Published *online* on the journal's web page: http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Enhancing Maintenance Efficiency Through K-Means Clustering at PT Semen Indonesia

Muhammad Fadhil Alviano[1], Amalia Nur Alifah[2], Calista Ghea Ardhani[3], Helga Fadhil Raditya[4], Harashta Tatimma Larasati[5,6]

[1, 2, 3, 4]*School of Computing, Telkom University, Telekomunikasi Street, Terusan Buah Batu, Bandung, 40257, Indonesia*
[5]*School of Electrical Engineering and Informatics, Bandung Institute of Technology, Ganesha Street, Bandung, 40132, Indonesia*
[6]*School of Computer Science and Engineering, Pusan National University, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241, South Korea*

[1]*fadhilalviano@student.telkomuniversity.ac.id*, [2]*amalialifah@telkomuniversity.ac.id*, [3]*calistaghea@student.telkomuniversity.ac.id*, [4]*helgaafr@student.telkomuniversity.ac.id*, [5,6]*harashta@pusan.ac.kr*

ARTICLE INFORMATION

Article History:
Received: August 27, 2024
Last Revision: September 26, 2024
Published *Online*: September 30, 2024

KEYWORDS

K-Means,
Clustering,
Maintenance Management,
Data Mining,
PT Semen Indonesia

CORRESPONDENCE

Phone: +6285710100017
E-mail: *amalialifah@telkomuniversity.ac.id*

ABSTRACT

PT Semen Indonesia, a leading cement manufacturer, faces the challenge of optimizing maintenance costs and improving operational efficiency across its various plants. Maintenance efficiency is critical in minimizing downtime, reducing operational costs, and maintaining competitiveness in the highly demanding industrial sector. This study addresses these challenges by applying K-Means clustering to the company's maintenance data. By analyzing patterns in maintenance frequency, total costs, and maintenance duration across their various plants, the company can identify work units that require more intensive attention or that can be optimized for greater efficiency. To achieve this, K-Means clustering analysis applied to gain deeper insights into the maintenance data, identifying patterns that can help improve operational efficiency and develop more targeted maintenance strategies based on the identified clusters. The analysis revealed two distinct clusters: Cluster 1, which contains 29 planner groups, and Cluster 2, which consists of 11 planner groups. Planner groups in Cluster 1 exhibit fewer maintenance activities, lower maintenance costs, and shorter task durations, indicating more efficient and consistent maintenance processes. In contrast, Cluster 2 includes planner groups with higher maintenance frequencies, greater costs, and longer task durations, suggesting inefficiencies in their operations. PT Semen Indonesia can use these insights to focus on improving the performance of groups in Cluster 2 by optimizing resource allocation, reducing operational inefficiencies, and enhancing overall maintenance strategies.

1. INTRODUCTION

PT Semen Indonesia (Persero) is an industrial company that was inaugurated on August 7, 1957, in Gresik by Ir. Soekarno, originally named Semen Gresik. The company initially had only one plant in Gresik, East Java, but has since grown rapidly into a parent company overseeing several subsidiaries. PT Semen Indonesia is committed to implementing healthy and ethical business practices and consistently upholding Good Corporate Governance (GCG) in all its business activities and operations. As part of its commitment to good governance, the company operates several plants across various regions in Indonesia, including Gresik, Tuban, Padang, and Tonasa. The main products of PT Semen Indonesia include Portland Cement Type I, Portland Composite Cement (PCC), and Portland Pozzolan Cement (PPC). Additionally, the company offers various specialized cement products for more specific construction needs [1].

PT Semen Indonesia (Persero), as one of the largest cement producers in Southeast Asia, operates across multiple plants with complex supply chains and production processes. The company faces the ongoing challenge of maintaining operational efficiency while managing extensive maintenance needs. Effective maintenance is critical for minimizing downtime, reducing operational costs, and ensuring the reliable functioning of machinery and equipment across various locations [2]. By optimizing maintenance processes, the company can not only improve its operational efficiency but also enhance its competitive edge in the highly demanding industrial sector [3].

However, managing maintenance at this scale requires not only robust maintenance strategies but also the ability to analyze large volumes of data to identify inefficiencies. With increasing pressure to reduce costs and improve resource allocation, PT Semen Indonesia needs an advanced, data-driven approach to tackle these challenges. This study focuses on applying K-Means clustering, a widely used data mining technique, to analyze maintenance data from PT Semen Indonesia. The aim is to identify patterns within the data that can be used to optimize maintenance processes, reduce costs, and improve overall operational performance. Through clustering, this research seeks to segment planner groups based on maintenance activities, costs, and task durations, providing actionable insights that will help PT Semen Indonesia streamline its operations [4].

