Innovations in Environmental Economics

www.iee.reapress.com

Innov. Environ. Econ. Vol. 1, No. 4 (2025) 271–280.

Paper Type: Original Article

Enhancing Sustainable Energy Practices through Innovation in Steel Industries: A Machine Learning Approach to Consumption Forecasting

Teuku Rizky Noviandy[1,*], Rahmat Sufri[1], Ryan Setiawan[1], Anisah Anisah[1]

[1] Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; rizky_si@abulyatama.ac.id; rahmatsufri_si@abulyatama.ac.id; ryan_si@abulyatama.ac.id; anisah_si@abulyatama.ac.id.

Citation:

Received: 10 June 2025 *Revised: 20 August 2025* *Accepted: 14 October 2025*	Rizky Noviandy, T., Sufri, P., Setiawan, P., & Anisah, A. (2025). Enhancing sustainable energy practices through innovation in steel industries: A machine learning approach to consumption forecasting. *Innovations in environmental economics, 1*(4), 271-280.

Abstract

The steel industry is one of the most energy-intensive manufacturing sectors, and small-scale producers often struggle to manage energy consumption efficiently while meeting sustainability goals. This study investigates the use of machine learning, specifically the Extreme Gradient Boosting (XGBoost) algorithm, to forecast energy usage in a small-scale steel manufacturing plant. A dataset comprising 35,040 instances and nine operational, temporal, and environmental features was pre-processed and used to develop a regression-based forecasting model. The XGBoost model demonstrated excellent predictive performance, achieving R^2 values of 0.999 on the training set and 0.997 on the test set, with low RMSE and MAE. Feature importance analysis revealed that reactive power factors and CO_2 emissions were the most influential variables affecting energy consumption patterns. The findings confirm that machine learning-driven forecasting can significantly enhance energy management by providing accurate predictions that support operational optimization, cost reduction, and environmental sustainability. This study underscores the potential of integrating intelligent predictive systems into small-scale steel manufacturing to promote more efficient, sustainable energy practices and to provide a foundation for future research on adaptive, real-time forecasting models.

Keywords: Energy consumption forecasting, Sustainable steel industry, Machine learning, Industrial energy management.

1 | Introduction

The steel industry plays a critical role in global manufacturing, supporting sectors such as infrastructure, transportation, construction, and industrial development [1–4]. Despite its importance, steel production

Corresponding Author: rizky_si@abulyatama.ac.id

https://doi.org/10.48313/iee.vi.57

remains one of the most energy-intensive industrial activities, relying heavily on electricity and fossil fuels for core processes such as melting, rolling, and casting [5], [6]. This high energy demand not only increases manufacturers' operational costs but also significantly contributes to greenhouse gas emissions, intensifying global climate challenges [7]. These issues are particularly burdensome for small- and medium-scale steel producers, who must balance production efficiency with compliance to increasingly stringent environmental regulations and rising societal expectations for sustainable practices [8].

The environmental implications of excessive energy consumption in steel manufacturing are profound. High levels of carbon dioxide and pollutant emissions contribute to climate change [9], [10], air degradation, and ecological instability [11], [12]. Additionally, the extraction and use of fossil fuels can damage ecosystems, reduce biodiversity, and contaminate water sources [13]. As global awareness of environmental sustainability grows, industries face mounting pressure to reduce their carbon footprints and transition to cleaner, more responsible operations [13–15]. Addressing these challenges is essential not only for environmental preservation but also for maintaining competitiveness and meeting stakeholder demands in a rapidly evolving industrial landscape.

Improving energy efficiency is a central strategy for mitigating environmental impacts and enhancing industrial sustainability [16], [17]. However, many traditional energy management systems used in steel plants lack the analytical intelligence needed to forecast energy demand accurately or optimize real-time usage [18], [19]. This limitation often leads to inefficiencies, resource waste, and unnecessary emissions. To overcome these issues, technological innovation, particularly in data analytics and machine learning, offers transformative potential [20–22]. Innovative digital tools can uncover complex, previously hidden patterns in industrial energy data, enabling more efficient and environmentally responsible decision-making.

