Article

Not peer-reviewed version

Mixed-Frequency Machine Learning for Nowcasting Energy-Related CPI in Turkey: Evidence from Explainable Models

Hamid Yeşilyayla *

Posted Date: 16 April 2026

doi: 10.20944/preprints202604.1196.v1

Keywords: inflation nowcasting; mixed-frequency forecasting; machine learning; XGBoost; LightGBM; Elastic Net; explainable forecasting; SHAP; energy-related CPI

Article

Mixed-Frequency Machine Learning for Nowcasting Energy-Related CPI in Turkey: Evidence from Explainable Models

Hamid Yeşilyayla

Pamukkale University, Denizli, Turkey; hyesilyayla@pau.edu.tr; Tel.: +905324323182

Highlights

What are the main findings?

1. Mixed-frequency machine learning improves nowcasting accuracy for the energy-related CPI component in the common sample.
2. Forecast gains are concentrated in headline inflation and EPİAŞ signals, while Brent remains a secondary cost channel.

What are the implications of the main findings?

3. Mixed-frequency nowcasting provides a usable framework for short-run inflation monitoring in Turkey.
4. Combining forecast evaluation with Elastic Net and SHAP interpretation improves the policy relevance of machine-learning forecasts.

Abstract

This study examines whether the monthly inflation rate of the energy-related consumer price index component in Turkey can be nowcast more accurately with mixed-frequency indicators. An expanding-window pseudo-out-of-sample design is used to compare a seasonal naive benchmark with Elastic Net, XGBoost, and LightGBM. The predictor set combines monthly macroeconomic indicators with features derived from daily Brent oil prices, daily USD/TRY exchange rates, and hourly EPİAŞ day-ahead electricity market data for 2012–2025. Forecast performance is evaluated with root mean squared error, mean absolute error, and symmetric mean absolute percentage error, while core-sample forecast differentials are assessed with the Diebold–Mariano test. All machine-learning models outperform the benchmark, and the lowest forecast errors are obtained from the core XGBoost specification. Explainability results from standardized Elastic Net coefficients and SHAP decompositions show that headline inflation and EPİAŞ variables provide the largest share of predictive content, Brent forms a secondary cost channel, inflation expectations are supportive, and exchange-rate variables do not emerge as an independently dominant block. The results support mixed-frequency machine learning as a useful framework for short-run monitoring of energy-related inflation in Turkey.

Keywords: inflation nowcasting; mixed-frequency forecasting; machine learning; XGBoost; LightGBM; Elastic Net; explainable forecasting; SHAP; energy-related CPI

---

1. Introduction

Timely inflation assessment matters for monetary policy, fiscal planning, and short-run macroeconomic surveillance. This need becomes more acute when the target series is released at a monthly frequency, while financial markets, commodity prices, and electricity markets generate

 doi:10.20944/preprints202604.1196.v1

information on a daily or hourly basis. In such settings, forecast accuracy depends not only on the model class, but also on how higher-frequency signals are aligned with the lower-frequency target.

This paper focuses on the nowcasting of an energy-related consumer price index component in Turkey. The target series is the monthly inflation rate of the CPI subgroup Housing, Water, Electricity, Gas and Other Fuels. Compared with headline inflation, this component is more directly exposed to imported energy costs, regulated price adjustments, and domestic electricity market conditions. For that reason, it provides a suitable setting for examining whether mixed-frequency indicators improve short-run inflation tracking.

The empirical design combines monthly macroeconomic controls with indicators derived from daily Brent oil prices, daily USD/TRY exchange rates, and hourly EPİAŞ day-ahead market data. The resulting dataset makes it possible to study whether short-run inflation forecasts benefit from information that is observed well before the monthly CPI release. This question is closely related to the broader literature on inflation forecasting in data-rich environments, mixed-frequency nowcasting, and explainable machine learning [1–4].

The contribution of the paper is twofold. First, it develops a mixed-frequency nowcasting framework for an energy-related inflation component in Turkey and compares seasonal naive, Elastic Net, XGBoost, and LightGBM specifications in an expanding-window pseudo-out-of-sample design. Second, it does not stop at forecast ranking. It combines standardized Elastic Net coefficients with SHAP-based decompositions for tree-based models in order to identify which variable families carry the predictive content of the nowcasting equation. This design makes it possible to distinguish between forecast performance and the underlying information structure.

The results point to a consistent forecasting pattern. All machine learning models improve on the seasonal naive benchmark, and the lowest forecast errors are obtained from the core XGBoost specification. The explainability evidence shows that headline inflation and EPİAŞ variables account for the largest share of predictive information, Brent indicators act as a secondary but persistent cost channel, inflation expectations play a supportive role, and exchange-rate variables do not emerge as an independently dominant block. These findings suggest that short-run monitoring of energy-related inflation in Turkey benefits from a mixed-frequency machine learning framework in which electricity market signals are treated as part of the forecasting information set.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data, feature construction, and forecasting design. Section 4 presents the forecasting models, evaluation metrics, and explainability framework. Section 5 reports the empirical results. Section 6 discusses the findings, and Section 7 concludes.

