{i=1}^N(a_i - a_i^*) + \sum_{i=1}^N y_i\,(a_i - a_i^*) \quad (11)$$

$$Subject\ to: \begin{cases} \sum_{i=1}^N (a_i - a_i^*) = 0 \\ (a_i - a_i^*) \in [0, C] \end{cases}$$

In Equation 11, a_i and a_i^* represent Lagrange factors, and *i* and *j* represent different examples. In this case, Equation 7 is expressed as follows (Li et al., 2009):

$$y = f(x) = \sum_{i=1}^N (a_i - a_i^*)\, K(x_i, x_j) + b \quad (12)$$

3.3. Decision Tree

Decision tree (DT) is frequently used in classification and regression problems because of their low computational costs and easy interpretation. Since short-term wind power prediction is a regression problem, the study focuses on the principle of regression trees. The DT is usually a binary tree in each node where a decision criterion is defined, taking into account a particular feature of the test model. A tag is assigned to each leaf node in the tree, so the practitioner can easily understand the tree's decisions (Hastie et al., 2009; Bishop, 2006).

Building regression trees is the process of creating an iterative tree by selecting the most appropriate features and split points based on the minimum square error (Breiman, 1984). Thanks to the tree's ability to be adjusted according to the properties of the data set, there is no need to pre-set the function structure, and it can be worked with both discrete and continuous variables. A simple decision tree is a binary tree consisting of two leaf nodes, a root node, and a branch. The decision tree

is denoted by $h(x; a_m)$, where a_m is the characteristic variable and split point in the m-th iteration. The examples given are $R = \{x_i , y_i\}_i^N$ and x_j are continuous variables. In the *R* set, x_j takes n different values. The values are written as $\{x_j^1, x_j^2, \ldots . x_j^n\}$ on an increasing level and the set *R* is divided into R^+ and R^- according to the division point s. If the value of x_{ij} is less than the s value, it is included in the R^- set, if it is large, it is included in the R^+ set (Equation 13).

$$R^+(j,s) = \{x_{ij} | x_{ij} \geq s\}$$
$$R^-(j,s) = \{x_{ij} | x_{ij} < s\} \quad (13)$$

The predicted value in each set must be equal to the output value or all samples' mean. The predicted value for the set R_m with the number of data N_m is calculated c_m with $Equation$ 14.

$$c_m = \frac{1}{N_m} \sum_{x_j \in R_m} y_i \quad (14)$$

The set consisting of all possible values of the division points s for each value of the continuous variable x_{ij} is shown by Equation 15.

$$S_{x_j} = \left\{ \frac{x_j^i + x_j^{i+1}}{2} \; | 1 \leq i \leq N-1 \right\} \quad (15)$$

To find the appropriate feature x_{ij} and the split point s, all the split points for all properties should be examined and selected as the final split point with minimum loss. Here the loss can be calculated by Equation 16.

$$\Psi(j+s) = \sum_{x_j \in R^+} (y_i - c_s^+)^2 + \sum_{x_j \in R^-} (y_i - c_s^-)^2 \quad (16)$$

Finally, the optimal feature variable and the split point are written as in Equation 17.

$$(j^*, s^*) = argmin_{j,s} \sum_{x_j \in R^+} (y_i - c_s^+)^2 + \sum_{x_j \in R^-} (y_i - c_s^-)^2 \quad (17)$$

3.4. Bayesian Optimization Algorithm

Hyperparameter selection plays an important role in the success of machine learning algorithms. Grid search, which is frequently used in hyperparameter optimization, becomes complicated when the parameter space size is high while performing a comprehensive search in simple models (Cornejo-Bueno et al., 2018; Alade et al.,2019). Since the random search algorithm works by randomly sampling the search field, it does not use the information obtained from previous experiments. This situation creates a problem, especially in hyperparameter selection problems where the function is unknown, and the cost of running an experiment is high. BOA comes into play at this point (Wang et al, 2012). BOA estimates the posterior distribution of the objective function using the Bayes theorem and determines the hyperparameter combination of the next example according to this distribution. Unlike the random search, it uses all the information obtained from the previous experiment and tries to find the parameter to bring the result to the global maximum. To avoid native optima, exploration and exploitation need to be changed. The gain function is defined to encrypt this exchange. The gain function returns the utility estimates of the candidate points for the next step of *f(x)* and selects $x_{(t+1)}$, which produces the maximum utility.

