Management Strategies and Engineering Sciences 2026; 8(2):1-19

Identification and Analysis of Factors Affecting Industrial Machinery Maintenance and Repair Costs Using the XGBoost Algorithm

Khalil Jaidari[1], Sadegh Feizollahi [2*] Alireza Irajpour [1]

[1] Department of Management, Qa.C., Islamic Azad University, Qazvin, Iran
[2] Department of Management, Il.C., Islamic Azad University, Ilam, Iran

*** Corresponding author email address**: sadegh.feizollahi@iau.ac.ir

Received: 2025-11-02 **Revised:** 2026-03-13 **Accepted:** 2026-03-20 **Published:** 2026-03-30

Abstract

The primary objective of this study was to scientifically identify and rank the factors affecting industrial machinery maintenance and repair costs in Iran's petrochemical industry. To achieve this objective, a mixed-methods approach comprising a systematic literature review, semi-structured interviews with 12 industry experts, and advanced statistical analyses was employed. The required quantitative data were collected from 10 industrial machines over a three-year period. During the analytical phase, statistical methods, including correlation coefficients and analysis of variance (ANOVA), were used to assess the relationships among the variables. Subsequently, the advanced XGBoost machine-learning algorithm and the Gain metric were employed to calculate the relative importance of the factors and establish their final ranking. The findings indicated that three factors—equipment age (18.9%), failure to implement preventive maintenance (17.2%), and adverse operating conditions (15.1%)—together accounted for more than 50% of the total importance. Managerial factors, including spare-parts procurement lead time (9.8%) and the level of CMMS utilization (8.5%), occupied the subsequent ranks.

Keywords: *Maintenance and repair costs, industrial machinery, petrochemical industry, XGBoost algorithm*

How to cite this article:

Jaidari, K., Feizollahi, S., & Irajpour, A. (2026). Identification and Analysis of Factors Affecting Industrial Machinery Maintenance and Repair Costs Using the XGBoost Algorithm. Management Strategies and Engineering Sciences, 8(2), 1-19.

1. Introduction

Maintenance and repair (M&R) activities constitute a critical component of industrial asset management because they directly influence equipment availability, production continuity, life-cycle cost, safety, and the economic performance of capital-intensive operations. In organizations that depend heavily on machinery, maintenance expenditure is not merely a technical cost center; rather, it is a managerial variable reflecting the quality of asset planning, resource allocation, operating discipline, spare-parts management, and long-term replacement strategies. Studies on machinery cost management have emphasized that maintenance costs should be evaluated alongside ownership, operating, downtime, and replacement costs if organizations are to achieve economically efficient asset utilization [1]. Technical and economic assessments of maintenance systems similarly indicate that the structure and quality of maintenance practices can substantially alter the total cost of machinery throughout its service life [2, 3]. In road-construction and civil-project settings, reductions in machinery maintenance expenditures have also been associated with improved planning, more systematic maintenance management, and better control of equipment utilization [4, 5]. These considerations make the identification of the principal cost drivers of M&R an important management problem, particularly in industries in which equipment failure can interrupt continuous production processes and generate cascading operational and financial losses.

The problem is especially important in the oil, gas, and petrochemical sectors, where rotating machinery, compressors, turbines, pumps, instrumentation systems, electrical equipment, and process-support systems operate under demanding conditions and are frequently exposed to high temperatures, vibration, corrosive media, hazardous atmospheres, and stringent safety requirements. Maintenance engineering in such industries therefore requires simultaneous coordination of technical reliability, workforce capability, inventory availability, regulatory requirements, and production targets [6]. Equipment obsolescence further increases this complexity because aging assets may exhibit declining reliability and efficiency, greater difficulty in obtaining original spare parts, and increased downtime risk [7]. Large-scale shutdown and turnaround projects create an additional financial burden because inadequate planning, delays, contractor dependency, procurement problems, and extended outages can translate into direct maintenance expenditures as well as lost production and business opportunities [8]. From a service-management perspective, the selection and evaluation of industrial maintenance services are also affected by multiple technical, economic, and organizational criteria, underscoring the multidimensional nature of maintenance decision-making [9]. Risk-based analysis likewise indicates that machinery repair decisions must consider not only the immediate failure event but also the broader operational consequences and uncertainties associated with repair activities [10].

A central challenge in controlling M&R costs is that the underlying cost drivers are heterogeneous and frequently interdependent. Equipment age, design complexity, loading patterns, operating temperature, vibration, environmental severity, maintenance history, spare-parts condition, and human intervention may act jointly rather than independently. Research on vibration monitoring demonstrates that deterioration in rotating machinery can often be detected through changes in dynamic behavior before functional failure occurs, making condition information highly relevant to maintenance planning and cost avoidance [11]. Research on sustainable manufacturing similarly shows that equipment operating settings can be predicted and optimized through data-driven approaches, indicating that operating conditions are not merely background variables but potentially controllable determinants of machine performance and resource consumption [12]. In oil and gas operations, machine-learning techniques have been proposed to improve instrumentation accuracy and support more reliable condition assessment under complex field conditions [13], while AI-driven equipment management has been introduced as a means of improving predictive maintenance and reducing avoidable equipment failures and downtime [14]. In parallel, batteryless Industrial Internet of Things (IIoT) architectures employing thermoelectric energy harvesting and Narrowband Internet of Things (NB-IoT) technologies have been investigated for harsh oil, gas, and refinery environments, demonstrating how environmental conditions, sensing architecture, energy requirements, and maintenance needs can be jointly addressed [15, 16].

The maintenance strategy adopted by an organization represents another major determinant of cost. Reactive maintenance may initially appear less expensive because intervention is postponed until failure; however, it can increase unplanned downtime, emergency labor requirements, secondary equipment damage, and urgent spare-parts procurement. Preventive maintenance seeks to mitigate these risks through scheduled intervention, whereas condition-based and predictive maintenance attempt to determine the appropriate timing of intervention according to actual equipment condition and estimated failure probability. Recent reviews indicate that predictive maintenance has been implemented across diverse industries using sensor data, machine learning, and condition-monitoring techniques to improve reliability and reduce unnecessary maintenance interventions [17]. Empirical work on industrial machinery likewise supports the use of machine learning to detect degradation patterns and facilitate earlier and more informed maintenance decisions [18]. Gaussian process modeling has been applied to predictive maintenance of mining machinery to characterize equipment condition under uncertainty [19], while hybrid reliability modeling has been employed in nuclear power plant circulating-water systems to support predictive maintenance decisions in safety-critical infrastructure [20]. Machine-learning-based predictive maintenance systems have also been developed for artificial-yarn machinery, illustrating the transferability of data-driven maintenance approaches across different manufacturing environments [21].

Preventive maintenance, however, cannot be managed independently of inventory and supply-chain decisions. Maintenance interventions depend on the timely availability of spare parts, and an otherwise optimal maintenance schedule may lose its economic value if required components are unavailable, delayed, obsolete, or excessively expensive. Joint optimization models have therefore integrated preventive maintenance and spare-parts ordering, demonstrating that maintenance and inventory decisions should be coordinated rather than treated as separate management problems [22]. Related research has addressed the joint optimization of preventive maintenance and spare-parts ordering under imperfect detection, further demonstrating the economic interdependence among equipment condition, maintenance timing, inspection effectiveness, and inventory policy [23]. Case-based research on spare-parts manufacturing strategy also indicates that local production, alternative sourcing, and prioritization of critical components may become strategically important when conventional supply channels are costly or unreliable [24]. These issues are particularly relevant to petrochemical operations, in which procurement lead times, original equipment manufacturer dependence, supply restrictions, foreign-exchange conditions, and inventory policies may collectively influence the final cost of maintenance.

The financial dimension of maintenance is equally shaped by broader cost-management and organizational practices. Evidence from manufacturing indicates that systematic cost-reduction strategies involve the integration of operational efficiency, resource control, process improvement, and managerial accounting rather than isolated reductions in individual expenditure categories [25]. In machinery-intensive operations, attempts to reduce maintenance expenditure should therefore avoid the false economy of under-maintenance, whereby short-term savings can eventually lead to higher failure frequencies, longer downtime, and greater long-term losses. Quantitative modeling of machine-repair systems has demonstrated that maintenance cost optimization can involve queueing, retrial, admission-control, and feedback mechanisms, illustrating that maintenance costs may exhibit nonlinear behavior arising from interacting operational policies [26]. At the project level, machine-learning techniques have also been reviewed as tools for estimating and forecasting construction costs, demonstrating the broader managerial value of data-driven prediction in situations where conventional linear assumptions may be insufficient [27]. Similarly, big-data analytics and artificial-intelligence approaches have been proposed for forecasting project schedules and costs, reinforcing the expanding role of advanced analytics in managerial planning and control [28].

The growing availability of operational data has consequently transformed the methodological landscape of maintenance-cost analysis. Traditional approaches based on descriptive statistics, expert judgment, maintenance records, and isolated engineering indicators remain valuable, but they may be insufficient when cost outcomes emerge from complex interactions, threshold effects, and nonlinear relationships among several technical and managerial variables. Machine-learning methods can complement conventional statistical analysis by learning such structures directly from historical data and by identifying variables that contribute most strongly to predictive performance. A systematic review of prognostics and health management for industrial mechanical systems indicates that machine learning has become an important component of fault diagnosis, remaining-useful-life estimation, health monitoring, and predictive maintenance because of its ability to process multidimensional equipment-condition data [29]. Comparisons of machine-learning approaches for predictive maintenance with limited datasets further suggest that algorithm selection, data characteristics, and validation procedures can materially affect predictive performance [30]. In oil-platform applications, artificial intelligence and machine learning have also been employed to predict equipment failures, illustrating the relevance of such methods to the operational realities of petroleum-sector assets [31].