The primary objective of this study is to apply K-Means clustering to analyze maintenance data at PT Semen Indonesia, with the aim of identifying and optimizing inefficient maintenance practices. By leveraging this data-driven approach, the research seeks to classify planner groups into high- and low-performing clusters based on variables such as maintenance frequency, total costs, and task duration. The expected outcomes include a clear identification of planner groups that excel in operational efficiency and those that require further improvement. These insights will provide a foundation for strategic enhancements in maintenance management, enabling PT Semen Indonesia to allocate resources more effectively, reduce operational costs, and improve overall maintenance performance

2. RELATED WORK

Data mining is the process of extracting knowledge by examining large databases to uncover new patterns [5]. This process involves six iterative and adaptive cycles: the Business Understanding Phase, which focuses on understanding business objectives and needs; the Data Understanding Phase, where relevant data is identified and explored; the Data Preparation Phase, which involves cleaning and processing the data; the Modeling Phase, where data models are built; the Evaluation Phase, which assesses the accuracy and effectiveness of the models; and the Deployment Phase, where the models are implemented in the operational environment [6]. The insights gained from data mining can be utilized by companies to make more accurate and strategic decisions.

Cluster analysis, or data clustering, is a technique used to identify patterns of relationships within data that share similarities, with the goal of uncovering patterns that are not immediately obvious [7]. Common clustering methods include k-means, hierarchical clustering, and density-based clustering. The primary benefit of clustering for businesses lies in its ability to provide a deep understanding of data by simplifying the analysis of patterns, trends, and relationships among data points. It also facilitates segmentation, which aids in the development of more effective marketing strategies [8].

Clustering methods can be categorized into six types: partitional clustering, which divides data into a set number of non-overlapping subsets; hierarchical clustering, which creates a hierarchy of clusters either through agglomerative (bottom-up) or divisive (top-down) approaches; density-based clustering, which forms clusters based on the density of data points; grid-based clustering, which transforms data into a finite grid structure for analysis; model-based clustering, which involves constructing a model for each cluster; and constraint-based clustering, which forms clusters by considering specific constraints such as time or geographic distance[9] [10]. The benefits of cluster analysis include identifying patterns of relationships within data, exploratory analysis, dimensionality reduction, anomaly detection, market segmentation, and customer profiling [11]. However, this technique also has drawbacks, such as sensitivity to initial conditions and the number of clusters, sensitivity to noise and outliers, challenges in interpretation, high computational costs, and analysis results that are influenced by the chosen algorithm [12].

K-Means is a straightforward approach for partitioning a dataset into K distinct and non-overlapping clusters. It is considered one of the most powerful and popular data mining algorithms [13]. The K-Means algorithm relies on the value of K to perform any clustering analysis. During the initialization process, the user must specify the number of clusters in the dataset. Clustering with different values of K will ultimately yield different results. Although K-Means is a widely used algorithm, it also has several limitations [14][15]. These include difficulties in determining initial clusters when data is sparse, the necessity of specifying the number of clusters before performing calculations, and the potential variability in clustering outcomes when the data is input differently, which means the true clusters may not be accurately identified [16].

The K-Means clustering method has not yet been widely applied in the field of industrial maintenance management. However, K-Means clustering has been effectively utilized in various other industrial applications. For instance, it has been used to optimize production processes, segment customer bases, and analyze supply chain performance [4], [17]. These examples demonstrate the versatility and potential of K-Means clustering in enhancing operational efficiency and decision-making within different industrial contexts. The successful implementation of K-Means clustering in these areas suggests that it could also offer significant benefits if applied to industrial maintenance management, providing new insights and strategies for improving maintenance operations.

3. METHODOLOGY

The methodology of this study is designed to systematically analyze maintenance data at PT Semen Indonesia using K-Means clustering, a robust and widely used data mining technique. This approach is employed to identify patterns and groupings within the maintenance data that can reveal inefficiencies and opportunities for optimization. The study begins by collecting and preprocessing relevant data, ensuring that it is clean, normalized, and suitable for analysis [18]. The K-Means clustering algorithm is then applied to segment the data into distinct clusters, each representing a group of planner units with similar maintenance performance characteristics. The clustering results are subsequently analyzed to derive actionable insights that can inform strategic decisions in maintenance management. This methodology not only provides a structured approach to data analysis but also aligns with the overall objective of enhancing operational efficiency within the company.

