Fırat Üniversitesi Müh. Bil. Dergisi
36(2), 859-867, 2024

Araştırma Makalesi
https://doi.org/10.35234/fumbd.1473145

Comparative Analysis of Machine and Deep Learning Methods in Estimating the Turkish Electricity Market Clearing Price

Ahmed İhsan ŞİMŞEK[1*]
[1] İşletme, İktisadi ve İdari Bilimler Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye
[*1] aisimsek@firat.edu.tr

(Geliş/Received: 26/04/2024; Kabul/Accepted: 13/07/2024)

Abstract: The estimation of the clearing price in the electricity market holds significant strategic importance within the energy sector. Energy firms can enhance their operational efficiency by providing clients with more dependable price alternatives through precise estimation of the market clearing price. The precise determination of the market clearing price holds significant significance in facilitating strategic decision-making for decision makers and investors operating within the energy sector. Accurate pricing projections are crucial for ensuring stability in the energy market and enhancing energy reliability for consumers. Hence, it is imperative to employ novel methodologies and enhance the precision of predictions within the energy sector in order to ascertain precise price estimates. This study utilized hourly power data derived from various sources such as natural gas, dam, lignite, imported coal, wind, solar, geothermal, and biomass. Additionally, hourly electricity demand data was employed as input variables to estimate the clearing price of the electricity market. The study encompasses a total of 8772 hours of data collected between April 17, 2023, to April 16, 2023. The study employed linear regression, XGBoost, Random Forest, LSTM, and SVR techniques for prediction. The models were evaluated by comparing their performances using statistical coefficients such as RMSE, MSE, MAE, and R2. Based on the acquired performance measures, it was noted that the XGBoost approach exhibited the highest level of prediction performance.

Key words: Electricity clearing price, machine learning, deep learning, decision support, energy.

Türkiye Elektrik Piyasası Takas Fiyatının Tahmininde Makine ve Derin Öğrenme Yöntemlerinin Karşılaştırmalı Analizi

Öz: Elektrik piyasa takas fiyatının tahmini enerji alanında stratejik öneme sahiptir. Doğru bir şekilde piyasa takas fiyatının tahmin edilmesi ile enerji şirketleri müşterilerine daha güvenilir fiyat alternatifleri sunarak operasyonel verimliliğini artırabilmektedir. Piyasa takas fiyatının doğru bir şekilde tahmini enerji sektöründeki karar vericilerin ve yatırımcıların stratejik seçimler yapmalarına yardımcı olması açısından büyük önem taşımaktadır. Enerji piyasasında istikrarın sağlanması ve tüketiciler açısından enerji güvenilirliğini artırmak için fiyat tahminlerinin doğru bir şekilde yapılması gerekmektedir. Bu nedenle enerji endüstrisinde doğru fiyat tahminlerinin tapılması için yeni yöntemlerin kullanılması ve daha doğru tahminlerin yapılması oldukça önemlidir. Bu çalışmada elektrik piyasa takas fiyatının tahmin edilmesi için Doğalgaz, baraj, linyit, ithal kömür, rüzgâr, güneş, jeotermal ve biokütleden üretilen saatlik elektrik verileri ile saatlik elektrik talep verileri girdi değişkeni olarak kullanılmıştır. Çalışma 17.04.2023-16.04.2024 arasındaki 8772 saatlik veriyi kapsamaktadır. Çalışmada XGBoost, Random Forest, LSTM ve SVR yöntemlerinin yanı sıra doğrusal regresyon ile de tahmin yapılmıştır. Modellerin performansları RMSE, MSE, MAE ve R2 istatistik katsayıları kullanılarak karşılaştırılmıştır. Elde edilen performans metriklerine göre en iyi tahmin performansının XGBoost yöntemi tarafından üretildiği gözlemlenmiştir.

Anahtar kelimeler: Piyasa takas fiyatı, makine öğrenmesi, derin öğrenme, karar destek, enerji.

1. Introduction

Turkey relies heavily on foreign countries to supply its energy requirements. Currently, energy emerges as a significant expenditure. Electrical energy is a fast-growing global requirement that requires high-quality, efficient, and rapid delivery. One of the primary challenges lies in the inability to store the energy derived from various sources such as oil, coal, hydropower, natural gas, nuclear, and renewable energy sources (e.g., wind, solar, biofuels). The electricity demand is steadily rising due to factors such as population growth, urbanization, industrialization, and the rapid advancement of technology, which has become an essential aspect of human existence [1].

Each country possesses energy requirements, and the current natural resources may be insufficient to satisfy these requirements. Consequently, individuals may depend on alternative sources to compensate for the insufficiency of energy. The significance of demand forecasting cannot be emphasized, given the existing constraints on the storage capacity of electrical energy. The demand for energy in Turkey is seeing growth due to

* Sorumlu yazar: aisimsek@firat.edu.tr. Yazarların ORCID Numarası: [1] 0000-0002-2900-3032

the country's burgeoning economy and expanding population. Ensuring the equilibrium between supply and demand and accurately forecasting energy consumption holds significant importance within the framework of Turkey's external reliance [2].