2. Related Literature

2.1. Inflation Forecasting with Machine Learning

Recent work on inflation forecasting shows that machine learning models can improve forecast accuracy when the predictor set is large and highly correlated. Medeiros et al. [2] report that, in data-rich settings, machine learning methods can outperform standard benchmark models. For Turkey, Özgür and Akkoç [5] show that shrinkage-based specifications such as lasso and Elastic Net are especially useful when variable selection is part of the forecasting problem.

The literature has also moved beyond forecast accuracy alone. Aras and Lisboa [3] combine tree-based forecasting with SHAP-based interpretation and show that explainability can be built into inflation forecasting without separating interpretation from prediction. Naghi et al. [6] add an important caution: model rankings vary across countries and periods, so forecast evaluation should remain sample-specific rather than tied to a single preferred algorithm.

2.2. Mixed-Frequency Nowcasting and High-Frequency Indicators

A second strand of the literature studies nowcasting designs that combine low-frequency target variables with high-frequency information. Modugno [1] shows that monthly inflation can be tracked

 doi:10.20944/preprints202604.1196.v1

more effectively when weekly and daily indicators are incorporated into the information set. Borup et al. [4] extend this logic to machine learning and demonstrate that mixed-frequency forecasting can deliver measurable out-of-sample gains when higher-frequency indicators are informative.

Related evidence comes from real-time and density-based applications. Knotek and Zaman [7] show that mixed-frequency model combinations improve real-time inflation nowcasts, while Bolivar [8] proposes a two-step machine learning strategy in which daily and weekly indicators are used to improve monthly inflation forecasts. Together, these studies justify a forecasting design in which frequency alignment is treated as a core modelling step rather than a purely technical pre-processing choice.

2.3. Energy-Related Price Signals and Disaggregated Inflation Forecasting

A third line of research is directly related to the target series used in this paper. Özmen and Özşahin [9] and Çelik and Danışoğlu [10] document the importance of energy prices and exchange-rate dynamics for inflation in Turkey. At the same time, Joseph et al. [11] show that forecasting at the level of CPI components can add useful information that is diluted in headline measures. For electricity-market signals, Arifoğlu and Kandemir [12] provide evidence that EPİAŞ-related price series are forecastable with nonlinear models, while Heistrene et al. [13] show that explainable forecasting tools can identify which electricity-price features carry the greatest predictive content.

Taken together, the literature leaves a specific gap. Studies on Turkey mostly focus either on headline inflation or on pass-through and causality questions, whereas mixed-frequency forecasting studies rarely examine an energy-related CPI component with electricity-market indicators embedded in the same forecasting equation. This paper addresses that gap by combining monthly, daily, and hourly information in a unified nowcasting design and by evaluating not only forecast accuracy but also the source of predictive information across variable families.

A detailed row-by-row literature table is reported in Supplementary Table S1. To keep the main text concise, Table 1 presents a compressed summary of the literature clusters most closely related to the present forecasting problem.

Table 1. Condensed Summary of the Related Literature.

Cluster	Representative studies	Data and frequency focus	Relevance to this paper
Machine learning and explainable inflation forecasting	[2,3,5,6]	Monthly inflation, data-rich panels, shrinkage methods, and explainable machine learning	Provides the methodological foundation for the forecasting architecture and the explainability layer.
Mixed-frequency nowcasting	[1,4,7,8]	Daily or weekly predictors combined with monthly or weekly target variables	Motivates the alignment of daily Brent and exchange-rate indicators and hourly EPİAŞ signals with a monthly target.
Disaggregated inflation forecasting	[11]	CPI components and component-level forecasting	Supports the decision to model an energy-related CPI component rather than headline inflation alone.

 doi:10.20944/preprints202604.1196.v1

Energy-related price signals and electricity-market evidence	[9,10,12,13]	Energy prices, exchange rates, and electricity-market indicators	Supports the use of energy-cost and electricity-market variables in the nowcasting equation.

Note. The detailed row-based literature table is provided in Supplementary Table S1. The summary table in the main text groups the literature by forecasting problem, data structure, and relevance to the present study.

3. Data and Forecasting Design

3.1. Target Variable and Information Set

The forecasting target is the monthly inflation rate of the energy-related Consumer Price Index (CPI) component Housing, Water, Electricity, Gas and Other Fuels in Turkey. This component is more directly exposed than headline CPI to imported energy costs, regulated price adjustments, and domestic electricity-market conditions. For that reason, it provides a suitable setting for evaluating whether mixed-frequency indicators improve short-run inflation nowcasts.

