International Journal of Reliability, Risk and Safety: Theory and Application

Online ISSN: 2676-3346

Vol. 7/ Issue 2/ Year 2024/ pp. 62-70
DOI: 10.22034/IJRRS.2024.7.2.6
Received: 15 September 2024, Revised: 16 November 2024, Accepted: 16 November 2024

Available online at: https://www.ijrrs.com

Original Research Article

Developing a New Fuzzy Clustering Method for Equipment Maintenance

Armin Mokhtari[1], Seyed Hamed Moosavirad[1*], Siavash Bayat[2], and Alireza Eftekhari[3]

1. Department of Industrial Engineering, Faculty of Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

2. Electronics Research Institute, Sharif University of Technology, Tehran, Iran

3. Zarand Iranian Steel Company (ZISCO), Kerman, Iran

*** s.h.moosavirad@uk.ac.ir**

Abstract

Data can enhance equipment maintenance and asset management by providing predictive insights and minimizing downtime. Implementing data gathering and predictive maintenance systems is essential for improving reliability and cost efficiency. However, addressing challenges such as high implementation costs, data integration issues, and the need for skilled personnel is crucial for maximizing their benefits. Maintenance managers at a steel holding company in Iran, as a case study aimed to implement predictive maintenance but faced high costs for full implementation. Selecting a subset of equipment parts posed a complex decision-making problem, as eligibility needed to be based on maintenance criteria rather than traditional factors like price and location. To address this, we proposed a framework using machine learning to cluster equipment parts based on maintenance-related criteria. While clustering simplifies decision-making, it introduces uncertainty. To mitigate this, we represent each cluster with a trapezoidal fuzzy number. The Silhouette method is employed to determine the optimal number of clusters, followed by the K-means++ method for clustering. Our approach successfully grouped 201 equipment parts into seven clusters based on criteria such as importance, maintenance period, and daily working hours. Fuzzy logic is used to interpret the clusters, reducing uncertainty and ensuring that no equipment is overlooked.

Keywords: Clustering; Condition monitoring; Maintenance; Trapezoidal fuzzy numbers; Unsupervised machine learning.

1. Introduction

Maintenance is an important field that affects equipment downtime and efficiency [1]. Unfortunately, maintenance issues have attracted little attention in the manufacturing industries. Many factories consider repair costs unavoidable and carry out maintenance and repair operations in emergencies. Therefore, these maintenance activities can impose high costs on these factories, spiritual costs such as losing opportunities and credit, and material costs such as downtime, resource wasting, and machine lifetime reduction. Thus, repair expenses might rise from 15% to 70% of production costs depending on the industry [2]. Unnecessary and inappropriate activities incur one-third of all maintenance costs [3]. Therefore, maintenance and repair have great importance and application in the manufacturing industry.

The change in manufacturing circumstances and the increase in technologies developed many maintenance policies in the past few decades [4]. While different maintenance strategies, such as run-to-failure (R2F) and preventive maintenance (PM), are typically used, a newer method called predictive maintenance (PdM) can provide factual data on the actual state of machines [5].

Machine learning (ML) approaches can manage a large amount of multivariate data delivered by industrial systems [1]. Although sensors and sensor fusion techniques have had significant advantages, Internet of Things (IoT) applications brought new challenges in data analysis and data scale [6]. Therefore, strong predictive techniques for PdM applications could be provided using ML [1].

Some investigations utilized ML algorithms. For instance, a study determined the optimum maintenance policy using four multivariate methods, including

How to cite this article:
A. Mokhtari, S.H. Moosavirad, S. Bayat, and A. Eftekhari, “Developing a new fuzzy clustering method for equipment maintenance,” *International Journal of Reliability, Risk and Safety: Theory and Application*, vol. 7, no. 2, pp. 62-70, 2024, doi: 10.22034/IJRRS.2024.7.2.6.

numerical taxonomy (NT), principal component analysis (PCA), artificial neural network (ANN), and data envelopment analysis (DEA) [7]. Although ML algorithms such as support vector machine (SVM), random forest (RF), deep learning, and K-means have been utilized to design PdM applications, some areas in PdM and ML still need more attention [1]. Clustering has been used for a wide range of conditions, such as the largest gas field in Europe, considered a real-life instance of equipment maintenance [8]. It was reported that clustering algorithms could provide about 70% of savings in expenses in some industries [9].

Some researchers have considered clustering algorithms in maintenance. For instance, a taxonomy of PdM has been developed to allow the analysis and judgment of business models [10]. Another research project utilized a convolutional neural network (CNN) to cluster written maintenance records of mining excavator buckets [11]. Their approach recognized that similar clusters of maintenance records represent different equipment degradation states, and their projected model could improve maintenance [11].