In this study, the expected improvement (EI) is used as the gain function. The value x looking for the global minimum of the given function f (x) is obtained as in Equation 18:

$$x = argmax\, E(max\{0, f_{t+1}(x) - f(x^+)\} | D_t \quad (18)$$

Where *D* represents the number of x components. When a Gaussian process is used, it is expressed for EI as:

$$EI(x) \begin{cases} (\mu(x) - f(x^+))\varphi(z) + \sigma(x)\,\varphi(z), if\ \sigma(x) > 0 \\ 0, \qquad if \sigma(x) = 0 \end{cases} \quad (19)$$

$$z = \frac{\mu(x) - f(x)^+}{\sigma(x)} \quad (20)$$

3.5. Bagging

In the bagging approach is aimed to create independent predictors by using the samples of the training set and the average or voting of the output of the prediction methods. When more than one similar dataset is created by re-sampling with preloading, the variance of the output error decreases and overfitting is solved (Breiman, 1996). Breiman (1996) showed also that bagging method gives better results than the single tree in both classification and regression problems.

In bagging, in the first step, N new training data of n dimensions are created, in which n of the n samples are selected homogeneously by changing from the original data set. Then, each tree in the community is individually trained with relevant new training data. For example, in this study, 50 trees were used in bagging tree models. In the last step, the average of all predictions is calculated to make a final forecast. In the bagging trees model, the prediction is defined as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \quad (21)$$

Here f_i shows the generating trees, *i* shows the boot data and each tree model created is trained on the boot data.

3.6. Stacking

The stacking method, known as a batch generalization, is a method of different levels that combines different prediction models in a single model. The approach aims to minimize the errors by reducing the bias of the generalizers by introducing an optimal learning system with the concept of meta-learning (Wolpert, 1992; Van der Laan et al., 2007).

The stacking approach consists of two levels, level-0 and level-1. In Level-0, different prediction models are trained, and the output variable is predicted. The predictions obtained at Level-0 are used as input for Level-1. The model in Level-1 is called a meta-model and learns with the previous level models that give the best prediction of each model of the previous level (Shamaei & Kaedi, 2016; Serbes et al., 2015; Petropoulos et al., 2017). The number of levels is not limited to 2 in the stacking method. According to the method's working principle, the n-level model uses the prediction of the n-1 level models. In the stacking method, the change in the prediction results is due to the diversity of models at different levels. This is because models with different generalization rules tend to produce different results (Mendes-Moreira et al., 2012).

3.7. Model performance evaluation

Different performance criteria such as Normalized Absolute Mean Error (NMAE), Normalized Root of Mean Squares Error (NRMSE) and determination coefficient (R^2) have been used to test the success of models used in short-term wind prediction. The performance criteria used are given in Equation 22-24.

$$NRMSE\ (\%) = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2}}{C}\ 100 \tag{22}$$

$$NMAE\ (\%) = \frac{\frac{1}{N}\sum_{i=1}^{N}|f_i - y_i|}{C}\quad 100 \tag{23}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(f_i - y_i)^2}{\sum_{i=1}^{N}(f_i - \bar{f}_i)^2} \tag{24}$$

Performance criteria are proportioned to the amount of installed power to measure the accuracy of wind energy prediction models in percentage terms and not be affected by the characteristics of the wind turbine (Chen et al., 2013). In the equations, *N* is the number of data, f_i is the actual value, y_i is the predicted value, and $\bar{f}_i$ is the average actual value. In Equation 22 and Equation 23, *C* represents the installed power. The turbine capacity used in the study is 3600 kW.

4. Proposed Approach for Short Term Wind Power Prediction

In the study, a hybrid approach consisting of BO, GPR, DT, SVR, bagging and stacking methods is proposed for short term wind prediction. The phases of the proposed approach are as follows:

Step 1. Data Collection and Data Preprocessing

- The data set consisting of SCADA data and meteorological parameters is rearranged so that the time step is 1 hour.
- The data for 15 days randomly selected from January, April, July and October are defined as test data and the remaining 6981 data as training data.
- Missing data is completed using the k-NN algorithm.
- The data is brought to the 0-1 range by applying max-min normalization to the data set. The formula used for normalization is given in Equation 25:

$$x_{nor,i} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{25}$$

Step 2. Bayesian Optimization of base learners

- BO is applied to find optimum hyperparameters of SVR, GPR and DT algorithms.
- Prediction results of GPR, DT and SVR algorithms are obtained with optimized hyperparameter values and denormalization is applied.
- The algorithm or algorithms with the least error values are selected as base learner for ensemble learning according to NRMSE, NMAE and R^2 performance criteria.