Machine learning has emerged as a powerful innovation for forecasting energy consumption, identifying inefficiencies, and supporting proactive energy management (Mhlanga [23], Benti [24], Forootan [25]). By analyzing large, multivariate datasets, machine learning algorithms can detect subtle relationships that traditional statistical methods may overlook [26], [27]. Among these techniques, Extreme Gradient Boosting (XGBoost) stands out for its scalability, accuracy, and ability to handle complex feature interactions [28]. As an innovative boosting-based algorithm, XGBoost can generate highly precise predictions, making it exceptionally valuable for operational optimization in energy-intensive industries.

Despite its documented success in other domains, the application of XGBoost to energy consumption forecasting within small-scale steel industries remains limited. This gap presents an opportunity to integrate advanced digital technologies into traditional manufacturing environments, thereby promoting greater sustainability and operational efficiency. Therefore, this study aims to explore how machine learning innovation, specifically XGBoost, can be employed to forecast energy consumption in a small-scale steel manufacturing context. Using a comprehensive dataset of operational parameters, the research develops a predictive model that provides actionable insights to optimize energy use, reduce environmental impact, and lower operational costs. The findings demonstrate that integrating innovative machine learning solutions into industrial energy management systems can support compliance with environmental regulations while promoting a more sustainable and efficient steel production ecosystem.

2 | Methods

2.1 | Dataset

This study utilized a publicly available dataset from the UCI Machine Learning Repository, originally collected from a small-scale steel manufacturing facility in South Korea [29]. It comprises 35,040 instances with nine features and one target variable (Usage_kWh), encompassing both continuous and categorical data relevant to energy consumption patterns in steel manufacturing. The dataset is suitable for forecasting energy usage. The energy consumption data is maintained in a cloud-based system and includes perspectives on daily, monthly, and annual usage. This comprehensive dataset enables the development of machine learning models to forecast energy consumption, facilitating improved energy management and contributing to environmental sustainability in the steel manufacturing sector. The variables used in this study are listed in *Table 1*, along with their names, types, and descriptions.

Table 1. Variables used in the study.

Variable Name	Type	Description
date	Date	Date of the record
Lagging_Current_Reactive.Power_kVarh	Continuous	Lagging current reactive power (kVarh)
Leading_Current_Reactive_Power_kVarh	Continuous	Leading current reactive power (kVarh)
CO2(tCO2)	Continuous	Carbon dioxide emissions (ppm)
Lagging_Current_Power_Factor	Continuous	Lagging current power factor (%)
Leading_Current_Power_Factor	Continuous	Leading current power factor (%)
NSM	Integer	Number of seconds from midnight
WeekStatus	Categorical	Weekend (0) or weekday (1) indicator
Day_of_week	Categorical	Day of the week (Sunday to Saturday)
Usage_kWh	Continuous	Industry energy consumption (kWh)

2.2 | Data Preprocessing

Effective data preprocessing was essential to prepare the dataset for modelling [30]. Initially, the dataset was loaded, and the date column was converted to a datetime format to enable time-based analysis. Ensuring the correct date format enabled the extraction of temporal features and the chronological sorting of data, which are crucial for time series forecasting. The data were then sorted in ascending order by the date column to maintain temporal sequence. Any missing values in the dataset were set to zero.

Feature engineering involved extracting additional time-related features from the date column, including hour, month, year, and day. These features help the model learn daily, monthly, and yearly trends or seasonality in energy consumption. Categorical variables were label-encoded to convert them into numerical form suitable for modeling. Label encoding assigns a unique numerical value to each category, making the data compatible with machine learning algorithms.

Features not required for modeling, such as the original date column and the day feature, were dropped to prevent redundancy and potential information leakage. The target variable was identified as Usage_kWh, representing the industry's energy consumption in kilowatt-hours. A time-based split was performed to divide the dataset into training and test sets, with 80% of the data used for training and 20% for testing. This approach ensures that the model is trained on past data and tested on future data, which is appropriate for time series forecasting and helps prevent look-ahead bias [31].

2.3 | XGBoost Model

An Extreme Gradient Boosting (XGBoost) model was developed to predict energy consumption based on the preprocessed dataset [32]. XGBoost was selected for its strong performance on regression tasks involving nonlinear relationships, multicollinearity among features, and large-scale industrial datasets, making it well-suited for modeling complex energy consumption patterns in steel manufacturing systems.

The model objective function was set to reg:squarederror, which is appropriate for continuous-valued regression problems and provides stable gradient updates during training. The number of estimators was set to 100, balancing model expressiveness and computational efficiency. This value is commonly used in energy forecasting studies and is sufficient to capture complex feature interactions without excessive training time.

