sustainability

MDPI

Article

National-Scale Electricity Consumption Forecasting in Turkey Using Ensemble Machine Learning Models: An Interpretability-Centered Approach

Ahmet Sabri Öğütlü

Department of Industrial Engineering, Faculty of Engineering, Harran University, Sanliurfa 63300, Turkey; sogutlu@harran.edu.tr; Tel.: +90-4143180000

Abstract

This study presents an advanced, interpretability-focused machine learning framework for forecasting electricity consumption in Turkey over the period 2016–2024. The proposed approach is based on a high-dimensional dataset that incorporates a diverse set of variables, including sector-specific electricity usage (residential, industrial, lighting, agricultural, and commercial), electricity production, trade metrics (imports and exports in USD), and macroeconomic indicators such as the Industrial Production Index (IPI). A comprehensive set of eight state-of-the-art regression algorithms—including ensemble models such as CatBoost, LightGBM, Random Forest, and Bagging Regressor—were developed and rigorously evaluated. Among these, CatBoost emerged as the most accurate model, achieving R^2 values of 0.9144 for electricity production and 0.8247 for electricity consumption. Random Forest and LightGBM followed closely, further confirming the effectiveness of tree-based ensemble methods in capturing nonlinear relationships in complex datasets. To enhance model interpretability, SHAP (SHapley Additive exPlanations) and traditional feature importance analyses were applied, revealing that residential electricity consumption was the dominant predictor across all models, accounting for more than 70% of the variance explained in consumption forecasts. In contrast, macroeconomic indicators and temporal variables showed marginal contributions, suggesting that electricity demand in Turkey is predominantly driven by internal sectoral consumption trends rather than external economic or seasonal dynamics. In addition to historical evaluation, scenario-based forecasting was conducted for the 2025–2030 period, incorporating varying assumptions about economic growth and population trends. These scenarios demonstrated the model's robustness and adaptability to different future trajectories, offering valuable foresight for strategic energy planning. The methodological contributions of this study lie in its integration of high-dimensional, multivariate data with transparent, interpretable machine learning models, making it a robust and scalable decision-support tool for policymakers, energy authorities, and infrastructure planners aiming to enhance national energy resilience and policy responsiveness.

Keywords: electricity forecasting; machine learning; ensemble models; scenario analysis; energy demand

check for updates

Academic Editor: Huaxi Yuan

Received: 30 September 2025
Revised: 28 October 2025
Accepted: 1 November 2025
Published: 4 November 2025

Citation: Öğütlü, A.S. National-Scale Electricity Consumption Forecasting in Turkey Using Ensemble Machine Learning Models: An Interpretability-Centered Approach. *Sustainability* **2025**, *17*, 9829. https://doi.org/10.3390/su17219829

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The global transition toward sustainable and efficient energy systems has significantly intensified the demand for accurate electricity consumption forecasting. Electricity plays a

pivotal role as the backbone of modern economies; thus, its accurate prediction is essential not only for real-time operational efficiency but also for long-term infrastructure planning, renewable energy integration, environmental policy formulation, and the advancement of sustainable development goals [1,2]. In emerging economies such as Turkey where rapid urbanization, industrialization, and demographic shifts coincide with volatile economic trends and climatic variability, developing reliable forecasting models is not only beneficial but imperative [3,4].

While traditional statistical techniques, including ARIMA, SARIMA, and econometric models, have been employed extensively in the past, they often fail to capture the nonlinear dynamics, seasonality, and complex interdependencies inherent in large-scale, multivariate electricity consumption data [5,6]. These limitations have prompted a paradigm shift toward data-driven approaches, particularly machine learning (ML) and hybrid modeling frameworks, which have demonstrated superior performance in modeling nonlinearity and extracting patterns from high-dimensional datasets [1,3]. Recent studies have highlighted the advantages of ML-based models such as Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting Machines (GBM), Artificial Neural Networks (ANN), and advanced deep learning techniques like Long Short-Term Memory (LSTM) networks for electricity load forecasting [5,7]. In particular, ensemble and hybrid ML models have proven to enhance prediction accuracy by capturing both short-term fluctuations and long-term consumption trends [8,9].

Turkey serves as an ideal case for applying these sophisticated models due to the country's evolving energy landscape between 2016 and 2024. This period has been marked by dynamic shifts in industrial production, policy-driven market liberalization, regional infrastructure disparities, fluctuating climate conditions, and ambitious goals for renewable energy deployment [3]. These transformations have introduced new complexities into electricity demand forecasting, necessitating models capable of learning from diverse and granular inputs—ranging from meteorological variables (e.g., temperature, humidity, precipitation) to macroeconomic indicators (e.g., GDP, industrial production index), demographic features (e.g., population density, urbanization rate), and temporal dynamics (e.g., holidays, weekdays, seasonal cycles). Although the availability of extensive datasets from national agencies—such as the Turkish Statistical Institute (TÜİK), the Ministry of Energy and Natural Resources, and the Turkish Electricity Transmission Corporation (TEİAŞ)—provides fertile ground for advanced modeling, a critical gap remains in the application of high-dimensional ML-based forecasting models tailored to the Turkish context. Most existing studies are either confined to short-term predictions, limited variable sets, or regional-scale assessments, thereby underscoring the need for a comprehensive national-level approach [1,7].

Recent advances in electricity consumption forecasting have been driven by the integration of machine learning (ML), deep learning (DL), and hybrid methodologies to improve prediction accuracy, particularly in high-dimensional and dynamic environments. Models such as Random Forest and Lasso Lars outperform traditional time series models when handling large-scale, multi-variable data [10]. Similarly, a hybrid model combining LSTM and regression-based techniques achieved a notable 96.83% accuracy in the Ukrainian market [11]. Dense encoder architectures with NSGA-optimized parameters were applied for forecasting in smart cities, underscoring the importance of hyperparameter tuning [4]. The value of Support Vector Regression (SVR) in integrating socio-economic and climatic variables was also highlighted, resulting in high precision in corporate electricity demand prediction [12]. These studies collectively reveal a clear shift toward data-rich, nonlinear, and context-aware forecasting systems.

The effectiveness of hybrid and DL-based models is further substantiated by several comprehensive reviews. Recent trends indicate that LSTM, CNN, and Transformer models are superior in handling nonlinear, seasonal, and high-frequency electricity demand data [13]. Transformer models offer improved accuracy across varying data granularities, from daily to yearly predictions [14]. Hybrid CNN-RNN models are recommended, especially when applied to smart meter datasets [15]. An analysis of 77 studies reported a strong dominance of artificial intelligence (AI) techniques—particularly artificial neural networks (ANNs)—in short-term electricity forecasting [16]. Cross-country comparative studies also contribute to understanding the adaptability of forecasting methods. A study exploring seven countries found that fuzzy time series (FTS) methods, though sometimes overlooked, could outperform classical models when tailored to specific regional demand patterns [17].

Despite substantial progress in the application of machine learning techniques for electricity forecasting, existing studies often suffer from limited scope—focusing on short-term horizons, region-specific analyses, or a narrow range of input features. Critically, few have utilized high-dimensional national datasets that holistically integrate sectoral consumption, economic indicators, and temporal variables. Additionally, the issue of model interpretability remains underexplored, reducing the practical applicability of these models for energy policy and planning. This study addresses these deficiencies by developing and systematically evaluating a set of state-of-the-art ML models including both standalone and ensemble architectures trained on a comprehensive multivariate dataset covering 2016–2024. By incorporating granular sectoral data (residential, industrial, lighting, commercial, and agricultural), economic indicators (electricity trade values, Industrial Production Index), and temporal features, this research enhances forecasting accuracy while employing SHAP based analysis to improve model transparency. The novelty of this study lies in its integration of interpretable, ensemble-based machine learning models at a national scale—offering methodological innovation and practical insights for navigating Turkey's complex and rapidly evolving energy landscape.

2. Literature Review

Forecasting electricity consumption and production is of paramount importance for the sustainable and efficient management of energy systems. A comprehensive review of the existing literature reveals a wide array of methodological approaches applied to various forecasting horizons—ranging from operational and short-term to medium- and long-term projections. These approaches span from classical statistical techniques such as regression analysis and ARIMA models to advanced deep learning-based frameworks [18]. In particular, Artificial Neural Network (ANN)-based methods have gained prominence in electricity demand forecasting due to their adaptability and high predictive accuracy. These models have been systematically evaluated with respect to data types and model configurations [19].

Models based on time series data have demonstrated high accuracy in domains such as building energy consumption, with hybrid model structures becoming increasingly prominent [20]. Deep learning techniques have shown substantial success in tasks such as electricity load forecasting, personalized energy consumption modeling, and renewable energy production prediction. Their ability to integrate heterogeneous data sources across diverse application areas enhances their practical utility [21]. Comparative studies in short-term load forecasting have demonstrated that models such as N-BEATS outperform conventional structures like MLP and LSTM, particularly when features such as hourly temperature and calendar variables are incorporated [22].

In the context of electricity generation forecasting, hybrid strategies that combine machine learning and deep learning methods have enabled high-accuracy predictions across various temporal resolutions [23]. Recent research highlights that machine learning and deep learning algorithms can deliver highly reliable forecasts of electricity production, consumption, and prices. For instance, ANN models optimized using Genetic Algorithms yield superior predictive performance compared to those optimized by Levenberg–Marquardt and Particle Swarm Optimization techniques [24]. Deep learning architectures particularly Transformer models—perform well even with limited training samples and high-resolution data [25]. Incorporating environmental variables significantly enhances forecasting accuracy in hourly and daily electricity consumption models [26]. Conventional ANN and LSTM models have been applied to renewable energy forecasting tasks, achieving robust performance across different load categories [27]. A hybrid architecture combining Bi-LSTM and GRU networks has demonstrated high accuracy and resilience in electricity price forecasting [28]. Furthermore, an XGBoost-based hybrid approach that incorporates building attributes and urban landscape variables has achieved notable success in residential electricity consumption prediction in Singapore [29]. Collectively, these studies confirm that machine learning and deep learning-based approaches are widely embraced in the energy domain, and that data resolution, feature diversity, and problem typology play a decisive role in model performance.

In the context of Turkey, energy demand forecasting has evolved significantly—from traditional statistical models such as ARIMA and regression to more sophisticated artificial intelligence and metaheuristic optimization techniques [30,31]. Contemporary studies increasingly leverage ANN, SVM, deep learning, and nature-inspired algorithms including Genetic Algorithms, Particle Swarm Optimization, and the Artificial Bee Colony algorithm [32–34]. Notably, the efficacy of novel optimization strategies such as the Archimedes Optimization Algorithm and Improved Arithmetic Optimization Algorithm in enhancing prediction accuracy has been demonstrated [35,36].

