OKU Fen Bilimleri Enstitüsü Dergisi
9(4): 2384-2419, 2026

OKU Journal of The Institute of Science and Technology, 9(4): 2384-2419, 2026

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of the Institute of Science and Technology

# Dalgalı Elektrik Piyasalarında Fiyat Belirleyicileri Analizi: Açıklanabilir Yapay Zeka Çerçevesi

**Ömer POLAT[1], Saltuk Buğra SELÇUKLU[2*]**

[1,2]Erciyes Üniversitesi, Enerji Sistemleri Mühendisliği Bölümü, Kayseri

[1]https://orcid.org/0000-0001-7197-9138
[2]https://orcid.org/0000-0002-9295-3866
*Corresponding author: sbselcuklu@erciyes.edu.tr

**Araştırma Makalesi**

***Makale Tarihçesi:***
Geliş tarihi: 01.02.2026
Kabul tarihi:02.06.2026
Online Yayınlanma: 09.09.2026

***Anahtar Kelimeler:***
Elektrik fiyat tahmini
Açıklanabilir yapay zeka
Liyakat sırası etkisi
LightGBM
SHAP
Enerji ekonomisi

## ÖZ

Yenilenebilir enerji kaynaklarının entegrasyonu, elektrik piyasalarına önemli ölçüde oynaklık getirmiş ve geleneksel tahmin modellerini yetersiz kılmıştır. Gelişmiş makine öğrenmesi modelleri üstün doğruluk sunsa da bu modellerin kara kutu doğası, risk yönetimi için yorumlanabilir analizlere ihtiyaç duyan piyasa katılımcıları açısından kritik bir güven boşluğu yaratmaktadır. Bu çalışma, gelişmekte olan ekonomilerde Gün Öncesi Piyasası fiyatlarını tahmin etmek için yorumlanabilir ve açıklanabilir bir yapay zekâ çerçevesi geliştirerek bu sorunu ele almaktadır. Çalışmanın literatüre temel katkısı, elektrik fiyat tahminini yalnızca doğruluk odaklı bir problem olarak ele almaması; çok değişkenli tahmin yapısını temel ekonomik teorilerle ilişkilendiren açıklanabilir bir karar destek çerçevesi sunmasıdır. Bu yönüyle çalışma, özellikle yapısal oynaklığın, döviz kuru duyarlılığının ve ithal yakıt bağımlılığının yüksek olduğu gelişmekte olan elektrik piyasalarında şeffaflık eksikliğini gidermeye yönelik özgün bir katkı sağlamaktadır. Yüksek döviz kuru oynaklığına sahip piyasalar için temsil niteliğinde bir vaka çalışması olarak Türkiye'nin kullanıldığı bu araştırmada, yedi makine öğrenmesi algoritması ve bir klasik istatistiksel karşılaştırma modeli kıyaslanmıştır. Elde edilen bulgular, Light Gradient Boosting Machine modelinin hata ölçütleri bakımından en başarılı model olduğunu göstermektedir. Ayrıca Diebold-Mariano testi, model performansları arasındaki farkların yalnızca sayısal hata ölçütleri üzerinden değil, istatistiksel anlamlılık açısından da değerlendirilmesini sağlamıştır.Özellikle, modelin ekonomik teoriye uygunluğunu doğrulamak amacıyla SHapley Additive exPlanations yöntemi kullanılmış; böylece modelin doğrusal olmayan Liyakat Sırası Etkisini bağımsız olarak öğrenebilme yeteneği teyit edilmiştir. Buna ek olarak, model çıktıları döviz kuru baskısı, yakıt maliyeti geçişkenliği, yenilenebilir üretimin fiyat düşürücü etkisi ve stratejik hidroelektrik üretim davranışı açısından yorumlanmıştır. Şeffaf olmayan tahminleri açıklanabilir ekonomik sinyallere dönüştüren bu çerçeve, tahmin doğruluğu ile paydaş güveni arasındaki boşluğu doldurmakta ve yüksek riskli enerji ticareti ortamlarında karar verme süreçleri için güçlü bir araç sunmaktadır.

# Unveiling Price Drivers in Volatile Electricity Markets: An Explainable AI Framework

**Research Article**

***Article History:***
Received: 01.02.2026
Accepted: 02.06.2026
Published online: 09.09.2026

## ABSTRACT

The integration of renewable energy sources has introduced significant volatility into electricity markets, rendering traditional forecasting models insufficient. While advanced machine learning models offer superior accuracy, their black-box nature creates a critical trust gap for market participants who require

***Keywords:***
Electricity price forecasting
Explainable artificial intelligence
Merit order effect
LightGBM
SHAP
Energy economics

interpretable insights for risk management. This study addresses this challenge by developing an interpretable and explainable artificial intelligence framework for forecasting day-ahead market prices in emerging economies. The main contribution of this study to the literature is that it does not treat electricity price forecasting solely as an accuracy-oriented task; instead, it establishes a transparent decision-support framework that connects a multivariable forecasting structure with fundamental economic theory. In this respect, the study provides a novel perspective on price formation in electricity markets characterized by structural instability, exchange rate sensitivity, and dependence on imported fuels. Using Türkiye as a representative case study for markets with high currency volatility, we benchmark seven machine learning algorithms and one classical statistical benchmark. The empirical findings show that the Light Gradient Boosting Machine model achieves the best performance in terms of forecasting error metrics. In addition, the Diebold-Mariano test is employed to evaluate whether the observed differences in forecasting accuracy are statistically meaningful rather than being assessed only through numerical performance measures. Notably, we utilize SHapley Additive exPlanations to validate the model against economic theory, thereby confirming its ability to learn the non-linear merit order effect independently. Furthermore, the model outputs are interpreted in relation to exchange rate pressure, fuel-cost pass-through, the price-suppressing role of renewable generation, and strategic hydroelectric dispatch behavior. By transforming opaque predictions into explainable economic signals, this framework bridges the gap between predictive accuracy and stakeholder trust, offering a robust tool for decision-making in high-stakes energy trading environments.

**To Cite:** Polat Ö., Selçuklu SB. Unveiling Price Drivers in Volatile Energy Markets: An Explainable AI Framework. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2026; 9(4): 2384-2419.

## 1. Introduction

Electricity markets serve as the nervous system of modern energy infrastructure, where the Market Clearing Price (MCP) acts as a critical signal for generation dispatch, investment, and consumption behavior. In stable, mature markets, price formation is relatively predictable, driven by established demand cycles and fuel costs. However, the global energy transition has introduced profound complexity. The integration of intermittent renewable energy sources (RES) has increased volatility, while emerging economies face the compounded challenges of high import dependency and exchange rate fluctuations (Weron, 2014; Ziel et al., 2015; Lago et al., 2018; Chang et al., 2019).

The quest for accurate Electricity Price Forecasting (EPF) has generated a vast body of literature. Early research predominantly utilized statistical time series models. Niimura (2006), Haghi and Tafreshi (2007), and Weron (2014) have extensively reviewed these foundational techniques, which include Autoregressive (AR), Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA) models (Contreras et al., 2002; Nogales et al., 2002; Conejo et al., 2005; Xie et al., 2013; Nowotarski et al., 2014). While these linear models offer interpretability and stability during calm market periods (Misiorek et al., 2006; Gürtler and Paulsen, 2018), they often struggle to capture the non-linear dynamics and high-frequency volatility introduced by smart grids and renewable penetration (Fulcher et al., 2013; Grossi and Nan, 2019).

To address these limitations, the field has shifted its focus toward computational intelligence. A multitude of studies have demonstrated the efficacy of Deep Learning (DL) architectures, including Long Short-Term Memory (LSTM) networks (Lago et al., 2018; Peng et al., 2018; Yu et al., 2019;

Backe et al., 2023; Plakas et al., 2023) and Convolutional Neural Networks (CNN) (Marcjasz et al., 2020; Lehna et al., 2022; Silva et al., 2023). Similarly, Machine Learning (ML) algorithms such as Support Vector Machines (SVM) (Zhang and Chuntian, 2008; Che and Wang, 2010; Díaz et al., 2019; Tschora et al., 2022) and Random Forests (RF) (Mei et al., 2014; Ludwig et al., 2015) have become standard tools. Recent advancements have highlighted the performance of gradient boosting machines like XGBoost and LightGBM (Brownlee, 2016; Brownlee, 2021), which frequently outperform both statistical benchmarks and heavier DL models in tabular data tasks (Prokhorenkova et al., 2018). Although LSTM and hybrid deep learning models are powerful alternatives for sequential forecasting problems, the present study prioritizes tree-based gradient boosting algorithms because the dataset is structured as a multivariable tabular market dataset enriched with lagged prices, generation variables, market bid volumes, fuel prices, exchange rates, and temperature indicators. In this structure, temporal dependence is already represented through engineered lag variables, while tree-based ensemble models provide an effective balance between predictive accuracy, computational efficiency, and post-hoc interpretability. Therefore, the model selection is not based solely on algorithmic popularity, but on the suitability of each method for the data structure and the explainable decision-support objective of the study.

On the other hand, the pursuit of predictive accuracy has come at a cost: interpretability. As models become increasingly complex, they often become black boxes, obscuring the logic behind their outputs. In high-stakes environments like energy trading, a forecast without an explanation is a liability. Stakeholders need to distinguish whether a structural shift, such as currency devaluation, or a transient event, such as a temporary renewable generation shortage, causes a predicted price spike. This has led to the emergence of Explainable Artificial Intelligence (XAI) as a critical research frontier. Techniques like SHapley Additive exPlanations (SHAP) have begun to bridge the gap between performance and understanding (Liu et al., 2023; Pütz et al., 2023), yet their application in EPF remains nascent, particularly in verifying whether ML models align with fundamental economic theories.

This study utilizes the Turkish Day-Ahead Market as a strategic case study to address this gap. Türkiye represents a specific class of emerging markets characterized by rapid renewable energy expansion, paired with a heavy reliance on imported fossil fuels and currency volatility. Previous research on the Turkish market (Yıldırım et al., 2012; Hayfavi and Talasli, 2014; Ozyildirim and Beyazıt, 2014; Kölmek and Navruz, 2015; Taysi et al., 2015; Ozozen et al., 2016; Avci et al., 2018; Ugurlu et al., 2018) has largely focused on error minimization using older datasets or standard architectures, often neglecting the explicit quantification of price drivers via XAI. Details of EPF studies in the Turkish day-ahead market are presented in Appendix A. Accordingly, the contribution of this study moves beyond local price forecasting and focuses on building an interpretable decision-support framework for volatile and import-dependent electricity markets. The main contributions of the study are as follows:

1. Methodological Rigor: The framework benchmarks multiple forecasting approaches, including standard machine learning models, gradient boosting algorithms, and a neural network

architecture, under the same data structure. Model performance is evaluated not only through conventional error metrics but also through the Diebold-Mariano test, which provides a formal statistical comparison of predictive accuracy.

2. Economic Explainability: The framework utilizes SHAP to deconstruct model decisions, verifying the Merit Order Effect and revealing strategic bidding behaviors that linear models often miss. In this way, the model outputs are connected with electricity market mechanisms such as renewable generation effects, fuel-cost pass-through, exchange rate pressure, and hydroelectric dispatch behavior.
3. Practical Decision Support: The proposed framework transforms black-box forecasts into interpretable economic signals that can support market participants in risk management, bidding strategy, and operational decision-making. This is particularly important for emerging markets where both market fundamentals and macroeconomic instability shape price formation.
4. Transparency and Reproducibility: The study presents a transparent workflow covering data preprocessing, model benchmarking, statistical testing, and explainability analysis. This structure improves the reproducibility of the analysis and helps reduce the gap between high-accuracy forecasting models and accountable decision-making in electricity markets.

## 2. Material and Methodology

### *2.1. The Case Study: Characteristics of Import-Dependent Markets*

Türkiye serves as an ideal laboratory for analyzing price formation in volatile economies. Muğaloğlu et al. (2025) examine the price volatility dynamics of Türkiye's day-ahead electricity market from 2018 to 2024, using the same dataset as this paper. The Turkish market shares several structural characteristics with many emerging electricity markets: (1) a rapidly growing share of renewables (Hydro/Wind/Solar), (2) a heavy reliance on imported fossil fuels (Natural Gas and Coal) for baseload generation, and (3) significant sensitivity to foreign exchange rates. Detailed information on the Turkish day-ahead market (EXIST) is provided in Appendix B. These characteristics make Türkiye a suitable case for examining not only forecasting accuracy, but also how machine learning models interpret the economic and operational drivers of price formation. Understanding how ML models interpret specific drivers offers valid insights for other markets facing similar structural constraints.

### *2.2. Data Curation and Preprocessing*

The dataset spans the period from January 1, 2018, to December 31, 2022. Data prior to 2018 was excluded due to structural market immaturity and data quality issues. The dataset includes 33 independent variables and the target variable, market-clearing price (MCP), which are detailed in Appendix C. Since correlation indicates the direction and effect of linear relationships between variables, evaluating these interactions allows us to make logical inferences and identify features that

do not contribute to predictions. The resulting linear relationships among the variables are illustrated as a correlation heat map in Figure C. 1.

