Sosyal Bilimler
EKEVAKADEMİ
dergisi

Makalenin geliş tarihi: 10.02.2025
1.Hakem Rapor Tarihi: 26.06.2025
2.Hakem Rapor Tarihi: 28.06.2025
Yayına Kabul Tarihi: 01.07.2025

eISSN:2148-0710 - pISSN:1301-6229

MACHINE LEARNING IN ENERGY EFFICIENCY: COMPARISON OF ENERGY ESTIMATION MODELS

M. Akif YENİKAYA[*]-Onur OKTAYSOY[**]

Abstract

In this study, the functionality of machine learning models was tried to be determined in order to estimate the energy consumption needed in industrial production. In this context, Linear Regression, Decision Tree, Random Forest, K-Nearest Neighbors and Support Vector Machine algorithms were compared and evaluated in terms of usability and performance values. In the study, five different machine learning models were compared to estimate energy consumption. In order to make energy consumption estimates, historical production data, energy consumption data and other relevant parameters were used as input data. Data was obtained from UCI data repository, an open-source platform. The machine learning process structured as 80/20 training/testing was adapted to the form where the models can perform energy efficiency analysis with comprehensive data parameters. Error metrics such as coefficient of determination (R^2), root mean square error (RMSE), mean square error (MSE) and mean absolute error (MAE) were used to evaluate the performance of the models. According to the findings, the Random Forest model used in the study provided a higher accuracy rate compared to other models, and the R^2 value was obtained as 0.9989. This result reveals that machine learning models can be used as effective tools in estimating energy consumption and that these tools can turn into a strategic advantage for businesses, considering the importance of energy in production. The research provides significant contributions to literature by revealing that machine learning technology can be an important tool in energy consumption and, moreover, by comparing the performance of different models in energy estimation.

Keywords: *Energy consumption, Resource efficiency, Machine learning, Energy forecasting, Technology.*

Jel Codes: *C88, D24, M11.*

Enerji Verimliliğinde Makine Öğrenmesi: Enerji Tahmin Modellerinin Karşılaştırılması

Öz

Bu çalışmada endüstriyel üretimde ihtiyaç duyulan enerji tüketimini tahmin etmek amacıyla makine öğrenimi modellerinin işlevselliği belirlenmeye çalışılmış, bu bağlamda Linear Regression, Decision Tree, Random Forest, K-Nearest Neighbors ve Support Vector Machine algoritmaları kullanılabilirlik ve performans değerleri bakımından karşılaştırılarak değerlendirilmiştir. Enerji tüketim tahminlerinin yapılabilmesi için geçmiş üretim verileri, enerji tüketim verileri ve diğer ilgili parametreler giriş verisi olarak kullanılmıştır. Veriler bir açık kaynak platformu olan UCI veri deposundan elde edilmiştir. 80/20 eğitim/test şeklinde yapılandırılan makine öğretim süreci kapsamlı veri parametreleriyle modellerin enerji verimliliği analizi yapabilecekleri forma uyarlanmıştır. Modellerin performanslarını değerlendirmek için determinasyon katsayısı (R^2), kök ortalama kare hatası (RMSE), ortalama kare hatası (MSE) ve ortalama mutlak hata (MAE) gibi hata metrikleri kullanılmıştır. Elde edilen bulgulara göre, çalışma kapsamında kullanılan Random Forest modeli diğer modellere oranla daha yüksek doğruluk oranını sağlayarak R^2 değeri 0.9989 olarak elde edilmiştir. Bu sonuç, enerji tüketimine yönelik tahminlemede makine öğrenmesi modellerinin etkili araçlar olarak kullanılabileceğini ve bu araçların üretimde enerjinin önemi göz önüne alındığında işletmeler için stratejik avantaja dönüşebileceğini ortaya koymaktadır. Araştırma enerji tüketiminde makine öğrenmesi teknolojisinin önemli bir araç olabileceğini ortaya koyması ve dahası farklı modellerin enerji tahminlemede sergiledikleri performansı karşılaştırması bakımından alanyazına önemli kazanım yaratmaktadır.

Anahtar Kelimeler: *Enerji tüketimi, Kaynak verimliliği, Makine öğrenmesi, Enerji tahmini, Teknoloji.*

Jel Kodları: *C88, D24, M11.*

[*] Dr. Öğr. Üyesi, Kafkas Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, akif.yenikaya@kafkas.edu.tr, https://orcid.org/0000-0002-3624-722X

[**] Dr. Öğr. Üyesi, Kafkas Üniversitesi, İİBF, Yönetim Bilişim Sistemleri Bölümü, onuroktaysoy@kafkas.edu.tr, https://orcid.org/0000-0002-8623-614X

1. Introduction

In today's global economy, energy is one of the main resources that directly affect the sustainability and competitiveness of businesses. Energy is an indispensable input for businesses to sustain their operations, from production processes to office operations. According to data from the International Energy Agency, energy consumption in the industrial sector accounts for 40% of total global energy demand (International Energy Agency [IEA], 2021). This shows that ensuring energy efficiency and optimizing energy costs has become a strategic imperative for businesses. The secure supply of energy resources requires businesses not only to manage their costs, but also to develop more environmentally friendly approaches in line with operational reliability and sustainability goals. This critical dependence of businesses on energy resources encourages the activation of energy management and increased investments in renewable energy. As a matter of fact, energy efficiency is an important factor that determines the level of development of countries, and indicators such as the energy intensity consumed are considered at the point of development. High energy intensity means that energy efficiency in production processes is low, and more energy is consumed per unit of product or service. This indicates more energy use in production activities and is considered an indicator of inefficient energy use (Wang et al., 2019).