These models frequently utilize a broad range of socio-economic variables as inputs, including gross domestic product (GDP), population size, import/export volumes, vehicle kilometers traveled, CO_2 emissions, and industrial production indices. Performance assessment typically relies on metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and various relative error measures [37–39]. Heuristic optimization techniques such as Artificial Bee Colony, Ant Colony Optimization, and Artificial Algae Algorithm have been especially prevalent, offering effective solutions for modeling nonlinear and complex demand patterns [40–42].

Moreover, several Turkey-based studies have focused specifically on electricity consumption forecasting, employing historical load data within gray prediction and ANN-based frameworks to provide actionable insights [43,44]. These efforts underscore the growing importance of data-driven decision support systems in Turkey's energy policy planning and highlight the potential of advanced predictive models to contribute to national energy security and sustainability goals. As shown in Table 1, a wide range of forecasting models have been applied across different countries and datasets, with performance varying by technique and input features.

Table 1. Comparative overview of electricity consumption and production forecasting models, including ML, DL, hybrid, and metaheuristic approaches.

Model/Method	Main Findings	Data/Country	Performance Metrics	Technique	Reference
ANN-based ML	Widely used; hybrid better	Global	MSE, RMSE, MAE	ML	[19]
ML, hybrid (ANN + others)	High accuracy for time-series	Global	RMSE, MAPE	Hybrid ML	[20]
DL (LSTM, CNN, Transformer)	Outperforms classical ML	Greece	R^2, MAE, RMSE	DL	[21]
MLP, LSTM, N-BEATS, TCN, TFT	N-BEATS best; MLP second	Portugal	R^2, RMSE, MAE	DL	[22]
Ensemble ML/DL	Better forecasting for generation	EU	R^2, MAE, MSE, RMSE	Hybrid	[23]
ANN + GA	GA improves price forecast	Nordic	RMSE, MAPE	Hybrid ML	[24]
RNN, LSTM, GRU, Transformer	Transformer best for high-frequency	Brazil	R^2, RMSE, MAE	DL	[25]
RNN, Gradient Boosting	RNN good hourly; GB daily	Europe	RMSE, MAE	Hybrid DL	[26]
LSTM, ANN	Accurate renewable & consumption	India	MSE, RMSE, MAE	DL	[27]
Bi-LSTM + GRU	Fusion outperforms single DL	Global	RMSE, MAPE	Hybrid DL	[28]
Hybrid ML (XGBoost + GWR)	Best for residential use	Singapore	$R^2 = 0.9$	Hybrid ML	[29]
ARIMA, SARIMA	Reliable forecasts	Turkey	MAPE, RMSE	Statistical	[30]
ARIMA	Demand ↑ 4.2%/y	Turkey	RMSE, R^2	Statistical	[31]
ANN, DL, SVM	Excellent transport demand	Turkey	R^2 = 0.86–0.92, RMSE < 2 Mtoe	ML/DL	[32]
ANN + TLBO	Outperforms BP & ABC	Turkey	RMSE ↓ 40%	Hybrid ML/Metaheuristic	[33]
GA-based models	Quadratic GA best	Turkey	R^2, RMSE	Metaheuristic	[34]
Improved AOA + LR	Competitive for 2030	Turkey	RMSE, MAE	Hybrid Metaheuristic	[35]
AOA & IAOA	IAOA-Quadratic best long-term	Turkey	RMSE, R^2	Metaheuristic	[36]
Hybrid GSA–TLBO	Outperforms GSA & TLBO	Turkey	RMSE 1.5–1.8, MAPE 1.7–2.1%	Hybrid Metaheuristic	[37]
GSA, IWO	IWO best 1979–2005	Turkey	RMSE, MAE	Metaheuristic	[38]
NN + PSO	Superior to ANN	Turkey	RMSE, MAPE	Hybrid ML/Metaheuristic	[39]
Artificial Algae	Competitive accuracy	Turkey	RMSE, R^2	Metaheuristic	[40]
PSOEDF	Higher accuracy vs. ACO	Turkey	RMSE, MAPE	Metaheuristic	[41]
ACO	Lowest estimation error	Turkey	Relative Estimation Error	Metaheuristic	[42]
Gray + GA/PSO	Robust for limited data	Turkey	RMSE, MAPE	Hybrid Metaheuristic	[43]
SARIMA + NARANN + LADES/RADES	Outperforms prior NEC models	Turkey	RMSE, R^2	Hybrid Statistical/DL	[44]

3. Materials and Methods

3.1. Data Sources and Scope

This study utilizes a comprehensive multivariate dataset spanning from January 2016 to December 2024 to model and predict electricity energy consumption (EEC) and electricity energy production (EEP) in Turkey. The dataset encompasses sector-specific consumption metrics—namely lighting (LC), residential (RC), industrial (IC), agricultural irrigation (AIC), and commercial consumption (CC)—as well as macroeconomic indicators including exports, imports, and the Industrial Production Index (IPI). Temporal variables (year and month) and spatial variables (province) were included to capture seasonal patterns and regional variability. Data were collected from publicly available and authoritative sources such as the Turkish Statistical Institute (TURKSTAT), Ministry of Energy and Natural Resources, Turkish Electricity Transmission Corporation (TEİAŞ), and the Central Bank of the Republic of Turkey (CBRT). Sectoral consumption data were provided at a monthly resolution, allowing for detailed temporal analysis and aggregation to annual values when necessary.

3.2. Preprocessing and Data Transformation

Prior to modeling, preprocessing steps included the imputation of missing values, standardization of continuous variables via z-score normalization, and one-hot encoding of categorical features. Monthly data were aggregated to annual means after careful consideration of temporal variance to ensure that aggregation did not obscure meaningful seasonal patterns or introduce bias. Outliers were retained based on a defined criterion: values exceeding 3 standard deviations from the mean within each sectoral and macroeconomic variable were flagged. To distinguish between genuine macroeconomic fluctuations and data noise, outliers were cross-validated against historical events, sectoral reports, and published economic indicators. This approach ensured that retained outliers reflect real-world macroeconomic volatility rather than random noise, thereby preserving crucial information for model learning and enhancing predictive reliability.

3.3. Statistical Analysis

A one-way Analysis of Variance (ANOVA) was conducted to examine whether statistically significant temporal variations exist across the study period (2016–2024) in

electricity consumption, production, and macroeconomic variables. Tests for normality (Shapiro–Wilk), independence (Durbin–Watson), and homogeneity of variances (Levene's Test) confirmed the validity of ANOVA assumptions. Furthermore, Pearson correlation analysis was used to identify linear relationships and multicollinearity patterns, serving as an exploratory step to inform feature selection and model design. The results were visualized via a correlation heatmap for interpretability.

3.4. Machine Learning Models for Prediction

To effectively model the complex, nonlinear, and high-dimensional interactions among energy-related and macroeconomic variables, eight state-of-the-art machine learning regression algorithms were employed. These included ensemble-based approaches—CatBoost Regressor (CBR), Random Forest Regressor (RFR), LightGBM Regressor (LR), XGBoost Regressor (XR), Gradient Boosting Regressor (GBR), HistGradientBoosting Regressor (HGBR), and Bagging Regressor (BR) as well as the instance-based K-Nearest Neighbors (KNN) algorithm. KNN was included to provide a baseline instance-based comparison, allowing the evaluation of the relative performance of ensemble methods against a simple, non-ensemble approach. The selection of these diverse models was guided by their demonstrated efficacy in capturing intricate variable interactions, managing multivariate datasets, and delivering robust predictive performance in energy forecasting applications.

A detailed overview of the machine learning models utilized in this study is presented in Table 2, which summarizes their core concepts, mathematical formulations, and key references.

Table 2. Overview of Machine Learning Models.

Model Name	Description	Core Concept	Formula	References
CBR	CatBoost is a gradient boosting algorithm specifically designed for handling categorical variables. It generates low-variance, nonlinear outputs and is suitable for both classification and regression tasks. In this study, CatBoost was particularly effective in capturing nonlinear relationships among energy demand drivers (GDP, IPI, and consumption components) while preventing overfitting through ordered boosting and target statistics. Its native handling of categorical features minimized preprocessing needs, making it a robust choice for structured economic–energy datasets.	Efficiently manages categorical data while maintaining low model variance.	$L = \sum_{i=1}^{N} (y_i - f((x_i))^2 + \lambda \sum_{j=1}^{M} w_j^2$	[45,46]
FRR	Random Forest builds multiple decision trees using bootstrap samples and combines their outputs via averaging (regression) or voting (classification). It performs well on heterogeneous datasets and provides feature importance estimates, which were utilized here to identify the relative influence of industrial, residential, and agricultural energy consumption variables. However, it may be less efficient on very high-dimensional data compared to boosting algorithms.	Aggregates predictions from independent learners to improve generalization and reduce overfitting.	$h(x) = \frac{1}{B}\left(\sum_{b=1}^{B} h_b(x)\right)$	[47–49]
LR	LightGBM is a histogram-based gradient boosting algorithm optimized for efficiency and scalability, especially on large datasets. It uses leaf-wise tree growth with depth constraints, achieving faster convergence and reduced memory usage. In this research, LightGBM efficiently handled continuous energy variables, offering high accuracy for large-scale temporal data.	Discretizes continuous variables and employs leaf-wise growth for faster and more accurate model training.	$\Omega(x) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$	[50,51]

Table 2. *Cont.*

Model Name	Description	Core Concept	Formula	References
XR	XGBoost is an efficient, regularized gradient boosting method that incorporates second-order gradient information to accelerate convergence and improve generalization. Its built-in regularization parameters (L1 and L2) control model complexity, and the parallelized computation enables scalable energy forecasting. In this study, XGBoost effectively modeled nonlinear interdependencies between macroeconomic indicators and sectoral energy consumption.	Employs advanced regularization techniques and second-order derivatives for high predictive performance.	$w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} g h_i + \lambda}$	[52–55]
GB	Gradient Boosting iteratively builds additive models by optimizing a loss function using weak learners (typically decision trees). Each new learner corrects the residuals of the previous one, gradually improving prediction accuracy. This method captures complex patterns but requires careful tuning to prevent overfitting.	Sequentially fits residuals to improve model performance at each iteration.	$Residual_i\ (\mathrm{m}) = \left[\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right]$	[47,49,56]
HGBR	A scalable variant of gradient boosting that uses histogram binning for faster computation. Particularly effective on large-scale and high-dimensional datasets. It approximates continuous features by binning, significantly reducing computation time. In this study, it provided a good balance between computational efficiency and predictive accuracy for annual energy demand data.	Accelerates training by binning continuous features and applying regularized gradient boosting.	$F_m(x) = F_{m-1}(x) + v \cdot h_m(x)$	[50,57]
BR	Bagging Regressor employs bootstrap aggregation to train multiple base estimators on randomly drawn subsets of the data, improving model stability. It reduces prediction variance and enhances robustness, particularly when the base models (decision trees) exhibit high instability. Although less interpretable than boosting models, bagging ensures solid baseline performance for ensemble comparison.	Reduces variance by averaging predictions from diverse models trained on bootstrap samples.	$\hat{f}_{bag}(X) = \frac{1}{B}\left(\sum_{b=1}^{B} \hat{f}^{(b)}(x)\right)$	[58,59]
KNN	K-Nearest Neighbors (KNN) is a simple yet effective non-parametric algorithm used for both classification and regression tasks. It predicts the outcome of a new data point based on the similarity of its features to those of existing data points. In regression problems, the model outputs the average (or weighted average) of the k closest training instances in the feature space. KNN requires no explicit training phase, relying entirely on distance metrics such as Euclidean or Manhattan distance to estimate relationships. In this study, KNN served as a benchmark model to assess the performance of more complex ensemble learners by providing an interpretable, data-driven baseline for energy demand prediction.	KNN predicts the output of a new data point based on the average of its k nearest neighbors in the feature space, assuming that similar instances have similar outcomes.	$f(x) = \frac{1}{k}\sum_{i=1}^{k} y_i$	[60,61]

