International Journal of Economics, Business and Innovation Research
https://e-journal.citakonsultindo.or.id/index.php/IJEBIR
E-ISSN : 2964-0865
IJEBIR

Analysis of the Effectiveness of Employee Training Programs Using K-Means Algorithm Based on RapidMiner in an Industrial Catering Company in Jakarta

Gonggo Prio Trilaksono[1], Arief Wibowo[2]

[1]Master of Management Student, Budi Luhur University
[2] Professor of Intelligent Systems for Crisis & Disaster Management, Budi Luhur University

2531600589@student.budiluhur.ac.id[1], arief.wibowo@budiluhur.ac.id[2]

Abstract

This study aims to measure the effectiveness of employee training programs in an industrial catering company in Jakarta based on the variables of work tenure and job position. The analytical method used is computational data mining with the K-Means Clustering algorithm through RapidMiner software. The evaluation for determining the optimal number of clusters is based on the Davies-Bouldin Index (DBI) testing. The results show that the configuration of three clusters (K=3) is the best model for both operational units, with DBI values of 0.671 for the Hospital unit and 0.636 for the Factory unit. This segmentation successfully mapped employee characteristics into three levels of effectiveness: "Excellent", "Good", and "Fair". It was found that in the Factory unit, the highest training effectiveness was achieved by employees with mid-level tenure at the basic operational level. Meanwhile, in the Hospital unit, the best performance was dominated by senior employees at the mid-to-lower job position scale. Managerially, the clustering results provide an objective foundation for the company in formulating personalized, targeted, and actual-data-based human resource development strategies.

Keywords: Cluster Analysis, K-Means, RapidMiner, Training Effectiveness, Industrial Catering.

INTRODUCTION

The development of information technology and digitalization encourages organizations to adopt data-driven approaches in supporting decision-making processes.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Optimal utilization of data can improve operational effectiveness, work efficiency, and quality of supervision of human resources. One relevant approach in processing large-scale data is data mining, which is the process of extracting information and hidden patterns from a dataset to generate knowledge useful to organizations (Han et al., 2012).

In its implementation, data mining encompasses various analytical techniques, one of which is clustering. This technique is used to group data based on the degree of similarity of certain characteristics so that objects within the same group are homogeneous compared to other groups. One of the most popular clustering methods is the K-Means algorithm due to its simple computational process, ease of implementation, and ability to produce effective data groupings across various domains (Tan et al., 2016).

From a human resource management perspective, competency management and development through training is a crucial aspect in supporting the achievement of organizational goals. Systematic post-training evaluations can help organizations identify levels of material absorption, work effectiveness, and individual contributions after programs are conducted. According to Aguinis (2019), performance and competency management is a continuous process aimed at identifying, measuring, and developing individual capacities to align with organizational goals. In line with this, Armstrong (2021) states that an effective evaluation system can help organizations improve productivity, develop specific competencies, and support more objective decision-making.

As part of the large-scale food service industry, industrial catering companies in Jakarta play a crucial role in supplying mass food needs for various critical sectors, such as hospitals and manufacturing plants. The operational characteristics of this sector demand strict quality standardization, timely distribution, and high hygiene standards. Therefore, periodic employee training programs are mandatory to maintain service quality and minimize operational errors in the field.

In the implementation of post-training evaluations, each employee shows a different level of effectiveness in material absorption. These differences are suspected to be closely related to the employees' occupational demographic backgrounds, such as their work tenure (experience) and job position (level of structural responsibility) within the company. However, the training effectiveness evaluation process has often been conducted manually, partially, and conventionally. This situation causes management or the Human Resources Development (HRD) division to experience difficulty in seeing structured correlations and objectively, measurably, and systematically identifying training result segmentation as the basis for managerial decision-making.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

One data mining method that can be used to perform this data segmentation is the K-Means Clustering algorithm. The K-Means algorithm is a reliable method for grouping data based on the degree of similarity of characteristics between objects. This method has been widely applied in competency and employee training evaluations as it can produce objective and easily interpretable data segmentation with the aid of software such as RapidMiner Studio. Research by Regina et al. (2021) shows that K-Means is capable of grouping employee capacity quality based on their characteristics, thereby supporting the evaluation process of organizational development programs.

