Open Science and Technology
Volume 5, Issue 2, 76–89
E-ISSN :2776–1681
https://opscitech.com/journal
DOI: 10.33292/ost.v5i2.159

Accuracy of Learning Method Implementation in Higher Education Using K-Means Clustering and Silhouette Coefficient

Neni Purwati[1]*, M. Cahyo Kriswantoro[1]

[1]Department of Medical Informatics, Faculty of Health Sciences, Universitas Muhammadiyah Lamongan, Lamongan, Indonesia

*Corresponding author's email: nenipurwati@umla.ac.id

Article Info

Article history:

Received: 13 September 2025
Revised: 19 November 2025
Accepted: 24 November 2025
Published: 6 December 2025

Keywords:

Clustering,
CRISP-DM,
K-Means,
Silhouette Coefficient,
Statistics.

Abstract

Background: The Covid-19 pandemic significantly transformed the learning process in higher education, forcing institutions to quickly adapt to unprecedented challenges. Traditional face-to-face learning was no longer feasible due to health restrictions, and this condition accelerated the integration of technology into teaching and learning activities. As a result, online, offline, and hybrid learning methods emerged as the primary alternatives for sustaining academic activities. However, the rapid shift also highlighted a critical issue: the effectiveness of these learning methods varied widely depending on institutional readiness, available resources, and student adaptability. Determining the most effective method has therefore become essential to ensure quality outcomes, maintain student performance, and support the continuity of higher education in the post-pandemic era.
Aims: This study aims to identify the appropriate post-pandemic learning strategy in higher education by applying the CRISP-DM methodology and the k-means clustering algorithm.
Methods: The dataset consists of 65,778 student records collected from 2015–2020, preprocessed through data reduction, cleaning, and transformation. K-means clustering was applied using Orange, an open-source data mining tool. Furthermore, evaluated with the Silhouette Coefficient, and use additional analysis has been carried out using feature statistics.
Result: The results show that offline learning produced the highest total frequency, hybrid learning was in the medium range and online learning the lowest. Silhouette Coefficient scores indicated cluster quality in the medium structure category, with values of 0.47, 0.56, and 0.65 across three clusters. These findings suggest that offline learning remains the most effective method under normal conditions, hybrid learning is more suitable during pandemic or transitional periods, while online learning can serve as an alternative depending on institutional or governmental policies.
Conclusion: The study concludes that clustering-based analysis provides practical insights for designing adaptive, data-driven learning strategies in higher education. This study provides practical implications and benefits by guiding institutions to design adaptive, data-driven learning strategies and establish more precise, responsive educational policies.

To cite this article: Purwati, N., Kriswantoro, M.C. (2025). Accuracy of Learning Method Implementation in Higher Education Using K-Means Clustering and Silhouette Coefficient. *Open Science and Technology, 5*(2), 75–89.

1. Introduction

The Circular Letter of the Minister of Education and Culture Number 4 of 2020, which regulates the Implementation of Education Policy during the Emergency Period of the Spread of Covid-19, and the Decree of the Minister of Education and Culture Number 719/P/2020 concerning Guidelines for the Implementation of Curriculum in Educational Units under Special Conditions by the Minister of

Education and Culture of the Republic of Indonesia, requires Higher Education Institutions to maintain competitiveness by utilizing available resources even during the pandemic (Prayoga *et al.,* 2021).

Face-to-face classroom learning during the pandemic was not possible but could be carried out by integrating technology into the learning process, both online (in-network) and offline (out-of-network) (Ambarita *et al.,* 2020). Direct instructions from the government also had to be followed by students as a preventive measure against the spread of the coronavirus. Therefore, various alternative learning methods such as online learning were implemented (Widyastuti, 2021). In addition to online and offline learning methods, some schools and universities also applied hybrid learning, a combination of offline and online learning with a proportion of 50% offline and 50% online (Verawati & Desprayoga, 2019; TIM DPD ADRI Jawa Timur, 2021). Hybrid learning, also known as blended learning, utilizes the strengths of both face-to-face and online learning to cover the weaknesses of each. Face-to-face learning has advantages that cannot be replaced by Distance Learning (PJJ), and vice versa (Andiopenta & Aripudin, 2021). The main objective of hybrid learning is to provide opportunities for learners with diverse characteristics to be independent, sustainable, and continuously developing, so that the learning process becomes more effective, efficient, and engaging (Setianingsih, 2021). After the pandemic ended, this habit has continued, with many universities still applying online or hybrid learning for various reasons and considerations.

Learning methods are one of the key components of the learning system that must be measurable. Therefore, organizers of the Teaching and Learning Process (PBM) must be able to determine the right and accurate learning methods to be implemented so that the intended learning outcomes can be achieved. Conversely, if the learning methods applied are not appropriate, the results of the learning process will be low or unsatisfactory. In relation to these challenges, many previous studies have used data mining to address clustering problems across various fields. In several previous studies, data mining has been widely used to address problems in data clustering across various fields. Data mining is the science of collecting, cleaning, processing, analyzing, and extracting useful knowledge from data. Fields related to data mining include databases, information science, high-performance computing, visualization, machine learning, statistics, neural networks, mathematics, information retrieval and extraction, and pattern recognition (Hermawati, 2013).

