Prediction of the Gross Electricity Generation Amount with Energy Sources Using the M5P Decision Tree Algorithm

Enes FİLİZ[1]

Article Type

Research Article

Application Date

2025.12.16

Acceptance Date

2026.03.12

DOI

10.53306/klujfeas.1841997

JEL Code

C00, C44, C450

Öz

Energy is considered an important issue for countries. It is stated that countries that can provide sufficient energy generation offer their people a more prosperous life. Electricity, which is at the forefront of energy generation, is regarded as an indispensable element of daily life. The electricity generation sector is characterized by a dynamic and complex structure. It is seen as the cornerstone of modern life and social progress in all areas such as the economy, health, and sports. The insurance of sufficient electricity generation is viewed as an important indicator of development for countries. In this context, this study focuses on gross electricity generation. A prediction of Türkiye's gross electricity generation value is aimed to be presented. Based on data from 1985 to 2020, predictions for the period 2021-2024 were made. The M5P decision tree algorithm from machine learning algorithms was used in the study. The Pairwise correlation feature selection algorithm was used for selecting the variables that affect the prediction of gross electricity generation value. Approximately 71% accuracy was achieved in the predictions made with all variables, and 86% accuracy was achieved in the predictions made with the effective variables. The findings reveal that utilizing effective variables significantly enhances predictive performance, with the MAPE value dropping to a notable 2.97%. A key empirical contribution of this study is the demonstration that 'Renewable energy and waste' exerts a more substantial influence on Türkiye's gross electricity generation than traditional energy sources. Consequently, these results offer a robust strategic framework for energy policymakers in navigating the national energy transition. It was observed that performance was increased in the predictions made with effective variables. Furthermore, other effective variables for the prediction of gross electricity generation were found to be 'Net consumption', 'Total installed capacity', and 'Liquid fuels'.

Keywords: Prediction, Gross Electricity Generation, Energy Sources, M5P Decision Tree Algorithm

Citation: Filiz, E., (2026). Prediction of the gross electricity generation amount with energy sources using the M5P decision tree algorithm. *Kırklareli University Journal of the Faculty of Economics and Administrative Sciences, 15*(1), 1-18. https://doi.org/10.53306/klujfeas.1841997

Statement: This study was prepared in accordance with the values of "Research and Publication Ethics". All responsibility for the study belongs to the author(s). The statements of research and publication ethics of the study are given on the last page.

[1] Assist. Prof., Balıkesir University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Quantitative Methods, enes.filiz@balikesir.edu.tr, ORCID: 0000-0002-8006-9467

M5P Karar Ağacı Algoritması Kullanılarak Enerji Kaynakları ile Brüt Elektrik Üretim Miktarının Tahmini

Enes FİLİZ[2]

Öz

Enerji konusu ülkeler açısından önemli bir durumdur. Enerji üretimini yeteri düzeyde sağlayabilen ülkelerin halkına daha refah bir hayat sunduğu söylenebilir. Enerji üretiminde ön planda olan elektrik, günlük hayatın vazgeçilmez bir öğesidir. Elektrik üretim sektörü dinamik ve karmaşık bir yapıdadır. Ekonomi, sağlık, spor vb. gibi tüm alanlarda modern yaşamın ve toplumsal ilerlemenin temel taşı olarak görülmektedir. Elektrikte üretimi yeterli derecede sağlamak ülkeler açısından önemli bir gelişmişlik göstergesidir. Bu bağlamda çalışma brüt elektrik üretimi konusuna odaklanılmıştır. Türkiye'nin brüt elektrik üretim değerinin tahmininin ortaya konulması amaçlanmıştır. 1985-2020 yılları arasındaki verilere dayanarak, 2021-2024 yılları arasının tahminleri yapılmıştır. Çalışmada makine öğrenmesi algoritmalarından M5P karar ağacı algoritmasından yararlanılmıştır. Brüt elektrik üretim değerinin tahminine etki eden değişkenler için ikili korelasyon öznitelik seçim algoritması kullanılmıştır. Tüm değişkenler ile yapılan tahminlerde yaklaşık %71, etkili değişkenler ile yapılan tahminlerde %86 başarı elde edilmiştir. Etkili değişkenler ile yapılan tahminlerde performansın arttığı ve MAPE değerinin %2.97 seviyesine gerilediği görülmüştür. Çalışma sonucunda, Türkiye'nin brüt elektrik üretiminde 'Yenilenebilir enerji ve atıklar' değişkeninin geleneksel kaynaklardan daha yüksek bir etki düzeyine sahip olduğu ampirik olarak kanıtlanmıştır. Bu bulgular, enerji politika yapıcıları için stratejik bir rehber niteliğindedir. Ayrıca brüt elektrik üretim değerinin tahminine etki eden diğer değişkenler olarak 'Net tüketim', 'Toplam kurulu güç' ve 'Sıvı yakıtlar' bulunmuştur.

Makale Türü

Araştırma Makalesi

Başvuru Tarihi

16.12.2025

Kabul Tarihi

12.03.2026

DOI

10.53306/klujfeas.1841997

JEL Kodu

C00, C44, C450

Anahtar Kelimeler: Tahmin, Brüt Elektrik Üretimi, Enerji Kaynakları, M5P Karar Ağacı Algoritması

Atıf: Filiz, E., (2026). M5P karar ağacı algoritması kullanılarak enerji kaynakları ile brüt elektrik üretim miktarının tahmini. *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *15*(1), 1-18. https://doi.org/10.53306/klujfeas.1841997

Bilgilendirme: Bu çalışma "Araştırma ve Yayın Etiği" değerlerine uygun olarak hazırlanmıştır. Çalışmanın tüm sorumluluğu yazar(lar)a aittir. Çalışmaya ait araştırma ve yayın etiği beyanlarına son sayfada yer verilmiştir.

