J Inno Sci Eng, 2025, 10(1):189-200

DOI:https://doi.org/10.38088/jise.1738364

Research Article

Modeling Turkey's Hourly Electricity Market Clearing Prices Using Exponential Gaussian Process Regression

Visal Özdemir [1], Mehmet Yılmaz [1*]

[1] Department of Electrical-Electronic Engineering, Ataturk University, Erzurum-25240, Turkey

ARTICLE INFO

Received Date: 10/07/2025
Accepted Date: 18/11/2025

Cite this paper as:
Özdemir, V., & Yılmaz, M. (2026) Modelling Turkey's Hourly Electricity Market Clearing Prices Using Exponential Gaussian Process Regression. *Journal of Innovative Science and Engineering.* 10(1), 189-200.

*Corresponding author: Mehmet Yılmaz
E-mail:mehmet.yilmaz@atauni.edu.tr

Keywords:
Support Vector Machines
Gaussian Process Regression
Day-Ahead Market
Market Clearing Price

ABSTRACT

In the electricity market, the Market Clearing Price (MCP) holds strategic importance for both producers and consumers, as it represents the price formed at the equilibrium point of supply and demand. Accurate prediction of the MCP is critical for market participants in terms of production planning, cost estimation, and risk management processes. In this study, modeling was performed using four different machine learning (ML) methods, namely Exponential Gaussian Process Regression (EGPR), Rational Quadratic Gaussian Process Regression (RGPR), Quadratic Support Vector Machine (QSVM) and Medium Gaussian Support Vector Machine (MGSVM), in order to estimate the MCP in the Turkish Day Ahead Market (DAM). A total of 4001 data points were used for the training phase. Two methods were used to prove the robustness and generalizability of the models. These are: 10-fold cross-validation method, and testing with a data set (25 points) completely independent of the training data. The data used in the modeling process include real-time electricity consumption (MWh), hourly electricity generation by source (natural gas, hydroelectric power plants (DAM), lignite, run-of-river, solar energy), Purchase/Sale Bid Volume, Matched Purchase/Sale Quantity and U.S. Dollar Exchange Rate. The performances of machine learning methods were compared according to performance criteria such as Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2). It was observed that the model created with EGPR had the highest R^2 value of "0.908" and "0.913" for 10-fold cross-validation and an independent test dataset, respectively. In addition, it was obtained that the MSE, RMSE and MAE error metrics for the 10-fold cross-validation of the model created with EGPR were the lowest with 26026.20, 161.33 and 120.33, respectively. For the independent test dataset, the MSE, RMSE, and MAE error metrics were 7734, 87.94, and 61.56, respectively, resulting in lower prediction errors than other ML methods. It was observed from both the 10-fold cross-validation results and the independent test dataset results that the model created with EGPR made more successful predictions than the models created with RGPR, QSVM and MGSVM.

1. Introduction

Technological advancements and a growing human population are increasing the demand for electrical energy. In energy markets, the Market Clearing Price (MCP) is defined as the intersection of supply and demand. Accurate price forecasting is essential for effective planning and management of energy systems. In this context, there is a Day Ahead Market (DAM) in Türkiye, where energy supply and demand are balanced, electrical energy purchase and sale offers are collected and the market clearing price is determined. Institutions such as the Energy Markets Operations Corporation (EPİAŞ) and the Turkish Electricity Transmission Corporation (TEİAŞ) play an important role in the supervision and operation of this market. MCP determined in this market reflects the price at which electricity supply and demand come into balance between producers and consumers. Accurate forecasting of prices in the DAM plays a critical role for market actors to make strategic planning and manage risks (Ertaylan et al., 2021).

Electricity is now a commercial good that is subject to competition as a result of the Turkish electricity market's deregulation and reorganization. The transmission system operator's investment choices and private generating businesses' planning and investment preferences are now heavily influenced by market dynamics and economic indicators in this new market structure (Awad et al., 2010). Economic balance models are useful tools for assessing new investment trends, price signals that appear in a competitive market, how producers or investors react to these signals, and supply-demand balance uncertainties like load factor and capacity. These models can help market participants with their strategic planning and forecasting. Despite the fact that market pricing simulation models have been created for numerous nations and areas in the literature, no suitable model has been created that is tailored to the Turkish electricity market. These models are required to assess market circumstances, track market power, and examine the actions of private generation businesses. As a result, regulatory bodies including the Competition Authority, EPİAŞ, and Energy Market Regulatory Authority (EPDK) can use these tools to monitor the market and create intervention plans. Additionally, market participants will use these models to help them make short-, medium-, and long-term decisions. In contrast to the existing single price structure, market clearing models tailored to Turkey and created based on multiple capacity and load variables can simulate regional (node-based) pricing and production levels and enable a variety of scenario analyses. These models can be used to analyze how different pricing strategies (such as regional and single pricing) affect price-cost margins and social welfare.

