

Multi-Horizon Explainability in Energy Price Forecasting: Disentangling Market Dynamics and Renewable Intermittency in Romania's Liberalized Electricity Market

Alexandru-Victor ANDREI*

Bucharest University of Economic Studies, Romania

**Corresponding author, andrei1victor23@stud.ase.ro*

Daniel-Traian PELE

IDA Institute Digital Assets Bucharest University of Economic Studies

Institute for Economic Forecasting, Romanian Academy, Bucharest, Romania

danpele@ase.ro

Alexandru-Adrian CRAMER

Bucharest University of Economic Studies, Romania

crameralexandru12@stud.ase.ro

Abstract. *Accurate electricity price forecasting is essential for participants in liberalized electricity markets, yet the black-box nature of machine learning models limits their practical adoption. While explainable artificial intelligence methods such as SHAP (SHapley Additive exPlanations) address interpretability concerns, existing applications present global feature importance rankings that ignore how price drivers vary across forecast horizons and market conditions. This study introduces two novel contributions: a multi-horizon explainability framework analyzing feature importance across immediate (0–6 hours), short-term (6–24 hours), medium-term (1–7 days), and long-term (7–14+ days) horizons, and a market regime analysis examining how explanations shift across low-price, normal, and high-price conditions. We apply Random Forest and XGBoost models to hourly Romanian electricity market data spanning January 2018 to September 2024 (59,160 observations), with features engineered from data encompassing the 2022–2023 energy crisis and summer 2024 heat wave. The selected Random Forest model achieves an R^2 of 0.60 with a mean absolute error of 31.04 EUR/MWh on the test set. SHAP analysis reveals that short-term features contribute 41 percent of horizon-attributed predictive contribution, substantially exceeding other horizons. Market regime analysis demonstrates that price momentum becomes 4.25 times more important during high-price events compared to normal conditions ($p < 0.001$), validated through Mann–Whitney U tests. These findings provide targeted guidance for market participants and offer quantitative signals for identifying market stress periods requiring enhanced risk management.*

Keywords: electricity price forecasting, explainable artificial intelligence, SHAP, multi-horizon analysis, market regime analysis, renewable energy integration, machine learning.

Introduction

Electricity price forecasting has become increasingly critical as liberalized markets replace traditional regulated structures across Europe. In these competitive markets, accurate price predictions enable generators to optimize bidding strategies, allow retailers to manage procurement costs, and support traders in developing arbitrage strategies. The Romanian electricity market, fully liberalized in January 2021, exemplifies both the opportunities and challenges of competitive electricity trading. With substantial and growing renewable energy penetration, the market exhibits price volatility characteristic of systems balancing intermittent renewable generation with inflexible baseload nuclear and hydro resources. Day-ahead electricity prices in Romania have ranged from negative values during periods of renewable abundance to over 1,000 EUR/MWh

DOI: 10.2478/picbe-2026-0014

© 2026 A.-V. Andrei; A.-A. Cramer; A. Amza, published by De Gruyter.

This work is licensed under the Creative Commons Attribution 4.0 License.

during scarcity events, creating significant financial risks for all market participants. This price volatility has driven extensive research into forecasting methods, with machine learning models such as Random Forest, XGBoost, and neural networks demonstrating superior accuracy compared to traditional statistical approaches (Lago et al., 2021; El-Azab et al., 2024). However, these sophisticated models operate as black boxes, providing predictions without explaining which factors drove each forecast. This opacity limits practical adoption, as traders hesitate to act on predictions they cannot understand and risk managers cannot validate that models capture genuine market relationships rather than spurious correlations. The recent push toward explainable artificial intelligence (XAI) addresses this challenge, with SHAP (SHapley Additive exPlanations) emerging as the leading approach for quantifying feature contributions (Lundberg and Lee, 2017).

Despite growing interest in explainable electricity price forecasting, existing studies exhibit two significant limitations. First, they typically present global feature importance rankings that treat all predictions identically, implicitly assuming that the same factors drive prices across all time horizons. This assumption contradicts practical experience: intraday traders monitoring real-time conditions focus on fundamentally different information than long-term planners developing hedging strategies. Second, most explainability analyses ignore market regimes, applying the same interpretation framework during both normal operations and crisis periods, even though electricity markets behave fundamentally differently during scarcity events compared to typical conditions (Heistrene et al., 2023; Pesenti and O’Sullivan, 2025).

This study addresses these gaps by introducing two novel contributions. First, we develop a multi-horizon explainability framework that systematically analyzes how feature importance varies across immediate (0–6 hours), short-term (6–24 hours), medium-term (1–7 days), and long-term (7–14+ days) forecast horizons. Second, we implement a market regime analysis that examines how feature importance shifts across low-price, normal, and high-price conditions, with statistical validation using Mann–Whitney U tests. We apply these frameworks to the Romanian electricity market using six years and nine months of hourly data (January 2018 to September 2024, comprising 59,160 observations), including the 2022–2023 energy crisis and the summer 2024 heat wave. Our analysis employs Random Forest and XGBoost models combined with SHAP explainability, using a comprehensive dataset covering actual generation by source, load forecasts, temperature data, and commodity prices.

