applied sciences

MDPI

Article

Development of New Electricity System Marginal Price Forecasting Models Using Statistical and Artificial Intelligence Methods

Mehmet Kızıldağ [1,2], Fatih Abut [2,*] and Mehmet Fatih Akay [2]

[1] SABIS Bilgi Teknolojileri ve Sanayi Ticaret Ltd. Şti, 01330 Adana, Turkey; mehmet@sabisahbs.com
[2] Department of Computer Engineering, Faculty of Engineering, Çukurova University, 01330 Adana, Turkey; mfakay@cu.edu.tr
* Correspondence: fabut@cu.edu.tr

Abstract: The System Marginal Price (SMP) is the cost of the last unit of electricity supplied to the grid, reflecting the supply–demand equilibrium and serving as a key indicator of market conditions. Accurate SMP forecasting is essential for ensuring market stability and economic efficiency. This study addresses the challenges of SMP prediction in Turkey by proposing a comprehensive forecasting framework that integrates machine learning, deep learning, and statistical models. Advanced feature selection techniques, such as Minimum Redundancy Maximum Relevance (mRMR) and Maximum Likelihood Feature Selector (MLFS), are employed to refine model inputs. The framework incorporates time series methods like Multilayer Perceptron (MLP), Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Convolutional LSTM (ConvLSTM) to capture complex temporal patterns, alongside models such as Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Extreme Learning Machine (ELM) for modeling non-linear relationships. Model performance was evaluated using the Mean Absolute Percentage Error (MAPE) across regular weekdays, weekends, and public holidays. XGBoost combined with MLFS consistently achieved the lowest MAPE values, demonstrating exceptional accuracy and robustness. Among all of the models, XGBoost combined with MLFS consistently achieved the lowest MAPE values, demonstrating superior accuracy and robustness. The results highlight the inadequacy of traditional models like ARIMA and SARIMA in capturing non-linear and highly volatile patterns, reinforcing the necessity of using advanced techniques for effective SMP forecasting. Overall, this study presents a novel and comprehensive approach tailored for complex electricity markets, significantly enhancing predictive reliability by incorporating economic indicators and sophisticated feature selection methods.

Keywords: system marginal price; machine learning; feature selection; xgboost; extreme learning machine; ARIMA; SARIMA; predictive modeling

check for updates

Citation: Kızıldağ, M.; Abut, F.; Akay, M.F. Development of New Electricity System Marginal Price Forecasting Models Using Statistical and Artificial Intelligence Methods. *Appl. Sci.* **2024**, *14*, 10011. https://doi.org/10.3390/app142110011

Academic Editor: Grigorios Beligiannis

Received: 9 October 2024
Revised: 28 October 2024
Accepted: 30 October 2024
Published: 2 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The System Marginal Price (SMP) is a pivotal metric in the electricity market, serving as a crucial indicator of the dynamic balance between supply and demand. As the equilibrium price at which electricity is traded, SMP is a vital component in the decision-making processes of energy producers, distributors, consumers, and regulators [1]. It provides a transparent benchmark that reflects market conditions and determines the cost at which the marginal unit of electricity is delivered. Consequently, understanding and accurately forecasting SMP trends is essential for optimizing resource allocation, managing financial risks, and informing policy decisions [1,2].

In competitive electricity markets, SMP is determined by the intersection of supply bids and demand requirements for each trading interval, typically on an hourly or daily basis. The highest accepted bid sets the SMP for that period, establishing the price signal that guides market behavior. However, this price-setting mechanism is not static. It is influenced

by a complex web of factors, including fluctuations in energy demand, variability in supply sources, policy interventions, and broader macroeconomic conditions. Such factors introduce significant volatility into SMP values, making precise forecasting a challenging yet indispensable task for energy market participants [3,4].

Accurate SMP forecasting is crucial for several reasons. For energy producers, it enables optimal resource allocation by determining when to deploy different generation units, thereby maximizing profitability. Producers can schedule their lower-cost baseload plants during periods of stable demand and activate higher-cost peaking units only during anticipated price surges, based on SMP predictions [5]. Similarly, distributors and consumers can leverage SMP forecasts to adjust their procurement strategies, hedging against price volatility and reducing overall energy costs. From a regulatory perspective, SMP serves as a barometer of market efficiency and competitiveness. Deviations in SMP can signal market inefficiencies or the need for policy interventions, such as capacity payments or infrastructural upgrades, to maintain system stability. Thus, accurate SMP forecasting not only supports operational decisions but also aids in the strategic planning and regulation of the electricity market [6,7].

In the Turkish electricity market, SMP holds strategic importance due to the unique characteristics of the market structure and the rapid integration of renewable energy sources. With its growing reliance on imported natural gas and the ongoing transition toward a more sustainable energy mix, the Turkish electricity market is particularly susceptible to volatility in fuel prices and exchange rates [8]. Additionally, the implementation of policy measures such as the Renewable Energy Support Mechanism (YEKDEM) has introduced further complexities into SMP calculations. Consequently, accurate SMP forecasting in Turkey requires models that are tailored to these specific conditions, incorporating a diverse set of variables to capture the multifaceted nature of price fluctuations [9].

This study introduces a comprehensive forecasting framework that integrates a range of statistical, machine learning, and deep learning models, alongside advanced feature selection techniques. The primary objective is to enhance SMP prediction accuracy and reliability, providing market participants with critical insights for informed decision-making. To achieve this, the study incorporates feature selection algorithms like Minimum Redundancy Maximum Relevance (mRMR) and Maximum Likelihood Feature Selector (MLFS), which filter out redundant or less informative variables, thereby refining input features to enhance model interpretability and predictive performance. The modeling framework includes traditional statistical methods, such as the Autoregressive Integrated Moving Average (ARIMA) and its Seasonal variant (SARIMA), serving as baseline models for capturing linear trends and seasonal patterns within SMP data. However, to handle non-linear dependencies and high-dimensional interactions often present in electricity price data, the study incorporates machine learning models, including Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Extreme Learning Machine (ELM). Each of these models contributes unique strengths for handling complex non-linear data patterns.

In addition to machine learning techniques, the framework leverages advanced deep learning models to capture temporal dependencies within SMP data more effectively. Long Short-Term Memory (LSTM) networks are employed for their capability to handle sequential data and address challenges like the vanishing gradient problem, while Bidirectional LSTM (Bi-LSTM) provides a bidirectional perspective on temporal data, capturing both past and future dependencies. Convolutional LSTM (ConvLSTM) is also incorporated to model spatial and temporal dependencies jointly, enhancing the model's capability to understand complex patterns over time. To further improve model robustness and adaptability, tree-based ensemble methods like eXtreme Gradient Boosting (XGBoost) are used. XGBoost efficiently manages non-linear interactions and mitigates overfitting through iterative boosting and regularization techniques. Each model in the framework is fine-tuned with hyperparameter optimization, leveraging grid search and cross-validation to achieve optimal predictive performance. This integrated methodology, combining diverse forecasting models with advanced feature selection techniques (mRMR and MLFS), enables

a flexible and robust forecasting framework tailored to address the specific dynamics of SMP in the Turkish electricity market.

The findings of this study make several key contributions to the field of SMP forecasting. First, the integration of machine learning and statistical techniques, coupled with rigorous feature selection, enables the development of hybrid models that effectively capture the non-linear and dynamic nature of SMP. Second, the use of advanced feature selection algorithms significantly enhances model performance by reducing dimensionality and focusing on the most relevant predictors, thereby improving interpretability and predictive accuracy. Third, the application of these models to the Turkish electricity market addresses a critical gap in the literature, as previous research has largely focused on other regions with different market structures and regulatory environments. The results demonstrate that both the ELM and XGBoost models yield high prediction accuracy, underscoring their potential for real-world SMP forecasting applications.

The rest of this study is structured as follows: Section 2 provides an overview of the existing literature on SMP forecasting, identifying key gaps and areas for further research. Section 3 details the theoretical foundations of the machine learning and statistical models used in this study, as well as the feature selection methodologies applied. Section 4 describes the dataset and methodology used in the research, while Section 5 presents the results and provides an in-depth discussion of the findings. Finally, Section 6 outlines the study and suggests potential directions for future research.

2. Related Works

In recent years, significant advancements have been made in the field of SMP forecasting, with researchers leveraging a variety of computational methods to enhance predictive accuracy and reliability. Table 1 provides a detailed examination of notable studies conducted in recent years, showcasing the diverse approaches and methodologies employed to tackle the complexities of SMP forecasting. These studies utilize a range of techniques from traditional statistical models to advanced machine learning algorithms, each contributing uniquely to the evolving landscape of electricity price forecasting. The synthesis of findings from these studies highlights the continuous improvements and innovations that characterize this vital area of research.

Table 1. Review of the literature on SMP forecasting in recent years.

Authors	Year	Predictor Variables	Methods	MAPE Values (%)
Özgüner et al. [10]	2017	Historical SMP, load, SDAP, NVI	ANN	9.77–21.81
Pan et al. [11]	2017	Historical SMP, weather values	AdaBoost, SVM, BPNN	3.84–8.43
Jufri et al. [12]	2019	Historical SMP	ANN	5.36
Panapakidis and Moschakis [13]	2019	Historical SMP	FFNN	7.42–9.22
Akay et al. [14]	2021	Historical SMP	LSTM, SVM	2.97–19.02
Shim et al. [15]	2023	Forecasted load, date, fuel price, fuel capacity, Historical SMP	XGBoost, RF, LightGBM, LR	2.70–6.93
Kim et al. [16]	2023	Baltic Dry Index, natural gas, coal, crude oil, US dollar index, exchange rate	SVM, DNN	4.70–5.20
Razak et al. [17]	2023	Historical SMP, demand forecast, generation mix forecast	LSSVM, GA	8.55–9.18

In their research, Özgüner et al. (2017) [10] implemented Artificial Neural Networks (ANNs) to project day-ahead hourly market clearing prices and marginal prices in the Turkish electricity market. The models were trained on historical pricing data, aiming to minimize the daily mean absolute percentage error. During the forecasting process, key input variables such as historical load, forecasted load, historical SMP, historical System Day-Ahead Price (SDAP), forecasted SDAP, historical Net Volume Index (NVI), and

forecasted NVI were employed to enhance the accuracy and reliability of the predictions. The approach included enhancements like volatility analysis and the application of the Bienaymé–Chebyshev inequality, allowing for probabilistic forecasting within specified boundaries, which resulted in MAPE values ranging from 9.77% to 21.81%.