Electricity price forecasting is a highly challenging task. Machine learning (ML) methods are frequently used to estimate MCP. Some of these; Foruzan et al. (2015) developed two methods for estimating the electricity market clearing price. The first is Artificial Neural Network (ANN), and the second is Support Vector Machine (SVM). The authors used the ant colony optimization (ACO) algorithm to obtain the best feature set for ANN. Their results showed that ACO-ANN shortens training time compared to ANN, and that ACO-ANN and SVM produce more accurate predictions. Díaz et al. (2019) used the Gradient Boosted Regression Trees method to predict the day-ahead market clearing price in Spain. The authors' primary reasons for choosing this method were its increased forecast accuracy and its simpler calibration structure compared to other ML methods. From the results obtained, they observed that the proposed model produced less prediction error than other models. Yanar (2020) used ML and Deep Learning methods to predict the electricity market spot price. In the study, a market exchange price data point for the period from 01.01.2016 to 01.01.2018 was used, and the performances of different methods were evaluated by calculating the Mean Absolute Percentage Error (MAPE) value. The results generally show that Convolutional Neural Network (CNN) and LSTM-based models perform better than other methods. Estimating the hourly MCP found in the GEP run by EPİAŞ was the goal of Demirezen and Çetin (2021). Machine learning techniques called Random Forest Regression (RF) and SVM were applied. In the study, 10,440 pieces of data were used between January 1, 2019 and March 10, 2020. 84% of this data was divided into two groups: the train set and 16% of it as the test set.The analysis revealed that the RF model, which included the variable group including the transaction volume, provided the best prediction performance criteria. Transaction volume is a crucial determinant for PTF estimate, as this study has demonstrated (Demirezen&Çetin, 2021). Spiliotis et al. (2021) compared two different machine learning with traditional statistical methods. They used a neural network and a random forest algorithm as machine learning algorithms. They evaluated the performance of these two methods using Belgian electricity market data. The results showed that ML methods performed better in forecasting than other traditional statistical methods. The results showed that ML methods performed better in forecasting than other traditional statistical methods. Das et al. (2025) proposed a hybrid method combining GPR and SVM methods for hourly electricity price prediction. Hourly electricity price for the years 2021-2023 was used for model training. The results showed that the hybrid model, with its low Root Mean Square Error (RMSE) value, provides more accurate forecasts in energy markets experiencing sudden fluctuations. Das & Schlüter (2025) conducted an hourly MCP forecast for the German market. They used a hybrid method as their estimation model. They used Value at Risk and Conditional Value at Risk metrics in conjunction with GPR in developing the hybrid method. They observed that the use of hybrid metrics reduced the risk associated with sudden price increases. Their GPR-based hybrid model demonstrated that this model is effective in risk management and enables better decision making. In Şimşek's (2024) study, a comparative analysis of machine and deep learning methods was conducted for the purpose of estimating MCP. In the study, data collected over 8772 hours between 04/17/2023 and 04/16/2024 were used; production data from natural gas, dams, lignite, imported coal, wind, solar, geothermal, biomass sources and electricity demand data were taken as inputs for the model. As prediction models, Linear Regression (LR), Extreme Gradient Boosting XGBoost, RF, and LSTM were used; performance comparisons were made using RMSE, Mean Square Error (MSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2). statistics. According to the findings, the highest prediction accuracy was achieved with the XGBoost method, followed by RF. On the other hand, the LSTM and LR models have shown lower performance. The study emphasizes the importance of accurate price forecasts in the strategic decision-making processes in the energy market. (Şimşek, 2024).

In this study, modeling was performed using four different ML methods, namely Exponential Gaussian Process Regression (EGPR), Rational Quadratic Gaussian Process Regression (RGPR), Quadratic Support Vector Machine (QSVM) and Medium Gaussian Support Vector Machine (MGSVM), in order to estimate the MCP in the Turkish Day Ahead Market (DAM). A total of 4001 data points consisting of 24-hour data obtained between 07:00 on 15.04.2023 and 05:00 on 29.09.2023 were used as training data, incorporating variables such as real-time electricity consumption (MWh), hourly electricity generation by source (natural gas, hydroelectric power plants (DAM), lignite, run-of-river, solar energy), Purchase/Sale Bid Volume, Matched Purchase/Sale Quantity and U.S. Dollar Exchange Rate. Although the data set in our study has a similar initial period to the data set used by Şimşek (2024), the data set we used provides a more up-to-date and complete representation of the factors affecting the MCP.

In order to prove the robustness and generalizability of the models, a 10-fold cross-validation method was used and testing was performed with a data set (25 pieces) that was completely independent of the training data. While the internal consistency and statistical robustness of the model were ensured with the 10-fold cross-validation method, an attempt was made to demonstrate the generalization ability of the model in data that it had not encountered before with independent test data.

The rest of the paper is organized as follows:

In Section 2, the data acquisition process, detailed examination of EGPR, RGPR, QSVM and MGSVM methods and the performance criteria used in the study are explained. Section 3 provides separate 10-fold cross-validation results and independent dataset results. Section 4 provides a comparative analysis of the 10-fold cross-validation results and independent dataset results of the models generated using the machine learning methods used in the study.

The comparison of both models and which model is more efficient is emphasized. The original contributions of this study are given below;

- Four different machine learning methods were compared: Exponential Gaussian Process Regression (EGPR), Rational Quadratic Gaussian Process Regression (RGPR), Quadratic Support Vector Machine (QSVM), and Medium Gaussian Support Vector Machine (MGSVM).

- Eleven different variables were determined for the estimation of the market clearing price. These are: Real-time electricity consumption (MWh), hourly electricity generation by source (natural gas, hydroelectric power plants (DAM), lignite, run-of-river, solar energy), Purchase/Sale Bid Volume, Matched Purchase/Sale Quantity and U.S. Dollar Exchange Rate.

- The highest R^2 value on both 10-fold cross-validation (0.908) and independent test data set (0.913) was obtained with the model trained with Exponential Gaussian Process Regression (EGPR).

2. Material and Methods

2.1. Data Collection Process

In this study, the performance of QSVM, MGSVM, EGPR and RGPR methods are compared to predict the MCP in the DAM. The data set used in the modeling phase includes various variables such as Real-time electricity consumption (MWh), hourly electricity generation by source