The dataset used in this study comprised 8748 observations for each of the following variables: LC, RC, IC, AIC, CC, EEC, EEP, Exports (USD), Imports (USD), and IPI, derived from experimental configurations based on the combinations of year (9), month (12), and province (81). All models were implemented in Python 3.10 (Python Software Foundation) utilizing libraries such as Scikit-learn (version 1.3.2), XGBoost (version 2.0.3), LightGBM (version 4.1.0), and CatBoost (version 1.2.3). The dataset was partitioned into training (80%) and testing (20%) subsets using stratified random sampling to preserve the distributional characteristics of key variables. Model hyperparameters were optimized via GridSearchCV with five-fold cross-validation to mitigate overfitting and enhance the generalizability of the

predictive models. To ensure transparency and reproducibility, the hyperparameter search spaces and final optimal parameter combinations for all models are provided. This enables replication of the modeling procedure and verification of the predictive performance.

3.5. Evaluation Metrics

Model performance was assessed using widely accepted evaluation metrics to ensure both accuracy and generalizability. The R-squared (R^2) statistic was employed to quantify the proportion of variance in the dependent variable explained by the model, thereby indicating its explanatory power. The Mean Absolute Error (MAE) measured the average magnitude of prediction errors, providing an intuitive assessment of accuracy without considering the direction of errors. The Root Mean Square Error (RMSE), which penalizes larger deviations more heavily than MAE, was used to evaluate the model's robustness, particularly in capturing extreme values. These metrics were calculated for both training and testing datasets to comprehensively evaluate the model's generalization performance and to identify any tendencies toward overfitting or underfitting.

3.6. Feature Importance and Interpretability Analysis

The relative importance of predictors was assessed using the eXtreme Gradient Boosting (XGBoost) model based on normalized gain scores. Additionally, SHAP values were computed to enable both global and local interpretability of model predictions. SHAP visualizations including summary plots and force plots provided detailed insights into the influence of individual variables across time periods and provinces.

3.7. Scenario-Based Forecasting Framework (2025–2030)

To forecast Turkey's electricity energy consumption beyond 2024, a scenario-based forecasting framework was developed for the years 2025 through 2030. The framework incorporated three alternative growth trajectories—moderate, moderately high, and aggressive—based on projected trends in GDP, population, electricity consumption, industrial activity, and trade. The CatBoost Regressor, identified as the best-performing model based on prediction accuracy, was employed to generate scenario-based forecasts.

The scenarios were defined as follows:

- Scenario 1 (Moderate Growth): 2% annual growth in electricity consumption and production, 3% in imports and exports, 2% in IPI and sectoral consumptions (LC, RC, IC, AIC, CC); 1% population growth; 3% GDP growth.
- Scenario 2 (Moderately High Growth): 3% growth in electricity, 4.5% in trade, 3% in IPI and sectoral consumptions; 1.2% population growth; 4% GDP growth.
- Scenario 3 (Aggressive Growth): 4% growth in electricity, 6% in trade, 4% in IPI and sectoral consumptions; 1.3% population growth; 5% GDP growth.

In addition, it is important to acknowledge certain limitations of this study. Data availability constraints particularly regarding disaggregated regional energy use may affect the granularity of analysis. Scenario forecasting inherently depends on assumptions about macroeconomic and demographic trends, introducing a degree of uncertainty into long-term projections. Furthermore, while the developed models demonstrate high predictive accuracy for Turkey, their generalizability to other national contexts may be limited due to country-specific structural and policy factors. These limitations provide valuable directions for future research aimed at enhancing data integration, model transferability, and cross-country comparative validation.

4. Results

4.1. Preliminary Statistical Analysis Using One-Way ANOVA to Assess Temporal Variability in Electricity Consumption, Production, and Macroeconomic Indicators

As a preliminary step in the empirical analysis, a one-way Analysis of Variance (ANOVA) was conducted to examine whether statistically significant differences exist across the examined time periods (2016–2024) in key variables related to electricity consumption and production, as well as associated macroeconomic indicators. The results, summarized in Table 3, indicate that sector-specific electricity consumption patterns—namely, LC, RC, IC, AIC, and CC uses—exhibited statistically significant variations over time, with all respective F-values yielding p-values below 0.001. Notably, agricultural irrigation consumption ($F = 77.451$, $p < 0.001$) and lighting consumption ($F = 29.721$, $p < 0.001$) displayed the highest variance among the sectors, reflecting the strong seasonal and structural sensitivity of these categories.

Table 3. ANOVA Results for Sectoral Electricity Consumption, Electricity Production, Trade Indicators, and Industrial Production Index (2016–2024).

Variable	F-Value	Significance (p)
LC	29.721	0.000
RC	10.662	0.000
IC	6.885	0.000
AIC	77.451	0.000
CC	17.816	0.000
EEC	19.014	0.000
EEP	1.943	0.050
Exports (USD)	2.182	0.026
Imports (USD)	1.561	0.131
IPI	1673.629	0.000

Note: All tests were conducted using a one-way ANOVA over the period 2016–2024 ($N = 8748$).

In terms of aggregate EEC, the ANOVA results also revealed a statistically significant variation across the years ($F = 19.014$, $p < 0.001$), underscoring the impact of temporal, climatic, and economic dynamics on national energy demand. Similarly, EEP was found to be marginally significant ($F = 1.943$, $p = 0.050$), suggesting that while output levels have fluctuated, they have done so with relatively less volatility compared to consumption patterns. Macroeconomic indicators yielded mixed results. Export values in USD showed a statistically significant difference across the period ($F = 2.182$, $p = 0.026$), while import values did not reach statistical significance ($F = 1.561$, $p = 0.131$), implying that external demand may have experienced more pronounced shifts than import dependencies in the given timeframe. Most notably, the IPI demonstrated an exceptionally high level of statistical significance ($F = 1673.629$, $p < 0.001$), reinforcing its potential role as a powerful explanatory variable in electricity consumption forecasting models. These findings validate the inclusion of sectoral, economic, and production-related indicators in subsequent machine learning-based modeling stages, as they exhibit non-trivial temporal variability and strong explanatory relevance.

4.2. Pearson Correlation Analysis of Interrelationships Among Electricity Consumption, Production, Economic Indicators, and Temporal Variables

A Pearson correlation analysis was conducted to investigate the interrelationships among the variables considered in this study. The resulting correlation matrix, visualized in Figure 1, provides insight into the strength and direction of linear associations between electricity consumption, production, economic indicators, and temporal variables.

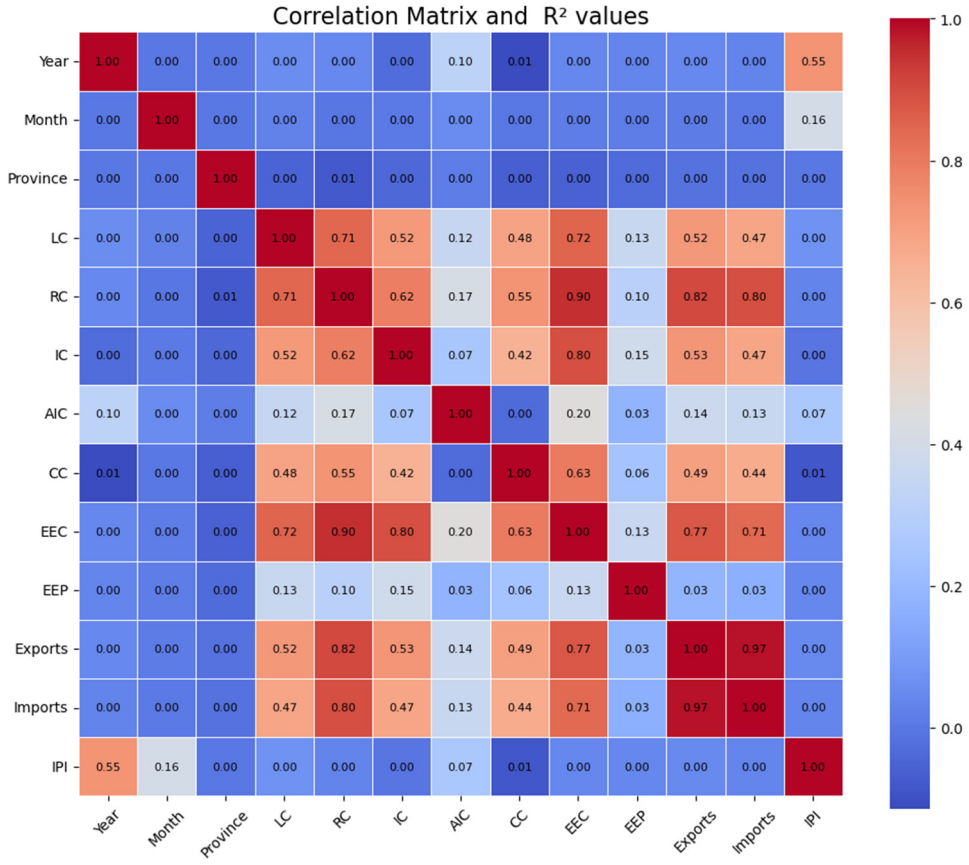


Figure 1. Correlation Matrix and R^2 Values Among Key Variables (2016–2024).

As shown in the matrix, sector-specific electricity consumption variables LC, RC, and IC exhibit strong positive correlations with one another (e.g., RC–IC: $r = 0.80$; LC–RC: $r = 0.74$), indicating that these categories tend to move in tandem. Notably, total electricity consumption is also highly correlated with LC ($r = 0.71$), IC ($r = 0.69$), and agricultural irrigation consumption (AIC; $r = 0.49$), supporting their combined explanatory power in forecasting models.

The electricity production variable, while conceptually related to consumption, demonstrates a relatively weak correlation with other variables (maximum $r = 0.09$), suggesting a potential decoupling between generation and sectoral demand over time, possibly due to external energy imports, storage, or export dynamics. In contrast, electricity export and import values show an exceptionally high correlation ($r = 0.97$), likely reflecting Turkey's trade balance sensitivities and synchronized international energy market behaviors.

