

Cluster Analysis Of Junior K3 Construction Training Participants Based On Competency Using K-Means Clustering Method

Nanda Fitria¹, Arief Wibowo², Qodariah³

¹Master of Management Study Program, Universitas Budi Luhur

²Lecturer, Master of Management Study Program, Universitas Budi Luhur

³Lecturer, Master of Management Study Program, Universitas Budi Luhur

Email: 2531600514@student.budiluhur.ac.id, arief.wibowo@budiluhur.ac.id,
qodariah@budiluhur.ac.id

Abstract

This study aims to analyze the characteristics and grouping of participants in the Junior K3 Construction Training based on competency using the K-Means Clustering method. Participant competency was measured using three variables: pre-test scores, post-test scores, and length of employment. The study employed a quantitative approach with a data mining method. The research population consisted of 54 participants of the Junior K3 Construction Training, all of whom were included as the sample using a saturated sampling technique. Data processing was carried out using RapidMiner Studio software through the stages of data cleaning, data selection, data normalization, clustering process, and centroid analysis. The results show that the K-Means algorithm successfully grouped participants into three clusters. Cluster 0 consists of 12 participants (22.22%) with high competency characteristics and relatively high work experience. Cluster 1 consists of 7 participants (12.96%) with the highest work experience but relatively lower initial competency compared to Cluster 0. Cluster 2 is the largest cluster with 35 participants (64.82%) who have developing competency levels and relatively lower work experience. Centroid analysis shows that the length of employment variable is the main differentiating factor between clusters, while post-test scores are relatively high across all clusters. The findings indicate that the training program was able to improve participant competency evenly despite participants having varying levels of work

881



experience. The research results can be used as a basis for formulating more effective competency development strategies through a participant-characteristic-based training approach.

Keywords: Competency, K3 Construction Training, K-Means Clustering, Data Mining, Human Resource Management.

INTRODUCTION

Human resources are a strategic asset that determines an organization's success in achieving its goals and maintaining competitive advantage. Developing human resource competency is one of the main factors that influences individual and organizational performance, as adequate competency can improve productivity, work effectiveness, and decision-making quality (Sakalessy et al., 2024). Therefore, organizations need to undertake continuous competency development efforts through education and training programs.

In the construction sector, workforce competency plays a very important role as it is directly related to workplace safety, work quality, project implementation efficiency, and the achievement of construction targets. The high risk of workplace accidents in the construction sector requires workers to possess knowledge, skills, and abilities that meet occupational health and safety (K3) standards. Improving workforce competency through training is one of the effective strategies for reducing the risk of workplace accidents and improving construction project performance.

One form of competency development that is widely implemented is the Junior K3 Construction Training. This training program aims to improve participants' ability to understand, apply, and supervise the implementation of occupational health and safety systems in construction projects. However, training participants generally have diverse characteristics, both in terms of work experience and initial competency levels. These differences cause participants' ability to absorb training materials and achieve learning outcomes to be uneven.

Identifying participant characteristics is important as it can serve as a basis for developing more effective and targeted competency development strategies. Participants with high initial competency require a different learning approach compared to those who are still in the competency development stage. Therefore, a method capable of grouping participants based on similarities in their characteristics is needed so that the training process can be tailored to the needs of each group.

882



The development of information technology and data mining provides opportunities for organizations to conduct more objective data analysis to support managerial decision-making. Data mining enables the identification of patterns, relationships, and certain characteristics from a dataset, thereby producing valuable information for organizations. One of the data mining methods widely used for data grouping is K-Means Clustering, as it is capable of effectively and efficiently grouping objects based on the similarity of their characteristics (Ezugwu et al., 2023). K-Means is also a widely used clustering algorithm in various studies due to its simple computational process and ability to produce easily interpretable groups.

Several previous studies have shown that the K-Means Clustering method is effective for competency grouping and human resource evaluation. Nababan (2023) applied K-Means Clustering to evaluate lecturer performance and found that the method was able to identify performance groups based on their characteristics. Furthermore, Hendrastuty (2024) showed that the K-Means algorithm can be used to group participant learning outcomes based on competency achievement levels, thereby facilitating the learning evaluation process. Another study conducted by Nursyi et al. (2024) also proved that the clustering method is capable of identifying participant characteristics in education and training in a more structured manner.