Energy costs have a significant impact on profitability, especially in energy-intensive sectors. Reducing energy costs for manufacturing and industrial enterprises is directly reflected in product prices, providing a competitive advantage. Increases in energy prices, especially in developing economies, can increase product costs and make it difficult to compete both in the domestic and international markets. In this context, by saving energy and investing in renewable energy sources, businesses can both control their costs and fulfill their environmental responsibilities by reducing their carbon footprint. On the other hand, controlling energy costs is also very important in terms of ensuring the long-term financial stability of enterprises. When energy costs, which are an important component of fixed costs, are optimized through strategic energy management practices, businesses can become more resilient to changing market conditions (Makridou et al., 2016).

In the case of Turkey, the importance of energy management is even more evident. As a rapidly developing and industrializing economy, Turkey is faced with ever-increasing energy demand. This increasing demand increases Turkey's dependence on foreign sources of energy, bringing with it not only an economic threat but also a political one. Approximately 75% of Turkey's total energy demand is met through imports, creating a need for foreign currency and putting pressure on the current account deficit (Ministry of Energy and Natural Resources, 2022). Under these circumstances, improving energy efficiency and making energy demand more predictable is of strategic importance for economic sustainability (Topcuoglu et al., 2024).

Energy demand forecasting plays a critical role in energy management in optimizing management processes to increase energy efficiency. Accurate and reliable forecasting of energy demand is critical for maintaining energy supply-demand balance, cost optimization and increasing energy efficiency. While traditional forecasting methods evaluate historical data based on certain patterns, they are limited in their ability to quickly analyze variable factors in today's dynamic energy markets. At this point, machine learning (ML) models stand out with their potential to provide high accuracy in energy demand forecasting (Vianna et al., 2019; Yenikaya et al., 2024).

Machine learning models have the potential to revolutionize energy forecasting with their ability to learn and infer complex relationships from large data sets. Various machine learning algorithms such as support vector machines (SVM), random forest (RF), artificial neural networks (ANN) have the potential to be widely used in short and long-term energy demand forecasting (Bahij et al., 2019). These algorithms can more accurately predict future demand trends by analyzing variables that affect energy demand, such as temperature, economic growth, time of day, and energy prices (Kong et al., 2017). The most important contribution of machine learning models in energy forecasting is their fast adaptation capabilities and impressive accuracy. Techniques such as artificial neural networks and deep learning models can capture seasonal variations in data and unpredictable demand spikes more precisely, enabling efficient use of energy resources.

The aim of this study is to address the challenges faced in today's energy sector by addressing the use of machine learning models in energy demand forecasting. In addition, this study aims to reveal how data analytics and artificial intelligence techniques can be used more efficiently to manage energy consumption more effectively. By evaluating the impact of machine learning algorithms on energy forecasting, the study aims to contribute to the development of sustainable solutions in energy management. In this context, Linear Regression, Decision Tree, Random Forest, K-Nearest Neighbors and Support Vector Machine, which are machine learning models, were evaluated in this study to predict energy consumption in production. As a result of the analysis, the energy consumption prediction performance of each model was compared, and the best model was determined. The findings show that these mixed methods can make significant contributions to the development of strategies for energy management and optimization.

1.1. Conceptual Framework

1.1.1. Energy Consumption and Forecast Need

Since the industrial revolution, there has been a rapid industrialization process across the world. Industrialization is considered as one of the driving forces of economic growth and plays a critical role in increasing the welfare of societies. However, this development brings with it an increasing demand for energy. Today, energy is recognized as a fundamental resource for the sustainability of economic activities and is used in a wide range of areas from industrial production to agriculture and transportation. This increasing dependence on energy has led countries and sectors to develop strategic policies to ensure energy supply security (IEA, 2023).

While modern societies are increasingly dependent on energy, the limited availability of energy resources makes this dependency even more complex. In particular, the exhaustibility of traditional energy sources such as fossil fuels have made sustainable energy management practices mandatory. In addition, carbon emissions generated during the use of these resources threaten environmental sustainability and contribute significantly to the global climate change problem (United Nations [UN], 2023). Accordingly, the need for energy demand forecasts is increasing day by day in order to ensure more efficient use of energy resources and to manage energy sustainably.

Energy forecasting has the potential to be a strategic tool that enables businesses to manage their resources effectively and sustainably (Zengin et al., 2021). With accurate forecasts, businesses can reduce energy costs by optimizing the energy supply-demand balance, increase operational efficiency, and minimize the financial risks of unexpected energy fluctuations (Kushwaha & Waoo, 2023). Being able to predict energy consumption gives businesses a competitive advantage, while at the same time enabling them to comply with their environmental responsibilities. Especially in large-scale production facilities, accurate forecasting of instantaneous changes in energy consumption is a critical requirement for both success in cost management and sustainable growth. Accurately forecasting the amount of energy consumption that will be needed in the future can lead to important gains such as balancing energy supply and demand, maintaining energy price stability, predicting the costs that may arise and preventing resource waste (Chahbi et al., 2022). Demand forecasting becomes even more important, especially for energy types that cannot be stored in a short time, such as electricity. Accurate demand forecasting not only contributes to more efficient use of energy resources, but also directs energy policies, reduces production costs and supports environmental sustainability (Gellings, 2020).

Machine learning technologies have the potential to be used as an important tool in building energy consumption models with their ability to analyze large data sets quickly and accurately. Beyond traditional statistical models, machine learning methods can demonstrate high performance in predicting energy consumption with the ability to model dynamic data such as time series and seasonal patterns. Accordingly, the number of studies on algorithms such as artificial neural networks, regression analysis, and long short-term memory (LSTM) in the context of energy demand forecasting is increasing day by day (Feng & Zhang, 2020).