**Preprocessing:** To ensure data integrity and model robustness, a multi-stage preprocessing pipeline was implemented. Initially, the dataset was examined for structural inconsistencies, erroneous entries, and discontinuities across the available time span. Since the earlier period contained substantial missing data and quality limitations, observations prior to 2018 were excluded from the analysis. Therefore, the final modelling period was restricted to the most reliable and contiguous interval between January 1, 2018, and December 31, 2022. This selection reduced the risk of introducing artificial patterns through extensive missing-value treatment and ensured that the forecasting models were trained on a consistent and comparable feature space. After this temporal filtering and data-cleaning process, the final dataset used for model training and testing did not contain missing values requiring imputation. For this reason, no interpolation-based or model-based imputation procedure was applied. For feature scaling, Min-Max Normalization was applied to distance-dependent architectures such as SVR and ANN to ensure equitable feature weighting, whereas tree-based models were trained on raw features to preserve interpretability and avoid unnecessary transformation bias.

### *2.3. Forecasting Models*

To ensure a robust evaluation, we compare four categories of models:

- **Classical Statistical Benchmarks:** Linear Regression (LR) was included to provide a traditional statistical baseline for comparison. This model is transparent and widely used as a reference method, but it is generally less flexible in capturing non-linear interactions among electricity market variables.
- **Machine Learning:** Random Forest (RF), Support Vector Regression (SVR), Decision Trees (DT).
- **Gradient Boosting:** XGBoost, CatBoost, and LightGBM.
- **ANN:** Sequential Artificial Neural Network.

The selected models represent different learning structures and therefore allow a balanced comparison between traditional, non-linear, ensemble-based, and neural approaches. Random Forest was trained through bootstrap aggregation of multiple decision trees, while the Decision Tree model was trained using recursive binary partitioning based on squared-error reduction. Support Vector Regression was trained using the epsilon-insensitive regression objective with regularization. XGBoost, CatBoost, and LightGBM were trained as gradient boosting decision tree models, where each new tree sequentially improves the residual errors of the previous learners. In particular, the LightGBM model was trained using a leaf-wise, best-first tree growth strategy rather than the traditional level-wise approach. This strategy splits the leaf that provides the maximum reduction in loss, enabling the model to capture sparse

but high-impact signals in the feature space. This property is especially relevant for the Turkish Day-Ahead Market, where price formation is affected by irregular but important movements in supply-demand conditions, fuel prices, exchange rates, and generation mix. The Sequential Artificial Neural Network was trained using the Adam optimizer with the mean squared error loss function. Hyperparameter tuning was conducted through grid search and Keras Tuner, where applicable. Details of model parameters and methodology are given in Appendix D.

### *2.4. Evaluation Strategy*

Performance is evaluated using Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), R-squared ($R^2$), Mean Squared Error (MSE), and Mean Absolute Error (MAE). Details on these metrics are given in Appendix E. Furthermore, we employ the Diebold-Mariano (DM) test to determine if the proposed model provides a statistically significant improvement.

The Diebold-Mariano test was specifically employed to move beyond basic error metrics such as RMSE and MAE. In electricity markets, even marginal differences in forecast errors may correspond to considerable financial exposure for market participants. However, a lower numerical error value alone does not necessarily prove that one model is statistically superior to another. The DM test addresses this issue by testing the null hypothesis of equal predictive accuracy between competing forecast series. Therefore, it provides a rigorous statistical basis for evaluating whether the observed performance differences among competing models reflect systematic differences in predictive accuracy rather than random variation in the test set. This is particularly important in the present study because the leading gradient boosting models produce relatively close forecasting results, making statistical validation necessary for a reliable model comparison.

## 3. Results

### *3.1. Predictive Performance*

This section outlines the performance evaluation of the candidate models, a crucial step in selecting the optimal predictive engine for our proposed decision support system. In addition to the machine learning models, Linear Regression (LR) was included as a classical statistical benchmark to provide a traditional reference point for model comparison. The forecast performance of the eight models is presented below. Detailed results of the error rates of all prediction models are given in Appendix H.

Since the market-clearing price (MCP) series contains zero or near-zero observations, MAPE values should be interpreted cautiously. In such cases, percentage-based errors may become disproportionately large even when the absolute prediction error is moderate. For this reason, the model comparison is primarily evaluated through $R^2$, MAE, MSE, and RMSE, while MAPE is retained as an additional percentage-based indicator.

Table 1 presents the error metrics ($R^2$, MAE, MSE, RMSE, and MAPE) obtained during the training phase. LightGBM and XGBoost demonstrated the highest fit with $R^2$ values of 0.996 and the lowest

error rates, indicating that they effectively captured the underlying patterns in the training data without excessive overfitting compared to simpler models, such as DT. The LR benchmark also produced a reasonable fit in terms of $R^2$; however, its substantially higher error values indicate that a purely linear specification is limited in capturing abrupt price movements and non-linear interactions in the electricity market.

**Table 1.** Training set error metrics (2018-2022)

| Model | $R^2$ | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|---|
| LR | 0.892 | 9.090 | 268.970 | 16.400 | 123.770 |
| RF | 0.990 | 2.889 | 24.953 | 4.995 | 20.654 |
| SVR | 0.956 | 6.088 | 108.723 | 10.427 | 43.632 |
| XGBoost | 0.996 | 2.238 | 10.744 | 3.278 | 13.827 |
| Sequential | 0.971 | 4.780 | 71.490 | 8.459 | 16.146 |
| CatBoost | 0.983 | 4.310 | 41.267 | 6.424 | 30.209 |
| LightGBM | 0.996 | 2.265 | 10.944 | 3.308 | 14.629 |
| DT | 0.950 | 6.367 | 125.300 | 11.194 | 54.977 |

Table 2 summarizes the performance on the test set, which is crucial for verifying model generalization. The LightGBM model consistently outperformed other models, achieving the highest $R^2$ (0.950) and the lowest MAE (5.981) and RMSE (11.248). While CatBoost and XGBoost also performed robustly, LightGBM provided the most accurate forecasts for the volatile period from 2018 to 2022. Compared with the advanced machine learning models, the LR benchmark showed weaker predictive performance, particularly in terms of MAE, RMSE, and MAPE. This result suggests that a simple linear structure does not adequately represent the relationship between market-clearing prices and explanatory variables. In particular, price spikes, threshold effects, renewable generation variability, exchange rate movements, fuel-cost changes, and bidding behavior create non-linear patterns that require more flexible forecasting models.

**Table 2.** Test set error metrics (2018-2022)

| Model | $R^2$ | MAE | MSE | RMSE | MAPE |
|---|---|---|---|---|---|
| LR | 0.890 | 9.390 | 288.850 | 17.001 | 178.940 |
| RF | 0.934 | 6.612 | 166.431 | 12.901 | 64.849 |
| SVR | 0.942 | 6.921 | 145.590 | 12.066 | 59.077 |
| XGBoost | 0.946 | 6.113 | 134.433 | 11.595 | 52.852 |
| Sequential | 0.944 | 6.478 | 141.787 | 11.908 | 33.128 |
| CatBoost | 0.948 | 6.512 | 131.601 | 11.472 | 48.796 |
| LightGBM | 0.950 | 5.981 | 126.528 | 11.248 | 49.627 |
| DT | 0.893 | 8.369 | 269.699 | 16.243 | 97.705 |

The inclusion of LR strengthens the empirical comparison by showing that classical statistical modelling can provide a useful baseline but remains insufficient for the volatile and structurally complex price

dynamics observed in the Turkish Day-Ahead Market. Although LR achieves an $R^2$ value close to the Decision Tree model, its higher MSE, MAE, RMSE, and especially MAPE indicate weaker robustness against extreme price movements. Therefore, the results support the use of advanced machine learning models, particularly gradient boosting algorithms, for electricity price forecasting in volatile markets.

The strong performance of LightGBM is crucial for our application, as even marginal improvements in forecast accuracy can have significant financial implications for market participants. However, the selection of LightGBM is not based solely on its numerical superiority. Its ability to process the 33-variable market dataset efficiently, capture non-linear interactions, and provide a suitable basis for SHAP-based interpretation makes it an appropriate core model for the proposed explainable decision-support framework.

### *3.2. Statistical Significance (Diebold-Mariano Test)*

To examine whether the observed forecasting performance differences are statistically meaningful and not merely a result of random variation, we conducted the Diebold-Mariano (DM) test using LightGBM as the reference model. The mathematical formulation is detailed in Appendix F. We compared the forecasts of LightGBM against those of the second and third best-performing models (CatBoost and XGBoost), as well as the Sequential ANN.

The DM test results, presented in Table 3, evaluate the null hypothesis ($H_0$) that there is no significant difference in forecast accuracy between the models.

**Table 3.** Diebold-Mariano test results (reference model: LightGBM)

| Comparison | DM Statistic | P-Value | Result |
|---|---|---|---|
| LightGBM vs CatBoost | -1.4465 | 0.148044 | Not Significant ($p > 0.05$) |
| LightGBM vs XGBoost | -3.7502 | 0.000177 | *** |
| LightGBM vs Sequential | -3.3141 | 0.000919 | *** |

Note: Significance levels are denoted as *** for $p < 0.01$. "Not Significant" indicates $p > 0.05$.

The analysis of the DM statistics reveals distinct performance tiers. As indicated in Table 3, the statistics for XGBoost and ANN yield p-values well below the 0.01 threshold, which allows us to strongly reject the null hypothesis and statistically confirm that LightGBM provides significantly more accurate forecasts than these competing models. In contrast, the comparison between LightGBM and CatBoost produces a p-value of 0.148, which is above the conventional 0.05 significance level. Therefore, the null hypothesis of equal predictive accuracy cannot be rejected for these two models. This means that LightGBM should not be interpreted as statistically superior to CatBoost, although it achieves slightly lower absolute error values in the test set.

This result provides a more nuanced basis for model selection. LightGBM is retained as the core model of the explainable AI framework because it combines the best numerical performance with computational efficiency and compatibility with SHAP-based interpretation. CatBoost is therefore recognized as a strong alternative model, while LightGBM is selected due to its overall balance of

accuracy, efficiency, and interpretability. This distinction is important because the purpose of the study is not only to identify the model with the lowest error values, but also to develop a reliable and explainable decision-support framework for volatile electricity markets.

## 4. Discussion

While accuracy is necessary, trust is paramount in forecasting. Using SHAP (SHapley Additive exPlanations), we deconstruct the LightGBM model to verify whether its learning aligns with established energy economics (SHAP graphs of all other models are given in Appendix I). This section moves beyond simple feature importance to analyze the direction and magnitude of each variable's impact. Run times of GridSearch and SHAP Analysis are given in Appendix G.

It should be emphasized at the outset that SHAP is used in this study as a post-hoc interpretability tool rather than as a causal identification method. SHAP values explain how the trained model distributes the predicted outcome across input variables, but they do not, by themselves, establish direct causal relationships. Therefore, the interpretations below are treated as model-based explanations and are evaluated together with electricity market theory, institutional knowledge, and observed market behavior. This distinction is particularly important in electricity markets, where many variables are jointly determined and may reflect simultaneous supply-demand responses.

### *4.1. Interpreting SHAP Values*

SHAP values provide a visual representation of the impact of individual features on model predictions. In the summary plot (Figure 1), the y-axis ranks features by their overall importance, while the x-axis quantifies the impact on the Market Clearing Price (MCP).

- **Color (Feature Value):** Red dots represent high values of a feature (e.g., high gas price), while blue dots represent low values.
- **Position (Impact):** The horizontal position indicates the SHAP value. Dots to the right of the center line (positive) suggest that the feature drives the price up. Dots to the left (negative) indicate the feature drives the price down.
- **Distribution:** The vertical spread of dots at a specific feature value shows the variability of the impact, often capturing non-linear interactions with other variables.

Accordingly, the SHAP summary plot is interpreted not only as a ranking of variables, but also as an explanation of how each variable contributes to upward or downward price movements under different market conditions. This provides a more informative perspective than conventional feature importance measures, which generally indicate only the relative importance of variables without showing the direction of their effects.

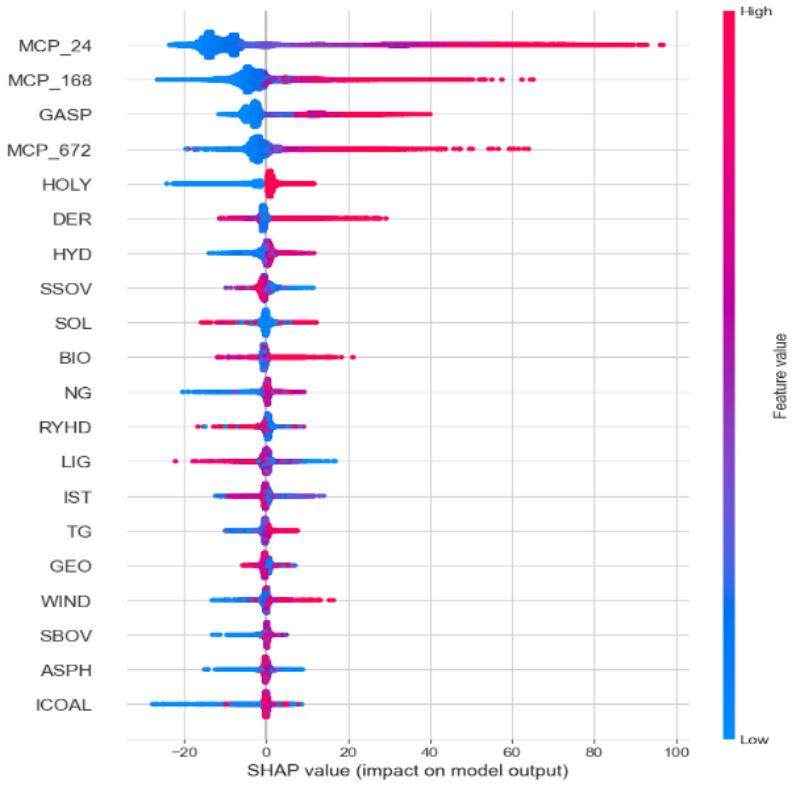


**Figure 1.** LightGBM SHAP values

*4.2. Historical Price Dependencies and Market Inertia*

The analysis reveals that lagged price variables, specifically MCP-24 (yesterday's price at the same hour) and MCP-168 (last week's price at the same hour), are overwhelmingly the most dominant predictors.