In data mining, one important function is cluster analysis, also known as clustering, which aims to group objects such that those within the same group share similarities, while objects in different groups show differences. Cluster analysis has several advantages, including the ability to group large volumes of data with many variables, and its applicability to ordinal, interval, and ratio scale data (Rachmatin, 2014; Ikhsan, 2021). Clustering is often used as an initial stage in the data mining process, with the resulting clusters serving as input for subsequent algorithms (Alasali & Ortakci, 2024). K-means clustering is commonly applied because the number of clusters required for item categorization is predetermined (Abdullah *et al.,* 2021; Ridzki *et al.,* 2023). K-means iteratively calculates clusters and centroids, applying a top-down approach to handle data grouping. It begins with a predetermined number of K clusters, such as three groups, with three random centroids created as the initial focus of the clustering process. The fundamental difference between clustering and classification algorithms lies in the absence of target/class/label in clustering, which places it in the category of unsupervised learning

Previous research has implemented k-means clustering in the financial sector, which aims to reduce the impact of fraud on the financial system and shows that machine learning-based k-means clustering is promising for fraud detection, thereby creating a safer and more trusted transaction environment in the financial sector (Huang *et al.,* 2024). Another study aims to group customers based on behavior and characteristics, with k-means clustering using the elbow method producing four optimal clusters, where the quantity and unit price variables were found to significantly influence customer behavior (Awalina & Rahayu, 2023). Other study that show identify more effective and accurate promotion strategies for prospective students, with results showing that the combination of k-means clustering and sentiment analysis increased new student admissions by up to 90% (Awaludin & Gani, 2024).

A related study revealed attributes overlooked by traditional approaches in tackling poverty, indicating that approximately 95% of Indonesia's impoverished population were classified within the 'Poor' cluster, with a Silhouette Score of 0.7416, thereby validating the efficacy of the clustering technique (Khalif *et al.,* 2024). Other study developed efficient inventory strategies and identified popular products, with Cluster 2 showing the best results (DBI -0.310), including 616 bestselling products and 8 less popular products (Pujiono *et al.,* 2024). Value-oriented and contextually grounded hybrid learning approaches have the potential to promote religious learning experiences that are authentic, adaptable, and transformative in the post-pandemic period (Negara, 2025).

Among these six studies, all employed the k-means clustering method; however, this research differs in terms of research objects, datasets, and study topics. *The study's novelty lies in its focus on delivering a data-driven evaluation of learning modalities across pandemic phases, supported by clustering validation and evidence-based insights that advance adaptive learning strategies and educational data analytics*. This study aims to identify the appropriate post-pandemic learning strategy in higher education by applying the CRISP-DM methodology and the k-means clustering algorithm. By clustering students based on learning patterns and academic performance, higher education institutions can design more adaptive and need-oriented learning strategies, while simultaneously supporting evidence-based decision making, improving learning effectiveness, and contributing to the advancement of learning analytics in the digital era of higher education.

2. Methods

The research methodology used is CRISP-DM (Cross-Industry Standard Process for Data Mining), a method that provides standardized procedures in data mining which can be applied to problem-solving strategies commonly found in the business sector or in research. It consists of several stages (Wiemer *et al.*, 2018; Shedriko & Firdaus, 2022) and represents the life cycle of a data mining project, with the stages illustrated in Figure 1 as follows (Fimawahib & Rouza, 2021):

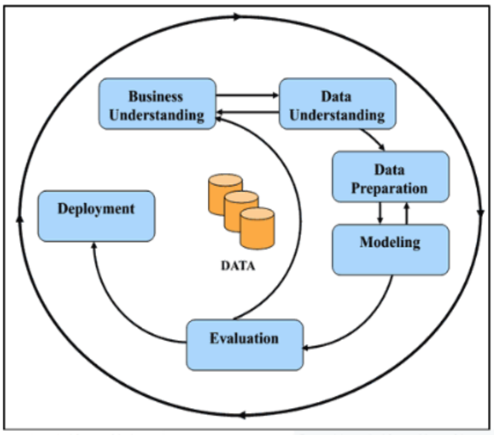

Figure 1. CRISP-DM Life Cycle (Plotnikova *et al.*, 2022)

2.1. **Business Understanding Stage:** This is the initial stage, which involves understanding the business objectives and identifying new information and knowledge to be discovered from past datasets using data mining techniques.
2.2. **Data Understanding Stage:** After defining the business objectives, this stage focuses on understanding the data. It begins with identifying the data collected, describing the data, exploring the data, and carefully examining the quality of the data to be processed.
2.3. **Data Preparation Stage:** At this stage, the process involves identifying and preparing the data, cleaning the data, checking for consistency, and selecting the data relevant to the research. The dataset obtained is then organized and grouped into predetermined clusters.
2.4. **Modeling Stage:** This stage determines the model to be used for grouping learning methods. K-means clustering is the model chosen in this study. The stages of the k-means model are illustrated in Figure 2 (Lestari *et al.*, 2022):