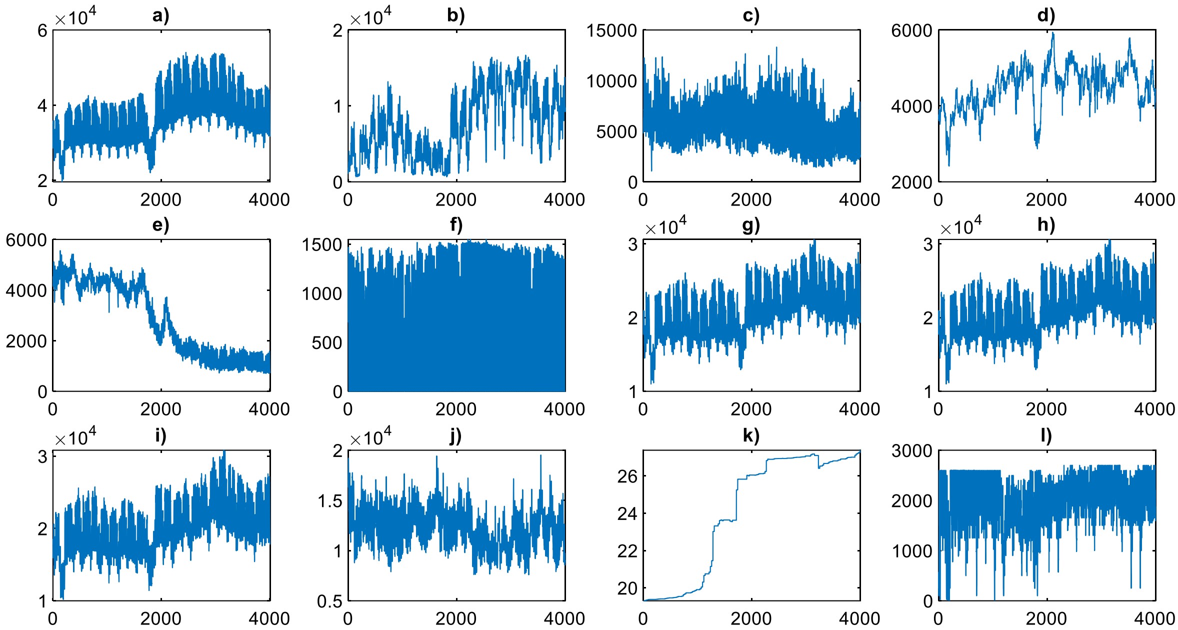

Figure 1: Dependent and independent variables used for model training

(natural gas, hydroelectric power plants (DAM), lignite, run-of-river, solar energy), Purchase/Sale Bid Volume, Matched Purchase/Sale Quantity and U.S. Dollar Exchange Rate. The relevant data are obtained from market and price data published by EPİAŞ and exchange rate data of the CBRT. The data were collected based on date-time information, and all features were normalized to eliminate potential scaling issues that could negatively impact the learning performance of the models. A total of 4001 data points consisting of 24-hour data obtained between 07:00 on 15.04.2023 and 05:00 on 29.09.2023 were used as training data. Before applying machine learning methods, z-score normalization method was used for data preprocessing step.

In order to prove the robustness and generalizability of the models, a 10-fold cross-validation method was used and testing was performed with a data set (25 pieces) that was completely independent of the training data. The independent variables and the visual of the dependent variable used in the study are presented in Figure 1, and the minimum and maximum values are presented in Table 1.

Table 1. Min-Max Value of Independent Variables and Dependent Variable

Variable	Minimum Value	Maximum Value
Real-time electricity consumption (a)	19595	54045
Natural gas (b)	663.71	16682
Hydroelectric power plants (DAM) (c)	1065.2	13310
Lignite (d)	2413.3	5943.2
Run-of-river (e)	703.74	5559.1
Solar energy (f)	0	1547.4
Purchase Bid Volume (g)	10984.4	30601
Sale Bid Volume (h)	10984.4	30601
Matched Sale Quantity (i)	10189	30818
Purchase Sale Quantity (j)	7589.5	19524
U.S. Dollar Exchange Rate (k)	19.3070	27.3752
MCP (l)	0	2700

2.2. Machine Learning and Methods

Machine learning generally began to gain prominence in the 1990s. It is a branch of artificial intelligence that identifies patterns from data and enables algorithms to improve through experience. Machine learning refers to systems that possess the ability to learn and make desired predictions based on input data. Rather than directly following a predefined model, such methods operate by providing predictions based on the input features. Machine learning can extract patterns and relationships from large data points to make predictions, support decision-making, and perform tasks. It is also capable of detecting subtle details that may be overlooked by humans.

There are several techniques used in machine learning:

Supervised Learning: This refers to the process of generating outputs from previously used machine learning results or collecting data from such outputs. These are interactive systems. Classification and regression can be given as examples.

Unsupervised Learning: There is no target information or training algorithm. It enables the discovery of unknown values in the inputs. All processes occur during the algorithm's testing phase.

Semi-Supervised Learning: This is a learning method that lies between supervised and unsupervised learning. It involves the simultaneous use of both labeled and unlabeled input data.

Machine learning has many areas of application, including financial services, healthcare, social media, fraud detection, retail, e-commerce, and sentiment analysis (Mutemi &Bacao, 2024).

In order to forecast the MCP using machine learning methods, multiple independent variables and one dependent variable will be used. The dependent variable refers to the value predicted by the model, while the independent variables are the input factors required to estimate the dependent variable.

2.2.1. Support Vector Machine (SVM)

The main objective of Support Vector Machine (SVM) is to find a function that closely fits all data points within a specified tolerance margin, called the ε-insensitive tube. (Tolun, 2008)

In this approach, only the data points lying outside the ε-insensitive margin influence the optimization process, and these points are referred to as support vectors. The solution of the SVM is modeled as a quadratic programming optimization problem and is solved using the Karush-Kuhn-Tucker (KKT) conditions. The primary objective here is to maximize prediction accuracy while controlling model complexity. (Tolun, 2008)

SVM is suitable not only for linear regression problems but also for data structures that are not linearly separable. In such cases, kernel functions are used to transform the input data into higher-dimensional feature spaces. This transformation enables nonlinear problems to be solved using linear methods. Commonly used kernel functions in practice include linear, polynomial, radial basis function (RBF), and sigmoid kernels. (Tolun, 2008)

The success of the SVM method relies on the principle of structural risk minimization based on statistical learning theory. In this context, the regularization parameter used ensures a balance between the model's error tolerance and its generalization capacity during the learning process. SVM is a preferred regression technique due to its ability to produce highly accurate results in various fields, ranging from financial forecasting to engineering applications (Tolun, 2008).