The paper is structured as follows. The next section reviews the literature on electricity price forecasting and explainable AI applications in energy markets. The methodology section describes data sources, feature engineering, model development, and the explainability frameworks. Results and discussions present the empirical findings, including model performance, global feature importance, multi-horizon analysis, and market regime comparisons. The conclusion synthesizes contributions and acknowledges limitations. Our findings demonstrate that short-term features (6–24 hours) contribute 41 percent of horizon-attributed predictive contribution and that price momentum becomes 4.25 times more important during high-price events, advancing both theoretical understanding of electricity price dynamics and practical tools for context-specific forecasting.

Literature review

Electricity price forecasting methods

Electricity price forecasting (EPF) has evolved substantially over the past two decades. Weron (2014) provided a foundational review classifying approaches into statistical, computational intelligence, and fundamental models, while Lago et al. (2021) established benchmarking best practices and demonstrated that regularized regression and deep neural networks consistently

outperformed naive baselines across multiple European markets. These reviews established that no single method dominates universally, and that performance depends critically on market characteristics, input features, and evaluation methodology.

Recent studies confirm the superiority of machine learning approaches for day-ahead forecasting. El-Azab et al. (2024) demonstrated that ensemble tree-based methods and deep learning architectures achieve state-of-the-art accuracy on real datasets. Ganczarek-Gamrot et al. (2025) examined seven European bidding zones and found that Random Forest consistently achieved the highest predictive accuracy, while also revealing that markets dominated by low-controllability renewable generation exhibit greater price volatility and reduced forecast accuracy. This finding is particularly relevant for Romania, where growing wind and solar penetration increases price variability. Nitsch et al. (2024) showed that machine learning models remain robust even under widely varying renewable energy penetration levels, with mean absolute errors below 1.40 EUR/MWh when forward-looking covariates such as residual load forecasts are incorporated.

The importance of feature engineering and training strategy has received growing attention. Bâra and Oprea (2024) examined Romanian day-ahead prices specifically, demonstrating that adjusting training periods according to price volatility and incorporating macroeconomic factors improved forecast performance by 48 percent in MAE during both stable and crisis periods. Michalakopoulos et al. (2026) proposed extremely short training windows of 45 to 60 days for day-ahead forecasting, finding that LightGBM outperformed deep learning models including LSTM in this data-scarce setting across Greek, Belgian, and Irish markets. Tschora et al. (2022) extended forecasting models by incorporating price histories from neighboring countries and used SHAP values to analyze feature contributions, showing that cross-border features significantly improve forecast quality. O'Connor et al. (2025) provided a comprehensive review across day-ahead, intraday, and balancing markets, concluding that hybrid and ensemble models perform best in day-ahead settings, while simpler statistical models remain competitive for balancing markets where data sparsity is more pronounced.

Explainable artificial intelligence in energy forecasting

Despite superior predictive accuracy, machine learning models' black-box nature presents barriers to practical adoption in electricity markets. The emergence of explainable artificial intelligence (XAI) addresses this gap, with SHAP (SHapley Additive exPlanations) becoming the dominant method due to its game-theoretic foundation and ability to provide both global and local explanations (Lundberg and Lee, 2017).

Heistrene et al. (2023) proposed a trust algorithm for EPF users based on XAI techniques, generating trust scores that reflect prediction quality for each new input. Their SHAP-based approach achieved over 85 percent accuracy in identifying reliable predictions under stable conditions and showed similar accuracy in flagging unreliable predictions during distribution shifts, validating the practical value of explainability for market participants. Pesenti and O'Sullivan (2025) applied SHAP and gradient-based explanations to deep neural network models across five electricity markets, introducing the novel concepts of SSHAP values and SSHAP lines to enhance the analysis of high-dimensional tabular models. Cramer et al. (2023) combined normalizing flows with XAI for German intraday price forecasting, finding that immediate price history and day-ahead price increments had the most substantial impact on price differences. Melgar-García and Troncoso (2024) developed a real-time explainable ensemble learning algorithm for the Spanish market that combines long and short-term behavior patterns through

incremental online learning, demonstrating that explainability and real-time prediction can coexist without sacrificing accuracy.

These studies demonstrate SHAP's versatility across different markets and model architectures. However, they share a common limitation: they present global feature importance rankings that aggregate explanations across all predictions, implicitly assuming that the same factors drive prices regardless of the forecast horizon or prevailing market conditions.

Research gap and contribution

The literature review reveals two specific gaps. First, while Heistrene et al. (2023), Pesenti and O'Sullivan (2025), and Cramer et al. (2023) provide valuable SHAP analyses, all three present global feature importance rankings without differentiating across forecast horizons. Second, although Borges et al. (2024) acknowledge that electricity markets behave differently during crises, none of the reviewed studies systematically compare SHAP-based feature importance across distinct price regimes with statistical validation. The Romanian electricity market, with its recent full liberalization in 2021, growing renewable integration, and exposure to the 2022–2023 energy crisis, provides an ideal setting for investigating these context-dependent dynamics. Our study fills these gaps by introducing a multi-horizon explainability framework and a statistically validated market regime analysis, advancing both theoretical understanding and practical tools for context-specific forecasting.