[2] Dr. Öğr. Üyesi, Balıkesir Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Sayısal Yöntemler Anabilim Dalı, enes.filiz@balikesir.edu.tr, ORCID: 0000-0002-8006-9467

Introduction

In the globalizing world order, different responses are given to events occurring worldwide. Countries must take immediate action in response to daily developments, and individuals organize their lives according to ongoing events. Under the banner of sustainability, all developments show that every action has a counterpart. The concept of sustainability is at the forefront of many areas, such as energy, technology, economy, and sports, and its popularity grows day by day. Sustainability issues are considered among the most critical agenda items for countries. The facilitation of life and the ability of people to live more comfortably are viewed as the direct results of sustainability. Additionally, sustainability presents significant opportunities for the revitalization of dwindling resources. The spread of recycling processes, particularly in renewable energy, demonstrates that vital investments are being made for future generations. Energy is currently regarded as the most important pillar of sustainability, as meeting the need for electricity is essential for the continuity of life.

Electricity, as the primary source of energy, plays a vital role in areas such as the economy, development, and sustainability. It is regarded as a cornerstone of modern life and social progress, meeting fundamental human needs such as shelter, heating, and lighting. In addition to daily comfort, essential sectors including healthcare, communication, and transportation depend heavily on a stable electricity supply. Recently, the popularity of electric vehicles has demonstrated that transportation development is progressing more robustly through electrification. From a broader perspective, a country's independence in electricity generation is viewed as a key indicator of its development level. In terms of sustainability, electricity generation from renewable sources—such as solar, wind, and geothermal—produces fewer carbon emissions compared to fossil fuels. This transition to clean energy is highlighted as a critical factor for environmental and climatic preservation. Given this framework, ensuring a safe and uninterrupted electricity supply from renewable sources directly affects both the quality of life and national development. In recent years, the energy sector has generated a wealth of data that has reached enormous proportions. Since the energy issue was brought to the forefront, data sets have been regularly collected and reported by countries. For the analysis of such large-scale data, machine learning, deep learning, and ensemble learning algorithms are considered the most appropriate tools. As the basic working principle of these algorithms relies on extensive datasets, they facilitate the analysis of complex results. Furthermore, previous analyses have revealed that these methods often achieve successful classifications and predictions.

Various studies have been conducted on the subject of energy and electricity generation in the literature. The predicted external costs of electricity generation in China were determined by Zhang et al. (2007) within the framework of different scenarios for long-term energy and environmental policies. The effects of price risks on policy design were examined by Gross et al. (2010) based on recent research by the UK Energy Research Centre. The predicted cost using different generation options was compared with the return distribution to which each was exposed, taking into account electricity price fluctuations. It was aimed by Kucukali and Baris (2010) to predict Türkiye's short-term annual gross electricity demand by applying fuzzy logic; general information about the country's economic, political, and electricity market conditions was also intended to be provided. Whether a pool market could achieve its goals of increasing competition and lowering electricity prices was investigated by O'Mahoney and Denny (2013). The causal relationship between economic growth and electricity generation from renewable energy sources in 20 OECD countries between 1990 and 2008 was aimed to be revealed by Ohler and Fetters (2014). A panel dataset covering 174 countries from 1980 to 2012 was used by Atems and Hotaling (2018) to determine the effect of electricity generation on economic growth. The relationship between population, gross domestic product (GDP) growth, electricity generation, electricity consumption, and carbon emissions was examined by Ali et al. (2020). The factors affecting carbon dioxide emissions from electricity generation in the European Union between 2000 and 2018 were investigated by Karmellos et al. (2021) using decomposition analysis and the LMDI-I in three time periods. The relationship between total renewable electricity generation, total hydroelectric generation, and carbon dioxide emissions for China was examined by Xiaosan et al. (2021). A study quantifying carbon emissions associated with the generation of electricity generated and consumed in European countries was presented by Scarlat et al. (2022). The development of gross nuclear electricity generation in terms of stability and predictability both globally and within the European Union was examined by Bórawski et al. (2024).

Various studies have been conducted in the literature on machine learning methods related to energy and electricity generation. An innovative hybrid approach to determine long-term electricity demand forecasts was presented by Mostafavi et al. (2013). It was aimed by Tüfekci (2014) to develop a forecasting model to predict the hourly full-load electrical power output of a cycle power plant, and several machine learning regression algorithms were compared for this purpose. An electricity generation forecasting system that could determine the amount of power required at a rate close to the electricity consumption of the United States was proposed by Rahman et al. (2016) using machine learning. Artificial neural networks (ANN), multiple linear regression (MLR), adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM) machine learning algorithms were used by Solyali (2020) to predict electricity demand in Cyprus and to determine important criteria for electricity generation.

It was aimed by Santarisi and Faouri (2021) to predict the full-load electricity power output of a combined-cycle power plant using different machine learning algorithms, including linear regression, ridge regression, lasso regression, elastic net regression, random forest regression, and gradient boosting regression. The seasonal and periodic characteristics of China's energy generation structure were examined by Lin and Shi (2022), and a machine learning algorithm was utilized to make monthly forecasts for the next five years. Long short-term memory (LSTM) was utilized by Bilgili and Pinar (2023) to forecast gross electricity consumption in Türkiye. The LSTM model was also aimed to be compared with the seasonal autoregressive integrated moving average (SARIMA) model to determine gross energy usage. A hybrid model was proposed by Zhang et al. (2024) to forecast the total electricity power load in combined cycle power plants using key input variables. Five different machine learning algorithms, including categorical boosting, histogram-based gradient boosting regression, extreme gradient boosting regression, light gradient boosting machine learning algorithm, and support vector regression, were used. It was aimed by Atalan et al. (2025) to accurately predict the renewable energy generation rate to meet Türkiye's electricity demand from renewable energy sources using machine learning algorithms such as Random Forest, Adaptive Boosting, and Gradient Boosting.

