Article history:
Received 10 May 2026
Revised 09 August 2026
Accepted 13 August 2026
Initial Publish 27 September 2026
Published online 01 March 2027

Journal of Resource Management and Decision Engineering

Volume 6, Issue 3, pp 1-16

Developing an Integrated Fuzzy Model for Ranking the Factors Causing Delay-Related Claims in Industrial Projects

Mohammad Ali. Niroomand[1], Rojeen. Mamershafai[2], Fereshteh. Khalili Peykani[3], Omid. Shabkhizan[4], Meysam. Mahmoudi[5*]

[1] Faculty of Engineering, Shahrood University, Shahrood, Iran
[2] Department of Energy Systems Engineering, Amirkabir University of Technology, Tehran, Iran
[3] Department of Economics and Management, SR.C., Islamic Azad University, Tehran, Iran
[4] Department of Management, SR.C., Islamic Azad University, Tehran, Iran
[5] Department of Mechanical Engineering, Faculty of Engineering, Velayat University, Iranshahr, Iran

*** Corresponding author email address**: m.mahmoudi@velayat.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Niroomand, M. A.., Mamershafai, R., Khalili Peykani, F., Shabkhizan, O., & Mahmoudi, M. (2027). Developing an Integrated Fuzzy Model for Ranking the Factors Causing Delay-Related Claims in Industrial Projects. *Journal of Resource Management and Decision Engineering*, *6*(3), 1-16.
https://doi.org/10.61838/kman.jrmde.404

ABSTRACT

This study aimed to develop an integrated fuzzy decision-making model for identifying and ranking the factors causing delay-related claims in industrial projects based on their probability of occurrence and effects on project time, cost, and quality. This applied, quantitative, cross-sectional study was conducted among 80 managers and specialists involved in industrial projects in Assaluyeh, Iran. Following a review of the literature and expert evaluation, 66 preliminary delay-related claim factors were identified and subsequently reduced to 48 factors. Data were collected using structured expert questionnaires assessing the probability of occurrence and the time, cost, and quality impacts of the identified factors. The analytical procedure combined statistical screening, probability-impact analysis, fuzzy evaluation, and multi-criteria decision-making. The relative importance of project-performance dimensions was determined through pairwise comparison, and the final prioritization of factors was performed using an integrated AHP-TOPSIS framework. Fuzzy logic was employed to accommodate uncertainty and imprecision in expert judgments. The final AHP-TOPSIS analysis assigned relative importance coefficients of 0.677, 0.674, and 0.655 to time, cost, and quality impacts, respectively. The highest-ranked factor was submission of an unrealistically low tender price solely to win the contract, followed by reduced foreign-contractor investment due to economic instability, prolonged activation of letters of credit, insufficient managerial focus caused by simultaneous involvement in multiple projects, and inadequate financial resources or diversion of project funds. Delayed contractor payments, governmental procurement regulations, delayed engineering responses, inaccurate estimation of work quantities, and delayed managerial problem resolution were also among the ten highest-ranked factors. The findings further showed that several high-priority factors simultaneously influenced more than one project-performance dimension, confirming the multidimensional nature of delay-related claims. Delay-related claims in industrial projects are primarily associated with interconnected contractual, financial, managerial, engineering, and procurement deficiencies, and the integrated fuzzy AHP-TOPSIS model provides a systematic mechanism for identifying and prioritizing these factors to support proactive risk and claim management.

Keywords: *Industrial Projects; Delay Claims; Claim Management; Risk Management; Fuzzy TOPSIS; AHP; Project Delay; Multi-Criteria Decision-Making.*

1. Introduction

The transition toward environmentally sustainable and resource-efficient industrial production has become a central challenge for contemporary energy systems. Industrial facilities are major consumers of electricity and thermal energy because manufacturing, processing, heating, cooling, material transformation, and auxiliary operations require substantial and often continuous energy inputs. The environmental consequences of this dependence extend beyond direct energy costs, as the quantity and composition of energy consumed influence greenhouse-gas emissions, resource depletion, and the broader environmental footprint of industrial production. Accordingly, improving energy efficiency is no longer merely an operational strategy for lowering production expenses; it represents an integral component of industrial decarbonization, sustainable development, and long-term competitiveness. Life-cycle investigations of emerging energy technologies further demonstrate that environmental performance cannot be evaluated independently of energy requirements throughout production and conversion processes, because energy consumption and environmental impacts remain closely interconnected across technological pathways (Zhang et al., 2025). At the macroeconomic level, improvements in energy efficiency, technological innovation, and environmentally oriented development have similarly been identified as important elements in advancing sustainability objectives (Feng & Xu, 2025). Thus, industrial energy management increasingly requires approaches capable of evaluating not only how much energy a facility consumes, but also how efficiently that energy is transformed into productive output and how the resulting operating profile affects environmental performance.

The concept of industrial energy efficiency is inherently multidimensional. Conventional energy analysis generally compares energy inputs with useful outputs and seeks opportunities for reducing avoidable consumption, whereas thermodynamic approaches provide deeper insight into the quality and irreversibility of energy transformations. Exergy analysis, in particular, has become an established framework for determining where useful energy potential is destroyed within thermal and industrial processes and for distinguishing between losses that are unavoidable and those that may be reduced through technological or operational modifications (Peng et al., 2025). The extension of energy and exergy assessment toward entropy-generation minimization and exergoenvironmental analysis has further strengthened the connection between thermodynamic performance and environmental sustainability (Ordonez et al., 2022). More comprehensive approaches integrate exergy with economic and environmental criteria, allowing inefficiencies to be considered in relation to both their financial consequences and environmental burdens (Azubuike et al., 2025). These developments illustrate the value of physically grounded efficiency analysis, particularly when detailed process information is available. Nevertheless, industrial systems are increasingly complex, heterogeneous, and data-intensive, and purely thermodynamic assessments may be difficult to implement continuously across large numbers of facilities with different technologies, production scales, operating schedules, and environmental conditions.

A further complication is that industrial sustainability cannot be reduced to energy consumption alone. A plant can decrease its gross energy demand while retaining an environmentally unfavorable fuel mix, or it can use comparatively cleaner energy while remaining inefficient in converting that energy into production output. Consequently, the simultaneous consideration of energy efficiency and environmental performance is essential. Broader research on sustainability demonstrates that environmental impacts are increasingly relevant not only to engineering decisions but also to societal and economic evaluations of production systems. Studies of environmentally preferable agricultural products, for example, have shown that consumer awareness regarding environmentally harmful inputs can shape perceptions and market behavior (Haghjou et al., 2020), while willingness to pay premiums for pesticide-free products illustrates the economic value that consumers may attach to reduced environmental and health burdens (Hayati et al., 2012). Although such studies address a different sector, they highlight an important principle for industrial sustainability: environmental performance increasingly represents an economically and socially meaningful dimension of production rather than an external consideration. Industrial organizations are therefore under growing pressure to identify strategies capable of improving production efficiency while simultaneously reducing emissions and other environmental burdens.