Data-driven prediction has expanded beyond equipment failure classification to encompass the estimation of maintenance and repair costs themselves. Deep-learning approaches based on long short-term memory (LSTM) architectures have been applied to estimate repair and maintenance costs in apartment buildings, demonstrating that historical and temporal data can support forward-looking cost estimation [32]. Artificial neural networks and computer-vision techniques have likewise been employed to estimate pothole repair costs from digital images, showing that maintenance costs can be linked to automatically extracted information concerning the physical condition and extent of deterioration of an asset [33]. Quantitative models have also been developed for estimating vehicle repair costs in insurance claims, where uncertainty, damage characteristics, and financial valuation must be incorporated into a predictive framework [34]. These studies are significant because they demonstrate that repair cost is not inherently restricted to traditional accounting estimation; rather, it can be modeled as an outcome associated with measurable technical, operational, and physical characteristics. In industrial energy systems, machine-learning approaches have additionally been employed to predict gas-turbine transients, supporting the broader premise that complex equipment behavior can be represented through data-driven models and subsequently incorporated into maintenance planning and cost-control decisions [35].

From a management perspective, however, prediction alone is insufficient. Maintenance managers require interpretable evidence regarding which variables exert the greatest influence on costs so that limited financial, technical, and human resources can be directed toward the interventions with the greatest potential impact. A model that accurately predicts maintenance expenditure but fails to distinguish the relative importance of equipment age, preventive-maintenance implementation, operating conditions, procurement delays, computerized maintenance management system (CMMS) utilization, workforce capacity, environmental severity, or budget constraints offers limited guidance for strategic prioritization. Feature-ranking methods are therefore particularly relevant to maintenance management because they transform predictive analysis into actionable managerial information. In this context, advanced ensemble learning methods such as Extreme Gradient Boosting (XGBoost) provide an appropriate analytical framework because they can accommodate complex and nonlinear associations and generate measures of feature importance during predictive modeling. Such capabilities are especially relevant when maintenance-cost determinants are mutually reinforcing—for example, when aging equipment is exposed to adverse operating conditions, when preventive maintenance is constrained by delays in spare-parts procurement, or when inadequate information systems diminish the effectiveness of planning and scheduling.

This analytical requirement is also connected to the broader transition from traditional maintenance toward intelligent maintenance systems. In conventional maintenance environments, decisions have often depended heavily on fixed schedules, operator experience, and post-failure responses. By contrast, contemporary maintenance increasingly relies on continuous condition monitoring, sensor-generated information, predictive algorithms, integrated information systems, and automated decision support. Advances in predictive maintenance demonstrate the potential of machine learning to distinguish patterns associated with degradation and impending failure before major operational disruption occurs [17, 18]. Likewise, the integration of intelligent sensing with industrial communication technologies may reduce the maintenance burden associated with monitoring infrastructure itself and facilitate more continuous observation of assets located in hazardous or difficult-to-access environments [16]. Accordingly, the economic value of digital maintenance transformation depends not only on adopting new technologies but also on understanding which underlying cost drivers should receive priority in data collection, managerial attention, and intervention.

The Iranian petrochemical industry provides a particularly relevant context for such an investigation. Petrochemical assets operate under technically demanding conditions, while maintenance managers must simultaneously confront aging equipment, imported-component dependence, procurement uncertainty, exchange-rate volatility, workforce capability, production-continuity requirements, budget limitations, and health, safety, and environmental obligations. Research conducted in other machinery-intensive Iranian settings has emphasized the importance of systematic maintenance management and the economic control of machinery costs [1, 4, 5]. Nevertheless, identifying a broad collection of possible maintenance-cost factors is only the first stage of managerial analysis. Decision makers also need evidence concerning the relative contribution of these factors so that maintenance budgets and organizational improvement programs can be prioritized rationally. The increasing application of intelligent equipment monitoring and AI-assisted maintenance in oil and gas environments further suggests that petrochemical maintenance management can benefit from combining engineering knowledge with data-driven analytical methods [6, 14].

An important research gap therefore exists between the qualitative identification of maintenance challenges and the quantitative prioritization of the factors that actually explain variation in maintenance and repair costs. Much of the existing literature addresses particular components of this problem, including machinery cost management, equipment failure prediction, predictive maintenance, condition monitoring, spare-parts optimization, repair-cost estimation, obsolescence management, or maintenance-project risk. Although these streams provide valuable knowledge, maintenance expenditure in a petrochemical complex is the outcome of a system in which technical, operational, managerial, human, logistical, economic, and environmental factors coexist. Examining any single dimension independently may consequently provide an incomplete representation of the cost-generation process. An integrated methodology that first derives candidate factors from prior research, supplements them with context-specific knowledge from experienced practitioners, statistically evaluates their relationships with maintenance expenditure, and finally applies machine learning to rank their predictive importance can provide a more comprehensive basis for maintenance decision-making.

Such an approach also has important implications for strategic asset management. Reliable prioritization of maintenance-cost drivers can help management distinguish factors that require immediate operational intervention from those that should be addressed through longer-term investment and policy. For example, a high contribution of equipment age would strengthen the case for replacement and life-extension planning; a high contribution of failure to implement preventive maintenance would indicate the need to improve maintenance compliance and scheduling; a strong role for operating conditions would justify greater investment in monitoring and operational control; and substantial effects of spare-parts lead time or CMMS utilization would emphasize the importance of supply-chain and information-system capabilities. In this manner, feature importance is not merely a statistical output but a mechanism for translating operational data into managerial priorities. Combining expert knowledge, statistical evidence, and machine-learning-based ranking can thus contribute to a more transparent allocation of maintenance resources and support the transition from reactive cost control toward proactive and evidence-based maintenance management.

The aim of this study is to identify, validate, and rank the factors affecting industrial machinery maintenance and repair costs in Iran's petrochemical industry by integrating a systematic literature review, expert interviews, statistical analysis, and the XGBoost machine-learning algorithm.

2. Methodology

This study was conducted using a mixed-methods (qualitative–quantitative) approach and a descriptive-survey design. The required data were collected through two methods: library research and a systematic literature review to identify the preliminary factors, and semi-structured interviews with 12 industry experts selected through purposive sampling based on the criterion of having at least 15 years of professional experience in the petrochemical industry. To ensure the content validity of the interview questionnaire, expert opinions were obtained, while inter-coder agreement was used to assess reliability, yielding an agreement coefficient of 0.87. Data analysis was conducted at three levels: qualitative analysis of the interview data using thematic analysis; statistical analyses, including correlation coefficients and analysis of variance (ANOVA), using SPSS; and final ranking of the factors using the XGBoost algorithm in the Python environment.

3. Findings and Results

The identification of factors affecting maintenance and repair costs was conducted in two stages. In the first stage, an initial list of factors was extracted through library research and a systematic review of the relevant literature. Subsequently, semi-structured interviews with industry experts were conducted to identify context-specific factors particular to Iran's petrochemical industry. Finally, by integrating these two perspectives, 13 final factors were identified as influential factors. In the second stage, quantitative methods and machine-learning algorithms were employed to determine the priority and relative contribution of each factor. Correlation coefficients and analysis of variance were used to assess the statistical relationships between the factors and the dependent variable (maintenance and repair costs). Subsequently, the advanced XGBoost algorithm and the Gain metric were applied to calculate the relative importance of each factor and determine their final ranking. This comprehensive methodological approach enabled the objective identification and precise ranking of the influential factors.

To comprehensively identify the factors affecting maintenance and repair costs, a systematic literature review was conducted by examining reputable databases and scientific articles published during the preceding five years. The reviewed studies demonstrated that cost-driving factors could be classified into five principal categories: maintenance and management approaches, operating and environmental conditions, equipment characteristics, major maintenance projects, and cost-reduction strategies. The most important factors identified through the literature review, together with descriptions of their effects and corresponding sources, are presented in Table 1.