The anticipation of power costs holds significant strategic significance within the energy industry. The precise prediction of electricity costs enables energy businesses to provide consumers with more consistent pricing options, hence enhancing operational efficiency. Hence, the prediction of electricity prices plays a crucial role in facilitating the strategic decision-making process for decision makers and investors operating within the energy market. Forecasting electricity prices aid decision makers in the energy sector in strategic planning and enhance the operational efficiency of energy organizations. The provision of precise price projections plays a crucial role in maintaining stability within the energy market and enhancing the dependability of energy services for consumers. Hence, the significance of study and advancements in the domain of power price forecasting within the energy sector necessitates continuous progress. Various mathematical models and artificial intelligence techniques are commonly employed in the estimation of electricity prices, as evidenced by the studies of [3-13].

ARIMA, multiple regression, and artificial neural network models were employed by [13] in their research to forecast the electricity price in Turkey. The input variables employed in the study were Natural Gas Production Amount, Wind Production Amount, Hydroelectric Production Amount (River + Dam), Thermal Production Amount (Lignite + Hard Coal), and Demand Amount. The output variable assessed the Market Demand Price. The study conducted an analysis on a total of 2928 hours of data collected over the months of March, April, and May 2020. The study concluded that the artificial neural networks method yielded the most precise forecasts. Unlike the research conducted by [13], this study includes the quantities of energy generated from imported coal, solar, geothermal, and biomass sources. Additionally, the number of input variables is expanded to enhance the precision of projections. Furthermore, the study augmented the dataset by incorporating 8772 hours of data spanning from 17.04.2023 to 16.04.2024, as opposed to the original 3-month data. Furthermore, instead of employing the ARIMA time series analysis method, the linear regression method was utilized. Furthermore, machine and deep learning approaches such as XGBoost, LSTM, Random Forest, and Support Vector Regression have been utilized to predict market clearing prices. The objective of this approach is to enhance the precision of estimating the market clearing price.

2. Literature Review

A variety of methodologies have been employed in scholarly works to approximate power prices. Historically, there has been a notable prevalence of research employing time series analysis and diverse optimization techniques [14-18]. Recently, research utilizing deep learning and machine learning techniques has gained prominence. The research employed an LSTM-based deep learning algorithm and utilized hourly data spanning five years, from January 1, 2015 to December 31, 2019. The effectiveness of the proposed model was evaluated using statistical coefficients including MAE, RMSE, MAPE, and SMAPE. The findings derived from the research demonstrated that the predictions generated by the suggested model were accurate [19].

A study was undertaken to assess the electricity price in the European market, utilizing parameters related to energy generation, meteorological conditions, and production inertia. SVM, RFR, DNN, and CNN models were utilized to make predictions. The predictive accuracy of the models was evaluated by comparing the statistical coefficients MAE, MAPE, SMAPE, DAAE, and RMAE. The study concluded that RF and CNN models were unable to generate precise predictions for the given dataset, whereas DNN and SVR algorithms had superior predictive performance [20].

A novel LSTM-NN based model was proposed for short-term electricity load and price forecasting in the conducted investigation. The evaluation of model performance in the study undertaken for the PJM and Spain electrical market involved the utilization of MAPE, MAE, RMSE, and VAR coefficients. The investigation has demonstrated the efficacy of the proposed new model's performance [21].

A separate study was conducted to predict short-term electricity prices for the Australian, Spanish, and PJM electricity markets. This study proposes a hybrid model that combines variational mode decomposition (VMD), self-adaptive particle swarm optimization (SAPSO), SARIMA, and deep belief network. The performance of the model was evaluated by measuring the coefficients of MAE, MAPE, and RMSE. The predictive accuracy of the suggested model has been evaluated against LSSVM, WNN, ARIMA, SAA, IAA, and MAH models. It has been demonstrated that the proposed model outperforms other approaches in terms of prediction performance [22].

A novel hybrid model was developed for electricity price prediction, comprising four modules: feature preprocessing, deep learning-based point prediction, error compensation, and probabilistic prediction module. The pre-processing module employed the Isolation Forest (IF) and Lasso algorithms. The point prediction module utilized LSTM, RNN DBN, and CNN models. Lastly, the probabilistic prediction module employed Quantile Regression (QR). The study utilized PJM energy market data to measure performance indicators, specifically employing MAPE, MAE, and RMSE coefficients. Empirical evidence has demonstrated that the suggested hybrid model has superior predictive capabilities in comparison to the outcomes achieved using LGBM, BPNN, SVR, and KNN techniques [23].

A novel hybrid algorithm utilizing Wavelet transform was developed in a separate study undertaken for the power market in the United States. The suggested WR-SAE-LSTM model has demonstrated accurate predictions in electricity price prediction for residential, commercial, and industrial sectors, with MAPE error coefficients of 0.86%, 0.47%, and 0.49% respectively. The outcomes derived from the suggested model exhibited a high degree of concordance with the predictions made by the EAI [24].