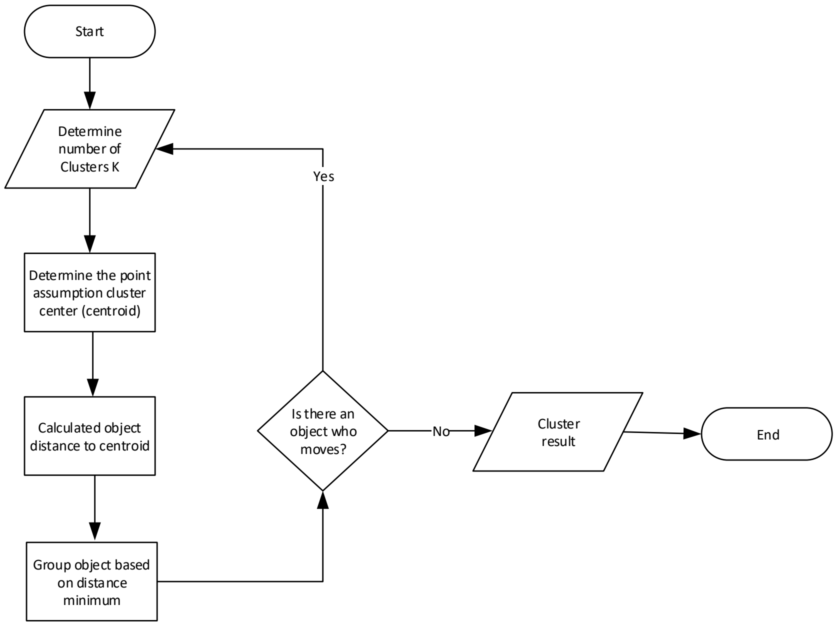

Figure 2. Flowchart K-Means Clustering (Hasdyna *et al.,* 2025)

The flow of the k-means clustering algorithm in Figure 2 above consists of several stages explained as follows:

a. **Determine the number of clusters (K):** Define the number of clusters in the dataset and specify the quantity labeled as *k* in the k-means algorithm.
b. **Determine the point assumption cluster center (centroid):** Identify the central point of the cluster, known as the centroid. The centroid value is determined by applying the midpoint calculation method, which follows the maximum value for the high cluster (C1), the average value for the medium cluster (C2), and the minimum value for the low cluster (C3) (Hutagalung *et al.,* 2021).
c. **Calculate the object distance to the centroid:** Compute the distance of each object to the centroid. The distance is calculated using the Euclidean distance or distance metric, with the following formula (Lubis & Tamam, 2022):

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

Description:
d(x,y) : The distance between data x and data y
xi : The i-th testing data
yi : The i-th training data

d. **Group objects based on the minimum distance:** Classify objects according to the closest or minimum distance.
e. **Is there any object that moves?** A check is performed to determine whether there are still data points that change. If *Yes,* the process is repeated (iteration) back to step (a). If the result is *No*, then the process can proceed to step (f). This process is also referred to as iteration, where steps (a) to (d) are repeated until the positions of the cluster centers show no significant changes or the data points no longer shift between clusters.

The termination of testing is carried out by calculating the difference between weights. If the process reaches convergence, then the testing can be stopped (Gunawan *et al.,* 2024). The process ends when the stopping criteria are met, or when the maximum number of iterations has been reached. This can be calculated using the following formula (Hendrastuty, 2024):

$$R_k = \frac{1}{N_k}(X_{1k} + \cdots X_{nk}) \quad (2)$$

Description:
Rk : Iteration termination
X1 : Value of the first data record
X2 : Value of the second data record
Nk : Total number of data records

f. **Cluster result**: Display the clustering outcome.
The k-means clustering algorithm has a weakness, namely its sensitivity to cluster initialization. The basic principle of the k-means algorithm can be explained as follows (Pratiwi & Ginting, 2024):
1. Cluster initialization.
2. Assign each data point to the most suitable cluster based on its proximity to the centroid. The centroid is a term vector considered as the midpoint of the cluster.
3. After all data points are assigned to clusters, recalculate the centroid of each cluster based on the documents contained within that cluster.
4. If the centroid does not change (within a certain threshold), then stop. If it still changes, return to step 2.

The tool used in this study is Orange, an open-source software specifically designed for data mining. It is quite powerful while also being relatively easy and simple to use. Some of the features provided by this tool/software are as follows (Shedriko & Firdaus, 2022):
a. Provides a wide variety of different algorithms for machine learning and data mining.
b. An open-source application that is freely available at no cost.
c. A standalone and independent platform.
d. Easy to use by anyone, not limited to data mining specialists.
e. Flexible facilities for conducting experiments.
f. Regularly updated with the addition of new features and algorithms.