Nevertheless, studies that specifically examine the grouping of Junior K3 Construction Training participants based on competency using pre-test scores, post-test scores, and length of employment variables remain relatively limited. Most previous studies have focused more on grouping employee performance, student learning outcomes, or general competency evaluation. Thus, there is a research gap in the form of limited studies that integrate the competency analysis of K3 construction training participants with a data mining approach to support decision-making in training management.

Based on the above description, this study aims to analyze the grouping of Junior K3 Construction Training participants based on competency using the K-Means Clustering method. The variables used include pre-test scores, post-test scores, and participants' length of employment. The results of this study are expected to provide an academic contribution to the development of data mining-based human resource management studies and to provide practical recommendations for training organizers in formulating more effective and targeted competency development strategies.



RESEARCH QUESTIONS

- A. What are the characteristics of Junior K3 Construction Training participants based on pre-test scores, post-test scores, and length of employment?
- B. What are the results of grouping Junior K3 Construction Training participants using the K-Means Clustering method?
- C. What are the characteristics of each cluster formed based on the pre-test score, post-test score, and length of employment variables?

RESEARCH METHOD

Type of Research

This study uses a quantitative approach with a data mining method. The quantitative approach was chosen because the study utilizes numerical data in the form of pre-test scores, post-test scores, and participants' length of employment to generate information about participants' competency characteristics. The data mining method is used to identify hidden patterns and characteristics in the data through clustering techniques.

Research Object

The object of this study is the participants of the Junior K3 Construction Training who attended the training program in 2025. The study focuses on analyzing participant competency based on training evaluation results and work experience.

Population and Sample

The population in this study consists of all 54 Junior K3 Construction Training participants. Given that the population is relatively small, the study uses saturated sampling (census), where all members of the population are used as the research sample. Thus, the number of samples used in this study is 54 participants.

Type and Source of Data

The study uses secondary data obtained from the training result documents of the Junior K3 Construction Training. The data used includes:

- a) Participants' pre-test scores.
- b) Participants' post-test scores.
- c) Participants' length of employment (months).



These three variables are used to describe initial competency, post-training competency, and participants' work experience.

Research Variables

The variables used in this study consist of three variables as presented in Table 1.

Table 1. Research Variables

Variable	Symbol	Indicator	Unit	Data Source
Initial Competency	X1	Pre-Test Score	Score (0–100)	Training Data
Final Competency	X2	Post-Test Score	Score (0–100)	Training Data
Work Experience	X3	Length of Employment	Months	Participant Data

Source: Data processed by Author 2026

Data Collection Technique

Data collection was carried out through documentation study on participant data obtained from the Junior K3 Construction Training organizer. The data was then verified to ensure completeness, consistency, and validity of information before the analysis process was conducted.

Data Analysis Technique

Data analysis was conducted using the K-Means Clustering method with the help of RapidMiner Studio software. This method is used to group participants based on the degree of similarity in their characteristics.

The stages of data analysis include:

- Data Cleaning: Data was examined to ensure the absence of duplicate data, missing values, or input errors that could affect the analysis results.
- Data Selection: At this stage, the attributes used in the clustering process were selected, namely pre-test scores, post-test scores, and length of employment.
- Data Transformation: Data was transformed into numerical format suitable for processing using the K-Means algorithm.



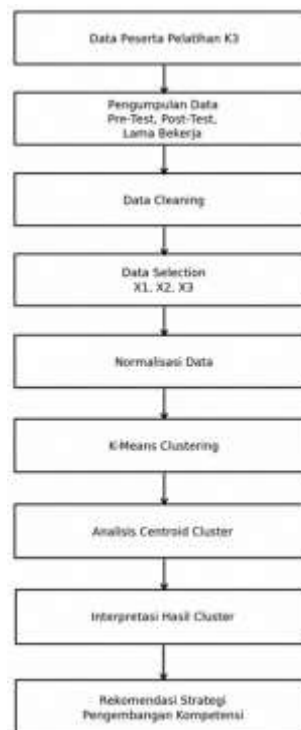
- d) Data Normalization: Normalization was performed using the Min-Max Normalization method to equalize the scale between variables so that no particular variable dominates the clustering process.