Another study used a combination of fuzzy clustering and semantic technologies to predict the time of machinery failures and identify their criticality [12]. Different unsupervised ML algorithms such as Hierarchical clustering, fuzzy C-means, principal component analysis (PCA), T^2 statistic, and model-based clustering have been applied to the vibration data of an exhaust fan in another study [13]. They also benchmarked the algorithms and chose the best one using a proposed procedure [13].

By reviewing ML methods applied to PdM, Carvalho et al. Showed that different datasets, such as vibration, sensor, and historical data, are analyzed using ML algorithms [1]. Table 1 shows some of the recent research projects that utilized clustering techniques in sensor or failure data in maintenance problems. However, the clustering of equipment parts, which could be beneficial, especially for decision-making, has been overlooked.

The manufacturing industry has also confirmed the importance of decision-making in maintenance [14]. ML has a notable effect on maintenance decision-making systems that are automated and intelligently supported [6]. To improve their decision-making, firms are also seeking some approaches to using the power of big data (BD) [15].

Table 1. Using clustering algorithms in PdM

Reference	Method(s)	Objective
[16]	A novel semi-supervised clustering method	Prioritize asset rehabilitation using 2 years of failure data
[17]	DBSCAN	Analyzed of sensor data for monitoring the time evolution of the health status of an industrial machine, specifically a freeze dryer in a pharmaceutical plant.
[18]	K-means	Determine outliers and identify strange behaviors of the wind turbines.
[19]	K-means	Clustered the data obtained from acceleration sensors to predict the point of failure of the robot.
[20]	Simple Linear Iterative Clustering (SLIC)	Wind turbine blade regional cluster in for maintenance purposes
[21]	K-means	Monitoring-based maintenance decision-making models for subgrade settlement
[22]	K-means	Anomaly trend analysis to estimate the remaining useful life of bearings
[10]	Ward’s algorithm	Clustering of PdM business models
[11]	Convolutional Neural Network (CNN)	Maintenance records clustering
[9]	A developed method	Maintenance activities clustering
[12]	Fuzzy clustering	Clustering of machine historical maintenance data
[23]	K-PdM, a cluster-based hidden Markov model	Clustered KPI data for machinery deterioration estimation and remaining useful life (RUL) prediction.
[13]	PCA, Hierarchical clustering, K-Means, Fuzzy C-Means, Model-based clustering	Vibration data clustering for maintenance purposes
[24]	K-means	Sensor data clustering in maintenance
[25]	K-Nearest Neighbor (k-NN), SVM, k-Means	Clustering of vibration signals for maintenance
[26]	Kernel Spectral Clustering (KSC)	KSC is applied to sensor data to distinguish between normal operating conditions and abnormal situations.
[27]	A greedy heuristic and a branch and bound algorithm	Clustering of maintenance jobs

Many manufacturers are experiencing a digital transformation due to the Industry 4.0 paradigm, which facilitates advanced data-driven decision-making processes [28]. Many studies integrated MCDM techniques with fuzzy logic to optimize maintenance policy [4]. For instance, a study simulated feasible maintenance scenarios and ranked them using Fuzzy Simulation (FS) and Fuzzy Data Envelopment Analysis (FDEA) [29]. Their approach was beneficial for decision-makers to find the optimal scenario for planning maintenance activities [29]. However, the present article utilizes fuzzy logic differently to visualize and represent clusters and reduce uncertainty.

This research aims to propose a novel framework that first uses machine learning to divide 201 equipment parts into distinct clusters based on maintenance criteria instead of common unrelated criteria such as price, location, and distance. Moreover, since judging a cluster instead of a single equipment part leads to uncertainty and some parts may be overlooked, this research goal is to represent each cluster using fuzzy logic to avoid neglecting certain equipment components and mitigate the subsequent uncertainty of judging clusters.

The paper is structured as follows for addressing this aim: Section 2 describes the applied methods for clustering and visualization considering the specific challenges of clustering equipment parts. Section 3 explains the case study and is followed by section 4 which presents the results of this research. Finally, section 5 offers the conclusion and recommendations for maintenance managers and scholars.

2. Methodology

The proposed method clusters equipment components using ML to streamline and accelerate the decision-making procedure and depicts clusters using trapezoidal fuzzy figures to lessen the inherent uncertainty associated with narrowing down the decision-making alternatives. As shown in
Figure **1**, firstly, a scope was selected among different departments in one of the subsidiaries of a steel holding company in Iran. Having data gathered, the preprocessing stage is done with some challenges specific to this issue. Subsequently, parts are clustered with K-means++, and clusters are presented with trapezoidal fuzzy numbers.