A learning rate of 0.1 was selected to control the contribution of each tree to the final ensemble. This moderate learning rate helps stabilize the boosting process by allowing gradual error correction while reducing the risk of overfitting. The maximum tree depth was limited to 5, constraining model complexity and preventing overly deep trees that may memorize noise rather than generalizable patterns, which is particularly important in time-dependent industrial data. To ensure reproducibility and consistency of results, the random state was fixed at 42. The selected hyperparameters follow widely accepted best practices in gradient boosting applications and have been shown to perform reliably across similar energy forecasting tasks [33].

The model was trained on the training dataset, where XGBoost builds an ensemble of decision trees sequentially [34], [35]. Each subsequent tree aims to correct the errors of the previous ones, allowing the model to capture complex patterns in the data. After training, the model was used to predict energy consumption on both the training and testing datasets. Predictions on the training set helped evaluate how well the model had learned from the data. In contrast, predictions on the testing set assessed the model's ability to generalize to new, unseen data.

2.4 | Model Evaluation

Model evaluation was conducted to assess the predictive accuracy and generalization capability of the XGBoost model. After generating predictions for both the training and testing sets, the predicted values were aligned with their corresponding timestamps to enable temporal analysis. To reduce short-term noise and emphasize broader consumption patterns, the actual and predicted energy usage values were aggregated to a daily level by grouping them by date and calculating the mean daily consumption.

Three evaluation metrics were used to assess model performance: the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). Together, these metrics provide a comprehensive understanding of the model's effectiveness by quantifying the variance explained, the magnitude of prediction errors, and the average absolute deviation between actual and predicted values (Roy and Kar, 2015).

The R^2 score evaluates the proportion of variance in the target variable that is captured by the model, serving as an indicator of overall goodness-of-fit:

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

The RMSE measures the square root of the average squared difference between actual and predicted values, indicating how widely predictions deviate from true observations:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}. \tag{2}$$

The MAE quantifies the average magnitude of prediction errors in absolute terms, providing a direct and interpretable error measure:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|. \tag{3}$$

where y_i represents the actual energy consumption, $\hat{y_i}$ is the predicted consumption, $\bar{y}$ is the mean of the actual values, and n is the number of observations. These metrics were calculated for the test dataset to evaluate the model's performance on unseen data. The combination of a high R^2 score and low RMSE and MAE values provided strong evidence of the XGBoost model's capability to accurately forecast energy consumption. To visually assess model behavior, line plots comparing actual and predicted daily averages were generated for both training and testing periods. This allowed identification of trends, deviations, and periods of strong or weak model performance. Additionally, residual plots were created to further examine error distributions and detect any systematic biases. Collectively, these evaluations confirmed the model's robustness and stability in forecasting energy consumption in the steel manufacturing context.

3 | Results and Discussion

The performance of the XGBoost model in forecasting energy consumption is summarized in *Table 2*. The results show that the model achieved highly accurate predictive performance on both the training and testing datasets. On the training set, the model achieved an R^2 of 0.999 and very low error metrics (RMSE = 0.299, MAE = 0.228), indicating that it effectively learned the underlying patterns in the multivariate energy consumption data.

Although a slight reduction in performance was observed on the test set, the model still maintained strong generalization, with an R^2 score of 0.997, RMSE of 0.667, and MAE of 0.564. This small gap between training and test performance suggests minimal overfitting and that the selected hyperparameters and preprocessing procedures were appropriate for the prediction task.

The high R^2 values for both datasets show that XGBoost successfully captured the nonlinear dependencies among the input features, including reactive power parameters, CO_2 emission levels, temporal indicators (hour, day, month), and categorical variables such as day-of-week and weekend status. These features collectively enhanced the model's ability to identify detailed energy consumption patterns across different operational conditions. Furthermore, the relatively low RMSE and MAE values confirm that the differences between predicted and observed energy usage were small, demonstrating the model's stability and accuracy in practical forecasting scenarios.

Table 2. Performance metrics of the XGBoost model for forecasting energy consumption on the training and testing datasets.