Methodology

This section presents the data, feature engineering, model development, and explainability frameworks used in this study. This study shares its dataset and preprocessing pipeline with a companion paper (Andrei et al., 2026) that applies counterfactual explanations to the same Romanian electricity market data. The two studies address complementary aspects of explainability: the present paper examines feature attribution across horizons and regimes via SHAP, while the companion paper explores individual prediction explanations through counterfactual analysis.

Data description

Our analysis uses electricity market data from Romania spanning January 1, 2018 to September 30, 2024, providing six years and nine months of hourly observations (59,160 data points). Data were collected from four sources: (i) the ENTSO-E Transparency Platform, which provided actual generation by source (gas, hydro run-of-river, hydro reservoir, nuclear, solar, wind), actual total load, and official day-ahead forecasts for load and renewable generation; (ii) the European Wholesale Electricity Price Data repository for cleaned day-ahead prices; (iii) the Open-Meteo API for temperature measurements averaged across multiple Romanian locations; and (iv) Refinitiv for daily commodity prices (Brent crude oil, TTF natural gas, EU ETS CO2 emission allowances).

The target variable is the day-ahead electricity price in EUR/MWh. Over the full study period, prices averaged 101.76 EUR/MWh with a standard deviation of 103.01 EUR/MWh, ranging from −106.36 to 1,021.61 EUR/MWh. For computational efficiency in our SHAP explainability analysis, we focused on the most recent 5,000 observations (March 6 to September 30, 2024), capturing diverse conditions across spring, summer, and early autumn. Table 1 presents descriptive statistics for the target variable across dataset configurations.

Table 1. Descriptive statistics for day-ahead electricity price (EUR/MWh)

Dataset	Obs.	Mean	Std	Min	P25	Median	P75	Max
Full dataset	59,160	101.76	103.01	-106.36	41.30	66.37	120.58	1,021.61
Analysis subset	5,000	112.45	98.52	-106.36	52.89	82.13	136.62	1,021.61
Training set (80%)	4,000	107.18	85.34	-106.36	51.82	79.35	131.71	916.72
Test set (20%)	1,000	118.63	128.46	-3.95	57.68	89.40	145.21	1,021.61

Notes: Full dataset: Jan 2018–Sep 2024. Analysis subset: Mar 6–Sep 30, 2024. Training: Mar 6–Aug 20, 2024. Test: Aug 20–Sep 30, 2024. Negative prices occur when renewable supply exceeds demand.

Source: Authors' own calculations based on data from ENTSO-E, Open-Meteo, and Refinitiv.

Feature engineering

Raw variables rarely capture the full complexity of electricity price dynamics. We engineered 70 features from the original variables, organized into eleven categories: price history (14 features, including lagged prices at $t - 1$, $t - 7$, $t - 14$, moving averages, volatility measures, and price deviations), forecast features (10 features, including day-ahead load, solar, and wind forecasts, plus derived indicators such as residual load and renewable share), lagged forecasts (6 features), generation history (7 features capturing yesterday's actual output by source), commodity prices (3 features), time features (18 features with both categorical and cyclical sine/cosine encodings of hour, day, week, month, and season), interaction features (7 features such as load- temperature and peak-load interactions), balance and tension features (3 features including load capacity ratio and reserve margin), forecast error features (4 features measuring yesterday's forecast accuracy for solar and wind), rolling statistics (5 features capturing 7-day trends in forecast errors), and domain-specific features (6 features such as hydro reservoir share and renewable change rates). All features were engineered on the complete 59,160-observation dataset before subsetting, ensuring lagged and rolling calculations used the full historical context. Critically, only information available at forecast time was used, avoiding data leakage.

Model development

Model selection and training

We compared two ensemble tree-based models. Random Forest constructs B independent decision trees, each trained on a bootstrapped sample with a random feature subset, and averages their predictions. For a new observation x , the Random Forest prediction is:

$$\widehat{y}_{\text{RF}}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x), \quad (1)$$

where $T_b(x)$ is the prediction of the b -th tree and B is the total number of trees

XGBoost (eXtreme Gradient Boosting) builds trees sequentially, with each new tree correcting errors from previous iterations. The model minimizes an objective function combining a loss term and a regularization term:

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (2)$$

where $l(y_i, \hat{y}_i)$ is the loss function measuring prediction error for observation i , f_k is the k -th tree, and $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ is the regularization term penalizing tree complexity through the number of leaves T and leaf weight magnitude w , with γ and λ as regularization parameters.

For Random Forest, we used $B = 100$ trees with maximum depth of 10 levels. For XGBoost, we used 100 boosting rounds with maximum depth of 6, learning rate of 0.1, and L2 regularization. The 5,000-observation analysis subset was split chronologically (no shuffling) into a training set of 4,000 observations (March 6 to August 20, 2024) and a test set of 1,000 observations (August 20 to September 30, 2024). Additionally, five-fold time-series cross-validation with a growing window approach was used for robust performance estimation.