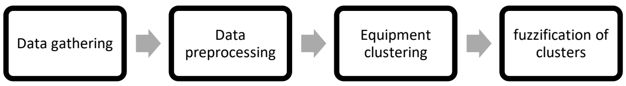

Figure 1. Methodology

2.1 Data Gathering

This stage includes defining maintenance-related criteria and gathering values of each equipment part in each criterion from the maintenance department of the company. Although different criteria, such as distance, price, and location, could be gathered for clustering, meaningful clusters are those that are distinguished by maintenance-related criteria.

Having maintenance-related criteria defined by the company's maintenance department, an actual dataset is prepared, including four columns for each part of the equipment, including (1) ID, including the equipment ID before the hyphen and the part ID after the hyphen, (2) the number of working hours per day, (3) the number of days between regular maintenance activities in current PM strategy, and (4) the importance of equipment in the production line in percentage. The head of the dataset can be seen in Table 2.

Table 2. Head of the dataset

ID	Working Hours (h)	PM period (Day)	Importance (%)
A2010-001	24	3650	80
A2010-002	10	365	80
A2015-001	8	3650	90
A2015-004	8	365	90
A2015-005	24	180	90

2.2 Data preprocessing

Preprocessing the data is an essential step in data mining, which often suffers from a lack of proper care [30]. In the preprocessing stage, we checked missing data and outliers and dealt with them accordingly. Finally, we scaled the dataset.

Since the company's maintenance department collaborated well on gathering data from all maintenance parts perfectly, there was no missing data in our dataset. Therefore, no action was needed and done to fill missing values.

The main challenge of this step was outliers. Although there are some significant outliers, our proposed method does not offer to change or eliminate outliers. The main reason is that clusters consist of parts that should not be eliminated later in maintenance decision-making. Hence, some equipment parts would not be considered in the decision-making stages if outliers were removed. If they had been manipulated, future maintenance decisions would have been based on unreal values that may have dire consequences. Ultimately, our decision was not to manipulate nor remove outliers to solve this challenge, and this was an essential difference between this research and other works.

The scikit-learn's MinMaxScaler can be used when attributes have various ranges of values in the dataset [31]. Thus, all values were scaled using Scikit-learn's MinMaxScaler in Python, which uses the equation (1).

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

2.3 Equipment clustering

In this stage, an ML clustering algorithm is used to divide equipment parts into distinct meaningful clusters based on maintenance-related criteria. As mentioned in section 2.2, outlier values should not be deleted. Thus, some algorithms, which are sensitive against outliers such as DBSCAN, are avoided in the clustering stage.

K-means is a popular unsupervised ML clustering algorithm to find k partitions of the dataset while close objects are in the same cluster and far ones are in different clusters [1]. This method is particularly useful when there is no prior knowledge about the data, making it suitable for various applications [32]. K-Means typically employs Euclidean Distance for calculations, which has been shown to provide better results compared to other distance measures [33]. Not only is K-means an easy method to understand and implement, but it also performs well and handles a large amount of data while the number of clusters is small [1].

K-means++, an improved version of K-means, is chosen to solve this problem because, in addition to all the mentioned benefits of K-means, it offers a simple, linear-time initialization algorithm that improves both the speed and accuracy of clustering compared to traditional K-means, making it a popular choice for various applications [34]. While the K-means method randomly selects the initial cluster centers, the K-means++ algorithm takes far data points as initial centers; thus, K-means++ is faster with fewer iterations [35].

In the K-means++, the number of clusters should be determined. Initially, the Elbow plot was used, yet as shown in Figure 2, the optimum number of clusters is impossible to recognize. Thus, the silhouette score (SSC) is chosen instead of the Elbow plot.

The SSC is a way to find the optimum number of clusters by comparing intra-cluster distances to inter-cluster distances. Consider a_i as the average variation of a member i to all other members inside its cluster and b_i as the average variation of member i to members inside the closest cluster [36]. Equation (2) shows how SSC is calculated [37].