The sustainability of industrial assets also depends on factors that extend beyond immediate energy demand. Equipment condition, degradation, maintenance practices, corrosion, and material durability can influence operational efficiency and determine the frequency with which equipment must be repaired or replaced. The relationship between corrosion protection and sustainability illustrates how extending equipment lifetime and preserving functional performance may reduce material demand, resource consumption, maintenance requirements, and environmental impacts over the life cycle of industrial systems (Prošek et al., 2025). From an energy-management perspective, these considerations reinforce the importance of equipment efficiency as a potentially actionable operational variable. Poorly maintained or degraded machinery can require more energy to provide the same productive service, whereas maintenance, retrofitting, and efficiency improvement may lower specific energy consumption without requiring reductions in production output. Therefore, a comprehensive industrial energy-efficiency framework should move beyond aggregate energy measurements and investigate the operational variables responsible for differences in consumption and environmental performance across facilities.

Traditionally, industrial energy management has relied on energy audits, engineering calculations, benchmarking, historical comparisons, and expert-defined operating rules. These methods remain important, particularly for diagnosing clearly identifiable equipment-level inefficiencies, but their ability to characterize complex relationships among numerous operational factors may be limited. Industrial energy consumption can respond simultaneously to production volume, production intensity, equipment efficiency, operating duration, fuel use, scheduling practices, and ambient environmental conditions. These relationships may be nonlinear, interactive, and context dependent. Advanced computational methods have therefore become increasingly important for the modeling, prediction, and optimization of engineering systems, particularly where conventional analytical representations are unable to capture complex interactions efficiently (Krzywanski et al., 2024). Data-driven approaches provide an alternative paradigm in which patterns are learned directly from observed operating data, enabling organizations to shift from retrospective assessments of past consumption toward predictive estimates of future or counterfactual performance.

Machine learning is particularly well suited to this transition because industrial facilities increasingly generate large quantities of structured operational data through sensors, supervisory-control systems, energy meters, production-management platforms, and environmental monitoring technologies. Research on power systems has demonstrated substantial interest in machine-learning and data-driven approaches for predictive analysis, while also emphasizing challenges involving data quality, interpretability, computational requirements, and generalizability (Strielkowski et al., 2023). In smart manufacturing environments, machine-learning models have likewise been proposed as mechanisms for optimizing energy use and resource consumption by exploiting relationships that may not be readily identifiable through conventional analysis (Nozari et al., 2025). Such capabilities are particularly relevant when energy consumption reflects simultaneous changes in multiple operating variables. Rather than considering production output or operating hours independently, predictive algorithms can estimate their combined and potentially nonlinear effects and can therefore offer a more realistic representation of industrial energy behavior.

A broad range of machine-learning algorithms can be used for energy prediction, including conventional regression techniques, artificial neural networks, support-vector methods, decision trees, Random Forests, and gradient-boosting algorithms. Comparative research on electricity consumption has demonstrated the applicability of decision-tree and artificial neural-network methods alongside conventional statistical techniques for explaining heterogeneous energy-use patterns (Nsangou et al., 2022). Tree-based approaches are especially attractive for structured energy datasets because they can accommodate nonlinearities and interactions without requiring researchers to specify a rigid functional relationship between predictors and outcomes in advance. Recent work using optimized tree-based learning combined with explainable artificial intelligence has further demonstrated the potential of such methods for predicting energy-related outcomes while simultaneously improving understanding of the variables underlying model predictions (Tang, 2026). The incorporation of explainability is particularly important in applied industrial settings, where managers require more than accurate forecasts: they need evidence regarding which operating variables are responsible for inefficient performance and which factors can realistically be modified.

Ensemble-learning approaches offer additional advantages because they combine multiple learners to increase predictive stability and reduce the limitations associated with individual models. Random Forest algorithms aggregate numerous decision trees and are comparatively robust in datasets containing interactions and complex variable structures, while gradient-boosting frameworks iteratively construct models that correct prediction errors and often achieve strong performance with structured tabular data. Boosted ensemble regression trees have been successfully employed to analyze the energy and environmental performance of building systems, demonstrating their usefulness for identifying opportunities to reduce energy demands associated with air-conditioning and lighting operations (Zaki et al., 2025). Research on environmental influences on renewable-energy prediction has similarly shown that tree-based ensemble learning can accommodate the effects of changing environmental conditions (Doğan et al., 2025). Such findings are relevant for industrial facilities because variables such as ambient temperature may alter heating, ventilation, cooling, refrigeration, or process-energy requirements and therefore contribute to fluctuations in consumption even when production characteristics remain relatively stable.

Support Vector Regression provides a conceptually different machine-learning alternative that is valuable for comparative model assessment. Through kernel functions, SVR can represent nonlinear relationships while controlling model complexity and is therefore useful when energy outcomes depend on complex combinations of predictors. Comparing algorithms based on multiple performance criteria rather than relying on a single technique can provide stronger evidence regarding the most appropriate predictive model for a given industrial dataset. Metrics such as the coefficient of determination, root-mean-square error, mean absolute error, and mean absolute percentage error provide complementary information concerning explanatory ability and prediction error. Cross-validation further allows investigators to determine whether model performance is stable across different subsets of observations rather than being specific to a single training-and-testing partition. This comparative approach is consistent with the broader computational literature emphasizing that model selection should reflect predictive performance, robustness, interpretability, and suitability for the structure of the underlying engineering problem (Krzywanski et al., 2024).

Prediction alone, however, is insufficient for effective industrial energy management. A highly accurate model may forecast energy consumption without revealing which variables account for excessive consumption or what operational modifications should subsequently be implemented. Feature-importance analysis provides an important bridge between predictive analytics and managerial decision-making by estimating the relative contribution of predictors to model performance. For industrial managers, knowing that energy demand can be accurately predicted is less actionable than knowing whether consumption is principally associated with production intensity, equipment condition, operating hours, ambient temperature, fuel use, or another operational characteristic. Explainability is becoming increasingly important as machine-learning systems are incorporated into energy decision-making, because transparent models can enhance confidence in predictions and facilitate the translation of algorithmic output into operational interventions (Tang, 2026). At the same time, the computational footprint of artificial intelligence itself deserves consideration. Work on sustainable artificial intelligence has emphasized that algorithmic energy efficiency should be recognized when designing data-driven solutions, particularly as the scale and frequency of computational modeling increase (Ferro et al., 2023). Accordingly, industrial analytics should pursue not only predictive accuracy but also transparent, efficient, and decision-relevant modeling.

The next step beyond prediction and interpretation is prescriptive analysis. Once relationships between operating characteristics and energy outcomes have been learned, predictive models can be incorporated into optimization procedures that search for feasible combinations of operating conditions associated with improved performance. Machine learning has increasingly been used not only for forecasting but also for optimizing complex energy and biological conversion processes. For example, machine-learning approaches applied to commercial-scale anaerobic digestion have demonstrated the potential to predict process performance under heterogeneous feedstocks and operating conditions and subsequently support performance optimization (Long et al., 2025). In industrial contexts, similar approaches can evaluate counterfactual scenarios involving changes in equipment efficiency, production scheduling, operating duration, demand management, or other controllable variables. The value of these methods lies in their ability to transform historical operational data into forward-looking recommendations rather than limiting analysis to retrospective identification of inefficiency.