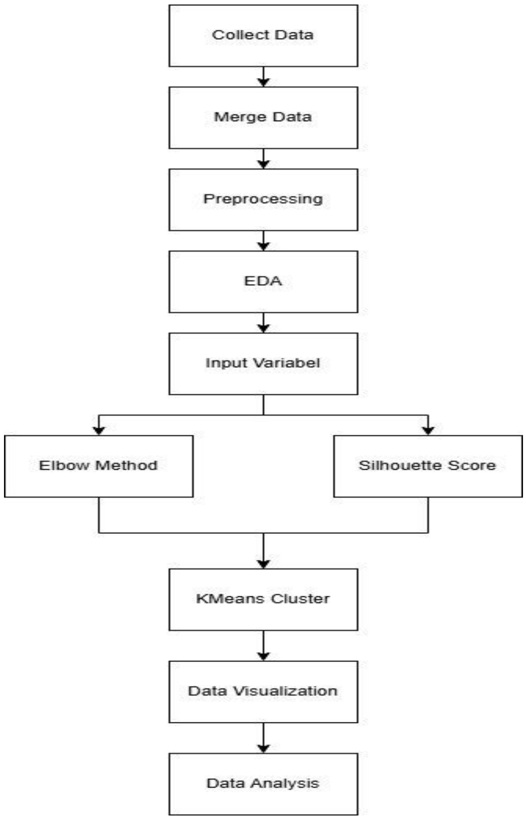

FIGURE 1. FLOWCHART OF DATA ANALYSIS

In figure 1, the flowchart presented outlines a systematic approach to data analysis using K-Means clustering. The process begins with Collect Data, where relevant data is gathered from PT Semen Indonesia. This is followed by the Merge Data step, which involves combining datasets into a cohesive structure suitable for analysis. Next, the Preprocessing phase is undertaken, where the data is cleaned, normalized, and prepared for further analysis [19]. Following this, Exploratory Data Analysis (EDA) is conducted to understand the underlying patterns, distributions, and relationships within the data. The identified Input Variables are then selected, which will be used as features in the clustering process.

During the Exploratory Data Analysis (EDA) stage, several techniques were employed to better understand the underlying patterns within the maintenance data, such as descriptive statistics, visualization, and correlation analysis. Descriptive statistics, such as mean, median, and standard deviation, were calculated to identify central tendencies and variability across variables like the number of maintenance tasks, total costs, and average task duration. Visualizations, including histograms and boxplots, were used to explore the distribution of the data and to detect any skewness or presence of outliers. Additionally, a correlation matrix was generated to assess the relationships between key variables, providing insights into potential interdependencies that could influence maintenance performance.

To determine the optimal number of clusters, two methods are employed: the Elbow Method and the Silhouette Score [19]. The Elbow Method helps to identify the point where adding more clusters does not significantly improve the model, while the Silhouette Score measures how well-separated the clusters are. Once the optimal number of clusters is determined, the K-Means Clustering algorithm is applied to group the data into distinct clusters. The results of this clustering process are then represented visually through Data Visualization, allowing for easier interpretation of the clusters. Finally, Data Analysis is performed to extract insights and inform strategic decisions based on the clustered data.

In this study, the K-Means clustering algorithm was selected due to its simplicity, efficiency, and effectiveness in handling large datasets, making it particularly suitable for analyzing the extensive maintenance data of PT Semen Indonesia. The K-Means clustering was implemented using Python, a powerful programming language for data analysis, with specific parameters such as the number of clusters (K) and maximum iterations carefully chosen to ensure the robustness and accuracy of the clustering results. This approach allowed for a clear segmentation of planner groups, facilitating more targeted and strategic maintenance management.

3.1 Data Preprocessing

The data analyzed originates from PT Semen Indonesia and pertains to maintenance spare parts, with a focus on four key columns. The first column, "Planner Group," identifies the work units responsible for ordering and utilizing the necessary spare parts. The second column, "Amount Maintenance”, records the total number of maintenance tasks performed by each planner group. The third column, "Total Amount," reflects the overall costs incurred for spare parts over a specific period. Lastly, the "Average Duration" column indicates the average duration of maintenance tasks conducted. This data, primarily consisting of integer values, is well-suited for clustering analysis using methods like K-Means. The goal of this analysis is to group planner groups based on emerging patterns in maintenance frequency, total costs, and average task duration. This approach aims to provide deeper insights into maintenance practices and spare part expenditures at PT Semen Indonesia, with the potential to enhance operational efficiency and management within this critical industry.

The selected features provide essential information for performing clustering analysis using methods such as K-Means [20]. The selection of variables for clustering was based on their relevance to the objectives of the study, which aimed to optimize maintenance processes. After conducting an initial correlation analysis, key features such as the number of maintenance tasks, total maintenance costs, and average task duration were chosen due to their significant impact on operational efficiency. Variables that showed little variation or weak correlation with the main performance metrics were excluded to simplify the model and improve clustering performance.

Outliers are data points that significantly deviate from the majority of the data in a dataset and can often skew results [21]. In the dataset from PT Semen Indonesia, certain values were identified as extreme outliers, which could negatively impact the clustering process. To address this, these outliers were removed from the data to ensure more accurate and reliable cluster formation.

Standardization is a process that ensures consistency and comparability across different data points. For the PT Semen Indonesia dataset, standardization was performed to normalize the data within a range of 0 to 1 [22]. This step is crucial for the clustering process, as it ensures that the data is neatly scaled and can be easily accessed and interpreted by others. The standardization method used in this research is min max standardization which given in equation 1 with:

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

X' : the standardized value of the original data point X. It is the value after it has been scaled to a range between 0 and 1.