$$\text{Silhouette score} = \frac{b-a}{\max(a,b)} \quad (2)$$

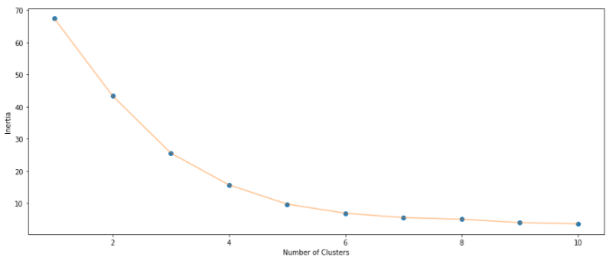

Figure 2. Elbow Plot

2.4 Fuzzification

Although ML clustering techniques can ease decision-making by converting many decision-making alternatives to a few distinct clusters, it is important to judge clusters fairly to reduce subsequent uncertainty of clustering. Judging a cluster of equipment parts based on a single crisp value is uncertain. The reason is that whether the value is mean, median, or mode, some equipment parts will be overlooked, which in maintenance can cause dire consequences. This article proposes utilizing trapezoidal fuzzy numbers to represent and visualize clusters. This approach, not only guarantees all equipment parts will be considered, but also the distribution of each cluster will not be overlooked.

A fuzzy subset of a real line is a fuzzy number where the maximum values of membership are clustered around the mean value, and the function of the membership is monotonic on both flanks [38]. In our visualization, the maximum membership values are between the first and third quartiles, called interquartile range (IQR), since the midspread of a dataset is the core part. The minimum value in each cluster is considered a1, and the maximum value is considered a_4. Therefore, clusters can be plotted, as shown in Figure 3.

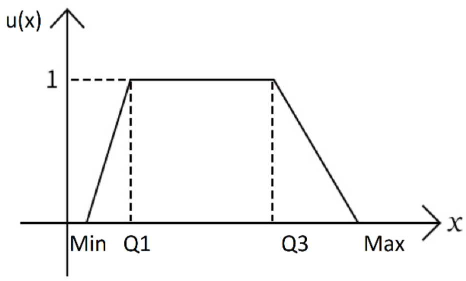

Figure 3. Plotting a cluster using a Trapezoidal Fuzzy Number

2.4.1 Definition

In order to represent a cluster with a trapezoidal fuzzy number, that number is defined as a fuzzy set like u when R→I = [0, 1] in which:

u is upper semi-continuous,
u(x) = 0 outside the range of the cluster,
u(x) is flat increasing on $[a_1, a_2]$,
$u(x) = 1,\ a_2 \leq x \leq a_3$,
u(x) is flat decreasing on $[a_3, a_4]$.
There are real numbers such that $a_1 \leq a_2 \leq a_3 \leq a_4$ and:
a1 is the minimum value of the cluster.
a2 is the lower quartile (Q_1) of the cluster.
a3 is the upper quartile (Q3) of the cluster.
a4 is the maximum value of the cluster.

Ultimately, when trapezoidal fuzzy numbers are calculated, clusters can also be explained using fuzzy linguistic variables.

3. Case Study

The case study of this research is a subsidiary of a steel holding company in Kerman, Iran, with more than 11000 shareholders and 22000 billion Rials of assets. In this case study, maintenance managers want to want to implement IoT and PdM.

However, maintenance costs must be considered since economic conditions typically limit the resources required for maintenance efforts [39]. Hence, the research is limited to the cooking process of the company, including 201 parts from 36 pieces of equipment. The process input is limestone, and the output is a form of lime called quicklime. The pieces of equipment are under traditional maintenance strategies, and implementation of IoT and PdM for all of them could be costly. Using machine learning to divide those parts into meaningful distinct clusters based on maintenance criteria makes it possible to determine the maintenance strategy for each group.

4. Results

The result of the clustering process utilizing K-means++ is illustrated in **Error! Reference source not found.**. As it is shown, there are seven distinctive clusters in the 3D environment.

The quantity of clusters is determined using the results shown in Table 3. The best number of clusters is seven, where the SSC is maximum.

Table 3. Silhouette score

Number of clusters	Silhouette score
2	0.527
3	0.512
4	0.576
5	0.658
6	0.685
7	**0.696**
8	0.644
9	0.656

Table 4 presents the values related to importance, PM period, and daily working hours of each cluster in the production line. Considering the importance criterion, clusters are clearly distinct except cluster 0, which has a wide range of importance from 10 to 85 percent. By utilizing fuzzy linguistic variables, it can be said that the importance of clusters 2, 4, and 5 is low (less than 40%), while that of clusters 1, 3, and 6 is high, up to 90 and 100 percent.

