J Inno Sci Eng, 2025, 9(2): 247-258

Research Article

DOI:https://doi.org/ 10.38088/jise.1596664

# Consideration of Environmental, Economic, and Oil Factors for Unit-based Estimation of Consumed Electrical Energy with ML Algorithms: A Case Study of Çanakkale Region

**Yasemin Ayaz Atalan** [1]

[1] *Çanakkale Onsekiz Mart University, Çanakkale Faculty of Applied Sciences, Energy Management Department, 17020, Çanakkale, Türkiye*

ARTICLE INFO

Received Date: 5/12/2024
Accepted Date: 10/07/2025

Cite this paper as:
Atalan Ayaz,Y. (2025) Consideration of Environmental, Economic, and Oil Factors for Unit-based Estimation of Consumed Electrical Energy with ML Algorithms: A Case Study of Çanakkale Region, *Journal of Innovative Science and Engineering.* 9(2), 247-258.

*Corresponding author: Yasemin Ayaz Atalan
E-mail:Yasemin.atalan@comu.edu.tr

*Keywords:*
Machine learning
Electricity pricing
Random forest
k-nearest neighbor
Energy forecasting
Environmental indicators
Economic indicators
Oil Price



**ABSTRACT**

This study focuses on predicting electricity unit prices in the Çanakkale region by analyzing the effects of environmental, economic, and oil-related factors through machine learning (ML) algorithms. The research addresses the accurate prediction of energy costs amid fluctuating market dynamics by applying Random Forest (RF) and k-nearest neighbor (kNN) algorithms to monthly data from 2015 to 2024. The independent variables used in the models include exchange rate (USD/TRY), oil price (TL/liter), Producer Price Index (PPI), Consumer Price Index (CPI), and average temperature. The RF algorithm achieves superior predictive accuracy with an MSE of 0.013, RMSE of 0.112, MAE of 0.081, MAPE of 0.087, and an $R^2$ of 0.919, outperforming the kNN model across all metrics. The findings reveal that exchange rate and PPI have the most significant influence on electricity pricing. This study provides empirical evidence supporting the use of ML methods in energy price prediction and contributes to developing more accurate and robust forecasting tools for regional energy management and policy-making.

## 1. Introduction

Electrical energy is one of the basic needs of modern societies and plays a critical role in economic development, industrial production, transportation, and daily life. Electrical energy is mainly produced by burning fossil fuels, nuclear reactions, or using renewable energy sources (solar, wind, hydroelectric) [1]. While using fossil fuels increases greenhouse gas emissions that lead to environmental pollution and

climate change, utilizing renewable energy sources provides advantages in terms of sustainability. The demand for electrical energy is increasing in parallel with economic growth, population growth, and technological developments, which requires the effective and efficient use of energy resources [2]. Therefore, developing innovative solutions and technologies in the energy sector is vital to ensure energy supply security and reduce environmental impacts [3].

The sustainable and reliable provision of electrical energy supply depends on diversifying energy production methods and increasing energy efficiency. Today, new-generation solutions such as smart grids and energy storage technologies provide more effective management in electrical energy production, transmission, and distribution processes [4]. ML and artificial intelligence (AI) applications, especially in areas such as electrical energy demand forecasting and load balancing, increase the resilience of energy systems by optimizing the supply-demand balance [5]. In this context, understanding and managing the dynamic structure of electrical energy systems contribute to making strategic decisions to support the sustainable use of energy resources.

Forecasting electricity consumption is essential to planning and management processes in the energy sector. Accurately estimating electricity consumption is vital to ensuring the efficient operation of electricity production and distribution systems [6]. The electricity demand can often vary significantly depending on the time of day, seasonal changes, economic activity level, and weather conditions [7]. Therefore, accurate estimation of electricity needs is crucial to prevent sudden fluctuations in energy production, minimize energy costs, and ensure energy supply security. In addition, forecast data facilitates the integration of renewable energy sources by providing flexibility in the management of electricity networks, thus reducing the carbon footprint [8]. Suppose the balance between energy supply and demand is not achieved. In that case, serious consequences such as power outages, grid overloads, and high energy costs can occur, further emphasizing the importance of forecasting.

Various methods are used in the literature to forecast electricity consumption. Traditional methods are generally based on statistical models [9]. Techniques such as time series analysis, regression models, and the Box-Jenkins methodology (ARIMA models) are widely used to forecast future electricity demand using past consumption data [10]. These methods predict the data's past trends, cyclicality, and seasonal components. However, the disadvantage of these statistical models is that they may be inadequate in accounting for the effects of complex and dynamic systems, sudden weather changes, or unexpected economic shocks. Therefore, while traditional methods are often used for short-term predictions, more advanced analyses are needed for long-term predictions.

In recent years, the use of AI and ML methods in electricity consumption predictions has increased [11]. ML methods can learn complex patterns using past data and thus make more accurate predictions [12]. Methods such as support vector machines (SVM), artificial neural networks (ANN), decision trees, and deep learning models are quite effective in electricity consumption predictions [13]. These methods are preferred mainly due to their ability to model the effects of many variables on electricity demand. In addition, these techniques allow real-time processing of energy data, contributing to the development of more dynamic and adaptive energy management strategies. With the further development of ML methods, the accuracy and reliability of electricity consumption forecasts will increase, significantly contributing to the energy sector's sustainability and efficiency [14].