Metric Type	R^2 Score	RMSE	MAE
Training	0.999	0.299	0.228
Testing	0.997	0.667	0.564

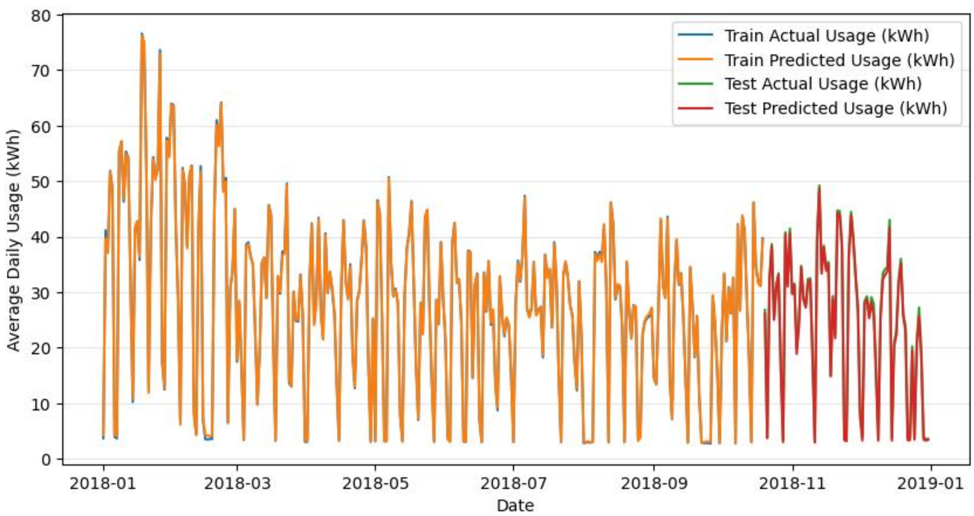

Fig. 1. Actual and predicted average daily energy consumption for the training and testing datasets using the XGBoost model.

Fig. 1 compares the actual and predicted average daily energy consumption for both the training and testing periods. The close alignment between the observed and forecasted values further supports the numerical findings. During the training period, the predicted line closely matches the actual consumption pattern, indicating the model's strong ability to learn complex nonlinear relationships. Similarly, during the testing period, the predictions closely track actual usage, capturing both short-term variations and longer-term trends. These results indicate that the model effectively utilizes the diverse feature set, including reactive power measurements, CO_2 emissions, temporal features, and categorical indicators such as week status, to learn detailed patterns in energy consumption behaviour. The model's high predictive accuracy suggests that machine learning–based forecasting is a viable approach for supporting energy optimization in small-scale steel production environments.

To further assess model performance and identify potential sources of prediction error, a residual analysis was conducted for both the training and testing datasets. Figure 2 illustrates the residuals, defined as the difference between actual and predicted energy consumption values, plotted over time. For the training period, the residuals are tightly clustered around zero, indicating that the model captured the majority of the underlying patterns in the data with minimal bias. Occasional fluctuations are observed but remain relatively small, suggesting stable, consistent learning. During the testing period, the residuals exhibit slightly larger variance, particularly toward the end of the timeline. This increase is expected as the model is exposed to previously unseen data, yet the error magnitudes remain within acceptable limits. The residuals do not exhibit systematic patterns or long-term drift, confirming that the model did not fall prey to temporal, structural, or seasonal biases. Overall, the residual analysis supports the conclusion that the XGBoost model generalizes well and maintains reliable predictive performance across different periods.

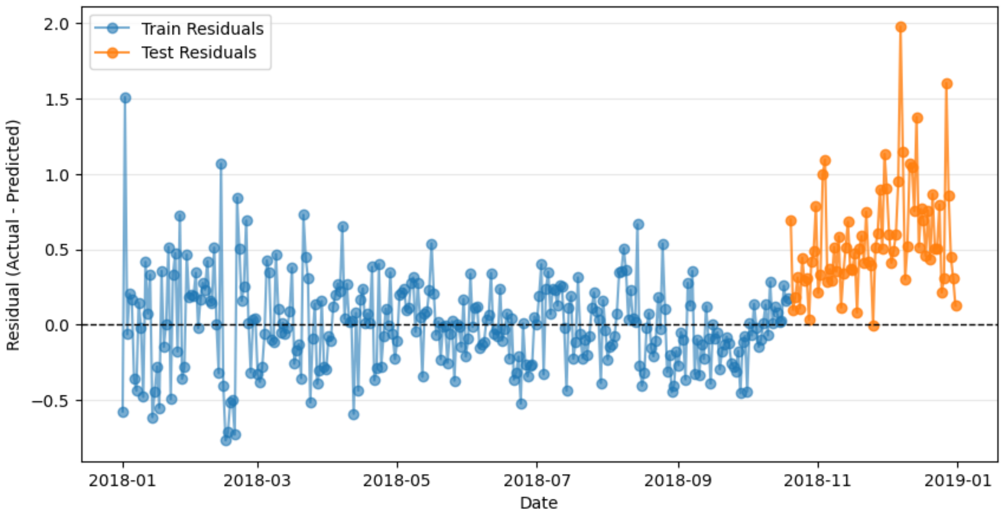

Fig. 2. Residual plot for training and testing datasets showing the difference between actual and predicted energy consumption over time.