The monthly information set contains headline monthly inflation, lags of the target series, the policy rate, and, in the extended specification, 12-month-ahead inflation expectations. These variables capture general pricing conditions, inflation persistence, and the monetary policy and expectations channels that may shape short-run movements in the energy-related CPI component.

Higher-frequency predictors are drawn from daily Brent oil prices, daily USD/TRY exchange rates, and hourly EPİAŞ day-ahead electricity-market data. From Brent and the exchange rate series, the feature set includes monthly mean, end-of-month, and dispersion measures, together with Turkish-lira-denominated oil indicators. From EPİAŞ, the feature set includes monthly average price, intra-month volatility, peak-hour and off-peak-hour averages, and the peak-off-peak gap.

The resulting information set is organized so that each variable block retains an economic interpretation. Headline inflation and target lags proxy for pricing inertia, Brent and exchange-rate indicators capture imported cost pressures, and EPİAŞ variables reflect domestic electricity-market conditions. Inflation expectations are kept only in the extended sample so that their incremental forecasting role can be evaluated directly.

3.2. Frequency Alignment and Feature Construction

All variables are aligned to a monthly forecasting panel indexed at month-end. Monthly series enter the panel at their observed frequency. Daily Brent and USD/TRY series are aggregated into monthly summary indicators, including monthly means, standard deviations, and end-of-month values. Hourly EPİAŞ prices are aggregated into monthly level, dispersion, and load-structure measures so that high-frequency market information can be incorporated into the same equation as the monthly target.

Feature construction follows a common lag structure of 1, 2, 3, 6, and 12 months. This structure is intended to capture short-run persistence, medium-run adjustment, and annual seasonal recurrence within a unified predictor set. Starting months with incomplete lag information are removed from the effective estimation sample, while the modelling pipeline retains a median-imputation step as a safeguard against incidental missing values.

The raw panel covers January 2012 to December 2025, corresponding to 168 monthly observations. After accounting for lag availability and the effective start dates of some series, the core sample spans January 2013 to December 2025 and contains 156 observations. The extended sample begins in January 2014 and contains 144 observations because the inflation-expectations series starts later than the remaining predictors.

3.3. Pseudo-Out-of-Sample Design

 doi:10.20944/preprints202604.1196.v1

Forecast evaluation is based on an expanding-window pseudo-out-of-sample design. At each forecast origin, the model is re-estimated using only the information available up to that date and then used to produce a one-step-ahead nowcast for the next monthly observation. This procedure preserves the time ordering of the data and avoids the leakage problems that can arise when random cross-validation is applied to dependent observations [14].

The initial estimation window is set to 60 months, which yields 96 pseudo-out-of-sample forecasts for the core sample and 84 for the extended sample. The benchmark forecast is the seasonal naive specification, which uses the value observed 12 months earlier as the comparison forecast. All machine learning models are therefore evaluated against the same seasonal baseline before they are compared with one another.

4. Forecasting Models and Evaluation

4.1. Seasonal Naive Benchmark

The seasonal naive model serves as the benchmark forecast. For monthly data, it uses the observation from the same month of the previous year as the one-step-ahead forecast. This benchmark is deliberately difficult to beat when seasonal persistence is strong, and it provides a transparent baseline against which machine learning models can be evaluated.

$$\hat{y}_t = y_{t-12} \tag{1}$$

This specification is retained throughout the empirical analysis so that all gains in forecast accuracy are measured relative to the same seasonal reference point.

4.2. Elastic Net

Elastic Net combines L1 and L2 penalties in a single regularized regression framework. The method is well suited to forecasting environments with many lagged and correlated predictors because it performs variable selection and coefficient shrinkage at the same time [15]. In the present setting, that property is useful because headline inflation, Brent indicators, exchange-rate measures, policy-rate variables, EPİAŞ features, and target lags generate a dense and potentially collinear predictor space.

$$\min_{\beta_0,\beta} \frac{1}{2n}\sum_{i=1}^{n} (y_i - \beta_0 - x_i^{\top}\beta)^2 + \lambda\left[\frac{1-\alpha}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right] \tag{2}$$

The estimation pipeline uses ElasticNetCV with a time-series cross-validation structure inside each training window. The tuning grid spans five alpha values between 0.2 and 1.0 and a logarithmic sequence of penalty values. All predictors are standardized before estimation so that coefficient magnitudes can later be compared across variables and across variable families.