Step 3. Ensemble Models

- Ensemble learning models (bagging and stacking) are created with different combinations of algorithms selected from among the optimized algorithms. The estimation results of hourly test data determined by the created ensemble learning models are obtained.
- Denormalization is applied to the obtained prediction results.

Step 4. Comparison of developed models

- Prediction results of Bayesian optimized GPR, SVR, DT, bagging and stacking models created by working with SVR, GPR, DT algorithms trained with default parameters are obtained.
- The denormalized prediction results are compared seasonally and annually in terms of NRMSE, NMAE and R^2 criteria.

5. Application

5.1. Data Collection and Data Preprocessing

In the study, a wind turbine located in Yalova was chosen for the application. The turbine characteristics are as follows: capacity 3600 kW, rotor diameter 116.8 meters, cut-in wind speed 3 m / s, cut-out wind speed 25 m / s. The data are wind speed, wind direction, temperature, air pressure, relative humidity, and wind power for 2018. The parameters used in the studies in the literature were taken into consideration while creating the data set in this study (Chen et al., 2014; Banik et al., 2020, Ahmad et al., 2018). The wind turbine scada dataset on the Kaggle website was used in the article for short term wind power prediction (Erisen, 2019). Meteorological data were obtained from on the National Weather Service Center Environmental Forecast Climate (NOAA) website (NOAA, 2018). The data obtained have been rearranged so that the time interval is 1 hour and consists of 8399 samples. The 228 missing data points in the data set were completed with the k-NN algorithm (Gao et al., 2020).

The fact that the parameters used have different value ranges affects the prediction accuracy and causes uncertainty. In order to prevent this situation, the normalization process was applied to the data set by using Equation 25. While normalizing the wind direction data, it was first converted to radians, then sine and cosine values were calculated (Li et al., 2020).

In wind power prediction studies, it is seen that the data have a periodic pattern between seasons. For this reason, in many studies in the literature, predictions were made by creating training and test data containing data for each season, considering seasonal characteristics (Esfetang & Kazemzadeh, 2018; Zheng et al., 2017; Acikgoz et al., 2020). In this study, randomly selected 15-day data for January, April, July and October were used for the test data, and the remaining data were accepted as training data.

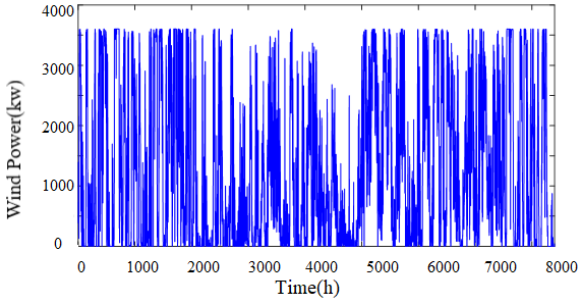

Figure 2. Wind power time series in 1-h time period in 2018

5.2. Bayesian Optimization of base learners

MATLAB 2020a program was used to create short-term wind power prediction models. GPR, DT and SVR, which are the selected prediction algorithms, were first applied with default parameters. The hyperparameter values that minimize the loss of 5- fold cross-validation with the BOA investigated. In the study, expected improvement is used as the acquisition function, which expresses how the parameter space should be investigated during BO.

In the first step of BO, initial hyperparameter values are assigned to the model. The model learns 4/5 of the data and tests it over 1/5. In each iteration, a new hyperparameter vector is created and the loss function value (log(1+loss)) is calculated. After the iterations are completed, the best hyperparameter vector that makes the loss function minimum is selected. With the selected hyperparameter vector, the final model is trained on the whole data set and the prediction results are obtained. Optimum values of sigma parameter in GPR, kernel function, box constraint, epsilon parameters in SVR and min leaf size and max num split parameters in DT were investigated.

Table 1 expresses the DT models in the BO iterations. Fig. 3 shows the minimum objective plot versus the number of function evaluations for the different decision tree models.