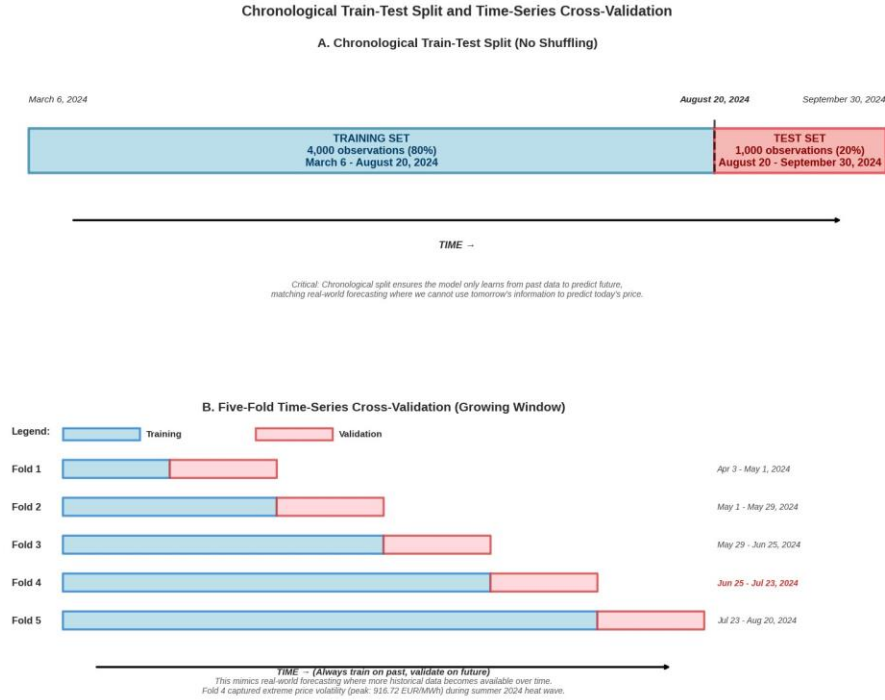


Figure 1. Chronological train-test split and five-fold time-series cross-validation

Source: Authors' own visualization.

Performance metrics

Model accuracy was evaluated using three standard metrics:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5)$$

where y_i is the actual price, $\hat{y}_i$ is the predicted price, $\bar{y}$ is the mean actual price, and n is the number of observations. R^2 ranges from 0 (no explanatory power) to 1 (perfect predictions). MAE measures the average absolute deviation in EUR/MWh. RMSE penalizes large errors more heavily. MAPE was not computed due to the presence of zero and negative prices in the test set.

SHAP explainability framework

SHAP (SHapley Additive exPlanations) decomposes each prediction into additive feature contributions based on Shapley values from cooperative game theory (Lundberg and Lee, 2017). For a prediction $f(x)$, SHAP expresses:

$$f(x) = \phi_0 + \sum_{j=1}^M \phi_j, \quad (6)$$

where ϕ_0 is the base value (mean prediction across the training set), ϕ_j is the SHAP value for feature j representing its contribution to this specific prediction, and M is the total number of features. Each ϕ_j quantifies how much feature j pushes the prediction above or below the average. We used TreeExplainer for exact SHAP value computation on tree-based models, calculated on the 1,000-observation test set.

Multi-horizon explainability

Our first contribution analyzes how feature importance varies across forecast horizons. We grouped the 70 features into four categories based on temporal distance: immediate (0–6 hours, including very recent prices and time-of-day indicators), short-term (6–24 hours, including day-ahead forecasts and today's prices), medium-term (1–7 days, including weekly lags, forecast errors, and 7-day rolling statistics), and long-term (7–14+ days, including commodity prices, seasonal patterns, and 30-day rolling statistics). We assigned 34 features with unambiguous temporal attribution to the four horizon categories, noting that price_{t-1} appears in both the immediate and short-term horizons due to its dual role. The remaining 36 features—including interaction terms (e.g., load–temperature interaction), composite indicators (e.g., gas cost proxy), generation history variables, and balance metrics—span multiple temporal scales and resist clean single-horizon classification. Rather than imposing arbitrary assignments that could distort the analysis, we restrict the horizon decomposition to features with clear temporal semantics. The percentages reported in Table 4 therefore represent the contribution shares within the horizon-assigned subset. A complete mapping of all 70 features is available in the replication repository. For each test set observation, we summed the absolute SHAP values within each horizon group, revealing each horizon's total contribution to predictive power.

Market regime analysis

Our second contribution examines how feature importance shifts across market conditions. We defined three regimes based on test set price percentiles. The low-price regime is $\leq P25 = 76.66$ EUR/MWh (250 observations). The normal regime spans $P25$ – $P75$ (500 observations). The high-price regime is $\geq P75 = 130.68$ EUR/MWh (250 observations). Regime thresholds are defined ex post using test set percentiles, making this analysis descriptive rather than operational. In a real-time deployment, practitioners would use rolling training-set quantiles to define regimes prospectively. For each regime, we computed mean absolute SHAP values per feature. To validate that importance differences are statistically significant, we applied Mann–Whitney U tests comparing SHAP value distributions between high-price and normal-price observations, using a significance threshold of $p < 0.001$.