In order to obtain more accurate and reliable results in energy demand forecasting, model development studies as well as data preparation processes should be carried out meticulously. Considering seasonal, geographical and social variables in energy consumption increases the accuracy of forecasts. In this context, the intensification of studies on processes such as data preparation, feature selection and model optimization in literature is among the important steps that will contribute to the accuracy of energy demand forecasting.

1.1.2. Use of Machine Learning in Energy Estimation

Machine learning is a field of data science that aims to predict future events by extracting patterns and relationships from large data sets (Goodfellow et al., 2016). In the field of energy forecasting, machine learning techniques can be used to discover energy consumption patterns, predict energy demand and develop energy management strategies from this data. While traditional statistical methods are limited in analyzing complex energy data, machine learning techniques can provide more accurate and flexible forecasts by processing large data sets (Zhang et al., 2021), making it an effective tool in energy demand management.

There are different algorithms in machine learning, and different machine learning algorithms can be used in energy forecasting according to different data types and forecast times. Below, the main algorithms commonly used in energy forecasting are described:

- **Regression Models:** This model is used to predict continuous and non-linear relationships in energy consumption. Algorithms such as linear regression, multiple linear regression, Lasso and Ridge allow modeling energy demand according to various environmental and social factors (Hyndman, 2018).
- **Time Series Models:** Time series models such as ARIMA and SARIMA are used to understand the changing patterns of energy consumption over time. These methods can identify trends and seasonal fluctuations in energy consumption by analyzing historical data. Modern time series models such as Prophet are highly effective in predicting complex seasonal patterns (Taylor & Letham, 2018).
- **Deep Learning Models:** Especially for very large data sets, deep learning provides neural network-based algorithms that can predict with high accuracy. Models such as long short-term memory (LSTM) and recurrent neural networks (RNN) can make more accurate predictions by learning the long-term dependencies of energy consumption. These models are particularly successful in learning complex patterns in energy consumption (Hochreiter, 1997).
- **Decision Trees and Ensemble Methods:** Decision trees and their ensembles are frequently used for energy forecasting. For example, Random Forest and XGBoost can combine the results of multiple decision trees to make more robust forecasts. These algorithms increase forecasting

accuracy by considering multiple factors (such as season, time of day, weather, etc.) that affect energy consumption (Breiman, 2001).

For machine learning models to work effectively, energy data needs to be properly prepared. Energy consumption data is usually large and can have different time scales such as hourly, daily, weekly or monthly consumption. These data can be influenced by social, economic, environmental and technical factors. The data preparation process includes the following steps:

- **Data Collection and Aggregation:** In energy consumption forecasting, data sets from different sources (weather, energy prices, population density, etc.) are brought together.
- **Feature Selection and Engineering:** The main variables affecting energy consumption are identified and new features are created to improve forecast accuracy (e.g., holidays, energy price fluctuations, etc.).
- **Data Cleaning:** Identifying and correcting missing or erroneous data, as well as performing outlier analysis.
- **Data Scaling:** Data scaling, especially for deep learning algorithms, enables the model to learn more stably.

The performance of energy forecasting models is evaluated using various successful metrics. These metrics are used to measure the accuracy, reliability and generalizability of the model. The main evaluation metrics are:

- **Mean Absolute Error (MAE):** It is the average of the absolute differences between the predicted values and the actual values.
- **Mean Square Error (MSE):** It is the mean of the squared errors and gives more weight to large errors.
- **R-Square (R^2)** and Weighted Accuracy Metrics: Used to analyze the model's predictive accuracy more comprehensively.

Machine learning-based energy forecasting models can help make energy management more sustainable and efficient. Accurate forecasts enable the optimization of energy production and distribution, reducing energy costs, avoiding unnecessary energy waste, and reducing carbon emissions, ultimately contributing significantly towards sustainable development goals in the energy sector (Gellings & Parmenter, 2016). In addition, it is possible to achieve gains beyond forecasting energy demand through machine learning. With this technology, it is also possible to gain benefits in various areas of energy management, such as managing the supply-demand balance in energy, optimizing energy distribution through smart grids, and supporting renewable energy integration. For example, it is also possible to make grid operations more flexible by forecasting energy from variable renewable sources such as wind and solar energy (Solyali, 2020).

1.1.3. Studies Examining Machine Learning in Energy Consumption Forecasting

Energy consumption forecasting is critical for energy management and optimization. Machine learning can make high accuracy predictions by using various models and techniques in this field. In the literature, various machine learning methods used for energy consumption forecasting stand out with their high accuracy rates. Some of the examples of these studies are as follows; Berriel et al. (2017) proposed a system that uses deep learning techniques to predict monthly energy consumption. Deep Fully Connected, Convolutional and Long Short-Term Memory Neural Networks, which are deep learning models, are examined and validated with real customer data.

Dan and Phuc (2018) developed machine learning models to predict energy loads based on the characteristics of building design. The study emphasizes the potential of buildings to reduce energy consumption and environmental impacts. It is stated that the model developed within the scope of the study can be used to improve energy efficiency and promote sustainable building designs. The researchers emphasize that thanks to these approaches, both reducing energy costs and reducing carbon footprint can have significant impacts when evaluated on a global scale.

Khan et al. (2020) proposed a hybrid energy forecasting model using machine learning models such as extreme gradient boosting, categorical boosting and random forest. The model was tested using hourly energy consumption data of South Korea and achieved high accuracy with an R^2 value of 0.9212. Ahmad et al. (2018) focused on Binary Decision Tree, Compact Regression Gaussian Process, Stepwise Gaussian Processes Regression and Generalized Linear Regression Model, which are machine learning models for the purpose of energy consumption estimation of buildings, and made short, medium and long-term estimations. In the study, the machine learning models used in the study were compared in order to increase the estimation accuracy. Mawson and Hughes (2020) compared "feed forward" and "recurrent", which are deep neural network models, for the purpose of estimating energy consumption in production facilities. They used simulation techniques to increase model accuracy. As a result of the research, they emphasized that machine learning technology can provide functional results in energy consumption estimation.