Table 1. The Bayesian optimization iterations for the DT model

Iter	Eval Result	Objective:log (1+loss)	Objective Runtime	BestSoFar (observed)	BestSoFar (Estim.)	MinLeafSize	MaxNumSplits
1	Best	0.011502	0.051844	0.011502	0.011502	467	15
2	Accept	0.023687	0.036668	0.011502	0.012333	1748	922
3	Best	0.0092397	0.063725	0.0092397	0.0092528	20	152
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
36	**Best**	**0.0088459**	**0.070322**	**0.0088459**	**0.0089**	**3**	**220**
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
48	Accept	0.0095316	0.056205	0.0088459	0.0088483	79	37
49	Accept	0.0093518	0.063937	0.0088459	0.0088493	28	504
50	Accept	0.010085	0.054857	0.0088459	0.0088499	2	17

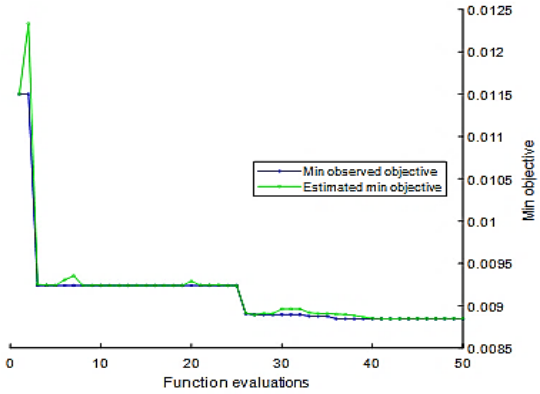

Figure 3. The minimum objective versus the number of function evaluations for the DT

According to the optimization results, the model reached the minimum objective function value by taking Min leaf size 3, and Max num split 220 in the 36-th iteration.

The Table 2, expresses the GPR models in the Bayesian optimization iterations. The Fig. 4 shows the plot of the minimum objective versus the number of function evaluations for the different GPR models.

Table 2. The Bayesian optimization iterations for the GPR Model

Iter	Eval Result	Objective: Log (1+loss)	Objective Runtime	BestSoFar (observed)	BestSoFar (Estim.)	Sigma
1	Best	0.0089561	60.936	0.0089561	0.0089561	0.030826
2	Accept	0.084427	73.087	0.0089561	0.013786	0.00016228
3	Accept	0.022513	66.347	0.0089561	0.015651	0.0017231
.	.	.	.	.	.	.
.	.	.	.	.	.	.
11	**Best**	**0.0068647**	**67.936**	**0.0068647**	**0.0069051**	**0.11505**
.	.	.	.	.	.	.
.	.	.	.	.	.	.
28	Accept	0.060073	65.317	0.0068647	0.0069193	0.011038
29	Accept	0.0069398	30.578	0.0068647	0.0069199	0.052232
30	Accept	0.022756	57.746	0.0068647	0.0069209	0.0029126

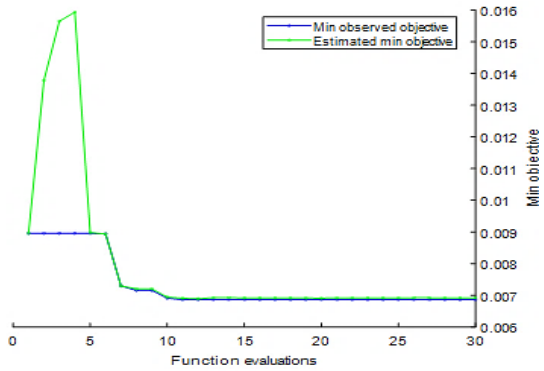

Figure 4. The minimum objective versus the number of function evaluations for the GPR

According to the results, the model reached the optimum result and minimum cross-validation loss in the 11-th iteration. Sigma took the value of 0.11505 in the optimum model.

The Table 3, represents the SVR models in the Bayesian optimization iterations. The Fig. 5 shows the plot of the minimum objective versus the number of function evaluations for the SVR model.

Table 3. The Bayesian optimization iterations for the SVR model

Iter	Eval result	Objective:log(1+loss)	Objective runtime	BestSoFar observed	BestSoFar estim.	BoxConstraint	Epsilon	KernelFunction
1	Best	0.12055	0.98747	0.12055	0.12055	0.082336	17.448	Linear
2	Best	0.018887	39.963	0.018887	0.02644	0.0039777	0.0013106	Polynomial
3	Accept	0.12055	0.24973	0.018887	0.024742	61.741	0.94565	Polynomial
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
25	**Best**	**0.0096151**	**1036.7**	**0.0096151**	**0.0096458**	**399.1**	**0.01481**	**gaussian**
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
28	Accept	0.0096409	826.89	0.0096151	0.009604	841.24	0.064184	gaussian
29	Accept	0.12055	0.67599	0.0096151	0.0096046	0.0010709	47.876	polynomial
30	Accept	0.12055	0.67059	0.0096151	0.0096049	0.0010016	45.804	gaussian