The main objective of this study is to present a prediction of Türkiye's gross electricity generation value. Based on data from 1985 to 2020, a prediction for the period 2021-2024 was calculated and compared with real values. Energy sources and power plant information were utilized for this purpose. The prediction performance of the M5P decision tree algorithm was investigated using both all variables and effective variables, and relevant comparisons were made. Additionally, the Pairwise Correlation feature selection algorithm was used to determine the effective variables.

1. Materials And Method

1.1. Data Set

To achieve the primary objective of this study and within the framework of the literature, the focus will be on Türkiye's gross energy generation. Gross energy generation will be predicted using Electricity generation and shares by energy resources and Power Plants Information. In this regard, the variables used in the study were obtained from the Turkish Statistical Institute (TÜİK) data portal website (TÜİK Web Page)(Turkstat, 2025). The experimental framework incorporated seven independent variables, as detailed in Table 1. The primary dataset encompassed the period between 1985 and 2020 for model calibration. Subsequently, the values for 2021–2024 were utilized as an out-of-sample test set to validate the predictive accuracy of the M5P algorithm against real-world observations. To account for the varying scales of the input variables (percentages vs. absolute values), a data normalization process was performed prior to the modeling stage to ensure numerical stability and prevent bias toward higher-magnitude features.

Table 1. Variables Used in the Study

Variable Name	Variable Property	Variable Type
Coal	Electricity generation and shares by energy resources (%)	Independent Variables
Liquid fuels		
Natural Gas		
Hydro		
Renewable Energy and wastes		
Total power installed (MW)	Power Plants Information	
Net consumption (GWh)		
Gross electricity generation (GWh)		Dependent Variables

1.2. M5P Decision Tree Algorithm

The aim of this study is to predict Türkiye's gross electricity generation values. To this end, machine learning algorithms, which have gained increasing popularity in recent years within the literature, will be utilized. Different machine learning algorithms have been applied to various topics, yielding successful results. The M5P decision tree algorithm, one of the machine learning algorithms that has achieved successful results in classification and prediction in the literature, will be used in this study (Behnood et al., 2017; Blaifi et al., 2018; Nhu et al., 2020; Mujammal et al., 2025).

The M5P decision tree algorithm is a machine learning algorithm based on a model tree structure. It is used to solve regression problems, perform classifications, and make future predictions. It implements the M5 model tree algorithm developed by Quinlan (1992). Wang and Witten (1996) developed the M5P algorithm based on the M5 algorithm. Unlike standard regression trees, this algorithm uses linear regression models instead of fixed values at the leaf nodes. This is where the main difference lies. The primary advantages of this algorithm include high predictive accuracy is interpretable and easy to apply, and can deliver fast results on large data sets (Quinlan, 1992; Wang and Witten, 1996; Behnood et al., 2017). The M5P decision tree algorithm, which is part of the Weka program used in machine learning analyses, also demonstrates high success in application. The M5P algorithm selects the feature that most reduces the standard deviation (SDR) of the data at each node when constructing the tree structure. The SDR value is calculated with the following formula:

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} * sd(T_i) \tag{1}$$

In this equation, $sd(T)$ represents the standard deviation of the samples within the node, while T_i denotes the subsets resulting from the branching process. This criterion ensures that the prediction error is minimized at each successive step of the tree construction. In addition, during the prediction phase, the models at the sub-nodes are passed through a smoothing filter towards the main nodes to increase the generalization ability of the model.

The M5P algorithm, implemented in the Weka workbench as a powerful version of the model tree architecture, operates by recursively partitioning the data space into subsets. Unlike standard decision trees that generate constant values at the leaves, Weka's M5P builds multivariate linear regression models within each segment. This hybrid structure is superior to traditional methods because it bridges the gap between the interpretability of decision trees and the mathematical precision of regression. By decomposing complex non-linear global patterns through its tree architecture and modeling each segment with specific linear equations, the algorithm ensures high consistency and robustness, particularly in forecasting dynamic data such as electricity generation.

1.3. Pairwise Correlation Feature Selection Algorithm

In line with the objective of the study, it is crucial to identify the factors that influence Türkiye's gross electricity generation values. Rather than using too many variables in large data sets, it is necessary to determine the variables that significantly the outcome of the study. Using fewer variables is significant in terms of both time and cost. Achieving success close to or better than that obtained with all variables using fewer variables demonstrates the great need for feature selection. In this context, the Pairwise Correlation Feature Selection Algorithm, one of the machine learning feature selection algorithms frequently used in the literature, will be used in this study (Bolón-Canedo et al., 2015; Gere et al., 2015; Jiménez et al., 2021).

The Pairwise Correlation Feature Selection Algorithm is a version of the Correlation Feature Selection Algorithm. It is used in machine learning to eliminate unnecessary features. This method, used in feature selection, works based on correlation. It calculates the value of a feature i by summing the values of the feature subsets formed by feature i and each of the other features. The key point here is that the ranking is based on low correlation with other features and high correlation with the class when making the selection. Another point is that most feature selection algorithms are based on algorithms designed to find pairwise correlations (Bolón-Canedo et al., 2015; Gere et al., 2015; Jiménez et al., 2021; AlNuaimi et al., 2022). In the Weka program, it is used as 'PairwiseCorrelationAttributeEval'.