Table 1. Factors Affecting Maintenance and Repair Costs and Cost-Reduction Strategies: Literature Review

No.	Main Category	Specific Factor / Strategy	Description of Effect / Mechanism	Source
1	Maintenance and management approaches	Reactive maintenance	Increases unplanned downtime and long-term costs	(Kolo, 2025)

		Lack of systematic predictive methods (TPM)	Reduces reliability and operational efficiency	(Musa, 2025)
		Inefficient inventory management and cataloging	Disrupts asset tracking and increases costs	(Musa, 2025)
		Lack of maintenance work orders and reports	Creates difficulties in planning and performance evaluation	(Musa, 2025)
2	Operating and environmental conditions	High-temperature environments and explosive atmospheres	Shortens the service life of conventional equipment and increases maintenance costs	(Aragonés, 2025)
		Environmental conditions, deterioration, and human error	Reduces instrumentation accuracy and increases maintenance requirements	(Jambol, 2024b)
3	Equipment characteristics	Asset complexity and sensitivity	Makes maintenance management more challenging	(Jambol, 2024)
		Equipment obsolescence	Reduces efficiency and reliability and results in spare-parts shortages	(Kolo, 2025)
		Specific parameters (vibration, temperature, power factor)	Directly affect the maintenance requirements of induction motors	(Zainol, 2022)
4	Major maintenance projects	Turnaround/shutdown maintenance	Results in production losses, lost business opportunities, and HSE-related costs	(Dhanaraj, 2024)
		Hiring external contractors	Increases labor costs during intensive maintenance periods	(Dhanaraj, 2024)
5	Cost-reduction strategies	Thermoelectric energy harvesting	Eliminates batteries and reduces maintenance costs of IIoT devices	(Aragonés, 2025)
		Artificial intelligence for predictive maintenance	Reduces downtime and costs and improves reliability	(Jambol, 2024)
		Machine learning for instrumentation accuracy	Enables predictive maintenance and prevents failures	(Jambol, 2024b)
		Computerized Maintenance Management Systems (CMMS)	Automate tasks, facilitate regulatory compliance, and support performance-indicator management	(Musa, 2025)
		Proactive obsolescence management	Reduces operational risk and life-cycle costs	(Kolo, 2025)
		Condition-Based Maintenance (CBM)	Prevents downtime and improves productivity	(Abbasi, 2020)

To validate and supplement the factors extracted from the literature, identify context-specific challenges, and prioritize the factors from an operational perspective, semi-structured interviews were conducted with 12 experts. The selection criterion was a minimum of 15 years of specialized professional experience in maintenance and repair within the oil, gas, and petrochemical industries, thereby ensuring that the perspectives obtained were grounded in extensive and credible experience. These interviews were designed according to a qualitative approach, with the key questions focusing on identifying cost-driving factors, evaluating the preliminary list of factors, determining the most critical categories, and examining barriers to implementing modern maintenance solutions. The questions were formulated as presented in the following table to maintain consistency across interviews and collect reliable data for subsequent analyses.

Table 2. Questions Asked During the Expert Interviews

No.	Interview Question
1	In your opinion, which factors contribute most substantially to increased maintenance and repair costs in your complex?
2	Do you agree with the preliminary list of factors extracted from the literature? Do you believe any important factor has been omitted?
3	Among technical, human, managerial, and environmental factors, which category is more critical?
4	What is the greatest challenge in implementing modern approaches such as predictive maintenance? For example, resistance to change, budget constraints, or difficulties in technology procurement.

Table 3. Demographic Characteristics of the Interviewees (12 Experts)

No.	Interviewee Code	Work Experience (Years)	Education Level	Field of Study	Area of Activity / Expertise
1	K-01	23	Master’s degree	Mechanical Engineering	Overall maintenance management
2	K-02	21	Bachelor’s degree	Mechanical Engineering	Turbine and compressor maintenance
3	K-03	19	Master’s degree	Industrial Engineering	Turnaround project planning
4	K-04	20	Master’s degree	Mechanical Engineering	Condition analysis and monitoring
5	K-05	18	Master’s degree	Electrical Engineering	Control systems and instrumentation
6	K-06	17	Master’s degree	Industrial Engineering	CMMS and scheduling

7	K-07	23	PhD	Chemical Engineering	Operations and maintenance supervision
8	K-08	22	Master's degree	Safety Engineering	Maintenance risk and safety management
9	K-09	21	Bachelor's degree	Electrical Engineering	High-voltage electrical equipment maintenance
10	K-10	20	Master's degree	Industrial Engineering	Spare-parts supply chain management
11	K-11	23	Bachelor's degree	Petroleum Engineering	Field operations supervision
12	K-12	19	Master's degree	Mechanical Engineering	Heat exchanger maintenance

Table 4. Results of the Expert Interviews: Coding of Interview Data

No.	Main Theme	Key Factor Extracted from Interviews	Description and Rationale Expressed by Experts
1	Managerial and strategic challenges	Lack of a long-term strategic maintenance plan	"Our maintenance activities are always in firefighting mode. There is no five-year roadmap for equipment upgrading and personnel training."
		Inefficient turnaround maintenance project management cycle	"The planning, procurement, and execution processes for turnarounds are extremely lengthy and problematic, resulting in exorbitant costs."
2	Resource and logistics constraints	Sanctions and difficulties in procuring original spare parts	"Access to OEM parts is extremely difficult and expensive. We are forced to use non-original parts, which have shorter service lives and fail again."
		Shortage of specialized and experienced personnel	"The older generation is retiring, while younger personnel lack sufficient experience and motivation to work in demanding environments."
3	Organizational culture and human factors	Resistance to change and new technologies	"To introduce a new CMMS, we have to spend months convincing personnel that the system is useful. A 'if it works, it is good enough' culture prevails."
		Declining organizational belonging and accountability	"Workers and engineers do not feel that they have a stake in the company's success. As a result, operational diligence declines."
4	Economic–political factors	Exchange-rate fluctuations and uncontrolled inflation	"The maintenance budget approved at the beginning of the year is no longer sufficient by the end of the year because of inflation."
5	Solutions proposed by experts	Implementation of an Integrated Maintenance Management System (IMS)	"We need an integrated software system that simultaneously handles planning, procurement, inventory, execution, and maintenance-data analysis."
		Emphasis on personnel training and empowerment	"Instead of purchasing new equipment, we should first invest in workforce training. A trained employee can be more valuable than an advanced machine."
		Prioritization and phased implementation of predictive maintenance	"CBM should not be implemented across all equipment overnight. We should start with critical equipment and demonstrate the results to senior management."

Following the identification of factors through the literature review and the extraction of industry experts' perspectives, the process of integrating and consolidating the factors was initiated. In this process, overlapping factors identified in both the literature and the interviews were merged, while new and context-specific factors emerging from the interviews were added to the final list. This approach resulted in a comprehensive set of factors affecting maintenance and repair costs that is both scientifically grounded and consistent with the specific conditions of Iran's petrochemical industry. Table 5 presents the final factors resulting from this integration.

Table 5. Final Factors Affecting Maintenance and Repair Costs: Integration of the Literature Review and Expert Interviews

No.	Category	Factor
1	Technical–equipment-related	Equipment deterioration and aging (equipment age)
		Technical complexity of equipment
		Adverse operating conditions (temperature, humidity, vibration)
		Failure to implement preventive maintenance
2	Managerial	Lack of long-term strategic planning
		Inefficient supply chain and inventory management
		Lack of implementation of maintenance management systems (CMMS)
		Inefficient turnaround maintenance project management cycle
3	Human	Shortage of specialized and experienced personnel
		Resistance to change and new technologies
		Declining accountability and organizational commitment
4	Economic–political	Sanctions and restrictions on spare-parts procurement
		Exchange-rate fluctuations and inflation
		Budgetary constraints
5	Environmental	Corrosive and explosive environmental conditions
		Health, Safety, and Environment (HSE) considerations

Table 6. Correlation Coefficient Matrix and Analysis of Variance Results

No.	Variable	Variable Type	Correlation with Cost	ANOVA *p*-value	Retained
1	Equipment age	Quantitative	0.81	0.0001	Yes
2	Technical complexity of equipment	Qualitative (categorical)	—	0.0001	Yes
3	Adverse operating conditions	Quantitative	0.79	0.0001	Yes
4	Failure to implement preventive maintenance	Quantitative	−0.76	0.0001	Yes
5	Lack of strategic planning	Qualitative (subjective)	Not computable	Not applicable	No—excluded
6	Inefficient supply chain management	Quantitative	0.58	0.0001	Yes
7	Lack of CMMS implementation	Qualitative (categorical)	—	0.0001	Yes
8	Inefficient turnaround maintenance cycle	Quantitative	0.62	0.0001	Yes
9	Shortage of specialized personnel	Quantitative	0.53	0.0001	Yes
10	Resistance to change	Qualitative (subjective)	Not computable	Not applicable	No—excluded
11	Declining accountability	Qualitative (subjective)	Not computable	Not applicable	No—excluded
12	Sanctions and supply restrictions	Quantitative	0.58	0.0001	Yes
13	Exchange-rate fluctuations and inflation	Quantitative	0.42	0.0001	Yes
14	Budgetary constraints	Quantitative	0.48	0.0001	Yes
15	Corrosive environmental conditions	Quantitative	0.45	0.0001	Yes
16	HSE considerations	Quantitative	0.45	0.0001	Yes