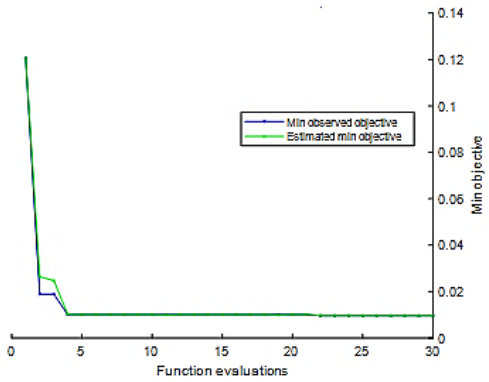

Figure 5. The minimum objective versus the number of function evaluations for the SVR

The SVR model has reached the optimum hyperparameter values in the 25-th iteration. The Kernel function of the optimum model is gaussian and Box constraint, Epsilon hyperparameters values are 399.1, 0.01481, respectively.

5.3.Selection and training of base learners

Table 4. Comparison of different regression algorithms used as base learner

Base Learner	Performance criteria		
	NRMSE (%)	NMAE (%)	R^2
BO-GPR	12.863	5.862	**0.873**
BO-SVR	13.821	5.700	0.845
BO-DT	**12.512**	**5.694**	0.866

BO has been used to improve base learner prediction performance. Table 4 shows a comparison of the BO-DT, BO-SVR and BO-GPR algorithms with BO as the base learner. According to NRMSE, NMAE, R^2 performance criteria, BO-GPR and BO-DT models have higher prediction performance than BO-SVR. Since no choice could be made between BO-GPR and BO-DT in terms of performance criteria, it was decided to use both as base learner in the study.

5.4. Ensemble Models

Table 5 lists the ensemble learning algorithms and parameters of the algorithms examined in this article. In the bagging algorithm, the BO- DT that generates fewer error predict compared to BO-GPR and BO-SVR was chosen as the basic learner. The bagging model prediction result was obtained by averaging 50 BO- DT. Bayesian optimized decision trees were trained with examples that are homogeneously selected from the original data set. In the stacking models, 3 different models were created by considering 2 levels. The first level BO-DT, BO-GPR and bagged tree model results are combined with the second level linear regression (LR) and BO-DT. It is aimed to create a model with better performance than both methods by combining BO-GRP and BO-DT models with meta learner in stacking models. For this purpose, the LR algorithm is also used as a meta learner, as it is a fast and less complex method, which increases the stacking performance, except for the BO-DT model.

Table 5. Parameters of all prediction models

Model	Parameters
BAG	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), Number of tree:50, Bagging size percent: 50%
STACK1	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR (Sigma: 0.11505) Meta learner: BO-DT
STACK2	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR (Sigma: 0.11505) Meta learner: LR (Lamda: 1.8584e-05, Learner: leastsquares)
STACK3	Base learner: BO-DT (Min Leaf Size: 3, MaxNumSplit: 220), BO-GPR, Bagged tree Meta learner: LR (Lambda: 1.8584e-05, Learner: leastsquares)

5.5. Experimantel Results

Table 6 contains seasonal performance criteria and the average values of all models used in the study.

When Table 6 is examined, it is seen that BO reduces the prediction error of the basic learners for each season. In addition, it has been demonstrated that the optimized GPR, SVR, DT of bagging and stacking algorithms improve the predictive performance of bagging and stacking models. The corresponding performance improvement was measured by the R^2, NRMSE and NMAE performance criteria. The ensemble learning algorithms created in the study were more successful in short-term wind power prediction than single algorithms for all four seasons. Considering the seasonal performance, it is seen that prediction errors are higher in winter and spring seasons. The reason for this is the fluctuations and sudden changes in the parameters during these seasons. Since this difference is less in summer and autumn seasons, the performance of the models is also better in this direction. When the annual and seasonal prediction results are examined, the average R^2, NRMSE, NMAE values of the bagging model with the best performance among the suggested methods were measured as 0.899, 11.045% and 4.880%, respectively.