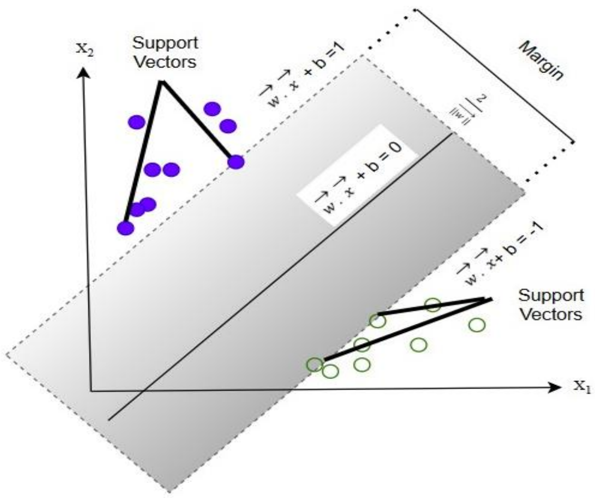

Figure 2: Support Vector Machines (SVM) and the Margin Concept

Figure 2 visually explains the conceptual structure of support vectors, the hyperplane, margin, and the decision boundaries between classes. SVM is a supervised learning algorithm that classifies by determining the optimal hyperplane to separate two classes. The main goal of SVM is to find a decision boundary (hyperplane) that best separates data points belonging to different classes while maximizing the margin between these two classes. This hyperplane is expressed by the equation $\vec{w}\cdot\vec{x}+b = 0$ while the margin-defining boundary lines are described by $\vec{w}\cdot\vec{x}+b = 1$ and $\vec{w}\cdot\vec{x}+b = -1$. The margin is the distance between these two lines and is given by $\frac{2}{\|\vec{w}\|}$,which is aimed to be maximized. The data points closest to these margin boundaries that directly influence the learning process are called "support vectors." In the figure, data points belonging to different classes (e.g., green and purple points), their support vectors, and margins are shown, conceptually illustrating the decision mechanism of the SVM algorithm. This structure enables SVM to provide effective solutions to complex classification problems with high generalization capability.

The mathematical formulation of the SVM model is given in Equation 1 (Şimşek, 2024).

$$f(x) = w \cdot \Phi(x) + b \quad (1)$$

In Equation 1, w represents the weight vector, $\Phi(x)$ is a nonlinear function that maps the data into a higher-dimensional space, and b denotes the bias term of the model.

SVM aims to ensure that the difference between the predicted values and the actual values does not exceed a certain error margin (ε). The loss function used to achieve this is defined in Equation 2 (Şimşek, 2024)

$$L(f(x), y, \mathcal{E}) = f(x) = \begin{cases} 0, & \text{if } |y - f(x)| \leq \mathcal{E} \\ |y - f(x)| - \mathcal{E}, & \text{if } |y - f(x) > \mathcal{E} \end{cases} \quad (2)$$

In Equation 2, Ɛ represents the maximum allowable error margin that the model can ignore, and *L(f(x), y, Ɛ)* denotes the loss function. This loss function is used to calculate the difference between the model's prediction *f(x)* and the actual value *y*. Here, *f(x)* represents the predicted value by the model, *y* is the true (target) value, and Ɛ is the error tolerance, also known as the epsilon tube. The model does not consider any error when the difference between the predicted *f(x)* and the actual *y* remains within the Ɛ boundaries.

According to Equation 2, there are two cases. (Şimşek, 2024)

1. The case where $|\mathbf{y} - \mathbf{f(x)}| \leq \mathcal{E}$; if the predicted value is at most ε\varepsilonε away from the true value, the model considers the error to be zero. In other words, small errors are ignored.
2. If $\boldsymbol{y - f(x) > \mathcal{E}}$ meaning the error exceeds the ε\varepsilonε value, this indicates that the model only considers large errors. Small errors are not penalized by the model, thus preventing overfitting.

Quadratic Support Vector Machine (QSVM)

QSVM is an advanced version of SVM. SVM is a linear classifier. QSVM is also used for nonlinear classifications. QSVM is used for large-scale classifications. It has fast prediction capabilities for binary classifications. Its flexibility is better than SVM.

$$\boldsymbol{K(y_s, y_t) = (1 + y_s y_t)^2} \quad (3)$$

In Equation 3, the parameters $\boldsymbol{y_s}$ and $\boldsymbol{y_t}$ are data points (Bhati, 2020).

Medium Gaussian Support Vector Machine (MGSVM)

MGSVM is a different version of SVM. While not sufficient for multiple classifications, it is sufficient for binary classifications. Its flexibility is better than SVM's. It is more balanced than SVM.

2.2.2. Gaussian Process Regression (GPR)

Gaussian processes are a supervised learning technique used for solving both regression and probabilistic classification problems. GPR relies on a non-parametric Bayesian approach for inference. Instead of estimating a distribution over parameters, Gaussian processes are used to directly model the distribution of the target function.

GPR is a probabilistic learning method structured using kernel functions. In the model, inputs x_i and outputs y_i are assumed to be drawn from an unknown probability distribution. GPR aims to predict the distribution of the output variable given an input vector. This approach models the statistical dependencies of the output values via kernel functions by assigning a random variable to each input. As shown in Equation 4, the fundamental structure of GPR is based on the framework of LR but extends beyond linear relationships, providing flexibility to model complex, nonlinear patterns (Yılmaz, 2023).

$$y = x^T \beta + \epsilon \quad (4)$$

In Equation 4, y represents the output (dependent variable) to be predicted, x is a column vector representing the input variables (feature vector), β is the vector of coefficients (parameters) corresponding to each input variable, and ε denotes the error term (noise) in the model, which is assumed to follow a zero-mean normal distribution. This formulation indicates that each observation is composed of a linear combination of x and β vectors plus a small random error ε. In other words, while the model attempts to directly estimate the true relationship, it also accounts for random deviations present in the data.

$$p(f) = N(f \mid 0, K) \quad (5)$$

In Equation 5, considering Gaussian processes, the term $p(f)$ is treated as a zero-mean process. The matrix K, representing the kernel function, is a covariance matrix generated by the kernel function that defines the covariance structure among the inputs.

$$K(x_i, x_j) = \exp - \frac{||x - x_i||^2}{2\sigma^2} \quad (6)$$

In Equation 6, $K_{ij} = K(x_i, x_j)$ is defined such that the element in the ith row and jth column of the matrix K represents the value of the kernel function between the inputs x_i and x_j.

$$p(y) = \int p(y|f)p(f)df = N(f \mid 0, K_y) \quad (7)$$

In Equation 7, the variable y is modeled based on this structure, incorporating the kernel-based covariance and the assumed noise in the observations. This formulation reflects how the target values are generated as a combination of the underlying function values and the noise component.