X : the original value of the data point before standardization.

X_{min}: the minimum value in the dataset for the feature being standardized.

X_{max}: the maximum value in the dataset for the feature being standardized.

3.2 Elbow Method and Silhouette Score

Elbow Method is a technique used to determine the optimal number of clusters (K) in a clustering algorithm, such as K-Means [23]. The method involves plotting the Within-Cluster Sum of Squares (WCSS) against the number of clusters and then finding the "elbow point" where the WCSS starts to diminish at a slower rate. The idea is that adding more clusters beyond this point does not significantly improve the model, indicating the optimal number of clusters. The Within-Cluster Sum of Squares (WCSS) is calculated as equation 2 follows:

$$WCSS = \sum_{i=1}^{K} \sum_{x \in C_i} (x - \mu_i)^2 \quad (2)$$

Where:

K : the number of clusters

C_i : the set of points in the i-th cluster

x : the data points within cluster C_i

μ_i : the centroid of cluster C_i

$(x - \mu_i)^2$: the squared Euclidean distance between a data point x and its cluster centroid μ_i

While Silhouette Score is a measure of how well-defined the clusters are. It quantifies how similar a data point is to its own cluster compared to other clusters [24]. The score ranges from -1 to 1, where a higher score indicates that the data points are well-clustered and separated from other clusters, while a score close to 0 indicates overlapping clusters. Negative values suggest that data points might have been assigned to the wrong cluster. The Silhouette Score for a data point i is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

Where:

$a(i)$: the average distance between the data point i and all other points in the same cluster (intra-cluster distance).

$b(i)$: the average distance between the data point i and all points in the nearest cluster that is not the point's own cluster (inter-cluster distance).

$$Silhouette\ Score = \frac{1}{n} \sum_{i=1}^{n} s(i) \quad (4)$$

The overall Silhouette Score for the clustering solution is the average of the Silhouette Scores of all individual points as equation 4 where n is the total number of data points.

3.3 K-Means Clustering

The K-Means clustering algorithm begins with the Initialization phase, where the number of clusters, K, is determined by the user. Initially, K data points are selected randomly from the dataset to serve as the starting centroids. These centroids represent the initial center of each cluster. Following initialization, the algorithm enters the Assignment Step. In this phase, the algorithm calculates the distance between each data point and the K centroids. The distance metric used in this research is the Euclidean distance, which measures the straight-line distance between two points in space. For two points $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ in an n-dimensional space, the Euclidean distance is given by equation 3 [22].

$$Distance(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (3)$$

Once all data points have been assigned to a cluster, the algorithm proceeds to the Update Step. Here, the centroid of each cluster is recalculated by taking the mean of all data points within that cluster. This new centroid represents the updated center of the cluster and is expected to be more accurate as it reflects the average position of the cluster's members. The algorithm continues to iterate between the Assignment and Update steps, with each iteration refining the clusters and their centroids. This process repeats until a Convergence Criterion is met. Convergence typically occurs when the centroids no longer move significantly between iterations, indicating that the clusters have stabilized. Alternatively, convergence can be declared if the assignments of data points to clusters do not change or if a predetermined maximum number of iterations is reached [25]. The final output of the K-Means algorithm consists of the identified centroids and the corresponding clusters, where each data point is grouped according to its proximity to the nearest centroid. This results in a partitioned dataset where each cluster is distinct and non-overlapping, providing valuable insights into the underlying structure of the data.

4. RESULT AND DISCUSSION

This section presents the key findings from the K-Means clustering analysis applied to the maintenance data of PT Semen Indonesia. The analysis aimed to identify distinct planner groups based on their maintenance practices, with the goal of optimizing operational efficiency and resource management. By exploring the characteristics of the resulting clusters, we can gain insights into the performance of different planner groups, highlighting both strengths and areas for improvement. The discussion will delve into the implications of these findings for maintenance management, providing strategic recommendations for enhancing efficiency.

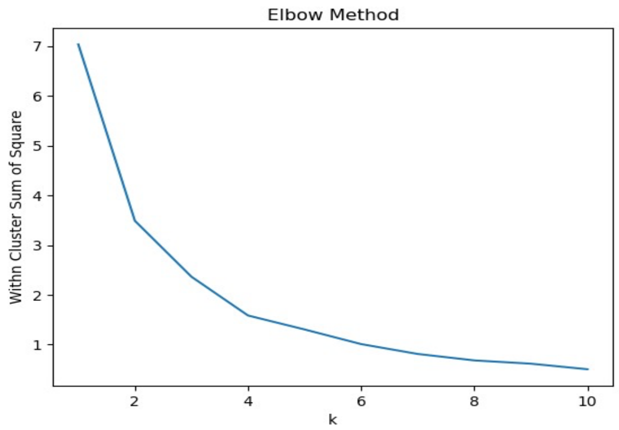

FIGURE 2. ELBOW METHOD

The optimal number of clusters, K, for the K-Means clustering analysis was determined to be 2. This selection was based on a combination of the Elbow Method and the Silhouette Score. The Elbow Method, as shown in Figure 2, was applied to plot the Within-Cluster Sum of Squares (WCSS) against different values of K, revealing a distinct "elbow" at K=2, where the rate of decrease in WCSS begins to level off. This indicates that adding more clusters beyond this point yields diminishing returns in terms of improved clustering.