Table 6. Seasonal and average results of models for 1-h prediction

		Methods									
Seasons	Performance criteria	GPR	DT	SVR	BO-GPR	BO-DT	BO-SVR	BAG	STACK1	STACK2	STACK3
Winter	R^2	0.868	0.707	0.502	0.869	0.809	0.723	**0.870**	0.801	0.782	0.823
	NRMSE(%)	13.563	21.902	27.518	13.460	15.940	21.459	**13.356**	17.666	17.039	15.875
	NMAE(%)	8.499	9.773	18.913	8.299	9.065	10.820	**8.259**	9.041	9.359	8.889
Spring	R^2	0.730	0.776	0.770	0.803	0.824	0.842	**0.888**	0.882	0.874	0.887
	NRMSE(%)	19.872	17.832	17.745	19.174	15.592	13.856	**13.830**	13.953	13.954	13.894
	NMAE(%)	8.709	7.916	10.354	8.512	6.320	5.407	**5.388**	5.636	5.408	5.980
Summer	R^2	0.881	0.919	0.801	0.915	0.925	0.910	**0.926**	0.910	0.922	0.918
	NRMSE(%)	10.203	8.124	14.931	9.694	8.051	10.578	**8.001**	9.067	8.089	8.231
	NMAE(%)	3.876	3.992	8.950	3.864	3.481	3.513	**3.320**	3.802	3.321	3.689
Fall	R^2	0.907	0.876	0.861	0.908	0.908	0.905	**0.912**	0.900	0.909	0.906
	NRMSE(%)	9.202	11.890	12.373	9.124	10.464	9.390	**8.992**	9.342	9.034	10.034
	NMAE(%)	2.790	4.963	6.861	2.771	3.912	3.059	**2.650**	3.564	4.139	3.366
Average	R^2	0.846	0.819	0.733	0.873	0.866	0.845	**0.899**	0.873	0.872	0.883
	NRMSE(%)	13.210	14.937	18.141	12.863	12.512	13.821	**11.045**	12.507	12.029	12.009
	NMAE(%)	5.968	6.661	11.270	5.862	5.694	5.700	**4.880**	5.511	5.531	5.482

6. Conclusions

In the study, prediction models based on BO, machine learning and ensemble learning algorithms were created for short term wind power prediction. In the models, analyzes were carried out by considering both statistical (SCADA records) and physical (meteorological parameters) data. The missing data in the data set created in the first step of the study were completed with the K Nearest Neighbor algorithm, and test and training data sets containing data for each season were created. For this purpose, the months of January, April, July and October were chosen to represent the winter, spring, summer and autumn seasons. In the second step, hyperparameters of the DT, SVR and GPR algorithms selected as base learner for wind power prediction were optimized with the Bayesian algorithm. The Bayesian optimized decision tree (BO-DT) that reaches the least prediction error was chosen in the bagging model. Stacking models include the Bayesian optimized GPR (BO-GPR) in addition to the BO-DT. When the results of the analysis are analyzed seasonally and annually, it is seen that the ensemble learning algorithms are more successful in wind power prediction than single methods. The annual average R^2, NMAE and NRMSE values are 0.899, 4.880%, 11.045%, respectively, and the average calculation time of the bagging model performs better than other wind energy prediction models, is less than 10 seconds.

The conclusions of the study can be summarized as follows:

- BOA was used for the first time in short-term wind energy prediction and has increased the prediction performance of regression algorithms.
- In the study, GPR and DT algorithms, which gave successful results in wind power prediction studies, were combined with the stacking algorithm for the first time and a model with higher performance than both methods was provided.
- Successful results were obtained by adding BO to the bagged tree model, which is also included in wind power prediction studies in the literature.
- It has been shown that the prediction errors of the models change depending on the seasons. All models created have fewer prediction errors in the summer and autumn seasons compared to the winter and spring seasons.

In future studies, the proposed model will be developed as follows:

- Optimum hyper parameter values of bagging and stacking algorithms, which are meta-learners, will also be investigated with BOA.
- In order to improve the data quality, which directly affects the prediction performance of the model, outliers in the data set will be detected and cleaned.

References

Acikgoz, H., Yildiz, C., Sekkeli, M., (2020). An extreme learning machine based very short-term wind power forecasting method for complex terrain. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 42(22), 2715-2730, DOI:10.1080/15567036.2020.1755390.

Ahmad, M.W., Mourshed, M., Rezgui, Y., (2018). Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy, 164, 465-474.