1.4. Prediction Model Performance Measurement Criteria

The M5P decision tree algorithm will be used to predict Türkiye's gross electricity generation values for the purpose of this study. To evaluate the success of the results provided by this algorithm, certain prediction model performance criteria will be utilized. The performance of the prediction models obtained using both all variables and the effective variables obtained through the Pairwise Correlation Feature Selection Algorithm will be compared using these criteria. A review of the literature reveals that the criteria frequently used in studies examining the performance of machine learning prediction models are Direction Accuracy (DAC), Relative Absolute Error (RAE), and Mean Absolute Percentage Error (MAPE). DAC is a criteria that measures the performance of the prediction model and is considered to yield better results the higher it is. It mostly indicates the prediction direction (Meade, 2002; Costantini et al., 2016). RAE is one of the criteria that measures the performance of the prediction model. Since this criterion is error-oriented, it is expected to be as low as possible. It reveals the absolute error between the real value and the predicted value (Makridakis and Hibon, 2000; Reich et al., 2016). MAPE is a measure that assesses the performance of the predict model based on errors and shows this as a percentage. The lower it is, the more successful the results are (De Myttenaere et al., 2016; Liantoni and Agusti, 2020). The formulas are generally given below:

$$= DAC\frac{1}{n}\sum_{t=1}^{n} d_t * 100 \tag{2}$$

where d_t is an indicator function defined as:

$$d_t = \begin{cases} 1 & if\ (y_{t+1} - y_t)(\hat{y}_{t+1} - y_t)\ > 0 \\ 0 & otherwise \end{cases}$$

$$RAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{\sum_{i=1}^{n} |\bar{y}_i - y_i|} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |\frac{y_t - \hat{y}_t}{y_t}| * 100 \tag{4}$$

2. Application

The application section was designed to predict Türkiye's gross electricity generation values in line with the objective of the study. In this context, the application process was carried out in three steps. In the first step, a prediction model was created using the M5P decision tree algorithm with all variables employed, and the initial predicted values were determined. In the second step, the effective variables for gross electricity generation were identified through the Pairwise Correlation Feature Selection Algorithm. This attribute selection process was performed using only the 1985-2020 training period. In the third step, a refined prediction model was developed using the M5P decision tree algorithm based on the identified effective variables, and the final prediction values were determined. These applications were performed using the Weka software, which incorporates machine learning algorithms and covers tasks such as classification and prediction. The parameters of the M5P decision tree algorithm are given in Table 2. The data were normalized within the scope of the M5P algorithm. For the training and testing of the data, the k-fold cross-validation procedure was utilized. It is a method used to measure the performance of a machine learning model more reliably. The dataset is divided into k equal parts. At each step, one part is reserved for testing, while the remaining k-1 parts are used for training. This process is repeated k times (Filiz, 2023). The final results are obtained by taking the average of the performance values achieved. The model is trained on data from the 1985–2020 period to predict values for 2021, 2022, 2023, and 2024, with performance validated through a 10-fold cross-validation approach. No changes will be made to the hyperparameters. The Research Flowchart of the Study is provided below:

Chart 1. Research Flowchart

Step 1: Initial Modeling (All Variables) • **Input:** All Independent Variables (1985-2020) • **Process:** M5P Decision Tree Algorithm • **Output:** Model Generation → Prediction → Initial Results
Step 2: Feature Engineering (Feature Selection) • **Input:** All Independent Variables • **Process:** Pairwise Correlation Feature Selection Algorithm • **Output:** Determination of **Effective Variables**
Step 3: Optimized Modeling (Effective Variables) • **Input:** Selected Effective Variables • **Process:** M5P Decision Tree Algorithm (Re-trained) • **Output:** Optimized Model → Final Prediction → Improved Results

Table 2. Parameters of the M5P Decision Tree Algorithm

Parameter	Value
Batch Size	100
Build Regression Tree	False
Debug	False
Do Not Check Capabilities	False
Min Num Instances	4.0
Num Decimal Places	4
Save Instances	False
Unpruned	False
Use Unsmooted	False

All parameters were selected to utilize the Model Tree function. In this way, linear equations were generated at the leaf nodes instead of fixed values.

3. Findings

The aim of the study and the designed application section is to determine the prediction of Türkiye's gross electricity generation values. The M5P decision tree algorithm was used for this purpose. In addition, the effective variables for the prediction of Türkiye's gross electricity generation values were determined using the Pairwise Correlation Feature Selection Algorithm.

In line with the first step of the study, the prediction model was determined using the M5P decision tree algorithm with the help of all variables and is given in Equation 4;

$$Estimated\ Value\ (gross\ electricity\ generation\ ) =$$
$$114{,}4967 * Natural\ gas - 493{,}649 * Renewable\ Energy\ and\ wastes - 0{,}1589 * Total\ power\ installed + 1{,}2451 * Net\ consumption\ + 1228{,}9855 \quad (4)$$

The performance metrics resulting from the initial modeling phase are summarized in Table 3.

Table 3. Prediction Performance Criteria of the M5P Decision Tree Algorithm When All Variables Are Used

	DAC	RAE	MAPE
M5P Algorithm	70,88%	104,81	7,71

According to Table 3, the analysis performed using the M5P decision tree algorithm resulted in a DAC value of 70,88%, an RAE value of 104,81, and a MAPE value of 7,71.