4.3. XGBoost

XGBoost is a regularized gradient-boosting algorithm that builds an additive ensemble of decision trees. At each iteration, the model fits a new tree to the remaining forecast error, while a complexity penalty discourages overly deep or unstable trees [16]. This structure is attractive in small samples because it can capture nonlinear thresholds and interactions without relying on a fully unrestricted specification.

$$\begin{gathered} \mathrm{Obj}^{(m)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(m-1)} + f_m(x_i)\right) + \Omega(f_m) \\ \Omega(f) = \gamma T + \frac{\lambda}{2}\sum_{j=1}^{T} w_j^2 \end{gathered} \tag{3}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2026 doi:10.20944/preprints202604.1196.v1

The core hyperparameters are set to n_estimators = 300, max_depth = 3, learning_rate = 0.05, subsample = 0.9, and colsample_bytree = 0.8. This parameterization is intended to balance flexibility and regularization in a relatively short pseudo-out-of-sample forecasting exercise.

4.4. LightGBM

LightGBM belongs to the same gradient-boosting family as XGBoost, but it uses a different tree-growing strategy and computational shortcuts such as Gradient-based One-Side Sampling and Exclusive Feature Bundling [17]. In this paper, LightGBM plays the role of a second tree-based comparator so that the forecasting results do not hinge on a single boosting implementation.

The LightGBM specification uses n_estimators = 300, num_leaves = 15, learning_rate = 0.05, subsample = 0.9, colsample_bytree = 0.8, and reg_lambda = 1.0. The feature set and the expanding-window evaluation protocol are kept identical to those of XGBoost, so performance differences can be attributed to the learning algorithm rather than to the forecasting design.

4.5. Forecast Evaluation Metrics

Forecast accuracy is evaluated with root mean squared error (RMSE), mean absolute error (MAE), and symmetric mean absolute percentage error (sMAPE). RMSE places greater weight on large forecast misses, MAE provides a more robust average loss measure, and sMAPE offers a scale-free complement to the level-based metrics [18].

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{5}$$

$$\text{sMAPE} = \frac{100}{n}\sum_{t=1}^{n}\frac{2|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \tag{6}$$

Although sMAPE is not fully symmetric in a strict statistical sense [19], it remains useful here as a supplementary metric because the target series varies in both sign and magnitude over time.

4.6. Statistical Comparison of Forecast Accuracy

To assess whether forecast differences are statistically meaningful, the core models are compared with the Diebold–Mariano test. The test evaluates whether the mean difference between two forecast-loss series is equal to zero [20]. Separate comparisons are reported for absolute-error loss and squared-error loss.

$$\text{DM} = \frac{\bar{d}}{\sqrt{\frac{\text{Var}(d_t)}{T}}} \tag{7}$$

Here, T denotes the number of forecasts and $\bar{d}$ is the sample mean of the loss differential. In the empirical section, a negative Diebold-Mariano statistic indicates that the first model listed in the comparison delivers a lower average forecast loss than the second model.

4.7. Explainability Framework

Explainability is built into the forecasting design rather than treated as a separate post-hoc add-on. For Elastic Net, interpretation is based on standardized coefficients, selection frequencies, and the evolution of variable-family shares over time. This makes it possible to distinguish between variables that enter the model frequently and variables that carry large average weights once selected.

 doi:10.20944/preprints202604.1196.v1

$$\text{Share}_{g,t} = \frac{\sum_{j\in g} \left|\beta_{j,t}^{\text{std}}\right|}{\sum_k \left|\beta_k^{\text{std}}\right|} \tag{8}$$

For XGBoost and LightGBM, explainability is based on SHAP values, which decompose each fitted prediction into a baseline term and variable-level contributions. The analysis is reported at two levels: individual predictors and broader variable families, namely headline inflation, EPİAŞ, Brent, exchange rate, policy rate, own lags, and inflation expectations. This structure allows the paper to compare forecast accuracy and predictive content within the same empirical framework.

$$f(x_i) = \phi_0 + \sum_{j=1}^{p} \phi_{ij} \tag{9}$$

5. Empirical Results

5.1. Forecast Accuracy Across Models

Table 2 reports forecast accuracy for the benchmark and the machine learning models. In the core sample, the RMSE ranking is XGBoost (3.191), LightGBM (3.205), Elastic Net (3.614), and seasonal naive (4.601). The MAE ranking is identical, with XGBoost again producing the lowest loss (1.979), followed by LightGBM (2.088), Elastic Net (2.317), and seasonal naive (2.745). All machine learning specifications therefore improve on the benchmark in the common core sample.

The extended sample yields the same broad result. LightGBM records the lowest RMSE (3.299), followed by XGBoost (3.534), Elastic Net (3.702), and seasonal naive (4.861). These gains, however, are obtained over a shorter forecast span and under a wider information set that includes inflation expectations. For that reason, the model ranking across samples should be read with care. When forecast accuracy, sample coverage, and the common explainability design are considered jointly, the core XGBoost specification remains the main reference model.

Table 2. Final model comparison across samples.