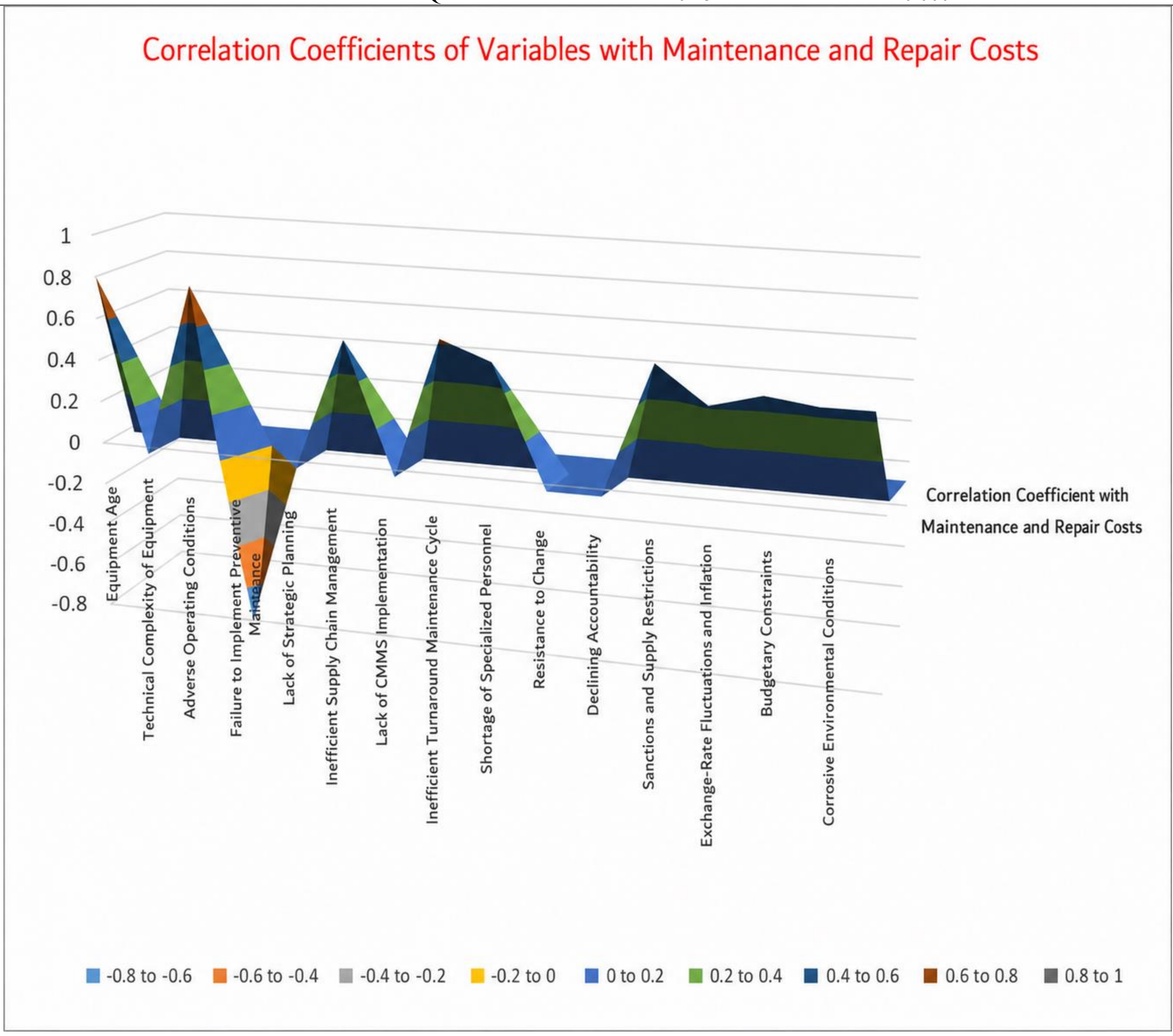

Figure 1. Correlation Coefficients of Variables with Maintenance and Repair Costs

Based on the findings obtained from the statistical analyses, three factors—"lack of strategic planning," "resistance to change," and "declining accountability"—were excluded from the final set of factors. The primary reason for excluding these factors was the fundamental challenges associated with their operationalization and measurability. Specifically, these factors could not be directly quantified, were based on subjective measures, and lacked reliable historical data for measurement. These limitations made it impossible to determine an objective and independent contribution for these variables in the modeling process.

In contrast, the remaining 13 factors were selected for inclusion in the final modeling stage. Of these, 10 quantitative variables demonstrated significant correlation coefficients ranging from 0.42 to 0.81, with significance levels below .001. In addition, three categorical qualitative variables were assessed using analysis of variance, all of which exhibited statistically significant *p*-values below .001. This final set provides a balanced combination of objective and measurable indicators, thereby enabling the scientifically rigorous and practically applicable implementation of machine-learning algorithms for ranking purposes.

In the subsequent stage of the study, and to prepare the data for final ranking using machine-learning algorithms, it was necessary to examine the interrelationships among the key variables demonstrating the strongest correlations with maintenance and repair costs. Accordingly, a partial correlation matrix was calculated specifically for variables with absolute correlation coefficients greater than 0.70. On the one hand, this analysis identifies variables that are relatively independent and influential for inclusion in the ranking models. On the other hand, by revealing strong intercorrelations, it helps prevent the inclusion of variables containing redundant information that could adversely affect the accuracy and interpretability of the ranking results. This evaluation provides an optimized and reliable data input for the machine-learning algorithms.

Table 7. Partial Correlation Matrix for Key Variables (|r| > 0.70)

Variable	Equipment Age	Operating Conditions	Failure to Implement PM	Maintenance and Repair Costs
Equipment age	1.00	0.68	−0.72	0.81
Operating conditions	0.68	1.00	−0.58	0.79
Failure to implement PM	−0.72	−0.58	1.00	−0.76
Maintenance and repair costs	0.81	0.79	−0.76	1.00

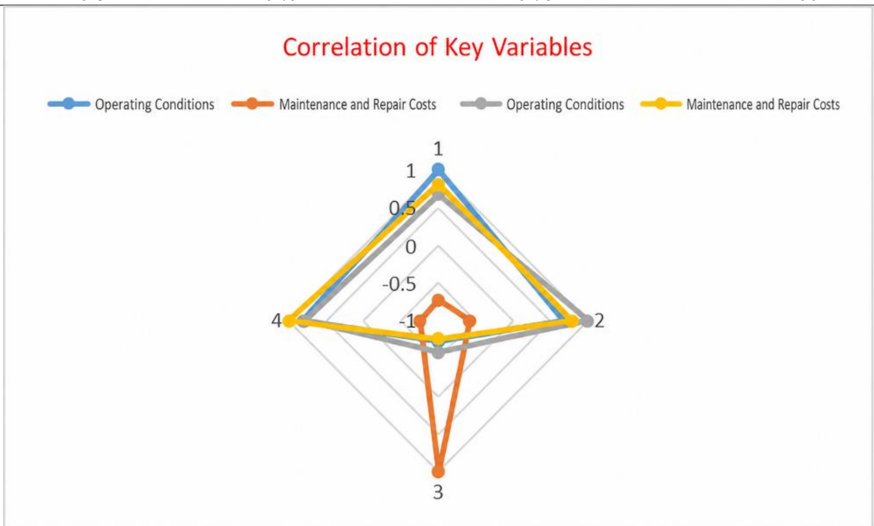

Figure 2. Correlation of Key Variables

Based on the calculated partial correlation matrix, the three key variables—equipment age, operating conditions, and failure to implement preventive maintenance—not only exhibit the strongest relationships with the target variable (maintenance and repair costs), but also show statistically meaningful interrelationships. This finding is important from two perspectives. First, it confirms that these three variables constitute the core factors affecting costs and are therefore expected to account for a substantial share of the final predictive model. Second, the relatively high correlation (0.68) between equipment age and operating conditions indicates that these two variables are partially interdependent and may exert combined effects on maintenance and repair costs.

This partial correlation analysis provides a clear roadmap for the machine-learning-based ranking stage. Machine-learning ranking algorithms can distinguish, with high accuracy, the contribution of each of these interdependent variables to cost prediction and determine their final priority. Furthermore, the negative sign of the correlation coefficient for the variable “failure to implement PM” (−0.76) indicates an inverse relationship, and the algorithm can appropriately quantify the inverse effect as well as the positive effect of adherence to preventive maintenance programs on cost reduction. Overall, this matrix indicates that the data entering the ranking stage possess an appropriate structure and quality for generating reliable and interpretable outputs.

XGBoost (Extreme Gradient Boosting) is an advanced and optimized implementation of the Gradient Boosting Machine framework introduced by Tianqi Chen in 2016. The algorithm operates through an ensemble structure of decision trees and applies the boosting technique to combine weak learners into a strong predictive model. The principal distinction between XGBoost and other boosting algorithms lies in its computational optimizations, use of regularization, and efficient parallelization capabilities, which have contributed to its superior performance in numerous data-mining competitions.

The XGBoost algorithm operates through an iterative and incremental process known as additive training. At each stage (t), a new decision tree (f_t) is added to the model to correct the errors produced by the model at the preceding stage ($t - 1$). This process is performed by optimizing an objective function composed of two components: first, a loss function that measures prediction error; and second, a regularization term that controls model complexity and prevents overfitting.

One of the key innovations of XGBoost is the use of a second-order Taylor approximation to achieve more efficient optimization of the loss function. The algorithm employs gradients (first-order derivatives) and Hessians (second-order derivatives) to obtain a more accurate approximation of the loss function. In addition, the histogram-based algorithm implemented in XGBoost enables more efficient calculation of split points. This method stores the data in discrete bins and performs calculations on these bins, thereby significantly increasing computational speed.

XGBoost provides several capabilities that distinguish it from other algorithms. These include support for L1 and L2 regularization, automatic handling of missing values, built-in cross-validation, and support for parallel processing. The algorithm also supports multiple loss functions and can be applied to various tasks, including regression, classification, and ranking. The feature-importance calculation functionality in XGBoost is another valuable capability and can be based on metrics such as Gain, Cover, and Frequency.

The Gain metric in XGBoost is used as a principal indicator for evaluating feature importance. It measures the average reduction in the loss function following each split based on a particular feature. Mathematically, the Gain associated with a particular feature is calculated by aggregating the improvement resulting from the use of that feature across all nodes in all trees. This metric not only measures the predictive power of each feature individually but also captures its ability to interact with other features to improve model accuracy. In the present study, Gain was therefore used to rank the factors affecting maintenance and repair costs.

