KARYA ILHAM PUBLISHING
CONNECTING MINDS, BRIDGING IDEAS, INSPIRING READERS

Journal of Advanced Research in Computing and Applications

Journal homepage:
https://karyailham.com.my/index.php/arca/index
ISSN: 2462-1927

Machine Learning Approaches for Analyzing Computerized Maintenance Management System Data Mining

Mohd Shukri Abdul Wahab[1], Syed Tarmizi Syed Shazali[1,*], Noor Hisyam Noor Mohamed[1], Abdul Rani Achmed Abdullah[2], Muhamad Fadzli Ashari[1], Mohd Syahmi Jamaludin[1]

[1] Faculty of Engineering, Universiti Malaysia Sarawak,Sarawak, Malaysia
[2] CWorks Technologies Sdn Bhd, Kinrara Straits One, Bandar Kinrara, Selangor. Malaysia

ARTICLE INFO

Article history:
Received 19 October 2025
Received in revised form 21 November 2025
Accepted 28 November 2025
Available online 10 December 2025

Keywords:
Maintenance management; Computerized Maintenance Management System (CMMS); machine learning; data mining; unsupervised learning; K-Means clustering

ABSTRACT

In the realm of maintenance management, the growing complexity and volume of Computerized Maintenance Management System (CMMS) data necessitate innovative approaches. Traditional methods fail to address contemporary CMMS datasets, compelling the adoption of advanced methodologies. This research focuses on leveraging machine learning approaches to comprehensively analyse CMMS data mining, enhancing predictive capabilities and interpretability. The study's objectives encompass addressing limitations in traditional data mining and employing unsupervised learning methods by using clustering techniques through K-Means. The results reveal heightened accuracy in predicting maintenance needs, improved interpretability, and identifying involved relationships within extensive CMMS datasets. This research contributes by demonstrating practical applications, offering insights for organisations enhancing CMMS analytics, and proposing future studies to advance machine learning integration further.

1. Introduction

In the dynamic landscape of maintenance management, the importance of leveraging advanced analytical methodologies to handle the increasing complexity and data volume of CMMS cannot be overstated [1]. This has become a critical imperative in ensuring the efficiency and resilience of maintenance operations.

A comprehensive review of the literature underscores the limitations of traditional data mining approaches in adequately addressing the intricacies of contemporary CMMS datasets [2]. Existing methodologies struggle to cope with the nuanced patterns and relationships within the growing amount of maintenance data.

The problem at hand lies in the evident gap between the growing demands placed on maintenance management and the capacity of current analytical techniques to deliver actionable

* *Corresponding author.*
E-mail address: starmizi@unimas.my

https://doi.org/10.37934/arca.40.1.280286

insights [3]. As organizations grapple with increasingly complex CMMS data, the need for advanced frameworks, especially those based on machine learning, becomes paramount.

The purpose of this research is to bridge this analytical gap by systematically investigating the application of machine learning approaches to CMMS data mining [4]. The integration of advanced techniques aims to enhance predictive capabilities, improve interpretability, and provide a robust framework for navigating the complexities inherent in contemporary maintenance datasets.

This study's objectives include applying both supervised and unsupervised learning techniques to CMMS data and evaluating the impact on predictive maintenance analytics [5]. These objectives are structured to address the challenges identified in the literature and provide actionable insights for practitioners.

Beyond identifying and addressing contemporary issues, this research aims to contribute significantly to the field [6]. By establishing a foundation for the integration of machine learning into CMMS analytics, the study provides practical insights that can be readily applied by organizations looking to optimize maintenance operations.

The article follows a structured approach with advanced methodologies. This comprehensive exploration promises to improve the understanding and application of machine learning in the context of CMMS data mining.

2. Methodology

To improve the analytical capabilities of CMMS data mining, a systematic methodology is used, including the selection of machine learning techniques and an overview of the data set.