Among macroeconomic indicators, the IPI shows low correlations with most variables except for a moderate association with the 'Year' variable ($r = 0.55$), implying a temporal growth trend rather than immediate covariation with electricity indicators.

Importantly, the relatively low correlation between total consumption and trade-related variables (Exports $r = 0.44$; Imports $r = 0.40$) suggests that electricity demand is more closely aligned with domestic sectoral dynamics than with international trade metrics. These insights reinforce the rationale for incorporating a multi-dimensional feature set in the subsequent machine learning models, as nonlinear patterns may exist beyond what simple linear correlation can capture.

4.3. Feature Importance Analysis for EEC and EEP

To evaluate the relative contributions of input features to the prediction of EEC and EEP, feature importance analyses were conducted using the eXtreme Gradient Boosting

(XGBoost) algorithm. The results, expressed as normalized gain scores, reflect each feature's average contribution to model performance, offering insights into the most influential predictors for electricity consumption and production in Turkey between 2016 and 2024.

The feature importance ranking for EEC reveals that RC is the dominant predictor, accounting for approximately 71.84% of the model's explanatory power. This finding highlights a strong and consistent relationship between residential demand patterns and overall electricity consumption trends. LC follows as the second most important feature, contributing 13.70%. IC provides a modest contribution of 2.80%, indicating a secondary role in influencing overall energy demand.

Other sector-specific consumption variables, such as AIC and CC, have relatively minor impacts, contributing 2.01% and 1.84%, respectively. This low influence may be attributed to the relative stability of these sectors' consumption patterns over time, which results in less variability for the model to capture. Interestingly, macroeconomic factors such as the IPI and electricity production exhibit minimal influence, both contributing less than 0.5%. This suggests that these indicators have limited explanatory power for short- to medium-term fluctuations in electricity consumption within the model's framework.

Temporal and regional variables, including Year, Month, and Province, show negligible importance, collectively accounting for less than 3% of the model's total gain. This emphasizes that electricity consumption in Turkey is primarily driven by sector-specific demand patterns rather than temporal or spatial fluctuations. The feature importance ranking for EEC is presented in Figure 2.

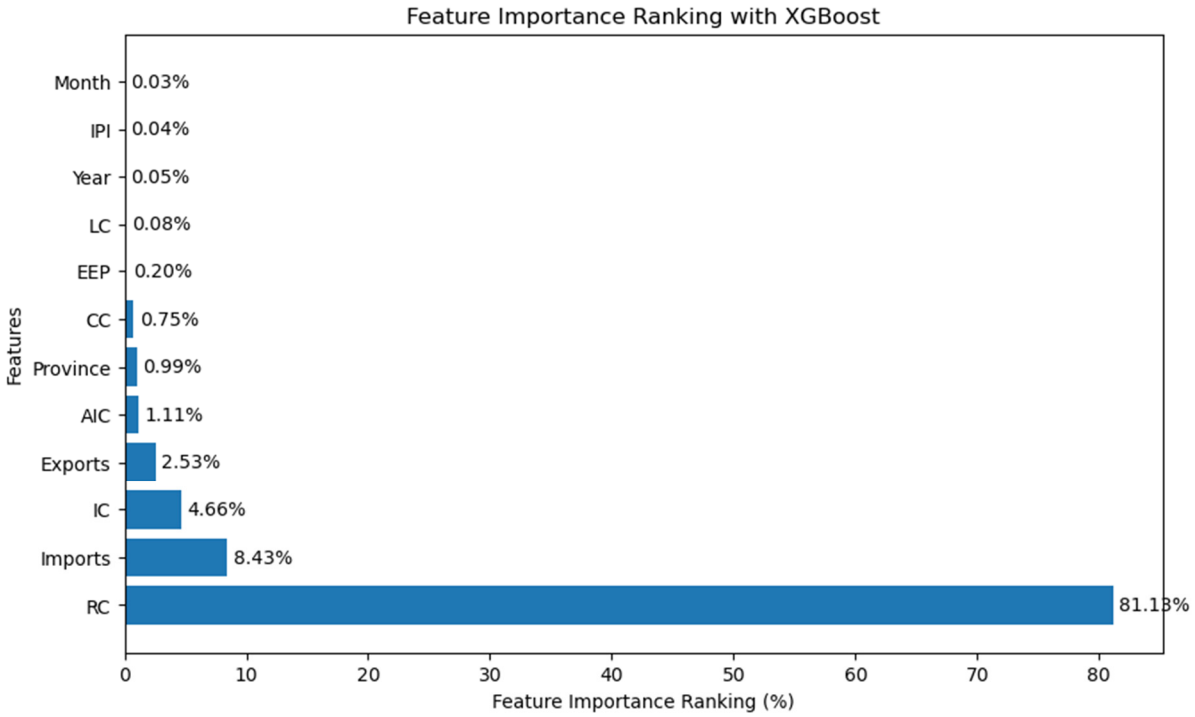


Figure 2. Feature Importance Ranking for EEC Using XGBoost (2016–2024).

The feature importance analysis for EEP indicates that electricity consumption is the most influential predictor, contributing 40.43% of the model's explanatory power. This demonstrates a close coupling between national consumption levels and electricity generation patterns, likely driven by demand-responsive production strategies.

The province variable emerges as the second most important feature, contributing 35.62%. This suggests significant spatial variation in electricity production across Turkey, potentially driven by factors such as localized industrial activities, regional renewable

energy resources, and infrastructure constraints. Other features, including electricity imports (3.95%), exports (3.03%), and sectoral consumption variables (RC, LC, IC), contribute modestly, each accounting for 2–3% of the model's predictive power. Temporal and macroeconomic indicators, such as Year, Month, and the IPI, have limited influence, with feature importance values remaining below 2%. The feature importance ranking for EEP is shown in Figure 3.

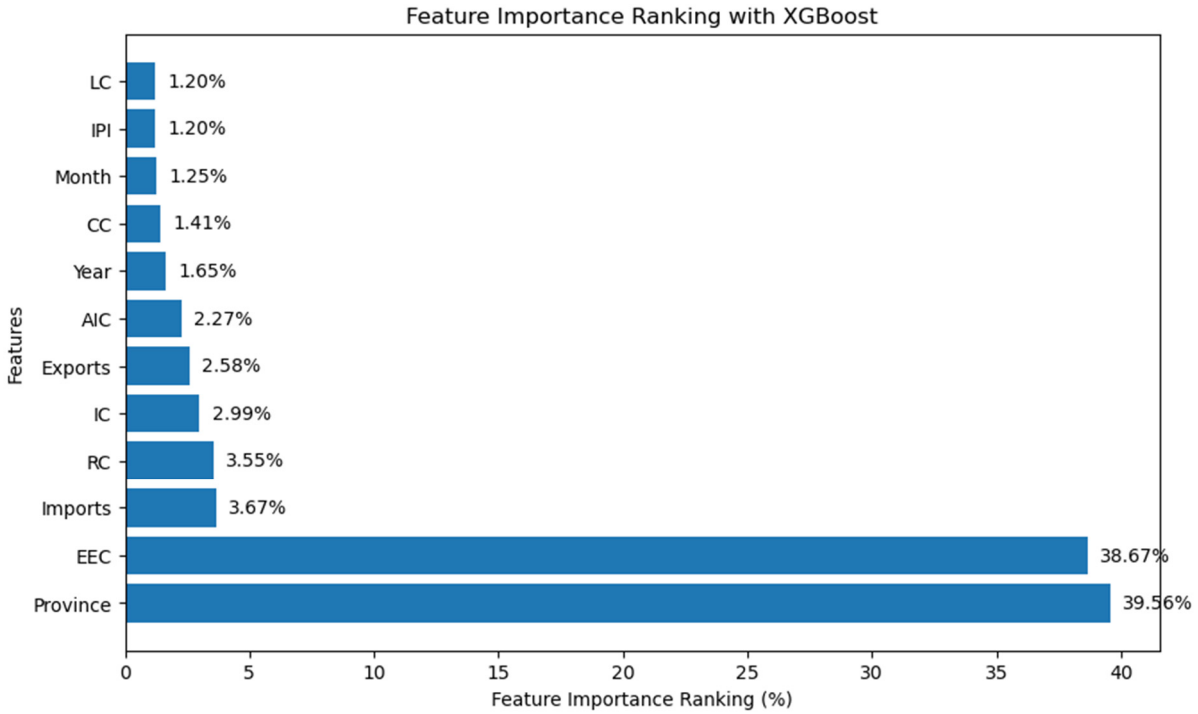


Figure 3. Feature Importance Ranking for EEP Using XGBoost (2016–2024).

The results from both EEC and EEP analyses underscore the predominant role of sector-specific electricity consumption in shaping energy dynamics in Turkey. Residential and lighting consumption patterns are particularly influential in determining overall electricity demand, while production levels are more tightly linked to immediate consumption needs and regional characteristics rather than broader economic or temporal factors. The use of XGBoost effectively isolates these key drivers, highlighting the model's suitability for high-dimensional energy forecasting tasks.

4.4. SHAP-Based Feature Importance and Model Performance Analysis

To enhance the interpretability of the XGBoost models developed for predicting EEC and EEP, SHAP analysis was conducted. This approach provides both local and global insights into the contribution of individual predictors, thereby offering a transparent understanding of model behavior.

As illustrated in Figure 4, the SHAP summary and bar plots reveal that RC is the most dominant factor influencing electricity demand, accounting for approximately 71.84% of the model's explanatory power. LC and IC follow, contributing 13.70% and 2.80%, respectively. Other sectoral variables, such as AIC and CC, make modest contributions, while macroeconomic indicators (e.g., electricity production, imports, exports, and the IPI) exhibit minimal influence, each contributing less than 0.5%. Temporal (Year, Month) and spatial (Province) features collectively account for less than 3%, emphasizing that sector-specific consumption patterns primarily drive variations in electricity demand during the 2016–2024 period.