Industrial energy optimization is also fundamentally a multi-objective problem. Minimizing energy consumption without considering production could generate recommendations that simply reduce operating activity, while minimizing emissions independently of cost or throughput could produce solutions that are technically desirable but operationally impractical. Energy, environmental, economic, and productive objectives must therefore be evaluated simultaneously. Advanced exergy, exergoeconomic, and exergoenvironmental analyses have demonstrated the importance of integrating technical, financial, and environmental dimensions when assessing sustainable thermal-energy systems (Azubuike et al., 2025), while the broader literature on energy, exergy, and environmental assessment emphasizes the importance of identifying trade-offs among efficiency losses and environmental consequences (Ordonez et al., 2022). A multi-objective framework can consequently search for operating conditions in which energy consumption and CO_2 emissions are reduced subject to constraints that preserve an acceptable level of production. This is especially important for real-world managerial decisions, because an intervention is unlikely to be adopted if its environmental benefits are obtained at the expense of substantial production losses or operational instability.

Despite considerable advances in machine learning, computational optimization, and environmental-performance analysis, these components are frequently investigated in isolation. Some studies focus primarily on forecasting energy demand, others identify thermodynamic inefficiencies, and others optimize selected process parameters. This fragmentation creates a methodological gap between descriptive assessment, predictive modeling, explanation, and operational prescription. Smart-factory research has emphasized the potential of machine learning for improving both energy optimization and resource use (Nozari et al., 2025), while broader computational research demonstrates the expanding ability of advanced methods to integrate modeling, prediction, and optimization within engineering applications (Krzywanski et al., 2024). Nevertheless, an integrated framework is required to connect these functions systematically: first characterizing facility performance, then predicting energy and environmental outcomes, identifying influential operational factors, and finally translating these relationships into feasible optimization scenarios. Such integration is particularly important across heterogeneous industrial facilities because a single aggregate indicator may conceal substantial differences in operating intensity, equipment efficiency, fuel consumption, environmental exposure, and production scale.

An additional methodological consideration concerns the availability of industrial operational data. Proprietary restrictions, confidentiality requirements, inconsistencies in measurement systems, and differences in process configuration can make sufficiently large and harmonized multi-facility datasets difficult to obtain. Methodological studies may therefore use synthetic datasets as controlled environments for developing and testing integrated analytical pipelines before validation using genuine plant records. In the present framework, the use of facility-level observations encompassing energy consumption, production output, electricity expenditure, fuel consumption, CO_2 emissions, operating hours, ambient temperature, equipment efficiency, and production intensity enables the simultaneous representation of productive, economic, operational, and environmental dimensions. This structure is especially appropriate for evaluating whether machine-learning methods can identify realistic associations among industrial variables and whether these associations can subsequently support optimization. The importance of incorporating environmental conditions is supported by evidence that environmental variability can materially influence energy-related predictive performance (Doğan et al., 2025), while the integration of energy and environmental criteria reflects the broader movement toward more comprehensive sustainability assessment (Prošek et al., 2025; Zhang et al., 2025).

Taken together, the literature indicates a transition from static and predominantly descriptive energy assessment toward integrated data-driven energy management. Thermodynamic and exergy methods remain essential for understanding physical inefficiencies (Peng et al., 2025), but machine-learning techniques extend analytical capability by identifying nonlinear patterns across heterogeneous operational datasets (Nsangou et al., 2022; Strielkowski et al., 2023). Ensemble models can improve prediction of energy and environmental outcomes (Zaki et al., 2025), explainable tree-based frameworks can provide greater transparency regarding influential factors (Tang, 2026), and machine-learning-assisted optimization can translate predictive relationships into improved operating conditions (Long et al., 2025). At the same time, sustainability requires balancing energy savings with environmental consequences, production requirements, equipment condition, and resource efficiency rather than optimizing any single indicator in isolation (Azubuike et al., 2025; Feng & Xu, 2025). An integrated approach combining these analytical stages can therefore provide a stronger foundation for evidence-based industrial decision-making by connecting measurement, prediction, interpretation, and optimization within a unified framework.

Accordingly, the aim of this study was to develop and demonstrate a data-driven framework integrating machine-learning prediction, feature-importance analysis, and multi-objective optimization to assess energy efficiency, identify the principal operational and environmental drivers of energy consumption, and optimize energy use and CO_2-emission performance in heterogeneous industrial systems while maintaining production output.

2. Methods and Materials

2.1. Data Description

The analysis was conducted on a panel dataset of 1,250 monthly observations collected from 125 industrial facilities, corresponding to an average of ten monthly records per facility. Each record captures the energy, production, cost, and environmental state of a facility for a single calendar month. The facilities span a range of production scales and operating regimes, producing a heterogeneous sample in which both efficient and inefficient operating patterns are represented. Because the dataset is synthetic, it was generated to preserve realistic marginal distributions and inter-variable correlations — for example, the positive association between operating hours and energy consumption, and between fuel consumption and CO_2 emissions — so that the modelling exercise remains representative of genuine industrial behaviour.

2.2. Variables and Feature Engineering

Eight primary variables were recorded for each observation, from which two derived performance indicators and one derived driver (production intensity) were computed. Table 1 summarizes the variable set, together with units and definitions. All predictors were standardized to zero mean and unit variance prior to model training so that kernel- and distance-based methods were not dominated by high-magnitude features; consequently, error metrics reported for the standardized targets (RMSE, MAE) are expressed in standardized units unless otherwise stated.

Table 1

Definition of variables used in the data-driven framework.

Variable	Unit	Definition / role
Energy consumption	MWh/month	Total electrical + thermal energy used by the facility (primary prediction target).
Production output	tonnes/month	Quantity of finished product manufactured during the month.
Electricity cost	10^3 USD/month	Monthly expenditure on purchased electricity.
Fuel consumption	10^3 m^3/month	Natural-gas-equivalent volume of on-site fuel combusted.
CO_2 emissions	tCO_2/month	Direct + indirect carbon dioxide emissions attributed to the facility.
Operating hours	h/month	Total hours of active production during the month.
Ambient temperature	°C	Mean outdoor temperature at the facility location.
Equipment efficiency	%	Ratio of rated to actual specific energy use of core equipment.
Production intensity	tonnes/h	Production output per operating hour (throughput rate).
Energy efficiency*	tonnes/MWh	Derived: production output ÷ energy consumption.
CO_2 emission intensity*	tCO_2/MWh	Derived: CO_2 emissions ÷ energy consumption.

* Derived performance indicators. Predictors were standardized before model training.

2.3. Energy-Efficiency and Environmental-Performance Metrics

Energy efficiency was quantified as the ratio of production output to total energy consumption (tonnes per MWh); higher values indicate that more product is obtained per unit of energy. Environmental performance was quantified through CO_2 emission intensity, defined as CO_2 emissions per unit of energy consumed (tCO_2 per MWh); lower values indicate a cleaner energy profile. These two indicators are complementary: a facility can be energy-efficient yet carbon-intensive if it relies on carbon-heavy fuels, or comparatively clean yet energy-inefficient if it wastes low-carbon energy. Treating them jointly is therefore essential to a balanced assessment.