To understand which variables contributed most to the model's predictive accuracy, a feature importance analysis was conducted using XGBoost's built-in importance scores. Figure 3 presents the ranked importance of all input features. The results indicate that Lagging_Current_Power_Factor and Lagging_Current_Reactive.Power_kVarh were the most influential predictors, reflecting their strong association with fluctuations in energy usage. These features are directly tied to electrical load behavior, making them highly relevant for forecasting consumption in steel production processes. The third most important feature, CO_2 (tCO_2), also played a significant role, suggesting a meaningful relationship between emission levels and operational energy demand. Other features, such as Leading_Current_Power_Factor, month, and NSM (Number of Seconds from Midnight), contributed moderately, capturing temporal trends and variations in load characteristics throughout the day and year. Categorical indicators, including day of the week and weekend status, exhibited lower importance scores, indicating that energy consumption patterns in this steel manufacturing context are driven more by operational factors than by weekly scheduling differences.

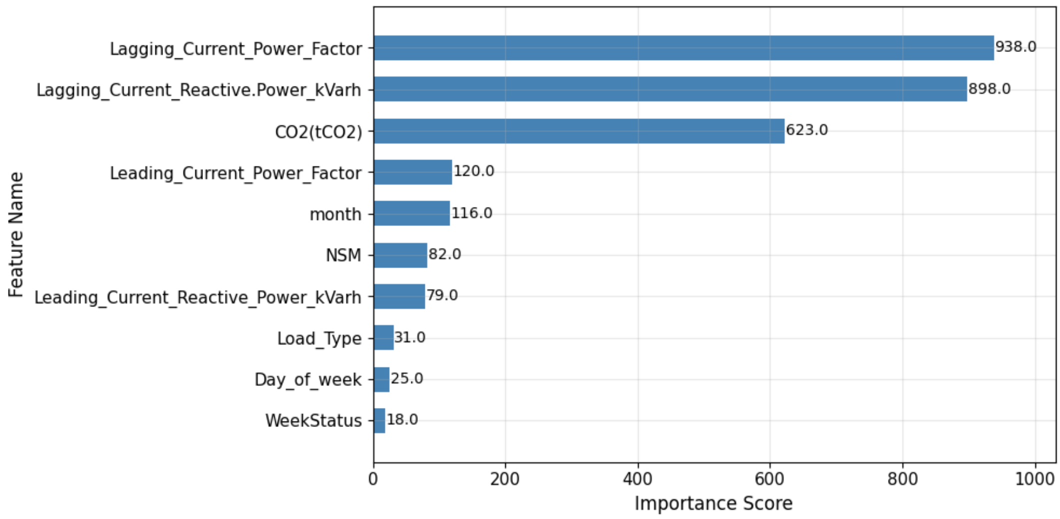

Fig. 3. Feature importance ranking produced by the XGBoost model.

The results of this study demonstrate that the XGBoost model is highly effective for forecasting energy consumption in a small-scale steel industry environment. The exceptionally high R^2 scores on both training (0.999) and testing (0.997) datasets indicate that the model successfully captures the complex nonlinear relationships inherent in industrial energy usage. This level of accuracy highlights XGBoost's strong capability to model multivariate time-series data, especially when intertwined operational, environmental, and temporal factors influence consumption. A notable observation is the model's strong generalization performance, evidenced by the relatively small gap between training and testing errors. While a slight increase in RMSE and MAE on the test set is expected, the overall performance confirms that overfitting was minimized through the chosen preprocessing strategy and hyperparameter configuration. This suggests that the model can reliably support real-world consumption forecasting tasks without substantial degradation when exposed to unseen data. The feature importance results further emphasize the dominance of electrical power-related parameters such as Lagging_Current_Power_Factor and Lagging_Current_Reactive.Power_kVarh, reflecting their direct correlation with operational energy demand in steel manufacturing. The importance of CO_2 emissions as a contributing factor implies a meaningful dependency between production load intensity and emission behavior, underscoring the value of integrating environmental indicators into predictive analytics. Temporal features also contributed significantly, indicating the presence of daily cycles and seasonal effects in the plant's energy usage patterns.