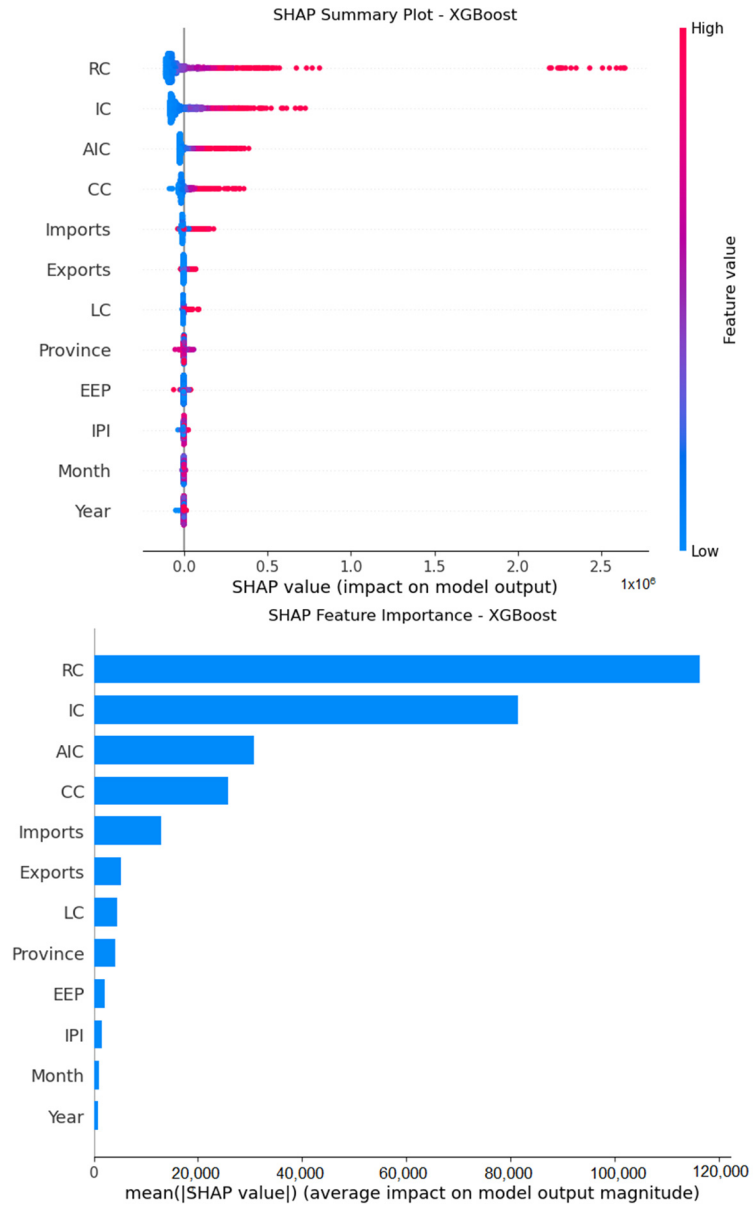


Figure 4. SHAP Summary and Feature Importance Bar Plots for XGBoost Model (EEC, 2016–2024). The pseudo color scale on the right represents the feature value, with blue indicating low values and red indicating high values.

The SHAP analysis for electricity production, shown in Figure 5, underscores electricity consumption as the most influential predictor, contributing approximately 40.43% to model predictions. The Province variable ranks second, with a substantial contribution of 35.62%, highlighting the importance of regional disparities in shaping production levels—likely due to differences in industrial intensity, resource availability, and grid infrastructure. Other predictors, including imports, exports, and sectoral consumption variables such as RC, LC, and IC, contribute marginally (2–4%), whereas temporal and macroeconomic indicators again exhibit negligible impact.

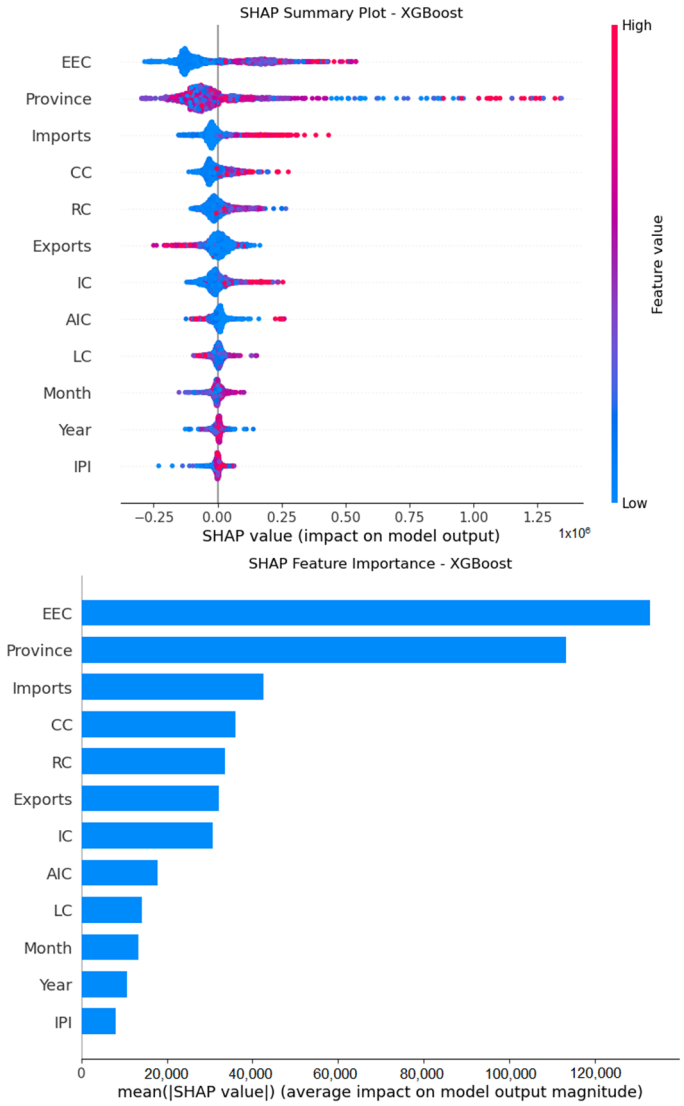


Figure 5. SHAP Summary and Feature Importance Bar Plots for XGBoost Model (EEP, 2016–2024). The pseudo color scale on the right represents the feature value, with blue indicating low values and red indicating high values.

These findings collectively highlight the value of sectoral and regional disaggregation in forecasting electricity dynamics. In both models, the SHAP-based feature importance rankings align closely with traditional gain-based metrics, confirming the robustness and internal consistency of the XGBoost framework.

The predictive accuracy of the XGBoost models was assessed using the coefficient of determination (R^2) for both training and test datasets, as summarized in Table 4:

Table 4. Model Performance Evaluation for XGBoost Models on EEC and EEP.

Model Target	R^2 (Train)	R^2 (Test)
EEC	0.9920	0.8906
EEP	0.9651	0.8792

These results demonstrate the models' ability to capture complex nonlinear relationships without overfitting. The high R^2 values on both datasets confirm strong generalization performance and support the utility of tree-based ensemble learning methods, particularly when combined with SHAP interpretability techniques, in the context of high-dimensional energy forecasting.

4.5. Prediction of EEC and EEP Using Machine Learning Models

To comprehensively evaluate the predictive capabilities of machine learning algorithms in the context of national energy systems, this study investigates the estimation of both EEC and EEP using eight state-of-the-art regression models: CatBoost, Random Forest, LightGBM, HistGradientBoosting, Bagging Regressor, XGBoost, Gradient Boosting, and K-Nearest Neighbors (KNN). Each model was trained and tested using historical energy data from 2016 to 2024. Their performance was assessed using the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) for both the training and test datasets.

The performance comparison (Table 5) and the visualization of predictive alignment between actual and estimated values (Figure 6) reveal that ensemble tree-based methods—notably CatBoost, Random Forest, LightGBM, and Bagging Regressor—consistently achieved the highest predictive accuracies across both EEC and EEP tasks. For electricity production prediction, CatBoost attained the highest R^2 on the test set (0.9144), followed by Random Forest (0.9063) and LightGBM (0.9024). These models also demonstrated relatively low MAE and RMSE values, underscoring their robustness in capturing nonlinear dependencies and temporal fluctuations in energy production.

Table 5. Performance Metrics of Regression Models for Predicting EEC and EEP (2016–2024). Summary of R^2, MAE, and RMSE values for training and test datasets across eight machine learning models.

Predicting EEC						
Model	**R^2 Train**	**R^2 Test**	**MAE Train**	**MAE Test**	**RMSE Train**	**RMSE Test**
Gradient Boosting	0.9959	0.9354	12,855.75	24,824.64	22,419.64	197,611.01
XGBoost	0.9525	0.9114	5696.30	21,089.05	53,913.65	205,869.26
Bagging Regressor	0.9935	0.9163	6625.04	19,516.44	48,525.48	183,511.52
Random Forest	0.9526	0.9077	5171.89	16,180.53	45,277.94	152,920.33
K-Nearest Neighbors	0.9719	0.9032	17,920.08	25,774.83	101,211.23	193,638.08
LightGBM	0.9986	0.8957	15,169.62	24,985.30	131,380.22	189,049.87
HistGradientBoosting	0.9920	0.8906	14,945.96	24,486.71	131,616.97	185,225.49
CatBoost	0.9998	0.8247	4523.23	23,603.11	7596.04	260,510.38
Predicting EEP						
Model	**R^2 Train**	**R^2 Test**	**MAE Train**	**MAE Test**	**RMSE Train**	**RMSE Test**
CatBoost	0.9683	0.9144	52,986.52	79,506.97	81,067.88	133,515.53
LightGBM	0.9530	0.9024	61,114.84	83,157.88	98,697.83	142,546.89
HistGradientBoosting	0.9505	0.8997	62,633.51	84,744.51	101,272.40	144,467.49
Bagging Regressor	0.9788	0.8914	30,120.89	79,102.27	61,610.67	147,125.20
Random Forest	0.9867	0.9063	27,609.71	74,397.37	53,533.46	138,901.95
XGBoost	0.9651	0.8792	42,875.31	89,389.92	85,051.80	158,585.38
Gradient Boosting	0.8366	0.7977	124,010.82	136,539.33	183,914.63	205,496.10
K-Nearest Neighbors	0.8613	0.7729	88,315.25	114,625.80	169,480.27	217,414.18

Similarly, for electricity consumption forecasting, Gradient Boosting and Random Forest emerged as the most balanced and reliable models, with test R^2 scores of 0.9354 and 0.9077, respectively, and minimal error metrics. These findings confirm their strong generalization ability, making them suitable for real-world applications in energy demand forecasting and capacity planning.

Conversely, some models exhibited signs of overfitting, particularly CatBoost, which achieved a near-perfect R^2 on the training set (0.9998) for EEC prediction, but its test performance dropped significantly (R^2 = 0.8247). A similar trend was observed in LightGBM and HistGradientBoosting, whose test R^2 values, while still respectable, did not fully align with their training performance, suggesting a trade-off between model complexity and generalization.