TABLE 1. SILHOUETTE SCORE

K	Silhouette Score
2	0.52275
3	0.52175
4	0.40603
5	0.36990
6	0.36103

Additionally, the Silhouette Score, detailed in Table 1, showed that K=2 provided the highest average silhouette score, indicating well-defined and separated clusters. These combined results suggest that K=2 is the most appropriate choice, providing a clear and meaningful segmentation of the planner groups into two distinct clusters.

TABLE 2. ALL DATA AFTER CLUSTERING

Planner Group	Number of Maintenance	Total Amount (Rupiahs)	Average Duration (Hours)	Cluster
441	22	587208002	24.318182	1
4H9	6	48209535	0.000000	1
506	501	4865075384	41.738523	2
542	11	161829608	6.181818	1
543	483	3439626219	19.331263	2
544	6	200853891	25.666667	1
550	431	3219577281	101.649652	2
551	529	7573787781	64.298677	2
552	165	2989972570	61.193939	1
553	522	2710243823	3.871648	2
554	348	9394216491	63.560345	2
555	65	877767845	36.815385	1
556	187	1587564584	1.550802	1
557	145	2411915580	31.696552	1
558	503	5332857371	2.457256	2
559	148	1490250484	5.540541	1
560	153	650615698	3.405229	1
561	234	2495160455	6.752137	1
568	6	3329661	44.000000	1
569	106	990188021	21.150943	1
5750	108	302968144	16.416667	1
5A1	79	321985252	14.430380	1
5A2	543	1587726091	5.697974	1
603	38	537497568	6.289474	1
604	154	2062659779	1.064935	1
609	705	3759749320	22.069504	2
610	171	1538808131	0.192982	1
619	126	1713686818	11.277778	1
620	58	151138053	30.275862	1
622	501	4384252919	18.249501	2
635	176	156234859	0.000000	1
636	186	603454039	0.758065	1
637	14	572382837	60.285714	1
638	463	1205296589	1.730022	2
662	184	732422347	27.668478	1
663	161	2290991008	54.248447	1
666	6	12881973	2.000000	1
667	131	541661176	5.717557	1
668	319	1760479470	4.103448	1
686	220	1216434473	42.281818	1

After obtaining the optimal K value, Table 2 summarizes the clustering results for various planner groups at PT Semen Indonesia, based on key maintenance metrics. The table includes columns for the planner group identifier, the number of maintenance activities performed (The Number of Maintenance), the total amount spent on maintenance (Total Amount), and the average duration of maintenance tasks (Average Duration). The "Total Amount" refers to the total maintenance costs incurred in Indonesian Rupiah (IDR), while the "Average Duration" represents the average time taken to complete maintenance tasks, measured in hours. The final column, Cluster, indicates the cluster to which each planner group has been assigned using the K-Means clustering algorithm.

Figure 3 presents a scatter plot that effectively illustrates the clustering of planner groups at PT Semen Indonesia, using two primary variables: the "Amount of Maintenance" (x-axis) and the "Total Maintenance Cost" (y-axis). The data points are distinctly color-coded to reflect their respective cluster assignments. Cluster 1 is represented by green dots, while Cluster 2 is shown with blue dots, allowing for easy visual differentiation. Additionally, the red points indicate the centroids, marking the central positions of each cluster and providing a clear reference for the cluster centers and overall data distribution.

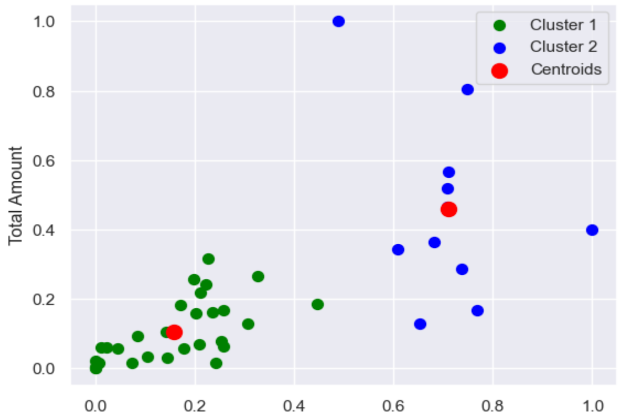

FIGURE 3. SCATTER PLOT BASED ON CLUSTERING RESULT

From the plot, it is evident that Cluster 1 generally consists of planner groups with lower values of both maintenance frequencies and costs, indicating more efficient maintenance operations. In contrast, Cluster 2 groups planner units with higher maintenance frequencies and costs, suggesting potential inefficiencies. The distinct separation between the two clusters highlights the effectiveness of the K-Means clustering algorithm in differentiating between high-performing and low-performing planner groups. The centroids provide a reference point for the average behavior of each cluster, further illustrating the differences in maintenance practices between the two groups.