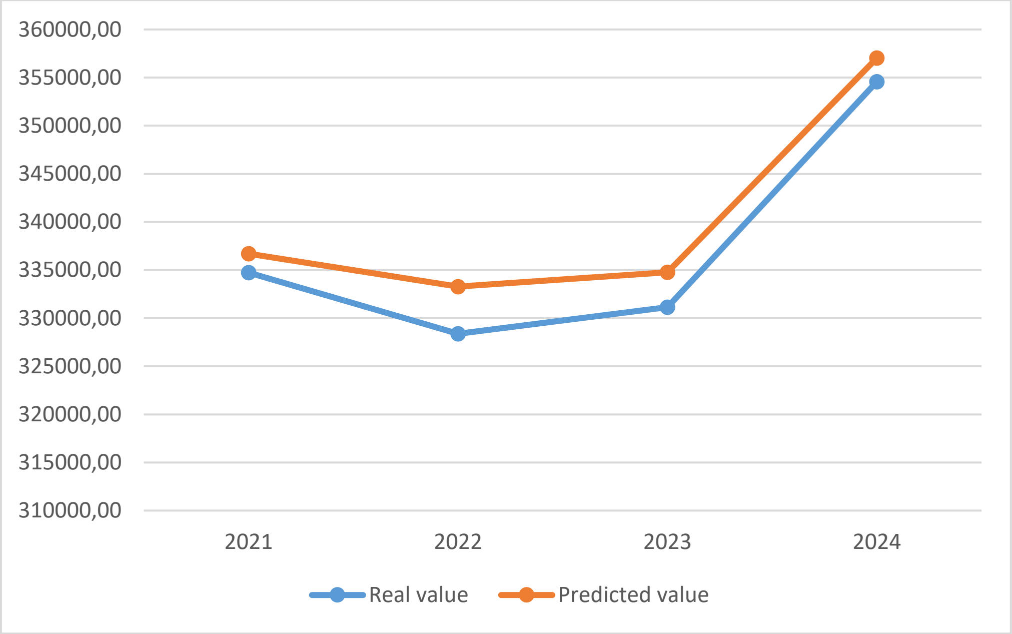
In line with the first step of the study, the prediction model was determined using the M5P decision tree algorithm with the help of all variables, and the obtained prediction values are given in Table 4.

Table 4. Predicted Values of the M5P Decision Tree Algorithm When All Variables Are Used

	Real Value	Predicted Value
2021	334,723,11	336,698,65
2022	328,379,34	333,282,48
2023	331,148,90	334,760,78
2024	354,570,21	357,023,63

In accordance with the first step of the study, a prediction model was determined using the M5P decision tree algorithm with all variables, and a graphical representation of the obtained prediction values and real values is provided in Figure 1.

Figure 1. Display of Real and Predicted Values in Line With The Model Created With All Variables



In accordance with Step 2 of the study, effective variables gross electricity generation are presented in Table 5 using the Pairwise Correlation Feature Selection Algorithm.

Table 5. Effective Variables Gross Electricity Generation and Their Levels of Importance

Effective variables	Impact level
Net consumption (GWh)	5,9926
Total power installed (MW)	5,8863
Renewable Energy and wastes	5,2747
Liquid fuels	5,0918

According to Table 5, the effective variables for predicting Türkiye's gross electricity generation value are determined as ‘Net consumption’, ‘Total power installed’, ‘Renewable Energy and wastes’, and ‘Liquid fuels’, respectively.

In line with the third step of the study, the prediction model was determined using the M5P decision tree algorithm with the help of effective variables and is given in Equation 5;

$$Estimated\ Value\ (gross\ electricity\ generation\ ) = \\ -1239{,}9423 * Renewable\ Energy\ and\ wastes + 1{,}2418 * Net\ consumption + 666{,}2701 \quad (5)$$

In accordance with the third step of the study, the prediction model was determined using the M5P decision tree algorithm with the help of effective variables, and the performance criteria values are given in Table 6.

Table 6. Prediction Performance Criteria of the M5P Decision Tree Algorithm in the Case of Using Effective Variables

	DAC	RAE	MAPE
M5P Algorithm	86,01%	42,66	2,97

According to Table 6, the analysis performed using the M5P decision tree algorithm resulted in a DAC value of 86,01%, an RAE value of 42,66, and a MAPE value of 2,97.

In line with the third step of the study, the prediction model was determined using the M5P decision tree algorithm with the help of effective variables, and the obtained prediction values are given in Table 7.

Table 7. Predicted Values of the M5P Decision Tree Algorithm in the Case of Using Effective Variables

	Real Value	Predicted Value
2021	334,723,11	332,989,47
2022	328,379,34	329,197,24
2023	331,148,90	331,016,82
2024	354,570,21	352,918,48

In line with the third step of the study, a prediction model was determined using the M5P decision tree algorithm with the help of effective variables, and a graphical representation of the obtained prediction values and real values is provided in Figure 2.

Figure 2. Display of Real And Predicted Values In Line With The Model Created With Effective Variables

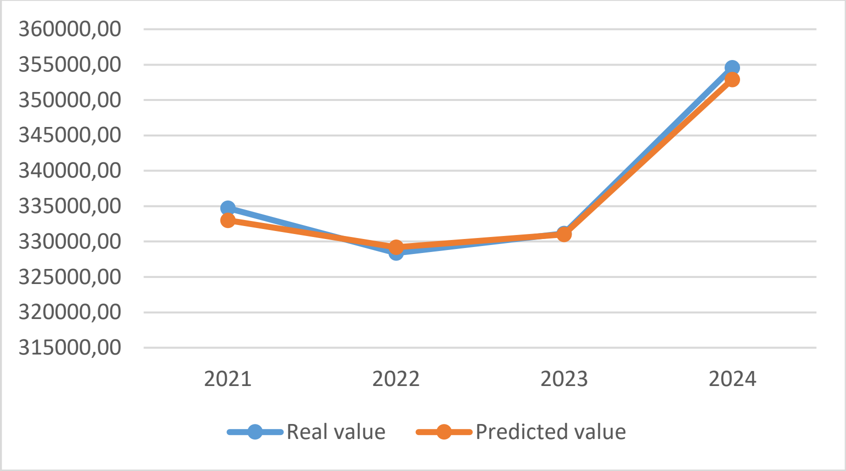


Table 8. Comparison Table of Step 1 and Step 3

	DAC	RAE	MAPE
Step 1 Results (All Variables)	70,88%	104,81	7,71
Step 3 Results (Effective Variables)	86,01%	42,66	2,97