Permutation importance is a simple yet powerful technique for identifying the most influential variables in a machine-learning model. The method first measures model performance using the original data and obtains a baseline performance metric, such as accuracy. The values of a specific variable are then randomly permuted across observations so that the relationship between that variable and the target variable is effectively disrupted. In the next step, the model is reevaluated using the modified dataset. If permuting the variable causes a substantial decline in model performance, this indicates that the variable is highly important for the model’s predictions. Conversely, if performance remains almost unchanged, the model is considered to have relatively little dependence on that particular variable. This method is considered reliable because it assesses the importance of each variable based not on the internal structure of the model, but on its observable impact on predictive performance.

SHAP, or SHapley Additive exPlanations, is a mathematical framework based on game theory that is designed to explain the outputs of machine-learning models. SHAP calculates the contribution of each feature to the model’s final prediction separately for each observation. The central idea is to calculate the average contribution of a feature to the prediction when that feature appears in different combinations with other features. By averaging these values across the entire dataset, a global and equitable measure of feature importance is obtained. One of the main advantages of SHAP is its ability to provide both local interpretability for individual predictions and global interpretability for the entire model, making it one of the most robust and theoretically grounded methods for allocating feature importance.

XGBoost offers several advantages for ranking the factors examined in this study. First, during the training process, the algorithm inherently calculates the importance of each variable using the Gain metric, which represents the average reduction in the loss function following splits based on that variable. Consequently, the measurement of factor importance is directly linked to improvements in model predictive performance. In addition, XGBoost can automatically identify and account for complex and nonlinear interactions among factors, which is particularly important in the operational environment of the petrochemical industry.

Furthermore, superior computational efficiency and greater robustness when handling noisy data are among the key advantages of XGBoost. Through regularization mechanisms and gradient boosting, the algorithm reduces the risk of overfitting and produces a model with strong generalizability. Compared with methods such as permutation importance, which require additional computations, XGBoost extracts factor importance directly and integrally during model training, thereby increasing computational efficiency while also promoting consistency in the results. These characteristics led to the selection of XGBoost as the optimal method for ranking factors affecting maintenance and repair costs in the petrochemical industry. The performance of these three approaches was subsequently evaluated according to specific criteria, as presented in the following table.

Table 8. Comparison of the Performance of Three Factor-Ranking Methods

Evaluation Criterion	XGBoost	Permutation Importance	SHAP	Conclusion
Execution speed	Very high (inherent calculation)	Moderate (requires model reevaluation)	Low (computationally complex)	XGBoost superior
Sensitivity to data noise	Low (robust)	Moderate	High	XGBoost superior
Consideration of factor interactions	Yes (comprehensive)	No (individual)	Yes (comprehensive)	XGBoost and SHAP superior
Interpretability	Moderate	High	Very high	SHAP superior
Alignment with prediction objective	Yes (direct)	Yes (indirect)	Yes (direct)	XGBoost and SHAP superior
Stability of results	Very high	High	Moderate	XGBoost superior
Computational considerations	Suitable for large datasets	Requires additional computational resources	Highly computationally intensive	XGBoost superior

The comparison of the three factor-ranking approaches indicates that XGBoost outperformed the alternatives across most of the evaluated criteria, including greater execution speed, stronger robustness to data noise, and direct alignment with the predictive objective, and was therefore selected as the optimal algorithm. Because the algorithm calculates factor importance inherently during model training and can account for complex interactions among variables, it represents the most appropriate option for the primary ranking of the factors.

Given that the relative advantages of the XGBoost algorithm were established in the preceding stage, XGBoost was employed in this stage to determine the final ranking of the factors affecting maintenance and repair costs. The implementation process consisted of four principal steps. First, the data corresponding to the 13 final factors were prepared and standardized. Second, the XGBoost model was trained and validated using optimized parameters. Third, the importance of each factor was calculated using the Gain metric, which represents the average reduction in the loss function following each feature-based split. Finally, the resulting values were normalized, and the factors were ranked according to their relative influence on the target variable. This approach enables the precise and objective identification of the most important predictors of maintenance and repair costs.

Step 1: Data Preparation

The data used in this study were based on actual information collected from 10 representative industrial machines operating in petrochemical complexes in Iran during the three-year period from March 2022 to March 2025. The dataset contained observations for the 13 final factors identified in the preceding stage. These variables were recorded monthly, producing a total of 360 observations (10 machines × 36 months). The data-preparation process was conducted as follows:

During data preparation, all quantitative variables were standardized using Z-score standardization to eliminate the effects of differences in measurement scales. Qualitative variables were converted into numerical values using numerical encoding to enable their inclusion in the model. In addition, outliers and missing values were identified and replaced using a hierarchical averaging method. Ultimately, an integrated and standardized dataset was produced, enabling the XGBoost model to be trained and evaluated with a high degree of reliability.

Scale standardization in this study was performed using Z-score standardization. For each quantitative variable, the population mean (μ) and standard deviation (σ) were calculated, after which each observation (x) was transformed into a standardized score using the equation $z = (x - \mu)/\sigma$. This mathematical transformation results in all variables having a mean of zero and a standard deviation of one. It provides two principal advantages: first, the effects of differences among measurement scales, such as years, days, and millions of Iranian rials, are eliminated; and second, all variables are placed on a common statistical scale suitable for machine-learning algorithms. This process was separately applied to the 10 quantitative variables in the study and provided an appropriate statistical basis for training the XGBoost model.

Numerical encoding was also employed to standardize the qualitative variables, whereby each category of a qualitative variable was converted into a unique numerical value. Specifically, the variable "equipment type," comprising five classes, was encoded using values from 0 to 4; "spare-parts condition," comprising three levels, was encoded using values from 0 to 2; and "level of CMMS utilization," comprising four levels, was encoded using values from 0 to 3. This method not only enabled the qualitative variables to be incorporated into the XGBoost model but also preserved the ordinal structure present in the levels of some of these variables.

Step 2: Algorithm Training

In the second step, the XGBoost model was trained using the standardized data. During this stage, key model parameters, including the learning rate, maximum tree depth, and number of trees, were optimized. Based on the gradient boosting algorithm, the model progressively constructed an ensemble of decision trees by correcting the errors generated by preceding models. Subsequently, the relative importance of each of the 13 factors in predicting maintenance and repair costs was calculated using the Gain metric, which represents the average reduction in the loss function following splits based on each feature. This metric directly indicates the extent to which each variable contributes to improving the model's predictive accuracy.

Table 9. Rationale for Selecting the Final XGBoost Parameter Values

Parameter	Final Value	Evaluated Range	Selection Rationale	Effect on Model Performance
n_estimators	300	50–500	Selected at the point beyond which improvements in accuracy became negligible	Prevents overfitting while ensuring sufficient learning
max_depth	6	3–12	Balances predictive power and generalizability	Provides sufficient depth to identify nonlinear relationships without unnecessary complexity
learning_rate	0.1	0.01–0.30	Optimal rate for relatively rapid and stable convergence	Prevents fluctuations and instability during training
subsample	0.8	0.6–1.0	Reduces variance without substantially increasing bias	Improves model generalization through partial sampling
colsample_bytree	0.9	0.6–1.0	Maintains tree diversity while using most features	Ensures that information from all variables is adequately represented across trees
reg_alpha	0.1	0–1	Applies a mild L1 penalty to reduce overfitting	Simplifies the final model and improves interpretability
reg_lambda	0.5	0–1	Controls model complexity through L2 regularization	Produces an appropriately complex and stable model

The final parameter values were selected at the point that produced the highest coefficient of determination ($R^2 = .89$) and the lowest mean squared error (MSE) on the test data. This combination of parameters provided the best attainable balance between predictive accuracy and model generalizability without exhibiting evidence of overfitting.

Steps 3 and 4: Calculation of the Gain Metric and Ranking

In the third step, the importance of each of the 13 factors was calculated using the Gain metric within the trained XGBoost model. This metric represents the average reduction in the loss function generated by each variable across the trees in the model. In other words, Gain quantifies the improvement in predictive accuracy that can be directly attributed to each factor.

The Gain values were subsequently normalized and expressed as percentages, after which the factors were ranked according to their percentage importance. In the XGBoost algorithm, the Gain metric is defined on the basis of the average reduction in the loss function following each split performed using a specific feature. Its mathematical formulation is as follows:

$$\text{Gain} = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma$$

In calculating Gain, the first- and second-order derivatives of the loss function are used, representing the gradient and curvature of the loss function, respectively. These quantities are calculated for each observation. When a node in a tree is divided into left and right branches, the Gain metric measures the improvement in predictive accuracy resulting from that split. Regularization parameters are also incorporated as control factors to prevent excessive model complexity. Ultimately, the average Gain values calculated for each feature across all trees are regarded as the final importance of that feature in the model.