2.4. Predictive Modelling

Three supervised-learning algorithms were trained to predict the two targets (energy consumption and CO_2 emission intensity). Random Forest aggregates many decorrelated regression trees and is robust to noise and multicollinearity. XGBoost builds an additive ensemble of trees through gradient boosting with regularization, and typically excels on structured tabular data. Support Vector Regression fits a margin-based kernel model and provides a strong non-ensemble baseline. The data were split 80/20 into training and testing partitions, and hyperparameters were tuned by grid search with five-fold cross-validation on the training partition. The final configurations are reported in Table 2.

Table 2

Tuned hyperparameter configurations for each model.

Model	Key hyperparameters	Search range
Random Forest	n_estimators = 500; max_depth = 18; min_samples_leaf = 3	100–800; 6–24; 1–10
XGBoost	n_estimators = 600; max_depth = 6; learning_rate = 0.05; subsample = 0.8	200–900; 3–10; 0.01–0.20; 0.6–1.0
SVR	kernel = RBF; C = 12.0; ε = 0.08; γ = scale	0.1–50; 0.01–0.20; scale/auto

2.5. Feature-Importance Analysis

To interpret the predictive models and identify actionable drivers, feature importance was computed from the best-performing model using the mean decrease in impurity, cross-checked against permutation importance to guard against bias toward high-cardinality features. Importances were normalized to sum to 100% to ease comparison across features.

2.6. Multi-Objective Optimization

The prescriptive stage searched for operating adjustments that jointly minimize energy consumption and CO_2 emissions subject to maintaining production output within one percent of baseline. Three controllable interventions were considered: (i) improved equipment efficiency, achieved through maintenance and retrofit of core equipment; (ii) production scheduling, which redistributes load toward more efficient operating windows; and (iii) energy-demand management, which trims peak demand and idle consumption. The optimization was framed as a constrained multi-objective problem and solved with an evolutionary search over feasible intervention magnitudes, using the trained predictive models as surrogate evaluators of energy and emissions.

2.7. Evaluation Metrics

Predictive accuracy was assessed using the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). R^2 measures the proportion of variance explained; RMSE and MAE quantify average error magnitude (in standardized units for the standardized targets); and MAPE expresses error relative to the observed value, aiding interpretation across facilities of different scale.

3. Findings and Results

This section presents the empirical findings of the framework in the order in which they were produced: descriptive characterization of the sample, correlation structure, efficiency and environmental-performance overview, comparative predictive accuracy for both targets, cross-validation stability, feature-importance attribution, facility-level efficiency tiers, and finally the multi-objective optimization outcomes with an accompanying sensitivity analysis.

Table 3 reports the distributional summary of the eight primary variables and the derived indicators across all 1,250 monthly observations. The average monthly energy consumption was 402 MWh, equivalent to 4,820 MWh per facility-year, with a coefficient of variation of roughly 36% reflecting the heterogeneity of the fleet. Mean CO_2 emission intensity was 0.74 tCO_2/MWh, and mean energy efficiency was 7.84 tonnes/MWh. The wide spread between the minimum and maximum of most variables confirms that the sample captures both lean, well-managed operations and comparatively wasteful ones, which is desirable for training models that must generalize across operating regimes.

Table 3

Descriptive statistics of study variables (n = 1,250 monthly observations).

Variable	Mean	Std. Dev.	Min	Median	Max

Energy consumption (MWh/mo)	401.7	145.2	84.6	388.4	981.3
Production output (tonnes/mo)	3,150	1,182	540	3,038	7,420
Electricity cost (10^3 USD/mo)	48.6	19.2	9.8	46.1	121.4
Fuel consumption (10^3 m^3/mo)	132.4	55.1	22.0	127.8	349.6
CO_2 emissions (tCO_2/mo)	297.6	118.3	58.4	283.2	742.7
Operating hours (h/mo)	620.3	94.8	352	628	731
Ambient temperature (°C)	21.3	8.4	−2.1	21.8	41.0
Equipment efficiency (%)	78.4	9.6	51.2	79.1	96.8
Production intensity (t/h)	5.08	1.79	1.21	4.94	10.62
Energy efficiency (t/MWh)	7.84	2.31	2.90	7.71	15.44
CO_2 intensity (tCO_2/MWh)	0.74	0.14	0.39	0.73	1.18

Facility-year energy consumption = 12 × monthly mean ≈ 4,820 MWh.

Table 4 presents the pairwise Pearson correlations among the principal drivers and outcomes. Energy consumption was most strongly associated with production intensity ($r = 0.79$) and operating hours ($r = 0.71$), and negatively associated with equipment efficiency ($r = −0.58$), consistent with the expectation that higher throughput and longer runtimes raise consumption while better equipment lowers it. CO_2 emissions tracked energy consumption ($r = 0.84$) and fuel consumption ($r = 0.80$) closely, underscoring the fuel-driven nature of the facilities' carbon profile. The moderate correlations among predictors indicate meaningful but not collinear relationships, supporting the use of models able to disentangle interacting effects.

Table 4

Pearson correlation matrix among key variables.

	EC	PI	OH	EE	FC	CO_2	AT
Energy cons. (EC)	1.00	0.79	0.71	−0.58	0.63	0.84	0.22
Prod. intensity (PI)	0.79	1.00	0.41	−0.33	0.48	0.66	0.14
Operating hrs (OH)	0.71	0.41	1.00	−0.19	0.52	0.61	0.09
Equip. eff. (EE)	−0.58	−0.33	−0.19	1.00	−0.36	−0.49	−0.07
Fuel cons. (FC)	0.63	0.48	0.52	−0.36	1.00	0.80	0.18
CO_2 emissions	0.84	0.66	0.61	−0.49	0.80	1.00	0.20
Amb. temp. (AT)	0.22	0.14	0.09	−0.07	0.18	0.20	1.00

Values are Pearson coefficients; $|r| \geq 0.06$ significant at $p < 0.05$ for $n = 1{,}250$.

Table 5 aggregates the two performance indicators across production-scale bands. Larger facilities achieved somewhat higher energy efficiency, reflecting economies of scale, but did not systematically achieve lower CO_2 intensity, which was governed more by fuel mix than by scale. The best-performing decile of facilities combined energy efficiency above 11 tonnes/MWh with CO_2 intensity below 0.60 tCO_2/MWh, demonstrating that the two objectives can be met simultaneously and providing an empirical target for the optimization stage.

Table 5

Performance indicators by facility production-scale band.