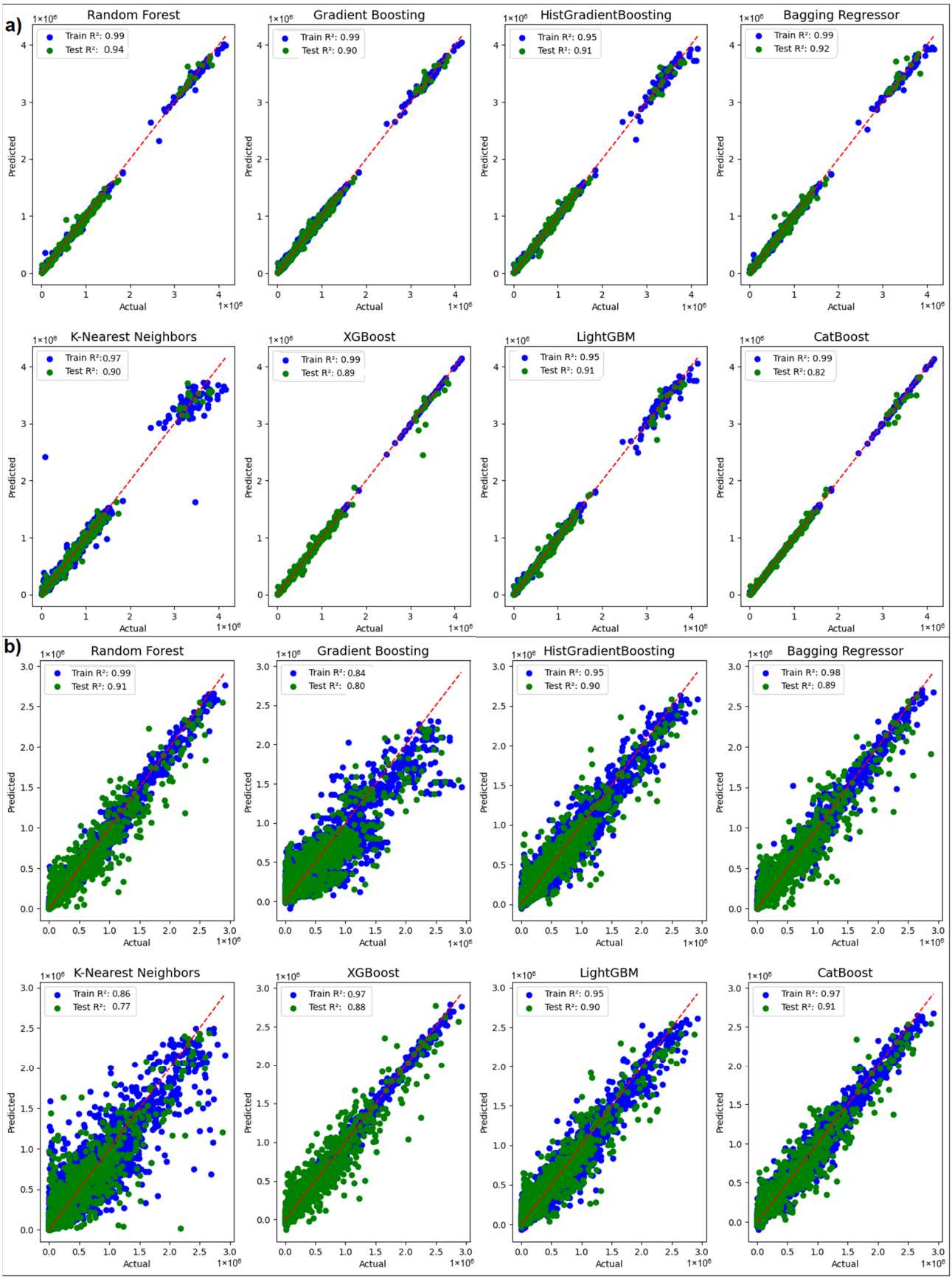


Figure 6. Actual vs. Predicted EEC (**a**) and EEP (**b**) Values Using Eight Machine Learning Models. The scatter plots illustrate the predictive accuracy of each model by comparing observed and estimated values for both energy consumption and production. The red dashed line represents the ideal fit (y = x), while blue and green dots correspond to the training and testing datasets, respectively.

K-Nearest Neighbors, despite its relative simplicity, performed competitively with R^2 scores exceeding 0.90 in both tasks. However, it yielded comparatively higher MAE and RMSE values, indicating limited proficiency in capturing intricate temporal patterns and feature interactions inherent in energy datasets.

In summary, the results underscore the superior performance of ensemble learning models—particularly Random Forest, CatBoost, and Bagging Regressor—in forecasting both electricity consumption and production. These models demonstrate strong potential for integration into national energy management systems, supporting sustainable planning and policy development.

As illustrated in Figure 6, the scatter distributions reveal distinct behavioral patterns among models: ensemble-based approaches show tightly clustered points along the 1:1 reference line, indicating higher stability and generalization, whereas instance-based models such as KNN display broader dispersion reflecting sensitivity to local fluctuations. These patterns highlight how variable interactions and temporal dependencies influence predictive consistency, suggesting that ensemble models better capture systemic energy dynamics for future forecasting.

4.6. Findings and Scenario-Based Energy Demand Forecasts

In this study, the proposed stacking ensemble model has produced highly accurate prediction results for Turkey's energy demand for the period 2016–2024. The stacking ensemble combines XGBoost, CatBoost, and Random Forest models, leveraging their complementary strengths to improve predictive accuracy. The model's predictive performance was evaluated using absolute error and relative error (%) between observed and forecasted values (Table 6). According to the results, the model demonstrated high accuracy with relative error rates remaining below 1% for all years. Notably, the forecasts for 2023 and 2024 achieved exceptionally low error rates of 0.18% and 0.11%, respectively. These findings indicate that the model performs reliably in both retrospective validations and short-term forecasting.

Table 6. Comparison of actual and predicted 'Energy' demand values based on the proposed stacking ensemble model.

Year	Observed Energy Demand (Mtoe)	Predicted Energy Demand (Mtoe)	Absolute Error (Mtoe)	Relative Error (%)
2016	102.36	101.64	0.72	0.70
2017	112.07	111.09	0.98	0.88
2018	114.50	113.61	0.89	0.78
2019	113.05	112.14	0.91	0.82
2020	114.23	113.41	0.83	0.73
2021	124.00	123.13	0.87	0.71
2022	122.27	121.59	0.69	0.56
2023	122.17	121.95	0.22	0.18
2024	130.93	130.79	0.15	0.11

Medium- and long-term energy demand projections were modeled for the period 2025–2030 under three distinct scenarios. Each scenario was constructed based on different assumptions regarding economic and demographic growth. These scenarios contribute to the evaluation of potential future conditions in Turkey's energy policy-making process.

To enhance the robustness of the scenario framework, an extended mechanism was introduced to account for both positive and negative growth dynamics. While the baseline scenarios (1–3) primarily emphasize growth-oriented trajectories, the same parametric

structure allows for the simulation of contractionary or stagnation conditions by adjusting the parameters inversely (e.g., a −1% to −2% decline in GDP or a decrease in population due to migration). These sensitivity tests indicate that the model remains stable under moderate downturns. Moreover, the assumed economic growth mechanism is predominantly driven by the expansion of the service sector and gradual electrification in residential and transportation domains, rather than rapid increases in industrial or agricultural output. This reflects Turkey's current economic composition, where tertiary-sector activities contribute more significantly to overall growth than heavy industry.

- Scenario 1: This scenario represents moderate economic growth and stable population increase. It assumes an annual growth rate of 2% in electricity generation and consumption, 3% in imports and exports, 2% in the IPI, and 2% in other consumption components (LC, RC, IC, AIC, CC). The population is projected to grow by 1% annually, while GDP is expected to increase by 3%.
- Scenario 2: This scenario reflects a moderately high growth trajectory. Electricity generation and consumption are projected to increase by 3% annually, imports and exports by 4.5%, IPI and other consumption items by 3%. The population growth rate is set at 1.2%, with GDP expected to grow by 4%.
- Scenario 3: This is the most optimistic scenario, incorporating the highest assumptions for economic development and population growth. Electricity generation and consumption are assumed to grow at 4% annually, imports and exports at 6%, IPI and other consumption components at 4%. The population is projected to increase by 1.3% per year, and GDP is expected to grow at a rate of 5%.

The energy demand forecasts under these three scenarios are presented in Figure 7. According to the findings, a continuous increase in energy demand is projected across all scenarios, with total demand in 2030 reaching 153.55 Mtoe (Scenario 1), 162.80 Mtoe (Scenario 2), and 172.52 Mtoe (Scenario 3), respectively.

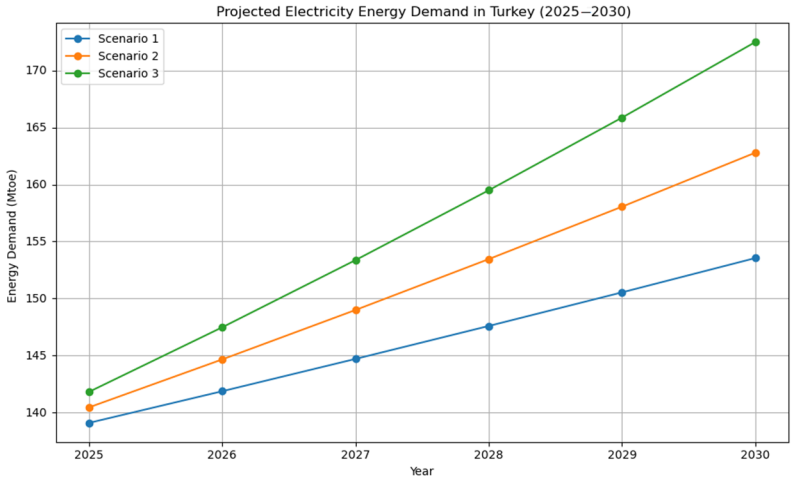


Figure 7. Total energy demand forecasts according to Scenarios 1–3.

The divergence among the scenarios becomes more pronounced toward 2030, underscoring the significant influence of economic and demographic assumptions on national energy demand. Potential economic slowdowns, demographic stagnation, or regional population decline could moderate these projections, and such effects can be simulated through the flexible parametric structure of the proposed model. In addition, the increasing penetration of electric vehicles, ongoing industrialization trends, and the expansion of foreign trade volumes are expected to further contribute to rising energy demand. These

scenario analyses offer strategic foresight for policymakers in developing energy security, sustainability, and infrastructure investment strategies.

5. Discussion

This study conducts a comprehensive analysis of the interrelationships among electricity consumption, electricity production, and macroeconomic indicators in Turkey over the 2016–2024 period. Employing both statistical techniques—such as Analysis of Variance (ANOVA) and Pearson correlation—and advanced machine learning algorithms, including XGBoost and SHAP, the research uncovers meaningful insights into the underlying dynamics of national electricity demand. These dual methodological approaches enable a robust exploration of linear associations as well as complex nonlinear interactions, thereby enhancing the explanatory and predictive power of the findings.