In another study in which a new hybrid model based on VMD, CNN and GRU was proposed for short-term electricity price prediction, the prediction performance of the proposed model was measured with MAPE and RMSE methods. The results obtained from the proposed model were compared with LSTM, CNN, VMD-CNN, BP and VMD-ELMAN methods. As a result of the analysis, it was revealed that the proposed hybrid model made more successful predictions than other compared methods [25].

Upon examining the literature, one may encounter research that provides estimations of power prices for Türkiye. The study employed artificial neural networks to anticipate short-term electricity prices in Turkey. The effectiveness of the models was evaluated using the Mean Absolute Percentage Error (MAPE) statistical coefficient. The findings of the study conducted in Istanbul province indicate that the artificial neural network model described in this research achieved accurate predictions, as evidenced by an error coefficient of 7.52%. [26].

The study employed multiple regression and artificial neural networks, utilizing data from 2019 [27]. In this study, the input factors utilized for estimating the electricity price included natural gas, hydroelectricity, terminus, and wind power output. The study's findings indicate that artificial neural networks yielded more accurate predictions, while the multiple regression model had greater success in identifying the variables.

A comparative analysis was conducted on the predictive performance of Random Forest and Support Vector Regression models, utilizing a dataset including 10.440 observations over a duration of 15 months from 2019 to 2020. 84% of the data acquired from the EPA database is assigned for training purposes, while the remaining 16% is given for testing. The models' prediction performances were evaluated using statistical coefficients such as MAE, MAPE, and RMSE. The study revealed that the Random Forest model outperformed the SVR model in terms of prediction accuracy [28].

3. Data&Methodology

A total of 8772 hours of data were utilized in this study, spanning from 17.04.2023 to 16.04.2024. The outcome variable utilized in this study was the Market clearing price, which represents the point of intersection between unit supply and demand. The input variables consist of hourly data that indicate the amount of electricity produced (measured in MWh) from various energy sources, such as Natural Gas, Dam, Lignite, River Water, Imported Coal, Wind, Solar, Geothermal, and Biomass. These factors also influence the market clearing price. Additionally, the variables utilized are the total electricity consumption amount (measured in MWH). Each model in the study used a dataset that was partitioned into 80% training data and 20% testing data.

The study utilized input and output variables derived from the research conducted by [13]. This study employed multiple regression analysis to examine the elements influencing the trading price of the electricity market. The objective was to identify the components that impact the price. This research incorporates additional input factors, namely imported coal, geothermal, biomass, and solar energy, in addition to the variables utilized in the [13]. Furthermore, it is worth noting that in the present study, the dataset was augmented by incorporating 8772 hours of data, as opposed to the previous study conducted by [13], which utilized 2928 hours of 3-month data. The research included six distinct methodologies. The methods under consideration for prediction include LSTM, XGBoost, SVR, RF, and Linear Regression. Various statistical indicators were employed to evaluate the efficacy of the model, encompassing RMSE, MAE, MSE, and R^2. The statistical parameters are calculated with mathematical equations represented as Equations 1, 2, 3, and 4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{N}} \tag{1}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{N} \tag{2}$$

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{N} \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu)^2} \tag{4}$$

The input and output variables used in the study and the methods used for analysis are given in Figure 1.

Figure 1. Input and Output Variables Used in the Study.

In the study, XGBoost, LSTM, Linear Regression, Random Forest and Support Vector Regression models were used to estimate the electricity market clearing price.

3.1. XGBoost Model

The XGBoost model was first introduced by Chen & Guestrin in 2016 [29]. The goal function employs normalization to decrease model complexity, mitigate overfitting, and expedite the learning process. XGBoost is a notable ensemble model that incorporates an efficient implementation of decision trees, resulting in a composite model that exhibits superior prediction performance compared to individual techniques employed in isolation [30]. The execution of the XGBoost technique is represented by equations 5-12.

$$\hat{y}_i = \emptyset(x_i) = \sum_{k=1}^{K} f_k(y_i), f_k \in \mathcal{F} \tag{5}$$
$$min\, L^{(t)}(y_i, \hat{y}_i{}^{(t)}) = min\left(\sum_{i=1}^{n} \iota\left(y_i, \hat{y}_i{}^{(t)}\right) + \sum_{k=1}^{t} \Omega(f_k)\right) \tag{6}$$
$$\Omega(f) = \gamma T + \frac{1}{2}\lambda w^2 \tag{7}$$
$$min\, L^{(t)} = min\left(\sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)\right] + \Omega(f_t)\right) \tag{8}$$
$$g_i = \partial_{\hat{y}_i{}^{(t-1)}} l(y_i, y_i^{t-1}) \tag{9}$$
$$h_i = \partial^2_{y_i^{t-1}}\, l(y_i, y_i^{t-1}) \tag{10}$$
$$w_j^* = -\frac{\sum g_i}{\sum h_i + \lambda} \tag{11}$$
$$obj^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum g_i)^2}{\sum h_i + \lambda} + \gamma.T \tag{12}$$