In the study conducted by Milicevic and Marinovic (2024), machine learning models were used in the production estimation of renewable resources, especially solar energy, and the amount of production that could occur in the future was tried to be estimated based on daily historical data and meteorological expectations, and the obtained data revealed that machine learning was successful. These models can analyze the impact of climate factors on forecasting, and artificial neural networks can achieve over 97% accuracy in this area.

Studies in the literature clearly demonstrate how machine learning techniques are used in energy consumption forecasting and how these forecasts contribute to energy management strategies. Machine learning models have become an important tool for improving energy efficiency thanks to their ability to predict energy consumption with high accuracy. These techniques enable more effective implementation of energy management and resource optimization strategies by predicting potential changes in energy consumption. In this way, energy costs are reduced, resources are used more efficiently, and environmental impacts are minimized.

This study makes a significant contribution to literature by providing a comprehensive comparison of various machine learning models for energy consumption forecasting. Unlike other studies, instead of focusing on deep learning techniques and optimization methods, it compares the raw performance of basic machine learning models. It also provides a broader perspective on energy management by addressing the usefulness and interpretability of the models in practical applications

2. Material

The dataset used in the research was obtained from the UCI data repository, an open skiing platform. The dataset (Sathishkumar et al., 2023) provides comprehensive information for energy consumption analysis. In this dataset, the energy consumption data (Usage_kWh) of an industrial manufacturing process was measured every 15 minutes between January 1, 2018, and December 31, 2018, and recorded as time series data. This data set, which is recorded in very short periods, provides an opportunity to analyze energy consumption in the industrial sector in depth. The high resolution of energy consumption data makes time series analysis more meaningful and provides a great advantage in identifying patterns and anomalies in energy consumption and in formulating energy efficiency strategies. Moreover, additional data such as reactive power, carbon emissions and power factors help to examine in detail the various factors affecting energy consumption and enable more effective optimization of energy management strategies. The change in electricity usage over time for the dataset used in the study is graphically shown in Figure 1.

Figure 1

Electricity Usage Over Time

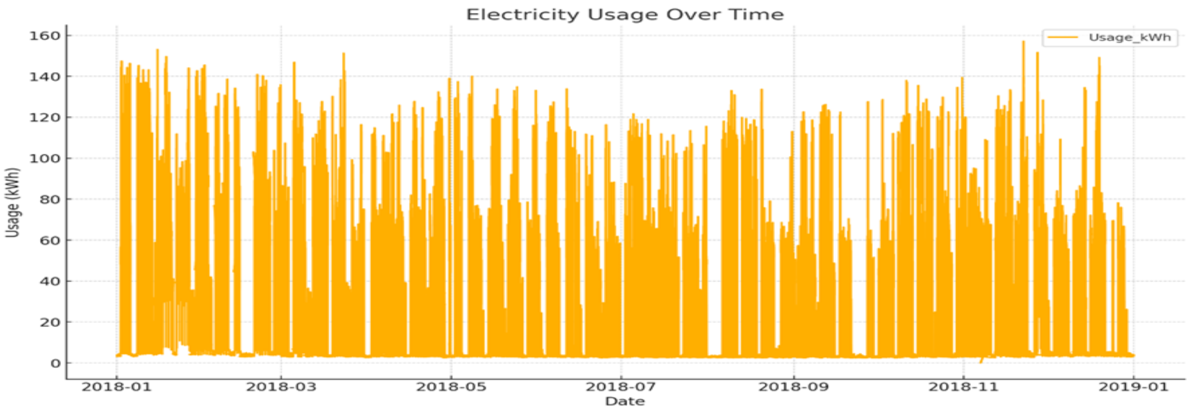


Table 1 shows the attributes and types of the dataset used in the study. The attributes in the dataset consist of both categorical and numeric (continuous and integer) variables. Categorical variables are date attribute which contains date and time information, WeekStatus which indicates weekday or weekend distinction, Day_of_week which indicates the day of the week and Load_Type which indicates energy load types. Usage_kWh, Lagging_Current_Reactive, Power_kVarh, Leading_Current_Reactive_Power_kVarh, CO2 (tCO2), Lagging_Current_Power_Factor and Leading_Current_Power_Factor are continuous variables. They are used to analyze energy consumption and quality. The NSM, which indicates the second of the day, is a fully numeric variable and is important in time series analysis. This diversity offers a wide range of analyses to analyze the energy consumption patterns of the dataset, environmental impacts and efficiency of electricity systems.