Results and discussions

This section presents the empirical findings, organized into model performance evaluation, global feature importance, multi-horizon explainability, and market regime analysis.

Model performance

Five-fold time-series cross-validation revealed variable but consistent performance across market conditions. The average R^2 across folds was 0.51 (std = 0.18) with average MAE of 34.44 EUR/MWh (std = 24.28). Performance ranged from $R^2 = 0.28$ during a volatile period (Fold 2) to $R^2 = 0.75$ during post-crisis stabilization (Fold 5). The poorest performance occurred in Fold 4 ($R^2 = 0.41$, MAE = 77.67 EUR/MWh), which captured the summer 2024 heat wave with prices peaking at 916.72 EUR/MWh, confirming that sudden volatility spikes substantially increase forecast difficulty. Table 2 presents the final model comparison on the held-out test set. Both models showed strong training performance, with XGBoost achieving near-perfect fit ($R^2 = 0.998$). However, on unseen data, Random Forest slightly outperformed XGBoost ($R^2 = 0.60$ vs. 0.57) with a smaller train-test gap (0.35 vs. 0.43), indicating better generalization. We selected Random Forest as our final model based on these generalization characteristics. The test MAE of 31.04 EUR/MWh represents approximately 26 percent of the test set mean price (118.63 EUR/MWh), a reasonable level given the extreme volatility in the test period (std = 128.46 EUR/MWh).

Table 2. Final model performance comparison

Model	Dataset	R^2	MAE (EUR/MWh)	RMSE (EUR/MWh)	Train-Test R^2 Gap
Random Forest	Training	0.954,6	9.13	17.25	0.3539
Random Forest	Test	0.600,7	31.04	71.23	—
XGBoost	Training	0.997,9	2.78	3.73	0.4284
XGBoost	Test	0.569,5	31.57	73.96	—

Notes: Training set: 4,000 observations (Mar 6–Aug 20, 2024). Test set: 1,000 observations (Aug 20–Sep 30, 2024). MAPE not reported due to zero and negative prices. Random Forest selected as final model due to better generalization.

Source: Authors' own research results.

SHAP analysis was conducted exclusively on the Random Forest model, as it demonstrated better generalization. Computing SHAP values for both models would double computational cost without substantially altering explainability insights, given that tree-based ensemble models with similar accuracy tend to identify comparable feature importance hierarchies (Lundberg and Lee, 2017). We acknowledge that XGBoost's sequential boosting structure could potentially yield different SHAP decompositions, which represents a direction for future investigation.

Table 3. Top 10 features ranked by global SHAP importance

Rank	Feature	Category	Mean SHAP
1	price_t-1	Pricelags	7.81
2	price_t-7	Pricelags	6.56
3	price_t-14	Pricelags	6.52
4	residual_load_forecast	Forecasts	5.41
5	price_t-2	Pricelags	4.66
6	price_deviation_7d	Pricelags	3.93
7	renewable_share_forecast	Forecasts	3.71
8	price_t-3	Pricelags	3.69
9	day_of_year	Time features	2.68
10	hour_sin	Time features	2.43

Notes: Importance measured as mean absolute SHAP value across 1,000 test set observations. Higher values indicate greater contribution to predictions.

Source: Authors' own SHAP analysis.

Global feature importance

Table 3 presents the top ten features ranked by mean absolute SHAP value across all test set predictions. Price history features dominate, with yesterday's price (price_t-1, importance 7.81) showing the strongest influence, followed by weekly lags price_t-7 (6.56) and price_t-14 (6.52). This confirms substantial price persistence in the Romanian market, with weekly patterns reflecting recurring industrial demand cycles. The fourth-ranked feature, residual load forecast (5.41), validates economic theory: prices rise when the electricity that must come from controllable generation increases. Temporal features (day_of_year, hour_sin) capture seasonal and daily demand cycles, while renewable share forecast (3.71) reflects the fundamental supply-demand balance. Notably absent from the top ten are commodity prices (oil, gas, CO2), suggesting that short-term electricity prices respond more to immediate supply-demand conditions than to fuel costs.

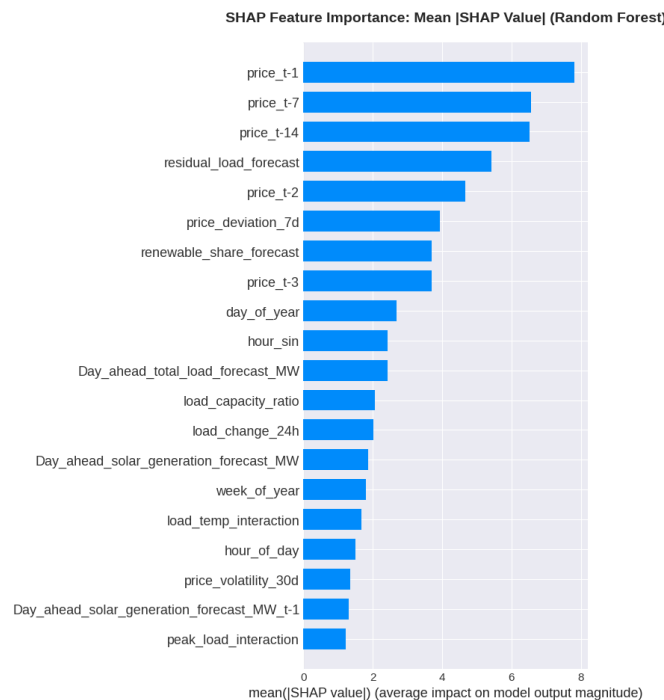


Figure 2. Global feature importance: top features ranked by mean absolute SHAP value