A tree ensemble model is employed to forecast the result of a dataset comprising n samples and m features. This model utilizes K additive functions, denoted as $\boldsymbol{D} = \{(\boldsymbol{x_i}, \boldsymbol{y_i})\}(|\boldsymbol{D}| = \boldsymbol{n}, \boldsymbol{x_i} \in \boldsymbol{\mathcal{R}^m}, \boldsymbol{y_i} \in \boldsymbol{\mathcal{R}}$. Equation 5 represents the space of regression trees, with F indicating this space. The variable $\boldsymbol{f_k}$ represents the quantity of learners who are not performing well, while K represents the overall number of learners who are not performing well. The target function of the algorithm at time t, abbreviated as $\boldsymbol{L^{(t)}}$, is formally defined by Equation 6. The parameter denoted as $\boldsymbol{l}(\boldsymbol{y_i}, \boldsymbol{\hat{y}_i}^{(t)})$ comprises a range of loss functions that are employed to address specific problems. Equation 7 introduces a commonly utilized approach for quantifying the degree of deviation between the observed value $(\boldsymbol{y_i})$ and the expected value $(\boldsymbol{\hat{y}_i}^{(t)})$, together with the overall intricacy of the model, denoted as $\sum_{k=1}^{t} \boldsymbol{\Omega}(\boldsymbol{f_k})$. In the tth iteration, the objective function is evaluated by replacing the expected value $(\boldsymbol{\hat{y}_i}^{(t)})$ for the ith sample. Equation 8 presents the execution of the calculation using the second-order approximation of the Taylor expansion at the estimated value of y from the previous iteration, represented as $(\boldsymbol{\hat{y}_i}^{(t-1)})$. Equation 8 represents the first and second derivatives of the loss function $\boldsymbol{l}(\boldsymbol{y_i}, \boldsymbol{\hat{y}_i}^{(t)})$, denoted as $\boldsymbol{g_i}$ and $\boldsymbol{h_i}$, respectively. The derivative can be computed by substituting the formulas labeled as Equation 18, Equation 9, and Equation 10 into Equation 6, as previously described. Equations 11 and 12 can be used to derive solutions. The variable obj*, denoted by equations 11 and 12, represents the numerical value of the score of the loss function. A lower score indicates a tree structure that is more proximate to an ideal state. The sign $\boldsymbol{w_j^*}$ denotes the efficient solution for the weights in the scenario being examined in the XGBoost model, n_estimators is 787, the learning rate is 0.01, max_depth is 16, min_child_weight is 3, subsample is 0.53 and gamma is 0.002.

3.2. LSTM Model

The Long Short-Term Memory (LSTM) was initially introduced by Hochreiter and Schmidhuber [31]. This model was meticulously built and incorporated memory features derived from the Recurrent Neural Network (RNN) in order to address the issue of long-term reliance. The Long Short-Term Memory (LSTM) model effectively preserves the neural network's long-term memory, making it suitable for analyzing carbon prices. The vanishing gradient problem arises in recurrent neural networks (RNNs) due to the unrestricted updating of information at the network layer, resulting in chaotic and easily disappearing information. Nevertheless, the Long Short-Term Memory (LSTM) network incorporates a forgetting unit and a memory unit within the hidden layer. This enables the elimination of insignificant information upon the introduction of new data while preserving crucial information in long-term memory. The units in Long Short-Term Memory (LSTM) are referred to as gates, and the core components of gating mechanisms consist of a single cell and three gates, namely the input gate, output gate, and forget gate [32]. The procedure for implementing the Long Short-Term Memory (LSTM) technique is outlined in Equation 13-19 subsequently:

$$\mathrm{x_{scaled}} = \frac{\mathrm{x}-\mathrm{x_{min}}}{\mathrm{x_{max}}-\mathrm{x_{min}}} \tag{13}$$