- **Market Inertia:** The SHAP plot reveals a clear trend, where high past prices strongly predict high current prices, while low past prices predict price drops, which reflects the strong inertia in electricity markets, where supply-demand conditions often persist across 24-hour cycles.

- **Non-Linearity:** The model captures the non-linear nature of this influence; extreme price events in the past have a disproportionately amplified effect on the forecast compared to average price fluctuations.

The reliance on lagged features is statistically supported by Autocorrelation (ACF) and Partial Autocorrelation (PACF) analyses performed on the dataset (Figure 2). The ACF analysis reveals a strong positive correlation with short-term past data, confirming that past values serve as crucial benchmarks for forecasting. The PACF analysis further refines this by showing high values for the first few lags that decrease rapidly, emphasizing that the recent history (the last few hours or days) contains the most critical predictive signal, which diminishes over time.

This finding is consistent with the structure of day-ahead electricity markets, where recent clearing prices, repeated demand cycles, and persistent fuel-cost conditions strongly shape price expectations.

However, the SHAP results should be interpreted as evidence that the model uses lagged prices as strong predictive signals, not as proof that past prices mechanically cause current prices.

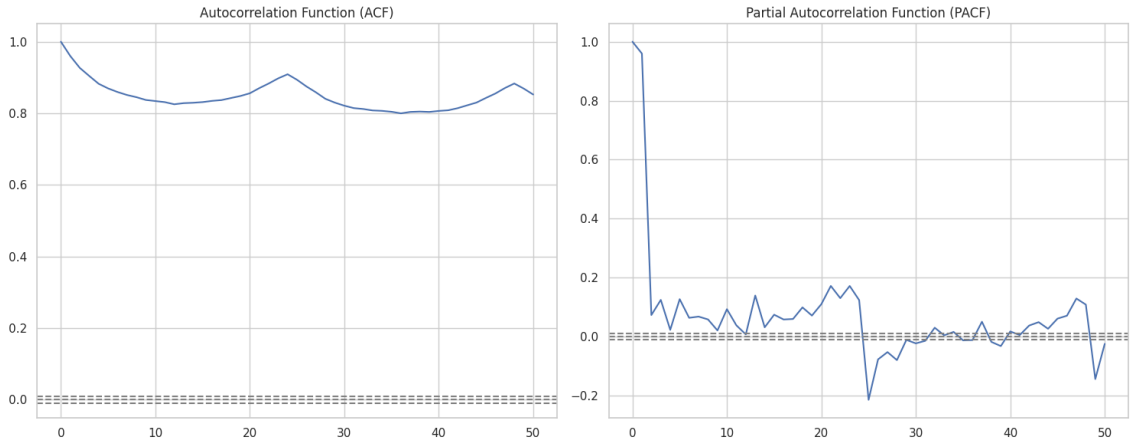


**Figure 2.** Autocorrelation (ACF) and partial autocorrelation (PACF) analysis

*4.3. Validation of the Merit Order Effect*

The Merit Order Effect posits that electricity sources with low marginal costs (such as wind and solar) displace more expensive thermal generation, thereby lowering prices. Our SHAP analysis confirms that the model captures this theory:

- **Renewables (Solar/Wind):** High values of Solar generation correspond to negative SHAP values, indicating a price-suppressing effect, which confirms the model has learned that zero-marginal-cost generation shifts the supply curve to the right. While Wind generation generally follows this trend, the model occasionally detects anomalies where high wind correlates with price increases, likely capturing periods where wind co-occurs with extreme demand shocks.
- **Fossil Fuels (Gas/Coal):** Conversely, Natural Gas prices (GASP) show a strong positive correlation. As an import-dependent market, the marginal price in Türkiye is frequently set by gas-fired plants; thus, global gas price fluctuations directly pass through to electricity prices.

The consistency between these SHAP patterns and the Merit Order Effect strengthens the economic credibility of the model. In other words, the model is not only producing accurate forecasts but is also relying on relationships that are broadly consistent with established electricity market theory. Nevertheless, this should be understood as economic validation of the model's learned associations, not as a causal estimate of the exact price effect of each generation source.

*4.4. The Counterintuitive Role of Hydro: Strategic Bidding*

A notable finding in the SHAP analysis is the positive relationship between Dam Hydro generation (HYD) and MCP. Intuition suggests that more hydro generation should lower prices. However, the model reveals a more sophisticated market dynamic: Strategic Bidding. Unlike run-of-river plants, a dam hydro acts as energy storage. Operators rationally withhold water during low-price hours and

discharge it during peak-price hours to maximize revenue. Therefore, high hydro generation volumes coincide with high-price periods (peak load), not because hydro causes high prices, but because high prices induce hydro generation.

This interpretation is important for avoiding a misleading reading of the SHAP output. The positive SHAP contribution of dam hydro generation should not be interpreted as a direct causal statement that hydro generation increases market prices. Rather, it reflects the model's ability to capture the timing of hydro dispatch under market conditions in which prices are already elevated. The model therefore distinguishes, to some extent, between the physical availability of hydro capacity and the economic decision to dispatch that capacity during high-value hours.

### *4.5. Calendar Effects: The Ambiguity of Holiday*

The HOLY variable acts as a binary indicator, distinguishing between business days (1) and weekends/holidays (0).

- **Business Days (Red):** A dense cluster of red dots appears on the positive side of the SHAP axis, which indicates that being a business day consistently exerts a minor upward pressure on prices due to higher industrial and commercial demand.
- **Holidays/Weekends (Blue):** The blue dots are distributed widely across the negative side. This variance suggests that not all "off days" are equal; the downward pressure on price varies significantly depending on whether the day is a standard Sunday or a major national holiday, likely interacting with other demand variables.

This result shows that calendar effects operate as conditional market signals rather than as uniform price determinants. In particular, the effect of holidays and weekends may vary depending on demand intensity, weather conditions, renewable generation, and the position of the day within the weekly consumption cycle.

### *4.6. Macro-Economic Pressures*

The Daily Exchange Rate (DER) emerges as a top-tier predictor. In import-dependent economies, currency depreciation increases the cost of imported coal and gas. The SHAP dependence plot for DER shows a sharp increase, quantifying the "pass-through" rate of currency devaluation into electricity prices. This insight is critical for policymakers monitoring inflationary pressures.

For the Turkish market, this finding is particularly relevant because imported fuel costs are closely linked to foreign exchange movements. The model's reliance on DER therefore reflects a structural characteristic of import-dependent electricity systems: macroeconomic instability can be transmitted into electricity prices through the fuel-cost channel.

*4.7. Analysis of Secondary Market Drivers*

Beyond the primary drivers, the model's analysis of secondary factors reveals nuanced market behaviors:

- **Biomass (BIO):** Unlike other renewables, high biomass generation is associated with a slight increase in MCP, which reflects the economic reality that biomass has a higher marginal cost compared to wind or solar, often acting as a price setter.
- **Lignite and Import Coal:** High generation from Lignite (LIG), a domestic resource, consistently lowers the MCP. Interestingly, Import Coal (ICOAL) exhibits a strong effect only when generation is low; the low output from import coal plants is strongly correlated with price drops, suggesting that these plants are often marginal units that set higher prices when active.
- **Supply Bids (SSOV):** The Submitted Sales Order Volume (SSOV) presents a paradox. While basic economics suggests high supply should lower prices, the model shows that high SSOV values do not automatically lead to price drops because high supply volumes often occur during peak demand hours. The model has learned that the cost structure of the supply (merit order) matters more than the raw volume offered.
- **Temperature:** The model includes Istanbul temperature data, but its impact is mixed. Extreme temperatures (both hot and cold) tend to increase prices due to cooling/heating demand, while moderate temperatures have a negligible effect.

Taken together, these secondary drivers show that the model captures more than simple one-directional relationships. In several cases, the impact of a variable depends on its interaction with demand conditions, generation mix, fuel prices, and market timing. This is one of the main advantages of using SHAP with a non-linear model: it reveals how the same variable may contribute differently across market regimes.

*4.8. Interpretability Limits and Economic Validation*

Although SHAP provides critical insights into model logic, it is important to note that these explanations are post-hoc additive feature attributions and do not inherently imply direct causality. A positive or negative SHAP value indicates how a variable contributes to the model's prediction under the observed data structure, but it does not prove that the same variable independently causes a price increase or decrease. This limitation is especially relevant in electricity markets, where generation, demand, fuel prices, bids, and exchange rates may interact simultaneously.

To reduce the risk of overinterpreting spurious correlations, the SHAP results in this study are evaluated against established economic mechanisms. The observed price-suppressing role of solar and wind generation is consistent with the Merit Order Effect, while the positive contribution of natural gas prices and exchange rates is consistent with the cost structure of an import-dependent electricity system.

Similarly, the positive association between dam hydro generation and MCP is interpreted through strategic dispatch behavior rather than as a direct causal effect. Therefore, the contribution of SHAP in this study is not to provide causal inference, but to examine whether the forecasting model learns relationships that are economically plausible and useful for decision support.

This distinction strengthens the reliability of the interpretation. The consistent alignment of the main SHAP attributions with electricity market theory suggests that the LightGBM model is not relying only on arbitrary statistical patterns. Instead, it appears to capture several established economic drivers of price formation. Nevertheless, future research may extend this framework by combining SHAP with causal inference methods, regime-switching models, or counterfactual simulations to separate stable economic mechanisms from temporary correlations more explicitly.

## 5. Conclusion

### *5.1. Implications for Decision Support*

This study presents a comprehensive framework for transparent EPF. By integrating LightGBM with SHAP, we provide a tool that not only forecasts prices with high accuracy but also diagnoses the market conditions driving those prices. For traders, this distinction enables them to differentiate between a price spike caused by a temporary windfall or renewable generation shortage versus one caused by currency devaluation or fuel-cost pressure, allowing for more nuanced hedging strategies.

A concrete application can be illustrated from the perspective of an energy trader operating in the day-ahead and intraday markets. Suppose that the model predicts a price spike for 14:00 on the following day. A conventional forecasting model would only provide the expected price level, leaving the trader uncertain about the underlying reason for the increase. In contrast, the proposed framework allows the trader to decompose the forecast through SHAP values. If the explanation indicates that the predicted spike is mainly driven by a decline in wind generation rather than an increase in natural gas prices, the trader can interpret the expected volatility as weather-dependent and short-term. In this case, intraday hedging or short-term position adjustment may be more appropriate. However, if the same price increase is mainly explained by natural gas prices or the daily exchange rate, the trader may interpret the signal as a structural cost-pressure event and adopt a more cautious hedging strategy.

The framework is also useful for policymakers and market monitors. Repeated explanations showing that high prices are driven by imported fuel costs and exchange rate movements may indicate vulnerability to external cost shocks. Similarly, if price increases are frequently associated with low renewable output or limited flexible generation, the results may support policy discussions on storage, demand response, reserve capacity, or diversification of the generation mix. Therefore, the proposed approach not only improves forecasting accuracy; it also translates model outputs into interpretable market signals that can support operational, financial, and regulatory decision-making.

### *5.2. Limitations and Structural Breaks*

A limitation of this study, and ML models in general, is the handling of regulatory structural breaks. The analysis period includes intervals of regulatory price caps, namely maximum price limits, imposed to shield consumers from volatility. While tree-based models can adapt to these distinct regimes more effectively than linear models, they cannot accurately predict the onset of a regulatory intervention. Future work will integrate advanced algorithms to adjust the forecasting horizon when structural breaks occur dynamically.

Another limitation concerns the interpretation of post-hoc explanations. Although SHAP improves transparency by showing how each variable contributes to a model prediction, it does not provide causal identification. Therefore, the explanations in this study should be interpreted as model-based associations that are evaluated in light of electricity market theory, rather than as direct causal effects. Future research may extend the proposed framework by combining explainable machine learning with causal inference methods, regime-switching models, probabilistic forecasting, or rolling-window validation to distinguish stable economic mechanisms from temporary correlations better.

### *5.3. Conclusion*

The findings show that in volatile, import-dependent electricity markets, black-box accuracy alone is insufficient. By employing an explainable artificial intelligence framework, we showed that the proposed model captures fundamental economic dynamics, including the merit order effect, fuel-cost pass-through, exchange rate pressure, and strategic hydro dispatch. The results indicate that LightGBM provides strong predictive performance, while SHAP transforms the model outputs into interpretable economic signals. This combination allows the framework to serve not only as a forecasting tool but also as a practical decision-support mechanism for market participants and policymakers.

Overall, the study contributes to the electricity price forecasting literature by linking predictive accuracy with economic interpretability in a volatile emerging market context. The transition from pure forecasting to explainable decision support is essential for building trust in artificial intelligence applications within critical energy infrastructure.

**Acknowledgment**

This work has received funding from the Scientific and Technological Research Council of Türkiye (TUBITAK) ARDEB 3501 program under grant agreement 222M440.