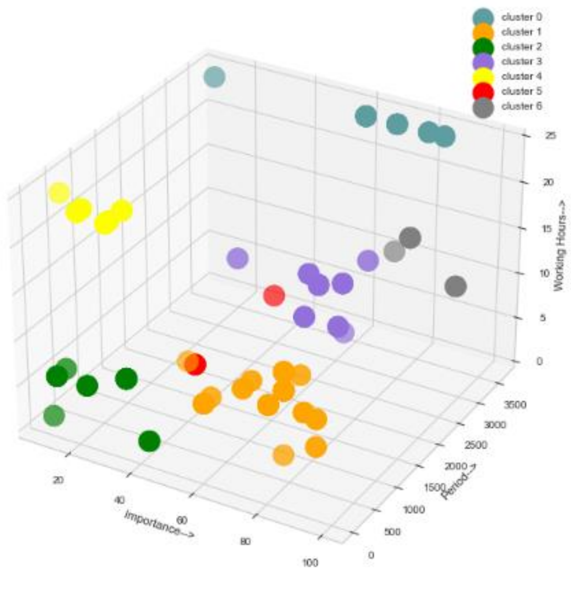

Figure 4. Clusters' 3D plot

Table 4. Trapezoidal Fuzzy Values of Clusters

Cluster	Importance (Percentage)	PM Period (Days)	Daily Working Hours (Hours)
0	(10, 60, 72.5, 85)	(3650, 3650, 3650, 3650)	(24, 24, 24, 24)
1	(50, 70, 80, 90)	(30, 30, 365, 730)	(3, 8, 10, 12)
2	(10, 10, 30, 40)	(90,180, 365, 365)	(1, 5, 6, 6)
3	(70, 90, 100, 100)	(30, 90, 180, 730)	(18, 20, 24, 24)
4	(10, 20, 30, 30)	(30, 90, 365, 365)	(24, 24, 24, 24)
5	(30, 30, 30, 30)	(1825, 1825, 3650, 3650)	(1,1,1,1)
6	(70, 70, 90, 90)	(3650, 3650, 3650, 3650)	(8,8,12,12)

As PM period values show, equipment parts of clusters 2 and 4 are undergoing maintenance activities every year or sooner, while the value goes up to two years for clusters 1 and 3. With more than five years of PM period, clusters 0, 5, and 6 do not need frequent maintenance activities.

Regarding daily working hours, with one hour per day, cluster 5 working hours is extremely low. Cluster 1 has 3 to 12 working hours which can be considered low. Cluster 6 has average working hours from 8 hours to 12 hours. Finally, the working hours of cluster 3 are high (18-24 hours) while those of cluster 4 are extremely high at 24 hours a day which means cluster 4 members are working non-stop.

The fuzzy values of importance, PM period, and daily working hours are depicted in

Figure **4**, Figure 5, and Figure 6.

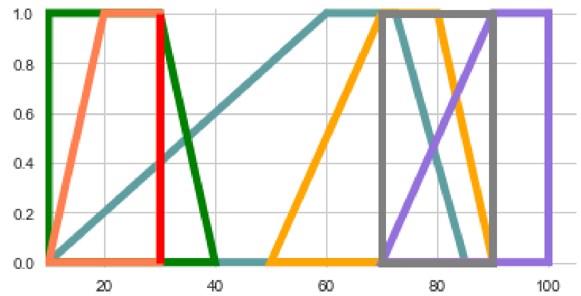

Figure 4. Importance of Clusters

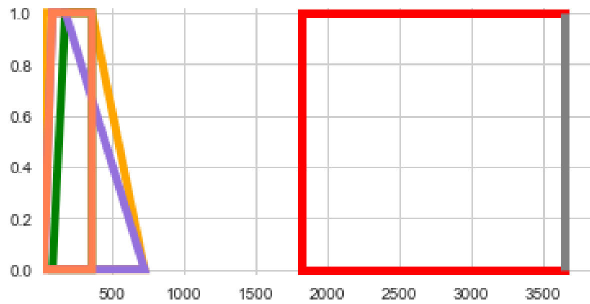

Figure 5. PM period of Clusters

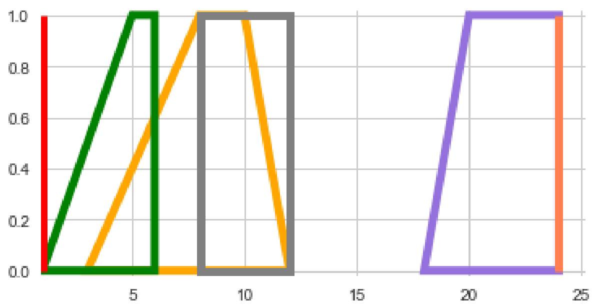

Figure 6. Daily Working Hours of Clusters

As a result, given information can be helpful to explain and solve the problem with fuzzy linguistic variables. Taking cluster 6 as an instance, the importance of the parts is (70, 70, 90, 90) percent, their maintenance period is (3650, 3650, 3650, 3650) days, and they work (8, 8, 12, 12) hours. For the sixth cluster, parts' importance and working hours are respectively high and average, while the PM period is extremely high. Thus, it can be concluded that since those parts can work for ten years without maintenance activities, they may not need PdM, even though they are highly important.