Table 1

Data Attributes and Types

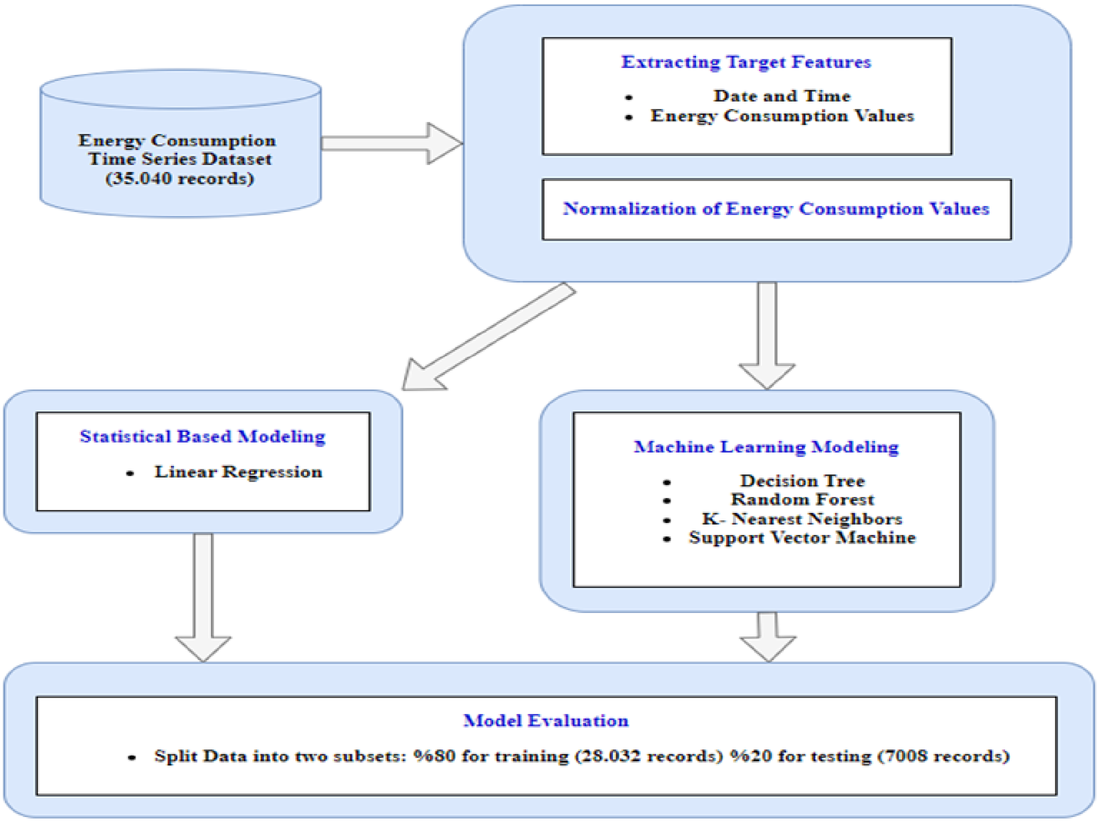
Attributes	Types
date	Categorical
Usage_kWh	Continuous
Lagging_Current_Reactive.Power_kVarh	Continuous
Leading_Current_Reactive_Power_kVarh	Continuous
CO2 (tCO2)	Continuous
Lagging_Current_Power_Factor	Continuous
Leading_Current_Power_Factor	Continuous
NSM	Integer
WeekStatus	Categorical
Day_of_week	Categorical
Load_Type	Categorical

3. Modeling Methodology

As shown in Figure 2, the modeling process of this research consists of three stages: data preprocessing, model creation, and model evaluation.

Figure 2

Modeling Process



In the first stage of data preprocessing, feature extraction and data normalization are the two steps to be performed. The feature extraction step is the process of selecting two variables from the original dataset. The selected attributes in the dataset used in this study are date-time and energy consumption in steel production. The data normalization step is performed to bring the selected features to a common scale in cases where they may be on different scales. Missing data was imputed, and outliers were identified and handled appropriately. It was checked whether the data set was balanced. This is important for the model to provide more reliable results. The cleaned and balanced data set was divided into two parts: training and testing. 80% of the data set was used for training the models, while 20% of the data set was used for testing to evaluate the performance of the models.

In the second phase of this research, the model development process was studied. In this stage, two different types of modeling algorithms were used: statistical methods and machine learning techniques. Linear regression used for statistical methods was the basis for the performance comparison of the prediction models used in this study.

Linear regression is the first type of regression method usually encountered when talking about prediction through a model. Linear regression refers to the simplest relationship, which is a straight-line relationship between the x variables and the dependent variable y. The main advantages of this approach are its simplicity and universality. It also provides an adequate but easy to understand explanation in most cases (Hastie et al., 2001).

In the study of machine learning modeling, decision tree, random forest, K nearest neighbor and support vector machine methods, which are widely used in the literature, were used. Decision tree is a class used for classification and prediction in machine learning. It involves dividing the data into subsets that are internally more homogeneous than the entire data set. A decision is made about a specific feature of the data at each node. Due to this decision, the data is divided into two or more groups. This process is repeated for all data points. Decision trees can perform extremely well in high-resolution data sets (Krahwinkler et al., 2011). Random forest, an ensemble learning method, is formed by combining multiple decision trees. This method is based on the principle of training each tree independently and averaging the results. Unlike decision trees, this method reduces the problem of over-learning and provides relatively higher accuracy compared to decision trees (Mathur & Badone, 2019). The K-nearest neighbor algorithm, which is used in regression and classification problems, looks at the nearest k neighbors, which is an important parameter that determines the performance of the algorithm, to determine the class of a new data point. The K-nearest neighbor algorithm, which can be computationally costly in large data sets, also appears as a simple and effective method (Sathe & Aggarwal, 2019). Support vector machine, a powerful algorithm that creates a hyperplane to classify data, aims to find the widest margin separating two classes. It does this by creating a line or plane that provides the largest distance between data points. Support vector machine can also be extended using kernel functions for nonlinear classifications (Jahangiri & Rakha, 2015).

Each of these methods has different advantages and disadvantages for different data sets and problems. For this reason, it is important to try more than one method and compare their performance when working on machine learning problems.

This research was conducted in accordance with the principles of scientific research and publication ethics. The data used in the study were obtained from the UCI Machine Learning Repository, an open-access source. Since the dataset does not contain any personal or private information, no ethics committee approval was required. Throughout the research process, accuracy, transparency, and impartiality of the data were strictly observed, and the findings were reported within the framework of academic integrity. Furthermore, no conflict of interest exists in this study, and all procedures were carried out in compliance with ethical standards.