A. Machine Learning Techniques

To systematically analyze CMMS data mining, a set of machine learning techniques is selected. This includes supervised learning methods such as classification and regression algorithms, facilitating predictive maintenance analysis. Additionally, unsupervised learning methods, including clustering techniques are used to uncover hidden patterns in CMMS datasets. Figure 1 shows the flow of the Machine Learning Approach (adapted from [5]).

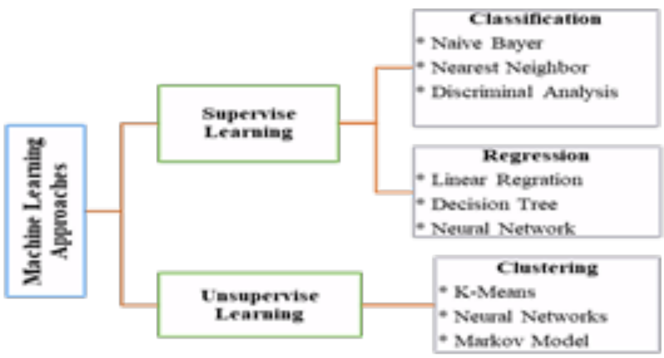


Fig. 1. Machine learning approaches

B. CMMS Dataset

The CMMS dataset serves as the basis for this study, representing a true picture of maintenance operations. An overview of the data set was provided, historical maintenance data mining from a representative sample of fuel stations was collected and analyzed to clustering through K-Means of size, structure, and related characteristics.

K-means clustering is a widely used clustering algorithm that aims to partition a given dataset into k groups. It does so by minimizing an objective function that is defined in terms of the total within-group sum-of-squares. However, the algorithm has a few limitations. For instance, it is sensitive to initialization, which means the success of the algorithm in finding globally optimal partitions depends on the starting values. Additionally, it can easily get stuck at a local minimum regarding the measurement (the sum of squared errors) used in the model. There are several initialization methods proposed for the k-means algorithm to overcome these issues.

Preprocessing steps are carefully applied to ensure data quality and improve the effectiveness of machine learning models. This includes handling missing values, standardizing numerical characteristics, and coding categorical variables. The dataset preparation process is executed with precision to create clean and representative input for machine learning algorithms.

In this study, we have identified two distinct attribute characteristics which is nominal and numerical. Our analysis involves the use of clustering methods to determine the multivariate characteristics and relations. To ensure better analysis and elimination of missing values, we have selected 5 different types of attributes and characterized 461 instances. The dataset related to CMMS can be found in Table 1.

Table 1
Attribute characteristic

Attribute	Attribute Type	Mean	Standard Deviation	Number of distinct
Month	Nominal	-	-	3
Location	Nominal	-	-	3
Maintenance Types	Nominal	-	-	3
Work Type	Nominal	-	-	3
Number of Report	Numeric	3.384	2.799	11

C. *WEKA Software*

In this study, we utilized the WEKA software tool, which provides machine learning algorithms specifically designed for data mining tasks. WEKA offers a range of algorithms that can be directly applied to a dataset, or alternatively, you can use your own Java code. The WEKA workbench also includes a set of powerful visualization tools and algorithms that can be utilized for better decision-making through data analysis and predictive modelling. Additionally, WEKA provides a graphical user interface (GUI) that makes it user-friendly. With WEKA, we can classify data, cluster it, preprocess it, perform K-Means analysis, discover association rules, and visualize it.

3. Results and Discussions

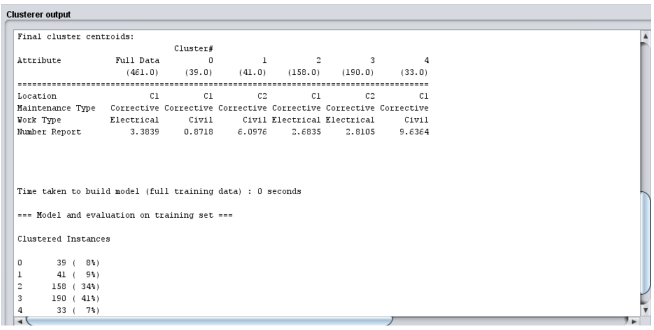
The convergence of empirical findings and analytical insights unfolds in the results and discussion section, illuminating the transformative impact of machine learning on the analysis of CMMS data.