2.5 Evaluation Stage: This evaluation is carried out to determine whether the implemented modeling is appropriate and suitable for the research case, and whether it aligns with the intended objectives. The results of this evaluation can then be used to decide the next steps whether the process can be continued or needs to be repeated from the beginning due to not meeting the objectives (Fahmi *et al.,* 2021).

This stage is the testing process of the clusters formed using the k-means clustering algorithm with the Silhouette Coefficient (SC). The Silhouette Coefficient method, developed by Rousseeuw in 1987, is used to assess the quality of a cluster. Another function of this method is to identify the degree of membership of each object in a cluster (Paramartha *et al.,* 2017). The Silhouette Coefficient integrates clustering validation methods, namely cohesion, which measures how closely related objects are within a cluster, and separation, which measures how far apart the clusters are from each other (Pramesti *et al.,* 2017).

The steps to calculate the Silhouette Coefficient are as follows (Lubis & Tamam, 2022):
a. For each object *i*, calculate the average distance from object *i* to all other objects within the same cluster. This average value is called *a*.
b. For each object *i*, calculate the average distance from object *i* to objects in other clusters. From all these average distances, take the smallest value. This value is called *b*.
c. Then, the Silhouette Coefficient for object *i* is obtained.

To calculate the value of the Silhouette Coefficient, also known as the Silhouette Score, the following formula can be used:

$$S = \frac{(b-a)}{Max\ (a,b)} \quad (3)$$

Description:

S : Silhouette Coefficient

a : The average distance from object i to all other objects within the same cluster

b : The smallest average distance to objects in other clusters

The dataset used in the calculation of the Silhouette Coefficient only applies sample data, while the sample size is determined using Slovin's formula as follows (Aryanto *et al.,* 2024):

$$n = \frac{1}{1+N(e)^2} \tag{4}$$

Description:
n : Sample
N : Population
e : Standard error

2.6. Deployment Stage: This stage is carried out once all the previous stages have been completed, or in other words, it is the final stage of the research (Pratiwi *et al.,* 2024). The result of this stage is a report on the k-means clustering calculations of the completed data, presented as the knowledge gained and the recommendations provided at the end of this study.

3. Results dan Discussion

3.1 Results

Based on the method used in this study, namely the CRISP-DM method, the following process stages were carried out:

3.1.1. Business Understanding

Private universities are required to be adaptive in facing the rapid demands of change, both in terms of information technology and government policy shifts in the field of education in Indonesia. The purpose of this research is to determine the appropriate learning method to be applied. The observation activity was carried out by directly observing behaviors, interactions, and situations relevant to the study, with the aim of understanding the existing social and cultural context (Nartin *et al.,* 2024). The expected output of this research is the benefit of identifying the most suitable and effective learning method, which will positively impact students' academic performance.

3.1.2. Data Understanding

This dataset originates from primary data, obtained from records of lecture implementation prior to COVID-19, during the COVID-19 period, and after COVID-19. The data used is in the form of Excel documents with the *.xlsx* format. The dataset consists of learning method implementation results from 2015 to 2020, described with 16 attributes, namely: NPM, Name, Academic Year Code, Semester, Course Code, Course Name, Class, Time, Room, Midterm Exam, Practical Midterm Exam, Final Exam, Practical Final Exam, Assignment, Total Score, and Letter Grade, amounting to 262,075 records. Attribute selection based on data tailored to research needs. The attributes of exam scores (midterm, final, practical) and assignments were selected because they directly reflect students' academic performance, while the total score serves as the main aggregate indicator in clustering. The letter grade functions as a validation variable to assess the consistency of the cluster results. Attributes such as academic year, semester, course, class, time, and room provide administrative context, whereas NPM and name act only as identifiers and are not used in the clustering process. Data exploration was then carried out by selecting attributes relevant to the research and verifying data quality through manual checks using Microsoft Excel and automated checks with Orange. The attribute Semester, course, class, time, room, and assessment scores reflect how learning modes influence implementation and student performance, while total score and letter grade show overall outcomes, only NPM and name serve as identifiers and are not analytically relevant. The Silhouette Coefficient is used to assess the quality of clustering results by evaluating how well an object fits within its assigned cluster compared to other clusters. In this context, the NPM is relevant as a unique identifier that ensures each student can be traced back when evaluating the Silhouette value, since the presence of NPM allows clustering results tested with the Silhouette Coefficient to be mapped back to specific individuals, thereby facilitating the analysis of differences among students within the same cluster as well as across clusters .If there are no missing values and no data duplication occurs, the dataset is considered to be of good quality (Fimawahib & Rouza, 2021).

3.1.3. Data Preparation

Data preparation was conducted through preprocessing to ensure that the data processed with k-means clustering was of high quality. The preprocessing included data reduction, data cleaning, and data transformation. Preprocessing is carried out using a widget called Preprocessing. Data reduction was performed by removing unnecessary attributes, leaving five attributes: NPM, Name, Academic Year Code, Total Score, and Letter Grade. Data cleaning was carried out by eliminating missing values in attributes with many empty or null entries, ensuring greater accuracy in k-means processing. Data transformation was applied to inconsistent entries, such as the use of commas and periods in the *Total Score* attribute. Furthermore, only data from undergraduate (S1) programs were selected (Table 1). After cleaning, 68,289 records remained. To ensure fairness in the distribution of courses from the curriculum of each S1 program over the three years, only odd-semester data were selected, resulting in a final dataset of 65778 records.