K-Means Clustering Process: The clustering process was carried out with the following steps: (a) Determine the number of clusters (K). (b) Determine the initial centroid randomly. (c) Calculate the distance of each data point to the centroid using Euclidean Distance. (d) Group data into the cluster with the nearest distance. (e) Calculate the new centroid based on the average of cluster members. (f) Repeat the process until no changes in cluster membership occur or the centroid has converged.

Cluster Interpretation: The final stage involves analyzing the characteristics of each cluster based on centroid values to obtain a competency profile of training participants that can be used as a basis for developing competency development strategies.

Research Flowchart

The research stages can be illustrated as follows:



[Figure 1. Research Flowchart]

Analysis Tool

Data processing was performed using RapidMiner Studio software. This software is used to carry out the data normalization process, clustering using the K-Means algorithm, cluster result visualization, and cluster centroid analysis.

RESULTS AND DISCUSSION**Description of Research Data**

This study uses data from 54 Junior K3 Construction Training participants. The variables used in the clustering process consist of pre-test scores, post-test scores, and length of employment. These three variables are used to describe initial competency, post-training competency, and participants' work experience.

Table 2. Descriptive Statistics of Research Variables

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	54	40	100	72.96	15.98
Post-Test	54	80	100	95.37	5.39
Length of Employment	54	36	384	122.52	70.60

Source: Data processed by Author 2026

Based on Table 2, the pre-test score has a mean of 72.96, indicating that participants already had an initial level of understanding of K3 construction materials before attending the training. The standard deviation of 15.98 indicates a fairly high variation in initial competency among participants. After attending the training, the average post-test score increased to 95.37 with a standard deviation of 5.39. These results indicate that the training made a positive contribution to improving participants' competency while also resulting in a more uniform level of competency.

The length of employment variable has a mean of 122.52 months, or approximately 10.21 years. The standard deviation of 70.60 indicates that participants have diverse work experience, ranging from those with relatively low experience to those who have been working for more than 30 years. This variation provides a strong basis for conducting a grouping process using the K-Means Clustering method.



K-Means Clustering Results

The clustering process was carried out using the K-Means algorithm with three groups as the number of clusters. The selection of three clusters aims to identify participant characteristics based on their competency levels and work experience. The grouping results produced a cluster member distribution as shown in Table 3.

Table 3. Cluster Member Distribution

Cluster	Number of Participants	Percentage
Cluster 0	12	22.22%
Cluster 1	7	12.96%
Cluster 2	35	64.82%
Total	54	100.00%

Source: Data processed by Author 2026

Based on Table 3, the majority of participants belong to Cluster 2 with 35 participants, or 64.82% of the total training participants. Cluster 0 consists of 12 participants or 22.22%, while Cluster 1 has the fewest members with 7 participants or 12.96%.

The dominance of the number of members in Cluster 2 indicates that the majority of participants have relatively similar characteristics based on the variables used in the study. Conversely, Cluster 0 and Cluster 1 represent groups of participants with more specific characteristics, resulting in fewer members. These results indicate that training participants have a fairly high level of heterogeneity and thus cannot be treated as a single homogeneous group.

Cluster Centroid Analysis

To determine the characteristics of each group of participants, an analysis was conducted on the centroid values produced by the K-Means algorithm. Centroid values describe the average characteristics of members in each cluster. The centroid results are presented in Table 4.



Table 4. Centroid Values of Each Cluster

Variable	Cluster 0	Cluster 1	Cluster 2
Pre-Test	79.167	68.571	71.714
Post-Test	96.667	95.714	94.857
Length of Employment (months)	167.333	262.286	79.200

Source: Data processed by Author 2026

Cluster 0: High Competency and Experienced

Cluster 0 has the highest average pre-test score of 79.167 and a post-test score of 96.667. In addition, the average length of employment reaches 167.333 months, or approximately 13.9 years. These characteristics indicate that participants in this cluster have good initial competency supported by considerable work experience.