Various previous studies have applied the K-Means algorithm for employee segmentation, performance quality, and employee competency assessment. Regina et al. (2021) applied K-Means to group employee performance quality in a chemical company, while Pramitasari & Nataliani (2021) compared the K-Means and Fuzzy C-Means algorithms in employee clustering based on achievement values. The results of these studies indicate that clustering methods can help organizations identify competency data patterns not easily recognized through conventional statistical analysis.

Other studies have also demonstrated the effectiveness of the K-Means algorithm in segmentation and internal capacity improvement evaluation. Pasricha & Prihandoko (2018) showed that the K-Means algorithm is capable of effectively grouping employee discipline levels and adaptation based on their characteristics. Setiawan (2019) also applied the K-Means algorithm to identify employee capabilities in operational fields and obtained groupings that can serve as the basis for organizational training decision-making. Furthermore, Intan et al. (2023) showed that applying K-Means to employee learning outcome assessments can produce more objective segmentation and help organizations map training efficiency.

Although various previous studies have demonstrated that the K-Means algorithm is effective for data segmentation, most studies still focus on general performance evaluation, attendance rates, or consumer segmentation. Research specifically examining training effectiveness by correlating it with work tenure and job position variables in specific industries such as industrial catering in Jakarta remains relatively limited. Yet, the mapping of employee groups post-training in the food logistics and hospital/factory service sectors is highly strategic to minimize operational loss risks and maintain client trust.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Problem Statement

The problem statement of this study focuses on how the implementation of the K-Means Clustering algorithm using RapidMiner Studio performs data segmentation of training performance data for employees in the Hospital unit and Factory unit at an industrial catering company in Jakarta. Additionally, the study explores the characteristics of each employee cluster formed based on the variables of work tenure, job position, and training score to measure the effectiveness of the training program. Finally, this study formulates a problem regarding how the utilization of the employee data segmentation results from the Hospital and Factory units can serve as an instrument supporting managerial decision-making in the evaluation of training programs, competency mapping, and Human Resource Development (HRD) planning.

Research Objectives

This study specifically aims to implement the K-Means Clustering algorithm through RapidMiner Studio to segment employee data placed in the Hospital and Factory sectors. This data mining approach is intended to measure the level of effectiveness of training programs by identifying and mapping the specific characteristics of each employee cluster based on the interrelationship among the variables of work tenure, job position, and training score. This study offers a novelty in the form of integrating occupational demographic variables to evaluate training effectiveness in two operational environments with different characteristics.

METHOD

This study uses a quantitative approach with the data mining method through the K-Means Clustering algorithm. The research object specifically focuses on employees assigned to the Hospital and Factory sectors at an industrial catering company in Jakarta. The analysis was conducted using profile and competency data encompassing three main variables: work tenure, job position, and training score.

All research stages were designed systematically, including data collection, data preprocessing, data normalization, the clustering process, cluster evaluation, and result interpretation. In the preprocessing stage, operational data was filtered through completeness and consistency checks to clean biased information, followed by data normalization transformation to eliminate scale range imbalances between variables—particularly between work tenure (in years), job position (level/weight scale), and training score (scale of 0–100).

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

The research population consisted of all employees who participated in the training program at the industrial catering company in Jakarta. The research sample comprised 75 respondents selected using purposive sampling technique, consisting of 38 Hospital unit employees and 37 Factory unit employees, with the criteria of having complete records related to the three main variables studied.

The clustering procedure was executed using the K-Means algorithm in RapidMiner Studio software. This procedure was carried out through facets of determining the number of groups (K), initial centroid initialization, allocation of employee data to the nearest cluster based on Euclidean distance, and iterative updating of centroid coordinates until convergence was achieved. To test and determine the most optimal number of clusters, this study uses the Davies-Bouldin Index (DBI) method, which compares the level of internal group cohesion (homogeneity) and inter-group separation (heterogeneity). The best configuration indicator is shown by the DBI value closest to zero (the smallest absolute value).