Tabel 1. Dataset

NPM	Name	Academic Year Code	Total Score	Letter Grade
…..	…..	…..	…..	…..
1211010054	Rifki Ahmad Maulana	20181	74.5	B
1211019003	Sony Adriansyah	20181	67	C
1311010032	Arby Kurniawan	20181	60.25	C
1511010021	Bobby Pramulia Putra	20191	75.68	B
1511050033	Richardo Martinus	20191	75.79	B
…..	…..	…..	…..	…..
1611050018	Ronaldy Dwi Putra	20201	67.55	B
1612110323	Shenda Vita Dewi	20201	82.23	A
1512120084	Agustinus Frando Suherwanto	20201	85	A
…..	…..	…..	…..	…..

3.1.4. Modelling

The algorithm used in this study is the K-Means Clustering algorithm. The tool employed is the Orange application version 3.37, which is open-source and can be easily installed independently through the following website link: https://orangedatamining.com/. The version of Orange available on the website is always up to date.

The first process carried out is determining the initial centroids, which is done randomly and automatically in Orange using the Interactive K-Means widget**,** producing a visualization as shown in Figure 3 below:

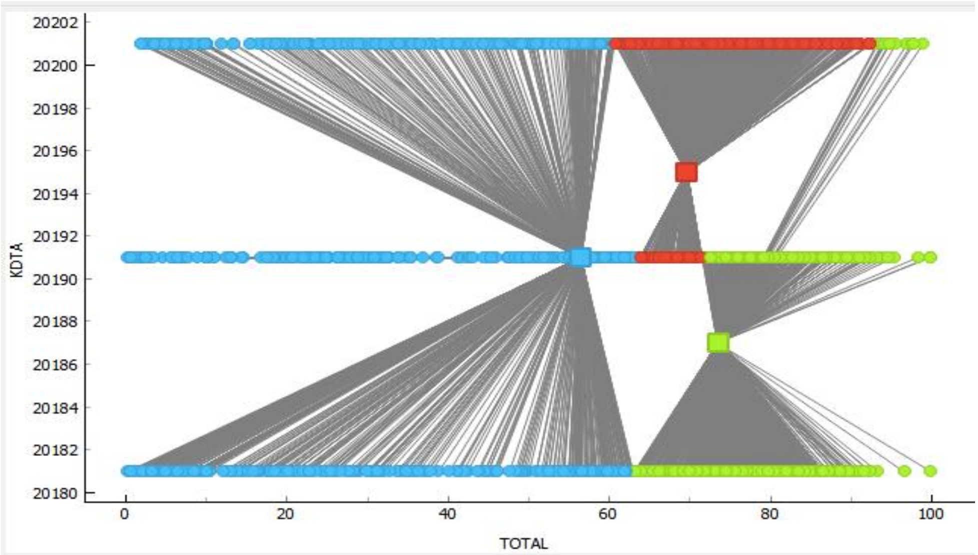

Figure 3. Determination of Initial Centroid

The model design uses widgets in the Orange data mining application for implementing the k-means clustering process which can be seen in Figure 4 below:

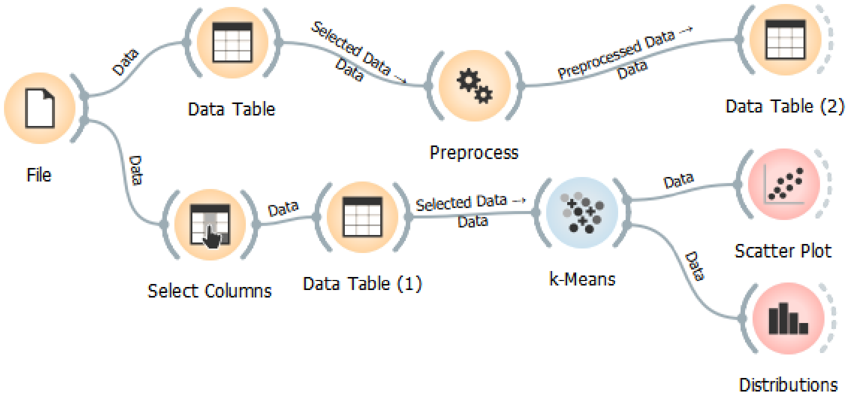

Figure 4. K-Means Method

Several widget functions in the Orange application or tool for the modeling stage used in this study include:

1. *File* is used to call or connect the dataset with the application so that it can be read, allowing the dataset to be processed and its patterns identified.
2. *Data Table* is used to view the details of the dataset used, such as attributes, number of records, data content, and so on.
3. *Select Columns* is used to select attributes that will be used or processed, since some attributes may not be required.
4. *Preprocess* is used to perform dataset preprocessing, such as cleaning the data, removing missing values, and ensuring that the dataset has good quality and is ready for further processing.
5. *K-Means* functions to perform the clustering process using the K-Means algorithm.
6. *Scatter Plot* functions to generate a dataset visualization in the form of a scatter plot.
7. *Distributions* functions to display visualizations in the form of distributions with bar charts.
8. *Silhouette Plot* is used to display a dataset visualization of silhouette samples based on selected attributes in the form of a plot (used in the evaluation stage).