5. Conclusion

Unlike previous studies focusing on clustering maintenance records, signals, or activities, this article clusters equipment parts in an industrial case study based on maintenance-related criteria rather than traditional factors such as price or location. A novel aspect of this approach is the representation and visualization of clusters using trapezoidal fuzzy numbers. This method not only provides more detailed information about each cluster, including the maximum, minimum, and range of values but also reduces uncertainty by ensuring that no cluster members are overlooked.

Long implementation and learning time in maintenance systems is a reason that reduces the practical value for real utilization [40]. The current approach significantly reduces the number of alternatives from 201 to just 7. This change decreases the time required for analysis, simplifies comparisons, and enhances the uniqueness of data visualizations. Additionally, this framework is not limited to Predictive Maintenance (PdM); it can also be applied to other areas such as Project Management (PM) and Return to Function (R2F). As long as the criteria are relevant to the problem, the proposed method can effectively handle all data types, even when the data is not numerical.

Mathematical programming in maintenance is becoming more complicated due to the expanding system size and the number of machines [41]. The current approach simplifies mathematical programming and decision-making in maintenance, particularly when dealing with numerous components akin to big data challenges. Maintenance managers can use this method to make various strategic decisions more effectively. By comparing different aspects of a few selected clusters, they can easily prioritize tasks. Additionally, maintenance engineers can better visualize the components since only a limited number of clusters must be represented instead of an overwhelming number of individual parts.

Maintenance managers can hire the proposed framework to find the most eligible subset of equipment for different decision-making problems based on related criteria. The framework can be integrated with mathematical programming and decision-making techniques for future studies since one of our objectives was to create a framework that facilitates the use of these methods. Researchers can utilize this proposed approach to tackle challenges related to big data and high-dimensional datasets. Additionally, new clustering algorithms can be replaced and benchmarked by researchers in the future. Moreover, ranking fuzzy numbers can have a significant role in linguistic decision-making [42]. Future studies could also explore different methods for ranking resulting trapezoidal fuzzy numbers.

Acknowledgments

We want to express our sincere gratitude to Mr. Abdolreza Pourvaziri Nasab for his invaluable assistance in collecting the data for this research.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interests

The authors have expressed no conflict of interest.

6. References

[1] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, 2019, Art. no. 106024, doi: https://doi.org/10.1016/j.cie. 2019.106024.

[2] M. Bevilacqua and M. Braglia, "The analytic hierarchy process applied to maintenance strategy selection," *Reliability Engineering and System Safety*, vol. 70, no. 1, pp. 71-83, 2000, doi: https://doi.org/10.1016/S0951-8320(00)00047-8.

[3] R. K. Mobley, *An Introduction to Predictive Maintenance*, 2nd Ed. Oxford, UK: Butterworth-Heinemann, 2002. doi: https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3.

[4] S.-H. Ding and S. Kamaruddin, "Maintenance policy optimization—literature review and directions," *The International Journal of Advanced Manufacturing Technology*, vol. 76, pp. 1263-1283, 2015, doi: https://doi.org/10.1007/s00170-014-6341-2.

[5] R. K. Mobley, *An Introduction to Predictive Maintenance*, USA: Elsevier, 2002.

[6] C. J. Turner, C. Emmanouilidis, T. Tomiyama, A. Tiwari, and R. Roy, "Intelligent decision support for maintenance: An overview and future trends," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 10, pp. 936-959, 2019, doi: https://doi.org/10.1080/0951192X.2019.1667033.

[7] A. Azadeh, M. Sheikhalishahi, and F. Monshi, "Selecting optimum maintenance activity plans by a unique simulation-multivariate approach," *International Journal of Computer Integrated Manufacturing*, vol. 29, no. 2, pp. 1-15, 2015, doi: https://doi.org/10.1080/0951192X.2014.1003409.

[8] B. De Jonge, W. Klingenberg, R. Teunter, and T. Tinga, "Reducing costs by clustering maintenance activities for multiple critical units," *Reliability Engineering and System Safety*, vol. 145, pp. 93-103, 2016, doi: https://doi.org/10.1016/j.ress.2015.09.003.

[9] S. Bazeli and M. S. Fallahnezhad, "Clustering of condition-based maintenance considering perfect and imperfect actions," *International Journal of Reliability, Risk and Safety: Theory and Application*, vol. 3, no. 1, pp. 69-76, 2020, doi: https://doi.org/10.30699/IJRRS.3.1.8.