**Conflict of Interest Statement**

The authors declare no conflict of interest.

**Declaration of generative AI and AI-assisted technologies in the writing process**

During the preparation of this work, the authors used Google Gemini to enhance the manuscript's readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

**Authors Contribution Declaration**

Authors contributed equally.

**Abbreviations**

| | |
|---|---|
| ACF | Autocorrelation |
| AI | Artificial intelligence |
| ANFIS | Adaptive-Network-based Fuzzy Inference System |
| ANN | Artificial Neural Network |
| AR | Autoregressive Model |
| ARIMA | Autoregressive Integrated Moving Average |
| ARMA | Autoregressive Moving Average |
| ARX | AutoRegressive with eXogenous |
| BI | Best Individual |
| BPM | Balancing Power Market |
| CMARS | Multivariate Adaptive Regression Splines |
| CNN | Convolutional Neural Network |
| DAM | Day-ahead Market |
| DTR | Decision Trees Regression |
| DL | Deep Learning |
| DM | Diebold-Mariano |
| EEX | European Energy Exchange |
| EMOIC | Energy Markets Operation Inc. |
| EPF | Electricity Price Forecasting |
| EXIST | Energy Exchange Istanbul |
| FNN | Feed-forward Neural Network |
| GDM | General Directorate of Meteorology |
| GRU | Gated Recurrent Units |
| IDM | Intra-Day Market |
| LightGBM | Light Gradient Boosting Machine |
| LR | Linear Regression |
| LSTM | Long Short-Term Memory |
| LSTNet | Long- and Short-Term Time-series Network |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MCP | Market Clearing Price |
| ML | Machine Learning |
| MSE | Mean Squared Error |
| MLR | Multiple Linear Regression |
| Nord Pool | Nordic Power Exchange |
| PACF | Partial Autocorrelation |
| PJM | Pennsylvania-New Jersey-Maryland Interconnection |
| PPC | Petroleum Pipeline Corporation |
| RF | Random Forest |
| RMSE | Root Mean Square Error |
| RNN | Recurrent Neural Network |
| SARIMA | Seasonally Autoregressive Integrated Moving Average |
| SETAR | Self-exciting threshold auto-regressive regression |
| SHAP | SHapley Additive exPlanations |
| SVR | Support Vector Regression |
| XAI | Explainable Artificial Intelligence |
| XGBoost | Extreme Gradient Boosting |

**Appendix A**

**Table A. 1.** MCP forecast studies in the Turkish Day-Ahead market

| Authors (Date) | Methodology | Variables Used | Period Covered | Results (Error Metric Values) |
|---|---|---|---|---|
| Şenocak and Kahveci (2016) | ANN, ANFIS | The production quantities of natural gas, hydro, and natural gas-liquid fuels-lignite energy resources | 2010- 2015 | MAPE, ANFIS |
| Demirezen and Çetin (2021) | Random Forest, Regression Support, Vector Regression | One day before MCP, a week ago MCP, a month ago MCP, day of the week (Weekdays 0; Weekends 1), holiday (holidays and special days 1; other days 0), MCP one hour ago, MCP two hours ago, MCP three hours ago, trade volume | 01.01.2019-10.03.2020 | MAE, MAPE, RMSE |
| Ertaylan et al. (2021) | Artificial Neural Network | Date, Time, One day before MCP, A week ago MCP, Total consumption amount, Total production amount, Gas reference price Holidays and Ramadan days (Independent variable) | 01.01.2019-31.12.2019 | MSE |
| Akay et al. (2021) | LSTM, SVM | System Marginal Price | 01.01.2018-30.06.2021 | MAPE |
| Arslan and Ertuğrul (2022) | Multiple Regression, Artificial Neural Networks | Energy production quantity, Demand quantity | 2020 (March, April, May, June) | MAPE |
| Karatekin and Başaran (2022) | Linear Regression, Polynomial Regression, ANN, XGBoost | Day-ahead hourly Market Clearing Price (MCP), Hourly production for each electricity generation source | 2015-2020 | RMSE |
| Sirin and Yilmaz (2020) | Quantile regression | Demand, MCP, Wind, River Hydro, Gas | May 2016 – May 2019 | least squares error rate |
| Avci et al. (2018) | Time series, Artificial neural network, Seasonality models | Forecasted demand, Margin, Day-specific dummy, Holiday dummy | 01.01.2012 – 31.12.2015 | MAPE |
| Taysi et al. (2015) | SARIMA, ANN | One day ago MCP at the same hour, One week ago MCP at the same hour, Average MCP in the last 24 hours | December 2012 – November 2014 | MAPE |
| Ugurlu et al. (2018) | Markov, Naive, SETAR, ANN, CNN, SARIMA, LSTM, GRU | 24-hour delayed price, 168-hour delayed price, 1-hour delayed price, 48-hour delayed price, 23-hour delayed price, 72-hour delayed price, 336-hour delayed price, Demand forecast instead of supply, Temperature, Demand/supply realized with a 24-hour delay, Balancing of the market price with a 24-hour delay | 01.01.2013 - 21.12.2016 | MAE |
| Ozyildirim and Beyazıt (2014) | Linear Regression, Radial basis function | MCP lags, Temperature, Square of temperature, Hourly, Daily, Monthly, Holiday dummies | 2010 - 2013 | MAE, MSE |
| Kölmek and Navruz (2015) | ANN, AR | Historical day-ahead prices, Demand forecast Average temperature, Consumption forecast, Amount of bilateral contracts, Day code | December 2009 – November 2010 | MAPE |
| Ozozen et al. (2016) | SARIMA, ANN | Day-ahead hourly Market Clearing Price (MCP), Demand/ Supply | 2014 - 2015 | MAPE, MSE |
| Arifoğlu and Kandemir (2022) | MLP, CNN, LSTM | Days of the week and public holidays, Real-time consumption, Hydroelectric power generation, System marginal price, Load forecast plan, Wind speed, Temperature | 13.06.2016 – 28.03.2021 | MAPE |

## Appendix B

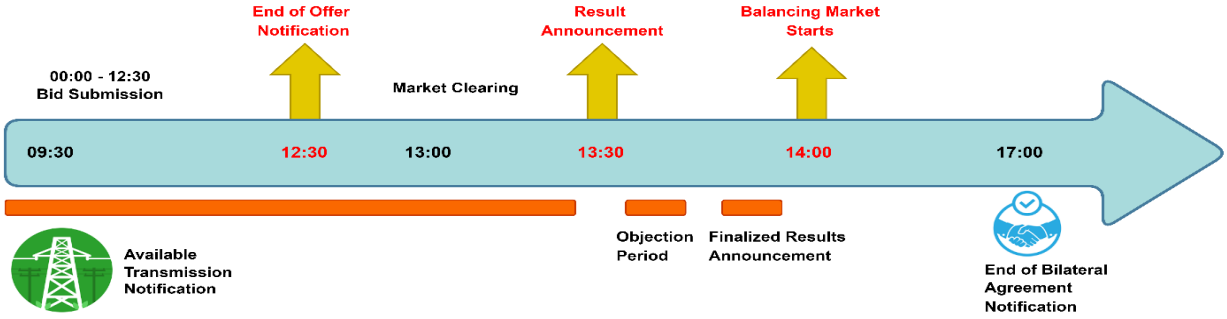


**Figure B. 1.** Turkish Day-Ahead market process

Electricity prices in the Turkish Day-Ahead Market (DAM) exhibit hourly, daily, and seasonal fluctuations (Figure B. 1). Prices tend to rise during high-demand hours and decline on weekends and public holidays when electricity consumption is lower (Figure B. 2). Seasonally, prices increase in the winter due to heating needs and in the summer due to cooling requirements. In contrast, prices generally decrease in the spring, likely as a result of the higher contribution of hydroelectric power plants. General statistics and further analyses of the Turkish DAM are presented in Figure B. 3, while the positive and negative effects of the parameters influencing the market-clearing price are illustrated in Figure B. 4.

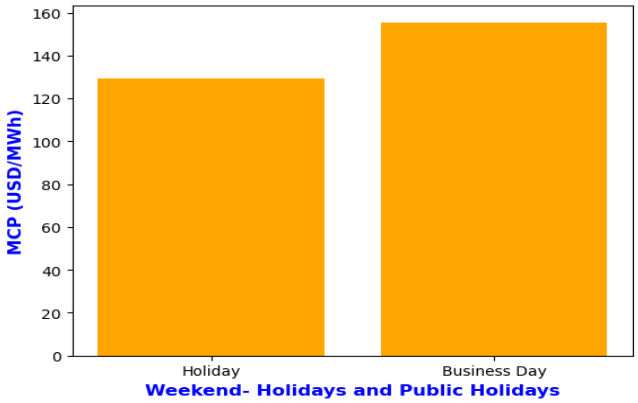


**Figure B. 2.** Average electricity prices in the Turkish Day-Ahead Market by Weekend - Holidays and public Holidays in 2022

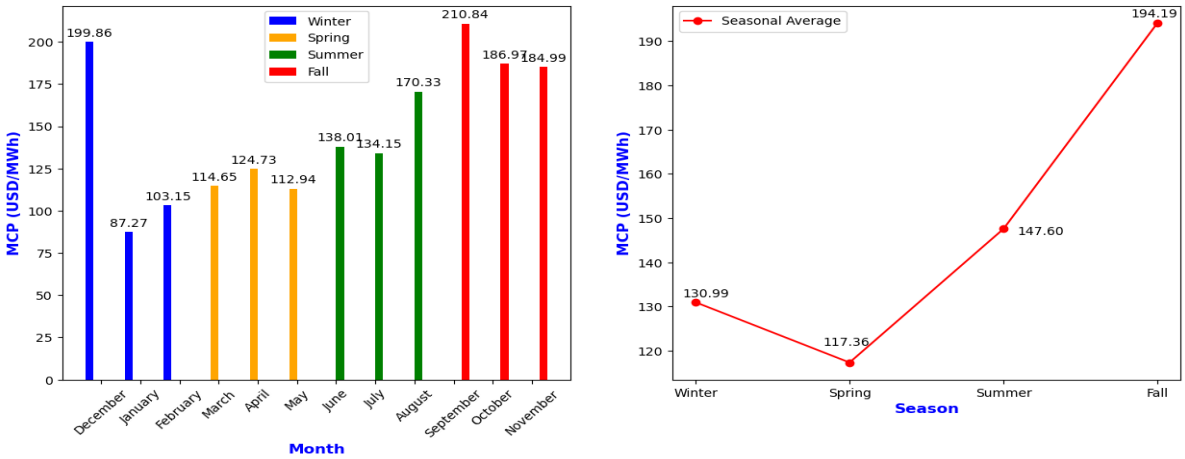


**Figure B. 3.** Average monthly and seasonal electricity prices in the Turkish Day-Ahead market in 2022

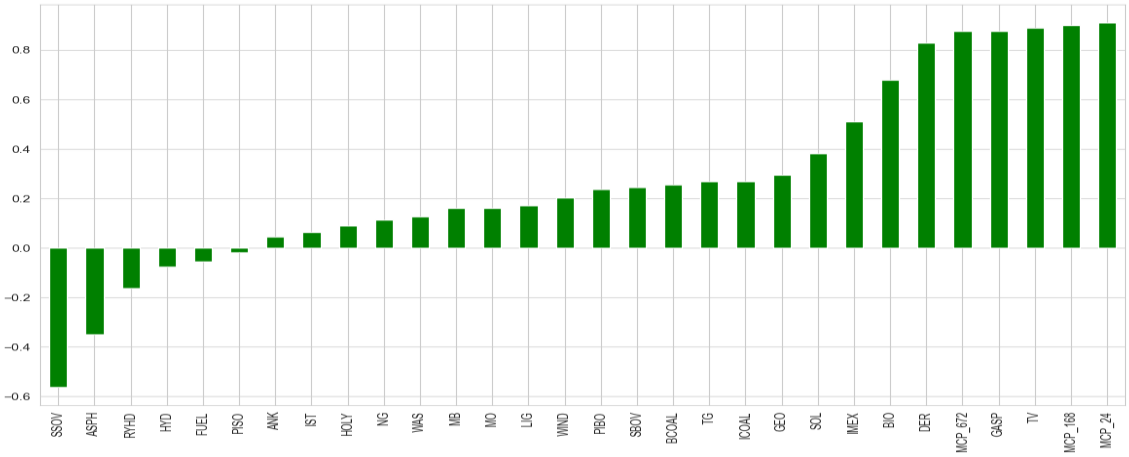


**Figure B. 4.** Independent variables affecting MCP

In the electricity market, prices can vary above or below the annual average during specific hours. As seen in Figure B. 5, the average MCP is above the annual average between 00:00-08:00, 11:00-14:00, and 14:00-22:00. However, between 01:00-07:00, 12:00-13:00, and at 23:00, the MCP is below the annual average. As depicted in Figure B. 6, on weekdays, prices are relatively consistent, while weekend prices exhibit variability. Saturday morning prices are similar to other weekdays, with a decrease observed on Saturday night and Sunday.