After the K-Means calculation process, the visualization results using the Scatter Plot widget from the dataset in the application of the learning method based on *Total (Axis X)* and *Grade (Axis Y)* can be seen in Figure 5 below:

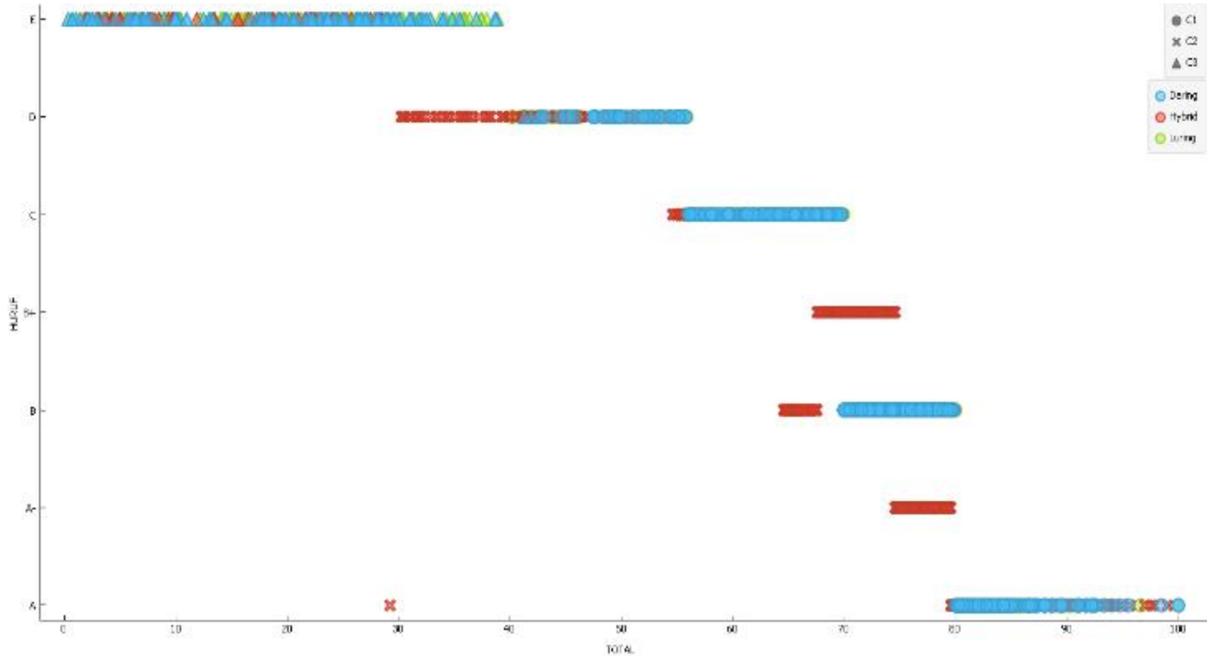

Figure 5. K-Means Scatter Plot

Figure 5 above shows student grades range from 0 to 100, and The letter grades consist of A, A-, B, B+, C, D, and E. That many students received an E grade with a total score range of 0–40, and the fewest received an A- grade with a total score range of 75–80, across the implementation of all learning methods, namely online, hybrid, or offline. The Academic Year Code represents a specific period within the academic calendar. This period is closely related to institutional policies in implementing learning methods, such as offline before the pandemic, online learning during the Covid-19 pandemic, and hybrid learning in the post-pandemic period. Thus, this attribute indirectly reflects the external context that determines the learning method applied in a given period. Therefore, the Academic Year Code is relevant as a determinant of the type of learning method, since each academic year code correlates with the instructional model adopted during that period. C1, C2, and C3 are determined as the optimal clusters based on clustering analysis with a Silhouette ≈ 0.57, indicating fairly clear cluster separation. Each cluster is distinguished by key student characteristics, such as study program, learning method, and scores: C2 is the largest cluster because most students study offline with scores ≥79.44 (grade B), while C1 and C3 represent groups with different score ranges or learning methods.