Source: Authors' own SHAP analysis.

Multi-horizon explainability analysis

Table 4 presents the aggregate contribution of each forecast horizon to overall predictive power. Short-term features (6–24 hours) dominate with 41 percent of horizon-attributed predictive contribution, followed by immediate features (23 percent), while medium-term and long-term features each contribute 18 percent. The short-term horizon's dominance makes economic sense for day-ahead forecasting: tomorrow's price depends primarily on today's conditions and official forecasts. The high per-feature contribution of immediate features (mean |SHAP| = 2.10 with only 7 features) reflects the strong influence of very recent prices and time-of-day patterns.

Table 4. Feature importance by forecast horizon (selected horizon-diagnostic features)

Horizon	No. of features	Mean SHAP	Total contribution	% of total
Immediate (0–6h)	7	2.10	14,729	23%
Short-term (6–24h)	8	3.30	26,412	41%
Medium-term (1–7 days)	9	1.24	11,136	18%
Long-term (7–14+ days)	10	1.16	11,587	18%
Total	34	1.88	63,864	100%

Notes: Total contribution summed across 1,000 test observations. 34 of 70 features are included; 36 features with ambiguous temporal attribution are excluded (see text). Percentages represent shares within the 34 assigned features, not total model output.

Source: Authors' own multi-horizon analysis.

Within each horizon, different features dominate. Immediate predictions rely on yesterday's price (7.81) and hour-of-day indicators. Short-term predictions increasingly weight residual load forecast (5.41) and recent price lags ($t-2$, $t-3$). Medium-term predictions shift toward the weekly cycle through price_{t-7} (6.56) and forecast error patterns. Long-term predictions incorporate price_{t-14} (6.52), 30-day volatility measures, and seasonal indicators. This hierarchy provides actionable guidance: day-ahead traders should prioritize monitoring today's prices and load forecasts, which drive 41 percent of horizon-attributed predictive contribution, while long-term planners should focus on commodity trends and seasonal patterns.

Market regime analysis

Table 5 presents the top five features in each price regime, revealing striking differences in what drives prices under different market conditions. During low-price periods, residual load forecast (7.96) emerges as the most important feature, as the supply-demand balance directly determines marginal generation costs when renewables are abundant. During normal conditions, multiple factors contribute relatively equally, with no single dominant driver. During high-price periods, yesterday's price (price_{t-1}) explodes to an importance of 16.86, more than three times its normal-period value (4.41), indicating that price momentum dominates during scarcity conditions. We note that the test set spans 41 days (August 20–September 30, 2024), capturing summer and early autumn conditions. The regime-specific findings therefore reflect this particular seasonal context, and the high-price regime (250 observations) is characterized by summer heat-driven demand rather than winter scarcity or energy crisis dynamics.

Table 5. Top 5 features by market regime

Rank	Low price (≤ 76.66)	Imp.	Normal (76.66–130.68)	Imp.	High price (≥ 130.68)	Imp.
1	residual_load_forecast	7.96	price _{t-14}	5.06	price _{t-1}	16.86
2	price _{t-1}	5.55	price _{t-1}	4.41	price _{t-7}	14.24
3	renewable_share_forecast	5.12	price _{t-7}	4.19	price _{t-14}	11.73
4	price _{t-14}	4.24	residual_load_forecast	4.07	price _{t-2}	11.55
5	price _{t-7}	3.65	renewable_share_forecast	3.17	price_deviation_7d	8.96

Notes: Low price: $\leq P25$, 250 obs. Normal: $P25-P75$, 500 obs. High price: $\geq P75$, 250 obs. Importance = mean|SHAP| within each regime. Price thresholds in EUR/MWh.

Source: Authors' own regime analysis.

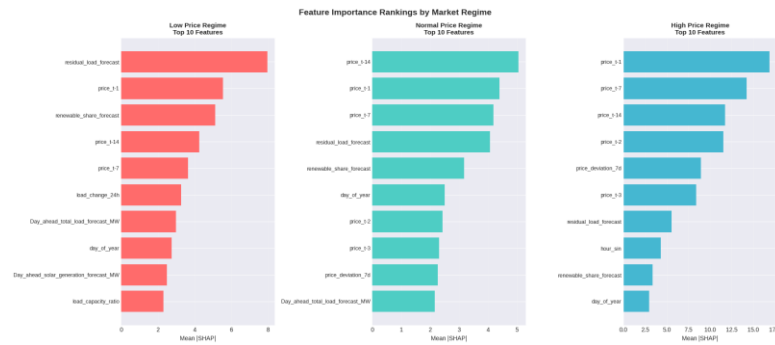


Figure 3. Market regime analysis: feature importance shifts across low, normal, and high price conditions

Source: Authors' own regime analysis.