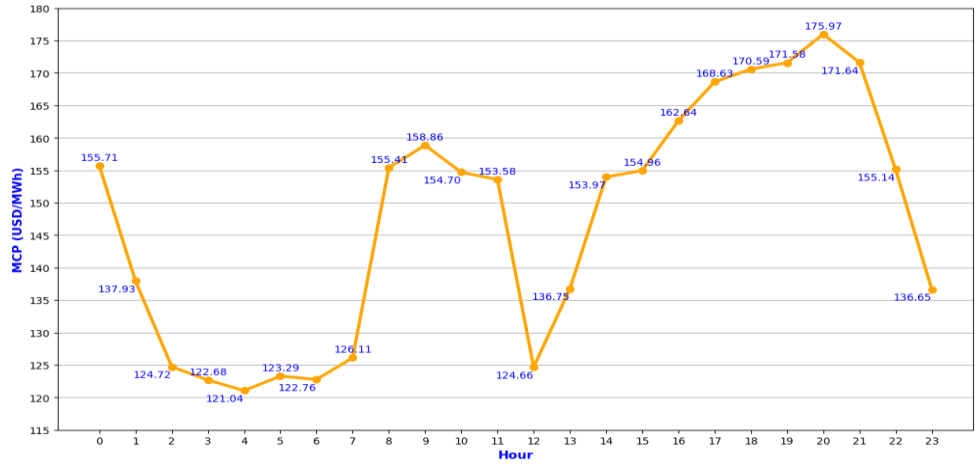


**Figure B. 5.** Average hourly electricity prices in the Turkish Day-Ahead market in 2022

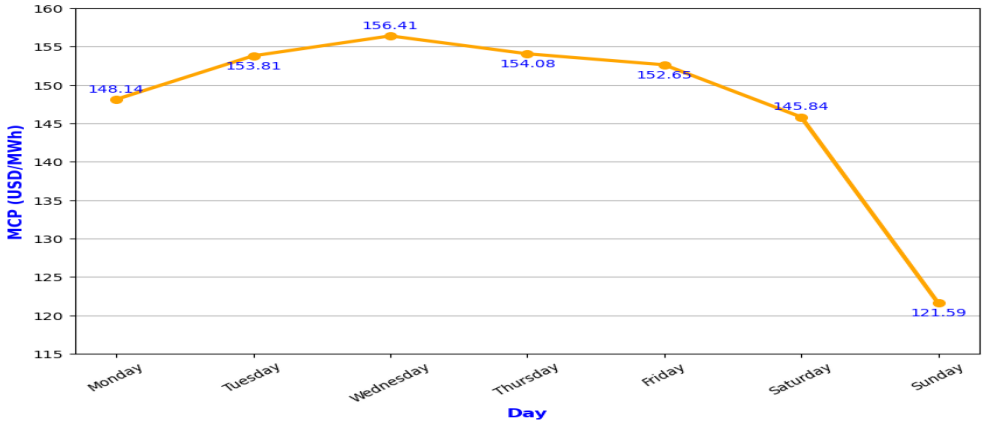


**Figure B. 6.** Average Daily Turkish Day-Ahead market electricity prices in 2022

Figure B. 7 shows time series lines depicting daily and weekly average MCP values. Daily data exhibits more fluctuations, while weekly data smoothens these fluctuations, reflecting more general trends. There is a general upward trend from 2018 to 2023, indicating an increase in prices in the day-ahead electricity market during this period.

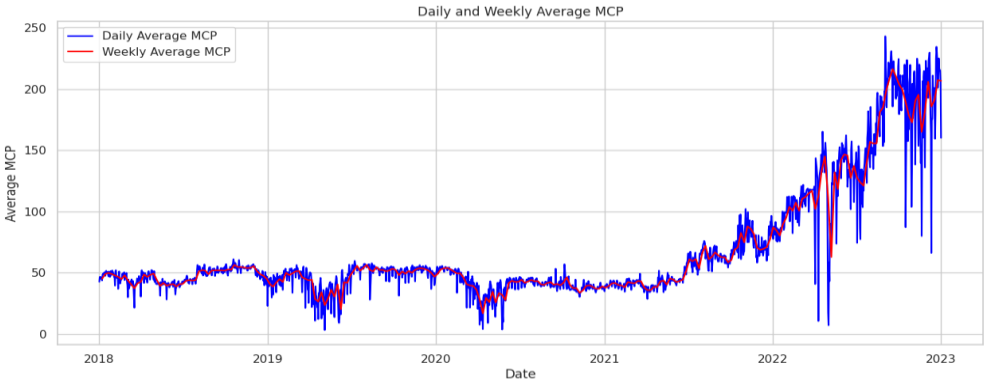


**Figure B. 7.** Average Daily and Weekly Turkish Day-Ahead Market Electricity Prices

Time series components are decomposed into three elements: trend, seasonality, and residuals (Figure B. 8). The trend component illustrates the long-term tendency in the data, showing a slowly increasing

trend over time. The seasonality component displays regular fluctuations throughout the year, even though there might not be a distinct annual pattern. The residuals component represents variations that cannot be explained by trend and seasonality. These variations are typically random and irregular.

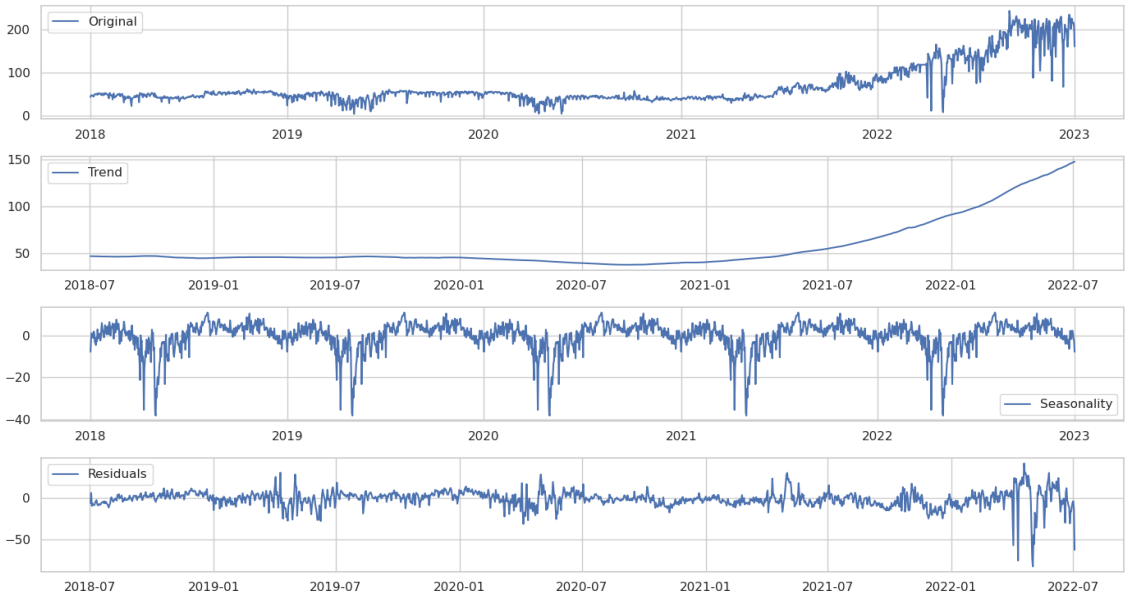


**Figure B. 8.** Original, trend, seasonality, residual

## Appendix C

**Table C. 1.** Dataset

| Feature | Explanation | Mean | Std | Min | Max |
|---|---|---|---|---|---|
| MCP (Target) | Observed Market Clearing Price (MCP) of the hour (USD/MWh) | 67.449 | 49.874 | 0.000 | 264.170 |
| HOLY | Binary value that represents whether an hour corresponds to a weekday, weekend, or holiday | 0.690 | 0.462 | 0.000 | 1.000 |
| MCP-24 | Observed MCP on the previous day of the prediction (24 hours before) (USD/MWh) | 67.380 | 49.814 | 0.000 | 264.170 |
| MCP-168 | Observed MCP one week (168 hours) before the prediction (USD/MWh) | 66.797 | 49.076 | 0.000 | 264.170 |
| MCP-672 | Observed MCP one month (672 hours) before the prediction (USD/MWh) | 65.014 | 46.667 | 0.000 | 264.170 |
| PISO | Price Independent Sales Order (Sum of sale order volumes at 0 price step) (MWh) | 11334.934 | 2856.462 | 3721.1 | 23000.6 |
| PIBO | Price Independent Bid Order (Sum of bid order volumes at price cap, 4800 TL/MWh) | 18811.540 | 4594.161 | 3721.3 | 43135.4 |
| SSOV | Submitted Sales Order Volume (Sum of sale order volumes at price cap, 4800 TL/MWh) | 29642.208 | 4767.823 | 10444.2 | 44737.7 |
| SBOV | Submitted Bid Order Volume (Sum of bid order volumes at 0 price step) (MWh) | 24578.563 | 4623.185 | 10715.2 | 47542.8 |
| MO | Matched Offers Quantity (MWh) | 19744.961 | 3776.736 | 9640.1 | 32222.5 |
| MB | Matched Bids Quantity (MWh) | 19745.167 | 3776.621 | 9640.2 | 32223.7 |
| TV | Trade Value (USD) | 1364707.72 | 1074959.50 | 0.000 | 6442106.8 |
| DER | Daily Exchange Rate (TL/USD) | 8.593 | 4.398 | 3.732 | 18.703 |
| TG | Total Generation (MWh) | 34207.754 | 5208.552 | 15333.3 | 51285.6 |
| NG | Generation from Natural Gas (MWh) | 8954.022 | 3999.599 | 496.66 | 19453.6 |
| HYD | Generation from Dam Hydro (MWh) | 5727.298 | 2921.057 | 296.690 | 15891.78 |
| LIG | Generation from Lignite Coal (MWh) | 4953.004 | 614.350 | 2189.10 | 6984.15 |

| | | | | | |
|---|---|---|---|---|---|
| RHYD | Generation from River Hydro (MWh) | 2217.594 | 1324.942 | 379.960 | 6103.59 |
| ICOAL | Generation from Import Coal (MWh) | 6913.349 | 1771.430 | 956.350 | 10138.84 |
| WIND | Generation from Wind (MWh) | 2992.833 | 1824.538 | 42.530 | 9327.74 |
| SOL | Generation from Solar (MWh) | 114.069 | 245.311 | -1.700 | 1289.82 |
| FUEL | Generation from Fuel Oil (MWh) | 84.487 | 59.370 | 0.000 | 237.60 |
| GEO | Generation from Geothermal (MWh) | 1021.936 | 177.328 | 525.110 | 1357.30 |
| ASPH | Generation from Asphaltite Coal (MWh) | 246.673 | 84.611 | 0.000 | 368.720 |
| BCOAL | Generation from Black Coal (MWh) | 403.430 | 116.198 | 77.500 | 692.620 |
| BIO | Generation from Biomass (MWh) | 523.189 | 219.352 | 202.580 | 1011.230 |
| IMEX | Import-Export Electricity Volume (MWh) | -27.012 | 315.928 | -3130.50 | 1648.540 |
| WAS | Generation from Waste Heat (MWh) | 82.855 | 13.630 | 30.690 | 127.760 |
| GASP | Natural Gas Price (USD/1000Sm$^3$) | 377.061 | 303.509 | 162.850 | 1444.510 |
| ANK | Hourly average temperature of Ankara (°C) | 13.700 | 9.299 | -13.000 | 38.600 |
| IST | Hourly average temperature of Istanbul (°C) | 16.390 | 7.577 | -2.600 | 35.900 |

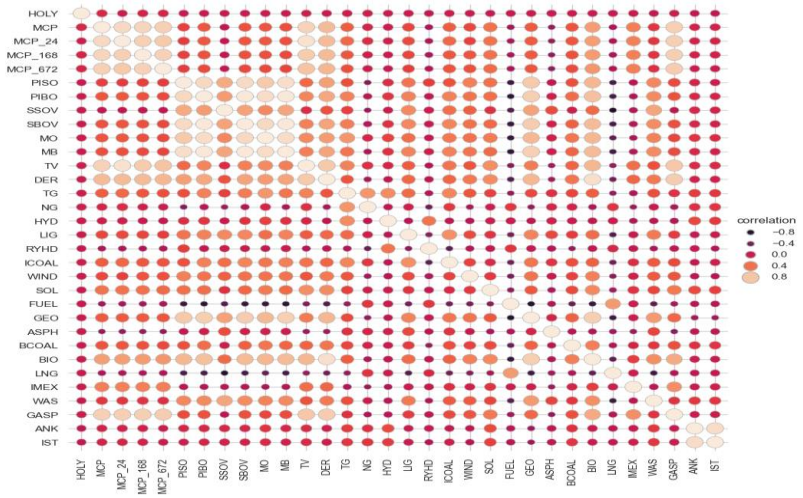


**Figure C. 1.** Correlation heat map

## Appendix D

Machine learning (ML) develops algorithms that learn from data to make accurate predictions. As illustrated in Figure D. 1, ML encompasses various subfields offering specialized techniques for diverse analytical tasks. For this study, models such as Linear Regression, Random Forest, Decision Tree, Support Vector Machine, CatBoost, XGBoost, LightGBM, and Sequential Artificial Neural Networks are utilized, with grid search and Keras Tuner applied for hyperparameter optimization. Figure D. 2 summarizes the methodology used for EPF by processing the data collected from EXIST through exploratory data analysis and visualizing the findings obtained. The analysis reveals the structural characteristics and basic statistics of the datasets, and effective prediction models are developed using various machine learning algorithms. This systematic approach ensures that our final expert system is not only built on a robust data foundation but is also methodologically sound, progressing from raw data to a validated and interpretable predictive tool.