$$f_t = \sigma_g\,(W_f x_t + U_f h_{t-1} + b_f) \tag{14}$$

$$i_t = \sigma_g\,(W_i x_t + U_i h_{t-1} + b_i) \tag{15}$$

$$C'_t = \sigma_c\,(W_c x_t + U_c h_{t-1} + b_c) \tag{16}$$

$$C_t = f_t\; x\; C_{t-1} + i_t\; x\; C'_t \tag{17}$$

$$o_t = \sigma_g\,(W_o x_t + U_o h_{t-1} + b_o) \tag{18}$$

$$h_t = o_t \,\mathrm{x}\, \tanh\,(C_t) \tag{19}$$

Normalization was performed on the data using Equation 13. The variables $\boldsymbol{x_t}$, $\boldsymbol{h_{t-1}}$, $\boldsymbol{f_t}$, and $\boldsymbol{\sigma_g}$ are included in Equation 14, representing the input of the time series, the previous hidden state, the output vector, and the activation function, respectively. The bias coefficient is commonly referred to as $\boldsymbol{b_f}$, whereas the forget gates are labeled as $\boldsymbol{W_f}$ and $\boldsymbol{U_f}$. The output vector is linked to the forget gate. Equation 14 represents this relationship. The relationship between the current point in the time series input, represented as $\boldsymbol{x_t}$, and the hidden state, represented as $\boldsymbol{h_{t-1}}$, from the previous time frame is described by Equations 15 and 16. The values of the coefficients $\boldsymbol{i_t}$ and $\boldsymbol{C'_t}$ within this gate are determined by these variables. These coefficients are calculated using the activation function. The variables $\boldsymbol{W_i}$, $\boldsymbol{U_i}$, $\boldsymbol{W_c}$, and $\boldsymbol{U_c}$ indicate the weight coefficients, while the symbols $\boldsymbol{\sigma_g}$ and $\boldsymbol{\sigma_c}$ represent the activation function. Equation 17 represents the update process of the cell state, denoted as $\boldsymbol{C_t}$. This process involves multiplying the output of the input gate, $\boldsymbol{i_t}$, with the cell candidate data, $\boldsymbol{C'_t}$, and the result of multiplying the prior cell state, $\boldsymbol{C_{t-1}}$, by the outcome of the forget gate, $\boldsymbol{f_t}$. The computation yields a description of the altered cellular condition, represented as $\boldsymbol{C_{t-1}}$. The equation represented as (18) illustrates the mechanism via which the output vector $\boldsymbol{\sigma_t}$ is produced by the transformation of the input vectors $\boldsymbol{h_{t-1}}$ and $\boldsymbol{x_t}$ using the activation function $\boldsymbol{\sigma_g}$. The input gate is connected to the bias coefficient, $\boldsymbol{b_o}$, as well as the weighted values of the cell state, $\boldsymbol{W_o}$ and $\boldsymbol{U_o}$. The value of the output gate, $\boldsymbol{o_t}$, is multiplied by the current sequential cell state, $\boldsymbol{C_t}$, once it is generated. The activation function tanh is applied to the output of the hidden layer, as shown in Equation (19). In the LSTM model, optimizer=adam, loss= MSE, epochs=100, batch_size=32, verbose=1.

3.3. Random Forest Model

In 2001, Breiman introduced the random forest model, which is an ensemble learning technique that integrates decision trees into random forests. This approach aims to enhance prediction accuracy and mitigate the issue of overfitting. The system functions by creating several decision trees during the training process and producing the mode of the categories for tasks such as classification or the mean forecast for regression analysis [33]. The random forest model is classified as an ensemble tree-based learning approach. The technique being offered is the generation of predictions by the calculation of the average of the outcomes generated by multiple individual trees. The construction of individual trees relies on the use of bootstrap samples rather than the original dataset. Bootstrap gathering, also referred to as bagging, is a technique employed to mitigate the issue of overfitting [34].

The structure of the RF method is depicted in Figure 2.

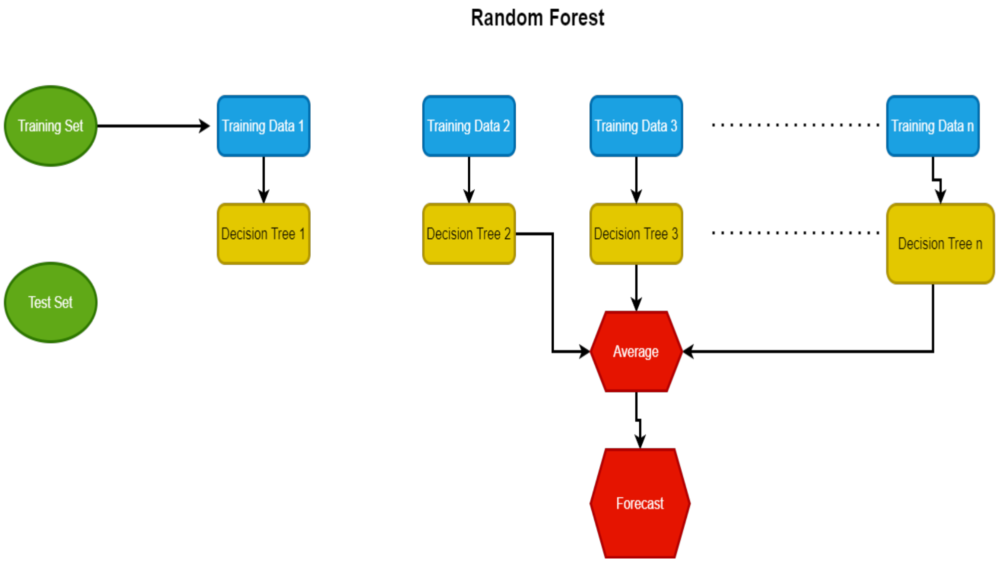

Figure 2. Random Forest Approach.

In the created Random Forest model, n_estimators=150, min_sample_splite=2, min_samples leaf =1, and random state=42 were taken as parameters. In the first stage of the Random Forest model, data is collected and divided into training and testing. After the necessary adjustments are made to the data, multiple decision trees are brought together. Afterwards, different data points are assigned to different trees by performing bootstrap sampling. A decision tree is created on each sampled dataset and decision and leaf nodes are created. Afterwards, the decision trees are combined. An assessment is conducted on the test set to evaluate the Random Forest model, and performance metrics are measured.