The discussion interprets these results within the domain of predefined research objectives, shedding light on their practical implications. A critical comparative analysis with traditional data mining approaches serves as a benchmark, showcasing the distinctive advantages and heightened efficacy of machine learning models.

A. Results

The finale of the machine learning analysis on the CMMS dataset exposes essential insights into maintenance operations. The presentation of key findings is structured to provide a comprehensive understanding of the impact of machine-learning technique. Unsupervised learning methods, particularly clustering algorithms, prove instrumental in uncovering latent patterns within the CMMS dataset. Visual representations, the identified clusters, shedding light on underlying structures in maintenance data.

Based on Figure 2, it is evident that Cluster 3 accounts for the largest portion, that is, 41%. Given the unique characteristics of Cluster 3, it is essential to focus on it. Cluster 3 is characterized by location C2, corrective maintenance type, electrical work type, and approximately 2.8 reports of downtime equipment. Therefore, it is necessary to make allocation decisions based on the average of maintenance and work type characteristics to make an informed decision. According to the full data set's average characteristics, corrective maintenance and electrical work types have around 3 records. Hence, there is a significant difference between the characteristics of Cluster 3 and those of the full data set.



Clusterer output

```
Final cluster centroids:
                               Cluster#
Attribute          Full Data          0          1          2          3          4
                     (461.0)     (39.0)     (41.0)    (158.0)    (190.0)     (33.0)
==================================================================================
Location                  C1         C1         C2         C1         C2         C1
Maintenance Type  Corrective Corrective Corrective Corrective Corrective Corrective
Work Type         Electrical      Civil      Civil Electrical Electrical      Civil
Number Report         3.3839     0.8718     6.0976     2.6835     2.8105     9.6364



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       39 (  8%)
1       41 (  9%)
2      158 ( 34%)
3      190 ( 41%)
4       33 (  7%)
```

Fig. 2. Clustered output screen in Weka of Simple K-means Algorithm

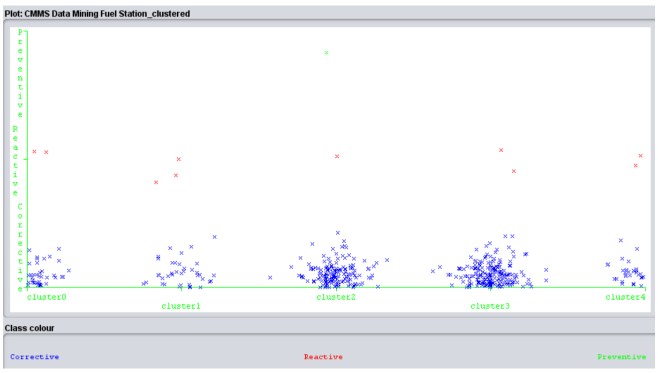


Fig. 3. Display plot of clusterer with maintenance type in Weka

Journal of Advanced Research in Computing and Applications
Volume 40, Issue 1 (2025) 280-286