DISCUSSION

Job Code Alignment

For computational processing purposes in RapidMiner Studio, the qualitative job position variable was transformed into numerical form (label encoding). The weighting structure of structural positions is presented in Table 1.

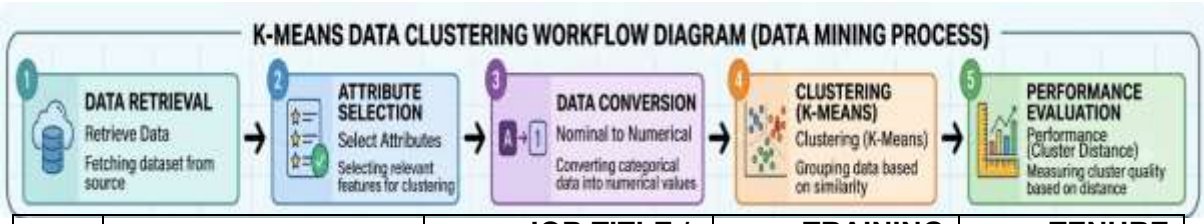
Table 1. Number Codification of Employee Job Types and Positions

No	Job Title /Position	No	Job Title / Position	No	Job Title / Position
1	Admin Officer	6	Jr. Cook	11	Steward
2	Nutritionist	7	Jr. Pastry Cook	12	Storeman
3	Captain	8	Senior Cook	13	Waiter
4	Chef	9	Senior Pastry Cook	14	Waitress
5	Cook Helper	10	Site Manager		

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

International Journal of Economics, Business and Innovation Research
https://e-journal.citakonsultindo.or.id/index.php/IJEBIR
E-ISSN : 2964-0865
IJEBIR

Prior to grouping the data using the K-Means Cluster method in RapidMiner, descriptive statistical analysis was first conducted to obtain a general overview of the data based on the variables of employee work tenure and job position. This analysis includes calculation of minimum, maximum, mean, and standard deviation values to evaluate the effectiveness of the training attended by hospital and factory employees.



NO	EMPLOYEE NAME	JOB TITLE / POSITION	TRAINING SCORE	TENURE (YEARS)
1	RMA	8	100	3
2	MIC	13	95	3
3	NHM	14	100	3
...	...	...	...	...
36	SFT	1	90	2
37	RNK	1	85	2
38	LDR	14	90	2

(Hospital Employee Data)

NO	EMPLOYEE NAME	JOB TITLE / POSITION	TRAINING SCORE	TENURE (YEARS)
1	RIN	8	90	3
2	SON	6	100	3
3	AHH	11	95	3
...	...	...	...	...
35	AFN	13	100	3
36	VDK	14	100	3
37	FPM	2	100	2

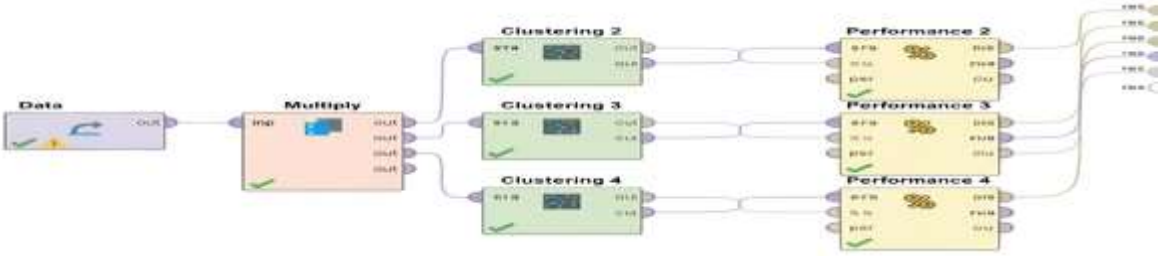
(Factory Employee Data)

Descriptive Statistical Analysis

Descriptive statistical analysis was conducted to provide a general overview of the minimum, maximum, mean, and standard deviation values of the sample objects prior to clustering.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

International Journal of Economics, Business and Innovation Research
https://e-journal.citakonsultindo.or.id/index.php/IJEBIR
E-ISSN : 2964-0865
IJEBIR



Before the clustering process, employee data from both operational sectors first underwent a preprocessing stage to ensure the data was in a clean and ready condition for processing. The preprocessing stage included checking for missing values in exam scores, adjustment and weighting of job position data types (label encoding), and data normalization to equalize the scale between variables. This normalization was critically performed so that the disproportionate scale differences—between the range of work tenure (in years), job position levels, and training scores (scale of 0–100)—would not influence or dominate the Euclidean distance calculation process in the K-Means algorithm.