TABLE 3. PLANNER GROUP IN EVERY CLUSTER

Cluster 1	Cluster 2
Unit Of Reliability Maintenance (441)	Section Of FM 3-4 Machine Maintenance (506)
Unit Of Quality Control (4H9)	Section Of CRSHR&CONV 1-2 Machine MAINT (543)
Section Of Crusher Operation (542)	Section Of RM 1-2 Machine Maintenance (550)
Section Of Heavy Equip & Coal Transport (544)	Section Of KC 1-2 Machine Maintenance (551)
Section Of KC 1-2 ELINS Maintenance (552)	Section Of RM 3-4 Machine Maintenance (553)
Section Of CRSHR&CONV 1-2 ELINS MAINT (555)	Section Of KC 3-4 Machine Maintenance (554)
Section Of FM 1-2 Operation (556)	Section Of FM 1-2 Machine Maintenance (558)
Section Of KC 3-4 ELINS Maintenance (557)	Section Of WHRPG Maintenance (5A2)
Section Of RM 3-4 ELINS Maintenance (559)	Section Of CRSHR&CONV 3-4 Machine MAINT (609)
Section Of FM 1-2 ELINS Maintenance (560)	Section Of EPDC Maintenance (622)
Section Of FM 3-4 ELINS Maintenance (561)	Section Of Utility Operation (638)
:	:

Table 3 above categorizes various sections and units within PT Semen Indonesia into two distinct clusters based on their maintenance practices, as identified through the K-Means clustering analysis. Cluster 1 contains 29 planner groups, significantly more than Cluster 2, which is consists of 11 planner groups. Cluster 1 consists of units and sections which have been identified as more efficient in their maintenance operations. These units generally exhibit better control over maintenance costs and durations, reflecting more optimized processes. On the other hand, Cluster 2 includes sections which have been categorized as less efficient. These sections tend to have higher maintenance costs and longer durations, indicating potential areas for improvement. The clustering results highlight key differences in maintenance performance across the organization, providing a clear basis for targeted interventions to enhance operational efficiency in the less efficient clusters.

This distribution suggests that most planner groups at PT Semen Indonesia are classified into Cluster 1, which likely represents more efficient units with lower maintenance costs and durations, as inferred from the previous analyses. On the other hand, the smaller number of planner groups in Cluster 2 may indicate that these are the less efficient units, characterized by higher costs and longer maintenance durations. The disparity in the number of planner groups between the two clusters highlights the varying levels of performance across the organization and underscores the potential for targeted improvements in Cluster 2.