Model	Sample	Start date	End date	Forecasts	RMSE	MAE	sMAPE	Improvement vs. naive (%)	DM result
Seasonal naive	Core	2018-01-31	2025-12-31	96	4.601	2.745	89.454	0.000	Reference
Elastic Net	Core	2018-01-31	2025-12-31	96	3.614	2.317	70.336	21.461	No significant difference
XGBoost	Core	2018-01-31	2025-12-31	96	3.191	1.979	70.007	30.653	Better than naive, p=0.037
LightGBM	Core	2018-01-31	2025-12-31	96	3.205	2.088	76.101	30.339	No significant difference
Seasonal naive	Extended	2019-01-31	2025-12-31	84	4.861	2.933	87.493	0.000	Reference
Elastic Net	Extended	2019-01-31	2025-12-31	84	3.702	2.459	73.850	23.838	Not tested
XGBoost	Extended	2019-01-31	2025-12-31	84	3.534	2.176	71.547	27.295	Not tested
LightGBM	Extended	2019-01-31	2025-12-31	84	3.299	2.111	71.642	32.128	Not tested

Note. The improvement rate is calculated as the percentage decline in RMSE relative to the seasonal naive benchmark. Diebold-Mariano tests are reported only for the core sample.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 April 2026 doi:10.20944/preprints202604.1196.v1

5.2. Diebold-Mariano Comparisons

Table 3 reports the pairwise Diebold-Mariano comparisons for the core sample. The core XGBoost specification outperforms the seasonal naive benchmark under both MAE loss (DM = -2.559, p = 0.010) and MSE loss (DM = -2.085, p = 0.037). By contrast, the comparisons among XGBoost, Elastic Net, and LightGBM are not statistically significant at conventional levels. The benchmark gap is therefore clear, whereas the distance among the advanced models is more limited.

Table 3. Diebold-Mariano test results for the core sample.

Model A	Model B	DM statistic (MAE loss)	p-value	DM statistic (MSE loss)	p-value
XGBoost	Seasonal naive	-2.559	0.010	-2.085	0.037
XGBoost	Elastic Net	-1.618	0.106	-1.289	0.197
XGBoost	LightGBM	-0.809	0.418	-0.089	0.929
LightGBM	Elastic Net	-1.144	0.252	-1.304	0.192

Note. The Diebold-Mariano test is reported only for the core sample. A negative DM statistic indicates that the first model listed produces a lower forecast loss than the second model.

Figure 1 reports the actual monthly energy-related inflation series together with the core XGBoost forecasts and the seasonal naive benchmark over the pseudo-out-of-sample evaluation period. The XGBoost path tracks the realized series more closely during the high-volatility episodes.

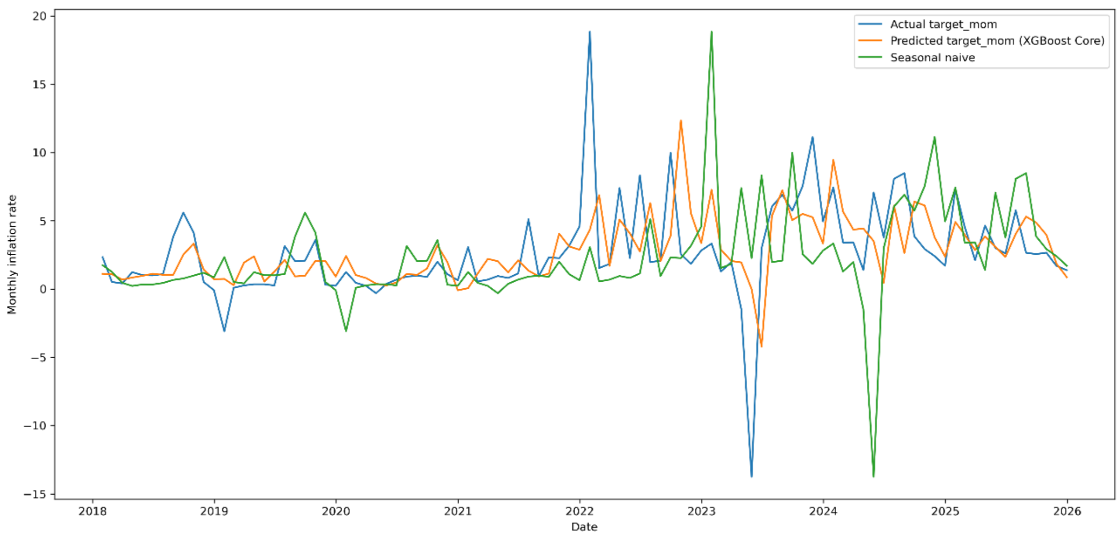

Figure 1. Actual and predicted energy-related monthly inflation under the core XGBoost specification.

5.3. Elastic Net Selection Patterns and Variable-Family Shares

Elastic Net evidence points to the same two leading blocks in both samples. The most frequently selected regressors are concentrated in headline inflation and EPİAŞ, with headline_mom, headline_mom_lag1, and epias_tl_mean appearing persistently across forecast origins. In the extended sample, inflation expectations enter the model, but they do not move into a dominant position within the leading predictor set.