Pan et al. (2017) [11] developed an ensemble learning model utilizing AdaBoost, enhanced with SVM and Back Propagation Neural Networks for short-term SMP forecasting. This method focused on analyzing primary factors affecting short-term electricity prices and proved to be more effective than traditional models, evidenced by MAPE values between 3.84% and 8.43%, demonstrating the method's substantial improvements in predictive accuracy.

The 2019 study by Jufri et al. [12] featured an ANN tailored for day-ahead SMP forecasting that integrated a combination of long-term and short-term historical data. Optimization was achieved through k-fold cross-validation. Input variables were meticulously selected using the Pearson correlation coefficient and discrete Fréchet distance, leading to a notably low MAPE of 5.36%, which marked an improvement over standard forecasting models.

Panapakidis and Moschakis (2019) [13] explored the potential of Feedforward Neural Networks (FFNNs) to forecast the System Marginal Price in the Greek energy market. Their research involved testing various FFNN configurations to understand the influence of network parameters on prediction outcomes. Their results indicated a MAPE between 7.42% and 9.22%, with Adaptive Neuro-Fuzzy Inference System models demonstrating particularly low error rates, suggesting their effectiveness for this application.

Akay et al. (2021) [14] explored SMP forecasting using LSTM and SVM models for one- and two-day predictions in the Turkish electricity market. Key predictor variables include historical SMP data with selected time lags for LSTM and periodic lookback values for SVM, optimizing predictive accuracy. The results revealed that the LSTM models yield MAPE values ranging from 2.97% to 19.02%, while SVM models show values between 3.93% and 15.23%, illustrating the impact of time lag and periodic lookback selection on forecast precision.

In a more recent study, Shim et al. (2023) [15] utilized a two-step SMP forecasting process beginning with load forecasting implemented through XGBoost and supported by Random Forest (RF), Light Gradient Boosting Machine, and Linear Regression. The research emphasized feature selection and hyperparameter optimization, yielding MAPE values from 2.70% to 6.93%. Notably, ensemble models combining LightGBM and RF delivered an outstanding performance, especially across different seasons, highlighting the robustness and accuracy of these combined approaches.

Kim et al. (2023) [16] presented an online machine-learning model for real-time SMP forecasting in South Korea using multiple economic indicators, including data from the transportation, energy, and monetary sectors. SVM, simple deep neural networks (DNNs), and DNN models were tested, with simple DNN performing best. Key predictors include the Baltic Dry Index, natural gas, coal, and oil prices. The model's MAPE values averaged 5.2%, demonstrating high accuracy. Real-time model updates improved forecast precision, beneficial for energy planning, and decision-making.

Finally, Razak et al. (2023) [17] proposed a hybrid model combining Least Squares Support Vector Machines (LSSVMs) with a Genetic Algorithm (GA) for short-term System Marginal Price (SMP) forecasting in Peninsula Malaysia. This model leverages GA to optimize both the LSSVM parameters and forecast inputs, enhancing accuracy. Predictor variables include historical SMP values from the previous three days and a week earlier and demand forecasts and generation mix forecasts covering different energy sources. The results revealed that daily forecasts achieve a MAPE of around 9.18%, while weekly forecasts show improved accuracy with a MAPE of 8.55%, demonstrating the model's potential in refining SMP prediction.

This work distinguishes itself from the existing body of the literature by employing a comprehensive methodological framework that integrates state-of-the-art feature selection techniques (i.e., mRMR and MLFS) with a diverse set of statistical and machine

learning-based models. Unlike previous studies that often focus on a single approach or limited feature set, this research evaluates the effectiveness of various combinations of models and feature selection methods to determine the optimal strategy for forecasting SMP under different market conditions. Moreover, this study systematically investigates the influence of macroeconomic variables and renewable generation sources, which are frequently overlooked in the literature, on SMP dynamics. The unique contributions of this work are threefold:

- Comprehensive Evaluation of Advanced Feature Selection Techniques: This study is among the first to systematically evaluate the impact of mRMR and MLFS on SMP forecasting, demonstrating how these methods can significantly improve predictive accuracy by refining the input space and enhancing model interpretability. This approach addresses a critical gap in the literature by showcasing the value of robust feature selection in electricity market forecasting;
- Detailed Comparison of Diverse Models and Market Scenarios: The research encompasses a broad range of models, including ELM, LSTM, Bi-LSTM, ConvLSTM, and traditional statistical approaches, applied to a complex dataset that spans regular days, public holidays, and periods of extreme market volatility. This comprehensive comparison provides a holistic view of model performance, highlighting the strengths and weaknesses of each approach under varying conditions;
- Integration of Macroeconomic and Renewable Generation Variables: By incorporating economic indicators such as exchange rates and stock market indices, as well as detailed generation data by source (e.g., hydro, wind, solar, and thermal), the study presents an analysis of the factors influencing SMP. This integration of broader contextual variables offers a more complete understanding of market dynamics, setting this work apart from others that typically rely on narrower datasets.

3. Methods

This section outlines the methodological framework, covering the diverse range of forecasting models and feature selection techniques employed in the study. It provides an in-depth explanation of the supervised learning-based, statistical, and time series methods used, along with the feature selection algorithms that optimize the model's performance.

3.1. Supervised Learning Based Methods

Supervised learning methods are designed to model the relationship between input variables and target outputs. This subsection introduces the SVM, XGBoost, and ELM models, explaining their relevance and applicability to SMP forecasting.

3.1.1. Support Vector Machine (SVM)

SVM models the relationship between input features and continuous targets using a linear function in a high-dimensional feature space [18]. The optimization problem is formulated as

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2} \parallel \mathbf{w} \parallel^2 + C\sum_{i=1}^{N} \left(\xi_i + \xi_i^*\right) \tag{1}$$

subject to

$$\begin{array}{l} y_i - \left(\mathbf{w}^T\phi(x_i) + b\right) \leq \epsilon + \xi_i \\ \left(\mathbf{w}^T\phi(x_i) + b\right) - y_i \leq \epsilon + \xi_i^* \end{array} \tag{2}$$

where w is the weight vector, b is the bias term, and $\phi(x_i)$ represents the nonlinear mapping of the input x_i to a higher-dimensional space. The parameters ξ and ξ^* are slack variables that allow some flexibility by tolerating prediction errors within a margin ϵ. The regularization parameter C controls the trade-off between minimizing model complexity (i.e., $\parallel w \parallel$) and the amount of tolerated error. This framework makes SVM robust against overfitting, especially when dealing with high-dimensional data.

3.1.2. Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced implementation of the gradient boosting framework, designed to enhance model performance and computational efficiency. XGBoost constructs an ensemble of weak learners, typically decision trees, in a sequential manner, where each new tree is trained to correct the errors made by the ensemble of previous trees. The model iteratively improves the predictive accuracy by minimizing a defined objective function, which consists of two main components: a loss function that measures the discrepancy between the actual and predicted values and a regularization term that controls the model complexity [19]. The objective function is minimized as follows:

$$L(\theta) = \sum_{i=1}^{N} l(y_i, \widehat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

where

- $l(y_i, \widehat{y}_i)$ is the loss function that quantifies the difference between the true value y_i and the predicted value $\widehat{y}_i$. Common loss functions used include the Mean Squared Error (MSE) for regression tasks and log-loss for classification tasks;
- $\Omega(f_k)$ is the regularization term that penalizes the complexity of each individual tree f_k in the model, helping to prevent overfitting. This term is defined as

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{4}$$

where T is the number of leaf nodes in the tree, w_j represents the weight of leaf j, γ controls the complexity of the model by imposing a penalty for adding more leaves, and λ is a regularization parameter that penalizes large weights.

XGBoost optimizes the objective function using second-order gradient descent, which takes into account both the first-order (gradient) and second-order (Hessian) information, making it highly effective at capturing complex patterns in the data. Furthermore, XGBoost incorporates additional enhancements, such as shrinkage (learning rate), column subsampling, and handling missing values, which contribute to its superior predictive performance and versatility across various machine learning tasks. These features make XGBoost a powerful tool for tackling both regression and classification problems in a wide range of applications [20].

3.1.3. Extreme Learning Machine (ELM)

ELM is a type of feedforward neural network designed to simplify the learning process by utilizing a unique training approach. Unlike traditional neural networks, which use iterative optimization methods like backpropagation to adjust weights and biases, ELM initializes the input weights and biases randomly and determines the output weights analytically. This characteristic significantly reduces the training time and complexity, making ELM highly efficient and well-suited for large-scale data applications [21].

The core idea of ELM is to approximate the desired target output using the activations of the hidden layer neurons. The process begins by establishing the hidden layer output matrix H, which represents the responses of all hidden neurons to the input data. Each entry in this matrix is calculated based on the randomly initialized input weights and biases. Once the hidden layer activations are obtained, they remain fixed throughout the training process, as ELM does not employ backpropagation to fine-tune these parameters [22].

Next, the relationship between the hidden layer output matrix H and the desired output, represented by the target matrix T, is established. The target matrix T contains the expected outputs for each training sample and serves as the ground truth for the network to learn. To find the best mapping from the hidden layer activations to the target outputs, ELM uses the output weight matrix β, which connects the hidden neurons to the output

layer [23]. The weight matrix β is computed analytically by solving a linear system of equations, which can be represented as

$$\beta = H^{\dagger} T \tag{5}$$

In Equation (5), $H^{\dagger}$ is the Moore–Penrose pseudo-inverse of the hidden layer output matrix H. The pseudo-inverse is a generalized form of the matrix inverse that is used to obtain a least-squares solution for β, even when H is not a square matrix or when it is singular (non-invertible). The use of the Moore–Penrose pseudo-inverse ensures that the calculated β minimizes the error between the predicted outputs of the network and the actual target values, making it an optimal solution in a least-squares sense.

3.2. Statistical Methods

Statistical methods like ARIMA and SARIMA provide robust baselines for time series forecasting, particularly when data exhibit linear patterns or seasonality. This subsection details these models' structure and their limitations in handling non-linear SMP dynamics.

3.2.1. Autoregressive Integrated Moving Average (ARIMA)

ARIMA models are used for time series forecasting by capturing linear dependencies in data. An ARIMA model is denoted as $ARIMA(p, d, q)$, where p is the number of autoregressive terms, d is the number of non-seasonal differences, and q represents the number of lagged forecast errors [24]. The ARIMA model is represented as

$$\Phi(B) Y_t = \theta(B) \epsilon_t \tag{6}$$

where $\Phi(B)$ is the autoregressive polynomial, $\theta(B)$ is the moving average polynomial, and ϵ_t is the error term.