Production band	Facilities	Energy eff. (t/MWh)	CO_2 int. (tCO_2/MWh)	Avg. output (t/mo)
Small (< 2,000 t/mo)	38	6.91	0.78	1,410
Medium (2,000–4,000)	54	7.86	0.74	3,020
Large (4,000–6,000)	23	8.42	0.71	4,880
Very large (> 6,000)	10	8.97	0.72	6,640
All facilities	125	7.84	0.74	3,150

Table 6 compares the three algorithms on the held-out test partition for energy-consumption prediction. XGBoost delivered the strongest performance across every metric, with an R^2 of 0.91 and an RMSE of 0.18 (standardized units), followed by Random Forest and then SVR. The consistent ordering across R^2, RMSE, MAE, and MAPE indicates that

XGBoost's advantage is robust and not an artefact of a single metric. In absolute terms, the XGBoost model's MAPE of 4.2% corresponds to an average error of roughly 17 MWh per month at the fleet mean, a level of accuracy suitable for planning and target-setting.

Table 6

Test-set performance for energy-consumption prediction.

Model	R^2	RMSE	MAE	MAPE (%)
XGBoost	0.91	0.18	0.13	4.2
Random Forest	0.88	0.21	0.16	5.1
Support Vector Regression	0.82	0.27	0.20	6.8

RMSE and MAE reported in standardized (z-score) units of the target; MAPE on original scale.

Table 7 reports the analogous comparison for CO_2 emission-intensity prediction. The ranking of models was preserved, with XGBoost again leading ($R^2 = 0.89$, RMSE = 0.20). Accuracy for the environmental target was marginally lower than for energy consumption, which is expected because emission intensity depends additionally on fuel-mix variation that is only partially observed in the feature set. Nonetheless, an R^2 near 0.90 confirms that environmental performance is highly predictable from routinely collected operational variables.

Table 7

Test-set performance for CO_2 emission-intensity prediction.

Model	R^2	RMSE	MAE	MAPE (%)
XGBoost	0.89	0.20	0.15	4.9
Random Forest	0.86	0.23	0.17	5.7
Support Vector Regression	0.80	0.29	0.22	7.4

To confirm that the reported accuracy was not specific to a single train/test split, five-fold cross-validation was performed on the full dataset. Table 8 lists the mean and standard deviation of R^2 and RMSE across folds for the energy-consumption target. The low fold-to-fold variance (R^2 standard deviation ≤ 0.02 for all models) indicates stable, generalizable performance, with XGBoost remaining the top model throughout.

Table 8

Five-fold cross-validation results (energy-consumption target).

Model	Mean R^2	R^2 Std. Dev.	Mean RMSE
XGBoost	0.905	0.012	0.185
Random Forest	0.876	0.016	0.214
Support Vector Regression	0.814	0.021	0.276

Table 9 presents the normalized feature importances derived from the XGBoost energy-consumption model, cross-validated against permutation importance. Production intensity was the dominant driver at 31.4%, followed by operating hours (22.7%), equipment efficiency (18.5%), and ambient temperature (11.2%). Together these four factors accounted for 83.8% of the model's explanatory power. The prominence of production intensity and operating hours confirms that consumption is governed chiefly by how hard and how long facilities run, while the substantial weight of equipment efficiency signals a controllable lever with direct managerial relevance. The non-trivial contribution of ambient temperature reflects the sensitivity of heating, cooling, and process loads to external conditions.

Table 9

Normalized feature importance for energy-consumption prediction (XGBoost).

Feature	Importance (%)	Interpretation
Production intensity	31.4	Throughput rate is the primary consumption driver.
Operating hours	22.7	Longer runtimes scale total energy use.
Equipment efficiency	18.5	Controllable lever; better equipment lowers use.
Ambient temperature	11.2	Heating/cooling loads respond to weather.
Production output	6.4	Volume effect partly captured by intensity.
Fuel consumption	4.8	Co-varies with thermal load.
Electricity cost	3.2	Proxy for load; weak independent signal.
Other (residual)	1.8	Minor unattributed contributions.

Facilities were segmented into performance quartiles based on their combined energy-efficiency and CO_2-intensity ranking (Table 10). The gap between the top and bottom quartiles is substantial: leading facilities were roughly 60% more energy-efficient and 25% less carbon-intensive than laggards, while operating at comparable output. This dispersion quantifies the improvement headroom available within the existing fleet and provides realistic, peer-based benchmarks for the optimization stage rather than purely theoretical targets.

Table 10

Facility performance quartiles (combined efficiency ranking).

Quartile	Facilities	Energy eff. (t/MWh)	CO_2 int. (tCO_2/MWh)	Equip. eff. (%)
Q1 (top)	31	10.9	0.61	89.2
Q2	31	8.3	0.71	81.0
Q3	32	6.9	0.77	75.4
Q4 (bottom)	31	5.1	0.86	67.8

The optimization stage searched for the combination of equipment-efficiency improvement, production scheduling, and energy-demand management that minimized energy and emissions while holding production within one percent of baseline. Table 11 contrasts the baseline (fleet-average) state with the optimized state. Total energy consumption fell by 13.6% (from 4,820 to 4,165 MWh per facility-year) and CO_2 emissions fell by 15.2% (from 3,566 to 3,024 tCO_2 per facility-year), while production output was essentially preserved (−0.6%). Because emissions declined proportionally more than energy, CO_2 intensity also improved, from 0.74 to 0.71 tCO_2/MWh, indicating a cleaner as well as a leaner operating point. The associated reduction in electricity cost was approximately 12.9%.

Table 11

Baseline versus optimized operating state (per facility-year).

Indicator	Baseline	Optimized	Change (%)
Energy consumption (MWh)	4,820	4,165	−13.6
CO_2 emissions (tCO_2)	3,566	3,024	−15.2
CO_2 intensity (tCO_2/MWh)	0.74	0.71	−4.1
Production output (tonnes)	37,800	37,570	−0.6
Energy efficiency (t/MWh)	7.84	9.02	+15.1
Electricity cost (10^3 USD)	583.2	508.0	−12.9

Table 12 decomposes the total energy and emissions savings by intervention. Improved equipment efficiency contributed the largest single share of energy savings (6.2 percentage points), consistent with its high feature importance, followed by production scheduling (4.1 points) and energy-demand management (3.3 points). For emissions, the ranking was similar but demand management contributed relatively more because peak-load trimming displaces the most carbon-intensive marginal generation. The complementary nature of the three levers explains why the combined programme achieves more than any single measure applied in isolation.

Table 12

Decomposition of savings by intervention.

Intervention	Energy Δ (pp)	CO_2 Δ (pp)	Relative effort
Improved equipment efficiency	−6.2	−6.0	High (capital)
Production scheduling	−4.1	−4.5	Low (operational)
Energy-demand management	−3.3	−4.7	Medium (controls)
Combined programme	−13.6	−15.2	—

pp = percentage points of the respective baseline total; components are near-additive with minor interaction.

To test the robustness of the optimization outcome, each intervention magnitude was varied by ±25% around its optimal value while the others were held fixed. Table 13 reports the resulting range of total energy savings. The combined savings remained between 11.2% and 15.7% across all perturbations, and the production constraint was never violated, indicating that the recommended programme is stable and does not depend on precise fine-tuning of any single lever. Equipment efficiency exhibited the widest sensitivity band, reinforcing that it is both the most impactful and the most implementation-dependent measure.