Electricity prices are determined by several complex factors that cause significant fluctuations in global energy markets today. In particular, environmental conditions, economic indicators, and oil prices, which are raw energy materials, play a critical role in changing the unit prices of electrical energy [15]. Local markets such as the Çanakkale region feel the effects of these factors more clearly. This study aims to estimate the factors affecting the unit prices of consumed electrical energy in Çanakkale province (exchange rate, oil prices, Producer Price Index (PPI), Consumer Price Index (CPI), and temperature) using ML algorithms. For this purpose, Random Forest and k-nearest neighbor algorithms were run using monthly data between 2015 and 2024, and the prediction performances of these algorithms were compared. Thus, it is aimed to contribute to both energy policy development processes and to develop strategies to reduce market uncertainties.

This study contributes to the literature in several important ways. First, it integrates environmental, economic, and oil price factors to predict electricity unit prices, addressing a multidimensional challenge not commonly explored in regional forecasting studies. Second, it compares the performance of two distinct ML algorithms using real-world data from the Çanakkale region. It provides empirical evidence on

model selection in the context of energy price forecasting. Third, error metrics, ANOVA, and Pareto analysis offer a comprehensive evaluation framework, enhancing methodological robustness. These contributions support more accurate and reliable forecasting models for energy policymakers, local energy providers, and researchers working on data-driven energy management strategies.

This study consists of five main sections. The introduction section of this study, which aims to estimate the environmental, economic, and oil price factors affecting the unit prices of electricity consumption in the Çanakkale region using ML algorithms, discusses the importance of predicting electricity prices and the difficulties encountered. The second section introduces the data set used to create forecasting models, dependent and independent variables, and the ML algorithms are explained in detail. The Result section includes a comparison of the models' performances and the evaluation of the forecasting results, and the Discussion and Conclusion section discusses the potential effects of the findings on energy policies and decision support systems.

## 2. Methodology

This study's methodology uses various ML algorithms to predict electricity unit prices in the Çanakkale region. The study considered various environmental and economic indicators such as exchange rate, oil prices, PPI, CPI, and average temperature as independent variables. In contrast, the dependent variable was determined as unit prices of electricity. Using these variables, models were developed to predict electricity prices with the help of Random Forest (RF) and k-nearest neighbor (kNN) algorithms.

In the data analysis, various performance metrics such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination ($R^2$) were used to evaluate the performance of ML algorithms. The study utilized hyperparameter optimization and cross-validation techniques to increase the accuracy and reliability of the prediction models. This methodological approach provides a comprehensive and data-driven framework for predicting energy prices and developing energy management strategies. The Python programming language and related libraries (scikit-learn, pandas, and numpy) were used to install and optimize the models. The dependent variable's data covers the time span from September 2015 to January 2024. The unit price data for the electrical energy consumed in this period are shown in Figure 1.

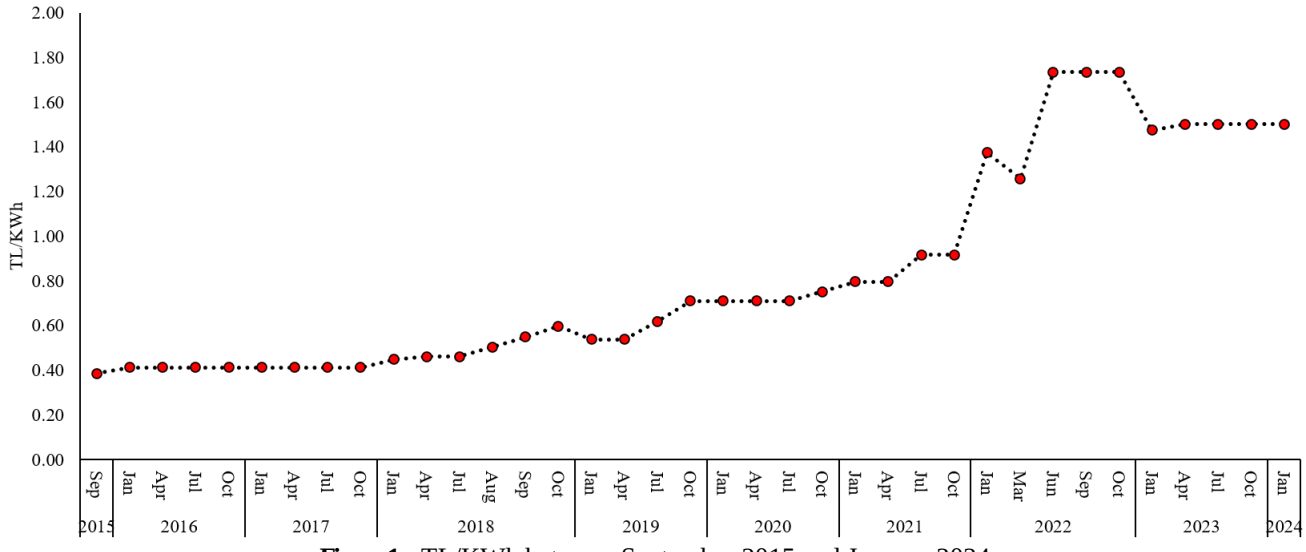


**Figure 1:** TL/KWh between September 2015 and January 2024

Before applying the machine learning algorithms, several data preprocessing steps were performed to ensure the quality and reliability of the input data. First, the missing values in the dataset were examined, and since the data was collected from monthly national and meteorological records with validation, significant missing values were completed. Second, all numerical variables were normalized using min-max scaling to bring them to a standard scale and eliminate potential bias due to different units and magnitudes. This normalization process improves the algorithm's performance and

ensures fair comparison between variables. The data was also checked for outliers using statistical metrics such as interquartile range, and extreme values were kept because they represent real-world fluctuations.