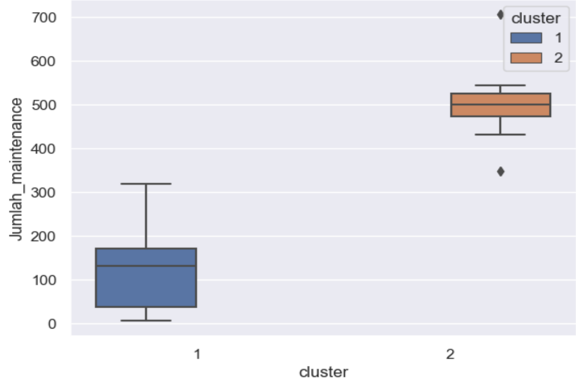

FIGURE 4. BOX CLUSTER VS AMOUNT MAINTENANCE

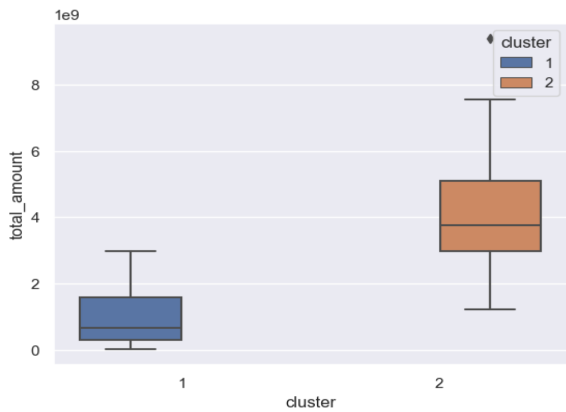

FIGURE 5. BOX PLOT CLUSTER VS AMOUNT MAINTENANCE

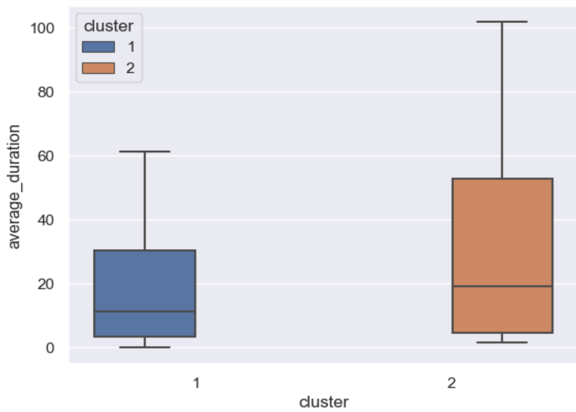

FIGURE 6. BOX PLOT CLUSTER VS AVERAGE DURATION

The three box plots in figure 4-6 presented offer a comparative analysis of the key variables Amount Maintenance, Total Amount, and Average Duration across the two clusters identified in the K-Means clustering analysis. The first box plot in figure 4 compares the number of maintenance activities between the two clusters. Cluster 1, represented in blue, shows a lower median and a narrower interquartile range, indicating fewer maintenance activities. Cluster 2, represented in orange, has a significantly higher median and a wider range, suggesting that planner groups in this cluster perform more maintenance tasks, possibly due to higher maintenance needs or inefficiencies. The second box plot in figure 5 highlights the total maintenance costs across the two clusters. Like the first plot, Cluster 1 shows a lower median cost with less variability, indicating more controlled and consistent spending on maintenance. In contrast, Cluster 2 shows a higher median cost and a wider interquartile range, reflecting greater variability and higher expenditures in maintenance activities.

The third box plot in figure 6 compares the average duration of maintenance tasks between the clusters. Cluster 1 again demonstrates lower values with less variability, suggesting more efficient and quicker maintenance processes. Cluster 2 shows significantly higher median and variability in maintenance durations, indicating longer and potentially less efficient maintenance operations. Overall, these box plots clearly illustrate the differences in maintenance performance between the two clusters, with Cluster 1 generally exhibiting more efficient maintenance practices, characterized by lower maintenance counts, costs, and durations. In contrast, Cluster 2 shows higher values across all metrics, highlighting areas where operational efficiencies may be lacking and where improvements could be targeted.

5. CONCLUSIONS

The K-Means clustering analysis has provided valuable insights into the maintenance practices at PT Semen Indonesia. By clustering planner groups based on key variables—such as the number of maintenance activities, total costs, and average task duration—two distinct clusters emerged. Cluster 1 is characterized by more efficient operations, with lower maintenance counts, costs, and durations, suggesting effective resource and schedule management. In contrast, Cluster 2 exhibited higher values, indicating inefficiencies and opportunities for improvement. The study suggests that PT Semen Indonesia should take actionable steps to address these inefficiencies, including targeted training for Cluster 2 planner groups and investing in predictive maintenance technologies. This could help forecast potential issues and optimize schedules, leading to enhanced operational efficiency and reduced costs. Additionally, the company should regularly review planner group performance, using benchmarking and key performance indicators (KPIs) to monitor variables such as maintenance frequency, costs, and duration. A continuous improvement approach will help ensure alignment with operational goals. However, the study has limitations. It relies solely on historical maintenance data and does not account for external factors like machine condition variability or supply chain disruptions. The use of K-Means clustering may also have limitations, and future research should explore other techniques such as density-based or hierarchical clustering. Incorporating real-time data and expanding the analysis across different sectors or plants could provide more comprehensive insights.

REFERENCES

[1] S. S. Rudiantara, S. Umar, A. A. Idat, A. P. Bhakt, L. S. Djama, A. P. Adi, “Pedoman Tata Kelola Perusahaan yang Baik (GCG Code),” 2022.

[2] B. S. Gandhare and M. M. Akarte, “Benchmarking maintenance performance in select agro-based industry,” *J. Qual. Maint. Eng.*, vol. 28, no. 2, pp. 296–326, 2022.

[3] P. Macnico, J. Christini, N. Sandra, Y. Nuraeni, N. B. Laulita, and F. Cuandra, “Analisa Implementasi Manajemen Rantai Pasok Berbasis Erp Pada Sistem Distribusi Pt Semen Indonesia Tbk,” *Transekonomika Akuntansi, Bisnis Dan Keuang.*, vol. 2, no. 3, pp. 145–164, 2022.

[4] V. UpendraReddya and S. J. J. Thangarajb, “Prediction of Likely Customers for Car Industries Using K-Means Clustering Compared with Logistic Regression,” *Adv. Parallel Comput. Algorithms, Tools Paradig.*, vol. 41, p. 225, 2022.

[5] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021.

[6] S. S. Aripin, I. Imlakiyah, and Y. Suharyat, “Transformasi Organisasi di Era Society 5.0: Inovasi, Adaptasi, dan Keterlibatan Manusia dalam Revolusi Teknologi,” *NUSRA J. Penelit. dan Ilmu Pendidik.*, vol. 5, no. 1, pp. 37–44, 2024.