To statistically validate that feature importance shifts across market conditions, we applied Mann–Whitney U tests to the SHAP value distributions from the cross-validation fold capturing the most extreme price behavior (Fold 4, which includes the summer 2024 heat wave with prices reaching 916.72 EUR/MWh). This fold provides the clearest separation between extreme and normal observations within a single training context. The ten features tested correspond to those with the highest absolute SHAP values within this fold's extreme-price subsample, which partially overlap with but are not identical to the global top ten features from Table 3. Extending statistical validation to the full test set SHAP decomposition across all globally important features represent a natural direction for future work.

Mann–Whitney U tests confirmed that these regime differences are statistically significant. Eight of the ten most important features exhibit highly significant differences between high-price and normal-price observations ($p < 0.001$). Yesterday's price shows the most dramatic shift, with a 4.25-fold increase in importance during extreme prices ($p < 0.0001$). Similarly significant increases were found for price_t-14 (2.84×), day_of_year (2.51×), residual_load_forecast (2.21×), and price_t-7 (2.17×). Notably, price_t-2 showed no significant regime difference (ratio 1.06, $p = 0.53$), suggesting that the two-day lag lacks the economic salience of either immediate persistence ($t - 1$) or weekly structural cycles ($t - 7$).

These patterns have clear economic interpretations. During oversupply, the marginal cost of generation sets the price, making the supply-demand balance the dominant driver. During normal operations, the market efficiently aggregates diverse information sources. During scarcity, capacity constraints bind and recent price levels become the primary signal of ongoing system stress, creating strong momentum effects. This finding implies that risk managers should trigger enhanced monitoring protocols when price persistence patterns intensify, as this signals a shift from normal to crisis market dynamics. The 4.25-fold increase in price_t-1 importance provides a quantitative benchmark for detecting such regime transitions.

Conclusion

This study developed an explainable electricity price forecasting model for the Romanian market and introduced two novel contributions: a multi-horizon explainability framework and a statistically validated market regime analysis. Using Random Forest and XGBoost models applied to 59,160 hourly observations (January 2018 to September 2024) with 70 engineered features, the selected Random Forest model achieved $R^2 = 0.60$ with MAE of 31.04 EUR/MWh on the test set.

The multi-horizon analysis revealed that short-term features (6–24 hours) contribute 41 percent of predictive power, nearly double the contribution of any other horizon. This finding provides actionable guidance for market stakeholders: day-ahead traders should prioritize monitoring today's prices, load forecasts, and renewable generation forecasts, while grid operators seeking to improve forecast accuracy should focus enhancement efforts on day-ahead load and renewable predictions, where marginal improvements yield maximum gains.

The market regime analysis demonstrated that feature importance shifts dramatically across price conditions. During high-price scarcity periods, price momentum (price_{t-1}) becomes

4.25 times more important than during normal conditions ($p < 0.0001$), while during low-price oversupply, the residual load forecast dominates. These statistically validated regime differences (eight of ten features significant at $p < 0.001$) imply that risk managers should implement regime-conditional monitoring protocols, intensifying attention to price persistence signals when markets enter crisis dynamics. A practical implementation could compute the rolling 7-day ratio of mean absolute SHAP values for price_{t-1} relative to its long-run average. When this ratio exceeds a threshold—for instance, $2.0\times$ based on the midpoint between normal ($1.0\times$) and crisis ($4.25\times$) levels observed in our data—it would signal a transition toward crisis dynamics warranting enhanced risk management. The precise threshold should be calibrated through backtesting on market-specific historical data, which we leave to future applied work.

Several limitations should be acknowledged. The 5,000-observation analysis window (March–September 2024) was selected for computational tractability of SHAP calculations. While the full 59,160-observation dataset spanning the 2022–2023 energy crisis was used for feature engineering—ensuring that all lagged and rolling calculations incorporate crisis-period dynamics—the model training and evaluation are limited to 2024 conditions. The energy crisis is referenced as contextual motivation for studying regime-dependent behavior, not as a directly modeled period. Consequently, the regime analysis reflects post-crisis market conditions, and direct transferability of the specific SHAP magnitudes to crisis-period dynamics requires separate validation. The findings are specific to the Romanian market during 2024, and direct transferability to markets with different generation mixes requires validation. The model showed moderate overfitting (train-test R^2 gap of 0.35), and performance degraded substantially during the summer 2024 heat wave ($R^2 = 0.41$), reflecting genuine uncertainty during extreme conditions rather than model failure.

Future research directions include extending the framework to other European markets for cross-market comparison, integrating deep learning architectures (LSTM, Transformers) while maintaining explainability, developing probabilistic forecasting with regime-dependent prediction intervals, and incorporating causal inference techniques to distinguish genuine causal effects from correlations. As electricity markets integrate increasing shares of variable renewable energy, understanding how price drivers shift across horizons and market regimes becomes ever more critical for the energy transition.