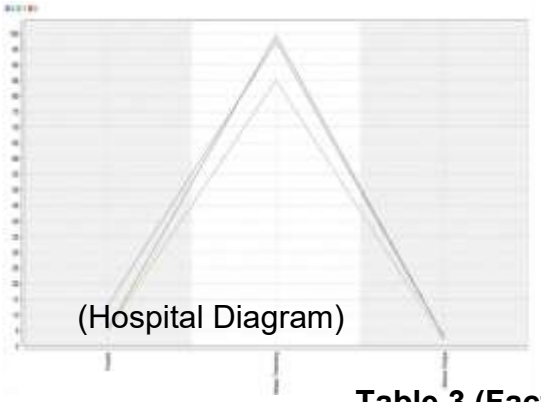
Table 2 (Hospital Employee Data Sample)

No	Employee Name	Job Title / Position	Training Score	Tenure (Years)
1	RMA	8	100	3
2	MIC	13	95	3
3	NHM	14	100	3
...	...	...	...	...
36	SFT	1	90	2
37	RNK	1	85	2
38	LDR	14	90	2

(Hospital Employee Data)

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

International Journal of Economics, Business and Innovation Research
https://e-journal.citakonsultindo.or.id/index.php/IJEBIR
E-ISSN : 2964-0865
IJEBIR



(Hospital Diagram)



(Factory Diagram)

Table 3 (Factory Employee Data Sample)

No	Employee Name	Job Title / Position	Training Score	Tenure (Years)
1	RIN	8	90	3
2	SON	6	100	3
3	AHH	11	95	3
...	...	...	...	...
35	AFN	13	100	3
36	VDK	14	100	3
37	FPM	2	100	2

(Factory Employee Data)

Table 4. Descriptive Statistics of Employee Operational Variables

Operational Unit	Research Variable	N	Minimum	Maximum	Mean	Std. Deviation
Hospital	Job Title / Position	38	1	14	8.13	4.55
	Training Score	38	80	100	94.58	6.80

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

	Tenure (Years)	38	2	3	2.69	0.46
Factory	Job Title / Position	37	1	14	9.70	4.03
	Training Score	37	80	100	95.81	5.75
	Tenure (Years)	37	2	3	2.68	0.46

Based on Table 4, the work tenure characteristics of employees in both units are very similar, with averages of 2.69 years (Hospital) and 2.68 years (Factory). In terms of job position distribution, the Factory unit has a higher average position index value (9.70) compared to the Hospital unit (8.13), indicating a different variation in field assignment structure between the two work environments.

Determination of Optimal Number of Clusters

The testing of cluster number scenarios was measured using the Davies-Bouldin Index (DBI) metric. In the data mining calculation system, negative index values closer to zero indicate an increasingly firm and valid group separation model.

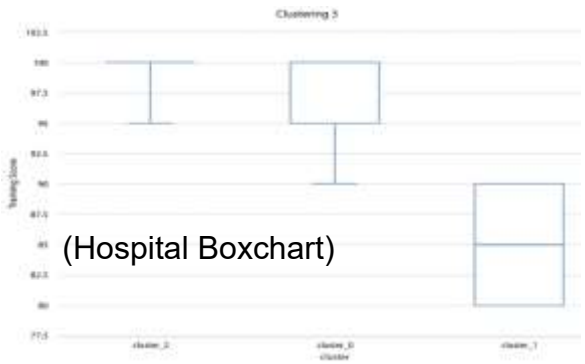
Table 5. Comparison of Davies-Bouldin Index (DBI) Evaluation Values

Number of Clusters Scenario (K)	DBI Value — Hospital Unit	DBI Value — Factory Unit	Model Quality Conclusion
K = 2	0.800	0.792	Suboptimal
K = 3 (Best)	**0.671**	**0.636**	**Most Optimal (Best)**
K = 4	0.723	0.798	Experiencing Decrease in Validity