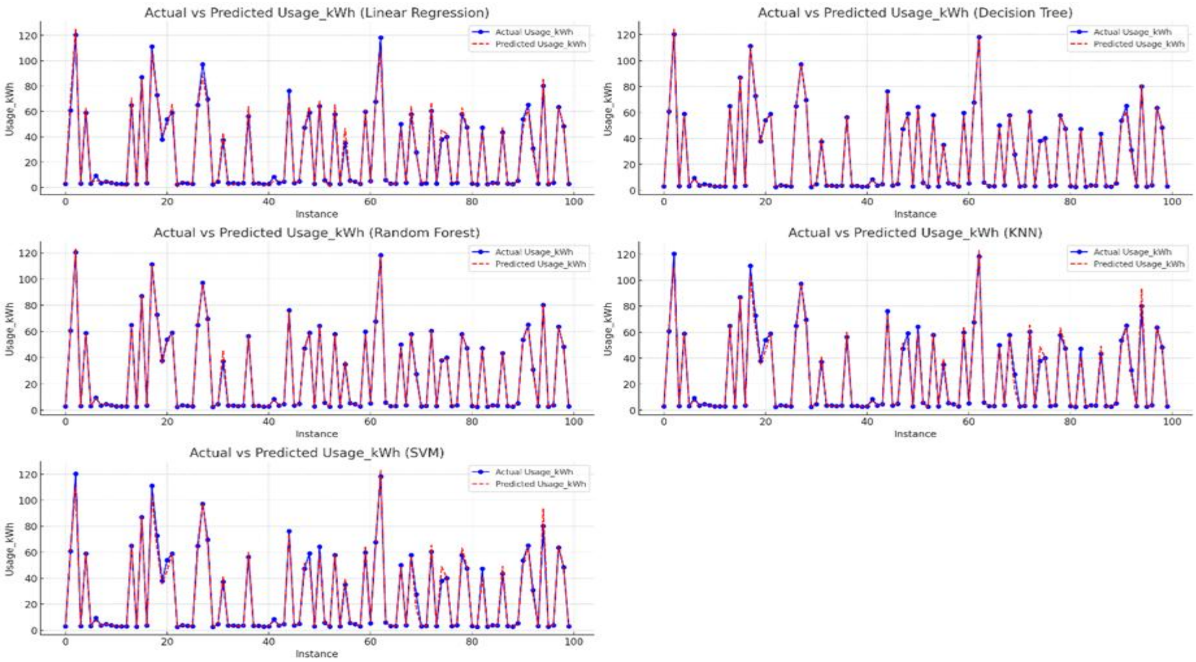
3.1. Performance Results of Models

The features in the dataset were used to estimate and classify energy consumption. Energy consumption was used as the target variable, lagging and leading current reactive power, CO_2 emissions, power factors, time and load type information were used as independent variables. Python Anaconda 3.12 was used as the interpreter in coding all algorithms. In the architectural implementation phase, various libraries were used to optimize the data analysis and visualization processes. In particular, Scikit played an important role in creating and evaluating machine learning models. Matplotlib and Seaborn were used in visualizing the data, while Numpy and Pandas libraries were effective in data manipulation and analysis. All these libraries were critical in increasing the accuracy and efficiency of the results.

In the statistical and machine learning analyses performed on the dataset used in this study, the specified coding interpreter and libraries were used. The graphics created to visualize the actual and estimated values obtained after the training processes are presented in Figure 3.

Figure 3

Actual and Predicted Value Graphs of the Models



In Figure 3 above, the blue stripe shows the actual energy consumption values, while the values shown with the red dashed stripe represent the estimated energy consumption values. When the mentioned values are examined, it is seen that there is a clear similarity between the actual consumption values and the estimated consumption values in the context of Linear regression, decision tree, random forest, k-nearest neighbor and support vector machine parameters. These obtained data reveal the potential of machine learning to be used as an effective tool for predicting the amount of energy needed in the steel industry and, moreover, in many different sectors with energy requirements, and thus for calculating activity costs accurately.

In order to compare the performances of different types of machine learning algorithms used in the study, 4 different evaluation criteria were used: R-squared (R^2), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). RMSE and MAE criteria were used to measure the difference between the real values and the values predicted by the models. The lower the values of these criteria, the better the performance. The R^2 criterion indicates the extent to which the data fits the regression model. The high value of R^2 indicates that the model used is that good. The MAPE criterion expresses the percentage of deviation of the predicted values from the real value. Low deviation values indicate high performance. The performance findings of the models used in the study are given in Table 2.

Table 2

Performance Findings of the Models Used

Model	R^2	RMSE	MAE	MAPE
Linear Regression	0.9843	4.220	2.596	19.339%
Decision Tree	0.9981	1.467	0.570	2.005%
Random Forest	0.9989	1.095	0.370	1.434%
K-Nearest Neighbors	0.9384	8.368	4.144	21.019%

Support Vector Machine	0.9170	9.713	4.583	21.782%

When the findings in the table are examined, it is seen that the model with the best performance among the models used in the study is Random Forest, which has a very high accuracy rate with values of R^2=0.9989, RMSE=1.095, MAE=0.370 and MAPE=1.434%.

4. Conclusion, Discussion, and Recommendations

This study provides a comprehensive analysis comparing various machine learning models for the steel industry to estimate energy consumption. The research findings reveal the variety of methods used in energy consumption estimation and that each method has its own advantages and disadvantages. The findings show that machine learning tools can play an important role not only in energy consumption estimation but also in long-term cost optimization strategies of enterprises. One of the main findings of the study is that the Random Forest model shows superior performance compared to other machine learning models with the obtained R^2=0.9989 value. This finding shows that the Random Forest model can be an effective tool in energy efficiency studies and has the potential to provide reliable estimates for energy management and optimization.