3.4. Support Vector Regression Model

The initial introduction of the support vector machine SVR technique in the 1990s is attributed to Cortes and Vapnik [35]. Subsequently, a regression methodology called support vector machine for regression was developed [36]. Support vector regression (SVR) is a highly efficient machine learning method that has been widely employed in several fields. The machine learning approach known as Support Vector Regression (SVR) employs Support Vector Machines (SVM) for the purpose of estimating functions [37].

The software application known as Support Vector Regression (SVR) is utilized to identify the optimal hyperplane that effectively segregates distinct variables. The hyperplane that is considered ideal is characterized by its possession of the maximum margin, hence ensuring a fair distance from all variables [38].

Equation 20-26 outlines the consecutive phases of the Support Vector Regression (SVR) methodology.

$$f\,(x) = \omega\,\Phi\,(x) + b \tag{20}$$

$$\mathrm{L}(f\,(\mathrm{x}), y, \varepsilon) = f(x) = \begin{cases} 0 & |\mathrm{y} - f\,(\mathrm{x})| \le \varepsilon \\ |\,y - f\,(\mathrm{x})| - \varepsilon & |\mathrm{y} - f\,(\mathrm{x})| > \varepsilon \end{cases} \tag{21}$$

$$\begin{cases} Min.\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \\ sub.t.\begin{cases} y_i - \omega\Phi(x_i) - b \le \varepsilon + \xi_i \\ -y_i + \omega\Phi(x_i) + b \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \end{cases} \tag{22}$$

$$\omega^* = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\,\Phi(x_i) \tag{23}$$

$$b^* = \frac{1}{N_{nsv}}\left\{\sum_{0<\alpha_i<C}\left[y_i - \sum_{x_i \in SV}(\alpha_i - \alpha_i^*)K(x_i, x_j) - \varepsilon\right] + \sum_{0<\alpha_i<C}\left[y_i - \sum_{x_j \in SV}(\alpha_j - \alpha_j^*)K(x_i, x_j) + \varepsilon\right]\right\} \tag{24}$$

$$K(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{25}$$
$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b^* \tag{26}$$

Support Vector Regression (SVR) aims to identify a linear regression function, represented as f(x), in a high-dimensional space. Let x denote an element that is a member of the set of real numbers, and let it function as the sample vector. The mapping of the function exhibits non-linear properties. Including a linear insensitivity loss function, denoted as L($\boldsymbol{f}$ (x), y, ε), improves the resilience of the optimization problem. Equation 22 provides a numerical representation of the loss function. Equation 22 involves the representation of the input vector and output value by the variables $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$, respectively. The variables in question are linked to a particular serial number, represented by the symbol i. Both variables $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ are members of the set of real numbers, which is represented as R. The input vector has a dimension of d. The variable d represents the cardinality of the elements contained in an input vector in this scenario. n represents the quantity of training samples. The symbol ε represents the measure of precision in regression analysis. The variable C denotes a punishment factor that measures the magnitude of the penalty imposed on a data sample when its error exceeds the threshold value ε. The slack variables $\boldsymbol{\xi_i}$ and $\boldsymbol{\xi_i^*}$ are employed to impose penalties on the complexity of the fitting parameters. To determine the estimation of and b, it is crucial to tackle the optimization problem as outlined in Equations 23 and 24. The variable $\boldsymbol{N_{nsv}}$ represents the number of support vectors that have been specifically recognized. The Lagrange multipliers, denoted as $\boldsymbol{\alpha_i}$ and $\boldsymbol{\alpha_i^*}$, are subject to the constraint of being non-negative. In this specific scenario, Equation 25 utilizes the kernel function, denoted as $\boldsymbol{K(x_i, x_j)}$. The Gaussian kernel function, renowned for its remarkable ability to generalize, is chosen. Equation 26 represents the final regression function. The SVR model utilized a kernel function. The following values were used: C=1, epsilon=0.1, and gamma: scale.

4. Results&Discussion

Table 1 and Table 2 present the training and test data acquired from LSTM, XGBoost, SVR, LR, and RF models, respectively. The model performances were evaluated using the MSE, RMSE, MAE, and R2 statistical factors. A decrease in the values of MSE, RMSE, and MAE suggests that the model exhibits reduced inaccuracy in its predictions. R2 is a measure of the extent to which the independent variables in a model account for the variation in the dependent variable. A high R2 value suggests that the model yields superior outcomes.

Table 1. Statistical coefficients of training data.

Training	XGBoost	RF	LSTM	SVR	LR
MSE:	0.0005	0.0011	0.0135	0.0092	0.0180
RMSE:	0.0216	0.0334	0.1161	0.0958	0.1342
MAE:	0.0165	0.0247	0.0899	0.0761	0.1060
R2:	0.9899	0.9759	0.7079	0.8014	0.6103

Table 2. Statistical coefficients of testing data.