Table 10. Final Ranking of Factors Affecting Maintenance and Repair Costs Based on the Gain Metric

Rank	Factor	Importance (Gain)	Percentage Importance
1	Equipment age	0.189	18.9%
2	Failure to implement preventive maintenance (PM)	0.172	17.2%
3	Adverse operating conditions	0.151	15.1%
4	Spare-parts procurement lead time	0.098	9.8%
5	Level of CMMS utilization	0.085	8.5%
6	Inefficient turnaround maintenance cycle	0.074	7.4%
7	Spare-parts condition	0.063	6.3%
8	HSE costs	0.057	5.7%
9	Exchange-rate fluctuations	0.048	4.8%
10	Environmental corrosivity index	0.035	3.5%
11	Number of specialized personnel	0.028	2.8%
12	Equipment type	0.021	2.1%
13	Budgetary constraints	0.019	1.9%

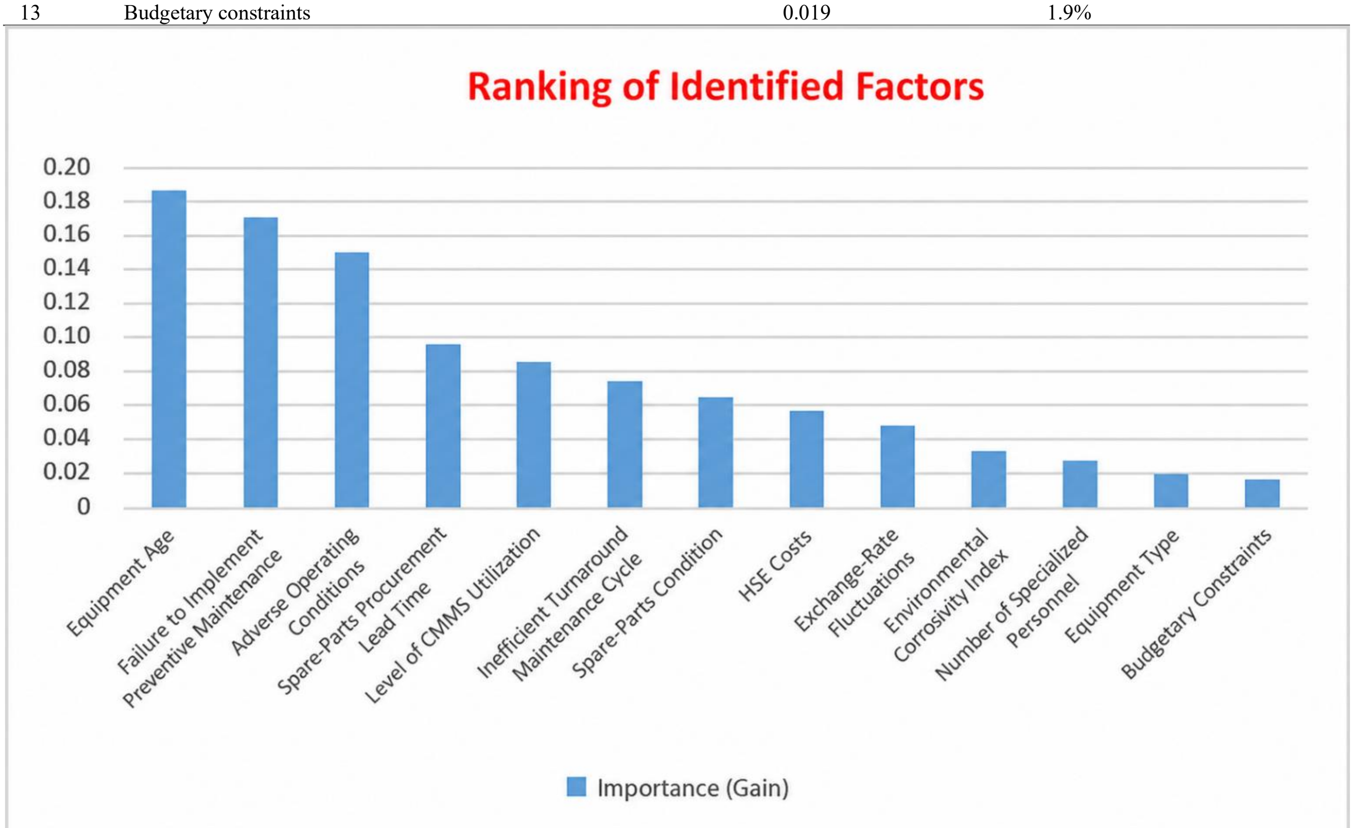

Figure 3. Ranking of Identified Factors

4. Discussion and Conclusion

The present study sought to identify and rank the factors affecting industrial machinery maintenance and repair costs in Iran's petrochemical industry by combining qualitative evidence, conventional statistical analysis, and machine-learning-based feature importance. The findings demonstrated that maintenance and repair costs are primarily driven by a relatively concentrated group of technical and operational variables. According to the XGBoost Gain metric, equipment age ranked first, accounting for 18.9% of total feature importance, followed by failure to implement preventive maintenance (17.2%) and adverse operating conditions (15.1%). Together, these three variables accounted for 51.2% of the total importance, indicating that more than half of the predictive contribution to maintenance and repair costs was concentrated in equipment deterioration, maintenance strategy, and operational stress. Spare-parts procurement lead time (9.8%) and the level of CMMS utilization (8.5%) occupied the fourth and fifth positions, respectively, followed by an inefficient turnaround maintenance cycle (7.4%), spare-parts condition (6.3%), HSE costs (5.7%), exchange-rate fluctuations (4.8%), environmental corrosivity (3.5%), number of specialized personnel (2.8%), equipment type (2.1%), and budgetary constraints (1.9%). The optimized XGBoost model achieved an R^2 of .89, suggesting that the selected factors collectively explained a substantial proportion of the variation in maintenance and repair costs.

The predominance of equipment age is theoretically and operationally plausible. Aging industrial equipment is generally associated with cumulative wear, fatigue, deterioration of mechanical and electrical components, declining energy efficiency, more frequent unplanned failures, and increasing difficulty in obtaining compatible replacement parts. The strong positive correlation observed between equipment age and maintenance cost (r = .81) therefore indicates that the financial consequences of asset aging extend beyond routine increases in repair frequency. Kolo emphasized that obsolescence of rotating equipment in the oil and gas sector creates substantial risks through reduced reliability, spare-parts scarcity, downtime, and escalating life-cycle costs [7]. Similarly, technical-economic evaluations of machinery maintenance systems have demonstrated that deteriorating equipment conditions can progressively increase repair requirements and reduce the economic efficiency of continued operation [2]. Studies of machinery-cost management also emphasize that aging and replacement decisions must be considered within total life-cycle cost rather than only immediate repair expenditure [1]. The present finding therefore suggests that petrochemical companies should view equipment age as a strategic asset-management variable rather than merely a descriptive characteristic of machinery.

The second-ranked factor, failure to implement preventive maintenance, accounted for 17.2% of total importance and exhibited one of the strongest statistical relationships with maintenance and repair expenditure. This result supports the central premise of modern maintenance management that postponing intervention until functional failure can substantially increase total maintenance costs. Preventive maintenance can reduce catastrophic failures, stabilize production schedules, improve spare-parts planning, and limit secondary damage to connected components. Research on predictive and preventive maintenance across industrial settings has consistently shown that earlier identification of deterioration enables interventions to be undertaken before breakdown-related costs escalate [17, 18]. Similarly, joint optimization of preventive maintenance and spare-parts ordering demonstrates that properly scheduled preventive interventions can improve both equipment availability and economic performance when coordinated with inventory decisions [22, 23]. The high feature importance observed in the current model is consequently consistent with the growing transition from reactive maintenance toward planned, condition-based, and predictive approaches.

The negative correlation reported for the preventive-maintenance-related variable should be interpreted in relation to the direction of coding and the underlying managerial concept. Regardless of the numerical sign, the XGBoost results demonstrated that this variable had the second-highest predictive contribution to maintenance costs. Conceptually, stronger adherence to preventive maintenance should be associated with lower emergency repair requirements and fewer unplanned failures, whereas inadequate implementation increases the probability of cost-intensive breakdowns. This interpretation is consistent with research applying machine-learning methods to predictive maintenance, where timely detection and intervention are intended to reduce unexpected failure and associated maintenance expenditure [21, 30]. Reliability modeling in critical infrastructure has similarly demonstrated that maintenance policies become more economically effective when they incorporate information about deterioration processes and likely future failure states [20]. Thus, the present results strengthen the argument that maintenance compliance and scheduling quality should be treated as direct cost-control mechanisms.

Adverse operating conditions ranked third with 15.1% importance and showed a strong correlation with maintenance costs (r = .79). This finding is particularly important in petrochemical environments, where machinery frequently operates under elevated temperatures, vibration, pressure variation, moisture, corrosive media, and potentially explosive atmospheres. Such conditions can accelerate mechanical wear, lubrication degradation, seal deterioration, instrumentation drift, insulation damage, and structural corrosion. Research in oil and gas environments has shown that equipment performance and instrumentation accuracy are strongly affected by harsh environmental and operational conditions, creating additional requirements for monitoring and maintenance [13]. AI-driven predictive maintenance has consequently been proposed as a means of continuously tracking equipment condition under complex field environments and identifying deterioration before failure [14]. Studies of thermoelectric energy harvesting and batteryless IIoT architectures in oil-refinery and oil-and-gas applications further underscore the importance of environmental severity in determining both monitoring requirements and maintenance architecture [15, 16]. The present results therefore suggest that operating-condition control can be as important as maintenance intervention itself.