Although the electricity unit price fluctuates in the period considered, the electricity unit price was calculated as a maximum of 1.73 TL/kWh for June-October 2022. Descriptive statistics such as mean, standard deviation, variance, minimum, maximum, kurtosis, and skewness of dependent and independent variables are given in Table 1.

Table 1 shows the descriptive statistics of various variables affecting the unit electricity prices in the Çanakkale region. The average (mean), standard deviation (StDev), variance, minimum and maximum values calculated for variables such as year, electricity unit price (TL/KWh), exchange rate ($/TL), oil price (TL/lt), Producer Price Index (PPI), Consumer Price Index (CPI) and average temperature (°C) are presented. The average value for the unit price of electricity is 0.8304 TL/KWh, the lowest value is 0.3851 TL/KWh, and the highest value is 1.7346 TL/KWh. While the average exchange rate is 9.42 $/TL, the average oil price is determined as 11.02 TL/lt. The data's standard deviation and variance values reveal the distribution and variability level in the relevant variables. The standard deviation of the PPI is calculated as 35.74, and the variance is 1277.21, which shows that the PPI has a wide distribution.
Quartiles (Q1, Median, Q3) and interquartile range (IQR) measures allow a more detailed analysis of the data's central tendency and distribution characteristics. While the median value for the electricity unit price is 0.7102 TL/KWh, the third quartile (Q3) value is 1.3151 TL/KWh, demonstrating a wide spread in the upper half of the distribution. Skewness and kurtosis measures evaluate whether the data is symmetrical and the effect of extreme values. While the exchange rate and oil price show positive skewness, the temperature data exhibits an almost symmetrical distribution (skewness: 0.12). Moreover, the skewness value for the exchange rate (0.88) and the skewness value for the oil price (1.43) indicate that these variables are skewed to the right, meaning that large values are more frequent. These statistics provide a comprehensive insight into the behavior of the analyzed variables over time and provide vital information for future forecasts.

In this study, ML algorithms, RF, and kNN models were run to estimate the unit prices of consumed electricity per kWh. The algorithm of the ML models is shown in Figure 2. RF and kNN algorithms were selected for this study due to their complementary strengths in regression tasks. RF is a robust ensemble learning method that can model complex, non-linear relationships and reduce overfitting using multiple decision trees and random feature selection. It is especially effective when working with heterogeneous data, such as mixed economic and environmental indicators. On the other hand, kNN is a simple yet powerful instance-based learning algorithm that performs well with smaller datasets and offers intuitive interpretability. By comparing these two algorithms, the study aims to evaluate a high-complexity ensemble model and a distance-based method, offering a balanced assessment of prediction performance for electricity pricing. The reason for using RF and kNN algorithms in this study is also due to the algorithms' advantages in both classification and regression problems. RF is one of the ensemble learning methods and allows many decision trees to come together to perform classification or regression [16]. This approach reduces the risk of overfitting the model, increasing overall performance and providing high accuracy. The kNN is a simple and effective method; it predicts the class of new examples using the distance between labeled data [17]. It gives good results, especially in small data sets and multi-dimensional areas. Combining both algorithms can increase the accuracy and generalizability of the model, so both methods are used in this study.

RF is an ensemble learning method and a robust classification and regression model formed by combining decision trees [16]. RF produces the final prediction by training multiple decision trees independently and taking the majority vote or average of the results from these trees. This method increases the model's overall performance and reduces the risk of overfitting. Each decision tree is trained on a random subset of the data, and each tree is created using a set of randomly selected features, which increases the generalizability of the model. This diversity makes the RF model robust. Hyperparameter information of the RF Model is given in Table 2.

**Table 1:** Descriptive statistics of variables (Year and month data are not considered)

| Variable | Year | TL/kWh | Currency ($/TL) | Oil Price (TL/lt) | PPI (%) | CPI (%) | Temperature ($^0$C) |
|---|---|---|---|---|---|---|---|
| Mean | 2019.5 | 0.8304 | 9.42 | 11.02 | 34.71 | 23.28 | 15.68 |
| StDev | 2.47 | 0.4626 | 7.57 | 9.12 | 35.74 | 20.17 | 6.68 |
| Variance | 6.09 | 0.214 | 57.27 | 83.19 | 1277.21 | 406.75 | 44.56 |
| Minimum | 2015 | 0.3851 | 2.84 | 4.21 | 3.83 | 7.69 | 6.3 |
| Q1 | 2017.5 | 0.4303 | 3.77 | 5.44 | 9.99 | 10.71 | 9.6 |
| Median | 2019 | 0.7102 | 5.93 | 6.88 | 19.44 | 13.75 | 16.3 |
| Q3 | 2022 | 1.3151 | 14.09 | 16.34 | 48.64 | 26.23 | 21.75 |
| Maximum | 2024 | 1.7346 | 30.09 | 37.08 | 125.53 | 72.45 | 25.1 |
| IQR | 4.5 | 0.8848 | 10.32 | 10.9 | 38.65 | 15.51 | 12.15 |
| Skewness | 0.03 | 0.88 | 1.43 | 1.65 | 1.46 | 1.4 | 0.12 |
| Kurtosis | -1.12 | -0.76 | 1.1 | 1.67 | 1.05 | 0.42 | -1.26 |