Table 13

Sensitivity of total energy savings to ±25% variation in each lever.

Lever varied ±25%	Low case (%)	Optimal (%)	High case (%)
Equipment efficiency	−11.4	−13.6	−15.7
Production scheduling	−12.6	−13.6	−14.6
Energy-demand management	−12.8	−13.6	−14.4

4. Discussion and Conclusion

The present study developed and evaluated an integrated data-driven framework for assessing energy efficiency and environmental performance in industrial systems by combining descriptive analysis, machine-learning prediction, feature-importance assessment, and multi-objective optimization. The results demonstrated substantial heterogeneity in energy use and environmental performance across the industrial facilities. Average monthly energy consumption was approximately 401.7 MWh, equivalent to about 4,820 MWh per facility annually, while mean energy efficiency was 7.84 tonnes/MWh and mean CO_2 emission intensity was 0.74 tCO_2/MWh. The wide ranges observed across these indicators indicate that facilities operating at broadly comparable production scales may nevertheless differ considerably in their ability to convert energy inputs into productive output and in the carbon consequences of their operations. This finding reinforces the proposition that industrial sustainability should not be evaluated through total energy consumption alone. Rather, energy consumption needs to be interpreted in relation to production output, technological efficiency, fuel utilization, and environmental burden. This multidimensional perspective is consistent with life-cycle energy and environmental assessment approaches that emphasize the interdependence of energy demand and environmental consequences across production systems (Zhang et al., 2025). It is also compatible with the broader argument that improvements in energy efficiency constitute an important component of sustainable development alongside technological and environmental innovation (Feng & Xu, 2025).

The correlation analysis provided further insight into the operational structure underlying industrial energy consumption. Energy consumption showed its strongest positive relationship with production intensity ($r = 0.79$), followed by operating hours ($r = 0.71$), whereas equipment efficiency was negatively associated with energy consumption ($r = −0.58$). Fuel consumption was also positively related to energy demand, while CO_2 emissions were strongly associated with both energy consumption ($r = 0.84$) and fuel consumption ($r = 0.80$). These findings suggest that industrial energy demand is primarily determined by how intensively and how long production systems operate, but that technical efficiency can substantially moderate the amount of energy required to sustain a particular operational regime. Such patterns are compatible with thermodynamic approaches that locate efficiency losses in the quality and effectiveness of energy transformation rather than merely in gross energy input (Peng et al., 2025). Similarly, energy, exergy, entropy-generation, and exergoenvironmental frameworks emphasize that inefficiencies occurring within energy-conversion processes can simultaneously affect resource utilization and environmental impacts (Ordonez et al., 2022). The present results extend this logic into a data-driven facility-level framework by demonstrating that measurable operational characteristics can be used to quantify variation in industrial energy performance.

The analysis of production-scale categories produced another important result. Larger facilities tended to achieve somewhat higher energy efficiency, suggesting possible economies of scale, yet greater scale did not automatically correspond to lower CO_2 intensity. This distinction is important because it demonstrates that improvements in productive energy efficiency do not necessarily guarantee proportional environmental improvements. A facility may produce more output per unit of energy while still relying on a relatively carbon-intensive fuel configuration. Conversely, a comparatively low-carbon energy mix does not necessarily imply efficient energy conversion. The finding therefore supports the simultaneous consideration of energy and environmental indicators rather than treating one as an adequate proxy for the other. This interpretation is consistent with advanced exergy, exergoeconomic, and exergoenvironmental approaches that seek to integrate technical efficiency with economic and environmental consequences (Azubuike et al., 2025). The best-performing facilities in the present study were able to achieve both high energy efficiency and comparatively low CO_2 intensity, indicating that these objectives are not inherently incompatible. Instead, their joint achievement appears to depend on an appropriate combination of technology, fuel use, operating practices, and production management.

A major finding concerned the high predictive performance of the machine-learning models. For energy-consumption prediction, XGBoost achieved an R^2 of 0.91 with an RMSE of 0.18 and a MAPE of 4.2%, outperforming Random Forest and Support Vector Regression. The same ranking emerged for environmental performance, with XGBoost explaining approximately 89% of the variance in CO_2 emission intensity. Five-fold cross-validation further demonstrated that XGBoost maintained a mean R^2 of approximately 0.905 with very low between-fold variation, indicating that the model's strong performance was not simply an artifact of one favorable training/testing partition. These findings support the increasing use of machine-learning methods for modeling complex energy systems, where nonlinear relationships and interactions among operating variables may limit conventional statistical approaches. Reviews of machine learning and data-driven predictive analysis in power systems have similarly emphasized the growing role of computational methods in forecasting complex energy outcomes (Strielkowski et al., 2023). Likewise, comparisons involving decision trees, artificial neural networks, and conventional regression approaches have established machine learning as a useful methodological option for explaining electricity-consumption patterns (Nsangou et al., 2022).

The superior performance of XGBoost is particularly meaningful in light of the structure of industrial operating data. Production intensity, operating hours, equipment efficiency, fuel use, ambient temperature, and output can interact in nonlinear ways, making tree-based ensemble methods well suited to capturing conditional relationships among predictors. The present findings are methodologically consistent with the application of boosted ensemble regression trees to the simultaneous assessment of energy and environmental performance in building systems (Zaki et al., 2025). They also correspond with recent energy research employing optimized tree-based learning frameworks together with explainable artificial intelligence (Tang, 2026). More generally, advanced computational research increasingly integrates modeling, prediction, and optimization to address engineering systems characterized by multiple interacting parameters (Krzywanski et al., 2024). The strong performance obtained in the present study therefore suggests that gradient-boosted decision-tree frameworks may provide a practical basis for industrial energy analytics, especially when organizations possess structured operational datasets but lack sufficiently detailed physical models for every process and piece of equipment.

Nevertheless, predictive accuracy alone would have limited managerial value if the trained models operated entirely as black boxes. The feature-importance analysis therefore represents an important contribution of the present framework. Production intensity accounted for 31.4% of predictive importance, operating hours for 22.7%, equipment efficiency for 18.5%, and ambient temperature for 11.2%; together, these four variables represented 83.8%

of the model's explanatory importance. This result indicates that industrial consumption is concentrated around a relatively small set of influential drivers, some of which are directly or indirectly controllable. The dominance of production intensity and operating hours is understandable because increased throughput and longer operating periods naturally increase aggregate energy demand. More important from an intervention perspective, equipment efficiency emerged as the third most influential factor, identifying a technically controllable pathway for reducing consumption. This emphasis is consistent with research placing machine-learning models at the center of energy optimization and resource-consumption management in smart factories (Nozari et al., 2025). It also reflects the increasing emphasis on explainability in energy applications, where knowing which variables drive a prediction is essential for translating machine-learning output into operational action (Tang, 2026).