These findings imply that machine learning–based forecasting, particularly with XGBoost, can provide substantial value to small-scale steel producers seeking to optimize energy use. Integrating such predictive systems into existing energy management workflows may support more informed scheduling, load balancing, and cost control. Furthermore, accurately forecasting consumption enables manufacturers to plan for demand-response programs, negotiate more favorable electricity tariffs, and reduce the environmental impact associated with inefficient energy practices.

Despite its strengths, this study presents several limitations. First, the dataset originates from a single steel plant in South Korea, which may limit the model's generalizability to other regions, production scales, or technological contexts. Differences in equipment, production schedules, and energy policies could alter consumption behavior and reduce model transferability. Second, the study relies exclusively on XGBoost. However, it performs well; a comparative analysis with deep learning models such as LSTMs or Temporal Convolutional Networks could provide deeper insights into the model's suitability for long-term forecasting. Additionally, missing values in the dataset were filled using simple imputation (zeros), which may not fully reflect the true variability in the underlying processes. Future research should explore integrating more granular operational data, such as furnace temperature, production batch size, and machine operating status, to enhance predictive performance. Incorporating real-time sensor data and Industrial Internet of Things (IIoT) systems would also enable the development of adaptive online learning models that dynamically respond to changing production conditions. Comparative studies across multiple steel plants would strengthen the generalizability of findings and facilitate the creation of a robust, industry-wide forecasting framework. Furthermore, future work could evaluate hybrid systems that combine prediction with optimization algorithms to support automated decision-making for energy efficiency and carbon-reduction initiatives.29

4 | Conclusion

This study demonstrated the effectiveness of machine learning, specifically the XGBoost algorithm, in forecasting energy consumption within a small-scale steel manufacturing environment. Using a comprehensive dataset comprising operational, temporal, and environmental variables, the model achieved outstanding predictive performance, with R^2 scores of 0.999 on the training set and 0.997 on the test set. These results confirm the model's ability to capture complex nonlinear interactions among reactive power factors, CO_2 emissions, and time-based features that drive energy use patterns in steel production processes.

The findings highlight the practical value of integrating advanced predictive analytics into industrial energy management systems. Accurate energy forecasts can support more strategic decision-making, including production scheduling, load optimization, cost allocation, and environmental compliance. For small- and medium-scale steel producers, such forecasting tools offer a cost-effective pathway toward improved efficiency and reduced emissions, aligning with global efforts to promote sustainable industrial operations.

Authors' Contributions

T. R. N.: Writing-original draft, Methodology, Data Curation, Conceptualization, Software, and Visualization, Writing-Review & Editing, and Validation. R. S.: Validation, Writing-Review & Editing, and Formal Analysis. R. S.: Writing-Review & Editing, Formal Analysis, and Investigation. A. A.: Writing-Review & Editing, Formal Analysis, and Investigation. The authors have read and agreed to the published version of the manuscript.

Data Availability

The dataset used in this study, Steel Industry Energy Consumption, is publicly available from the UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/851/steel+industry+energy+consumption.

Funding

No external funding was received for this research.

Conflict of Interest

There are no competing interests to declare.

Consent for Publication

The authors have given consent for the publication of this manuscript.

Ethics Approval and Consent to Participate

The authors confirm that this research did not involve human participants or animal subjects.

References

[1] Gajdzik, B., & Wolniak, R. (2021). Transitioning of steel producers to the steelworks 4.0—Literature review with case studies. *Energies*, *14*(14), 4109. https://doi.org/10.3390/en14144109

[2] Zhang, J., Shen, H., Chen, Y., Meng, J., Li, J., He, J., ... & Tao, S. (2023). Iron and steel industry emissions: a global analysis of trends and drivers. *Environmental science & technology*, *57*(43), 16477-16488. https://pubs.acs.org/doi/pdf/10.1021/acs.est.3c05474