Cluster 1: Very High Work Experience

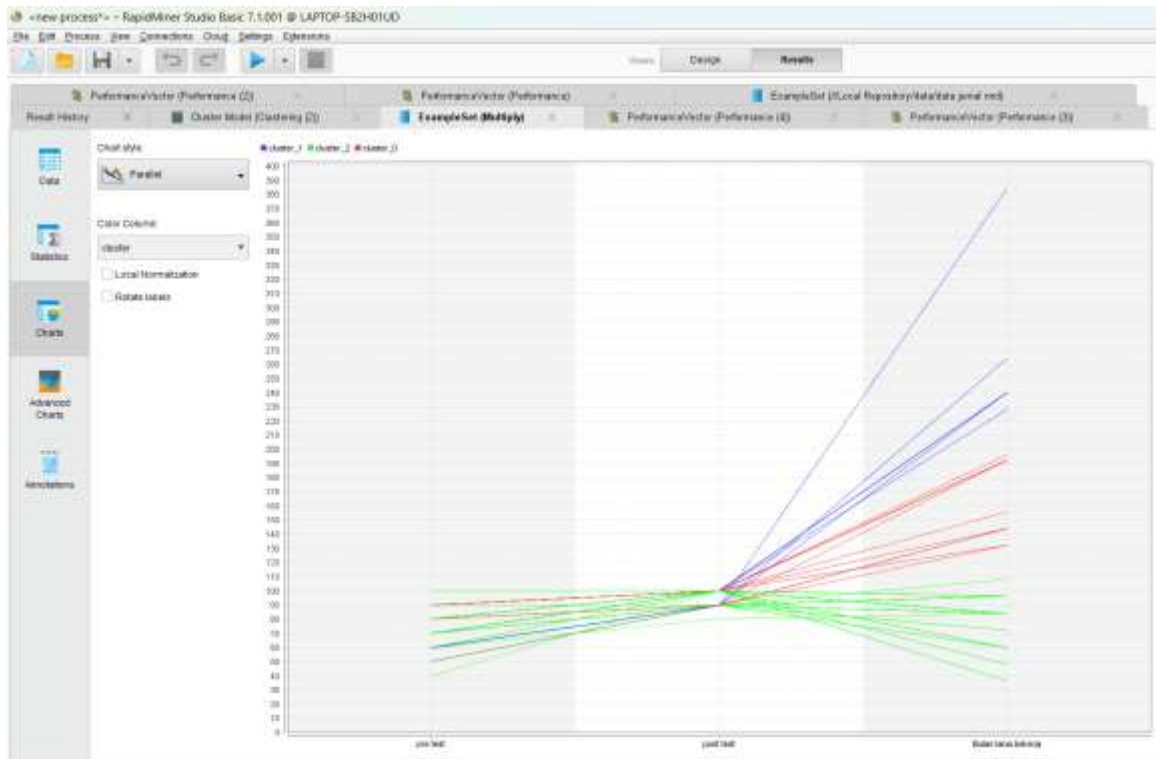
Cluster 1 has the highest average length of employment of 262.286 months, or approximately 21.9 years. However, the pre-test score in this cluster is lower compared to Cluster 0. These results indicate that long work experience is not always accompanied by a higher level of mastery of theoretical material. Nevertheless, post-test scores show that participants were able to improve their competency after attending the training.

Cluster 2: Developing Competency

Cluster 2 is the cluster with the largest number of members. This group has an average length of employment of 79.200 months, or approximately 6.6 years. Pre-test and post-test scores indicate that participants have reasonably good abilities but still require further competency development through continuous training programs.



Cluster Visualization



[Figure 2. Parallel Plot of K-Means Clustering Results for Junior K3 Construction Training Participants]

Figure 2 shows the cluster visualization results using a Parallel Plot. Each color represents a group of participants formed through the K-Means Clustering algorithm. It can be seen that the length of employment variable has the highest level of variation compared to the pre-test and post-test variables. Cluster 1 shows characteristics of participants with the highest work experience, while Cluster 2 is dominated by participants with relatively lower work experience. Meanwhile, post-test scores across all three clusters are in a relatively high range and do not show significant differences. These findings indicate that the training was able to improve participant competency relatively evenly despite participants having varying levels of work experience. These visualization results reinforce the centroid analysis showing that the length of employment variable is the main differentiating factor in cluster formation.

DISCUSSION

The research results show that Junior K3 Construction Training participants have heterogeneous characteristics both in terms of competency and work experience. This finding addresses the first research question, which states that there is variation in initial competency, final competency, and work experience among training participants.

The second research question regarding participant grouping results was successfully answered through the application of the K-Means Clustering method, which produced three clusters with different characteristics. This grouping proves that the K-Means method is capable of identifying patterns of participant similarity based on pre-test scores, post-test scores, and length of employment.