Training	XGBoost	RF	LSTM	SVR	LR
MSE:	0.0005	0.0011	0.0135	0.0092	0.0180
RMSE:	0.0216	0.0334	0.1161	0.0958	0.1342
MAE:	0.0165	0.0247	0.0899	0.0761	0.1060
R^2:	0.9899	0.9759	0.7079	0.8014	0.6103

The power market clearing price was estimated using five distinct models based on the acquired data set. Upon examination of the data, it becomes evident that the Linear regression model yields the most unfavorable forecast outcomes. The LR model yielded a mean squared error (MSE) of 0.018, root mean squared error (RMSE) of 0.134, and mean absolute error (MAE) of 0.106. The R^2 value yielded the most unfavorable outcome in comparison to the other models. Following the LR model, it is observed that the LSTM model yields the most unfavorable outcome. The study conducted by Arifoglu & Kandemir (2022) found that LSTM achieved the highest level of prediction accuracy [39]. In contrast, the findings of this study indicate that the LSTM model exhibits inferior prediction performance compared to the LR model. The support vector regression (SVR) model is ranked third in terms of the predictive outcomes it generates. The error coefficients for the SVR model were measured as 0.009 for MSE, 0.095 for RMSE, and 0.0761 for MAE. While the SVR model yielded higher statistical coefficients than the LR and LSTM models, it exhibited inferior performance in comparison to the other two models employed in the study. The investigation yielded the highest prediction performances from the RF and XGBoost models. The RF model yielded mean squared error (MSE), root means squared error (RMSE), mean absolute error (MAE), and R^2values of 0.0011, 0.0334, 0.0247, and 0.9756, respectively. The data indicates that the RF model exhibits much superior predictive ability compared to the LR, SVR, and LSTM models. The analysis reveals that the XGBoost model exhibits the highest level of prediction performance. The XGBoost model and RF models have superior prediction performances compared to the other three approaches. When conducting a comparison between XGBoost and RF techniques, it becomes evident that the XGBoost model exhibits superior prediction performance across all assessed performance parameters. The Max Mean Squared Error (MSE) coefficient for the XGBoost model was calculated to be 0.0005. This demonstrates that it yields highly successful outcomes. Furthermore, the Mean Absolute Error (MAE) value was calculated to be 0.0165. This value indicates that it generates forecasts with a lower margin of error compared to alternative methods.

There is a prevailing observation that the predictive capabilities of Linear Regression, LSTM, and SVR models are inadequate for the given dataset. This is confirmed by the high values of MSE, MAE, and RMSE, as well as the lower R^2 values in comparison to other approaches. The Random Forest and XGBoost models show strong predictive capabilities in forecasting the clearing price of the Turkish Electricity market, considering the specific dataset employed. The models have good prediction performance, as indicated by their low RMSE, MSE, and MAE values, as well as high R^2 value.

The study's error coefficients exhibited a comparatively smaller magnitude when compared to the findings provided by Arslan & Ertugrul (2022), particularly in the predictions generated by ensemble methodologies like Random Forest and XGBoost. The Random Forest and XGBoost strategies, which surpass the ARIMA, Multiple Regression, and Artificial Neural Networks methods employed in the study, are anticipated to yield more precise outcomes in predicting energy market clearing prices, even when dealing with a bigger dataset. Further study initiatives may involve the utilization of diverse machine learning and deep learning methodologies to determine the trading price of the electrical market. In addition, the study can contain hybrid models, which combine many models, and then compare them with typical hybrid models. Moreover, the dataset has the capacity to be expanded and employed in other domains.

References

[1] Haliloğlu EY, Tutu BE. Türkiye için kısa vadeli elektrik enerjisi talep tahmini. Yasar University EJ; 2018; 13(51): 243-255.

[2] Nebati EE, TAŞ M, Ertaş G. Türkiye'de elektrik tüketiminde talep tahmini: zaman serisi ve regresyon analizi ile karşılaştırma. Eur J Sci Technol; 2021; (31): 348-357.

[3] Contreras J, Espínola R, Nogales F, Conejo A. Arima models to predict next-day electricity prices. IEEE Trans Power Syst 2003; 18(3): 1014-1020.

[4] Amjady N, Daraeepour A, Keynia F. Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network. IET Gener Transm Distrib; 2010; 4(3): 432-444.

[5] Carpio KJE, Go AML, Roncal CKM. Forecasting day-ahead electricity prices of Singapore through ARIMA and wavelet-ARIMA. DLSU Bus Econ Rev; 2012; 22(1): 97-118.

[6] Voronin S, Partanen J, Kauranne T. A hybrid electricity price forecasting model for the Nordic electricity spot market. Int Trans Electr Energy Syst; 2013; 24(5): 736-760.

[7] Wang Z, Liu F, Wu J, Wang J. A hybrid forecasting model based on bivariate division and a backpropagation artificial neural network optimized by chaos particle swarm optimization for day-ahead electricity price. Abstr Appl Anal 2014; 2014: 1-31.

[8] Jiang P, Liu F, Song Y. A hybrid multi-step model for forecasting day-ahead electricity price based on optimization, fuzzy logic and model selection. Energies 2016; 9(8): 618.

[9] Gao G, Lo K, Lu J. Risk assessment due to electricity price forecast uncertainty in UK electricity market. In: 52nd International Universities Power Engineering Conference (UPEC); 2017; New York, NY, USA: IEEE. pp. 1-6.