[3] Fakher, H. A., & others. (2023). The impact of gross domestic product, financial development, energy consumption on environmental quality: with emphasis on six environmental indicators. *Journal of natural environment*, *76*(2), 345–363. (**In Persian**). https://www.sid.ir/fileserver/jf/721-285046-x-1370621.pdf

[4] Fakher, H. A. (2022). Threshold impact of financial development on the composite environmental quality index with emphasis on the role of research and development: using multi-criteria decision makin and principal component analysis. *Journal of decisions and operations research*, *6*(Special Issue), 1–25. https://doi.org/10.22105/dmor.2021.272043.1321

[5] Chen, Y., & Lin, B. (2021). Towards the environmentally friendly manufacturing industry--the role of infrastructure. *Journal of cleaner production*, *326*, 129387. https://doi.org/10.1016/j.jclepro.2021.129387

[6] Shao, Y., Li, J., & Zhang, X. (2022). The impact of financial development on CO2 emissions of global iron and steel industry. *Environmental science and pollution research*, *29*(29), 44954–44969. https://doi.org/10.1007/s11356-022-18977-7

[7] Mahmood, S., Misra, P., Sun, H., Luqman, A., & Papa, A. (2024). Sustainable infrastructure, energy projects, and economic growth: mediating role of sustainable supply chain management. *Annals of operations research*. https://doi.org/10.1007/s10479-023-05777-6

[8] Hui, L., Luo, Z., Liu, K., & others. (2024). Impact of pollution prevention practices and green environmental practices on sustainable performance: Empirical evidence from Chinese SMEs. *Environmental research*, *255*, 118991. https://doi.org/10.1016/j.envres.2024.118991

[9] Fakher, H. A., Panahi, M., Emami, K., Peykarjou, K., Zeraatkish, S. Y., & others. (2021). New insight into development of environmental-economic model based on a composite environmental quality index: an application of principal components analysis. *Journal of decisions and operations research*, *6*(2), 183–209. https://doi.org/10.22097/eeer.2021.280746.1192

[10] Fakher, H. A., Panahi, M., Emami, K., Peykarjou, K., & Zeraatkish, S. Y. (2021). New insight into examining the role of financial development in economic growth effect on a composite environmental quality index. *Environmental science and pollution research*, *28*(43), 61096–61114. https://doi.org/10.1007/s11356-021-15047-2

[11] Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., … Abd Rahman, S. (2023). Urban air quality classification using machine learning approach to enhance environmental monitoring. *Leuser journal of environmental studies*, *1*(2), 62–68. https://doi.org/10.60084/ljes.v1i2.99

[12] Castelli, M., Clemente, F. M., Popovič, A., Silva, S., & Vanneschi, L. (2020). A machine learning approach to predict air quality in California. *Complexity*, *2020*(1), 8049504. https://doi.org/10.1155/2020/8049504

[13] Kolawole, A. S., & Iyiola, A. O. (2023). Environmental pollution: threats, impact on biodiversity, and protection strategies. In *Sustainable utilization and conservation of africa's biological resources and environment* (pp. 377–409). Springer. https://doi.org/10.1007/978-981-19-6974-4_14

[14] Dhayal, K. S., Giri, A. K., Agrawal, R., Agrawal, S., Samadhiya, A., & Kumar, A. (2026). Do the innovative technological advancements foster the green transition pathways for industry 5.0? A perspective toward carbon neutrality. *Benchmarking: an international journal*, *33*(4), 1070–1098. https://doi.org/10.1108/BIJ-04-2024-0330

[15] Hilmi, E., Yandri, E., Uhanto, U., Saiful, R., & Hamja, N. (2024). Hybrid Energy Solutions for Sustainable Offshore Oil and Gas Operations: Leveraging Thermoelectric, Solar, and Wind Potential. *Leuser journal of environmental studies*, *2*(2), 52–61. https://doi.org/10.60084/ljes.v2i2.218

[16] Chen, X., & Chen, M. (2024). Energy, environment and industry: instrumental approaches for environmental regulation on energy efficiency. *Environmental impact assessment review*, *105*, 107439. https://doi.org/10.1016/j.eiar.2024.107439

[17] Fakher, H. A., Abedi, Z., Ahmadian, M., & Shaygani, B. (2018). Comparative examine the impact of financial development (based on money market and capital market) in the intensity of economic growth effects on the environmental performance. *Environmental researches*, *9*(17), 133–146. https://dor.isc.ac/dor/20.1001.1.20089597.1397.9.17.17.4