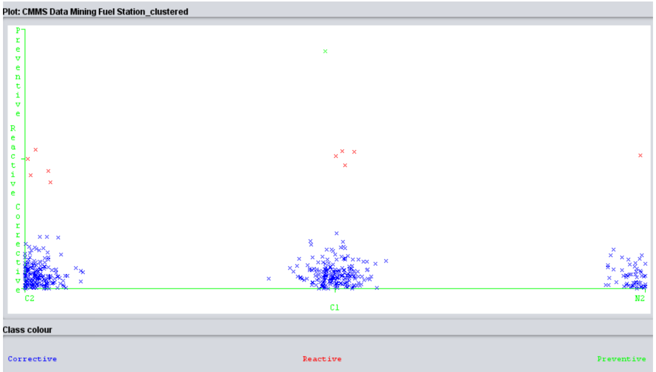


Fig. 4. Display plot of location with maintenance type in Weka

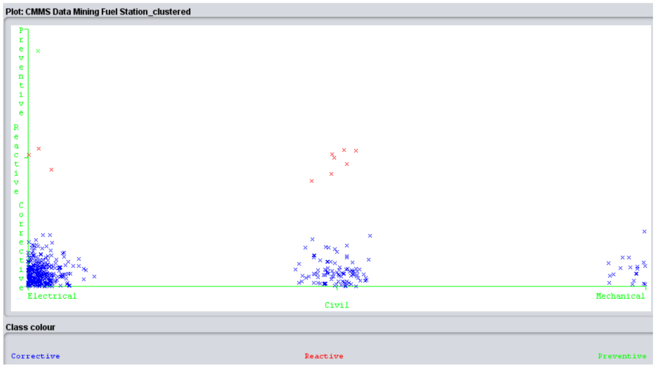


Fig. 5. Display plot of work with maintenance type in Weka

Figures 3, 4, and 5 shows the clustered output display for each algorithm plot for the cluster, location, and type of work for each maintenance operations. These figures represent each cluster obtained. The K-means algorithm is an effective method for getting the partitioning result that is closest to the original data set, as shown in Table 1. The screenshot figure demonstrates that machine learning clustering for simple K-Means accurately identifies the structure of the clusters that are similar to the original data set.

4. Discussion

The discussion section critically interprets the results obtained from the machine learning analysis of CMMS data, placing them within the context of the research objectives. Additionally, it involves a comparative analysis with traditional data mining approaches and explores the broader implications of integrating machine learning into CMMS analytics.

Electrical maintenance is the most important type of maintenance that is dedicated to the upkeep and care of the electrical infrastructure at fuel stations. It involves systematic inspection, repair, and maintenance of various electrical components that are critical to the station's operations. These components include control panels, lighting systems, sensors, and other electronic devices that contribute to the efficient functioning of the facility. Tasks within this cluster may includeeude:

- Conducting regular inspections of electrical panels and control systems.
- Repairing or replacing faulty electrical components to ensure continuous functionality.
- Performing proactive maintenance on lighting systems and electronic devices.
- Ensuring compliance with electrical safety standards and regulations.

They cater to the unique needs of electrical systems, covering routine maintenance, repairs, and proactive measures to ensure optimal performance and prevent unexpected failures.

The achieved results align closely with the predefined research objectives, affirming the efficacy of machine learning in advancing CMMS data analysis. The interpretative discussion delves into the nuances of predictive maintenance accuracy, uncovering latent patterns, and deriving holistic insights through ensemble models. This section elucidates how the machine learning outcomes contribute to meeting the study's specific goals and addressing the identified challenges in maintenance analytics.

A critical comparative analysis is undertaken, contrasting the outcomes of machine learning approaches with traditional data mining methods [7]. While traditional techniques have historically been foundational in uncovering patterns in maintenance data, the discussion articulates the distinctive advantages of machine learning, particularly in handling the intricacies of contemporary CMMS datasets. The nuanced comparisons highlight the superior predictive capabilities and adaptability of machine learning models, emphasizing their potential to surpass the limitations of traditional approaches [8].

The discussion extends to the broader implications of integrating machine learning into CMMS analytics. This includes considerations of enhanced efficiency in maintenance operations, improved decision-making processes. and the potential for proactive intervention based on predictive insights [9]. Furthermore, the discussion addresses the scalability of machine learning models, emphasizing their adaptability to diverse CMMS scenarios and their potential to revolutionize maintenance management practices [10].

4. Conclusion

The deduction of this study into machine learning approaches for analyzing CMMS data highlights key findings that bear profound significance for the landscape of maintenance management. As predictive maintenance accuracy achieves commendable heights, latent patterns within the CMMS dataset emerge, offering a more nuanced understanding of maintenance operations. Ensemble models contribute holistic insights, elevating the robustness of CMMS analytics. In conclusion, the significance lies not only in the immediate improvements observed but also in the potential for transformative change in maintenance practices.