[10] J. Passlick, S. Dreyer, D. Olivotti, L. Grützner, D. Eilers, and M. H. Breitner, "Predictive maintenance as an internet of things enabled business model: A taxonomy," *Electronic Markets*, vol. 31, pp. 67–87, 2020, doi: https://doi.org/10.1007/s12525-020-00440-5.

[11] Z. Yang, P. Baraldi, and E. Zio, "A novel method for maintenance record clustering and its application to a case study of maintenance optimization," *Reliability Engineering & System Safety*, vol. 203, 2020, Art. no. 107103, doi: https://doi.org/10.1016/j.ress.2020.107103.

[12] Q. Cao, A. Samet, C. Zanni-Merk, F. D. B. De Beuvron, and C. Reich, "An Ontology-based Approach for Failure Classification in Predictive Maintenance Using Fuzzy C-means and SWRL Rules," *Procedia Computer Science*, vol. 159, pp. 630–639, 2019, doi: https://doi.org/10.1016/j.procs.2019.09.218.

[13] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *5th International Conference on Industrial Engineering and Applications*, 2018, pp. 355–361, doi: https://doi.org/10.1109/IEA.2018.8387124.

[14] J. Thor, S.H. Ding, and S. Kamaruddin, "Comparison of multi criteria decision making methods from the maintenance alternative selection perspective," *International Journal of Engineering and Science*, vol. 2, no. 6, pp. 27-34, 2013.

[15] M. Janssen, H. van der Voort, and A. Wahyudi, "Factors influencing big data decision-making quality," *Journal of Business Research*, vol. 70, pp. 338–345, 2017, doi: https://doi.org/10.1016/j.jbusres.2016.08.007.

[16] R. Beig Zali, M. Latifi, A. A. Javadi, and R. Farmani, "Semisupervised clustering approach for pipe failure prediction with imbalanced data set," *Journal of Water Resources Planning and Management*, vol. 150, no. 2, 2024, Art. no. 04023078, doi: https://doi.org/10.1061/JWRMD5.WRENG-6263.

[17] E. Oliosi, G. Calzavara, and G. Ferrari, "On sensor data clustering for machine status monitoring and its application to predictive maintenance," *IEEE Sensors Journal*, vol. 23, no. 9, pp. 9620–9639, 2023, doi: https://doi.org/10.1109/jsen.2023.3260314.

[18] P. C. Rodriguez, P. Marti-Puig, C. F. Caiafa, M. Serra-Serra, J. Cusidó, and J. Solé-Casals, "Exploratory analysis of scada data from wind turbines using the k-means clustering algorithm for predictive maintenance purposes," *Machines 2023, Vol. 11, Page 270*, vol. 11, no. 2, p. 270, 2023, doi: https://doi.org/10.3390/machines11020270.

[19] S. Predictive Maintenance *et al.*, "Predictive maintenance system for wafer transport robot using k-means algorithm and neural network model," *Electronics, Vol. 11, Page 1324*, vol. 11, no. 9, 2022, Art. no. 1324, doi: https://doi.org/10.3390/electronics11091324.

[20] [20] Jack. W. Barker, N. Bhowmik, and Toby. P. Breckon, "Semi-Supervised Surface Anomaly Detection of Composite Wind Turbine Blades From Drone Imagery," in *17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Volume 4: VISAPP*, 2022, pp. 868-876, doi: https://doi.org/10.5220/0010842100003124.

[21] Z. Duan, C. Gui, and Y. Hou, "Monitoring-based maintenance decision-making models for subgrade settlement," in *Green and Intelligent Technologies for Sustainable and Smart Asphalt Pavements*, X. Liu et al. Ed. London: CRC Press, 2021, pp. 655-661.

[22] P. V. Kamat, R. Sugandhi, and S. Kumar, "Deep learning-based anomaly-onset aware remaining useful life estimation of bearings," *PeerJ Computer Science*, vol. 7, p. e795, 2021, Art. no. e795, doi: https://doi.org/10.7717/peerj-cs.795.

[23] Z. Wu *et al.*, "K-PdM: KPI-Oriented Machinery Deterioration Estimation Framework for Predictive Maintenance Using Cluster-Based Hidden Markov Model," *IEEE Access*, vol. 6, pp. 41676-41687, 2018, doi: https://doi.org/10.1109/ACCESS.2018.2859922.

[24] E. Uhlmann, R. P. Pontes, C. Geisert, and E. Hohwieler, "Cluster identification of sensor data for predictive maintenance in a Selective Laser Melting machine tool," *Procedia Manufacturing*, vol. 24, pp. 60-65, 2018, doi: https://doi.org/10.1016/j.promfg.2018.06.009.