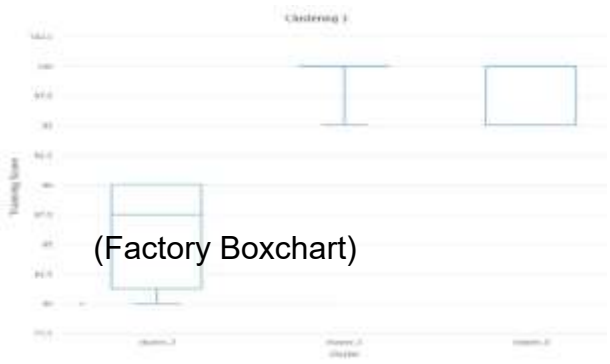
Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

International Journal of Economics, Business and Innovation Research
https://e-journal.citakonsultindo.or.id/index.php/IJEBIR
E-ISSN : 2964-0865
IJEBIR

The results in Table 5 confirm that the formation of three clusters (K=3) is the best scenario for both operational units. This model produces the most accurate separation of group boundaries and the highest level of internal similarity among cluster members.



(Hospital Boxchart)



(Factory Boxchart)

K-Means Cluster Segmentation Results

The computational process groups the total sample into three categories of training effectiveness: "Excellent", "Good", and "Fair". The distribution of members and the description of basic characteristics of each group are presented in the following tables.

Table 6. Characteristics of K-Means Clustering Results for the Hospital Unit

Cluster Code	Number of Members	Effectiveness Category	Group Characteristics and Typology
Cluster 0	15 Employees	Good	Stable performance above average; possesses a solid level of material compliance but requires interactive space to prevent burnout.
Cluster 1	10 Employees	Fair	Passing scores are at the minimum threshold (scale of 80); requires intensive mentoring and post-training supervision.
Cluster 2	13 Employees	Excellent	Highly mature competence (scores > 95); dominated by senior employees and holds high potential to serve as internal instructors.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Table 7. Characteristics of K-Means Clustering Results for the Factory Unit

Cluster Code	Number of Members	Effectiveness Category	Group Characteristics and Typology
Cluster 0	20 Employees	Good	Possesses strong technical mastery in maintaining the rhythm of industrial culinary production; moderate-to-high performance.
Cluster 1	9 Employees	Excellent	Passing scores nearly reach perfection; highly responsive to work quality standardization training targets.
Cluster 2	8 Employees	Fair	Lowest training material absorption level; adaptation to factory operational standardization remains suboptimal.

Final Centroid Value Analysis

Centroid values represent the central point or average value of each variable that forms the main characteristics of the identified clusters.

Table 8. Final Centroid Values Based on Work Unit

Work Unit	Analysis Variable	Cluster 0	Cluster 1	Cluster 2
Hospital	Job Title / Position	12.867	5.400	5.154
	Training Score	97.333	85.000	99.231
	Tenure (Years)	2.933	2.000	3.000
	Group Category	Good	Fair	Excellent

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Factory	Job Title / Position	12.500	4.222	9.250
	Training Score	98.000	99.444	86.250
	Tenure (Years)	2.850	2.778	2.125
	Group Category	Good	Excellent	Fair

Note: The cluster index number naming labels in the RapidMiner K-Means algorithm between the Hospital unit and the Factory unit have different default output category sequences (Table 8 shows the details of the actual real center values for each variable).

Sectoral Characteristic Analysis

"Excellent" Performance Cluster

Based on the centroid data in Table 8, this top-performing group is represented by two contrasting demographic characteristics between the two operational units. In the Hospital Unit, the "Excellent" category was achieved by a group of employees who were fully senior with an absolute work tenure of 3 years at the mid-to-lower job position scale (centroid value of 5.154), coupled with a training score of 99.231. Meanwhile, in the Factory Unit, the "Excellent" category was instead dominated by employees with mid-level tenure (averaging 2.7 years) occupying basic operational job levels (position scale of 4.222), with a maximum average training score of 99.444. This reflects a phenomenon whereby field workers in the factory area tend to be more adaptive and responsive to daily operational technical training materials compared to the dynamics of material absorption in the hospital medical service environment.