3.2.2. Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA extends ARIMA by incorporating seasonal components. The SARIMA model is denoted as $ARIMA(p, d, q)(P, D, Q)_s$, where P, D, and Q are the seasonal parameters and s is the number of periods in a season [25]. The general formula is given as

$$\Phi_P(B^s)\Phi_p(B)\Big(1 - B\Big)^d (1 - B^s)^D Y_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t \tag{7}$$

In Equation (7), the seasonal autoregressive polynomial $\Phi_P(B^s)$ captures the influence of previous seasonal values, while the non-seasonal autoregressive polynomial $\phi_p(B)$ models the effect of past values on the current value Y_t. To make the series stationary, the non-seasonal differencing operator $\Big(1 - B\Big)^d$ and seasonal differencing operator $(1 - B^s)^D$ are applied to remove trends and seasonal variations, respectively. The seasonal moving average polynomial $\Theta_Q(B^s)$ describes the relationship between seasonal errors at seasonal lags, while the nonseasonal moving average polynomial $\theta_q(B)$ models how previous errors influence the current value. Finally, ϵ_t represents the white noise error term, which accounts for random variations that cannot be explained by the model [26].

By integrating these components, SARIMA is capable of modeling both short-term dependencies and long-term seasonal patterns, making it an effective tool for time series forecasting in scenarios where data exhibit seasonal behavior, such as monthly sales figures or annual energy consumption trends. This model allows for a comprehensive analysis of the time series, capturing the complex interplay between seasonal and non-seasonal factors that influence the observed values over time [27].

3.3. Time Series-Based Methods

Time series methods like LSTM, Bi-LSTM, ConvLSTM, and MLP are designed to capture complex temporal dependencies in sequential data. This subsection describes how

these models are structured to handle varying temporal patterns and relationships within SMP data.

3.3.1. Long Short-Term Memory (LSTM)

LSTM networks are designed to address the vanishing gradient problem in traditional RNNs by incorporating memory cells and gating mechanisms. The LSTM unit comprises several key components that work together to manage information flow over time [28].

The Input Gate is responsible for controlling how much of the current input x_t should be passed to the cell state. The mathematical formulation of the input gate is given by

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

In Equation (8), i_t represents the activation of the input gate. The sigmoid function, denoted as σ, squashes the values into the range $[0, 1]$, determining the extent to which each component of the input contributes to the cell state. W_i is the weight matrix for the input gate, h_{t-1} is the hidden state from the previous time step, x_t is the input at the current time step, and b_i is the bias term associated with the input gate.

The Forget Gate plays a crucial role in deciding which information from the previous cell state should be retained or discarded. The forget gate is computed as follows:

$$f_t = \sigma\Big(W_f \cdot [h_{t-1}, x_t] + b_f\Big) \tag{9}$$

Here, f_t is the activation of the forget gate, which, like the input gate, uses the sigmoid function σ. The weight matrix W_f and the bias term b_f control the influence of the previous hidden state h_{t-1} and the current input x_t on the forget gate's decision.

The Cell State Update mechanism is used to update the cell state C_t based on the candidate cell state $\widetilde{C}_t$ and the output of the forget gate. The updated cell state is calculated as

$$C_t = f_t * C_{t-1} + i_t \times \widetilde{C}_t \tag{10}$$

In Equation (9), C_t is the cell state at the current time step, C_{t-1} is the cell state from the previous time step, and $\widetilde{C}_t$ is the candidate cell state generated by applying the hyperbolic tangent function tanh to the input and previous hidden state, as follows:

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{11}$$

W_c is the weight matrix used to compute $\widetilde{C}_t$, and b_c is the associated bias term.

Finally, the output gate determines the final output of the LSTM unit. The output gate is defined as

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{12}$$

where o_t is the activation of the output gate. The output h_t of the LSTM unit is then obtained by applying the output gate's activation to the updated cell state C_t using the hyperbolic tangent function, as follows:

$$h_t = o_t * \tanh(C_t) \tag{13}$$

In Equation (13), h_t is the hidden state (or output) of the LSTM block, which is subsequently passed on to the next time step or the next network layer.

3.3.2. Bi-Directional Long Short-Term Memory

Bi-LSTM, unlike traditional LSTM models, considers bidirectional dependencies within the temporal structure of input data, allowing it to capture intricate relationships and patterns present in sequential data. By processing both past and future information simultaneously, Bi-LSTM gains a holistic understanding of the context within time series data, enhancing its predictive accuracy and feature extraction capabilities. Consequently,

Bi-LSTM emerges as a powerful tool across various domains such as natural language processing, time series forecasting, and sequential data analysis [29].

The bidirectional nature of Bi-LSTM enables it to analyze sequences in both forward and backward directions, facilitating a comprehensive comprehension of temporal dynamics and intricate dependencies within the data. This dual perspective empowers Bi-LSTM to uncover nuanced patterns and correlations that might be missed by unidirectional models, thereby improving its ability to model complex sequential data and make accurate predictions. As a result, Bi-LSTM stands out as a versatile and effective solution for tasks requiring nuanced understanding and prediction of sequential patterns [30].

3.3.3. Convolutional Long Short-Term Memory

ConvLSTM processes data by utilizing convolutional operations to update the states of the network. Let the input at time step t be denoted as $\mathbf{X}_t \in \mathbb{R}^{h \times w \times c}$, where h and w are the height and width of the input, and c is the number of channels. The hidden state and cell state at time step t are represented as $\mathbf{H}_t$ and $\mathbf{C}_t$, respectively, each with dimensions $h \times w \times d$, where d is the number of filters (or hidden units). The ConvLSTM equations for the input gate, forget gate, cell state, and output gate are defined as follows [31]:

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_{xi} * \mathbf{X}_t + \mathbf{W}_{hi} * \mathbf{H}_{t-1} + \mathbf{W}_{ci} \circ \mathbf{C}_{t-1} + \mathbf{b}_i) \\ \mathbf{f}_t &= \sigma\Big(\mathbf{W}_{xf} * \mathbf{X}_t + \mathbf{W}_{hf} * \mathbf{H}_{t-1} + \mathbf{W}_{cf} \circ \mathbf{C}_{t-1} + \mathbf{b}_f\Big) \\ \mathbf{C}_t &= \mathbf{f}_t \circ \mathbf{C}_{t-1} + \mathbf{i}_t \circ \tanh(\mathbf{W}_{xc} * \mathbf{X}_t + \mathbf{W}_{hc} * \mathbf{H}_{t-1} + \mathbf{b}_c) \\ \mathbf{o}_t &= \sigma(\mathbf{W}_{xo} * \mathbf{X}_t + \mathbf{W}_{ho} * \mathbf{H}_{t-1} + \mathbf{W}_{co} \circ \mathbf{C}_t + \mathbf{b}_o) \\ \mathbf{H}_t &= \mathbf{o}_t \circ \tanh(\mathbf{C}_t) \end{aligned} \tag{14}$$

In Equation (14), $\mathbf{W}_{xi}, \mathbf{W}_{xf}, \mathbf{W}_{xc}$, and $\mathbf{W}_{xo}$ are the convolutional kernels for the input-to-state connections, while $\mathbf{W}_{hi}, \mathbf{W}_{hf}, \mathbf{W}_{hc}$, and $\mathbf{W}_{ho}$ are the convolutional kernels for the state-to-state connections. The weight matrices $\mathbf{W}_{ci}, \mathbf{W}_{cf}$, and $\mathbf{W}_{co}$ are associated with the peephole connections, and the terms $\mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_c$, and $\mathbf{b}_o$ are the bias terms. The convolution operation is represented by $*$, while $\circ$ denotes the Hadamard (element-wise) product. Equation (14) shows how ConvLSTM incorporates spatial information through convolutional operations rather than matrix multiplications. The input gate $(\mathbf{i}_t)$ controls how much of the current input information is passed to the cell state, while the forget gate $(\mathbf{f}_t)$ determines which information from the previous cell state $(\mathbf{C}_{t-1})$ should be retained or forgotten. The cell state update equation combines the effects of the input and forget gates, allowing the network to maintain long-term dependencies. The output gate $(\mathbf{o}_t)$ determines how much of the updated cell state contributes to the hidden state $(\mathbf{H}_t)$, which is passed on to the next time step.

3.3.4. Multi-Layer Perceptron

An MLP maps the input time series data $\mathbf{X}$ to an output $\mathbf{Y}$ through a series of transformations and non-linear operations. For a single neuron j in a hidden layer, the weighted sum of inputs is expressed as $z_j = \sum_{i=1}^{n} w_{ji} \cdot x_i + b_j$, where z_j is the weighted sum of inputs for neuron j, w_{ji} is the weight between input i and neuron j, x_i is the input value from the previous layer, b_j is the bias term for neuron j, and n is the number of inputs. The weighted sum z_j is then passed through an activation function $f(\cdot)$ to introduce non-linearity, producing the neuron's output $a_j = f(z_j)$ [32].

Common activation functions used in MLPs include the sigmoid, hyperbolic tangent (tanh), and Rectified Linear Unit (ReLU). The sigmoid function is defined as $f(z) = \frac{1}{1+e^{-\varepsilon}}$, the hyperbolic tangent as $f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, and the ReLU function as $f(z) = \max(0, z)$. These functions enable the MLP to learn and represent complex patterns in the input data by adding non-linearity to the network [33].

The final prediction of the MLP, denoted as $\widehat{y}$, is computed in the output layer by taking the activations from the last hidden layer and applying a linear transformation: $\widehat{y} = \sum_{j=1}^{m} w_{oj}.\ a_j^{(L)} + b_o$, where w_{oj} is the weight connecting neuron j in the last hidden

layer to the output layer, b_o is the bias term for the output neuron, $a_j^{(L)}$ is the activation of the j-th neuron in the last hidden layer, and m is the number of neurons in the last hidden layer [34].

In time series forecasting, the input data **X** typically consist of lagged observations, such as $\{x_{t-1}, x_{t-2}, \ldots, x_{t-p}\}$, where p represents the number of lagged time steps. This structure enables the MLP to capture temporal dependencies and predict future values based on historical data. The network's weights and biases are optimized using algorithms like gradient descent, which minimize a loss function such as the Mean Squared Error (MSE), defined as MSE = $\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2$, where y_i is the actual target value at time step i, $\widehat{y}_i$ is the predicted value by the MLP, and N is the total number of observations [35].