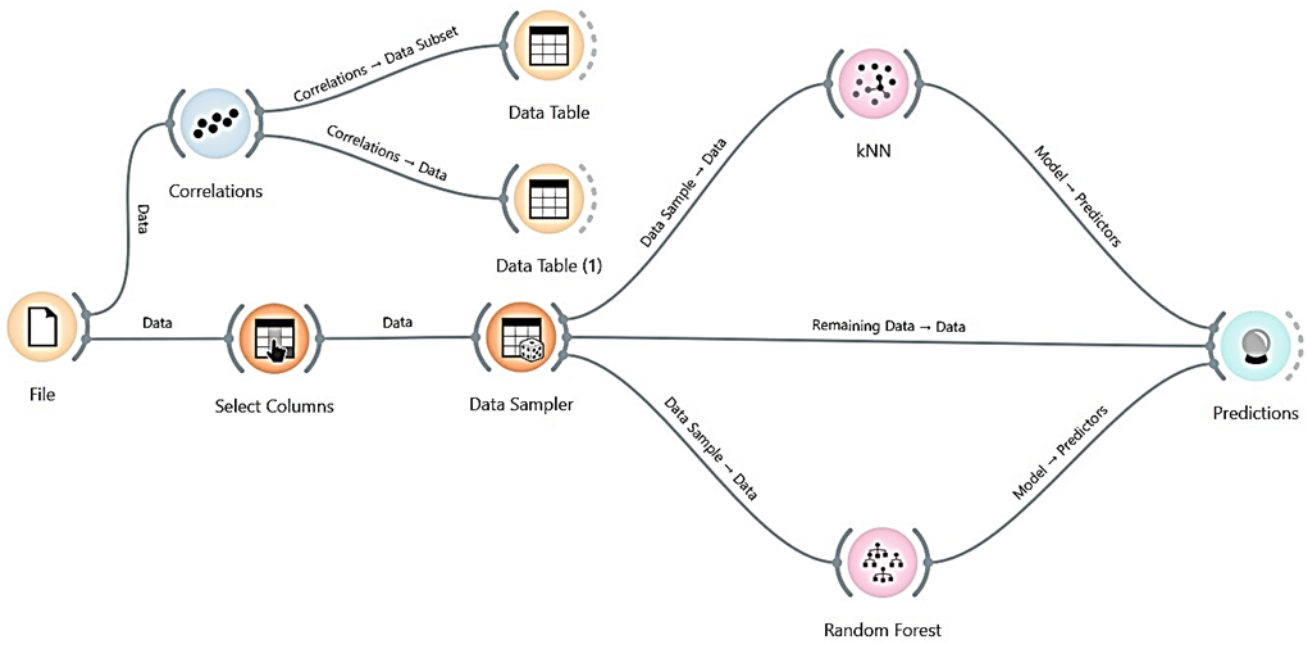


**Figure 2:** Workflow visual according to the working principles of ML algorithms

The hyperparameters of the RF model used in this study play an important role in determining the performance and features of the model. The model is structured with an ensemble of 10 decision trees. In this way, more stable and accurate predictions are made using the combination of each tree's classification or regression decisions. The number of features considered in the model is unlimited, allowing the model to use all the essential information in the dataset. The maximum tree depth is not limited, meaning that the trees can fully reflect the complexity of the dataset. In addition, the criterion for stopping nodes from being separated is that there should be at least five examples at each node. This allows the model to undergo a more general and robust learning process, avoiding over-discrimination. Replicability indicates that the results will not be the same every time due to randomness during model training. This configuration allows the model to work both flexibly and efficiently.

**Table 2:** Hyperparameter information of the RF algorithm

| |
|---|
| Name: Random Forest |
| Model parameters |
| Number of trees: 10 |
| Maximal number of considered features: unlimited |
| Replicable training: No |
| Maximal tree depth: unlimited |
| Stop splitting nodes with maximum instances: 5 |

The kNN algorithm is a simple but effective classification and regression method. kNN predicts the class or value of a data point by looking at the labels or values of its k nearest neighbors [17]. These

neighbors are usually selected using a specific distance metric, and the most common classification method is to determine the prediction by taking the majority vote. The main advantage of kNN is that it does not require a complex model training process and is relatively fast to predict new data [18]. However, the performance of kNN depends on the number of neighbors selected and the distance metric. Hyperparameter data of the kNN Model are given in Table 3.

**Table 3:** Hyperparameter information of the kNN algorithm

| |
|---|
| Name: kNN |
| Model parameters |
| Number of neighbours: 5 |
| Metric: Euclidean |
| Weight: Uniform |

The hyperparameters of the kNN model used in this study significantly affect the model's prediction performance. The model makes predictions using five neighbors, which means that the projections will be made according to the majority of the labels of the k nearest neighbors. The distances between neighbors are calculated using the Euclidean distance metric, which considers the actual physical distances between data points. In addition, the weights are set to Uniform. This means that all neighbors will be considered with equal weight; therefore, each neighbor will contribute equally to the prediction. These hyperparameters affect the similarities in the kNN model's dataset and the projections' accuracy, determining the model's overall success.

The values of MSE, RMSE, MAE, MAPE, and $R^2$ used to measure the performance of ML algorithms are critical in evaluating the model's predictive accuracy and overall success [19]. MSE and RMSE determine the error size of the model by measuring how far the predictions are from the actual values; RMSE presents the error measurement in the same unit by taking the square root of MSE, which makes it easier to interpret. MAE shows how much the predictions deviate from the actual values on average and expresses the error magnitude in absolute terms. MAPE evaluates the error performance of the model in relative terms by giving the mean percentage of errors. $R^2$, on the other hand, measures how much of the variance the model explains in the dataset, which indicates the overall explanatory power of the model. These performance metrics are essential for comparing different algorithms and selecting the most suitable model, providing a vital guide for evaluating the model's effectiveness on real-world data. In this study, MSE, RMSE, MAE, MAPE, and $R^2$ values were calculated to measure the performance of ML algorithms. The formulas for these performance measurement criteria are given below, respectively:

MSE measures the average of the squares of the errors, which are the differences between predicted and actual values. A lower MSE indicates that the model's predictions are closer to the actual values [20]. The MSE equation is as follows:

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

where, y_i represents the actual value, y^_i represents the predicted value and n is the number of observations in the dataset. RMSE is the square root of MSE, providing an error measure in the same unit as the original data. It reflects the average magnitude of the errors [21]. The RMSE equation is as follows:

$$RMSE = \sqrt{MSE} \quad (2)$$

MAE measures the average absolute difference between predicted and actual values. It provides a straightforward interpretation of prediction errors [22]. The MAE equation is as follows:

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

MAPE expresses the error as a percentage of the actual values, giving a relative measure of prediction accuracy [23]. The MAPE equation is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (4)$$

where the percentage errors are calculated based on non-zero actual values. $R^2$ measures the proportion of the variance in the dependent variable that is predictable from the independent variables [24]. An $R^2$ value close to 1 indicates a better fit of the model to the data, while 0 suggests no explanatory power. The $R^2$ equation is as follows:

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

**Table 4:** ANOVA data for variables

| Variables | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 13 | 7.75298 | 0.596383 | 96.13 | 0.000 |
| Currency ($/TL) | 1 | 0.09335 | 0.093352 | 15.05 | 0.001 |
| Oil Price (TL/lt) | 1 | 0.04041 | 0.040412 | 6.51 | 0.018 |
| PPI (%) | 1 | 0.08329 | 0.083289 | 13.43 | 0.001 |
| CPI (%) | 1 | 0.01566 | 0.015655 | 2.52 | 0.126 |
| Temperature ($^0$C) | 1 | 0.00184 | 0.001844 | 0.30 | 0.591 |
| Year | 1 | 0.00002 | 0.000017 | 0.00 | 0.959 |
| Month | 7 | 0.12254 | 0.017506 | 2.82 | 0.028 |

**Abbreviation**: DF, Degrees of Freedom; Adj SS, Adjusted Sum of Squares; Adj MS, Adjusted Mean Squares

where $\bar{y}_i$ is the mean of the actual values. These performance metrics—MSE, RMSE, MAE, MAPE, and $R^2$—comprehensively evaluate the model's predictive accuracy and fit to the data, allowing for the determination of how well the model performs and comparing various algorithms.

## 3. Results and Discussion

This section presents the results of RF and kNN models applied to predict electricity unit prices in the Çanakkale region. Detailed numerical results, including performance metrics, are presented for both models. These metrics are essential to evaluate the accuracy and effectiveness of the models in predicting electricity prices. The section includes a comparative analysis of the performance of RF and kNN algorithms, highlighting their strengths and weaknesses based on the calculated values. The results aim to show the predictive capabilities of each model and their suitability in predicting electricity prices in the context of changing environmental, economic, and oil price factors. First, an Analysis of variance (ANOVA) was performed to show the effect of independent variables on the dependent variable. In addition to ANOVA, Pareto analysis was performed to show the effect strengths of independent variables. ANOVA data are presented in Table 4, and the Pareto chart is in Figure 3.

Table 4 presents the ANOVA results examining the effects of various independent variables on the electricity unit prices. Table 4 includes various statistics measuring the contribution of each variable to the regression model. Overall, the total effect of the regression model was calculated as 7.75298, and the F-value of this model was 96.13, indicating a very high accuracy. This shows that the model explains the effects of the independent variables well enough and obtains a significant result. In addition, variables such as exchange rate ($/TL) and PPI have high F-values and low p-values, indicating that they significantly affect electricity prices. These variables are factors that increase the model's predictive power and are effective. Other variables show impact at different levels. While Oil Price (TL/lt) has a significant influence (p-value 0.018), the effects of the CPI and temperature variables are less significant and have high p-values (0.126 and 0.591, respectively). The effect of the year variable is almost negligible (p-value 0.959). In addition, the impact of the month variable is also noteworthy, indicating a significant effect with a p-value of 0.028. These results show that the most influential factors on electricity prices are the exchange rate and PPI, while others are less effective.

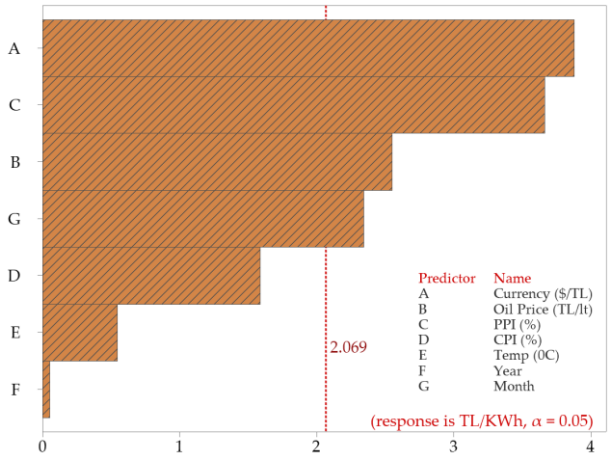


**Figure 3:** Pareto Chart of the Standardized Effects for kWh

Figure 3 lists the standardized effects that show how unit prices of electricity consumption are affected by different factors. In the graph, the impact of exchange rate and oil prices on electricity prices stands out with higher standardized effects than other variables. In particular, the exchange rate has the highest impact, while the effects of factors such as temperature and year are pretty low. The Pareto chart visually highlights the effects of these variables, indicating which factors have a more significant impact on electricity prices and, therefore, which variables should be considered more in policy and strategy development processes.