4. Discussion and Conclusion

In the section concerning the aim of the study and the designed application, the prediction of Türkiye's gross electricity generation value is addressed as the primary objective. The M5P decision tree algorithm was used for this purpose. The prediction performance criteria of the utilized algorithm were established, and real values were compared with predicted values. In addition, the effective variables for the prediction of Türkiye's gross electricity generation value were identified using the Pairwise Correlation Feature Selection Algorithm. Finally, the predicted value of Türkiye's gross electricity generation was determined using the M5P decision tree algorithm with the assistance of the effective variables, and the results obtained with all variables were examined.

According to the results of the first step of the study, the M5P decision tree algorithm was found to accurately predict Türkiye's gross electricity generation value by approximately 71%. This process is considered valid when all variables are used. In the third step of the study, it was revealed that Türkiye's gross electricity generation value was accurately predicted at approximately 86% by the M5P decision tree algorithm when using effective variables. This finding underscores the critical role of feature selection in enhancing model performance. Furthermore, the degree to which prediction accuracy increased is visually demonstrated in Figures 1 and 2. In particular, the extent to which the predicted values found using effective variables were close to the real values is presented. Similarly, it was determined that lower RAE and MAPE values (indicating a decrease in errors) were yielded from the prediction using effective variables compared to the prediction using all variables. The successful prediction performance of the utilized M5P decision tree algorithm is demonstrated by these results. Similarly, existing literature corroborates the high performance of the M5P algorithm (Akgündoğdu et al. 2019; Saha et al. 2023; Mujammal et al. 2025). In the second step of the study, the effective variables for the prediction of Türkiye's gross electricity generation were successfully determined using the Pairwise Correlation Feature Selection Algorithm. These variables are listed as ‘Net consumption’, ‘Total power installed’, ‘Renewable Energy and wastes’, and ‘Liquid fuels’. The most noteworthy result is considered to be that the ‘Renewable Energy and Wastes’ variable is more important than traditional energy generation sources. It was also one of the two variables included in the prediction model in Equation (5). The coefficients in Equations 4 and 5 represent a mathematical reflection of the strategic transformation in Türkiye's energy portfolio. In particular, the high weight assigned to 'Renewable Energy and Waste' and 'Net Consumption' variables proves the decisive impact of policy shifts driven by localization and sustainability goals on gross generation capacity. These coefficients indicate that the model does not merely summarize historical data; rather, it reflects a strategic planning framework that aligns Türkiye's increasing industrial demand with the ongoing green energy transition. Similarly, the importance of the ‘Renewable Energy and Wastes’ variable in energy generation is mentioned in literature (Laureti et al. 2023; Guidi et al. 2023). Another variable included in the prediction model in Equation (2), the ‘Net consumption’ variable, has also been found in the literature to be an effective variable for gross electricity generation value predictions (Pamuk, 2016; Bakay and Başarslan, 2025). In general, energy is regarded as a very important indicator of development for countries. In this regard, the success of the M5P decision tree algorithm in predicting Türkiye's gross electricity generation value was demonstrated by the results of the study. It is shown that the M5P decision tree algorithm can be used in this type of data set. The findings provide a strategic framework for energy policy makers. The most significant contribution of this study to the literature is the concrete demonstration of the critical relationship between feature selection and predictive accuracy. Furthermore, by empirically proving that 'Renewable Energy and Wastes' is a more decisive variable than traditional sources, a novel perspective for energy policy-making is offered by the study. Unlike many models, an interpretable structure is provided by the M5P algorithm as a decision tree with linear regression functions at its leaf nodes; what sets this study apart is the validation of this hybrid structure's high performance specifically within the context of Türkiye’s energy data.

The study is subject to certain limitations. The dataset contains data only for Türkiye for the period of 1985-2020, whereas the subsequent period of 2021-2024 was reserved exclusively for model evaluation. DAC, RAE, and MAPE values were utilized as the primary prediction performance criteria. To develop the prediction model, the M5P decision tree algorithm was employed. Furthermore, the Pairwise Correlation Feature Selection Algorithm was applied to identify the most effective variables for predicting Türkiye's gross electricity generation. All results were obtained using the Weka software, which is a comprehensive platform for machine learning applications.

References

Akgündoğdu, A., Öz, I., & Uzunoğlu, C. P. (2019). Signal quality based power output prediction of a real distribution transformer station using M5P model tree. *Electric Power Systems Research*, *177*, 106003. https://doi.org/10.1016/j.epsr.2019.106003

AlNuaimi, N., Masud, M. M., Serhani, M. A., & Zaki, N. (2022). Streaming feature selection algorithms for big data: A survey. *Applied Computing and Informatics*, *18*(1/2), 113-135. https://doi.org/10.1016/j.aci.2019.01.001

Atalan, Y. A., Şahin, H., Keskin, A., & Atalan, A. (2025). Strategic forecasting of renewable energy production for sustainable electricity supply: A machine learning approach considering environmental, economic, and oil factors in Türkiye. *PLoS One*, *20*(8), e0328290. https://doi.org/10.1371/journal.pone.0328290