Family-level coefficient shares lead to the same conclusion. In the core sample, headline inflation and EPİAŞ are each dominant in 48 forecast origins, while in the extended sample EPİAŞ is dominant in 44 periods and headline inflation in 40. Brent, own lags, and the policy rate become more visible in some subperiods, but they do not sustain comparable dominance.

5.4. SHAP-Based Evidence from Tree Models

The SHAP results point to the same general structure in the core tree models. In XGBoost, headline_mom receives the largest mean absolute SHAP value (0.932741), followed by epias_tl_mean, headline_mom_lag1, epias_tl_std_lag6, and brent_try_last_lag1. The ordering places headline inflation first at the single-variable level, EPİAŞ second, and Brent among the most visible secondary contributors.

The core LightGBM results yield a related pattern. Headline_mom again ranks first (0.887086), followed by policy_rate_lag1, epias_tl_std_lag3, brent_usd_last_lag3, and brent_usd_last. EPİAŞ volatility and load-structure measures become more visible in LightGBM than in XGBoost, while Brent also gains weight. Taken together, the tree models point to the same family-level ordering: headline inflation and EPİAŞ carry the largest share of predictive content, Brent remains the clearest secondary cost channel, and exchange-rate variables do not enter the top SHAP rankings.

5.5. Core versus Extended Sample Evidence

Comparing the core and extended samples yields two points. First, the addition of inflation expectations changes the ranking within the extended sample but does not overturn the main forecasting result. Extended LightGBM records the lowest RMSE within that shorter sample, yet the combined evidence does not displace core XGBoost as the main reference specification. The extended design covers fewer forecast origins and introduces a later-starting predictor block, so its gains should not be read as a clean improvement over the core setup.

Second, inflation expectations appear to play a supportive rather than defining role. In the extended models, inflation_expectation_12m and its first lag become more visible, especially in LightGBM, but they do not emerge as a dominant family in the Elastic Net profiles or in the broader ranking of predictive blocks. The exchange-rate family is even weaker as an independent source of forecast gains. Exchange-rate variables do not become dominant in the Elastic Net profiles and do not enter the leading SHAP ranks, whereas Brent-based indicators retain a clearer and more persistent presence. This pattern suggests that exchange-rate effects are absorbed through lira-denominated oil costs and the broader pricing block rather than operating as a separate leading family.

6. Discussion

The first point concerns forecast performance relative to the benchmark. Across both samples, all machine-learning specifications reduce forecast loss relative to the seasonal naive model, and the core XGBoost specification yields the lowest errors within the common sample. This result is consistent with evidence that machine-learning methods can improve inflation forecasts in data-rich settings, but it also supports the caution in Naghi et al. [6] that model rankings remain sample-specific. The present evidence therefore supports a limited claim: for this target variable, sample period, and pseudo-out-of-sample design, mixed-frequency tree models extract more useful short-run information than the seasonal benchmark.

A second point concerns the composition of predictive content. The Elastic Net and SHAP results both place headline inflation and EPİAŞ at the center of the forecasting equation. This pattern is consistent with two features of the target series. Headline inflation carries general pricing conditions and inertia, whereas EPİAŞ variables proxy domestic electricity-market conditions that are close to the energy-related CPI component by construction. In that sense, the findings complement studies that link electricity-market variables to forecastable price dynamics in Turkey and studies that favor component-level inflation modelling over exclusive reliance on headline CPI [11,12].

Brent plays a secondary but persistent role. It does not dominate the forecasting system, yet it appears repeatedly in the tree-based rankings and remains visible in the family-level summaries. This is consistent with the view that global energy costs matter for Turkish inflation, while their effect on the present target is filtered through regulated pricing, domestic electricity conditions, and the timing of pass-through [9,21]. The empirical pattern is therefore better described as a stable cost channel than as the primary source of short-run forecast variation.

The weaker position of the exchange-rate family requires careful interpretation. The Turkish inflation literature assigns a central role to exchange-rate pass-through, yet the present models do not identify exchange-rate variables as an independently dominant family. One plausible explanation is collinearity across economically related blocks. Part of the information carried by the exchange rate is already embedded in lira-denominated Brent indicators, headline inflation, and other variables that absorb broad pricing pressure. Another explanation is that the target series is a specific energy-related CPI component rather than aggregate inflation, so the exchange-rate channel may reach the dependent variable through more indirect routes than in headline inflation equations [10,22,23].

Finally, the explainability layer changes the contribution of the paper. A pure forecast comparison would show that one specification performs better than the others, but it would leave open the question of why. By combining standardized Elastic Net coefficients with SHAP decompositions, the analysis traces the predictive content back to interpretable variable families and shows how their relative importance changes across forecast origins. This makes the results easier to relate to the economic structure of the target series and places the paper closer to the recent literature on explainable forecasting rather than to a black-box model comparison alone [3,24].