**Table 5:** Performance measurement values of RF and kNN algorithms.

| Model | MSE | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| RF | 0.013 | 0.112 | 0.081 | 0.087 | 0.919 |
| *kNN* | 0.022 | 0.150 | 0.126 | 0.172 | 0.856 |

Table 4 presents the performance metrics of the RF and kNN algorithms for estimating electricity prices. The RF algorithm exhibits lower values across all error metrics than the kNN algorithm. Specifically, RF has an MSE of 0.013 and an RMSE of 0.112, notably lower than kNN's MSE of 0.022 and RMSE of 0.150. This demonstrates that RF's predictions are closer to the actual values with less error variance. The MAE for RF is 0.081, while kNN shows a higher MAE of 0.126. Similarly, the MAPE for RF is 0.087, compared to kNN's 0.172, reflecting that RF's predictions are more accurate than kNN's values. These lower error metrics suggest that RF's predictions are more precise and less variable than kNN's.

Regarding explanatory power, the $R^2$ value for RF is 0.919, significantly higher than kNN's $R^2$ of 0.856. The $R^2$ value indicates how well the model explains the variance in the dependent variable, with a value closer to 1 representing better explanatory power. RF's higher $R^2$ value describes the high significance proportion of the variance in electricity prices, thus offering a more reliable and accurate prediction model. The overall superior performance of RF is attributed to its ensemble approach, which combines multiple decision trees to make predictions, thus reducing overfitting and improving generalizability. In contrast, kNN's performance is more sensitive to the choice of neighbors and distance metrics, which may contribute to its relatively higher error values and lower $R^2$. Therefore, based on these performance metrics, the RF algorithm demonstrates better accuracy and reliability than kNN in predicting electricity prices.

The prediction data in Figure 4 was created to compare the performance of RF and kNN algorithms with actual data. In general, the Random Forest algorithm makes more accurate predictions than kNN. While the first prediction data has an actual value of 0.796131, RF's prediction is 0.798039, and kNN's prediction is 0.555623, which shows that RF's predictions are closer than kNN's. In addition, kNN'spredictions have a higher error margin than RF's for some values. While the actual value of the 4th observation is 0.751068, RF's prediction is 0.798039, and kNN's prediction is 0.471734. These results show that the RF algorithm generally provides more accurate predictions, and kNN may sometimes lead to significant deviations.

In Figure 5, it is possible to compare the performances of RF and kNN algorithms based on error values. While the error values of the RF algorithm generally fluctuate positively, the kNN algorithm shows both positive and negative error values. The error values of RF vary between small positive values and a few negative values, indicating that the model mostly experiences minor deviations in its predictions. Still, sometimes it can make more significant errors. In particular, the highest error value of RF is an essential positive value of 0.2429, while the lowest is a significant negative value of -0.1512. This indicates that RF can make substantial errors, but generally gives small error values.