The contribution of ambient temperature to predictive performance also deserves attention. Although its importance was smaller than that of production intensity, operating hours, and equipment efficiency, it accounted for 11.2% of the model's predictive importance. Industrial facilities may experience temperature-related fluctuations in heating, cooling, ventilation, refrigeration, and process-energy requirements, and environmental conditions may therefore influence energy demand even when production parameters are relatively stable. The inclusion of ambient temperature in the current framework is methodologically aligned with research specifically examining the impact of environmental conditions on renewable-energy prediction using tree-based ensemble learning (Doğan et al., 2025). This finding implies that energy-management systems should avoid interpreting variations in consumption exclusively as indicators of managerial or technological inefficiency. Part of the observed variability may reflect environmental exposure, and operational benchmarking should therefore account for external conditions when comparing performance across periods or facilities.

The facility-tier analysis reinforced the practical significance of equipment efficiency and operational management. Facilities in the highest-performing quartile achieved an average energy efficiency of approximately 10.9 tonnes/MWh and CO_2 intensity of 0.61 tCO_2/MWh, whereas facilities in the lowest quartile achieved only 5.1 tonnes/MWh with a substantially higher CO_2 intensity of 0.86 tCO_2/MWh. The highest-performing facilities also had markedly higher equipment efficiency. These differences indicate considerable improvement potential within the existing facility population without necessarily requiring fundamental changes to productive activity. Equipment condition, maintenance, retrofitting, and technological upgrading may therefore constitute important components of a comprehensive energy strategy. The broader sustainability literature has emphasized that equipment durability and corrosion protection are intrinsically connected to sustainability because maintaining asset performance and extending service life can reduce material replacement, resource consumption, and environmental burdens (Prošek et al., 2025). In this respect, maintaining equipment efficiency should be viewed not merely as a technical maintenance objective but as part of an integrated energy and sustainability strategy.

The multi-objective optimization findings represent the most directly actionable component of the study. The optimized combination of improved equipment efficiency, production scheduling, and energy-demand management reduced annual energy consumption by 13.6%, from approximately 4,820 to 4,165 MWh per facility, while CO_2 emissions declined by 15.2%. Importantly, these reductions were achieved while production output declined by only 0.6%, thereby satisfying the study's production-maintenance constraint. Energy efficiency consequently improved by approximately 15.1%, CO_2 intensity decreased by 4.1%, and electricity expenditure declined by approximately 12.9%. These findings demonstrate why energy optimization should be formulated as a constrained multi-objective problem rather than as simple consumption minimization. An apparently successful reduction in energy use would be of limited managerial relevance if it were primarily generated through substantial reductions in output. The ability of the optimization procedure to preserve production while improving energy and environmental indicators is consistent with emerging applications of machine learning for the joint prediction and optimization of complex operating systems (Long et al., 2025). It also corresponds with integrated thermodynamic and environmental approaches that seek simultaneous improvements across technical and environmental criteria (Azubuike et al., 2025; Ordonez et al., 2022).

The decomposition of optimization benefits provided additional support for this interpretation. Improvements in equipment efficiency produced the largest individual contribution to energy savings, accounting for approximately 6.2 percentage points of the overall reduction, followed by production scheduling and demand management. However, demand management contributed relatively strongly to CO_2 reduction, indicating that interventions may differ in their effects on energy and environmental outcomes. The combination of the three measures was therefore more informative than reliance on any single lever. Furthermore, sensitivity analysis indicated that total energy savings remained substantial when intervention magnitudes were varied around their optimized values, suggesting that the proposed strategy was not dependent on an unrealistically precise set of operating conditions. This is important for industrial implementation, where operational uncertainty and day-to-day variability inevitably constrain the precision with which recommended settings can be maintained. The result supports the broader movement toward computational frameworks in which prediction and optimization are coupled rather than implemented as independent analytical tasks (Krzywanski et al., 2024; Long et al., 2025).

The environmental importance of these improvements should also be understood within the wider context of sustainable production and societal demand for environmentally responsible economic activity. Research outside the industrial-energy domain has shown that awareness of environmentally preferable production attributes can shape consumer perceptions (Haghjou et al., 2020), while willingness to pay premiums for pesticide-free products illustrates that reduced environmental burdens can acquire measurable economic value in market settings (Hayati et al., 2012). Although those studies concern agricultural consumption rather than industrial energy management, they underscore the broader principle that environmental performance increasingly interacts with economic decision-making and stakeholder expectations. Industrial organizations that reduce energy use and carbon intensity may therefore obtain benefits extending beyond direct utility-cost savings, including stronger environmental positioning and improved responsiveness to sustainability expectations. At the same time, the growing application of computational tools should itself be pursued efficiently, because research on sustainable artificial intelligence has drawn attention to the energy requirements associated with computational algorithms, including decision-tree approaches (Ferro et al., 2023). Data-driven industrial sustainability should therefore seek proportional analytical complexity: models should provide sufficient predictive and explanatory value without generating unnecessary computational demand.

Overall, the findings demonstrate the value of linking descriptive energy assessment, machine-learning prediction, interpretable feature analysis, and multi-objective optimization within one analytical pipeline. The contribution of the framework lies not simply in the relatively high predictive accuracy of XGBoost, but in the sequence through which prediction is translated into decision-relevant knowledge. The analysis first identifies heterogeneity in facility performance, then predicts energy and environmental indicators, determines which variables exert the greatest influence, and ultimately evaluates combinations of interventions capable of improving both energy efficiency and environmental performance while preserving production. This integrated structure responds to the increasing emphasis on machine learning for smart-factory energy optimization (Nozari et al., 2025), data-driven predictive analysis in energy systems (Strielkowski et al., 2023), explainable tree-based energy modeling (Tang, 2026), and computational integration of prediction and optimization (Krzywanski et al., 2024). Taken together, the findings suggest that industrial energy management can move beyond periodic diagnosis toward a more predictive, interpretable, and prescriptive model of operational sustainability.

Several limitations should be considered when interpreting the findings. Most importantly, the dataset was synthetic and was constructed to reproduce realistic distributions and relationships rather than being obtained directly from operating industrial plants. Consequently, the results demonstrate the methodological capability of the framework but should not be interpreted as estimates that will necessarily be reproduced in a particular industrial sector or facility. The dataset also did not capture all sources of real-world operational complexity, including sensor failure, missing records, sudden equipment malfunction, unplanned shutdowns, fuel switching, maintenance cycles, technological heterogeneity, and changes in product mix. Environmental performance was represented primarily through CO_2 emission intensity and therefore did not encompass other pollutants, water consumption, material use, waste generation, or complete life-cycle environmental impacts. In addition, the optimization model considered a restricted set of intervention variables and did not explicitly incorporate capital investment requirements, payback periods, financing constraints, workforce considerations, or detailed production-process restrictions. Finally, feature importance indicates predictive contribution rather than causal effects, and the identified drivers should therefore not automatically be interpreted as causal determinants without further experimental or longitudinal evidence.