7. Conclusions

This paper examines whether an energy-related CPI component in Turkey can be nowcast more accurately with mixed-frequency indicators. Within the present design, the answer is positive. All machine-learning models outperform the seasonal naive benchmark, and the core XGBoost specification yields the lowest forecast errors in the common sample. The predictive content of the nowcasting equation is concentrated in headline inflation and EPİAŞ variables, while Brent forms a secondary cost channel.

The paper also contributes in methodological terms. It combines monthly, daily, and hourly information in a single pseudo-out-of-sample framework and links forecast evaluation to an explicit explainability design. The combination of Elastic Net coefficient profiles and SHAP-based tree interpretation makes it possible to compare models on forecast loss and, at the same time, to identify which information blocks drive the forecasts. This is the main value added of the study relative to a standard model-ranking exercise.

The results should be read with the limits of the data and design in mind. The target variable is one CPI component, the extended sample is shorter because the expectations series starts later, and the feature set does not include regulated tariff adjustments, natural-gas pricing details, or real-time data vintages. Future research may extend the framework by adding such variables, by testing alternative forecast horizons and tuning schemes, and by examining whether the same explanatory structure holds for other CPI components or for headline inflation.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary Table S1: Extended literature summary; Supplementary Table S2. Variable construction details - Panel A. Raw series, aggregation rules, and monthly variables - Panel B. Lag structure and model-side inclusion; Figure S1: Actual versus core XGBoost predictions (core); Figure S2: Actual versus seasonal naive predictions (core); Figure S3: Top 20 mean absolute SHAP values – XGBoost (extended sample); Figure S4: Top 20 mean absolute SHAP values – LightGBM (extended sample); Figure S5: Variable-family coefficient shares over time in Elastic Net (core sample); Figure S6: Variable-family coefficient shares over time in Elastic Net (extended sample); Figure S7: Number of dominant periods by variable family in the Elastic Net specifications; Figure S8: Top 20 mean absolute SHAP values – XGBoost (core sample); Figure S9: Top 20 mean absolute SHAP values – LightGBM (core sample).

Funding: This research received no external funding.

 doi:10.20944/preprints202604.1196.v1

Data Availability Statement: The data used in this study were obtained from publicly available sources, namely the Electronic Data Delivery System of the Central Bank of the Republic of Turkey, the one-week repo rate page of the Central Bank of the Republic of Turkey, the U.S. Energy Information Administration, and the EPİAŞ transparency platform. The processed monthly panel, model outputs, and code used in the empirical analysis are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the author used OpenAI ChatGPT for structural revision, OpenAI ChatGPT, DeepL and Quillbot for language-editing support. The author reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

CPI, Consumer Price Index.
EVDS, Electronic Data Delivery System.
EPİAŞ, Enerji Piyasaları İşletme A.Ş.
MAE, Mean Absolute Error.
RMSE, Root Mean Squared Error.
SHAP, Shapley Additive Explanations.
sMAPE, Symmetric Mean Absolute Percentage Error.
USD/TRY, U.S. dollar/Turkish lira exchange rate.

References

1. Modugno, M. Now-casting inflation using high frequency data. Int. J. Forecast. 2013, 29, 664–675. https://doi.org/10.1016/j.ijforecast.2012.12.003
2. Medeiros, M.C.; Vasconcelos, G.F.R.; Veiga, Á.; Zilberman, E. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. J. Bus. Econ. Stat. 2021, 39, 98–119. https://doi.org/10.1080/07350015.2019.1637745
3. Aras, S.; Lisboa, P.J.G. Explainable inflation forecasts by machine learning models. Expert Syst. Appl. 2022, 207, 117982. https://doi.org/10.1016/j.eswa.2022.117982
4. Borup, D.; Rapach, D.E.; Schütte, E.C.M. Mixed-frequency machine learning: Nowcasting and backcasting weekly initial claims with daily internet search volume data. Int. J. Forecast. 2023, 39, 1122–1144. https://doi.org/10.1016/j.ijforecast.2022.05.005
5. Özgür, Ö.; Akkoç, U. Inflation forecasting in an emerging economy: Selecting variables with machine learning algorithms. Int. J. Emerg. Mark. 2022, 17, 1889–1908. https://doi.org/10.1108/IJOEM-05-2020-0577
6. Naghi, A.A.; O'Neill, E.; Danielova Zaharieva, M. The benefits of forecasting inflation with machine learning: New evidence. J. Appl. Econom. 2024, 39, 1321–1331. https://doi.org/10.1002/jae.3088
7. Knotek, E.S., II; Zaman, S. Real-time density nowcasts of US inflation: A model combination approach. Int. J. Forecast. 2023, 39, 1736–1760. https://doi.org/10.1016/j.ijforecast.2022.04.007
8. Bolivar, O. High-frequency inflation forecasting: A two-step machine learning methodology. Lat. Am. J. Cent. Bank. 2026, 7, 100172. https://doi.org/10.1016/j.latcb.2025.100172
9. Özmen, İ.; Özşahin, Ş. Effects of global energy and price fluctuations on Turkey's inflation: New evidence. Econ. Chang. Restruct. 2023, 56, 2695–2728. https://doi.org/10.1007/s10644-023-09530-8
10. Çelik, M.Y.; Danışoğlu, F. The relationship between energy inflation and exchange rate: A study on Turkey. İnsan Toplum Bilim. Araşt. Derg. 2023, 12, 2650–2667. https://doi.org/10.15869/itobiad.1314986
11. Joseph, A.; Potjagailo, G.; Chakraborty, C.; Kapetanios, G. Forecasting UK inflation bottom up. Int. J. Forecast. 2024, 40, 1521–1538. https://doi.org/10.1016/j.ijforecast.2024.01.001
12. Arifoğlu, A.; Kandemir, T. Electricity price forecasting in Turkish day-ahead market via deep learning techniques. Mehmet Akif Ersoy Üniv. İİBF Derg. 2022, 9, 1433–1458. https://doi.org/10.30798/makuiibf.1097686