Spare-parts procurement lead time ranked fourth, with a Gain importance of 9.8%, while spare-parts condition accounted for a further 6.3%. These findings highlight the central role of maintenance logistics and supply-chain management. Even when a failure is accurately diagnosed and maintenance personnel are available, an unavailable component may extend downtime, increase production losses, force emergency procurement, or lead to the temporary use of lower-quality substitutes. Joint maintenance-inventory models have demonstrated that spare-parts availability and preventive maintenance decisions are economically interdependent [22, 23]. El Mola likewise showed that spare-parts manufacturing and sourcing strategies can become critical when organizations face supply uncertainty and long procurement cycles [24]. This issue is especially relevant in the Iranian petrochemical context, where international supply restrictions, reliance on original equipment manufacturer components, exchange-rate volatility, and logistical uncertainty can amplify procurement costs. The finding that sanctions and supply restrictions were statistically significant is also consistent with this interpretation, even though their effects were subsequently reflected through measurable operational indicators such as procurement lead time and spare-parts condition.

The level of CMMS utilization was the fifth most important factor, accounting for 8.5% of feature importance. This result suggests that digital maintenance management is not simply an administrative support function but an important determinant of economic performance. A computerized maintenance management system can integrate work orders, preventive maintenance schedules, equipment histories, spare-parts inventories, labor allocation, failure records, and maintenance-performance indicators. Musa identified weaknesses in maintenance documentation, inventory management, systematic planning, and asset information as important problems in oil and gas maintenance management and emphasized the role of structured maintenance-management systems in overcoming such inefficiencies [6]. The present results extend this argument by quantitatively demonstrating that the extent of CMMS utilization has a meaningful association with maintenance and repair costs. In practice, a CMMS can reduce duplication, improve scheduling accuracy, prevent missed preventive tasks, and provide the historical data required for predictive modeling.

The inefficient turnaround maintenance cycle ranked sixth, with 7.4% importance. Turnarounds and shutdowns represent some of the most financially intensive maintenance events in petrochemical operations because they combine direct maintenance expenditure with production interruption, contractor mobilization, equipment inspection, parts replacement, and HSE requirements. Dhanaraj reported that large-scale maintenance projects are exposed to financial risks related to planning, procurement, execution delays, contractor management, and other sources of uncertainty [8]. The present findings align closely with this evidence. An inefficient turnaround cycle can extend outage duration, increase labor and contractor expenses, generate urgent procurement requirements, and postpone the return of production units to service. The statistical association observed in this study therefore supports the need to manage turnaround activities as integrated projects with clearly coordinated engineering, procurement, inventory, scheduling, and risk-control processes.

The importance attributed to HSE costs and environmental corrosivity further reflects the particular characteristics of petrochemical operations. HSE costs accounted for 5.7% of the total Gain, whereas environmental corrosivity accounted for 3.5%. In hazardous industrial environments, maintenance tasks frequently require permit-to-work systems, isolation procedures, gas testing, specialized protective equipment, scaffolding, inspection, and additional supervision. Although such expenditures are indispensable for safe operation, they increase the direct cost of maintenance activities. At the same time, corrosive environments may increase failure rates and shorten component life, thereby generating additional maintenance demand. Research addressing harsh industrial environments has emphasized the need for specialized monitoring architectures capable of operating reliably under high-temperature and hazardous conditions [15, 16]. Likewise, vibration and condition-monitoring studies demonstrate that environmental and mechanical stress indicators can provide early evidence of degradation in rotating machinery [11]. Thus, the findings indicate that environmental severity contributes to cost through both direct protective requirements and indirect acceleration of deterioration.

Exchange-rate fluctuations accounted for 4.8% of feature importance, whereas budgetary constraints were ranked last at 1.9%. The relatively greater contribution of exchange-rate variability suggests that external economic instability may influence maintenance expenditure more directly than the nominal size of the approved maintenance budget. In industries dependent on imported components, specialized services, or foreign technology, exchange-rate fluctuations can rapidly change replacement-part and contractor costs. Cost-control research in manufacturing has shown that sustainable reductions in expenditure depend on integrated operational and financial management rather than simple budget restriction [25]. The relatively low importance of budget constraints in the present model may therefore indicate that limiting expenditure does not itself remove the technical need for maintenance. Indeed, inadequate budgets may postpone intervention temporarily, but the resulting deterioration can eventually produce even greater costs. This result supports an asset-management perspective in which budget allocation should be based on equipment criticality and expected life-cycle risk rather than across-the-board expenditure reduction.

The number of specialized personnel and equipment type showed comparatively smaller feature importance, at 2.8% and 2.1%, respectively. Their lower ranking should not, however, be interpreted as evidence that workforce expertise or equipment characteristics are unimportant. Rather, part of their effect may be mediated by more proximal variables such as maintenance execution quality, downtime, condition monitoring, and technical complexity. Empirical research on maintenance-service selection indicates that technical competence, service capability, and organizational characteristics remain important components of industrial maintenance performance [9]. Similarly, different classes of equipment can have different operating patterns and failure modes, but model-based analysis may attribute much of this variability to directly measured factors such as age, condition, operating stress, and preventive-maintenance performance. The ranking produced here is therefore best interpreted as the relative predictive contribution of variables within the present dataset, rather than an absolute judgment regarding their managerial significance.

Another noteworthy result was the exclusion of lack of strategic planning, resistance to change, and declining accountability from the final quantitative model because they could not be operationalized using reliable historical measurements. Their qualitative importance was nevertheless evident in the expert interviews. This distinction demonstrates the value of the mixed-methods approach adopted in the study. Organizational and cultural factors may substantially influence maintenance performance but can be difficult to quantify using routine maintenance databases. The expansion of big data and artificial intelligence in project planning and cost prediction may eventually allow some of these less tangible factors to be represented through behavioral or process indicators [28]. For the present analysis, however, restricting the XGBoost model to objectively measurable variables improved the transparency and reproducibility of the ranking process.

Overall, the findings demonstrate that maintenance and repair costs in petrochemical machinery are generated through interconnected technical, operational, managerial, logistical, and economic mechanisms rather than a single cost driver. The high predictive performance of the XGBoost model ($R^2 = .89$) suggests that machine learning can provide a useful complement to conventional maintenance-cost analysis by distinguishing the relative importance of variables even when their relationships are nonlinear and interconnected. Previous work has demonstrated the effectiveness of machine learning for equipment-failure prediction, machinery-condition assessment, maintenance forecasting, and cost estimation across different contexts [27, 29, 31, 32]. Machine-learning approaches have also successfully modeled complex machinery behavior and repair-related outcomes in applications ranging from gas turbines and motor operating conditions to vehicle and infrastructure repair costs [12, 33-35]. Accordingly, the present study contributes to the literature by moving beyond failure prediction alone and applying XGBoost feature importance to the managerial prioritization of maintenance and repair cost drivers in a petrochemical setting.

This study has several limitations that should be considered when interpreting its findings. First, the quantitative dataset was obtained from 10 industrial machines over a three-year period, resulting in 360 monthly observations; although this structure provided repeated operational information, the number and diversity of machines remained limited. Second, the research was conducted within the context of Iranian petrochemical operations, and some identified economic and supply-chain conditions may differ in other industries or national settings. Third, three qualitatively important organizational variables were excluded from the machine-learning stage because valid historical quantitative measures were unavailable. Fourth, numerical coding of categorical variables may not fully represent all differences among their underlying categories. Finally, feature importance indicates predictive contribution rather than causal effect; therefore, a high Gain value should not automatically be interpreted as proof that changing a factor by itself will produce a proportional change in maintenance expenditure.

Future research should replicate the model using larger datasets that include a greater number of machines, petrochemical complexes, equipment classes, and longer observation periods. Comparative studies across oil, gas, refining, power-generation, mining, and manufacturing environments would help establish the external validity of the proposed ranking framework. Researchers should also develop validated quantitative scales or proxy indicators for organizational variables such as resistance to change, maintenance culture, strategic planning maturity, accountability, and workforce engagement so that these factors can be incorporated into predictive models. Future studies could compare XGBoost with Random Forest, LightGBM, CatBoost, neural networks, and hybrid ensemble approaches and evaluate model interpretation through SHAP and permutation importance. Longitudinal causal modeling could additionally be used to determine whether changes in highly ranked factors are followed by measurable reductions in maintenance costs. Integrating real-time sensor, CMMS, inventory, procurement, and financial information would provide an especially promising basis for dynamic cost prediction.

Petrochemical managers should prioritize maintenance investment according to the ranked factors rather than applying uniform cost-reduction measures across all assets. Aging and critical equipment should be identified through structured asset-health and replacement programs, while preventive-maintenance compliance should be continuously monitored and linked to equipment criticality. Operating conditions should be tracked through condition-monitoring systems capable of identifying abnormal temperature, vibration, pressure, and corrosive exposure before failure occurs. Organizations should reduce spare-parts procurement lead times by classifying critical inventory, developing alternative suppliers, evaluating localized manufacturing, and establishing minimum stock levels for high-risk components. CMMS utilization should be expanded beyond work-order recording to include preventive scheduling, failure histories, inventory integration, labor planning, and performance analytics. Turnaround projects should be managed through integrated planning, procurement, scheduling, and risk-control structures. Finally, maintenance-cost dashboards should incorporate the highest-ranked predictors identified in this study so that managers can continuously monitor changes in cost risk and allocate financial and technical resources to interventions with the greatest expected impact.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

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

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

[1] M. Zoghi, M. Ghasemlou, and M. Jafari Fesharaki, "Machinery Costs and Their Management," presented at the 9th National Conference on Sustainable Development in Civil Engineering, Tehran, Iran, 2022. [Online]. Available: https://civilica.com/doc/1523994.