3.4. Feature Selectors

Feature selection techniques play a critical role in improving model interpretability and reducing computational complexity. This subsection introduces mRMR and MLFS as key methods to refine the input feature set, enhancing predictive accuracy.

3.4.1. Minimum Redundancy Maximum Relevance Feature Selector

The methodology behind mRMR is based on two main rationales. The first rationale is that highly correlated features should not be used, i.e., the redundancy between features should be taken into account, thus keeping features that are maximally dissimilar to each other [36]. A possible solution to determine redundancy among the variables in S is

$$W_I(S) = \frac{1}{|S|^2}\sum_{X_i, X_j \epsilon S} MI(X_i, X_j) \tag{15}$$

where $MI(X_i, X_j)$ gives the measure of mutual information between the variables X_i and X_j.

The second rationale of mRMR is that minimum redundancy should be complemented by a maximum relevance criterion of the features in regard to the class variable. A measure of global relevance of the variables in S in terms of C is

$$V_I(S) = \frac{1}{|S|}\sum_{X_i \epsilon S} MI(C, X_i) \tag{16}$$

A straightforward but efficient formula of combining redundancy and relevance to obtain a good subset of features is

$$S^* = argmax_{S \subseteq U}(V_I(S) - W_I(S)) \tag{17}$$

Because it is not practical to find the best subset using an exhaustive search, the selected subset is obtained in an incremental way, starting with the feature having maximum value of $MI(C; X_i)(S_0 = \{X_{i_0}\})$ and gradually adding, to the current subset S_{m-1}, the feature that maximizes

$$\underset{X_j \epsilon U \setminus S_{m-1}}{max}\left(MI(C, X_j) - \frac{1}{m-1}\sum_{X_i \epsilon S_{m-1}} MI(X_j, X_i)\right) \tag{18}$$

3.4.2. Maximum-Likelihood Feature Selector

MLFS ranks input features based on their mutual information (MI) with the target variable, using a maximum likelihood estimator. MI quantifies the dependency between variables and is defined as

$$I(X, Y) = \iint p(x, y)\log\frac{p(x, y)}{p(x)p(y)}dxdy \tag{19}$$

where $p(x, y)$ is the joint probability density of the input variable x and the target variable y, and $p(x)$ and $p(y)$ are their respective marginal densities. MLFS estimates MI through the density ratio $w(x, y)$, which represents the relationship between the joint and marginal densities [37], as follows:

$$w(x, y) = \frac{p(x, y)}{p(x)p(y)} \tag{20}$$

Given n paired samples (x_i, y_i), the MI can be approximated as

$$I(X, Y) \approx \frac{1}{n}\sum_{i=1}^{n} \log w(x_i, y_i) \tag{21}$$

where $w(x, y)$ is estimated by maximizing the log-likelihood function $L(\alpha)$. In Equation (22), α is a parameter vector used to scale the feature values, and $\phi(x, y)$ is a basis function representing the feature transformations:

$$L(\alpha) = \sum_{i=1}^{n} \log\left(\alpha^{\top}\phi(x_i, y_i)\right) \tag{22}$$

Normalization constraints ensure that the estimated density ratio $w(x, y)$ sums to 1 and remains non-negative, providing an accurate approximation of the relationship between the variables. This optimization helps identify the most relevant features by focusing on those that maximize MI, improving model performance while reducing redundancy.

4. Dataset and Methodology

This section outlines the dataset used for model development, preprocessing techniques applied, and the methodology for model training and evaluation. It also discusses the feature selection, cross-validation strategies, and hyperparameter optimization methods implemented to enhance model performance.

4.1. Dataset

The dataset employed in this study spans from 1 January 2021, to 14 September 2023, and is provided by the Energy Exchange Istanbul (EXIST). It includes hourly SMP values, which serve as the primary target variable for prediction. A wide array of predictor variables were incorporated into the models to capture the various factors influencing SMP fluctuations. These predictors include

- Date Information: Year, month, day, and hour, capturing the time series dynamics;
- Calendar Effects: Categorical variables distinguishing weekdays, weekends, and special days, such as holidays, to capture demand shifts during these periods;
- Macroeconomic Indicators: Financial indicators like USD/TRY and EUR/TRY exchange rates, and the BIST 100 index.

The dataset was preprocessed to handle missing values, detect and correct outliers, and standardize the scale of predictor variables. The statistics of the dataset show that the SMP ranged from a minimum of 0 TL/MWh to a maximum of 4800 TL/MWh, with an average value of 1713.34 TL/MWh. The standard deviation was 1301.16 TL/MWh.

4.2. Model Development and Feature Selection

In this study, a comprehensive forecasting framework was developed, combining machine learning, deep learning, and statistical models with advanced feature selection techniques to enhance the predictive accuracy and efficiency of SMP forecasts in Turkey's electricity market. Given the complex and volatile nature of SMP, this framework was meticulously designed to capture both linear and non-linear relationships in the data, addressing the various interdependencies that significantly influence price fluctuations.

To effectively model the SMP data, a diverse set of machine learning and deep learning models was employed. This included ELM, SVM, XGBoost, LSTM, and Bi-LSTM,

each selected for its strengths in managing non-linear dynamics and high-dimensional data, which are prominent in SMP forecasting. ELM and SVM were leveraged for their high performance in pattern recognition and handling complex data structures, while XGBoost provided an effective solution for managing non-linear interactions and preventing overfitting through boosting and regularization. LSTM and Bi-LSTM were incorporated specifically to capture temporal dependencies within the SMP data, with Bi-LSTM offering a bidirectional view that accounts for both past and future price dynamics, crucial for modeling the temporal complexities present in electricity pricing.

In addition to machine learning and deep learning methods, traditional statistical models such as ARIMA and SARIMA were utilized as foundational baselines. These models were chosen for their robustness in capturing clear seasonal patterns and trends within the data. Their inclusion provided a stable framework for forecasts during periods when SMP exhibits relatively consistent predictable behavior, thus balancing the more complex machine learning models in scenarios where market volatility is lower and historical patterns remain stable.

Feature selection was a vital aspect of model development, significantly improving the models' predictive accuracy by narrowing down the dataset to the most impactful variables while reducing noise and multicollinearity. Two advanced feature selection techniques, mRMR and MLFS, were employed to refine the feature set systematically. The mRMR method was used to prioritize features that held high relevance to SMP while minimizing redundancy among predictors, ensuring that the selected variables were not excessively correlated, which could introduce unnecessary noise and impact predictive quality. For example, mRMR isolated specific calendar effects, such as weekends and public holidays, over more generic time indicators, allowing the model to capture unique demand patterns associated with these periods without redundant overlap from other time variables.

Following the mRMR process, MLFS was applied iteratively to select features that maximized the likelihood of observed SMP outcomes. Unlike mRMR, MLFS was particularly effective in identifying non-linear relationships within the data by choosing features that made the most substantial contributions to predictive performance. This iterative approach allowed the models to adapt to complex interactions between SMP and its predictors, capturing influential macroeconomic factors, such as fluctuations in the USD/TRY exchange rate, that may not exhibit a simple linear relationship with SMP. By emphasizing features that most directly impacted SMP, MLFS further refined the dataset, enhancing both model accuracy and adaptability to Turkey's electricity market dynamics.

This integrative approach, combining machine learning, deep learning, and statistical models with a robust feature selection process, resulted in a highly adaptable and accurate SMP forecasting framework. By focusing on the most relevant and distinct predictors, the framework not only maximized predictive performance across varying market conditions but also established a benchmark for SMP forecasting accuracy. This methodology supports informed decision-making for Turkey's electricity market and offers potential adaptability to similar markets where dynamic price forecasting is essential for strategic planning and policy development.

4.3. Model Training and Evaluation

To ensure robustness, generalizability, and the effective mitigation of overfitting risks in integrating diverse models for SMP forecasting, each model underwent training using a 10-fold cross-validation process. This method divides the dataset into 10 equal parts, or "folds," with each iteration training the model on 9 folds while testing on the remaining one. This iterative process repeats until each fold has been used for validation, allowing for a comprehensive assessment of the model's performance across multiple data segments. By systematically exposing models to varied data subsets, this technique mitigates overfitting and enhances the model's ability to generalize to unseen data, which is particularly essential in complex forecasting scenarios where overfitting risks are pronounced due to the integration of models with both linear and non-linear dynamics.

Beyond cross-validation, the study employed additional regularization and optimization methods tailored to each model to manage complexity and prevent overfitting. For instance, models such as XGBoost included L_1 and L_2 regularization penalties to control model complexity by limiting the growth and impact of decision trees, reducing overfitting by constraining excessive feature reliance and interaction effects. In deep learning models like LSTM and Bi-LSTM, dropout regularization was applied, randomly dropping units during training to prevent the model from becoming overly reliant on specific patterns in the data and ensuring that it learns more generalized representations. Dropout is particularly effective in deep learning contexts, as it reduces co-adaptation of neurons, enhancing the model's resilience against overfitting.

To fine-tune each model's performance, a combination of grid search and Bayesian optimization was used for hyperparameter tuning. This two-pronged approach provided both systematic and efficient exploration of the hyperparameter space, enhancing the model's performance while balancing complexity. Key parameters for each model were identified and optimized, as detailed below:

- SVM: The penalty parameter C and the kernel coefficient γ were optimized to balance the trade-off between margin maximization and error tolerance, enhancing model simplicity and generalization;
- XGBoost: Parameters including the learning rate, maximum tree depth, and subsample ratio were fine-tuned. The learning rate controlled how quickly the model adapted to the training data, while maximum depth constrained tree complexity, and the subsample ratio limited the fraction of data used in each boosting iteration to reduce variance and prevent overfitting;
- LSTM and Bi-LSTM: The number of hidden units, learning rate, and dropout rates were carefully adjusted. The hidden units allowed the model to capture temporal dependencies in the data, while the learning rate and dropout settings provided control over training intensity and regularization to avoid over-reliance on specific training patterns;
- ELM: The number of hidden neurons and activation function were optimized to minimize prediction error while controlling model simplicity, ensuring that the model could handle high-dimensional data effectively without excessive complexity.