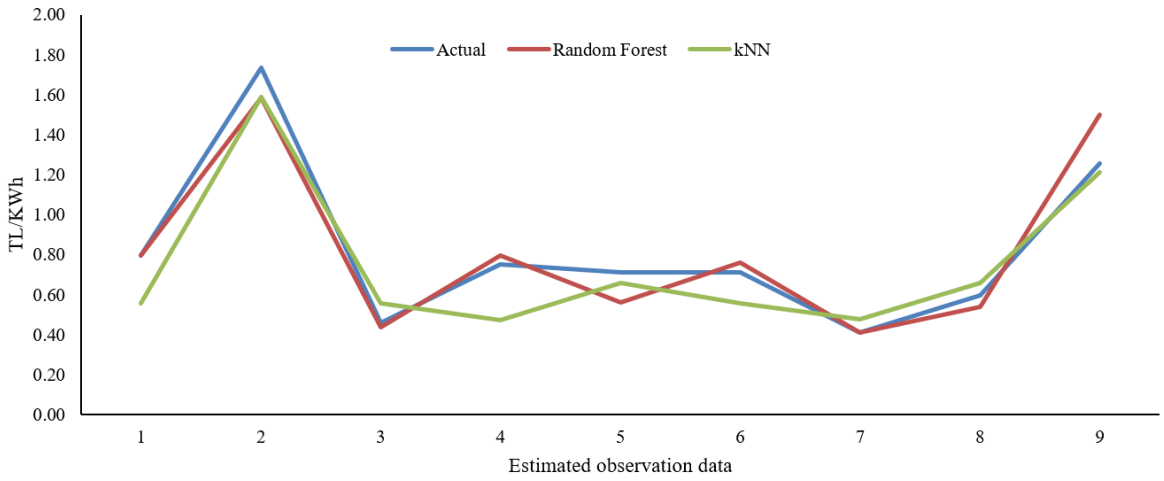


**Figure 4:** Comparison of RF and kNN algorithms based on actual data.

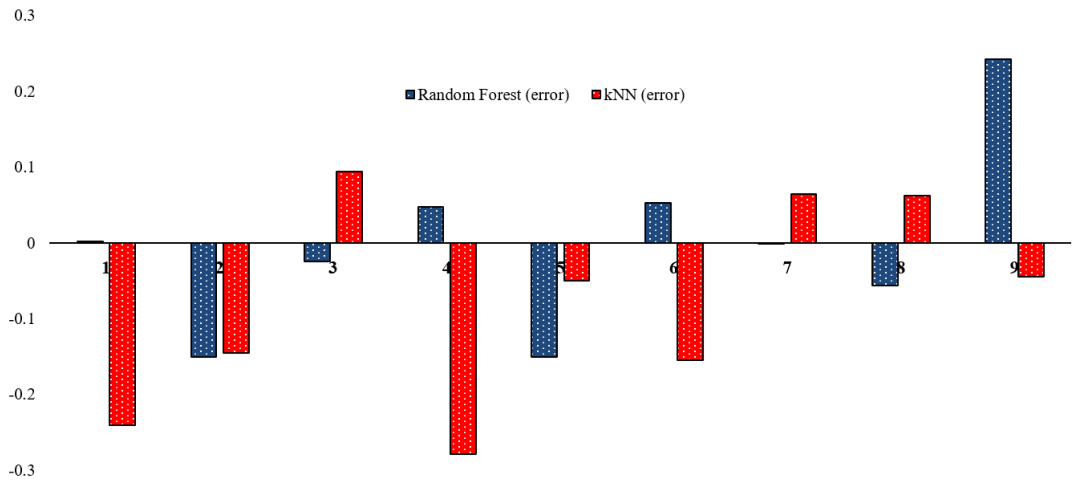


**Figure 5:** The margin of error between the estimated data of the dependent variable and the actual data.

In the kNN algorithm, the error values cover a more comprehensive range and include both positive and negative values. This represents that there is more variety and fluctuation in the predictions of kNN. In particular, the lowest error value of kNN is a tremendous negative value of -0.279334, while the highest is a much smaller positive value of 0.0944888. This shows that kNN can make substantial errors in some examples, but it results in smaller and more negative error values overall. In general, the error values of RF are less variable than kNN, but the error values of kNN are more diverse and extensive. This difference is due to the more stable performance of RF and the more variable performance of kNN.

The results show a significant difference between the performances of RF and kNN algorithms used to predict electricity unit prices in the Çanakkale region. RF algorithm performs significantly better than kNN in all error metrics. Specifically, the MSE and RMSE values of RF are 0.013 and 0.112, respectively, while the values of kNN are 0.022 and 0.150, respectively. In addition, MAE and MAPE values of RF are 0.081 and 0.087, respectively, while MAE and MAPE values of kNN are 0.126 and 0.172, respectively. $R^2$ value of RF is calculated as 0.919, which shows that the model explains a large portion of the variance in electricity prices and has high explanatory power. In contrast, the $R^2$ value of kNN is 0.856. The higher $R^2$ value of RF indicates that the model makes more accurate predictions and provides a more reliable model. The superior performance of RF is based on the ensemble method that combines multiple decision trees. This method reduces the risk of overfitting and increases generalizability. The performance of kNN, on the other hand, varies depending on the number of neighbors and distance measures, resulting in higher error values and lower $R^2$.

Analysis of the error values reveals that kNN's predictions are more variable than RF's. While the error values of RF generally contain minor positive deviations and some large negative values, the error values of kNN are distributed in a wide range, both positive and negative. The highest error value of RF is a significant positive value of 0.24286, while the lowest error value is -0.151205. RF can make significant errors, but generally provides small error values. The lowest error value of kNN is an essential negative value of -0.279334, while the highest error value is 0.0944888. This wide error range and high variability reveal that the predictions of kNN are less stable and may cause significant deviations in some cases. In general, it is seen that RF provides more reliable and precise predictions, while kNN results in higher error variability and more significant deviations. Therefore, RF is a more suitable model for predicting electricity prices than kNN, which gives less reliable results.

Finally, ML algorithms are important in predicting electricity unit prices. The comparison of RF and kNN algorithms showed that RF provides more reliable predictions with lower error values and higher explanatory power. The lower MSE, RMSE, MAE, and MAPE values of RF reveal that the model predicts electricity prices more precisely and accurately. This provides energy producers and distribution companies a significant advantage regarding cost estimations and strategic decisions. Accurate prediction of electricity prices reduces uncertainties in the energy sector, optimizes costs, and improves energy consumption strategies [25]. It also provides opportunities for consumers to save money and manage their energy consumption. Therefore, the superior performance of the RF algorithm in electricity price prediction provides significant contributions in terms of economic and

environmental sustainability by increasing the exact predictability in the energy market.

This study has certain limitations that must be acknowledged. The dataset is limited to the Çanakkale region and reflects regional economic and environmental dynamics, limiting the results' spatial generalizability. The RF model, which is effective in estimation, limits its practical application in decision-making processes requiring clear causal inferences regarding feature interpretability. In addition, the models are built based on quantitative variables only and do not include exogenous disruptions such as policy interventions, energy market regulations, or geopolitical events. These factors significantly affect electricity pricing, and their exclusion narrows the analytical scope of the study.

## 4. Conclusion

This study provides significant scientific and innovative contributions by examining environmental, economic, and oil price factors with ML algorithms to predict electricity unit prices in the Çanakkale region. The performances of RF and kNN algorithms were compared using the dataset between 2015 and 2024. The key findings of this study are listed below:

- The RF algorithm outperforms kNN in all performance metrics, achieving lower MSE (0.013 vs. 0.022) and higher $R^2$ (0.919 vs. 0.856), indicating superior prediction accuracy.
- Among the independent variables, exchange rate and PPI are the most influential factors on electricity unit prices based on ANOVA and Pareto analyses.
- The RF model demonstrates a more stable error distribution, while the kNN model shows higher variance and larger deviations from actual values.
- Machine learning techniques, particularly RF, are practical tools for forecasting electricity prices and supporting data-driven energy policy decisions in regional markets.

From a scientific perspective, this study demonstrates the applicability and effectiveness of ML methods to energy pricing problems. This reveals the potential of ML techniques to improve prediction and analysis processes in complex and dynamic systems such as the energy sector. In addition, this study examines the effects of environmental and economic factors on electricity prices in detail, providing vital information on how these factors should be considered in energy management strategies.