Alade, I.O., Abd Rahman, M.A., Saleh, T.A., (2019). Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. Solar Energy, 183, 74–82.

Bağcı, E., (2019). Türkiye'de Yenilenebilir Enerji Potansiyeli, Üretimi, Tüketimi ve Cari İşlemler Dengesi İlişkisi. R&S-Research Studies Anatolia Journal, 2 (4), 101-117. DOI: 10.33723/rs.501940.

Banik, R., Das, P., Ray, S., Biswas, A., (2020). Wind power generation probabilistic modeling using ensemble learning techniques. Materials Today: Proceedings, 26, 2157–2162.

Bishop, C., (2006). Pattern Recognition and Machine Learning. Information Science and Statistics. Springer.

Breiman, L., (1996). Bagging predictors. Machine learning, 24, 123–140.

Breiman. L., Friedman, J.H., Olshen, R.A., (1984). CART: Classification and Regression Trees. Biometrics, 40, 358–380.

Chen, N., Qian, Z., Nabney, I., Meng, X., (2014). Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Trans. Power Syst., 29(2), 656-665.

Chen, N., Qian, Z., Meng, X., (2013). Short-term wind power forecasting using Gaussian processes. Proceedings of 23rd international joint conference on artificial intelligence, 3-9 August 2013. pp. 2790-2796.

Cornejo-Bueno, L., Garrido-Merchán, E.C., Hernández-Lobato, D., Salcedo-Sanz, S., (2018). Bayesian optimization of a hybrid system for robust ocean wave features prediction. Neurocomputing, 275, 818–828.

Deng, H., Fannon, D., Eckelman, M.J., (2018). Predictive modeling for US commercial building energy use: a comparison of existing statistical and machine learning algorithms using CBECS microdata. Energy Build, 163, 34-43.

Don, B., Cao, C., Lee, S.E., (2005). Applying support vector machines to predictbuilding energy consumption in tropical region. Energy Build, 37 (5). 545-553.

Erisen, B., (2019). Wind turbine scada dataset, Version 3. Available Online: https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset (accessed on 10 August 2020).

Eroğlu, M., (2019). Enerji Hukuku. İstanbul: BAU.

Esfetang, N. N. & Kazemzadeh, R., (2018). A novel hybrid technique for prediction of electric power generation in wind farms based on WIPSO neural network and wavelet transform. Energy, 149, 662-674.

ETKB. (2019). Enerji ve Tabii Kaynaklar Bakanlığı, Available Online: https://enerji.gov.tr/bilgi-merkezi-enerji-elektrik (accessed on 4 January 2021).

Fu. C., Li. G.-Q., Lin. K.-P., Zhang. H.-J., (2019). Short-term wind power prediction based on improved chicken algorithm optimization support vector machine. Sustainability, 11, 512.

Gao, Z., Shi, J., Li, H., Chen, C., Tan, J., Liu, L., (2020). Substation Load Characteristics and Forecasting Model for Large-scale Distributed Generation Integration. IOP Conf. Series: Materials Science and Engineering, 782, 032044 doi:10.1088/1757-899X/782/3/032044

Hastie, T., Tibshirani, R., Friedman, J., (2009). The Elements of Statistical Learning: Data Mining. Inference. and Prediction. Second Edition. Springer Series in Statistics. Springer.

Heinermann, J. & Kramer, O., (2016). Machine learning ensembles for wind power prediction. Renewable Energy.89, 671-679.

Heo, Y. & Zavala, V.M., (2012). Gaussian process modeling for measurement and verification of building energy savings. Energy Build, 53, 7-18.

Karık, F., Sözen, A. & İzgeç, M.M., (2017). Rüzgâr gücü tahminlerinin önemi: Türkiye elektrik piyasasında bir uygulama. Politeknik Dergisi. 20(4), 851-861.

Kerem, A., (2018). Rüzgar Parametrelerinin Değişiminin İzlenmesi ve Yapay Zeka Algoritmaları Kullanarak Tahmini. Doktora Tezi. Gazi Üniversitesi

Lee, D. & Baldick, R., (2014). Short-term wind power ensemble prediction based on Gaussian processes and neural networks. IEEE Trans Smart Grid, 5(1), 501-510.