The grid search method systematically tested a predefined set of hyperparameters, providing a thorough but computationally intensive exploration. In contrast, Bayesian optimization used probabilistic models to evaluate a broader hyperparameter space more efficiently, predicting parameter combinations likely to enhance performance. By combining grid search with Bayesian optimization, the study achieved a balance between exhaustive exploration and computational efficiency, tailoring each model to its optimal configuration. The performance of each model was evaluated using MAPE, which is defined as follows:

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100 \tag{23}$$

where y_i and $\widehat{y}_i$ represent the actual and predicted values of SMP for the i-th observation, respectively, and n is the total number of observations. MAPE measures the average percentage error between actual and predicted values. This metric was chosen for its ability to effectively evaluate the accuracy of forecasts in the context of time series data with significant variability.

All models were implemented in Python using libraries such as scikit-learn, TensorFlow, and XGBoost. The effectiveness of each model was assessed on a series of selected dates, representing different market conditions, including regular weekdays, weekends, and holidays. By analyzing model performance under these varying scenarios, the study aims to assess the adaptability and robustness of each forecasting approach.

In more detail, 14 February 2023 represents a typical weekday with standard market activity and consistent electricity consumption patterns. This date was selected to establish

a baseline for model performance under normal market conditions, serving as a reference point against which results from more irregular scenarios can be compared. In contrast, 25 March 2023 was chosen as a weekend day, characterized by decreased electricity demand due to reduced industrial and commercial activities. The inclusion of this date allows the study to analyze how models perform when faced with lower demand and reduced market volatility, common to weekends. The 19 May 2023 was selected as a public holiday, a day that often exhibits irregular consumption patterns and unique market behaviors. Public holidays typically see sharp drops in electricity consumption and altered generation schedules, making it essential to test the models' capability to handle these deviations from typical trends.

Another date chosen to capture abnormal market behavior was 29 July 2023, which experienced high volatility and significant fluctuations in electricity prices. This date was used to assess how well the models adapt to rapid changes and non-linear interactions between variables, which are more challenging to predict. In contrast, 9 August 2023, was marked by relatively stable market conditions and minimal fluctuations, representing a low volatility day. The purpose of including this date was to evaluate the models' performance in scenarios where fewer abrupt changes occur, allowing them to leverage historical patterns effectively. The study further explored model consistency by selecting 30 August 2023, another public holiday. This date was chosen to compare with 19 May, assessing how the models handle different public holidays, each with its unique consumption patterns and potential market anomalies. Lastly, 14 September 2023 was included as a transitional period marking the end of the summer season. This date was used to test the models' ability to adapt to shifting seasonal dynamics and changing consumption trends, which often lead to gradual adjustments in electricity demand and supply patterns.

5. Results and Discussion

The results and discussion section presents the performance of different forecasting models under various market conditions. It includes a detailed comparison of model accuracy, the impact of feature selection, and the effectiveness of each method in handling SMP variations.

5.1. Results

Table 2 presents the MAPE values for all forecasting models evaluated on selected dates, providing a detailed comparison of their performance across various scenarios.

Table 2. MAPE values of all models.

Methods	MAPE (%) Values on Selected Dates						
	14.02.2023	25.03.2023	19.05.2023	29.07.2023	09.08.2023	30.08.2023	14.09.2023
MLP	18.47	15.21	11.51	12.00	10.37	18.04	15.02
LSTM	13.57	13.43	10.85	8.01	9.44	16.50	18.47
Bi-LSTM	13.04	14.40	21.96	13.65	11.62	19.37	13.26
ConvLSTM	13.43	16.55	10.92	7.12	8.25	16.89	15.28
ARIMA	18.75	19.28	11.27	8.59	9.30	19.91	19.72
SARIMA	18.03	19.36	10.55	8.96	9.62	20.72	19.11
SVM	15.52	16.72	11.14	8.72	7.33	15.88	14.44
SVM—mRMR	14.35	14.85	10.64	9.25	6.77	16.25	13.75
SVM—MLFS	12.01	15.62	9.99	8.95	7.05	15.92	14.62
XGBoost	7.72	11.76	12.09	5.22	5.08	14.65	10.95
XGBoost—mRMR	10.31	11.45	7.69	4.42	6.12	13.95	11.62
XGBoost—MLFS	8.91	13.21	5.87	4.71	5.11	12.40	10.52
ELM	16.45	13.47	11.48	8.12	6.84	15.07	12.10
ELM—mRMR	13.57	11.81	9.08	7.67	5.23	14.11	11.17
ELM—MLFS	12.09	12.47	9.66	6.97	6.29	14.20	11.31

Figure 1 illustrates the performance of various statistical and machine learning methods in predicting the hourly SMP for the Turkish electricity market, showcasing the average MAPE values derived from predictions made over seven distinct days.

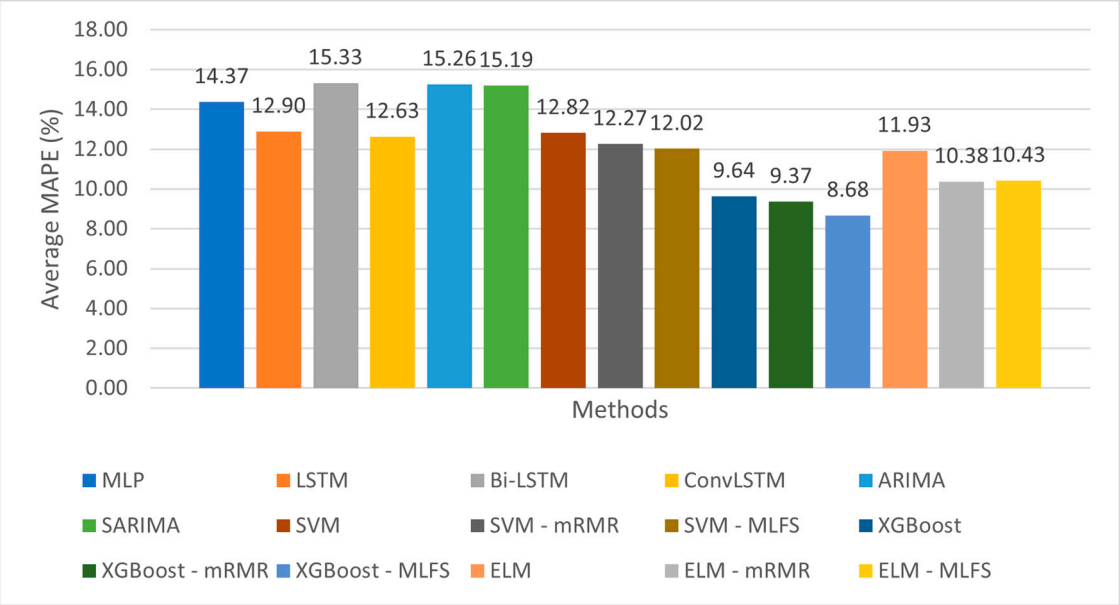


Figure 1. Method-based average MAPE values of the models.

5.2. Method-Based Analysis

The method-based analysis in this section focuses on evaluating the performance of different forecasting methods used in the study, specifically in terms of their MAPE values across various dates. The goal is to provide a comparative evaluation of each method, highlighting their strengths and limitations under different market scenarios. By doing so, this section aims to identify which models are most effective in capturing the complex dynamics of the Turkish electricity market.

Traditional statistical models, like ARIMA and SARIMA, consistently showed higher MAPE values during the chosen dates. This shows how poorly they can handle the naturally unstable and nonlinear SMP data. For example, the ARIMA model recorded an average MAPE of 15.26%, while SARIMA had a slightly lower average of 15.19%. These high error rates suggest that traditional models are not well-suited for SMP forecasting in highly dynamic markets like Turkey's, where rapid price fluctuations and complex interdependencies between variables are prevalent. The models' reliance on historical trends and linear assumptions restricts their ability to adapt to sudden shifts in market conditions, leading to suboptimal performance.

Machine learning models, such as SVM and ELM, outperformed traditional statistical models in most scenarios. The baseline SVM model yielded an average MAPE of 12.82%, indicating its moderate ability to capture non-linear relationships. Incorporating feature selection techniques, such as mRMR and MLFS, further improved SVM's performance. SVM combined with mRMR achieved an average MAPE of 12.27%, while SVM with MLFS further reduced the error to 12.02%. This improvement is because feature selection methods can clean up the set of features by focusing on the most useful predictors and getting rid of the rest. This lets SVM focus on the important drivers of SMP.

ELM models, known for their rapid training time and strong generalization ability, also demonstrated robust performance in this study. The baseline ELM model achieved an average MAPE of 11.93%, but when enhanced with feature selection methods, the results improved. ELM combined with mRMR resulted in a MAPE of 10.38%, while ELM with MLFS achieved a slightly higher MAPE of 10.43%. This steady rise in performance shows that ELM, when combined with feature selection, can effectively capture the complex and

nonlinear interactions found in SMP data. This makes it a strong candidate for tasks like forecasting in the electricity market.

Among all the methods evaluated, tree-based ensemble models, particularly XGBoost, showed the most promising results. The baseline XGBoost model consistently achieved the lowest MAPE values, averaging 9.64%. The integration of feature selection techniques further enhanced XGBoost's performance, with XGBoost combined with mRMR yielding an average MAPE of 9.37% and XGBoost with MLFS achieving the lowest average MAPE of 8.68%. This highlights XGBoost's ability to handle non-linear interactions and adapt to various market conditions, making it the top-performing method for SMP forecasting.

Deep learning models, such as LSTM, Bi-LSTM, and ConvLSTM, displayed varying levels of effectiveness depending on the market scenario. LSTM showed relatively strong performance with an average MAPE of 12.90%, while ConvLSTM performed similarly with a MAPE of 12.63%. Bi-LSTM, however, had the highest average error rate among the deep learning models, with a MAPE of 15.33%, indicating potential challenges in generalizing across datasets with irregular patterns.

5.3. Date-Based Analysis

The date-based analysis provides an in-depth evaluation of the performance of different models on selected dates to identify patterns and assess the robustness of each approach under varying market conditions. The dates chosen for this analysis—14 February, 25 March, 19 May, 29 July, 9 August, 30 August, and 14 September—represent a range of scenarios, including regular working days, weekends, and public holidays, where sudden shifts and high volatility are expected. The performance of each model is compared based on their MAPE values, highlighting their strengths and weaknesses.

On 14 February, Valentine's Day (Special Day), the traditional models ARIMA (18.75%) and SARIMA (18.03%) struggled. Machine learning models performed better. SVM recorded 15.52% without feature selection, improving to 14.35% with mRMR and 12.01% with MLFS. ELM also improved from 16.45% to 13.57% (mRMR) and 12.09% (MLFS). XGBoost, the top performer, achieved 7.72%, though MLFS slightly increased the MAPE to 8.91%. Among time-series-based methods, MLP showed higher error at 18.47%, while LSTM, Bi-LSTM, and ConvLSTM showed moderate performance with 13.57%, 13.04%, and 13.43%, respectively.