[18] Safari, A., Daneshvar, M., & Anvari-Moghaddam, A. (2024). Energy intelligence: A systematic review of artificial intelligence for energy management. *Applied sciences*, *14*(23), 11112. https://doi.org/10.3390/app142311112

[19] Aprizal, A., Wiranatakusuma, D. B., Rizki, M., & Anugrah, R. A. (2025, April). The role of technology in climate resilience: A systematic literature review and mapping study approach. In *Journal of Physics: Conference Series* (Vol. 2989, No. 1, p. 012037). IOP Publishing. https://doi.org/10.1088/1742-6596/2989/1/012037

[20] Zong, Z., & Guan, Y. (2025). AI-driven intelligent data analytics and predictive analysis in Industry 4.0: Transforming knowledge, innovation, and efficiency. *Journal of the knowledge economy*, *16*(1), 864–903. https://doi.org/10.1007/s13132-024-02001-z

[21] Fakher, H. A. (2021). The role of environmental sustainability, foreign direct investment and trade openness in economic growth: with emphasis on the causal linkage. *Big data and computing visions*, *1*(2), 57–70. https://doi.org/10.22105/bdcv.2021.142227

[22] Fakher, H. A. (2020). Analytical insights on the relationship between economic growth and environmental degradation in framework of EKC hypothesis and various environmental indicators. *Innovation management and operational strategies*, *1*(3), 252–268. https://doi.org/10.22105/imos.2021.272348.1032

[23] Mhlanga, D. (2023). Artificial intelligence and machine learning for energy consumption and production in emerging markets: A review. *Energies*, *16*(2), 745. https://doi.org/10.3390/en16020745

[24] Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. *Sustainability*, *15*(9), 7087. https://doi.org/10.3390/su15097087

[25] Forootan, M. M., Larki, I., Zahedi, R., & Ahmadi, A. (2022). Machine learning and deep learning in energy systems: A review. *Sustainability*, *14*(8), 4832. https://doi.org/10.3390/su14084832

[26] Shu, X., & Ye, Y. (2023). Knowledge discovery: Methods from data mining and machine learning. *Social science research*, *110*, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817

[27] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

[28] Fouladvand, S., Noshad, M., Goldstein, M. K., Periyakoil, V. J., & Chen, J. H. (2023). Mild cognitive impairment: Data-driven prediction, risk factors, and workup. *AMIA summits on translational science proceedings*, *2023*, 167. https://pmc.ncbi.nlm.nih.gov/articles/PMC10283085/pdf/2267.pdf

[29] VE, S., Shin, C., & Cho, Y. (2021). Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city. *Building research & information*, *49*(1), 127–143. https://doi.org/10.1080/09613218.2020.1809983

[30] Rahmanparast, A., Milani, M., Camci, M., Karakoyun, Y., Acikgoz, O., & Dalkilic, A. S. (2025). A comprehensive method for exploratory data analysis and preprocessing the ASHRAE database for machine learning. *Applied thermal engineering*, *273*, 126556. https://doi.org/10.1016/j.applthermaleng.2025.126556

[31] Noviandy, T. R., Maulana, A., Idroes, G. M., Suhendra, R., Adam, M., Rusyana, A., & Sofyan, H. (2023). Deep learning-based bitcoin price forecasting using neural prophet. *Ekonomikalia journal of economics*, *1*(1), 19–25. https://doi.org/10.60084/eje.v1i1.51

[32] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). https://doi.org/10.1145/2939672.2939785

[33] Ali, Y. A., Awwad, E. M., Al-Razgan, M., & Maarouf, A. (2023). Hyperparameter search for machine learning algorithms for optimizing the computational complexity. *Processes*, *11*(2), 349. https://doi.org/10.3390/pr11020349

[34] Noviandy, T. R., Idroes, G. M., Fauzi, F. M., & Idroes, R. (2024). Application of ensemble machine learning methods for QSAR classification of leukotriene A4 hydrolase inhibitors in drug discovery. *Malacca pharmaceutics*, *2*(2), 68–78. https://doi.org/10.60084/mp.v2i2.217

[35] Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., & Ray, S. (2024). A model-agnostic interpretability approach to predicting customer churn in the telecommunications industry. *Infolitika journal of data science*, *2*(1), 34–44. https://doi.org/10.60084/ijds.v2i1.199