[2] G. Iovlev, "Technical and Economic Assessment of Maintenance and Repair Systems of Agricultural Machinery," *Agrarian Science,* 2023, doi: 10.32634/0869-8155-2023-369-4-129-136.

[3] E. Gradov and E. Malykha, "Actual Issues of Maintenance of Equipment in the Agro-Industrial Complex," *Tekhnicheskiy servis mashin,* 2024, doi: 10.22314/2618-8287-2024-62-4-25-31.

[4] A. Akbari, "Examining Effective Methods for Reducing Maintenance and Repair Costs of Road Construction Machinery," presented at the 10th Annual International Congress on Civil Engineering, Architecture and Urban Development, Tehran, Iran, 2024. [Online]. Available: https://civilica.com/doc/2153806.

[5] M. Tahmasbi, "Development of Optimal Maintenance Management for Road Construction Machinery in Ilam Municipality Civil Projects," presented at the 10th International Conference on Interdisciplinary Research in Management, Accounting and Economics in Iran, Tehran, Iran, 2023. [Online]. Available: https://civilica.com/doc/1777429.

[6] A. I. N. Musa, "A Case Study of Maintenance Engineering and Management in Oil and Gas Industry," presented at the Full Text Proceedings Book, 2025. [Online]. Available: https://doi.org/10.71284/asc.2025134.

[7] U. K. Kolo, "Obsolescence Management of Rotating Equipment in the Oil and Gas Industry: A Strategic Approach to Minimizing Loss and Downtime," presented at the SPE Nigeria Annual International Conference and Exhibition, 2025. [Online]. Available: https://doi.org/10.2118/228668-ms.

[8] D. Dhanaraj, "A Study of the Sources and Mitigation Strategies of Financial Risks in Large-Scale Maintenance Projects," *International Journal of Scientific Research,* 2024, doi: 10.36106/ijsr/2822769.

[9] P. M. Chaudhari, "Empirical Analysis of Factors Affecting Industrial Maintenance Services Selection," *Journal of Informatics Education and Research,* 2024, doi: 10.52783/jier.v4i1.634.

[10] S. Shang, "Risk Analysis of Industrial Machinery and Equipment Repair," presented at the Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management, 2024. [Online]. Available: https://doi.org/10.1145/3718751.3718932.

[11] M. Romanssini, P. C. C. de Aguirre, L. Compassi-Severo, and A. G. Girardi, "A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery," *Eng,* vol. 4, no. 3, pp. 1797-1817, 2023, doi: 10.3390/eng4030102.

[12] K. Akkad, "Motor Operational Settings Prediction for Sustainable Manufacturing Facilities," *IEEE Access,* vol. 12, pp. 118185-118194, 2024, doi: 10.1109/ACCESS.2024.3447077.

[13] D. D. Jambol, "Leveraging Machine Learning to Enhance Instrumentation Accuracy in Oil and Gas Extraction," *Computer Science & IT Research Journal,* 2024, doi: 10.51594/csitrj.v5i6.1200.

[14] D. D. Jambol, "Transforming Equipment Management in Oil and Gas with AI-Driven Predictive Maintenance," *Computer Science & IT Research Journal,* 2024, doi: 10.51594/csitrj.v5i5.1117.

[15] R. Aragonés, "Transforming Industrial Maintenance with Thermoelectric Energy Harvesting and NB-IoT: A Case Study in Oil Refinery Applications," *Sensors (Basel, Switzerland),* vol. 25, 2025, doi: 10.3390/s25030703.

[16] R. Aragonés, "Enhanced Heat-Powered Batteryless IIoT Architecture with NB-IoT for Predictive Maintenance in the Oil and Gas Industry," *Sensors (Basel, Switzerland),* vol. 25, 2025, doi: 10.3390/s25082590.

[17] T. Lukito, R. Herlianti, M. Mayanti, and L. H. Kusumah, "Implementation of Predictive Maintenance in Various Industry: A Review," *TEKNOSAINS: Jurnal Sains, Teknologi Dan Informatika,* vol. 12, no. 1, pp. 133-144, 2025, doi: 10.37373/tekno.v12i1.1338.

[18] M. B. Shaikh, "Implementing Machine Learning for Predictive Maintenance in Industrial Machinery," presented at the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2024. [Online]. Available: https://doi.org/10.1109/ICCCNT61001.2024.10724004.

[19] Z. Shao and M. Kumral, "Implementing Gaussian Process Modelling in Predictive Maintenance of Mining Machineries," *Mining Technology,* vol. 133, no. 4, pp. 348-368, 2024, doi: 10.1177/25726668241275434.

[20] R. M. Spangler, V. Agarwal, and D. G. Cole, "A Hybrid Reliability Model Using Generalized Renewal Processes for Predictive Maintenance in Nuclear Power Plant Circulating Water Systems," *IEEE Access,* vol. 11, pp. 136726-136740, 2023, doi: 10.1109/ACCESS.2023.3338716.

[21] T. Akyaz and D. Engin, "Machine Learning-Based Predictive Maintenance System for Artificial Yarn Machines," *IEEE Access,* vol. 12, pp. 125446-125461, 2024, doi: 10.1109/ACCESS.2024.3454548.

[22] G. Ding, K. Kaliyaperumal, and X. Wang, "A New Mathematical Model for Economic Ordering with Preventive Maintenance and Reworking in a Supply Chain," *Discrete Dynamics in Nature and Society,* p. 8636482, 2023, doi: 10.1155/2023/8636482.

[23] Y. He and Z. Gao, "Joint Optimization of Preventive Maintenance and Spare Parts Ordering Considering Imperfect Detection," *Systems,* vol. 11, p. 445, 2023, doi: 10.3390/systems11090445.

[24] K. M. S. G. El Mola, "Development of a Manufacturing Strategy of Spare Parts: A Case Study," *International Journal of the Analytic Hierarchy Process,* vol. 15, no. 2, 2023, doi: 10.13033/ijahp.v15i2.1000.

[25] R. A. Fadjarenie, C. Rachmadani, and D. Tarmidi, "Cost Reduction Strategy in Manufacturing Industries: Empirical Evidence from Indonesia," *Jurnal Akuntansi,* vol. 28, no. 1, pp. 61-79, 2024, doi: 10.24912/ja.v28i1.1747.

[26] S. S. Sanga and G. Sai Charan, "Fuzzy Modeling and Cost Optimization for Machine Repair Problem with Retrial under Admission Control F-Policy and Feedback," *Mathematics and Computers in Simulation,* vol. 211, pp. 214-240, 2023, doi: 10.1016/j.matcom.2023.03.036.

[27] A. Raoufi and S. N. Bashi Azghadi, "Application of Machine Learning Techniques in Estimating and Predicting Construction Project Costs: A Review of Research Conducted from 2017 to 2022," presented at the 16th International Symposium on Advances in Science and Technology, Mashhad, Iran, 2023. [Online]. Available: https://civilica.com/doc/2019667.

[28] A. H. Sabzevari, "Examining Methods for Using Big Data and Artificial Intelligence to Predict Project Scheduling and

Costs," 2024. [Online]. Available: https://civilica.com/doc/2179970.
[29] L. Polverino, R. Abbate, and P. Manco, "Machine Learning for Prognostics and Health Management of Industrial Mechanical Systems and Equipment: A Systematic Literature Review," *International Journal of Engineering Business Management,* vol. 15, 2023, doi: 10.1177/18479790231186848.
[30] T. Ozkul and A. Topallı, "Comparison of Machine Learning Methods for Limited Predictive Maintenance," *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi,* vol. 14, no. 1, pp. 183-191, 2025, doi: 10.28948/ngumuh.1465282.
[31] A. Ghasemian and K. Khodayar, "Predicting Equipment Failures on Oil Platforms Using Artificial Intelligence and Machine Learning," presented at the 3rd International Congress of Chemical Engineering and Petroleum Industry Students, Tehran, Iran, 2024. [Online]. Available: https://civilica.com/doc/2182832.
[32] J. M. Kim, S. G. Yum, M. D. Adhikari, and J. Bae, "A LSTM Algorithm-Driven Deep Learning Approach to Estimating Repair and Maintenance Costs of Apartment Buildings," *Engineering, Construction and Architectural Management,* vol. 31, no. 13, pp. 369-389, 2024, doi: 10.1108/ECAM-11-2023-1194.
[33] E. Olaye, E. Owraigbo, and N. Bello, "Estimating Cost of Pothole Repair from Digital Images Using Stereo Vision and Artificial Neural Network," *International Journal of Applied Methods in Electronics and Computers,* vol. 12, no. 1, pp. 001-009, 2024, doi: 10.58190/ijamec.2024.77.
[34] B. Sagidolla, S. Ali, D. Aibolat, and N. Shayakhmetov, "Quantative Model for Estimaitng Vehicle Repair Costs in Insurance Claims," *Herald of the Kazakh-British Technical University,* vol. 21, no. 4, pp. 58-80, 2024, doi: 10.55452/1998-6688-2024-21-4-58-80.
[35] A. N. Fadja *et al.*, "Machine Learning Approaches for the Prediction of Gas Turbine Transients," *Journal of Computer Science,* vol. 20, no. 5, pp. 495-510, 2024, doi: 10.3844/jcssp.2024.495.510.