In the model training stage, each algorithm was trained according to its own learning mechanism to ensure a fair comparison across different model families. Linear Regression was used as a classical statistical benchmark. Random Forest was trained through bootstrap aggregation, while Decision Tree was trained using recursive partitioning based on squared-error minimization. Support Vector Regression was trained with an epsilon-insensitive loss function and regularization. XGBoost, CatBoost, and LightGBM were trained as gradient boosting decision tree algorithms, in which successive trees are added to reduce the residual errors of the previous learners. LightGBM differs from many conventional tree-based implementations by adopting a leaf-wise, best-first growth strategy. Instead of expanding trees level by level, it selects the leaf with the highest potential loss reduction and splits that leaf first. This allows the model to learn irregular but influential patterns in the dataset, which is useful for electricity price forecasting because market-clearing prices may react sharply to fuel-cost changes, exchange rate movements, renewable generation fluctuations, and bidding behavior. The Sequential Artificial Neural Network was trained using the Adam optimizer with mean squared error as the loss function.

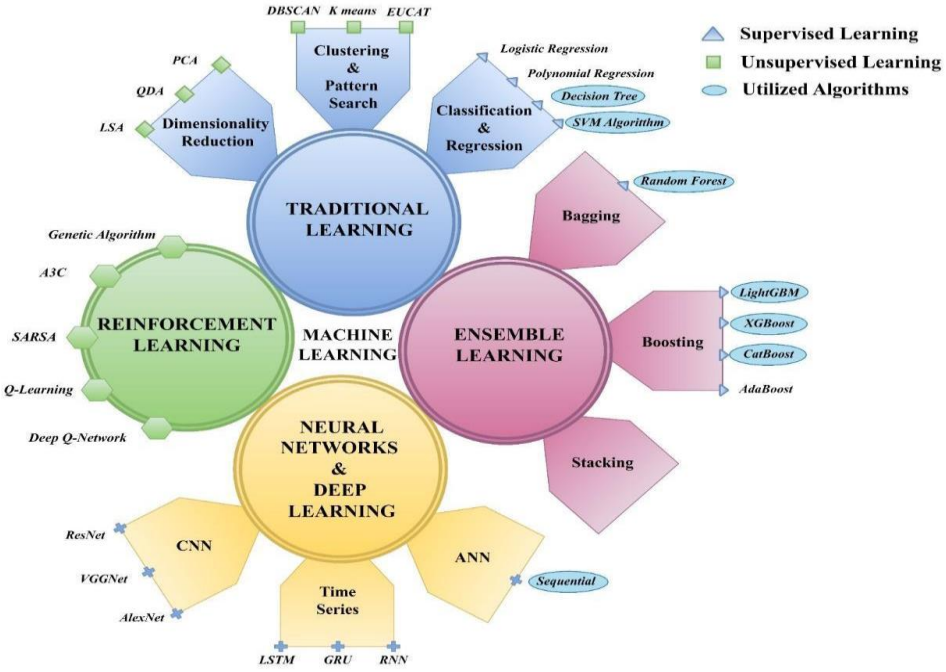


**Figure D. 1.** Machine learning models and the utilized algorithms in this study

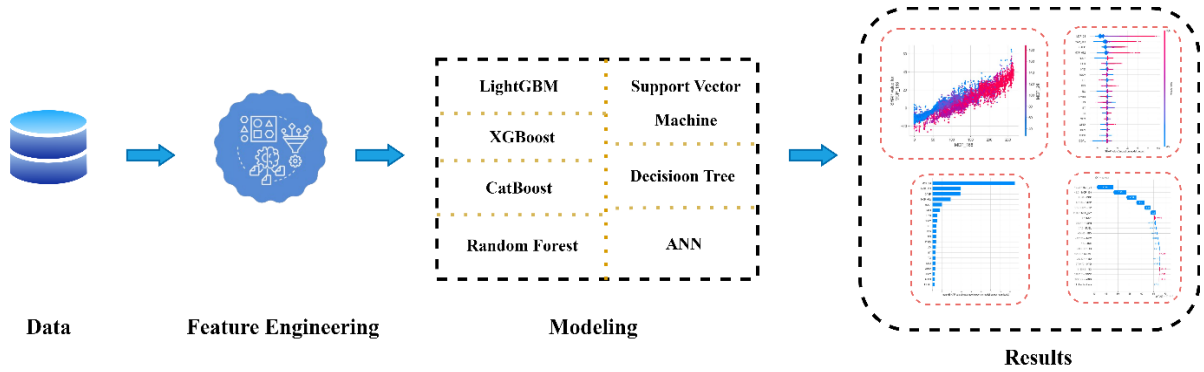


**Figure D. 2.** Methodology

**Table D. 1.** Model hyper-parameters

| | | |
|---|---|---|
| XGBoost | learning_rate: Controls the model's adaptation speed to new information | 0.05 |
| | max_depth: Determines the depth of the decision tree | 11 |
| | n_estimators: Sets the number of decision trees | 150 |
| | colsample_bytree: Specifies the ratio of randomly selected features for each decision tree | 0.5 |
| | objective: Determines the error criterion that the model is trying to optimize | reg:squarederror |
| | alpha: Controls the amount of L1 regularization (Lasso) | 10 |
| | Train MAE | %3 |
| | Test MAE | %9 |
| | Train RMSE | %4 |
| | Test RMSE | %17 |
| RF | n_estimators: Determines the number of decision trees | 300 |
| | max_depth: Specifies the depth of the decision tree | 20 |
| | max_features: Determines the number of randomly selected features for each decision tree | 20 |
| | Train MAE | %4 |
| | Test MAE | %9 |
| | Train RMSE | %7 |
| | Test RMSE | %19 |
| SVR | C: Regularization coefficient that affects the model's complexity and generalization ability | 900 |
| | Epsilon: Determines the error tolerance level | 7 |
| | Train MAE | %9 |
| | Test MAE | %10 |
| | Train RMSE | %15 |
| | Test RMSE | %17 |
| DT | max_depth: Specifies the depth of the decision tree | 10 |
| | criterion: Splitting criterion for regression based on squared error. | squared_error |
| | max_features: Determines the number of features considered at each split point | 15 |
| | Train MAE | %9 |

| | | |
|---|---|---|
| | Test MAE | %12 |
| | Train RMSE | %16 |
| | Test RMSE | %24 |
| CatBoost | iterations: Number of training iterations | 500 |
| | depth: Specifies the depth of the decision tree | 7 |
| | learning_rate: Controls the speed of the model's adaptation to new information | 0.3 |
| | loss_function: Determines the error criterion that the model is trying to optimize | RMSE |
| | silent: Suppresses information messages during training | True |
| | Train MAE | %6 |
| | Test MAE | %9 |
| | Train RMSE | %9 |
| | Test RMSE | %17 |
| LightGBM | learning_rate: Controls the speed of the model's adaptation to new information | 0.2 |
| | num_leaves: Specifies the number of leaf nodes in the decision tree | 60 |
| | n_estimators: Determines the number of decision trees | 350 |
| | Train MAE | %3 |
| | Test MAE | %8 |
| | Train RMSE | %4 |
| | Test RMSE | %16 |
| Sequential | model = Sequential()<br>model.add(Dense(64, activation = 'selu',<br>input_dim = X_train.shape[1]))<br>model.add(Dense(2048, activation = 'selu'))<br>model.add(Dense(1024, activation = 'selu'))<br>model.add(Dense(512, activation = 'relu'))<br>model.add(Dense(256, activation = 'selu'))<br>model.add(Dense(128, activation = 'relu'))<br>model.add(Dense(64, activation = 'selu'))<br>model.add(Dense(32, activation = 'relu'))<br>model.add(Dense(16, activation = 'selu'))<br>model.add(Dense(8, activation = 'relu'))<br>model.add(Dense(1))<br>opt = Adam(learning_rate = 0.001)<br>model.compile(optimizer=opt, loss = 'MSE') | Dense: Represents hidden layers, specifying the number of neurons and activation function<br>Adam: Optimization algorithm that controls the learning rate<br>MSE: Error function that measures the model's prediction errors. |
| | Train MAE | %7 |
| | Test MAE | %9 |
| | Train RMSE | %12 |
| | Test RMSE | %18 |

**Appendix E**

To rigorously validate the performance and reliability of our proposed forecasting system for this high-stakes application, a comprehensive suite of evaluation metrics was employed. Five error metrics are used to evaluate model performance. These are R2, MSE, MAE, RMSE, and MAPE. R-squared (R2) is the square of the correlation coefficient, measuring the proportion of the dependent variable's variation explained by the independent variables in the model. It indicates the strength of the relationship between independent and dependent variables on a scale from 0 to 1, where a value of 1 represents perfect correlation. R2 does not consider the issue of overfitting. It is calculated as follows, where SSResidual is the sum of squared errors, and SSTotal is the sum of mean differences:

$$R^2 = 1 - \frac{SS_{Residual}}{SS_{Total}}$$

Mean Squared Error (MSE) measures how close a regression curve is to a set of points. It is always a positive value, and a lower MSE value indicates better performance of the model. MSE is calculated as follows, where $y_i$ represents the actual values, $\hat{y}_i$ represents the predicted values, and N is the number of samples:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_{i-} \hat{y}_i)^2$$

Mean Absolute Error (MAE) is a measure of the difference between two continuous variables. MAE, for each example in the dataset, is the average of the absolute differences between the actual value and the predicted value. It is commonly used in regression and time series problems and is calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_{i-} \hat{y}_i|$$

Root Mean Square Error (RMSE) is a commonly used metric in machine learning to measure the distance between the values predicted by a model and the actual values. It is a second-degree metric that quantifies the magnitude of the error. RMSE represents the standard deviation of prediction errors (residuals). Residuals are a measure of how far the regression line is from the data points. RMSE indicates how spread out these residuals are around the line that best fits the data. It informs about the density of the data points around the line that fits them the best. The RMSE value can range from zero to infinity, with lower values indicating better performance of the predictor. An RMSE value of zero

implies that the model makes no errors. RMSE is the square root of Mean Squared Error (MSE), and it is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(y_{j-} \hat{y}_j)^2}{n}}$$

Mean Absolute Percentage Error (MAPE) is an error metric that utilizes absolute values to prevent positive and negative errors from canceling each other out. It is commonly used in time series models to compare prediction accuracy. MAPE values below 10% indicate high accuracy in prediction models, while models with MAPE values between 10% and 20% can still be classified as accurate. MAPE is calculated as follows:

$$MAPE = \frac{\sum_{j=1}^{N} \frac{|y_j|}{\hat{y}_j}}{N}$$

These are common evaluation metrics used in ML to assess the performance of models, particularly in the context of regression problems. Each metric serves a different purpose, and the choice of which one to use depends on the specific characteristics of the problem and the emphasis of the evaluation. R2 is used to understand the goodness of the fit. MSE and RMSE are used to measure error magnitude, with RMSE providing an error measure in the same units as the target variable. MAE is used when a simple and interpretable error measure that is less sensitive to outliers is desired. MAPE is used when an assessment prediction accuracy in percentage terms is needed, often in forecasting scenarios. We applied all the above measures so that comparisons with other studies could be available when needed. This multi-faceted evaluation ensures a holistic understanding of the model's strengths and weaknesses, a critical requirement for any deployable decision-support tool intended for a real-world production environment.

**Appendix F**

The Diebold-Mariano (DM) test serves to determine whether the difference in predictive accuracy between two competing forecasting models is statistically significant. The mathematical procedure is outlined as follows:

- **Loss Differential:** Let $e_{1,t}$ and $e_{2,t}$ denote the forecast errors of the benchmark model and the competing model at time t, respectively. The loss differential series, dt, is defined based on a loss function g(.) (typically squared error or absolute error):

$$d_t : g(e_{1,t}) - g(e_{2,t})$$

- **Null Hypothesis:** The test evaluates the null hypothesis ($H_0$), which posits that the expected loss differential is zero, implying equal predictive accuracy:

$$H_0 \ : \ E[d_t] = 0$$

- **DM Statistic:** The Diebold-Mariano test statistic is calculated as the ratio of the sample mean of the loss differential ($\bar{d}$**)** to its standard error:

$$DM = \frac{\bar{d}}{\sqrt{\widehat{VAR}\ (\bar{d})}}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$ and $\widehat{VAR}\ (\bar{d})$ represents the consistent estimate of the asymptotic variance of $\bar{d}$.

- **Decision Rule:** Under the null hypothesis, the DM statistic asymptotically follows a standard normal distribution. The null hypothesis is rejected if the absolute value of the DM statistic exceeds the relevant critical value, such as 1.96 at the 5% significance level.

## Appendix G

**Table G. 1.** Run times of gridSearch and SHAP analysis

| Models | SHAP uptime | GridSearch uptime |
|---|---|---|
| **LightGBM** | 50 seconds | 2 seconds |
| **CatBoost** | 7 seconds | 4 seconds |
| **XGBoost** | 5 minutes | 10 seconds |
| **Sequential** | 6 minutes | 20 minutes |
| **SVR** | 3154 minutes | 13 minutes |
| **RF** | 10 seconds | 10 minutes |
| **DT** | 10 seconds | 3 seconds |

## Appendix H

The error rates of all prediction models based on the $R^2$, MSE, MAE, RMSE, and MAPE metrics are provided in Figure H.1. In addition to the seven machine learning models, Linear Regression (LR) was included as a classical statistical benchmark, resulting in a total of eight evaluated models. The inclusion of LR provides a transparent baseline and enables a clearer comparison between classical statistical modelling and more flexible machine learning approaches. To ensure consistency between the numerical results and graphical diagnostics, prediction error plots are presented for all evaluated models, including LR and LightGBM.