In terms of an innovative approach, the study went beyond traditional statistical estimation methods and developed more dynamic and highly accurate models using ML algorithms. This innovative approach helps better understand the potential and benefits of ML in energy pricing and policy development processes. As a result, this study scientifically verifies the applicability and effectiveness of ML techniques in the energy sector and offers an innovative method for energy price prediction. The findings allow energy managers and policymakers to make more informed and effective decisions on increasing energy efficiency and reducing costs. Continuing such studies will contribute to developing more accurate and reliable forecasting models in the energy field and establishing sustainable energy policies.

## Article Information

**Conflict of Interest/Common Interest:** The author has declared no conflict of interest.

## References

**[1]** W. Strielkowski, L. Civín, E. Tarkhanova, M. Tvaronavičienė, and Y. Petrenko, "Renewable energy in the sustainable development of electrical power sector: A review," Energies, vol. 14, no. 24, p. 8240, 2021.

**[2]** B. Li, "Effective energy utilization through economic development for sustainable management in smart cities," Energy Reports, vol. 8, pp. 4975–4987, 2022.

**[3]** A. Q. Al-Shetwi, "Sustainable development of renewable energy integrated power sector: Trends, environmental impacts, and recent challenges," Sci. Total Environ., vol. 822, p. 153645, 2022.

**[4]** M. Kiasari, M. Ghaffari, and H. H. Aly, "A Comprehensive Review of the Current Status of Smart Grid Technologies for Renewable Energies Integration and Future Trends: The Role of Machine Learning and Energy Storage Systems," Energies, vol. 17, no. 16, p. 4128, 2024.

**[5]** I. Antonopoulos et al., "Artificial intelligence and machine learning approaches to energy

demand-side response: A systematic review," Renew. Sustain. Energy Rev., vol. 130, p. 109899, 2020.

**[6]** T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," Sustain. Cities Soc., vol. 55, p. 102052, 2020.

**[7]** N. Chabouni, Y. Belarbi, and W. Benhassine, "Electricity load dynamics, temperature and seasonality Nexus in Algeria," Energy, vol. 200, p. 117513, 2020.

**[8]** S. Shahzad and E. Jasińska, "Renewable Revolution: A Review of Strategic Flexibility in Future Power Systems," Sustainability, vol. 16, no. 13, p. 5454, 2024.

**[9]** M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and models for electric load forecasting: a comprehensive review," Logist. Sustain. Transp, vol. 11, no. 1, pp. 51–76, 2020.

**[10]** Z. Ceylan, "Assessment of agricultural energy consumption of Turkey by MLR and Bayesian optimized SVR and GPR models," J. Forecast., Mar. 2020, doi: 10.1002/for.2673.

**[11]** Y. A. Atalan and A. Atalan, "Integration of the Machine Learning Algorithms and I-MR Statistical Process Control for Solar Energy," Sustain., vol. 15, no. 18, Sep. 2023, doi: 10.3390/su151813782.

**[12]** P. Choudhury, R. T. Allen, and M. G. Endres, "Machine learning for pattern discovery in management research," Strateg. Manag. J., vol. 42, no. 1, pp. 30–57, 2021.

**[13]** R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," J. Build. Eng., vol. 45, p. 103406, 2022.

**[14]** T. Ahmad, R. Madonski, D. Zhang, C. Huang, and A. Mujeeb, "Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm," Renew. Sustain. Energy Rev., vol. 160, p. 112128, 2022.

**[15]** G. Wang et al., "The relationship among oil prices volatility, inflation rate, and sustainable economic growth: Evidence from top oil importer and exporter countries," Resour. Policy, vol. 77, p. 102674, 2022.

**[16]** M. M. Ghiasi and S. Zendehboudi, "Application of decision tree-based ensemble learning in the classification of breast cancer," Comput. Biol. Med., vol. 128, p. 104089, 2021.

**[17]** M. Suyal and P. Goyal, "A review on analysis of K-nearest neighbor classification machine learning algorithms based on supervised learning," Int. J. Eng. Trends Technol., vol. 70, no. 7, pp. 43–48, 2022.

**[18]** E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review," J. Data Anal. Inf. Process., vol. 8, no. 4, pp. 341–357, 2020.

**[19]** I. Ullah, K. Liu, T. Yamamoto, R. E. Al Mamlook, and A. Jamal, "A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability," Energy Environ., vol. 33, no. 8, pp. 1583–1612, 2022.

**[20]** T. O. Hodson, T. M. Over, and S. S. Foks, "Mean squared error, deconstructed," J. Adv. Model. Earth Syst., vol. 13, no. 12, p. e2021MS002681, 2021.

**[21]** D. P. Kiley, S. Haley, B. Saylor, and B. L. Saylor, "The Value of Evidence-Based Computer Simulation of Oral Health Outcomes for Management Analysis of the Alaska Dental Health Aide Program," Institute of Social and Economic Research, University of Alaska Anchorage, 2008. [Online]. Available: http://hdl.handle.net/11122/4459

**[22]** S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," PLoS One, vol. 18, no. 2, p. e0279774, 2023.

**[23]** S. Walker, W. Khan, K. Katic, W. Maassen, and W. Zeiler, "Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings," Energy Build., vol. 209, p. 109705, 2020.

**[24]** J. Gao, "R-Squared (R2)–How much variation is explained?," Res. Methods Med. Heal. Sci., vol. 5, no. 4, pp. 104–109, 2024.

**[25]** T. Ahmad et al., "Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities," J. Clean. Prod., vol. 289, p. 125834, 2021.