Exponential Gaussian Process Regression (EGPR)

EGPR is a widely used model in GPR. EGPR is preferred because of its simplicity and its ability to better detect connections between data points.

$$k(x, x') = \sigma^2 \exp\left(-\frac{|x - x'|}{l}\right) \quad (8)$$

In Equation 8, l is the length scale parameter indicating the relationship between points. σ^2 is the variance parameter (Danturti, 2025). This equation establishes a relationship in which similarity increases or decreases exponentially depending on the distance between points.

Rational Quadratic Gaussian Process Regression (RGPR)

RGPR is the most commonly used GPR model after EGPR. It is used to model data at different scales. It can be used in many areas, such as data requiring multivariate statistics and image analysis.

$$K(x_i, x_j) = \sigma_f^2\left(1 + \frac{(||x_i - x_j||)^2}{2\alpha l^2}\right) \quad (9)$$

In Equation 9, σ^2 represents the variance parameter, l represents the length scale, and α represents the scale mixing parameter. A smaller α value indicates a more multi-scale structure. The $x_i - x_j$ value represents the distance between the data. These parameters help RGPR to identify different data in the model and determine the flexibility of the model. As the model's flexibility increases, the risk of overfitting also increases.

2.3. Performance Metrics (PM)

RMSE, MSE, MAE, and R^2 were used to evaluate the performance of models created using machine learning methods. These performance criteria numerically express the accuracy of the models' predicted values. The mathematical equivalents of the performance criteria are given in Equations 10-13 (Yılmaz et al., 2025).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (13)$$

In Equation 10-13, y_i represents the actual value, and $\hat{y}_i$ represents the predicted value, $\bar{y}_i$ represents the mean value of the actual data.

3. Results and Discussion

In this study, the performance of QSVM, MGSVM, EGPR, and RGPR methods is compared to predict the MCP in the DAM. The reasons for selecting these methods in this study are not to replicate the classic SVM and GPR, but to evaluate the performance of their advanced versions, particularly in data structures with high volatility in energy markets. The reasons for selecting QSVM and MGSVM methods are their greater generalization ability by creating more flexible decision boundaries in time-series structures. EGPR is preferred because its better models' sharp changes over short distances and is more sensitive to sudden changes. RGPR is preferred because it predicts short- and long-term variables in a balanced manner.

The performance of the models created with the QSVM, MGSVM, EGPR, and RGPR methods was tested with both 10-fold cross-validation and an independent test dataset. The aim of the two-validation structure is to evaluate the consistency of the models in different data segments and their prediction performance in real data.

3.1. 10-fold cross-validation Results

Using 10-fold cross-validation, the entire dataset was divided into 10 equal parts. Nine of these 10 parts were used for training and one for validation. After repeating this process 10 times, the model's validation performance was calculated by averaging the values obtained from all trials.

The actual and predicted values resulting from the 10-fold cross-validation of the model created using the EGPR method are shown in Figure 3. The performance criteria obtained for each floor and the average of the out-of-fold (OFF) validation results obtained for each floor are detailed in Table 2.

Table 2. 10-fold cross-validation results for EGPR Model

Fold	R^2	MSE	RMSE	MAE
1	0.901	26150	161.71	120.08
2	0.908	24672	157.07	117.81
3	0.901	30173	173.70	132.78
4	0.905	25745	160.45	120.18
5	0.907	26190	161.83	119.05
6	0.910	24839	157.60	117.32
7	0.909	25176	158.67	118.86
8	0.918	28960	170.18	126.26
9	0.898	26973	164.23	118.93
10	0.920	21387	146.24	112.07
OOF/ Validation	**0.908**	**26026.2**	**161.33**	**120.33**

The 10-fold cross-validation test results of the model obtained with the GPR method are presented in detail in Table 3. Actual values and predicted values are presented in Figure 4.

Table 3. 10-fold cross-validation results for RGPR Model

Fold	R^2	MSE	RMSE	MAE
1	0.917	23918	154.65	114.12
2	0.921	24606	156.86	113.91
3	0.898	30140	173.61	129.73
4	0.890	30968	175.98	134.59
5	0.922	26378	162.41	120.15
6	0.905	24364	156.09	115.98
7	0.897	25161	158.62	122.62
8	0.872	28672	169.33	123.78
9	0.913	26112	161.59	117.84
10	0.918	23246	152.47	115.20
OOF/ Validation	**0.907**	**26356.2**	**162.35**	**120.79**

The actual and predicted values resulting from the 10-fold cross-validation of the model created with the MGSVM method are shown in Figure 5. The performance criteria obtained for each floor and the average of the out-of-fold (OFF) validation results obtained for each floor are given in detail in Table 4.

Table 4. 10-fold cross-validation results for MGSVM Model

Fold	R^2	MSE	RMSE	MAE
1	0.856	32873	181.31	132.26
2	0.896	31014	176.11	133.44
3	0.885	34713	186.31	143.57
4	0.852	36132	190.08	144.05
5	0.865	36158	190.15	146.30
6	0.865	44493	210.93	151.14
7	0.879	34407	185.49	141.70
8	0.878	40281	200.70	146.10
9	0.851	39505	198.76	149.68
10	0.848	42220	205.47	153.88
OOF/ Validation	**0.869**	**37178.0**	**192.82**	**144.21**

The 10-fold cross-validation test results of the model obtained with the QSVM method are presented in detail in Table 5. Actual and predicted values are given in Figure 6.