Between 2018 and 2022, the LR and DT models produced weaker predictive results compared to the advanced machine learning models. Although LR provides an interpretable and useful benchmark structure, its relatively higher MAE, MSE, RMSE, and MAPE values indicate that a purely linear model is limited in capturing the non-linear and volatile price movements observed in the Turkish Day-Ahead Market. Similarly, the weaker generalization performance of the DT model suggests that a single-tree

structure is insufficient for representing complex interactions among fuel prices, exchange rates, generation mix, market bid volumes, and lagged price variables.

Although CatBoost and XGBoost demonstrate competitive forecasting performance, the error rates and graphical analyses in Figure H.1 indicate that LightGBM achieves the strongest overall predictive performance. However, since the Diebold–Mariano test shows no statistically significant difference between LightGBM and CatBoost, LightGBM is not selected solely on the basis of numerical superiority. Rather, it is selected because it combines strong predictive performance with computational efficiency and compatibility with SHAP-based explainability analysis. Therefore, based on this comprehensive, quantitative, and interpretability-oriented assessment, LightGBM is selected as the core predictive component of the proposed decision-support framework.

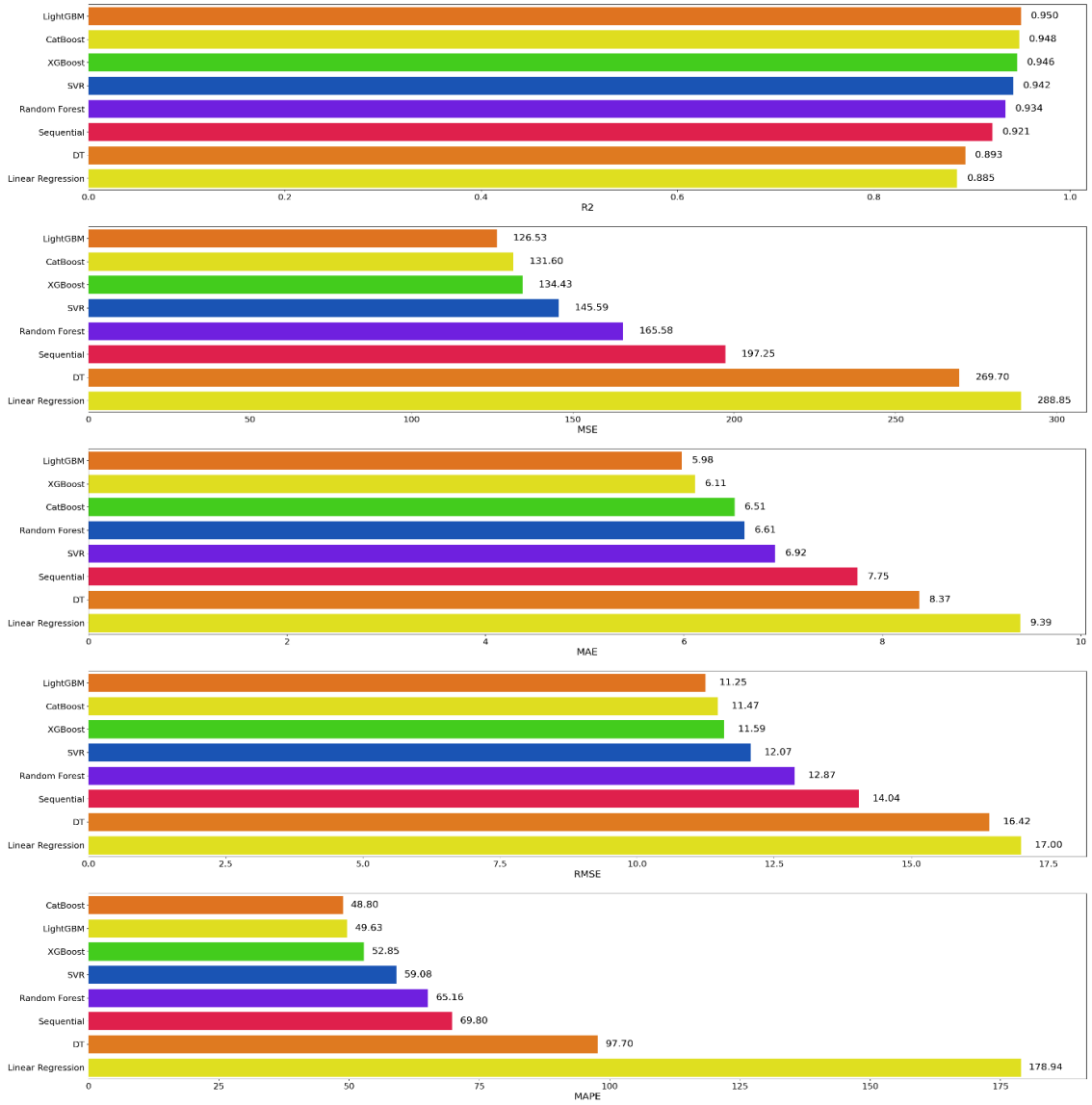


**Figure H. 1.** Result of the models

Machine learning and statistical models can be used to solve various forecasting problems; however, their performance may differ depending on the data structure, the forecasting task, and the complexity of the model. Therefore, during model selection and tuning, it is essential to evaluate how well each model performs under the same experimental framework. In this study, several metrics and visualization techniques are used to assess forecasting performance. Among these visual tools, prediction error plots are particularly useful because they show the relationship between actual values and model predictions. In this study, the prediction performance of eight different models is compared: LR, XGBoost, RF, SVR, Sequential ANN, DT, CatBoost, and LightGBM. Each model was trained on the same dataset and evaluated under the same performance framework. The purpose of this analysis is to visualize the prediction errors of the models and to identify which model provides the most reliable forecasting performance. For this purpose, the relationships between actual and predicted values are presented in Figure H.2 to Figure H.9. Actual values are shown on the x-axis, while predicted values are shown on the y-axis. Green dots represent the predictions generated by each model for individual observations. The red line represents the ideal prediction line, where predicted values are exactly equal to actual values. The black line represents the fitted trend between actual and predicted values. The closer the black line is to the red line, the stronger the predictive performance of the model.

The LR and DT models demonstrate weaker predictive performance compared to the gradient boosting models, confirming the limitations of simpler linear and single-tree structures in modelling volatile electricity prices. SVR and the Sequential ANN show moderate forecasting performance, while XGBoost, CatBoost, and LightGBM provide stronger predictive results. Among these models, LightGBM stands out as the most suitable core model for the proposed framework because it provides a strong balance between numerical accuracy, computational efficiency, and suitability for SHAP-based explainability analysis.