Atems, B., & Hotaling, C. (2018). The effect of renewable and nonrenewable electricity generation on economic growth. *Energy Policy*, *112*, 111-118. https://doi.org/10.1016/j.enpol.2017.10.015

Bakay, M. S., & Başarslan, M. S. (2025). Forecasting of Türkiye's net electricity consumption with metaheuristic algorithms. *Utilities Policy*, *95*, 101929. https://doi.org/10.1016/j.jup.2025.101929

Behnood, A., Behnood, V., Gharehveran, M. M., & Alyamac, K. E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Construction and Building Materials*, *142*, 199-207. https://doi.org/10.1016/j.conbuildmat.2017.03.061

Bilgili, M., & Pinar, E. (2023). Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. *Energy*, *284*, 128575. https://doi.org/10.1016/j.energy.2023.128575

Blaifi, S. A., Moulahoum, S., Benkercha, R., Taghezouit, B., & Saim, A. (2018). M5P model tree based fast fuzzy maximum power point tracker. *Solar Energy*, *163*, 405-424. https://doi.org/10.1016/j.solener.2018.01.071

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-based Systems*, *86*, 33-45. https://doi.org/10.1016/j.knosys.2015.05.014

Bórawski, P., Bełdycka-Bórawska, A., Klepacki, B., Holden, L., Rokicki, T., & Parzonko, A. (2024). Changes in gross nuclear electricity production in the European Union. *Energies*, *17*(14), 3554. https://doi.org/10.3390/en17143554

Costantini, M., Cuaresma, J. C., & Hlouskova, J. (2016). Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *Journal of Forecasting*, *35*(7), 652-668. https://doi.org/10.1002/for.2398

De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, *192*, 38-48. https://doi.org/10.1016/j.neucom.2015.12.114

Filiz, E. (2023). Evaluation of match results of five successful football clubs with ensemble learning algorithms. *Research Quarterly for Exercise and Sport*, *94*(3), 773-782. https://doi.org/10.1080/02701367.2022.2053647

Gere, A., Sipos, L., & Héberger, K. (2015). Generalized pairwise correlation and method comparison: Impact assessment for JAR attributes on overall liking. *Food Quality and Preference*, *43*, 88-96. https://doi.org/10.1016/j.foodqual.2015.02.017

Gross, R., Blyth, W., & Heptonstall, P. (2010). Risks, revenues and investment in electricity generation: Why policy needs to look beyond costs. *Energy Economics*, *32*(4), 796-804. https://doi.org/10.1016/j.eneco.2009.09.017

Guidi, G., Violante, A. C., & De Iuliis, S. (2023). Environmental impact of electricity generation technologies: a comparison between conventional, nuclear, and renewable technologies. *Energies*, *16*(23), 7847. https://doi.org/10.3390/en16237847

Jiménez, F., Sánchez, G., Palma, J., Miralles-Pechuán, L., & Botía, J. (2021). Multivariate feature ranking of gene expression data. *arXiv preprint* arXiv:2111.02357. https://doi.org/10.48550/arXiv.2111.02357

Karmellos, M., Kosmadakis, V., Dimas, P., Tsakanikas, A., Fylaktos, N., Taliotis, C., & Zachariadis, T. (2021). A decomposition and decoupling analysis of carbon dioxide emissions from electricity generation: Evidence from the EU-27 and the UK. *Energy*, *231*, 120861. https://doi.org/10.1016/j.energy.2021.120861

Kucukali, S., & Baris, K. (2010). Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach. *Energy Policy*, *38*(5), 2438-2445. https://doi.org/10.1016/j.enpol.2009.12.037

Laureti, L., Massaro, A., Costantiello, A., & Leogrande, A. (2023). The impact of renewable electricity output on sustainability in the context of circular economy: a global perspective. *Sustainability*, *15*(3), 2160. https://doi.org/10.3390/su15032160

Liantoni, F., & Agusti, A. (2020). Forecasting bitcoin using double exponential smoothing method based on mean absolute percentage error. *JOIV: International Journal on Informatics Visualization*, *4*(2), 91-95. http://dx.doi.org/10.30630/joiv.4.2.335

Lin, B., & Shi, L. (2022). New understanding of power generation structure transformation, based on a machine learning predictive model. *Sustainable Energy Technologies and Assessments*, *51*, 101962. https://doi.org/10.1016/j.seta.2022.101962

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451-476. https://doi.org/10.1016/S0169-2070(00)00057-1

Meade, N. (2002). A comparison of the accuracy of short term foreign exchange forecasting methods. *International Journal of Forecasting*, *18*(1), 67-83. https://doi.org/10.1016/S0169-2070(01)00111-X

Mostafavi, E. S., Mostafavi, S. I., Jaafari, A., & Hosseinpour, F. (2013). A novel machine learning approach for estimation of electricity demand: An empirical evidence from Thailand. *Energy Conversion and Management*, *74*, 548-555. https://doi.org/10.1016/j.enconman.2013.06.031

Mujammal, M. A. H., Moualdia, A., Boulkhrachef, S., Wira, P., Boudana, D., & Albasheri, M. A. (2025). Advancing wind energy conversion: smart maximum power point based on M5-Pruned algorithm for enhanced wind energy production. *Production Engineering*, *19*(2), 347-368. https://doi.org/10.1007/s11740-024-01315-w

Nhu, V. H., Shahabi, H., Nohani, E., Shirzadi, A., Al-Ansari, N., Bahrami, S., ... & Nguyen, H. (2020). Daily water level prediction of Zrebar Lake (Iran): A comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS International Journal of Geo-Information*, *9*(8), 479. https://doi.org/10.3390/ijgi9080479