Table 5. 10-fold cross-validation results for QSVM Model

Fold	R^2	MSE	RMSE	MAE
1	0.857	41129	202.8	157.61
2	0.857	46182	214.9	163.84
3	0.837	48387	219.97	172.13
4	0.832	54135	232.67	179.84
5	0.842	39322	198.3	155.86
6	0.790	48457	220.13	170.7
7	0.818	45415	213.11	167.47
8	0.847	48035	219.17	169.65
9	0.835	45307	212.86	169.65
10	0.842	44626	211.25	165.73
OOF/ Validation	**0.838**	**46099.6**	**214.71**	**166.65**

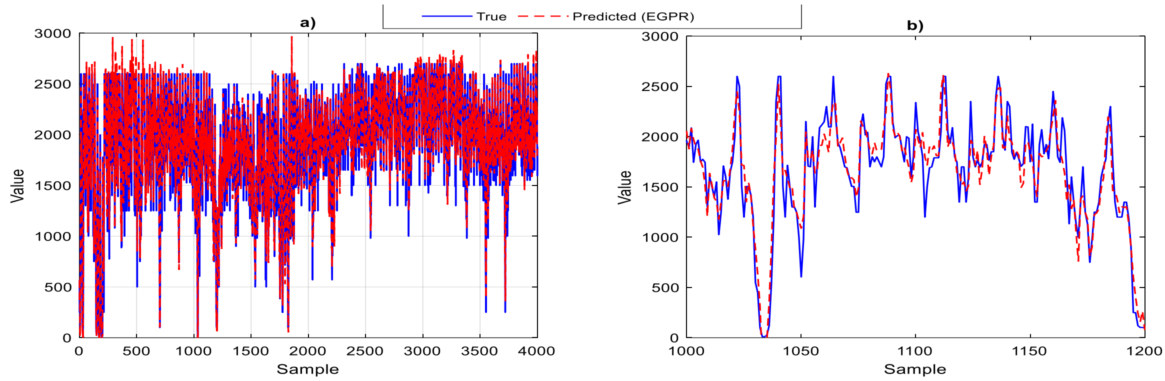

Figure 3: 10-fold cross-validation results for EGPR (a) Actual vs Predicted Values, (b) Zoomed View of (a)

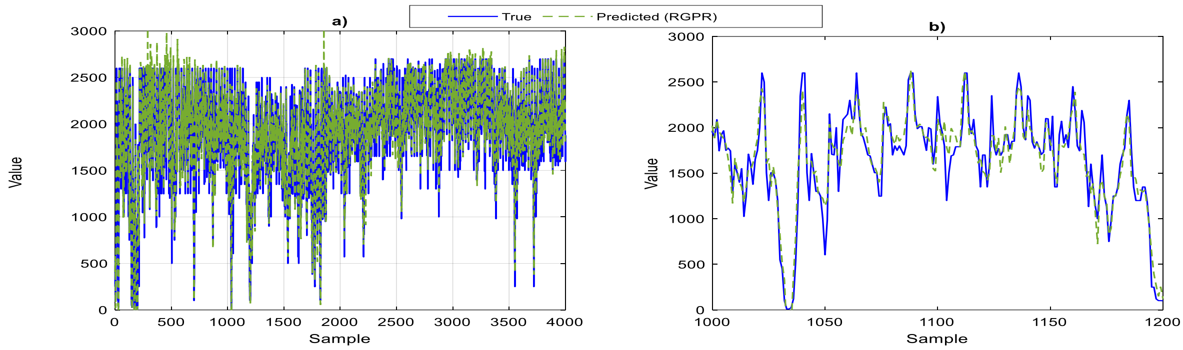

Figure 4: 10-fold cross-validation results for RGPR (a) Actual vs Predicted Values, (b) Zoomed View of (a)

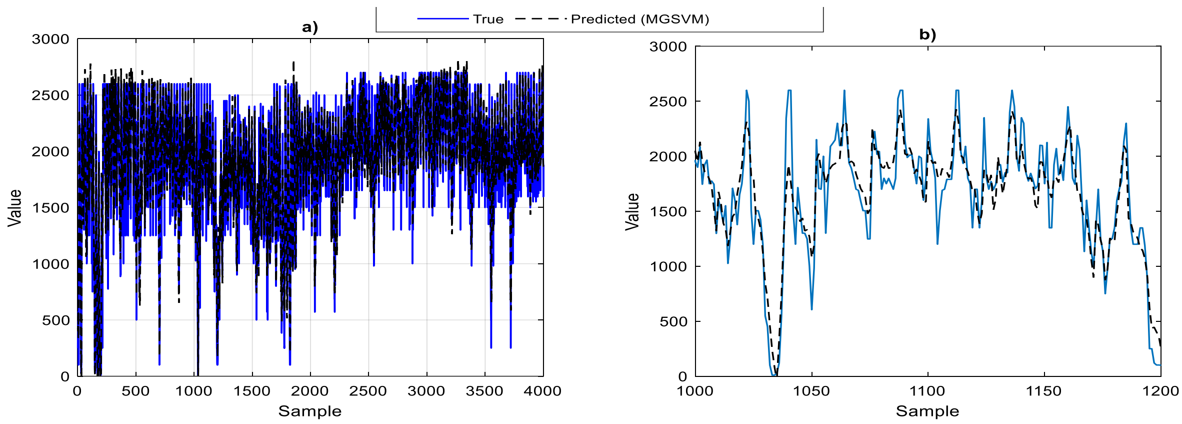

Figure 5: 10-fold cross-validation results for MGSVM (a) Actual vs Predicted Values, (b) Zoomed View of (a

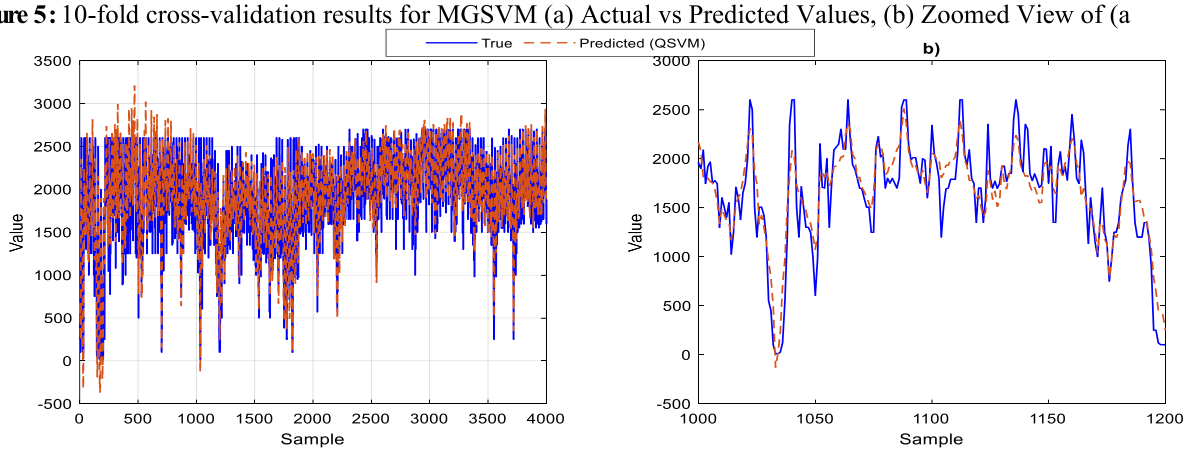

Figure 6: 10-fold cross-validation results for QSVM (a) Actual vs Predicted Values, (b) Zoomed View of (a)