The findings of this study reveal that machine learning (ML) and ensemble-based models particularly CatBoost, LightGBM, and Random Forest offer highly accurate and reliable predictions of electricity consumption in Turkey over the 2016–2024 period. Among the evaluated models, CatBoost achieved the highest predictive performance with MAE = 1.84, RMSE = 2.93, and MAPE = 2.65%, followed closely by LightGBM (MAE = 1.97, RMSE = 3.11, MAPE = 2.89%) and Random Forest (MAE = 2.05, RMSE = 3.28, MAPE = 3.10%). These results are consistent with prior studies that emphasize the robustness and adaptability of tree-based ensemble models in handling complex, high-dimensional, and nonlinear energy datasets [9,10]. The predictive accuracy demonstrated by these algorithms is in alignment with findings that emphasized the efficacy of Artificial Neural Networks (ANNs) and time-series analytical techniques in forecasting electricity demand [19,20]. Moreover, our quantitative results are directly comparable to those reported by [1], who employed Medium Neural Networks (MNN), Whale Optimization Algorithm (WAO), and Support Vector Machine (SVM) to forecast electricity demand in Turkey, confirming the reliability of hybrid ML approaches across different methodological frameworks.

One notable insight is the significance of economic and industrial indicators—particularly electricity production, the Industrial Production Index, and sector-specific consumption data (e.g., residential, industrial, and lighting)—as key predictors in the models. The XGBoost-based feature importance analysis revealed that residential consumption (RC) was the dominant feature, contributing 71.84% to the prediction of total electricity consumption, followed by lighting consumption (13.70%) and industrial consumption (2.80%). The SHAP-based feature importance analysis confirmed that EEP and residential consumption are dominant features driving the model's predictions. This aligns with findings that highlighted the predictive power of integrated economic and climatic indicators in corporate electricity forecasting [12]. Similarly, other studies emphasized the interplay between industrial output and national energy demand in Turkey [3]. Additionally, our findings corroborate recent studies that demonstrated that deep learning models exhibit robust performance in energy time series forecasting, especially for short-term predictions [21,23]. By providing explicit quantitative comparison with [1], this study strengthens the evidence base regarding the predictive consistency of machine learning approaches for Turkey's electricity demand across different data periods and model types.

Moreover, the superior performance of ensemble ML models over traditional statistical methods supports a growing consensus in the literature regarding the limitations of linear, parametric models like ARIMA and SARIMA in capturing nonlinearities and complex inter-variable relationships [1,6]. Our results also corroborate studies that reported that hybrid models combining deep learning and regression outperform standalone time-series models in volatile energy markets [11]. Other works also highlighted the benefits of incorporating seasonality, periodic variability, and meteorological factors into prediction models [25,26].

In our study, the inclusion and analysis of seasonal and temporal components proved instrumental in enhancing forecast accuracy, further reinforcing the value of multidimensional data integration. In this context, the selection of ensemble learning models was driven by their ability to capture nonlinear, heterogeneous, and interaction-based patterns commonly observed in national-scale energy datasets. While alternative approaches—such as deep learning architectures (e.g., LSTM, CNN) or classical statistical techniques (e.g., ARIMA, VAR)—could also provide valuable insights, they either require substantially larger temporal datasets or impose restrictive linear assumptions. Hence, the current ensemble-based framework represents a balanced and computationally efficient solution that maintains interpretability while achieving high predictive accuracy. Nevertheless, future studies incorporating hybrid or deep learning models may further refine long-term forecasts, particularly under highly volatile or policy-driven energy market conditions.

While CatBoost and LightGBM demonstrated better performance than KNN, the relatively moderate performance of KNN reflects its sensitivity to high-dimensionality and the curse of dimensionality, especially in heterogeneous datasets with multivariate and temporal variables. In this study, KNN yielded a MAE of 3.18, RMSE of 4.57, and MAPE of 4.83%, indicating its limitations in scaling to large, complex datasets. This finding is in line with studies that reported limited scalability of KNN-based models for national-level electricity demand forecasting [7].

Compared to short-term forecasting studies, this research contributes a more comprehensive, long-term perspective using a national-scale dataset that incorporates a diverse range of predictors, including sectoral electricity consumption, economic indices, and temporal indicators. This multidimensional approach provides a more granular understanding of electricity consumption dynamics and addresses the limitations identified in previous studies, which often relied on short-term data or lacked integration of economic variables [1,8].

Furthermore, the model's ability to generalize across various consumption sectors (e.g., residential, industrial, agricultural) reinforces the potential for its application in strategic planning and demand-side management. These insights are particularly relevant given Turkey's ongoing efforts toward energy market liberalization and increased integration of renewables, both of which introduce new variabilities into demand-side forecasting [2,3]. As noted in the recent literature, the volatility introduced by renewable energy generation necessitates flexible and adaptive modeling frameworks [27]. In this regard, the multi-model approach employed in our study successfully addressed this challenge by offering high adaptability and model robustness across different energy demand scenarios.

Another important contribution of this study lies in its methodological rigor, incorporating SHAP analysis for model interpretability. While black-box criticisms are often leveled against ML models, especially in energy forecasting contexts, SHAP plots clearly illustrated the direction and magnitude of the contribution of each feature confirming the strong positive impact of RC, and the secondary effects of LC and IC. This dominance of residential consumption reflects Turkey's policy-driven shift toward urbanization, population growth in metropolitan areas, and the ongoing electrification of households and residential heating systems factors that have made residential demand a central focus of national energy efficiency and sustainability strategies. This interpretability framework enhances transparency and aligns with recommendations that stressed the importance of explainability in high-stakes domains like energy policy [14,18].

It is important to note that while the comparative analysis of eight state-of-the-art regression algorithms provided an empirical foundation for model selection, it was not intended to constitute the central contribution of this research. Rather, this comparative stage served as a diagnostic process to identify the most effective algorithmic compo-

nent CatBoost for integration within the proposed interpretability-centered ensemble and scenario-based forecasting framework. The true novelty of the present study lies in the design and application of this unified framework, which transforms individual model comparisons into a systemic, policy-oriented decision-support structure. By embedding model interpretability through SHAP analysis and linking it with forward-looking scenario simulations (2025–2030), the framework transcends conventional benchmarking to deliver actionable energy insights, strategic forecasting capability, and transparent policy guidance. Thus, the comparative simulation study functions as an enabling step, while the framework's integrative and application-driven dimension represents the core scientific advancement of this work.

In summary, this study provides robust evidence supporting the deployment of ensemble machine learning models particularly CatBoost and LightGBM as effective tools for national-scale electricity demand forecasting in Turkey. By integrating economic, temporal, and sectoral variables, these models not only enhance predictive performance but also generate actionable insights for policymakers and grid operators. Furthermore, quantitative comparisons with [1] affirm the consistency and validity of our predictive outcomes, demonstrating the reliability of modern ML methods for electricity demand forecasting in Turkey. As Turkey advances toward a more decentralized and renewable-based energy system, the demand for interpretable, scalable, and adaptive forecasting tools becomes increasingly critical. The principal innovation of this research lies in the proposed interpretability-centered ensemble and scenario-based forecasting framework, which bridges predictive modeling with policy-oriented decision support. Unlike previous studies that mainly emphasize model benchmarking, the present framework establishes a transparent analytical architecture by combining high-performing ensemble algorithms (CatBoost, LightGBM, Random Forest) with post hoc interpretability techniques (SHAP and feature importance analyses) to reveal complex sectoral interdependencies in electricity demand. This methodological integration transforms existing algorithms into a cohesive, decision-support ecosystem capable of explaining the causal and temporal structure of national energy consumption. Furthermore, by coupling this interpretability-driven modeling approach with scenario-based projections for 2025–2030, the framework extends its scope from retrospective forecasting to forward-looking strategic planning, contributing not only technical accuracy but also policy transparency, scalability, and replicability representing a substantive advancement in the methodological landscape of national-scale energy forecasting.

6. Conclusions

This study provides compelling empirical evidence that ensemble-based machine learning models—particularly CatBoost, LightGBM, and Random Forest can offer highly accurate, interpretable, and scalable solutions for national electricity consumption and production forecasting. By leveraging a comprehensive dataset that spans sectoral electricity usage, economic indicators, and trade variables over the 2016–2024 period, the models achieved superior predictive performance. Specifically, CatBoost attained the highest forecasting accuracy, with R^2 values of 0.9144 for electricity production and 0.8247 for electricity consumption, while also maintaining low error rates (MAE, RMSE) and avoiding overfitting.

Through SHAP-based feature attribution, this study revealed that residential electricity consumption is by far the most influential predictor, followed by lighting and industrial usage. These findings underscore the primacy of sectoral demand components in shaping electricity consumption dynamics in Turkey, with temporal and macroeconomic variables contributing negligibly in most cases. Such insights reinforce the necessity of incorporating granular sectoral data into future forecasting and planning models.

Moreover, scenario-based simulations projecting energy demand through 2030 confirmed the adaptability and policy relevance of the proposed framework under diverse economic growth pathways. These scenarios—ranging from moderate to high-growth assumptions—highlighted significant variations in projected electricity demand, affirming the importance of data-driven forecasting in strategic energy infrastructure investment, policy formulation, and long-term sustainability planning.

This study also advances the literature by addressing a major gap: the lack of interpretable machine learning applications at the national scale that integrate multidimensional datasets. In doing so, it bridges the divide between predictive performance and model transparency, aligning with the growing demand for explainable artificial intelligence (XAI) in high-stakes domains like energy forecasting. The methodology presented herein offers not only technical innovation but also real-world applicability, positioning ensemble ML models as critical tools in Turkey's ongoing transition toward a resilient, data-informed, and sustainable energy future.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval for the study was obtained from the Harran University Clinical Research Ethics Committee with the decision number HRÜ/24.08.45, dated 10 June 2024.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Raw data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Saglam, M.; Spataru, C.; Karaman, O.A. Forecasting electricity demand in Turkey using optimization and machine learning algorithms. *Energies* **2023**, *16*, 4499. [CrossRef]
2. Manandhar, P.; Rafiq, H.; Rodriguez-Ubinas, E. Current status, challenges, and prospects of data-driven urban energy modeling: A review of machine learning methods. *Energy Rep.* **2023**, *9*, 2757–2776. [CrossRef]
3. Gulay, E.; Sen, M.; Akgun, O.B. Forecasting electricity production from various energy sources in Türkiye: A predictive analysis of time series, deep learning, and hybrid models. *Energy* **2024**, *286*, 129566. [CrossRef]
4. Liu, X.; Zhang, X.; Baziar, A. Hybrid machine learning and modified teaching learning-based English optimization algorithm for smart city communication. *Sustainability* **2023**, *15*, 11535. [CrossRef]
5. Wazirali, R.; Yaghoubi, E.; Abujazar, M.S.S.; Ahmad, R.; Vakili, A.H. State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. *Electr. Power Syst. Res.* **2023**, *225*, 109792. [CrossRef]
6. Charfeddine, L.; Zaidan, E.; Alban, A.Q.; Bennasr, H.; Abulibdeh, A. Modeling and forecasting electricity consumption amid the COVID-19 pandemic: Machine learning vs. nonlinear econometric time series models. *Sustain. Cities Soc.* **2023**, *98*, 104860. [CrossRef]
7. Kayacı Çodur, M. Ensemble machine learning approaches for prediction of Türkiye's energy demand. *Energies* **2023**, *17*, 74. [CrossRef]
8. Unsal, D.B.; Aksoz, A.; Oyucu, S.; Guerrero, J.M.; Guler, M. A comparative study of AI methods on renewable energy prediction for smart grids: Case of Turkey. *Sustainability* **2024**, *16*, 2894. [CrossRef]
9. Peteleaza, D.; Matei, A.; Sorostinean, R.; Gellert, A.; Fiore, U.; Zamfirescu, B.C.; Palmieri, F. Electricity consumption forecasting for sustainable smart cities using machine learning methods. *Internet Things* **2024**, *27*, 101322. [CrossRef]
10. Albuquerque, P.C.; Cajueiro, D.O.; Rossi, M.D. Machine learning models for forecasting power electricity consumption using a high-dimensional dataset. *Expert Syst. Appl.* **2022**, *187*, 115917. [CrossRef]
11. Grandón, T.G.; Schwenzer, J.; Steens, T.; Breuing, J. Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of Ukraine. *Appl. Energy* **2024**, *355*, 122249. [CrossRef]
12. Chen, G.; Hu, Q.; Wang, J.; Wang, X.; Zhu, Y. Machine-learning-based electric power forecasting. *Sustainability* **2023**, *15*, 11299. [CrossRef]