[25] G. K. Durbhaka and B. Selvaraj, "Predictive maintenance for wind turbine diagnostics using vibration signal analysis based on collaborative recommendation approach," in *International Conference on Advances in Computing, Communications and Informatics*, 2016, pp. 1839-1842. doi: https://doi.org/10.1109/ICACCI.2016.7732316.

[26] R. Langone, C. Alzate, B. De Ketelaere, J. Vlasselaer, W. Meert, and J. A. K. Suykens, "LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines," *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 268-278, 2015, doi: https://doi.org/10.1016/j.engappai.2014.09.008.

[27] G. Van Dijkhuizen and A. Van Harten, "Optimal clustering of frequency-constrained maintenance jobs with shared set-ups," *European Journal of Operational Research*, vol. 99, no. 3, pp. 552-564, 1997, doi: https://doi.org/10.1016/S0377-2217(96)00320-7.

[28] S. Arena, E. Florian, F. Sgarbossa, E. Sølvsberg, and I. Zennaro, "A conceptual framework for machine learning algorithm selection for predictive maintenance," *Engineering Applications of Artificial Intelligence*, vol. 133, part D, , 2024, Art. no. 108340, doi: https://doi.org/10.1016/j.engappai.2024.108340.

[29] A. Azadeh, M. Sheikhalishahi, S. M. Khalili, and M. Firoozi, "An integrated fuzzy simulation–fuzzy data envelopment analysis approach for optimum maintenance planning,", *International Journal of Computer Integrated Manufacturing*, vol. 27, no. 2, pp. 181–199, 2013, doi: *https://doi.org/10.1080/0951192X.2013.812804*.

[30] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Cham: Springer, 2015.

[31] M. Ghiassi, H. Saidane, and R. Oswal, "YAC2: An α-proximity based clustering algorithm," *Expert Systems with Applications*, vol. 167, 2021, Art. no. 114138, doi: https://doi.org/10.1016/j.eswa.2020.114138.

[32] A. Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Teknika*, vol. 18, no. 1, pp. 76–82, 2016, doi: https://doi.org/10.18196/st.v18i1.708.

[33] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, "Tinjauan pustaka sistematis pada data mining: studi kasus algoritma k-means clustering," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, pp. 52-61, 2023, doi: https://doi.org/10.21063/jtif.2022.V10.2.52-61.

[34] S. Vassilvitskii, "K-means: algorithms, analyses, experiments," Ph.D. dissertation, Stanford University, Department, University, California, USA, 2007., 2007.

[35] A. Kapoor and A. Singhal, "A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms," in *3rd International Conference on Computational Intelligence and Communication Technology (CICT)*, 2017, pp. 1-6. doi: https://doi.org/10.1109/CIACT.2017.7977272.

[36] T. Weißer, T. Saßmannshausen, D. Ohrndorf, P. Burggräf, and J. Wagner, "A clustering approach for topic filtering within systematic literature reviews," *MethodsX*, vol. 7, 2020, Art. no. 100831, doi: https://doi.org/10.1016/j.mex.2020.100831.

[37] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987, doi: https://doi.org/10.1016/0377-0427(87)90125-7.

[38] D. Dubois and H. Prade, "Operations on fuzzy numbers," *International Journal of Systems Science*, Vol. 9, no. 6, pp. 613-626, 1978, doi: http://dx.doi.org/10.1080/00207727808941724.

[39] Y.F. Niu, R.H. Zhou, X.Z. Xu, and H.Y. Xiang, "A reliability index to measure multi-state flow network considering capacity restoration level and maintenance cost," *Reliability Engineering and System Safety*, vol. 250, 2024, Art. no. 110209, doi: https://doi.org/10.1016/j.ress.2024.110209.

[40] Y. Peng, M. Dong, and M. J. Zuo, “Current status of machine prognostics in condition-based maintenance: a review,” *The International Journal of Advanced Manufacturing Technology*, vol. 50, pp. 297-313, 2010, doi: https://doi.org/10.1007/s00170-009-2482-0.

[41] Y.-S. Shum and D.-C. Gong, “The application of genetic algorithm in the development of preventive maintenance analytic model,” *The International Journal of Advanced Manufacturing Technology*, vol. 32, pp. 169-183, 2007, doi: https://doi.org/10.1007/s00170-005-0314-4.

[42] S. Abbasbandy and T. Hajjari, “A new approach for ranking of trapezoidal fuzzy numbers,” *Computers and Mathematics with Applications*, vol. 57, no. 3, pp. 413-419, Feb. 2009, doi: https://doi.org/10.1016/j.camwa.2008.10.090.