"Good" Performance Cluster

For this moderate category, both work units show a tendency toward uniform data patterns. This group is consistently filled by employees who hold the highest structural positions in each unit (scale of 12.867 in the Hospital and 12.500 in the Factory), with service tenures ranging between 2.8 and 2.9 years. This phenomenon confirms that those holding supervisory or managerial functions have stable training competency results in the range of 97 to 98, despite their field technical implementation not being as prominent as that of the worker groups below them.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

"Fair" Performance Cluster

In the group with the lowest level of effectiveness, the algorithm successfully identified patterns of material absorption weaknesses. In the Hospital Unit, employees with new work tenure (2 years) at the basic operational level (position scale of 5.400) recorded the lowest score of 85.000. Meanwhile, in the Factory Unit, employees with the shortest tenure (2.1 years) at mid-level coordination positions (position scale of 9.250) showed low training absorption scores of 86.250. This group becomes the top priority for periodic coaching interventions.

Managerial Discussion

The findings of this study prove that work tenure and job position do not always correlate linearly with the effectiveness of training material absorption. This indicates that program effectiveness is sociologically influenced by the characteristics of the work environment, personal motivation, and the relevance of the material to the employees' actual workload. These findings support the performance management principles of Aguinis (2019), who states that the success of knowledge transfer (transfer of training) is highly influenced by individual readiness and alignment of the work environment.

Through this data mining-based mapping approach, the industrial catering company in Jakarta can move away from subjective conventional evaluations and shift toward a personalized and data-driven HR management strategy. The "Excellent" performance group will be given retention strategies, recognition (reward), and projected as role models or internal instructors (trainers) for the standardization of kitchen operational work. The "Good" performance group will be directed toward more interactive refreshment training programs to reduce daily work burnout. Meanwhile, the "Fair" performance group becomes the priority for intensive coaching through direct on-the-job mentoring programs to ensure that theoretical understanding can be transformed into practical skills, thereby maintaining hygiene and culinary product quality standards of the company.

CONCLUSION

This study successfully implemented the K-Means Clustering algorithm assisted by RapidMiner Studio to map the effectiveness of employee training programs at an industrial catering company in Jakarta. The three-cluster configuration (K=3) proved to be the most valid model based on the Davies-Bouldin Index testing, with DBI values of 0.671 (Hospital) and 0.636 (Factory). The centroid analysis results objectively mapped the

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

distribution of employee competencies into three performance classifications (Excellent, Good, and Fair). These findings provide a practical contribution to company management in compiling a blueprint for training evaluation, career path planning, and allocation of more targeted human resource development budgets, thereby minimizing operational loss risks in the food service industry.

REFERENCES

Aguinis, H. (2019). Performance Management (4th ed.). Chicago Business Press.

Armstrong, M. (2021). Armstrong's Handbook of Human Resource Management Practice (15th ed.). Kogan Page.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Intan, T., Rahmawati, E., & Setiawan, A. (2023). Application of the K-Means Algorithm in Employee Learning Outcome Assessment for Training Efficiency Mapping. Journal of Information Systems and Business Management, 11(2), 89–98.

Pasricha, A., & Prihandoko, P. (2018). Analysis of K-Means Algorithm in Grouping Employee Discipline and Adaptation Levels. Journal of Science and Main Technology, 13(1), 34–42.

Pramitasari, D., & Nataliani, Y. (2021). Comparison of K-Means and Fuzzy C-Means Algorithms in Employee Data Grouping. Journal of Informatics, 15(2), 120–128.

Regina, R., Sutinah, E., & Agustina, S. (2021). Application of K-Means Algorithm for Employee Performance Segmentation. Journal of Information and Communication Technology, 8(1), 45–54.

Setiawan, B. (2019). Implementation of K-Means Clustering to Identify Employee Capabilities in the Framework of Organizational Training Decision-Making. Journal of Human Resource Management, 7(3), 201–212.

Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to Data Mining (2nd ed.). Pearson.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.