[7] GeeksforGeeks, “Data Mining – Cluster Analysis.” https://www.geeksforgeeks.org/data-mining-cluster-analysis/ (accessed Feb. 01, 2023).

[8] MySkill Blog, “Mengenal Clustering: Pengertian, Manfaat, Metode, Contoh & Syarat.” https://blog.myskill.id/istilah-dan-tutorial/clustering-cara-mengelompokkan-data-untuk-memahami-pola/ (accessed Oct. 30, 2023).

[9] A. E. Ezugwu *et al.*, “A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects,” *Eng. Appl. Artif. Intell.*, vol. 110, p. 104743, 2022.

[10] H. Shu *et al.*, “Density-based clustering for bivariate-flow data,” *Int. J. Geogr. Inf. Sci.*, vol. 36, no. 9, pp. 1809–1829, 2022.

[11] A. E. Satriatama *et al.*, “Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, pp. 172–182, 2023.

[12] P. A. Ariawan, “Optimasi pengelompokan data pada metode K-means dengan analisis outlier,” *J. Nas. Teknol. dan Sist. Inf*, vol. 5, no. 2, pp. 88–95, 2019.

[13] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means algorithm: A comprehensive survey and

performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.
[14] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023.
[15] S. Regina, E. Sutinah, and N. Agustina, "Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 573–582, 2021.
[16] M. Gul and M. A. Rehman, "Big data: an optimized approach for cluster initialization," *J. Big Data*, vol. 10, no. 1, p. 120, 2023.
[17] T. C. Kit, N. Firdaus, and M. Azmi, "Customer profiling for Malaysia online retail industry using K-Means clustering and RM model," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, pp. 106–113, 2021.
[18] I. Chahid, A. K. Elmiad, and M. Badaoui, "Data Preprocessing For Machine Learning Applications in Healthcare: A Review," in *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 2023, pp. 1–6.
[19] A. N. Alifah, H. N. Fadhilah, and T. M. Sianipar, "Klasterisasi KabupatenKota di Jawa Barat Berdasarkan Tingkat Kenyamanan dengan Metode K-Means Clustering," in *PROSIDING SEMINAR NASIONAL SAINS DATA*, 2022, vol. 2, no. 1, pp. 30–38.
[20] K. P. Sinaga, I. Hussain, and M.-S. Yang, "Entropy K-means clustering with feature reduction under unknown number of clusters," *Ieee Access*, vol. 9, pp. 67736–67751, 2021.
[21] A. Mazarei, R. Sousa, J. Mendes-Moreira, S. Molchanov, and H. M. Ferreira, "Online boxplot derived outlier detection," *Int. J. Data Sci. Anal.*, pp. 1–15, 2024.
[22] W. A. Prastyabudi, A. N. Alifah, and A. Nurdin, "Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering," *Procedia Comput. Sci.*, vol. 234, pp. 96–105, 2024.
[23] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *EURASIP J. Wirel. Commun. Netw.*, vol. 2021, pp. 1–16, 2021.
[24] Y. Januzaj, E. Beqiri, and A. Luma, "Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique.," *Int. J. Online Biomed. Eng.*, vol. 19, no. 4, 2023.
[25] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

AUTHORS

Muhammad Fadhil Alviano
is a student at Telkom University majoring in Data Science. He has a keen interest in applying data analytics and machine learning techniques to solve real-world industrial challenges. Through his experiences, he has gained practical experience in optimizing maintenance management and improving operational efficiency in industrial settings.

Amalia Nur Alifah
is a lecturer and researcher specializing in mathematics and data science. With a background in mathematics, she has worked on several projects aimed at optimizing operational efficiency through advanced data analysis techniques. She is particularly interested in leveraging machine learning algorithms, such as K-Means clustering, to enhance decision-making in industrial settings.

Helga Fadhil Raditya
is a student at Telkom University majoring in Data Science. He has a keen interest in applying data analytics and machine learning techniques to solve real-world industrial challenges. Through his experiences, he has gained practical experience in optimizing maintenance management and improving operational efficiency in industrial settings.

Calista Ghea Ardhani
is a student at Telkom University majoring in Data Science. He has a keen interest in applying data analytics and machine learning techniques to solve real-world industrial challenges. Through his experiences, he has gained practical experience in optimizing maintenance management and improving operational efficiency in industrial settings.

Harashta Tatimma Larasati
is a lecturer and researcher which focuses on quantum computing, cryptography, and adversarial attacks, with a growing portfolio of publications in leading journals. Harashta has contributed to various research areas including quantum natural language processing, blockchain security, and software-defined networking. She actively explores the intersection of quantum technologies and cybersecurity, aiming to enhance the security and efficiency of emerging systems.