On 29 July, a regular Saturday, traditional models continued to struggle, with ARIMA and SARIMA recording 8.59% and 8.96%, respectively. SVM improved from 8.72% to 9.25% with mRMR and 8.95% with MLFS. ELM showed 8.12% without selection, improving to 7.67% with mRMR and 6.97% with MLFS. XGBoost performed exceptionally well, with 5.22% without feature selection, and 4.42% with mRMR, achieving the best performance at 4.71% with MLFS. Among the time-series-based methods, MLP showed 12.00%, while LSTM performed at 8.01%, Bi-LSTM at 13.65%, and ConvLSTM at 7.12%.

On 9 August, a regular weekday, ARIMA and SARIMA recorded 9.30% and 9.62%. SVM performed steadily, achieving 7.33% without feature selection and improving to 6.77% with mRMR and 7.05% with MLFS. ELM showed 6.84% without selection, improving to 7.67% with mRMR and 6.29% with MLFS. XGBoost was consistent with 5.08% without feature selection and 5.11% with MLFS. Among the time-series-based methods, MLP performed at 10.37%, LSTM at 9.44%, Bi-LSTM at 11.62%, and ConvLSTM at 8.25%.

On 30 August, Victory Day (Public Holiday), traditional models performed poorly, with ARIMA at 19.91% and SARIMA at 20.72%. SVM improved with feature selection, recording 15.88% without selection, 16.25% with mRMR, and 15.92% with MLFS. ELM improved from 15.07% to 14.11% with mRMR and 14.20% with MLFS. XGBoost performed at 14.65% without feature selection, improving to 12.40% with MLFS. Among the time-series-based methods, MLP recorded 18.04%, LSTM at 16.50%, Bi-LSTM at 19.37%, and ConvLSTM at 16.89%.

On 14 September, a regular weekday, the traditional models struggled, with ARIMA recording 19.72% and SARIMA at 19.11%. SVM performed at 14.44% without selection,

improving to 13.75% with mRMR and 14.62% with MLFS. ELM improved from 12.10% to 11.17% with mRMR and 11.31% with MLFS. XGBoost maintained superior performance, recording 10.95% without selection and 10.52% with MLFS. Among the time-series-based methods, MLP showed 15.02% and LSTM performed at 18.47%, Bi-LSTM at 13.26%, and ConvLSTM at 15.28%.

5.4. Evaluation of Feature Selection Impact on Model Performance

The influence of feature selection on model accuracy is highlighted in Figure 2, which provides a comparative analysis of MAPE values for SVM, XGBoost, and ELM-based models, both with and without feature selection algorithms. MLFS emerged as the most effective feature selection method, significantly reducing error rates for the SVM and XGBoost models. The improved performance is attributed to MLFS's ability to prioritize highly informative features while discarding redundant or less relevant variables, thereby enhancing the overall predictive capacity of the models. In contrast, the mRMR algorithm did not consistently yield positive results and, in some cases, increased the error rates for XGBoost models. This variability indicates that while feature selection is a critical step in model development, its effectiveness is contingent on the specific algorithm and data characteristics.

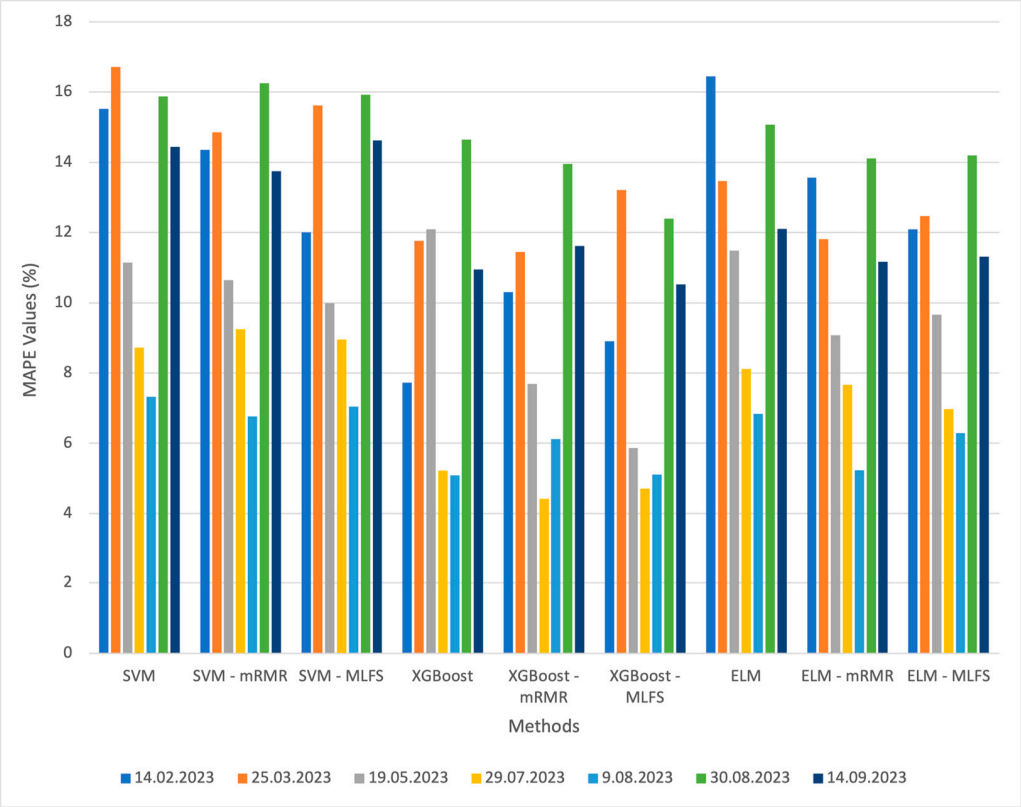


Figure 2. Comparison of feature selection algorithms.

ELM-based models also benefited from feature selection, with MLFS reducing MAPE values more effectively than mRMR. The consistent performance of MLFS across different models suggests its suitability for SMP forecasting tasks, where identifying key predictors is crucial for achieving high prediction accuracy. This underscores the importance of selecting an appropriate feature selection technique, as it can significantly influence model outcomes.

5.5. Comparison of Time Series-Based Models

Figure 3 presents a detailed comparison of MAPE values across various time series-based models, including MLP, LSTM, Bi-LSTM, ConvLSTM, ARIMA, and SARIMA. The

results indicate that LSTM and ConvLSTM models consistently outperformed other time series models in capturing complex temporal dependencies and accurately forecasting SMP trends. Their ability to integrate both long-term and short-term patterns within the data enabled them to maintain lower error rates, even on dates characterized by significant volatility. ConvLSTM, which combines convolutional and recurrent layers, further demonstrated strong performance on 14 February and 9 August, effectively handling both spatial and temporal relationships in the data.

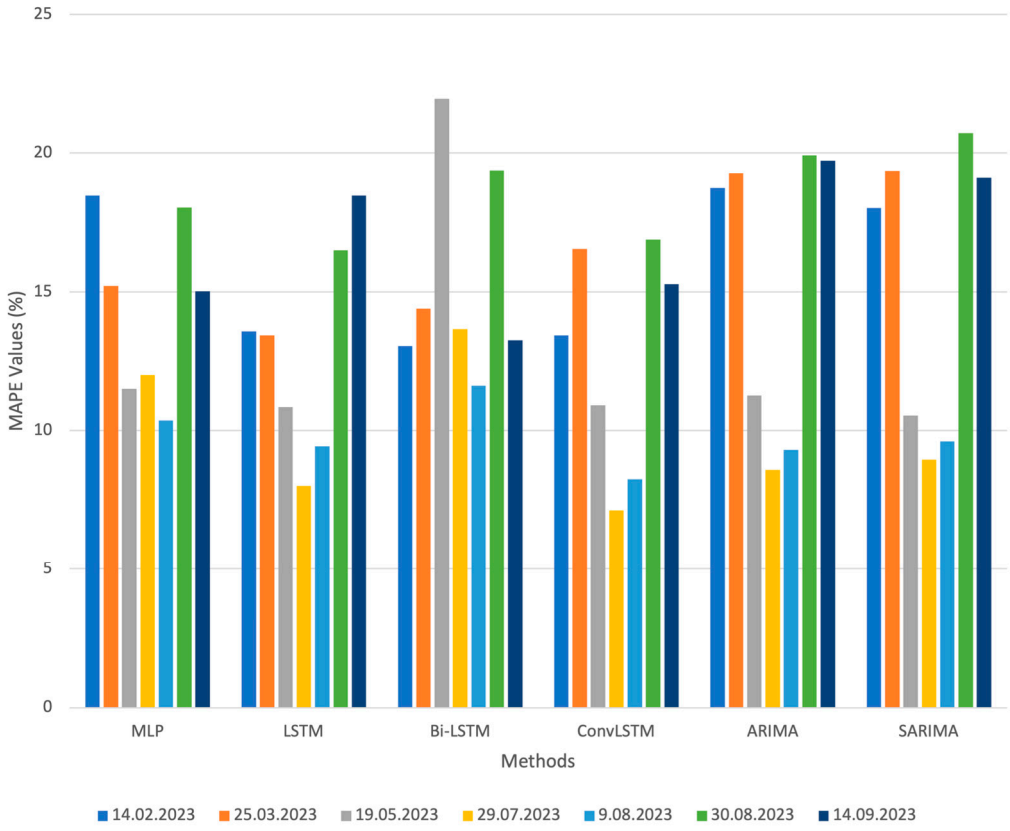


Figure 3. Comparison of time series-based methods.

In contrast, the ARIMA and SARIMA models, while effective for traditional time series forecasting, showed limitations in handling non-linear and abrupt changes in SMP values. Their moderate to high MAPE values on 30 August and 14 September suggest that these models may not be suitable for complex non-stationary data patterns typically observed in electricity markets.

5.6. Comparing the Results with the Literature

The results obtained in this study were also compared with the findings of existing research to evaluate the performance of the proposed models and methodologies. Although this comparison provides valuable insights into the effectiveness of the advanced machine learning and statistical methods employed, it is important to note that the datasets used by different studies vary significantly in terms of time periods, market conditions, and geographic regions. These differences limit the potential for direct comparison, and any conclusions drawn should be interpreted with caution.