3.2. Independent Dataset Results

During training, the models only worked on training data. The models' performance against new data, which they have not encountered in real life, was tested with an independent data set. The independent data set consists of 25 data sets between September 29, 2023, at 6:00 AM, and September 30, 2023, at 6:00 AM. The independent data set used in the study is shown in Figure 7. The minimum and maximum values for these variables are detailed in Table 6.

Table 6. Min-Max Value of Independent Variables and Dependent Variable for independent dataset

Variable	Minimum Value	Maximum Value
Real-time electricity consumption (a)	30851	42792
Natural gas (b)	7621.6	13913
Hydroelectric power plants (DAM) (c)	1661.2	7113.2
Lignite (d)	4344	4605.7
Run-of-river (e)	649.89	1348.8
Solar energy (f)	0	1376.4
Purchase Bid Volume (g)	18826	26702
Sale Bid Volume (h)	18826	26702
Matched Sale Quantity (i)	18017	26373
Purchase Sale Quantity (j)	9123.1	13422
U.S. Dollar Exchange Rate (k)	27.3752	27.3752
MCP (l)	1864	2700

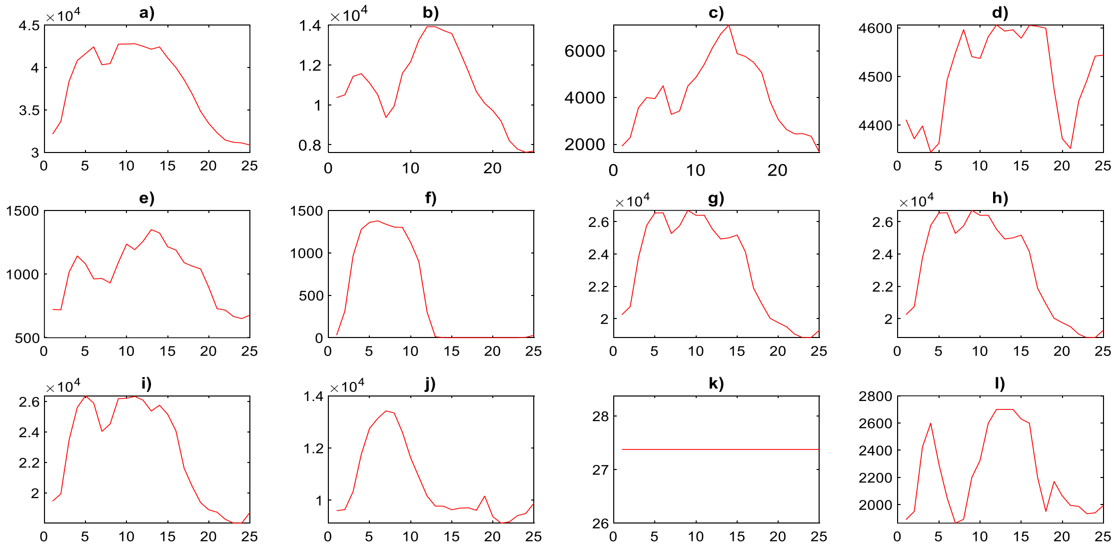

Figure 7: Dependent and independent variables used for independent dataset

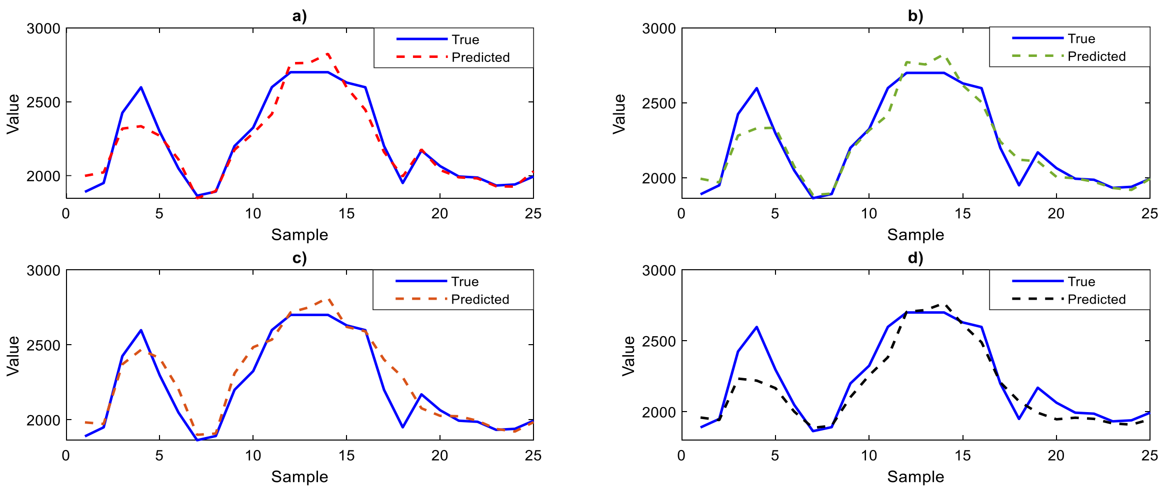

Figure 8: Independent dataset results for (a) EGPR, (b) RGPR (c) QSVM (d) MGSVM

Performance metrics for the independent test data of the model created with the EGPR, RGPR, QSVM, and MGSVM methods are given in Table 7. Figure 8 shows the success of the models created with the EGPR, RGPR, QSVM, and MGSVM methods in predicting the real test data.