Data and Code Availability Statement

The data used in this study are available from the ENTSO-E Transparency Platform (<https://transparency.entsoe.eu/>).

The code and the database underlying the analysis are available on Quantlet at: <https://github.com/QuantLet/Multi-horizon-explainability-in-Romania-s-energy-price-forecasting>.

Acknowledgments

The authors would like to thank the ENTSO-E Transparency Platform for providing access to the market data used in this study.

This paper is supported through the project “IDA Institute of Digital Assets”, CF166/15.11.2022, contract number CN760046/23.05.2023; the project “AI for Energy Finance (AI4EFin)”, CF162/15.11.2022, contract number CN760048/23.05.2023, financed under Romania’s National Recovery and Resilience Plan, Apel nr. PNRR-III-C9-2022-I8; and the Marie Skłodowska-Curie Actions under the European Union’s Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym DIGITAL, Project No. 101119635.

This paper was co-financed by The Bucharest University of Economic Studies during the PhD program

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this paper, the authors used Anthropic’s Claude for the purposes of interpreting results, generating equations, generating text and codes for figures representations and tables based on our results. Following the use of this tool, the authors meticulously reviewed and revised the content as necessary and assume full responsibility for the final content of the publication. The paper’s PDF was created using LaTeX code inside the Prism AI online tool (from OpenAI). The Word version of the paper was manually created afterwards using ICBE 2026 proceedings template.

References

- Bâra, A., & Oprea, S.-V. (2024). Predicting day-ahead electricity market prices through the integration of macroeconomic factors and machine learning techniques. *International Journal of Computational Intelligence Systems*, 17, 10. <https://doi.org/10.1007/s44196-023-00387-3>
- Borges, J., Maia, R., & Guerreiro, S. (2024). Forecasting electricity price during extreme events using a hybrid model of LSTM and ARIMA architecture. In J. Filipe, M. S’ mialek, A. Brodsky, & S.
- Hammoudi (Eds.), *Enterprise Information Systems. ICEIS 2023. Lecture Notes in Business Information Processing* (Vol. 518). Springer. https://doi.org/10.1007/978-3-031-64748-2_15
- Cramer, E., Witthaut, D., Mitsos, A., & Dahmen, M. (2023). Multivariate probabilistic forecasting of intraday electricity prices using normalizing flows. *Applied Energy*, 346, 121370. <https://doi.org/10.1016/j.apenergy.2023.121370>
- El-Azab, H. I., Swief, R. A., El-Amary, N. H., & Temraz, H. K. (2024). Machine and deep learning approaches for forecasting electricity price and energy load assessment on real datasets. *Ain Shams Engineering Journal*, 15(4), 102613. <https://doi.org/10.1016/j.asej.2023.102613>
- Ganczarek-Gamrot, A., Gorczyca-Goraj, A., Pilot, K., & Kania, K. (2025). Electricity price volatility and the performance of machine learning forecasting models in European energy markets. *Energies*, 18(24), 6535. <https://doi.org/10.3390/en18246535>

- Heistrene, L., Machlev, R., Perl, M., Belikov, J., Baimel, D., Levy, K., Mannor, S., & Levron, Y. (2023). Explainability-based trust algorithm for electricity price forecasting models. *Energy and AI*, 14, 100259. <https://doi.org/10.1016/j.egyai.2023.100259>
- Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293, 116983. <https://doi.org/10.1016/j.apenergy.2021.116983>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Melgar-García, L., & Troncoso, A. (2024). A novel incremental ensemble learning for real-time explainable forecasting of electricity price. *Knowledge-Based Systems*, 305, 112574. <https://doi.org/10.1016/j.knosys.2024.112574>
- Michalakopoulos, V., Menos-Aikateriniadis, C., Sarmas, E., Zakynthinos, A., Georgilakis, P. S., & Askounis, D. (2026). Less is more: Data-driven day-ahead electricity price forecasting with short training windows. *Energies*, 19(2), 376. <https://doi.org/10.3390/en19020376>
- Nitsch, F., Schimeczek, C., & Bertsch, V. (2024). Applying machine learning to electricity price forecasting in simulated energy market scenarios. *Energy Reports*, 12, 5268–5279. <https://doi.org/10.1016/j.egyr.2024.11.013>
- O'Connor, C., Bahloul, M., Prestwich, S., & Visentin, A. (2025). A review of electricity price forecasting models in the day-ahead, intra-day, and balancing markets. *Energies*, 18(12), 3097. <https://doi.org/10.3390/en18123097>
- Pesenti, A., & O'Sullivan, A. (2025). Explaining deep neural network models for electricity price forecasting with XAI. *Energy and AI*, 21, 100532. <https://doi.org/10.1016/j.egyai.2025.100532>
- Tschora, L., Pierre, E., Plantervit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, 313, 118752. <https://doi.org/10.1016/j.apenergy.2022.118752>
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030–1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>