The performance of machine learning models used to estimate energy consumption depends on many factors such as the structure of the model, the characteristics of the dataset, and the nature of the estimated parameters. For example, in the study conducted by Sathishkumar et al. (2020), it was determined that the Random Forest model performed lower than other traditional algorithms in energy consumption estimation. In this study, the Random Forest model had the highest accuracy rate with R^2 = 0.9989. This finding is remarkable in terms of revealing the importance of correct model selection, dataset characteristics, and model parameters in energy consumption estimation.

The literature review in the field of energy efficiency and sustainability shows that the machine learning models used in this study are consistent with the findings in the literature in terms of increasing energy estimation accuracy. For example, studies by Ahmad and Chen (2020) and Bahij et al. (2019) support that algorithms such as Random Forest and support vector machines provide superior performance in industrial energy consumption estimation. Similarly, the study by Milicevic and Marinovic (2024) achieved an accuracy rate of over 97% in solar energy production estimations using artificial neural networks and revealed that these models are also reliable in the field of renewable energy. It can be stated that the models that demonstrate a similar success in this study have the competence to create a sustainable infrastructure for energy efficiency.

The contribution of this study in terms of energy management is especially important when evaluated in the context of effective management of energy costs, energy saving and reducing environmental impact. The high accuracy of energy consumption estimates provides businesses with the opportunity to plan their energy consumption in advance and optimize their energy costs, thus allowing them to gain competitive advantage. Estimating energy consumption in energy-intensive sectors not only provides economic gain but also contributes to more efficient use of resources. In this context, as emphasized in the studies of Gellings and Parmenter (2016), it should be emphasized that energy demand management with machine learning has a great potential in terms of making more precise estimates in the future and ensuring the balance of energy supply and demand.

In addition, this research provides important results for the expansion of machine learning applications aimed at increasing energy efficiency in the industrial sector. The findings obtained in the study show how data analytics and machine learning technologies can play an effective role not only in predicting energy consumption but also in achieving environmental sustainability goals of energy management strategies. In particular, it is possible for businesses to fulfill their environmental responsibilities by reducing carbon emissions while controlling energy costs. Machine learning models that support strategic applications for the integration of renewable energy sources contribute to increasing environmental sustainability in the energy sector (Devaraj et al., 2021). In this context, it is very important to revise public policies in terms of benefiting from these technologies with an understanding beyond economic concerns, in order to ensure faster progress in terms of increasing efficiency and sustainable resource use.

The original contribution of this study does not lie in proposing a new model, but rather in systematically comparing existing machine learning algorithms on the same dataset to reveal, both qualitatively and quantitatively, which model performs better under which conditions in the context of energy consumption forecasting. While most of the existing literature tends to focus on the performance of a single model, this study: Utilizes high-resolution time series data based on real production processes, Applies multiple algorithms (LR, DT, RF, KNN, SVM) on the same dataset and compares all models using consistent evaluation metrics (R^2, RMSE, MAE, MAPE).

Thus, our study provides a clear benchmarking of different models in terms of their suitability for various decision-making scenarios. This represents a valuable contribution to the literature by offering practical guidance for model selection in energy management strategies. Moreover, the particularly high accuracy achieved by the Random Forest algorithm (R^2=0.9989) empirically supports its applicability in energy-intensive industries for practical forecasting purposes.

Future studies should aim to conduct more comprehensive analyses and comparisons to further improve the performance of machine learning models used in energy consumption estimation and optimize energy management strategies. In particular, studies to increase the interpretability of deep learning techniques and other advanced models can enable these models to be used more effectively in energy management applications. In this context, research on both basic machine learning models and deep learning techniques is extremely important in terms of their contribution to energy efficiency studies.

5. References

Ahmad, T., Chen, H., Huang, R., Yabin, G., Wang, J., Shair, J., ... & Kazim, M. (2018). Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. *Energy*, *158*, 17-32. https://doi.org/10.1016/j.energy.2018.05.169

Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society*, *54*, 102010. https://doi.org/10.1016/j.scs.2019.102010

Bahij, M., Cherkaoui, M., & Labbadi, M. (2019). Energy consumption forecasting in industrial sector using machine learning approaches. *Innovation in Information Systems and Technologies to Support Learning Research (Conference paper)*, 155-164. https://doi.org/10.1007/978-3-030-36778-7_17

Berriel, R., Lopes, A., Rodrigues, A., Varejão, F., & Oliveira-Santos, T. (2017). Monthly energy consumption forecast: A deep learning approach. *2017 International Joint Conference on Neural Networks (IJCNN)*, 4283-4290. https://doi.org/10.1109/IJCNN.2017.7966398

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Chahbi, I., Rabah, N., & Tekaya, I. (2022). Towards an efficient and interpretable machine learning approach for energy prediction in industrial buildings: A case study in the steel industry. *2022*

IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), 1-8. https://doi.org/10.1109/AICCSA56895.2022.10017816

Dan, T., & Phuc, P. (2018). Application of machine learning in forecasting energy usage of building design. *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, 53-59. https://doi.org/10.1109/GTSD.2018.8595595

Devaraj, J., Madurai Elavarasan, R., Shafiullah, G. M., Jamal, T., & Khan, I. (2021). A holistic review on energy forecasting using big data and deep learning models. *International journal of energy research*, *45*(9), 13489-13530. https://doi.org/10.1002/er.6679

Enerji ve Tabii Kaynaklar Bakanlığı (2022). *Türkiye ulusal enerji planı 2022*. https://enerji.gov.tr//Media/Dizin/EIGM/tr/Raporlar/TUEP/T%C3%BCrkiye_Ulusal_Enerji_Plan%C4%B1.pdf (Access Date: 08.09.2024).