The third research question regarding the characteristics of each cluster shows that the length of employment variable is the main differentiating factor between participant groups. This is evident from the significant differences in centroid values for the length of employment variable compared to other variables. Conversely, post-test scores show relatively high results across all clusters, indicating that the training was able to improve participant competency evenly.

From a Human Resource Management perspective, the clustering results provide important information that competency development strategies cannot be applied uniformly to all participants. A cluster-characteristic-based training approach is considered more effective than a general training approach.

Managerial Implications**Table 5. Recommended Competency Development Strategies**

Cluster	Characteristics	Development Strategy
Cluster 0	High competency and experienced	Advanced Training and K3 Leadership
Cluster 1	Very high work experience	Refreshment Training and K3 Regulatory Updates
Cluster 2	Developing competency	Coaching, Mentoring, and Advanced Training

Source: Data processed by Author 2026

The research results indicate that training organizers need to apply different competency development approaches for each group of participants. Participants in Cluster 0 can be prepared as prospective K3 mentors or supervisors. Participants in



Cluster 1 require updates on current regulations and technology, while participants in Cluster 2 require mentoring and advanced training to continuously improve their competency. Thus, the clustering results can be used as a basis for developing training programs that are more effective, efficient, and targeted in supporting the improvement of human resource quality in the K3 construction sector.

CONCLUSION

Based on the research results, it can be concluded that Junior K3 Construction Training participants have heterogeneous characteristics based on pre-test scores, post-test scores, and length of employment. Descriptive statistics results show that participant competency improved after attending the training, as indicated by the increase in the average post-test score compared to the pre-test score.

The application of the K-Means Clustering method successfully grouped participants into three clusters with different characteristics. Cluster 0 consists of participants with high competency and relatively high work experience. Cluster 1 has the highest work experience, but the initial competency level is relatively lower compared to Cluster 0. Meanwhile, Cluster 2 is the largest group, dominated by participants with developing competency and relatively lower work experience.

Centroid analysis results indicate that the length of employment variable is the factor that most differentiates the characteristics between clusters. Nevertheless, post-test scores across all clusters show relatively high results, indicating that the Junior K3 Construction Training program was able to improve participant competency evenly without being dominantly influenced by the level of work experience.

These research findings indicate that grouping participants using the K-Means Clustering method can provide useful information for training organizers in developing more targeted competency development strategies. Thus, training programs can be designed based on the characteristics of each group of participants so that the effectiveness of human resource development in the construction K3 sector can be improved.

RECOMMENDATIONS

- a) Training organizers need to implement cluster-based competency development strategies so that training programs can better suit participants' needs.
- b) Participants in Cluster 0 can be prepared as mentors or internal instructors through advanced training and K3 leadership programs.

892



- c) Participants in Cluster 1 require refreshment training programs and updates on the latest construction safety regulations and technologies.
- d) Participants in Cluster 2 require coaching, mentoring, and advanced training programs to accelerate competency improvement.
- e) Future research is recommended to add other variables such as education level, professional certification, age, and position to produce a more comprehensive participant segmentation.

REFERENCES

- Dessler, G. (2020). Human Resource Management (16th ed.). Pearson.
- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2023). A comprehensive survey of clustering algorithms. *Artificial Intelligence Review*, 56(3), 1921–1984.
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann.
- Hendrastuty, N. (2024). Penerapan K-Means Clustering untuk pengelompokan hasil belajar peserta pelatihan. *Jurnal Informatika dan Sistem Informasi*, 12(1), 45–55.
- Nababan, R. (2023). Analisis kinerja dosen menggunakan metode K-Means Clustering. *Jurnal Teknologi Informasi dan Manajemen*, 8(2), 101–110.
- Nursyi, A., Pratama, D., & Hidayat, R. (2024). Implementasi data mining untuk segmentasi peserta pelatihan berbasis kompetensi. *Jurnal Sistem Informasi Indonesia*, 9(1), 33–44.
- Sakalessy, M., Rumawas, W., & Tumbel, A. (2024). Human resource competency and organizational performance. *International Journal of Management Studies*, 11(2), 55–67.
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at Work: Models for Superior Performance*. John Wiley & Sons.