Li, C., Lin, S., Xu. F., Liu, D., Liu, J., (2018). Short-term wind power prediction based on data mining technology and improved support vector machine method: a case study in Northwest China. J. Cleaner Prod., 205(4), 909-922, 10.1016/j.jclepro.2018.09.143

Li, L.-L., Zhao, X., Tseng, M.-L., Tan, R.R., (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. J Clean Prod. 242, 118447. 10.1016/j.jclepro.2019.118447

Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A., (2009). Applying support vector machine to predict hourly cooling load in the building. Appl Energy, 86(10), 2249- 2256.

Ma, Y.-J. & Zhai, M.-Y., (2019). A dual-step integrated machine learning model for 24h-ahead wind energy generation prediction based on actual measurement data and environmental factors. Appl Sci., 9, 21-25. 10.3390/app9102125

Mendes-Moreira, J.A., Soares, C., Jorge, A.M., Sousa, J.F.D., (2012). Ensemble approaches for regression: A survey. ACM Comput. Surv., 45(1), 1-10. http://dx.doi.org/10.1145/2379776.2379786

NOAA, 2018. National Weather Service Center Environmental Forecast Climate. Avaliable Online: https://www.ncdc.noaa.gov/cdo-web/ (accessed on 15 October 2020).

Öz, S. & Alyürük, M., (2020). Energy Sector Overview and Future Prediction for Turkey. Journal of Industrial Policy and Technology Management. 3(1). 2020. 59-69

Petropoulos, A., Chatzis, S.P., Siakoulis, V., Vlachogiannakis, N., (2017). A stacked generalization system for automated forex portfolio trading. Expert Syst. Appl., 90, 290–302. http://dx.doi.org/10.1016/j.eswa.2017.08.011

Rasmussen. C. E. &Williams, C. K. I., (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts.

Serbes, G., Sakar, B.E, Gulcur, H.O., Aydin, N., (2015). An emboli detection system based on dual tree complex wavelet transform and ensemble learning. Appl. Soft Comput., 37 (2015) 87–94. http://dx.doi.org/10.1016/j.asoc.2015. 08.015

Shamaei, E. & Kaedi, M., (2016). Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. Appl. Soft Comput., 45, 187–196. http://dx.doi.org/10.1016/j.asoc.2016.03.009

Şenel, M.C. & Koç, E., (2015). Dünyada ve Türkiye'de Rüzgâr Enerjisi Durumu-Genel Değerlendirme. Mühendis ve Makina. 56(663), 46-56.

Tahir, M., El-Shatshat, R., Salama, M.M.A., (2018). Improved stacked ensemble based model for very short-term wind power forecasting. In: Proceedings – 53rd International universities power engineering conference (UPEC). Glasgow. 4–7 September.

TÜREB, (2021). Türkiye Rüzgar Enerjisi Birliği RES Veritabanı, Available Online: https://www.tureb.com.tr/ (accessed on 31 December 2020).

Van der Laan, M.J., Polley, E.C., Hubbard, A.E.,

(2007). Super learner. Statistical Applications in Genetics and Molecular Biology, 6(1) http://dx.doi.org/10.2202/1544-6115.1309

Wan, Z.Y. & Sapsis, T.P., (2017). Reduced-space Gaussian Process Regression for data-driven probabilistic forecast of chaotic dynamical systems. PhysicaD, 345, 40–55.

Wang, C., Li, S., Zhu, M., (2012). Bayesian network learning algorithm based on unconstrained optimization and ant colony optimization. Syst Eng Electron, 5, 784–90.

Wolpert, D.H., (1992). Stacked generalization. Neural Netw. 5 (2), 241–259. http://dx.doi.org/10.1016/S0893-6080(05)80023-1.

YEKDEM, (2020). Yenilenebilir Enerji Kaynakları Destekleme Mekanizması (YEKDEM), Available Online:https://www.enerjiportali.com/wpcontent/uploads/2020/09/KSD_YEKDEM_2.09.2020-1-1.pdf (accessed on 8 January 2020).

Zendehboudi, A., Baseer, M.A., Saidur, R., (2018). Application of support vector machine models for forecasting solar and wind energy resources: a review. J. Clean. Prod., 199, 272-285.

Zheng, D., Shi, M., Wang, Y., Eseye, A. T., Zhang, J. (2017). Day-Ahead Wind Power Forecasting Using a Two-Stage Hybrid Modeling Approach Based on SCADA and Meteorological Information, and Evaluating the Impact of Input-Data Dependency on Forecasting Accuracy. Energies, 10(12), [1988]. https://doi.org/10.3390/en10121988