Table 7. Independent dataset results for EGPR, RGPR, QSVM and MGSVM Models

Methods	R^2	MSE	RMSE	MAE
EGPR	**0.913**	**7734**	**87.94**	**61.56**
RGPR	0.905	8381	91.55	62.01
QSVM	0.876	10925	104.52	73.20
MGSVM	0.844	13781	117.39	80.551

When comparing independent dataset results, it was observed that EGPR had the lowest error metrics overall. It was also found that EGPR had the highest coefficient of determination value.

4. Conclusion

In this study, the performance of the QSVM, MGSVM, EGPR and RGPR methods is compared to predict the MCP in the DAM. 4001 datasets were used for the training phase. All operations were performed on the MATLAB/2021b platform. A 10-fold cross-validation and a data set completely independent of the training data were used to test the generalizability and accuracy of the models. When comparing the 10-fold cross-validation results, the model trained with EGPR had the highest R squared value. The high R2 value of the model trained with EGPR indicates that its generalization ability is better than that of models trained with RGPR, MGSVM, and QSVM. Furthermore, the MSE, MAE, and RMSE error measurements of the model trained with EGPR are lower than those of models trained with RGPR, MGSVM, and QSVM. This indicates that the model trained with EGPR makes predictions closer to the actual data. Comparing the results obtained for independent test data, the coefficients of determination of the models trained with EGPR, RGPR, MGSVM, and QSVM are *0.913*, *0.905*, *0.844*, and *0.876*, respectively. The model trained with EGPR was observed to yield better results in MSE, MAE, and RMSE error measurements than models trained with other machine learning methods. Future studies are considering using a hybrid of machine learning and deep learning methods.

Article Information

Financial Disclosure: The author (s) has not received any financial support for the research, authorship or publication of this study.

Authors' Contrtibution: Concept: V.Ö., M.Y.; Design: V.Ö.; Resources: V.Ö., M.Y.; Supervision: M.Y.; Data Collection: V.Ö.; Analysis: V.Ö., M.Y.; Literature Search: V.Ö., M.Y.; Writing Manuscript: V.Ö., M.Y.; Critical Review: M.Y.

Conflict of Interest/Common Interest: No conflict of interest or common interest has been declared by the authors.

Ethics Committee Approval: This study does not require ethics committee permission or any special permission.

Artificial Intelligence Statement: The author(s) bear full responsibility for the content and accuracy of their work, including any use of artificial intelligence (AI) technologies, and confirm that they have read the AI Policy, which is accessible on the journal's website.

References

Awad, M. L., Casey, K. E., Geevarghese, A. S., Miller, J. C., Rahimi, A. F., Sheffrin, A. Y., ... & Wolak, F. A. (2010). Using market simulations for economic assessment of transmission upgrades: *Application of the California ISO approach. Restructured Electric Power Systems*, 241-270.

Bhati, B. S., & Rai, C. S. (2020). Analysis of support vector machine-based intrusion detection techniques. *Arabian Journal for Science and Engineering*, 45(4), 2371-2383.

Danturti, S., Mishra, P., Mishra, N., ST, S., Sahoo, B., & Bhowmik, P. (2025). Exponential GPR based Power Prediction for Solid Oxide Fuel Cell under H_2 and O_2 Flow Maloperation. *Results in Engineering*, 105412.

Das, A., & Schlüter, S. (2025). Gaussian Process Regression with a Hybrid Risk Measure for Dynamic Risk Management in the Electricity Market. *Risks*, 13(1), 13.

Das, A., Schlüter, S., & Schneider, L. (2024). Electricity Price Prediction Using Multi-Kernel Gaussian Process Regression combined with Kernel-Based Support Vector Regression. *arXiv preprint* arXiv:2412.00123.

Demirezen, S., & Çetin, M. (2021). Rassal Orman Regresyonu Ve Destek Vektör Regresyonu İle Piyasa Takas Fiyatinin Tahmini. *Nicel Bilimler Dergisi*, 3(1), 1-15.

Díaz, G., Coto, J., & Gómez-Aleixandre, J. (2019). Prediction and explanation of the formation of the Spanish day-ahead electricity price through machine learning regression. *Applied Energy*, 239, 610-625.

Enerji Piyasaları İşletme A.Ş. (EPİAŞ). (2025, January 25). Şeffaflık Platformu. Erişim adresi: https://seffaflik.epias.com.tr/home

EPİAŞ., (2016), Gün İçi Piyasası Kullanıcı Kılavuzu.

Ertaylan, A., Aktaş, Ö., & Doğan, Y. (2021). Yapay sinir ağları ile piyasa takas fiyatı tahminlemesi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 23(67), 93-105.

Foruzan, E., Scott, S. D., & Lin, J. (2015*). A comparative study of different machine learning methods for electricity prices forecasting of an electricity market*. In 2015 north American power symposium (NAPS) (pp. 1-6). IEEE

Mutemi, A., & Bacao, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. *Big Data Mining and Analytics*, 7(2), 419-444.

Şimşek, A. İ. (2024). Comparative Analysis of Machine and Deep Learning Methods in Estimating the Turkish Electricity Market Clearing Price. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 36(2), 859-867.

Spiliotis, E., Doukas, H., Assimakopoulos, V., & Petropoulos, F. (2021). Forecasting week-ahead hourly electricity prices in Belgium with statistical and machine learning methods. *In Mathematical modelling of contemporary electricity markets* (pp. 59-74).

Tolun, S. (2008). *Destek vektör makineleri: Banka başarısızlığının tahmini üzerine bir uygulama*. İktisadî Araştırmalar Vakfı.

Yanar, A., & Akay, M. F. (2020). Prediction of electricity market clearing price using machine learning and deep learning. *Ç. Ü Fen ve Mühendislik Bilimleri Dergisi*, 39(9), 137-141.

Yılmaz, M., Çorapsız, M. R., & Çorapsız, M. F. (2025). A novel maximum power point tracking approach based on fuzzy logic control and optimizable Gaussian Process Regression for solar systems under complex environment conditions. *Engineering Applications of Artificial Intelligence*, 141, 109780.

Yılmaz, M., Kaleli, A., & Çorapsız, M. F. (2023). Machine learning based dynamic super twisting sliding mode controller for increase speed and accuracy of MPPT using real-time data under PSCs. *Renewable Energy*, 219, 119470.