The findings of this study indicate that the machine learning models, particularly XGBoost and ELM, achieved higher predictive accuracy compared to traditional statistical models like ARIMA and SARIMA. This observation is consistent with existing research, which often highlights the limitations of traditional models in capturing non-linear relationships and dealing with high levels of volatility inherent in electricity market data. For

instance, Özgüner et al. (2017) implemented ANN to predict day-ahead SMP in the Turkish electricity market and reported MAPE values ranging from 9.77% to 21.81%. In contrast, this study's XGBoost model combined with the MLFS consistently achieved lower MAPE values, such as 4.42% on 29 July 2023 and 5.87% on 19 May 2023. This significant reduction in error rates underscores the advantages of using ensemble learning methods like XGBoost, particularly when integrated with advanced feature selection techniques. However, it is essential to consider that Özgüner et al.'s dataset may have included different temporal and market conditions, which could influence model performance and the comparability of results.

Similarly, Pan et al. (2017) [11] developed an ensemble learning model using AdaBoost, SVM, and Back Propagation Neural Networks (BPNNs), achieving MAPE values between 3.84% and 8.43%. The results of the current study indicate that XGBoost, when paired with MLFS, achieved comparable and even superior MAPE values, such as 4.42% on 29 July 2023, and 5.08% on 9 August 2023. This suggests that the inclusion of sophisticated feature selection techniques significantly reduces model complexity and enhances predictive accuracy. Nevertheless, the specific dataset characteristics and market dynamics considered by Pan et al. may differ from those of the current study, affecting the generalizability of these findings across diverse market contexts.

Jufri et al. (2019) [12] employed an ANN-based approach for day-ahead SMP forecasting and reported MAPE values as low as 5.36%. The proposed XGBoost and ELM models in this study achieved MAPE values below 5% on several occasions, such as 4.71% on 29 July 2023 and 5.11% on 9 August 2023, for XGBoost-MLFS. The improvement can be attributed to the integration of advanced feature selection methods like mRMR and MLFS, which refined the input variables and improved the model's ability to capture complex patterns in SMP data. However, Jufri et al.'s dataset likely covered a different set of variables and timeframes, making it difficult to establish a direct comparison.

Panapakidis and Moschakis (2019) [13] used FFNN for SMP forecasting in the Greek energy market, reporting MARNE values between 7.42% and 9.22%. The XGBoost and ELM models presented in this study achieved MAPE values significantly lower than this range (e.g., 4.71% on 29 July 2023, for XGBoost-MLFS). This reduction in error rates suggests that tree-based ensemble methods and ELMs are more adept at handling the complexities of SMP forecasting in highly dynamic electricity markets. Again, differences in the market structure and temporal scope of the datasets used in these studies should be considered when evaluating relative model performance.

Shim et al. (2023) [15] utilized XGBoost, RF, and LightGBM for SMP forecasting and achieved MAPE values as low as 2.70%. While the best MAPE values for XGBoost-MLFS in this study were slightly higher, ranging between 4.42% and 5.87%, the inclusion of additional variables such as macroeconomic indicators and feature selection techniques enhanced the model's robustness under varying market conditions. This study demonstrated greater adaptability across regular, volatile, and holiday scenarios, making it more versatile for practical applications in real-world settings. However, since Shim et al.'s dataset was from a different region with its own unique market dynamics, the findings may not be directly comparable.

One of the distinguishing features of this study is the use of feature selection techniques like mRMR and MLFS, which contributed to substantial improvements in model performance. While several previous studies did not employ such techniques or used simpler selection methods, the application of MLFS in this study significantly reduced the dimensionality of the input space and allowed the models to focus on the most relevant variables, thereby enhancing interpretability and predictive accuracy. This impact is particularly evident when comparing the performance of XGBoost-MLFS to models like SVM and ANN in studies by Pan et al. (2017) [11] and Shim et al. (2023) [15].

Overall, the results of this study reaffirm findings from the literature that traditional statistical models such as ARIMA and SARIMA exhibit limited performance in environments characterized by non-linear dependencies and sudden shifts in market dynamics.

In this study, ARIMA and SARIMA consistently showed higher MAPE values compared to machine learning models, particularly on dates with irregular market behavior. This outcome is consistent with the observations made in the studies by Özgüner et al. (2017) [10] and Panapakidis and Moschakis (2019) [13]. Conversely, machine learning models like XGBoost and ELM demonstrated resilience and adaptability, achieving superior predictive accuracy even under challenging market conditions.

6. Conclusions and Future Work

This section concludes the key findings of the study and outlines potential avenues for future research.

6.1. Conclusions

This study focused on developing and evaluating advanced SMP forecasting models for the Turkish electricity market by leveraging traditional statistical and machine learning methods. The research aimed to overcome the limitations of traditional forecasting approaches, which often struggle with capturing non-linear dependencies and sudden shifts in market dynamics due to their linear assumptions and reliance on historical trends. By integrating sophisticated feature selection techniques, this study demonstrated that machine learning models, particularly XGBoost and ELM, consistently outperformed traditional methods like ARIMA and SARIMA under various market conditions.

The results show that XGBoost, when combined with advanced feature selection methods such as MLFS and mRMR, achieved significantly lower MAPE values compared to other models, demonstrating superior robustness and predictive accuracy. The application of these feature selection techniques played a crucial role in refining the input feature set, eliminating redundant information, and focusing on the most relevant variables, thereby enhancing model performance while reducing computational complexity. These findings align with the broader literature, which suggests that the inclusion of sophisticated feature selection and hybrid approaches can significantly improve the predictive power of machine learning models.

The performance of deep learning models, such as LSTM and ConvLSTM, further supports the notion that advanced models can capture complex temporal patterns and dependencies within time series data. Although these models performed well in scenarios with high volatility and long-term dependencies, their overall accuracy and consistency were still lower than XGBoost, especially on public holidays and transitional periods where sudden shifts in market conditions occur. This confirms that while deep learning models have strong potential in time series forecasting, they may struggle with overfitting and adapting to abrupt changes in data patterns, a limitation commonly noted in the literature.

The date-based analysis provides further evidence of the superiority of machine learning models over traditional statistical methods. On regular weekdays and low volatility days, XGBoost models exhibited exceptional stability and predictive power, maintaining low MAPE values even without feature selection. In contrast, during high volatility periods and public holidays, where market behavior is more unpredictable, the performance of all models generally declined. However, XGBoost remained the most reliable method, consistently outperforming other approaches. This ability to adapt to diverse market conditions highlights the robustness of XGBoost and further supports its application for SMP forecasting in electricity markets, where sudden changes and non-linear dependencies are prevalent.

The results of this study are consistent with broader findings in the literature, which indicate that traditional models like ARIMA and SARIMA, despite their popularity, often fail to handle the complexities of electricity markets. Their linear structure and sensitivity to non-stationary data lead to higher error rates, particularly during periods of volatility or irregular market behavior. This research highlights the limitations of traditional models and demonstrates that modern machine learning approaches, equipped with advanced feature selection techniques, can offer substantial improvements in predictive performance.

The ability of XGBoost to maintain high accuracy across regular weekdays, weekends, and public holidays, while achieving MAPE values as low as 4.42% on high volatility days, underscores its effectiveness in addressing the challenges of SMP forecasting in a complex market like Turkey's.

In applying advanced machine learning models like XGBoost alongside feature selection techniques (i.e., MLFS and mRMR), some limitations must be acknowledged to present a balanced view of their effectiveness in forecasting. The scope, specific to Turkey's electricity market with its unique regulatory and economic dynamics, may limit external validity, as findings may not generalize to other markets. Moreover, models like XGBoost and ELM pose interpretability challenges, with XGBoost's complex structure complicating transparency, a potential hindrance in contexts where clear model insights are crucial for regulatory or operational decisions. The high computational demands, especially for hyperparameter tuning, can further challenge real-time applications in smaller markets. Addressing these issues through feature attribution for interpretability and approximate optimization to reduce computational load could enhance the models' broader usability. Scalability also presents concerns, as models like XGBoost and LSTM require substantial resources when applied to large datasets. Techniques like batch processing and parallelization are employed to handle larger datasets efficiently, while optimized tree structures in XGBoost reduce iteration time. Additionally, mRMR and MLFS streamline input variables, improving processing speed. These strategies collectively enhance scalability, though highly granular datasets still demand significant resources and extended processing times.

This study's findings could be strengthened by linking them more explicitly to economic theories and market behavior frameworks, thereby deepening the theoretical contributions and interpretive value of the forecasting model. For instance, the observed effects of peak demand periods and fluctuations in renewable energy inputs on SMP align with the demand–supply equilibrium theory, which suggests that prices adjust dynamically to shifts in supply and demand, seeking a balance that reflects current market conditions. Additionally, the volatility observed on public holidays and high-demand periods can be contextualized within theories of market efficiency and rational expectations, where price fluctuations represent the collective response of market participants to anticipated future states and newly available information. By embedding these results within established economic principles, the study not only enhances the relevance of its predictive model but also provides a broader theoretical basis for understanding and anticipating price movements in electricity markets, thus offering insights with applicability beyond the empirical scope of this research.

6.2. Future Work

Despite the substantial contributions of this research, several opportunities for future work remain. One potential avenue for exploration is the development of hybrid models that integrate multiple forecasting techniques, such as combining deep learning architectures with ensemble-based approaches. Hybrid models could leverage the strengths of individual methodologies to further improve the predictive accuracy and robustness of SMP forecasts.

Another promising direction is the application of more advanced deep learning frameworks, such as Transformers and Graph Neural Networks (GNNs). These architectures have shown exceptional performance in various domains, including natural language processing and network analysis, and could be adapted to capture complex spatiotemporal dependencies in electricity markets. The integration of these architectures with traditional models like ARIMA and SARIMA could also provide a more comprehensive approach to time series forecasting.

In addition to methodological innovations, future research should focus on expanding the scope and quality of datasets used for SMP prediction. Incorporating additional variables related to renewable energy integration, such as solar and wind generation data, could significantly enhance the models' ability to account for fluctuations in energy

supply. Furthermore, the inclusion of socio-economic factors and policy interventions could provide a more holistic view of market dynamics, leading to more accurate and context-aware forecasts.

Another area of future work involves real-time forecasting and dynamic model updating. Developing models capable of adapting to rapidly changing market conditions and updating their predictions in real time would be invaluable for both market participants and regulators. Implementing such adaptive models would require the use of advanced reinforcement learning techniques and continuous learning frameworks, enabling the models to evolve with changing market dynamics.