Ohler, A., & Fetters, I. (2014). The causal relationship between renewable electricity generation and GDP growth: A study of energy sources. *Energy Economics*, *43*, 125-139. https://doi.org/10.1016/j.eneco.2014.02.009

O'Mahoney, A., & Denny, E. (2013). Electricity prices and generator behaviour in gross pool electricity markets. *Energy Policy*, *63*, 628-637. https://doi.org/10.1016/j.enpol.2013.08.098

Pamuk, N. (2016). Empirical analysis of causal relationship between electricity production and consumption demand in Turkey using Cobb-Douglas model. *Journal of Polytechnic*, *19*(4), 415-420. https://izlik.org/JA84ZR83BW

Quinlan, J. R. (1992). Learning with continuous classes. *In 5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348). https://doi.org/10.1142/1897

Rahman, M. N., Esmailpour, A., & Zhao, J. (2016). Machine learning with big data an efficient electricity generation forecasting system. *Big Data Research*, *5*, 9-15. https://doi.org/10.1016/j.bdr.2016.02.002

Reich, N. G., Lessler, J., Sakrejda, K., Lauer, S. A., Iamsirithaworn, S., & Cummings, D. A. (2016). Case study in evaluating time series prediction models using the relative mean absolute error. *The American Statistician*, *70*(3), 285-292. https://doi.org/10.1080/00031305.2016.1148631

Saha, S., Kundu, B., Paul, G. C., & Pradhan, B. (2023). Proposing an ensemble machine learning based drought vulnerability index using M5P, dagging, random sub-space and rotation forest models. *Stochastic Environmental Research and Risk Assessment*, *37*(7), 2513-2540. https://doi.org/10.1007/s00477-023-02403-6

Santarisi, N. S., & Faouri, S. S. (2021). Prediction of combined cycle power plant electrical output power using machine learning regression algorithms. *Eastern-European Journal of Enterprise Technologies*, *6*(8), 114. https://doi.org/10.15587/1729-4061.2021.245663

Scarlat, N., Prussi, M., & Padella, M. (2022). Quantification of the carbon intensity of electricity produced and used in Europe. *Applied Energy*, *305*, 117901. https://doi.org/10.1016/j.apenergy.2021.117901

Sharif Ali, S. S., Razman, M. R., & Awang, A. (2020). The nexus of population, GDP growth, electricity generation, electricity consumption and carbon emissions output in Malaysia. *International Journal of Energy Economics and Policy*, *10*(3), 84-89. https://doi.org/10.32479/ijeep.8987

Solyali, D. (2020). A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. *Sustainability*, *12*(9), 3612. https://doi.org/10.3390/su12093612

Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, *60*, 126-140. https://doi.org/10.1016/j.ijepes.2014.02.027

Turkish Statistical Institute (TurkStat). (2025). Web page, https://data.tuik.gov.tr/. Access date: 05.12.2025.

Wang, Y. & Witten, I. H. (1996). *Induction of model trees for predicting continuous classes*. (Working paper 96/23). Hamilton, New Zealand: University of Waikato, Department of Computer Science.

Xiaosan, Z., Qingquan, J., Iqbal, K. S., Manzoor, A., & Ur, R. Z. (2021). Achieving sustainability and energy efficiency goals: assessing the impact of hydroelectric and renewable electricity generation on carbon dioxide emission in China. *Energy Policy*, *155*, 112332. https://doi.org/10.1016/j.enpol.2021.112332

Zhang, J., Zhang, M., Yang, J., & Zheng, X. (2024). Prediction of electricity load generated by combined cycle power plants using integration of machine learning methods and HGS algorithm. *Computers and Electrical Engineering*, *120*, 109644. https://doi.org/10.1016/j.compeleceng.2024.109644

Zhang, Q., Weili, T., Yumei, W., & Yingxu, C. (2007). External costs from electricity generation of China up to 2030 in energy and abatement scenarios. *Energy Policy*, *35*(8), 4295-4304. https://doi.org/10.1016/j.enpol.2006.12.026

Araştırma ve Yayın Etiği Beyanları
Research and Publication Ethics Statements

Benzerlik Oranı *Similarity Rate*	Çalışmanın intihal.net tarafından gerçekleştirilen taramada benzerlik oranı %7 olarak tespit edilmiştir. *The similarity rate of the article performed by intihal.net was determined as 7%.*
Hakem Değerlendirmesi *Peer-Review Evaluation*	Bu çalışma Editör Kurulu tarafından belirlenen ve çalışma konusunda uzmanlığı bulunan 4 hakem tarafından çift yönlü kör hakemlik prensibiyle değerlendirilerek yayıma uygun görülmüştür. *This study has been evaluated by 4 referees determined by the Editorial Board and having expertise in the field of study with the principle of double-blind peer-reviewing and deemed suitable for publication.*
Tekrar Kullanım *Reuse*	Herhangi bir bildiri veya tezden üretilmemiştir. *This study/work has not been derived from any previous paper or thesis.*
Yapay Zekâ Kullanımı *Use of Artificial Intelligence*	Çalışmanın dil kontrolünde destek alınmıştır. *The study was proofread for language.*
Katkı Oranı *Contributions*	Yazarın çalışmadaki katkı oranı %100'dür. *The author's contribution rate to the study is 100%.*
Çıkar Çatışması *Conflict of Interest*	Yoktur. *None.*
Destek/Teşekkür *Support/Acknowledgement*	Yoktur. *None.*
Etik Kurul Onayı *Ethics Committee Approval*	Etik Kurul onayına gerek olmayan çalışmadır. *This study does not require Ethics Committee approval.*
Ölçek Kullanım İzni *Scale Use Permission*	Ölçek iznine gerek yoktur. *Scale permission is not required.*