[10] Pourdaryaei A, Mokhlis H, Illias H, Kaboli S, Ahmad S. Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach. IEEE Access 2019; 7: 77674-77691.

[11] Huang C, Shen Y, Chen Y, Chen H. A novel hybrid deep neural network model for short-term electricity price forecasting. Int J Energy Res; 2020; 45(2): 2511-2532.

[12] Karatekin C, Başaran T. Forecasting the day ahead electricity energy price by using data analysis methods. Iğdır Univ J Inst Sci Technol; 2022; 12(4): 2075-2084.
[13] Arslan B, Ertuğrul İ. Çoklu regresyon, ARIMA ve yapay sinir ağı yöntemleri ile Türkiye elektrik piyasasında fiyat tahmin ve analizi. J Manag Econ Res; 2022; 20(1): 331-353.
[14] Misiorek A, Trueck S, Weron R. Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. Stud Nonlinear Dyn Econom; 2006; 10(3).
[15] Wang R, Fu-xiong W, Ji W. Particle swarm optimization based GM(1,2) method on day-ahead electricity price forecasting with predicted error improvement. In: 2nd International Workshop on Database Technology and Applications; 2010; Wuhan, China. pp. 1-6.
[16] Martínez-Álvarez F, Troncoso A, Riquelme J, Aguilar-Ruiz J. Energy time series forecasting based on pattern sequence similarity. IEEE Trans Knowl Data Eng 2011; 23(8): 1230-1243.
[17] Ozozen A, Kayakutlu G, Ketterer M, Kayalica O. A combined seasonal ARIMA and ANN model for improved results in electricity spot price forecasting: case study in Turkey. In: Proceedings of Portland International Conference on Management of Engineering and Technology (PICMET); 2016; Portland, OR, USA. pp. 2681-2690.
[18] Costa e Silva E., Borges A., Teodoro MF., Andrade MA., Covas R. Time series data mining for energy prices forecasting: an application to real data. In: 16th International Conference on Intelligent Systems Design and Applications (ISDA 2016); 2016; Porto, Portugal: Springer International Publishing. pp. 649-658.
[19] Li W, Becker DM. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. Energy 2021; 237: 121543.
[20] Tschora L, Pierre E, Plantevit M, Robardet C. Electricity price forecasting on the day-ahead market using machine learning. Appl Energy 2022; 313: 118752.
[21] Memarzadeh G, Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electr Power Syst Res; 2021; 192: 106995.
[22] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for short term electricity price forecasting. Appl Energy 2020; 258: 114087.
[23] Zhang R, Li G, Ma Z. A deep learning based hybrid framework for day-ahead electricity price forecasting. IEEE Access 2020; 8; 143423-143436.
[24] Qiao W, Yang Z. Forecast the electricity price of US using a wavelet transform-based hybrid model. Energy; 2020; 193: 116704.
[25] Huang CJ, Shen Y, Chen YH, Chen HC. A novel hybrid deep neural network model for short-term electricity price forecasting. Int J Energy Res; 2021; 45(2): 2511-2532.
[26] Var H, Türkay BE. Yapay sinir ağları kullanılarak kısa dönem elektrik yükü tahmini short term electric load forecasting using artificial neural networks. In: Elektrik–Elektronik–Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu; 2014; Bursa, Türkiye. pp. 34-37.
[27] Kalfa VR, Arslan B, Ertuğrul İ. Determining the factors affecting the market clearing price by using multiple linear regression method. Alphanumeric J; 2021; 9(1): 35-48.
[28] Demirezen S, Çetin M. Rassal Orman Regresyonu ve Destek Vektör Regresyonu ile Piyasa Takas Fiyatının Tahmini. J. Quant Sci; 2021; 3(1): 1-15.
[29] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; New York, NY, USA. pp. 785-794.
[30] Jabeur SB, Mefteh-Wali S, Viviani JL. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann Oper Res; 2024; 334(1): 679-699.
[31] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997; 9(8): 1735-1780.
[32] Zhou F, Huang Z, Zhang C. Carbon price forecasting based on CEEMDAN and LSTM. Appl Energy 2022; 311: 118601.
[33] Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Wang X, Bian H, Zuang S, Pradhan BB, Ahmad BB. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Sci Total Environ; 2020; 701: 134979.
[34] Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J; 2020; 20(1): 3-29.
[35] Cortes C, Vapnik V. Support vector networks. Mach Learn; 1995; 20: 273-297.
[36] Vapnik VN. An overview of statistical learning theory. IEEE Trans. Neural Netw.; 1999; 10(5): 988-999.
[37] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput; 2004; 14: 199-222.
[38] Zouzou Y, Citakoglu H. Reference evapotranspiration prediction from limited climatic variables using support vector machines and Gaussian processes. Eur J Sci Technol; 2021; 28: 346-351.
[39] Arifoğlu A, Kandemir T. Electricity price forecasting in Turkish day-ahead market via deep learning techniques. Mehmet Akif Ersoy Univ J Econ Admin Sci; 2022; 9(2): 1433-1458.