Moreover, the findings of this research suggest that feature selection plays a crucial role in improving model performance. Future studies could explore the development of new feature selection algorithms or the refinement of existing methods to identify even more relevant features for SMP forecasting. Automated feature engineering techniques, such as AutoML, could also be integrated into the model development process to streamline feature selection and hyperparameter optimization.

Lastly, future work could extend the findings of this research to other markets and regions, testing the applicability and generalizability of the proposed models and methods in different contexts. Comparative studies across multiple electricity markets, including those with varying levels of deregulation and renewable energy integration, would provide valuable insights into the strengths and limitations of each approach, informing the development of more versatile and globally applicable SMP forecasting models.

Author Contributions: Conceptualization, F.A. and M.K.; methodology, F.A. and M.K.; Formal analysis, M.K.; Investigation, M.K. and M.F.A.; Project administration, M.F.A.; Supervision, M.F.A.; Writing—original draft, F.A.; Writing—review and editing, F.A. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Mehmet Kızıldağ was employed by the company SABIS Bilgi Teknolojileri ve Sanayi Ticaret Ltd. Şti. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Manfre Jaimes, D.; Zamudio López, M.; Zareipour, H.; Quashie, M. A Hybrid Model for Multi-Day-Ahead Electricity Price Forecasting Considering Price Spikes. *Forecasting* **2023**, *5*, 499–521. [CrossRef]
2. Lee, T.H.; Lee, K.J.; Jo, B.W.; Kim, L.H.; Yeo, Y.K. Assessment of Three Forecasting Methods for System Marginal Prices. *Korean J. Chem. Eng.* **2011**, *28*, 1331–1339. [CrossRef]
3. Stanković, Z.Z.; Rajic, M.N.; Božić, Z.; Milosavljević, P.; Păcurar, A.; Borzan, C.; Păcurar, R.; Sabău, E. The Volatility Dynamics of Prices in the European Power Markets during the COVID-19 Pandemic Period. *Sustainability* **2024**, *16*, 2426. [CrossRef]
4. Wang, D.; Gryshova, I.; Kyzym, M.; Salashenko, T.; Khaustova, V.; Shcherbata, M. Electricity Price Instability over Time: Time Series Analysis and Forecasting. *Sustainability* **2022**, *14*, 9081. [CrossRef]
5. Durmus Senyapar, H.N.; Aksoz, A. Empowering Sustainability: A Consumer-Centric Analysis Based on Advanced Electricity Consumption Predictions. *Sustainability* **2024**, *16*, 2958. [CrossRef]
6. Chen, Q.; Guo, H.; Zheng, K.; Wang, Y. Day-Ahead Electricity Price Forecasting. In *Data Analytics in Power Markets*; Chen, Q., Guo, H., Zheng, K., Wang, Y., Eds.; Springer Nature: Singapore, 2021; pp. 113–134, ISBN 9789811649752.
7. Liu, J.; Hu, H.; Yu, S.S.; Trinh, H. Electricity Pricing and Its Role in Modern Smart Energy System Design: A Review. *Designs* **2023**, *7*, 76. [CrossRef]
8. Balakrishnan, R.; Geetha, V.; Kumar, M.R.; Leung, M.-F. Reduction in Residential Electricity Bill and Carbon Dioxide Emission through Renewable Energy Integration Using an Adaptive Feed-Forward Neural Network System and MPPT Technique. *Sustainability* **2023**, *15*, 14088. [CrossRef]

9. Guven, D.; Ozozen, A.; Kayakutlu, G.; Kayalica, M.O. Predicting Power Deviation in the Turkish Power Market Based on Adaptive Factor Impacts. In *Proceedings of the Artificial Intelligence for Knowledge Management*; Mercier-Laurent, E., Kayalica, M.Ö., Owoc, M.L., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 213–234.
10. Özgüner, E.; Tör, O.; Güven, A. Probabilistic Day-Ahead System Marginal Price Forecasting with ANN for the Turkish Electricity Market. *Turk. J. Electr. Eng. Comput. Sci.* **2017**, *25*, 4923–4935. [CrossRef]
11. Pan, K.; Shi, W.; Wang, X.; Li, J. A Short-Term Marginal Price Forecasting Model Based on Ensemble Learning. In Proceedings of the 2017 International Conference on Progress in Informatics and Computing (PIC), Nanjing, China, 15–17 December 2017; pp. 81–85.
12. Jufri, F.H.; Oh, S.; Jung, J. Day-Ahead System Marginal Price Forecasting Using Artificial Neural Network and Similar-Days Information. *J. Electr. Eng. Technol.* **2019**, *14*, 561–568. [CrossRef]
13. Panapakidis, I.P.; Moschakis, M.N. Comparison of Machine Learning Models for the Prediction of System Marginal Price of Greek Energy Market. *Int. J. Energy Environ. Eng.* **2019**, *13*, 148–152.
14. Akay, M.; Abut, F.; Bedir, B.; Erdoğan, M.; Kanar, E. Forecasting the System Marginal Price Using Long Short-Term Memory and Support Vector Machine. In Proceedings of the 6th International Mediterranean Science and Engineering Congress (IMSEC 2021), Antalya, Turkey, 25–27 October 2021; pp. 628–633.
15. Shim, S.W.; Lee, D.H.; Roh, J.H.; Park, J.-B. A Machine Learning-Based Algorithm for Short-Term SMP Forecasting Using 2-Step Method. *J. Electr. Eng. Technol.* **2023**, *18*, 1493–1501. [CrossRef]
16. Kim, T.; Ha, B.; Hwangbo, S. Online Machine Learning Approach for System Marginal Price Forecasting Using Multiple Economic Indicators: A Novel Model for Real-Time Decision Making. *Mach. Learn. Appl.* **2023**, *14*, 100505. [CrossRef]
17. Razak, I.A.W.A.; Abdullah, W.S.W.; Sulaima, M.F. Enhanced Short-Term System Marginal Price (SMP) Forecast Modelling Using a Hybrid Model Combining Least Squares Support Vector Machines and the Genetic Algorithm in Peninsula Malaysia. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 289–298.
18. Awad, M.; Khanna, R. Support Vector Machines for Classification. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Awad, M., Khanna, R., Eds.; Apress: Berkeley, CA, USA, 2015; pp. 39–66, ISBN 978-1-4302-5990-9.
19. Ahn, J.M.; Kim, J.; Kim, K. Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. *Toxins* **2023**, *15*, 608. [CrossRef]
20. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
21. Ding, S.; Zhao, H.; Zhang, Y.; Xu, X.; Nie, R. Extreme Learning Machine: Algorithm, Theory and Applications. *Artif. Intell. Rev.* **2015**, *44*, 103–115. [CrossRef]
22. Wang, J.; Lu, S.; Wang, S.-H.; Zhang, Y.-D. A Review on Extreme Learning Machine. *Multimed. Tools Appl.* **2022**, *81*, 41611–41660. [CrossRef]
23. Lv, Y.; Li, B.; Yu, J.; Ding, Y. Reduce Training Error of Extreme Learning Machine by Selecting Appropriate Hidden Layer Output Matrix. *J. Syst. Sci. Syst. Eng.* **2021**, *30*, 552–571. [CrossRef]
24. Kontopoulou, V.I.; Panagopoulos, A.D.; Kakkos, I.; Matsopoulos, G.K. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet* **2023**, *15*, 255. [CrossRef]
25. Khashei, M.; Bijari, M.; Hejazi, S.R. Combining Seasonal ARIMA Models with Computational Intelligence Techniques for Time Series Forecasting. *Soft Comput.* **2012**, *16*, 1091–1105. [CrossRef]
26. Korstanje, J. The SARIMA Model. In *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*; Korstanje, J., Ed.; Apress: Berkeley, CA, USA, 2021; pp. 115–122, ISBN 978-1-4842-7150-6.
27. Szostek, K.; Mazur, D.; Drałus, G.; Kusznier, J. Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production. *Energies* **2024**, *17*, 4803. [CrossRef]
28. Van Houdt, G.; Mosquera, C.; Nápoles, G. A Review on the Long Short-Term Memory Model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955. [CrossRef]
29. Fan, Y.; Tang, Q.; Guo, Y.; Wei, Y. BiLSTM-MLAM: A Multi-Scale Time Series Prediction Model for Sensor Data Based on Bi-LSTM and Local Attention Mechanisms. *Sensors* **2024**, *24*, 3962. [CrossRef] [PubMed]
30. Cheng, Q.; Chen, Y.; Xiao, Y.; Yin, H.; Liu, W. A Dual-Stage Attention-Based Bi-LSTM Network for Multivariate Time Series Prediction. *J. Supercomput.* **2022**, *78*, 16214–16235. [CrossRef]
31. Song, T.; Xu, G.; Yang, K.; Li, X.; Peng, S. Convformer: A Model for Reconstructing Ocean Subsurface Temperature and Salinity Fields Based on Multi-Source Remote Sensing Observations. *Remote Sens.* **2024**, *16*, 2422. [CrossRef]
32. Madhusudhanan, K.; Jawed, S.; Schmidt-Thieme, L. Hyperparameter Tuning MLP's for Probabilistic Time Series Forecasting. In *Proceedings of the Advances in Knowledge Discovery and Data Mining*; Yang, D.-N., Xie, X., Tseng, V.S., Pei, J., Huang, J.-W., Lin, J.C.-W., Eds.; Springer Nature: Singapore, 2024; pp. 264–275.
33. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing* **2022**, *503*, 92–108. [CrossRef]
34. Emmert-Streib, F.; Yang, Z.; Feng, H.; Tripathi, S.; Dehmer, M. An Introductory Review of Deep Learning for Prediction Models With Big Data. *Front. Artif. Intell.* **2020**, *3*, 4. [CrossRef]

35. Surakhi, O.; Zaidan, M.A.; Fung, P.L.; Hossein Motlagh, N.; Serhan, S.; AlKhanafseh, M.; Ghoniem, R.M.; Hussein, T. Time-Lag Selection for Time-Series Forecasting Using Neural Network and Heuristic Algorithm. *Electronics* **2021**, *10*, 2518. [CrossRef]
36. Zhao, Z.; Anand, R.; Wang, M. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. In Proceedings of the 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Washington, DC, USA, 5–8 October 2019; pp. 442–452.
37. Suzuki, T.; Sugiyama, M.; Sese, J.; Kanamori, T. Approximating Mutual Information by Maximum Likelihood Density Ratio Estimation. *J. Mach. Learn. Res.-Proc. Track* **2008**, *4*, 5–20.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.