Tasks within a cluster may include routine facility inspections, predictive maintenance, repairs and corrections, replacement of components, safety measures, environmental compliance, documentation, reporting, and coordination with other clusters.

This study serves as a catalyst, urging a proactive stance towards the adoption of machine learning in CMMS data mining. The call to action resonates with industry stakeholders, inviting them to embrace these advanced methodologies as integral tools for optimizing maintenance operations. The adoption of machine learning in CMMS analytics not only promises enhanced efficiency and reliability but sets the stage for a future where maintenance strategies are not just responsive but anticipatory, ushering in a new era of resilience and efficacy in the management of critical assets. As this research journey concludes, it invites the industry to embark on a new trajectory, where the synergy of machine learning and CMMS data mining becomes the cornerstone of progressive maintenance management practices.

Acknowledgement
The authors would like to thank the Faculty of Engineering and The Research, Innovation and Enterprise Centre (RIEC), Universiti Malaysia Sarawak for the Industrial Grant with CWorks Technologies Sdn Bhd (Project Code: IRG/F02/CWORKS/85394/2022) for their continuous support in conducting this study.

References

[1] Cohen, Danielle, Nikola Visnjic, Dominic Akaateba, and Kelly Hadfield. "Narrative literature review of facilitators and barriers to implementing computerized maintenance management systems in low-middle-income countries." *Health and Technology* 13, no. 3 (2023): 373-378. https://doi.org/10.1007/s12553-023-00743-5

[2] D'Orazio, Marco, Gabriele Bernardini, and Elisa Di Giuseppe. "Predict the priority of end-users' maintenance requests and the required technical staff through LSTM and Bi-LSTM recurrent neural networks." *Facilities* 41, no. 15/16 (2023): 38-51. https://doi.org/10.1108/F-07-2022-0093

[3] Chan, Albert PC, Francis KW Wong, Carol KH Hon, and Tracy NY Choi. "Construction of a Bayesian network model for improving the safety performance of electrical and mechanical (E&M) works in repair, maintenance, alteration and addition (RMAA) projects." *Safety science* 131 (2020): 104893. https://doi.org/10.1016/j.ssci.2020.104893

[4] D'Orazio, Marco, and Gabriele Bernardini. "Towards a technical sentiment lexicon for the maintenance of human-centred buildings." *TEMA* 9, no. 01 (2023). https://doi.org/10.30682/tema0901e

[5] Aery, M., and Chet Ram. "A review on machine learning: Trends and future prospects." *Research Cell: An International Journal of Engineering Sciences* 25, no. 63019 (2017): 89-96.

[6] Ventikos, Nikolaos P., and Konstantinos Louzis. "Developing next generation marine risk analysis for ships: Bio-inspiration for building immunity." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 237, no. 2 (2023): 405-424. https://doi.org/10.1177/1748006X221087501

[7] Tarımer, İ., and B. C. Karadağ. "Comparison with Classification Algorithms in Data Mining of a Fuel Automation System's Sales Data." *Global Economics Review* 1 (2020): 1-10. https://doi.org/10.31703/ger.2020(V-I).20

[8] Abdullah, Abdul Rani Achmed, Siti Zura A. Jalil, and Nik Nadzirah Nik Mohamed. "Determining Synonymy of Risk and Priority in Maintenance Prioritization.

[9] Gül, Muhammed, Ali Fuat Güneri, Fatih Yılmaz, and Oğuzhan Çelebi. "Analysis of the relation between the characteristics of workers and occupational accidents using data mining." *The Turkish Journal of Occupational/Environmental Medicine and Safety* 1, no. 4 (2016).

[10] Rafael, Camila, Mateus Vicente Peternella, Beatriz Lavezo dos Reis, Gislaine Camila Lapasini Leal, Rodrigo Clemente Thom de Souza, and Edwin Vladimir Cardoza Galdamez. "Occupational health and safety and data mining: a bibliometric analysis." *Revista Gestão da Produção Operações e Sistemas* 16, no. 2 (2021): 168-168. https://doi.org/10.15675/gepros.v16i2.2784