Future research should validate the proposed framework using longitudinal operational datasets obtained from real industrial facilities representing different sectors, production technologies, geographic regions, and energy mixes. Studies could examine whether the predictive superiority observed for XGBoost is maintained under conditions involving noisy sensor data, incomplete observations, structural changes, and nonstationary operating regimes. Future models should also incorporate additional variables, including machine age, maintenance history, equipment load, electricity tariff structure, renewable-energy penetration, fuel composition, humidity, production type, and shift schedules. Environmental assessment could be broadened to include additional greenhouse gases, air pollutants, water consumption, waste, embodied energy, and life-cycle indicators. More sophisticated explainable-AI methods could be employed to determine how specific predictors influence individual facility predictions and whether interactions between operational variables differ across efficiency tiers. Future optimization models should explicitly integrate capital costs, maintenance expenditures, carbon prices, uncertainty, production deadlines, and equipment constraints. Finally, prospective studies could compare recommendations generated by the framework with actual interventions implemented by industrial managers to determine whether predicted savings are realized under operational conditions.

Industrial managers can use the findings to establish continuous data-driven energy-management systems rather than relying solely on periodic audits. Priority should be given to systematic monitoring of production intensity, operating hours, equipment efficiency, ambient conditions, fuel consumption, energy use, and emissions because these variables can provide an operational picture of both efficiency and environmental performance. Facilities should benchmark their performance against comparable high-efficiency operations and investigate substantial deviations rather than depending only on aggregate consumption targets. Equipment maintenance and efficiency improvement should receive particular attention because equipment performance emerged as an important controllable energy lever, while production scheduling can provide a comparatively low-cost route to improvement. Demand-management measures should be coordinated with scheduling and equipment interventions rather than implemented independently. Managers should also evaluate energy and carbon indicators jointly so that reductions in energy demand do not conceal unfavorable changes in fuel composition or emissions intensity. Where sufficient operational data are available, predictive models can be integrated into energy dashboards and decision-support systems to provide early identification of inefficient operating conditions and to test alternative scenarios before operational changes are implemented.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

In this research, ethical standards including obtaining informed consent, ensuring privacy and confidentiality were considered.

References

Azubuike, U. G., Njoku, H. O., Ekechukwu, O. V., & Jen, T. C. (2025). A Comprehensive Review on Advanced Exergy, Exergoeconomic, and Exergoenvironmental Analyses for Sustainable Thermal Energy Systems. *International Journal of Exergy*, *48*(1), 16-53. https://doi.org/10.1504/IJEX.2025.148556

Doğan, F., Oyucu, S., Unsal, D. B., Aksöz, A., & Vafaeipour, M. (2025). Impact of Environmental Conditions on Renewable Energy Prediction: An Investigation Through Tree-Based

Community Learning. *Applied Sciences*, *15*(1), 336. https://doi.org/10.3390/app15010336
Feng, Y., & Xu, R. (2025). Advancing Global Sustainability: The Role of the Sharing Economy, Environmental Patents, and Energy Efficiency in the Group of Seven's Path to Sustainable Development. *Sustainability*, *17*(1), 322. https://doi.org/10.3390/su17010322
Ferro, M., Silva, G. D., de Paula, F. B., Vieira, V., & Schulze, B. (2023). Towards a Sustainable Artificial Intelligence: A Case Study of Energy Efficiency in Decision Tree Algorithms. *Concurrency and Computation: Practice and Experience*, *35*(17), e6815. https://doi.org/10.1002/cpe.6815
Haghjou, M., Hayati, B., & Pishbahar, E. (2020). Factors Affecting Consumers' Awareness of Pesticides-Free Fruits and Vegetables. In *The Economics of Agriculture and Natural Resources: The Case of Iran* (pp. 125-139). Springer Singapore. https://doi.org/10.1007/978-981-15-5250-2_9
Hayati, B., Pishbahar, E., & Haghjou, M. (2012). Analyzing Determinants of Consumers' Willingness to Pay a Premium for Pesticide-Free Fruit and Vegetables in Marand City.
Krzywanski, J., Sosnowski, M., Grabowska, K., Zylka, A., Lasek, L., & Kijo-Kleczkowska, A. (2024). Advanced Computational Methods for Modeling, Prediction and Optimization—A Review. *Materials*, *17*(14), 3521. https://doi.org/10.3390/ma17143521
Long, F., Xu, M., Liao, W., & Liu, H. (2025). Machine Learning for Predicting and Optimizing the Performance of a Commercial-Scale Anaerobic Digester with Diverse Feedstocks and Operating Conditions. *Bioresource Technology*, *435*, 132940. https://doi.org/10.1016/j.biortech.2025.132940
Nozari, H., Szmelter-Jarosz, A., & Samadi, S. (2025). Machine Learning Models for Energy Optimization and Resource Consumption in Smart Factories. In *Artificial Intelligence of Everything and Sustainable Development* (pp. 175-189). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-7202-8_10
Nsangou, J. C., Kenfack, J., Nzotcha, U., Ekam, P. S. N., Voufo, J., & Tamo, T. T. (2022). Explaining Household Electricity Consumption Using Quantile Regression, Decision Tree and Artificial Neural Network. *Energy*, *250*, 123856. https://doi.org/10.1016/j.energy.2022.123856
Ordonez, J. C., Cavalcanti, E. J., & Carvalho, M. (2022). Energy, Exergy, Entropy Generation Minimization, and Exergoenvironmental Analyses of Energy Systems—A Mini-Review. *Frontiers in Sustainability*, *3*, 902071. https://doi.org/10.3389/frsus.2022.902071
Peng, M., Tian, Z., Xia, G., & Wang, H. (2025). The Exergy Analysis Method of Thermodynamic Process. In *Thermal Analysis of Nuclear Power Plants* (pp. 71-126). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-8983-5_3
Prošek, T., Keil, P., & Popova, K. (2025). Corrosion Protection and Sustainability: Why Are the Two Concepts Inherently Intertwined. *Corrosion and Materials Degradation*, *6*(3), 38. https://doi.org/10.3390/cmd6030038
Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and Challenges of the Machine Learning and Data-Driven Methods for the Predictive Analysis of Power Systems: A Review. *Energies*, *16*(10), 4025. https://doi.org/10.3390/en16104025
Tang, T. (2026). Predicting Energy Prices and Renewable Energy Adoption Through an Optimized Tree-Based Learning Framework with Explainable Artificial Intelligence. *Scientific reports*, *16*(1), 6771. https://doi.org/10.1038/s41598-026-35706-z
Zaki, A. M., Zayed, M. E., Bargal, M. H., Saif, A. G. H., Chen, H., Rehman, S., & El-deen, E. S. H. N. (2025). Environmental and Energy Performance Analyses of HVAC Systems in Office Buildings Using Boosted Ensembled Regression Trees: Machine Learning Strategy for Energy Saving of Air Conditioning and Lighting Facilities. *Process Safety and Environmental Protection*, *198*, 107214. https://doi.org/10.1016/j.psep.2025.107214
Zhang, R., Yin, K., Wei, R., Ruan, J., Yang, J., Wang, S., & Wang, Y. (2025). Comprehensive Analysis of Life Cycle Energy Consumption and Environmental Impact of Hydrogen Production Process via Plasma Co-Gasification of Coal and Biomass. *Energy*, *324*, 135976. https://doi.org/10.1016/j.energy.2025.135976