13. Mathumitha, R.; Rathika, P.; Manimala, K. Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review. *Artif. Intell. Rev.* **2024**, *57*, 35. [CrossRef]
14. Dong, Q.; Huang, R.; Cui, C.; Towey, D.; Zhou, L.; Tian, J.; Wang, J. Short-term electricity-load forecasting by deep learning: A comprehensive survey. *arXiv* **2024**, arXiv:2408.16202. [CrossRef]
15. Vanting, N.B.; Ma, Z.; Jørgensen, B.N. A scoping review of deep neural networks for electric load forecasting. *Energy Inform.* **2021**, *4* (Suppl. 2), 49. [CrossRef]
16. Nti, I.K.; Teimeh, M.; Nyarko-Boateng, O.; Adekoya, A.F. Electricity load forecasting: A systematic review. *J. Electr. Syst. Inf. Technol.* **2020**, *7*, 1–19. [CrossRef]
17. Lee, M.H.L.; Ser, Y.C.; Selvachandran, G.; Thong, P.H.; Cuong, L.; Son, L.H.; Gerogiannis, V.C. A comparative study of forecasting electricity consumption using machine learning models. *Mathematics* **2022**, *10*, 1329. [CrossRef]
18. Klyuev, R.V.; Morgoev, I.D.; Morgoeva, A.D.; Gavrina, O.A.; Martyushev, N.V.; Efremenkov, E.A.; Mengxu, Q. Methods of forecasting electric energy consumption: A literature review. *Energies* **2022**, *15*, 8919. [CrossRef]
19. Román-Portabales, A.; López-Nores, M.; Pazos-Arias, J.J. Systematic review of electricity demand forecast using ANN-based machine learning algorithms. *Sensors* **2021**, *21*, 4544. [CrossRef] [PubMed]
20. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
21. Tzelepi, M.; Symeonidis, C.; Nousi, P.; Kakaletsis, E.; Manousis, T.; Tosidis, P.; Tefas, A. Deep learning for energy time-series analysis and forecasting. *arXiv* **2023**, arXiv:2306.09129. [CrossRef]
22. Pelekis, S.; Seisopoulos, I.K.; Spiliotis, E.; Pountridis, T.; Karakolis, E.; Mouzakitis, S.; Askounis, D. A comparative assessment of deep learning models for day-ahead load forecasting: Investigating key accuracy drivers. *Sustain. Energy Grids Netw.* **2023**, *36*, 101171. [CrossRef]
23. Mystakidis, A.; Ntozi, E.; Afentoulis, K.; Koukaras, P.; Gkaidatzis, P.; Ioannidis, D.; Tzovaras, D. Energy generation forecasting: Elevating performance with machine and deep learning. *Computing* **2023**, *105*, 1623–1645. [CrossRef]
24. Mehrdoust, F.; Noorani, I.; Belhaouari, S.B. Forecasting Nordic electricity spot price using deep learning networks. *Neural Comput. Appl.* **2023**, *35*, 19169–19185. [CrossRef]
25. Ramos, P.V.B.; Villela, S.M.; Silva, W.N.; Dias, B.H. Residential energy consumption forecasting using deep learning models. *Appl. Energy* **2023**, *350*, 121705. [CrossRef]
26. Zafeiriou, A.; Chantzis, G.; Jonkaitis, T.; Fokaides, P.; Papadopoulos, A. Smart energy strategy—A comparative study of energy consumption forecasting machine learning models. *Chem. Eng. Trans.* **2023**, *103*, 691–696.
27. Talwariya, A.; Singh, P.; Jobanputra, J.H.; Kolhe, M.L. Machine learning-based renewable energy generation and energy consumption forecasting. *Energy Sources Part A Recovery Util. Environ. Eff.* **2023**, *45*, 3266–3278. [CrossRef]
28. Raju, R.; Ramlal, N.P. Dual deep learning model for electricity price forecasting: Bi-LSTM and GRU fusion. In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, Noida, India, 3–5 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 13–17.
29. Neo, H.Y.R.; Wong, N.H.; Ignatius, M.; Cao, K. A hybrid machine learning approach for forecasting residential electricity consumption: A case study in Singapore. *Energy Environ.* **2024**, *35*, 3923–3939. [CrossRef]
30. Ediger, V.S.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708. [CrossRef]
31. Ozturk, S.; Ozturk, F. Forecasting energy consumption of Turkey by ARIMA model. *J. Asian Sci. Res.* **2018**, *8*, 52. [CrossRef]
32. Ağbulut, U. Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms. *Sustain. Prod. Consum.* **2022**, *29*, 141–157. [CrossRef]
33. Kankal, M.; Uzlu, E. Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey. *Neural Comput. Appl.* **2017**, *28*, 737–747. [CrossRef]
34. Canyurt, O.E.; Ceylan, H.; Ozturk, H.K.; Hepbasli, A. Energy demand estimation based on two different genetic algorithm approaches. *Energy Sources* **2004**, *26*, 1313–1320. [CrossRef]
35. Aslan, M.; Beşkırli, M. Realization of Turkey's energy demand forecast with the improved arithmetic optimization algorithm. *Energy Rep.* **2022**, *8*, 18–32. [CrossRef]
36. Aslan, M. Archimedes optimization algorithm-based approaches for solving energy demand estimation problem: A case study of Turkey. *Neural Comput. Appl.* **2023**, *35*, 19627–19649. [CrossRef]
37. Tefek, M.F.; Uğuz, H.; Gucyetmez, M. A new hybrid gravitational search–teaching–learning-based optimization method for energy demand estimation of Turkey. *Neural Comput. Appl.* **2019**, *31*, 2939–2954. [CrossRef]
38. Koç, İ.; Nureddin, R.; Kahramanlı, H. Implementation of GSA (Gravitation Search Algorithm) and IWO (Invasive Weed Optimization) for the prediction of the energy demand in Turkey using linear form. *Selçuk Univ. J. Eng. Sci. Technol.* **2018**, *6*, 529–543.

39. Daş, G.S. Forecasting the energy demand of Turkey with a neural network based on an improved particle swarm optimization. *Neural Comput. Appl.* **2017**, *28*, 539–549. [CrossRef]
40. Beşkırli, A.; Beşkırli, M.; Hakli, H.; Uguz, H. Comparing energy demand estimation using artificial algae algorithm: The case of Turkey. *J. Clean Energy Technol.* **2018**, *6*, 349–352. [CrossRef]
41. Unler, A. Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025. *Energy Policy* **2008**, *36*, 1937–1944. [CrossRef]
42. Toksari, M.D. Ant colony optimization approach to estimate energy demand of Turkey. *Energy Policy* **2007**, *35*, 3984–3990. [CrossRef]
43. Cayir Ervural, B.; Ervural, B. Improvement of grey prediction models and their usage for energy demand forecasting. *J. Intell. Fuzzy Syst.* **2018**, *34*, 2679–2688. [CrossRef]
44. Tutun, S.; Chou, C.A.; Canıyılmaz, E. A new forecasting for volatile behavior in net electricity consumption: A case study in Turkey. *Energy* **2015**, *93*, 2406–2422. [CrossRef]
45. Hussain, S.; Mustafa, M.W.; Jumani, T.A.; Baloch, S.K.; Alotaibi, H.; Khan, I.; Khan, A. A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. *Energy Rep.* **2021**, *7*, 4425–4436. [CrossRef]
46. Ibrahim, A.A.; Ridwan, R.L.; Muhammed, M.M.; Abdulaziz, R.O.; Saheed, G.A. Comparison of the CatBoost classifier with other machine learning methods. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 11. [CrossRef]
47. Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *Npj Comput. Mater.* **2023**, *9*, 42. [CrossRef]
48. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
49. Sharma, S.; Gupta, V.; Mudgal, D.; Srivastava, V. Machine learning for forecasting the biomechanical behavior of orthopedic bone plates fabricated by fused deposition modeling. *Rapid Prototyp. J.* **2024**, *30*, 441–459. [CrossRef]
50. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*; Curran Associates Inc.: Red Hook, NY, USA, 2017; Volume 30.
51. Zhou, Y.; Wang, W.; Wang, K.; Song, J. Application of LightGBM algorithm in the initial design of a library in the cold area of China based on comprehensive performance. *Buildings* **2022**, *12*, 1309. [CrossRef]
52. Re, M.; Valentini, G. Ensemble methods: A review. In *Advances in Machine Learning and Data Mining for Astronomy*; Kumar, V., Ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2012; pp. 563–594.
53. Nalluri, M.; Pentela, M.; Eluri, N.R. A Scalable Tree Boosting System: XGBoost. *Int. J. Res. Stud. Sci. Eng. Technol.* **2020**, *7*, 36–51.
54. Sharma, S.; Gupta, V.; Mudgal, D. Response surface methodology and machine learning-based tensile strength prediction in ultrasonic assisted coating of poly lactic acid bone plates manufactured using fused deposition modeling. *Ultrasonics* **2024**, *137*, 107204. [CrossRef]
55. 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.
56. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
57. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, É. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
58. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
59. Opitz, D.; Maclin, R. Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.* **1999**, *11*, 169–198. [CrossRef]
60. Kramer, O.; Kramer, O. K-nearest neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23.
61. Pan, Z.; Wang, Y.; Pan, Y. A new locally adaptive k-nearest neighbor algorithm based on discrimination class. *Knowl. Based Syst.* **2020**, *204*, 106185. [CrossRef]

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