Feng, C., & Zhang, J. (2020). Assessment of aggregation strategies for machine-learning based short-term load forecasting. *Electric Power Systems Research*, *184*, 106304. https://doi.org/10.1016/j.epsr.2020.106304

Gellings, C. W. (2020). *The smart grid: Enabling energy efficiency and demand response*. River Publishers. https://doi.org/10.1201/9781003151524

Gellings, C. W., & Parmenter, K. E. (2016). Demand-side management. *In Energy management and conservation Handbook*, 399-420. CRC. https://doi.org/10.1201/9781315374178-20

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Regularization for deep learning. *Deep learning*, 216-261. https://mnassar.github.io/deeplearninghandbook/slides/07_regularization.pdf (Access Date: 23.08.2024).

Hastie, T., Friedman, J., & Tibshirani, R. (2001). Linear methods for regression. In: The elements of statistical learning. Springer Series in Statistics. Springer. https://doi.org/10.1007/978-0-387-21606-5_3

Hochreiter, S. (1997). Long short-term memory. *Neural Computation MIT-Press*. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J. (2018). *Forecasting: principles and practice*. OTexts. Monash University, Australia,

IEA, 2023 - International Energy Agency (2023). *World Energy Outlook 2023*. International Energy Agency. https://www.iea.org/reports/world-energy-outlook-2023 (Access Date: 10.09.2024).

IEA, 2021 - International Energy Agency, (2021). *Global Energy Review 2021*. https://www.iea.org/reports/global-energy-review-2021?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosgenerate&stream=top (Access Date: 04.09.2024).

Jahangiri, A., & Rakha, H. (2015). Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 2406-2417. https://doi.org/10.1109/TITS.2015.2405759

Khan, P., Byun, Y., Lee, S., & Park, N. (2020). Machine learning based hybrid system for imputation and efficient energy demand forecasting. *Energies*, *13*(11), 2681. https://doi.org/10.3390/en13112681

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE transactions on smart grid*, *10*(1), 841-851. https://doi.org/10.1109/TSG.2017.2753802

Krahwinkler, P., Roßmann, J., & Sondermann, B. (2011). Support vector machine based decision tree for very high resolution multispectral forest mapping. *2011 IEEE International Geoscience and Remote Sensing Symposium*, 43-46. https://doi.org/10.1109/IGARSS.2011.6048893

Kushwaha, N., & Waoo, A. (2023). Energy consumption prediction by using machine learning. *International Journal For Multidisciplinary Research*, *5*(6), 1-12. https://pdfs.semanticscholar.org/dd9e/5db5ca2529fe92a46c53cb7082ceba01f0d2.pdf

Makridou, G., Andriosopoulos, K., Doumpos, M., & Zopounidis, C. (2016). Measuring the efficiency of energy-intensive industries across European countries. *Energy Policy*, *88*, 573-583. https://doi.org/10.1016/j.enpol.2015.06.042

Mathur, S., & Badone, A. (2019). A methodological study and analysis of machine learning al gorithms. *International Journal of Advanced Technology and Engineering Exploration, 6*(51), 45-9. https://doi.org/10.19101/IJATEE.2019.650020

Mawson, V. J., & Hughes, B. R. (2020). Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector. *Energy and Buildings*, *217*, 109966. https://doi.org/10.1016/j.enbuild.2020.109966

Milićević, M. M., & Marinović, B. R. (2024). Machine learning methods in forecasting solar photovoltaic energy production. *Thermal Science*, *28*(1 Part B), 479-488. https://doi.org/10.2298/TSCI230402150M

Sathe, S., & Aggarwal, C. (2019). Nearest neighbor classifiers versus random forests and support vector machines. *2019 IEEE International Conference on Data Mining (ICDM)*, 1300-1305. https://doi.org/10.1109/ICDM.2019.00164

Sathishkumar, V. E., Lim, J., Lee, M., Cho, K., Park, J., Shin, C., & Cho, Y. (2020). Industry energy consumption prediction using data mining techniques. *International Journal of Energy, Information and Communications*, *11*(1), 7-14. http://dx.doi.org/10.21742/ijeic.2020.11.1.02

Sathishkumar, V. E., Changsun, S., & Yongyun, C. (2023). *Steel industry energy consumption*. UCI Machine Learning Repository. https://archive.ics.uci.edu/datasets

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

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45. https://doi.org/10.1080/00031305.2017.1380080

Topcuoglu, E., Oktaysoy, O., Kaygin, E., Kosa, G., Uygungil-Erdogan, S., Kobanoglu, M. S., & Turan-Torun, B. (2024). The potential of the society 5.0 strategy to be a solution to the political and structural problems of countries: The case of Türkiye. *Sustainability, 16*(22), 9825. https://doi.org/10.3390/su16229825

UN, 2023 – *United Nations Environment Programme – 2023*. https://www.unep.org/resources/global-environment-outlook-6 (Access Date: 10.09.2024).

Vianna, V., Celeste, W., & Freitas, R. (2019). Energy efficiency in the context of Industry 4.0. *International Journal of Advanced Engineering Research and Science*, *6*(12), 1-16. https://doi.org/10.22161/ijaers.612.1

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*. 198, 111799. https://doi.org/10.1016/j.enconman.2019.111799

Yenikaya, M. A., Kerse, G., & Oktaysoy, O. (2024). Artificial intelligence in the healthcare sector: comparison of deep learning networks using chest X-ray images. *Frontiers in Public Health*, *12*, 1386110. https://doi.org/10.3389/fpubh.2024.1386110

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., & Livingood, W. (2021). A review of machine learning in building load prediction. *Applied Energy*, 285, 116452. https://doi.org/10.1016/j.apenergy.2021.116452

Zengin, Y., Naktiyok, S., Kaygın, E., Kavak, O., & Topçuoğlu, E. (2021). An investigation upon industry 4.0 and society 5.0 within the context of sustainable development goals. *Sustainability*, *13*(5), 2682. https://doi.org/10.3390/su13052682