 doi:10.20944/preprints202604.1196.v1

13. Heistrene, L.; Belikov, J.; Baimel, D.; Katzir, L.; Machlev, R.; Levy, K.; Mannor, S.; Levron, Y. An improved and explainable electricity price forecasting model via SHAP-based error compensation approach. IEEE Trans. Artif. Intell. 2025, 6, 159–168. https://doi.org/10.1109/TAI.2024.3455313
14. Bergmeir, C.; Benítez, J.M. On the use of cross-validation for time series predictor evaluation. Inf. Sci. 2012, 191, 192–213. https://doi.org/10.1016/j.ins.2011.12.028
15. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 2005, 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
16. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785
17. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154. https://doi.org/10.5555/3294996.3295074
18. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. Int. J. Forecast. 2006, 22, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
19. Goodwin, P.; Lawton, R. On the asymmetry of the symmetric MAPE. Int. J. Forecast. 1999, 15, 405–408. https://doi.org/10.1016/S0169-2070(99)00007-2
20. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. J. Bus. Econ. Stat. 1995, 13, 253–263. https://doi.org/10.1080/07350015.1995.10524599
21. Yilmazkuday, H. Drivers of Turkish inflation. Q. Rev. Econ. Finance 2022, 84, 315–323. https://doi.org/10.1016/j.qref.2022.03.005
22. Anderl, C.; Caporale, G.M. Nonlinearities in the exchange rate pass-through: The role of inflation expectations. Int. Econ. 2023, 173, 86–101. https://doi.org/10.1016/j.inteco.2022.10.003
23. Mirza, N.; Naqvi, B.; Rizvi, S.K.A.; Boubaker, S. Exchange rate pass-through and inflation targeting regime under energy price shocks. Energy Econ. 2023, 124, 106761. https://doi.org/10.1016/j.eneco.2023.106761
24. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2020, 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9
25. Barkan, O.; Benchimol, J.; Caspi, I.; Cohen, E.; Hammer, A.; Koenigstein, N. Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. Int. J. Forecast. 2023, 39, 1145–1162. https://doi.org/10.1016/j.ijforecast.2022.04.009
26. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774. https://doi.org/10.5555/3295222.3295230
27. Nguyen, T.-T.; Nguyen, H.-G.; Lee, J.-Y.; Wang, Y.-L.; Tsai, C.-S. The consumer price index prediction using machine learning approaches: Evidence from the United States. Heliyon 2023, 9, e20730. https://doi.org/10.1016/j.heliyon.2023.e20730
28. Öğünç, F.; Akdoğan, K.; Başer, S.; Gülenay Chadwick, M.; Ertuğ, D.; Hülagu, T.; Kösem, S.; Özmen, M.U.; Tekatlı, N. Short-term inflation forecasting models for Turkey and a forecast combination analysis. Econ. Model. 2013, 33, 312–325. https://doi.org/10.1016/j.econmod.2013.04.001
29. Şengüler, H.; Kara, B. Forecasting the inflation for budget forecasters: An analysis of ANN model performance in Turkey. Ekon. Polit. Finans Araşt. Derg. 2025, 10, 58–91. https://doi.org/10.30784/epfad.1588423
30. Şimşek, A.İ. Forecasting consumer price index using macroeconomic variables: A comparative analysis of machine learning and deep learning approaches. Bingöl Üniv. Sos. Bilim. Enst. Derg. 2024, 28, 15–29. https://doi.org/10.29029/busbed.1394983