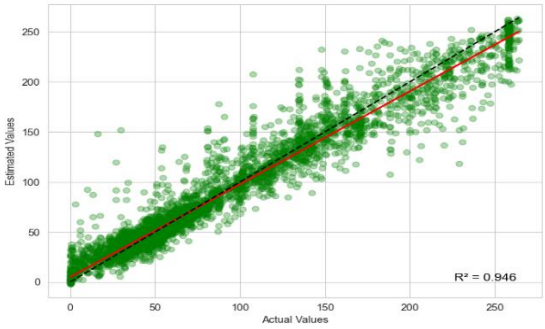


**Figure H. 2.** Prediction error for XGBoost

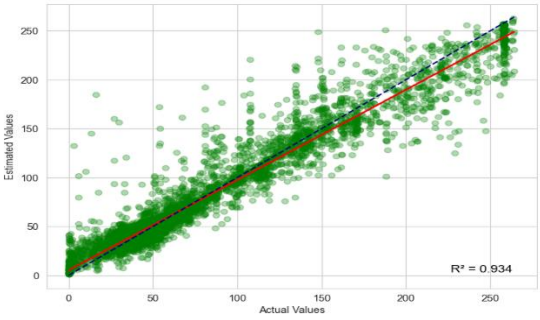


**Figure H. 3.** Prediction error for RF

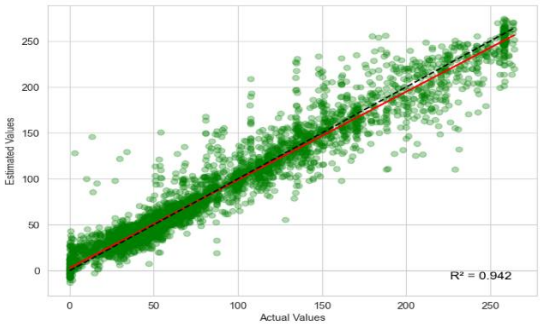


**Figure H. 4.** Prediction error for SVR

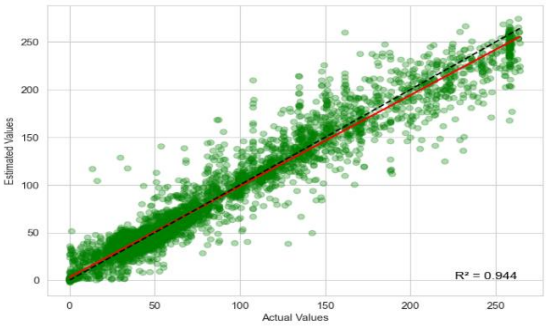


**Figure H. 5.** Prediction error for Sequential

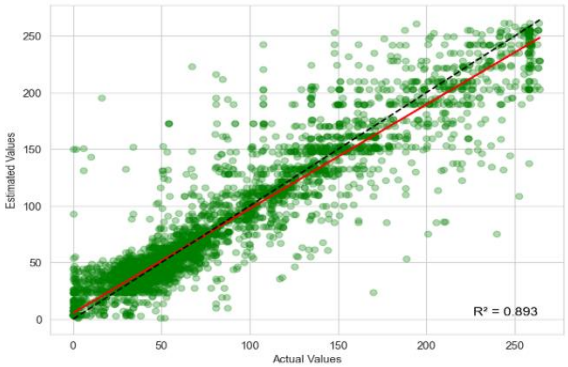


**Figure H. 6.** Prediction error for DT

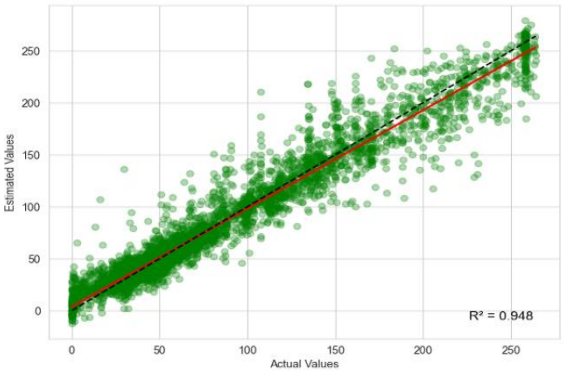


**Figure H. 7.** Prediction error for CatBoost

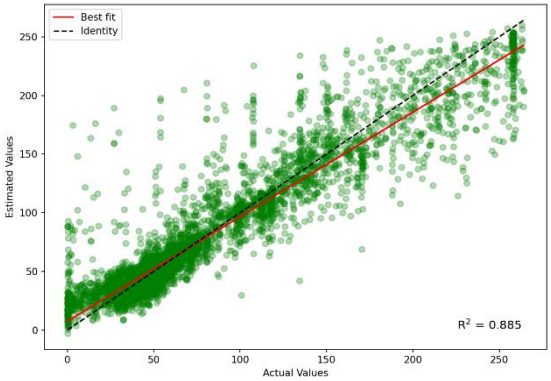


**Figure H. 8.** Prediction error for LR

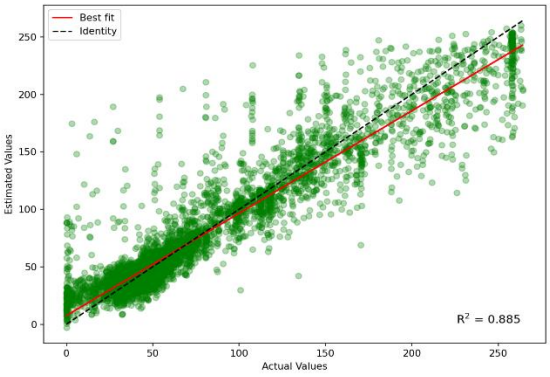


**Figure H. 9.** Prediction error for LightGBM

## Appendix I

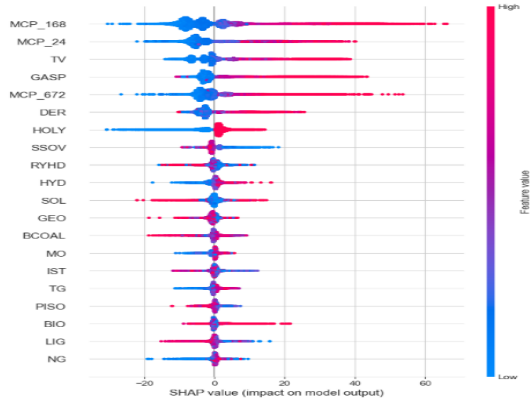


**Figure I. 1.** CatBoost SHAP value

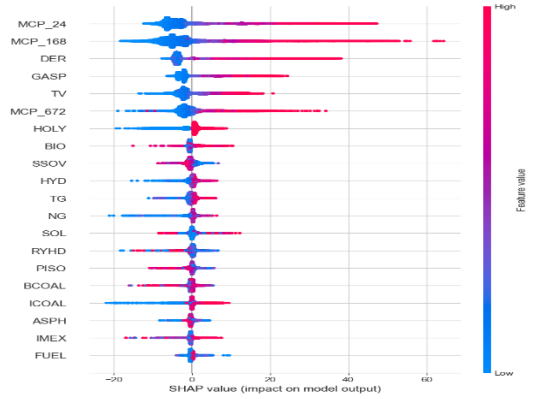


**Figure I. 2.** XGBoost SHAP value

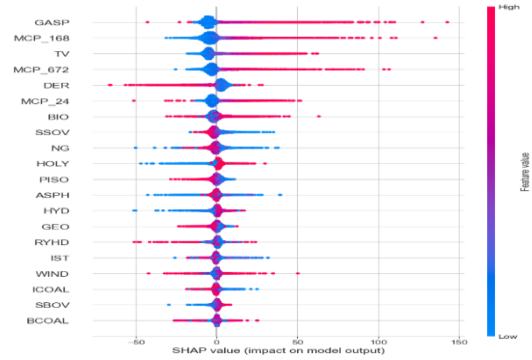


**Figure I. 3.** Sequential SHAP value

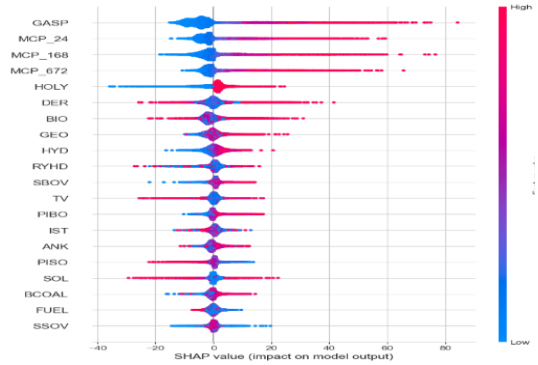


**Figure I. 4.** SVR SHAP value

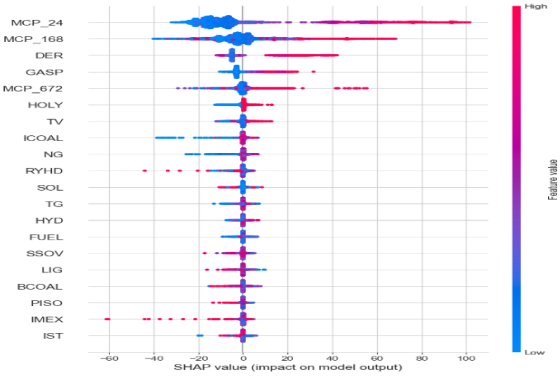


**Figure I. 5.** RF SHAP value

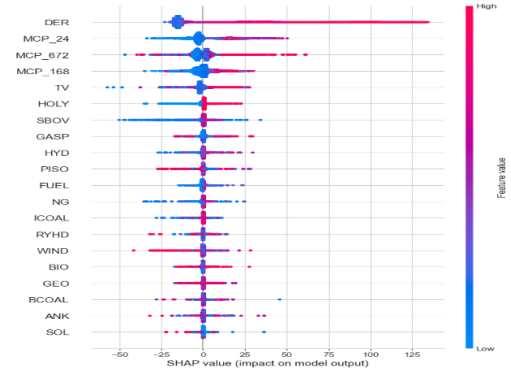


**Figure I. 6.** DT SHAP value

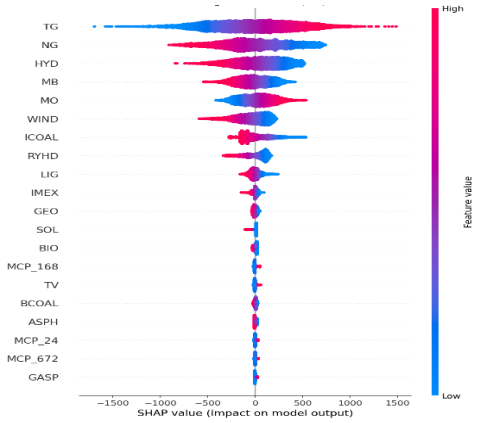


**Figure I. 7.** LR SHAP value

**References**


Akay MF., Abut F., Bedir B., Erdoğan M., Kanar E. Forecasting the system marginal price using long short-term memory and support vector machine. 6th International Mediterranean Science and Engineering Congress (IMSEC 2021), 25-27 October 2021, Sayfa no: 628-33, Alanya, Antalya, Türkiye.

Arifoğlu A., Kandemir T. Electricity price forecasting in Turkish day-ahead market via deep learning techniques. Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi 2022; 9(2): 1433-1458.

Arslan B., Ertuğrul İ. Çoklu regresyon, ARIMA ve yapay sinir ağı yöntemleri ile Türkiye elektrik piyasasinda fiyat tahmin ve analizi. Journal of Management Economics Research 2022; 20(1): 331-353.

Avci E., Ketter W., Van HE. Managing electricity price modeling risk via ensemble forecasting. The Case of Turkey Energy Policy 2018; 123: 390-403.

Backe S., Riemer-Sørensen S., Bordvik DA., Tiwari S., Andresen C. Predictions of prices and volumes in the nordic balancing markets for electricity. 19th International Conference on the European Energy Market (EEM), 6–8 June 2023, Sayfa no: 1–6, Lappeenranta, Finland.

Brownlee J. How to develop a light gradient boosted machine (lightgbm) ensemble. https://machinelearningmastery.com/light-gradient-boosted-machine-lightgbm-ensemble/

Brownlee J. XGBoost with python: Gradient boosted trees with XGBoost and scikit-learn. San Juan, Puerto Rico: Machine Learning Mastery; 2016.

Chang Z., Zhang Y., Chen W. Electricity price prediction based on hybrid model of adam optimized lstm neural network and wavelet transform. Energy 2019; 187: Article no. 115804.

Che J., Wang J. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. Energy Conversion and Management 2010; 51(10): 1911-1917.

Conejo A., Plazas MA., Espínola R., Molina A. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Transactions on Power Systems 2005; 20(2): 1035-1042.

Contreras J., Espínola R., Nogales F., Conejo A. ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems 2003; 18(3): 1014-1020.

Demirezen S., Çetin M. Rassal orman regresyonu ve destek vektör regresyonu ile piyasa takas fiyatının tahmini. Nicel Bilimler Dergisi 2021; 3(1): 1-15.

Díaz J., Romero Á., Dorronsoro JR. Day-ahead price forecasting for the Spanish electricity market. International Journal of Interactive Multimedia and Artificial Intelligence 2019.

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

Fulcher BD., Little MA., Jones NS. Highly comparative time-series analysis: the empirical structure of time series and their methods. Journal of the Royal Society Interface 2013; 10(83): Article no. 20130048.

Grossi L., Nan F. Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources. Technological Forecasting and Social Change 2019; 141: 305-318.

Gürtler M., Paulsen T. Forecasting performance of time series models on electricity spot markets: A quasi-meta-analysis. International Journal of Energy Sector Management 2018; 12(1): 103-129.

Haghi HV., Tafreshi SMM. Modeling and forecasting of energy prices using non-stationary Markov models versus stationary hybrid models including a survey of all methods. IEEE Canada Electrical Power Conference (EPC 2007), 25–26 October 2007, Sayfa no: 429–434, Montreal, Quebec, Canada.

Hayfavi A., Talasli I. Stochastic multifactor modeling of spot electricity prices. Journal of Computational and Applied Mathematics 2014; 259: 434-442.

Karatekin C., Başaran T. Gün öncesi piyasasında elektrik enerjisi fiyatının veri analizi ile tahmin edilmesi. Journal of the Institute of Science and Technology 2022; 12(4): 2075-2084.

Kölmek MA., Navruz İ. Forecasting the day-ahead price in electricity balancing and settlement market of Turkey by using artificial neural networks. Turkish Journal of Electrical Engineering Computer Sciences 2015; 23(3): 841-852.

Lago J., De Ridder F., De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy 2018; 221: 386-405.

Lehna M., Scheller F., Herwartz H. Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. Energy Economics 2022; 106: Article no. 105742.

Liu H., Shen X., Tang X., Liu J. Day-ahead electricity price probabilistic forecasting based on SHAP feature selection and LSTNet quantile regression. Energies 2023; 16(13): Article no. 5152.

Ludwig N., Feuerriegel S., Neumann D. Putting big data analytics to work: Feature selection for forecasting electricity prices using the lasso and random forests. Journal of Decision Systems 2015; 24(1): 19-36.

Marcjasz G., Lago J., Weron R. Neural networks in day-ahead electricity price forecasting: single vs. multiple outputs. arXiv preprint arXiv:2008.08006; 2020.

Mei J., He D., Harley R., Habetler T., Qu G. A random forest method for real-time price forecasting in New York electricity market. IEEE PES General Meeting | Conference & Exposition, 27–31 July 2014, Sayfa no: 1–5, National Harbor, MD, USA.

Misiorek A., Trueck S., Weron R. Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. 2006; 10(3): 1-36.

Muğaloğlu E., Kılıç E., Keskin H., Selçuklu SB. Dynamics of electricity price volatility and its impacts on energy investments. Renewable Energy 2025; 256: Article no. 124383.

Niimura T. Forecasting techniques for deregulated electricity market prices - extended survey. IEEE PES Power Systems Conference and Exposition, 29 October–1 November 2006, Sayfa no: 51–56, Atlanta, GA, USA.

Nogales F., Contreras J., Conejo A., Espínola R. Forecasting next-day electricity prices by time series models. IEEE Power Engineering Review 2002; 17(2): 342-348.

Nowotarski J., Raviv E., Trück S., Weron R. An empirical comparison of alternative schemes for combining electricity spot price forecasts. Energy Economics 2014; 46: 395-412.

Ozozen A., Kayakutlu G., Ketterer M., Kayalica O. A combined seasonal ARIMA and ANN model for improved results in electricity spot price forecasting: case study in Turkey. Portland International Conference on Management of Engineering and Technology (PICMET), 4–8 September 2016, Sayfa no: 2681–2690, Honolulu, HI, USA.

Ozyildirim C., Beyazıt M. Forecasting and modelling of electricity prices by radial basis functions: Turkish electricity market experiment. İktisat İşletme ve Finans 2014; 29(344): 31-54.

Peng L., Liu S., Liu R., Wang L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 2018; 162: 1301-1314.

Plakas K., Karampinis I., Alefragis P., Birbas A., Birbas M., Papalexopoulos A. A predictive fuzzy logic model for forecasting electricity day-ahead market prices for scheduling industrial applications. Energies 2023; 16(10): Article no. 4085.

Prokhorenkova L., Gusev G., Vorobev A., Dorogush AV., Gulin A. Catboost: Unbiased boosting with categorical features, 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), 3–8 December 2018, Sayfa no: 6639–6649, Montréal, Canada.

Pütz S., Kruse J., Witthaut D., Hagenmeyer V., Schäfer B. Regulatory changes in German and Austrian power systems explored with explainable artificial intelligence. 14th ACM International Conference on Future Energy Systems (e-Energy 2023), 20–23 June 2023, Sayfa no: 26–31, Orlando, FL, USA.

Silva AR., Fidalgo JN., Andrade JR. Easing predictors selection in electricity price forecasting with deep learning techniques. 19th International Conference on the European Energy Market (EEM), 6–8 June 2023, Sayfa no: 1–6, Lappeenranta, Finland.

Sirin SM., Yilmaz BN. Variable renewable energy technologies in the Turkish electricity market: Quantile regression analysis of the merit-order effect. Energy Policy 2020; 144: Article no.111660.

Şenocak F., Kahveci H. Periodic price averages forecasting of mcp in day-ahead Market. National Conference on Electrical, Electronics and Biomedical Engineering (ELECO), 1–3 December 2016, Sayfa no: 664–668, Bursa, Turkey.

Taysi ZC., Biricik G., Bozkurt ÖÖ. Evaluation of price forecast systems for Turkish electric market. 23rd Signal Processing and Communications Applications Conference (SIU), 16–19 May 2015, Sayfa no: 620–623, Malatya, Turkey.

Tschora L., Pierre E., Plantevit M., Robardet C. Electricity price forecasting on the day-ahead market using machine learning. Applied Energy 2022; 313: Article no. 118752.

Ugurlu U., Oksuz I., Tas O. Electricity price forecasting using recurrent neural networks. Energies 2018; 11(5): Article no. 1255.

Weron R. Electricity Price Forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting 2014; 30(4): 1030-1081.

Xie M., Sandels C., Zhu K., Nordström L. A seasonal ARIMA model with exogenous variables for elspot electricity prices in Sweden. 10th International Conference on the European Energy Market (EEM), 27–31 May 2013, Sayfa no: 1–4, Stockholm, Sweden.

Yıldırım MH., Özmen A., Bayrak ÖT., Weber GW. Electricity price modelling for Turkey. International Conference on Operations Research (OR 2011), 30 August–2 September 2011, Sayfa no: 39–44, Zurich, Switzerland.

Yu R., Gao J., Yu M., Lu W., Xu T., Zhao M., Zhang J., Zhang R., Zhang Z. Lstm-EFG for wind power forecasting based on sequential correlation features. Future Generation Computer Systems 2019; 93: 33-42.

Zhang J., Chuntian C. Day-ahead electricity price forecasting using artificial intelligence. IEEE Canada Electric Power Conference, 6–7 October 2008, Sayfa no: 1–5, Vancouver, BC, Canada.

Ziel F., Steinert R., Husmann S. Efficient modeling and forecasting of electricity spot prices. Energy Economics 2015; 47: 98-111.