Meanwhile, the visualization using the Distribution widget from the dataset of the applied learning methods based on Total (Axis X) and Frequency (Axis Y) can be seen in Figure 6 below:

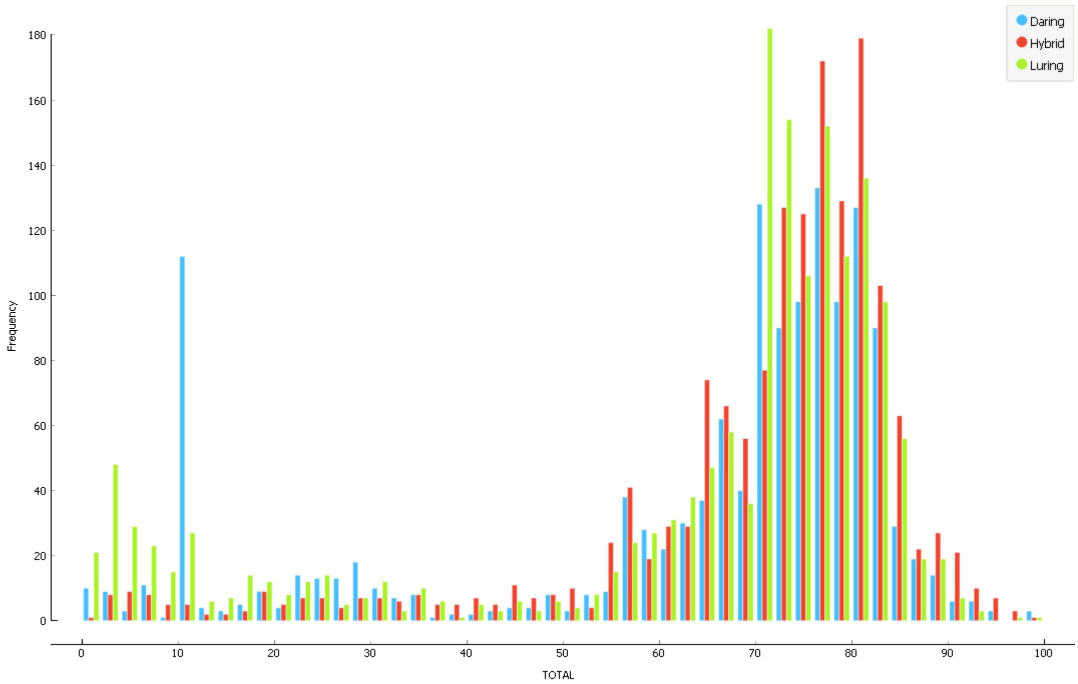

Figure 6. Distribution K-Means

From Figure 6 above, explain the distribution of k-means results it can be concluded that the highest frequency of Total scores is found in the offline learning method (shown in green), the medium frequency of Total scores is found in the hybrid learning method (shown in red), and the lowest frequency of Total scores is found in the online learning method (shown in blue). In addition to using the frequency of the data visualized in the image, additional analysis has been carried out using feature statistics which function to understand academic performance by looking at student grade trends, assessing the distribution of online/offline/hybrid learning methods, conducting initial analysis for clustering as a way to determine which attributes are the most varied and relevant to be used in grouping. The application of this additional analysis yielded the following data Table 2:

Table 2. Feature Statistics

Name	Mean	Median	Mode	Min	Max	Dispersion
Silhouette	0.572811	0.564436	0.645899	0.472664	0.645899	0.0632986
KDTA			<20186			1.1
TOTAL			≥79.44			1.39
HURUF			B			1.76
PRODI			TI			1.39
METODE			Offline			1.1
CLUSTER			C2			1.03

The table indicates that the clustering quality is fairly good (Silhouette ≈ 0.57), the majority of students come from the IT study program, follow offline learning methods, and the largest cluster is C2, representing a group of students with scores ≥79.44, equivalent to a letter grade B.

3.1.5. Evaluation

In order for the evaluation or testing stage in the Orange application to be carried out to calculate the Silhouette Score, data sampling using the Slovin formula shows that the minimum sample size that can be taken from this study is 7% of the dataset, resulting in 4,605 records. This was done to ensure the process could be completed more quickly. A sample is a subset of the population whose characteristics will be studied, selected because it can represent the entire population, thus its size may be smaller than the population (Tarigan *et al.*, 2024).

Then the process continues with the model design using widgets as shown in Figure 7 below:

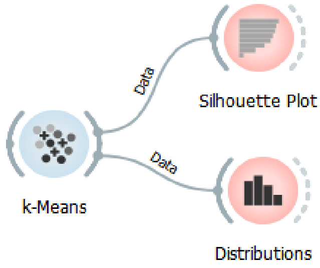

Figure 7. Evaluation K-Means Design

The visualization results of the dataset for applying the learning method with the Silhouette Plot widget based on Silhouette Score and Cluster can be seen in Figure 8 below:

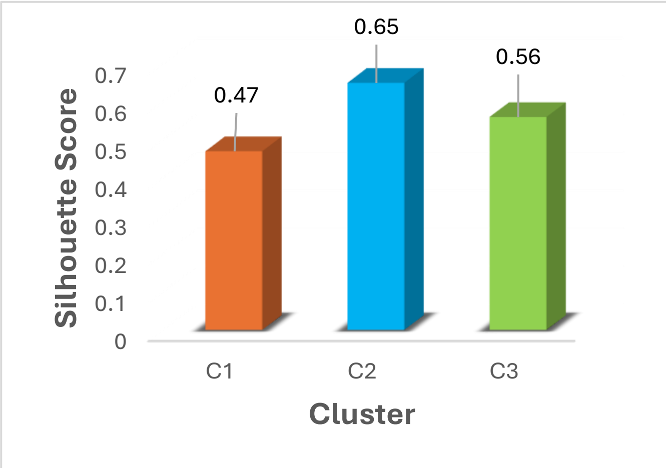

Figure 8. Silhouette Coefficient

The average Silhouette Coefficient value of each object within a cluster is a measure that indicates how tightly the data is grouped in that cluster. The interpretation of the Silhouette values, based on Kaufman and Rousseeuw, is presented in Table 1 below:

Table 3. Silhouette Coefficient Values

Scale	Description
0.7 < SC <= 1	*Strong Structure*
0.5 < SC <= 0.7	*Medium Structure*
0,25 < SC <= 0.5	*Weak Structure*
SC <= 0.25	*No Structure*

Based on Figure 8 and Table 3, the silhouette score indicates that Cluster C2 obtained the highest value of 0.65, with a description of medium structure or moderate value.

The visualization using the Distribution Silhouette widget can be seen in Figure 9. The test results with Silhouette Score in Figure 9 show that the minimum index value in cluster C1 is 0.47, the medium in cluster C3 is 0.56, and the maximum is in cluster C2 at 0.65.

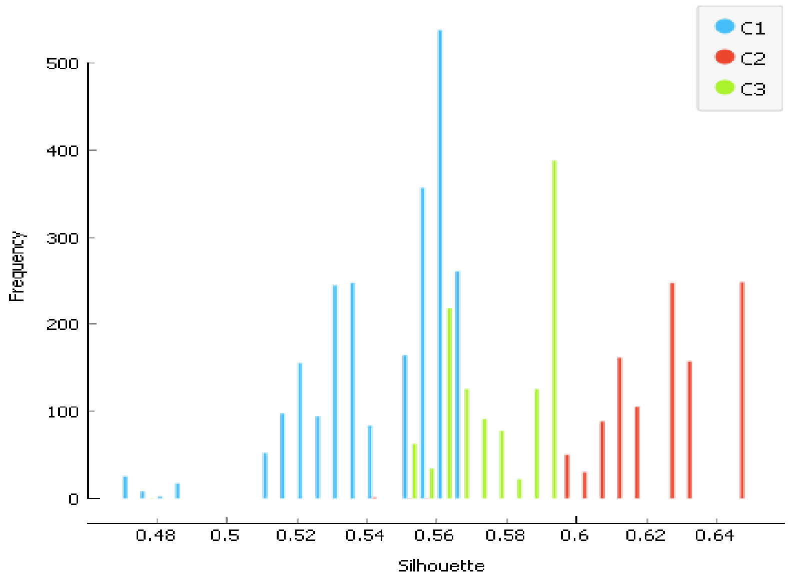

Figure 9. Distributions Silhouette

3.1.6. Deployment

From the implementation of clustering using k-means for each cluster formed (C1, C2, and C3) based on the Total value of each cluster, the Silhouette Coefficient results showed the lowest (minimum) score in cluster C1 of 0.47, medium in cluster C3 of 0.56, and the highest (maximum) in cluster C2 of 0.65. The quality of clusters obtained from the k-means algorithm falls into the medium structure category, which means it belongs to a reasonable cluster category. This category is considered good and aligns with the initial research objectives, therefore the results of this study can be implemented.

3.2. Discussion

Investigations explicitly examining the accuracy of applying online, hybrid, and offline learning methods have not yet been undertaken, although extensive research has addressed hybrid learning implementation (Gerawati *et al.,* 2025; Nazla *et al.,* 2025; Grahani & Priambudi, 2024). Prior to the COVID-19 pandemic, instructional activities were delivered exclusively through conventional offline formats. During the pandemic, however, the offline approach faced substantial limitations due to public health protocols aimed at mitigating viral transmission. As a result, hybrid learning evolved in the post-pandemic era as a blended instructional model that integrates face-to-face and online modalities within a single learning framework.

The results indicate that offline learning is the most effective because direct interaction and a more conducive learning environment enhance students' understanding and engagement. The medium-level cluster structure reflects that, although student characteristics vary, consistent grouping patterns can still be identified. The recommended sequence of offline–hybrid–online learning corresponds to differences in effectiveness, infrastructure readiness, and students' adaptability. Meanwhile, the medium-range Silhouette values suggest that the clusters are sufficiently distinguishable, although not completely separated.

The results of this study yield several significant implications for the formulation of learning strategies in higher education. First, the indication that offline learning attains the highest levels of frequency and effectiveness highlights the need for institutions to prioritize face-to-face instruction as the main approach under normal circumstances to enhance learning outcomes. Second, the medium structure cluster quality suggests that data driven segmentation can serve as a dependable basis for designing more responsive and adaptable learning policies. Third, the recommended hierarchy of learning modalities offline followed by hybrid and online provides policymakers with a framework for aligning instructional models with contextual demands, particularly during transitional or crisis periods. Lastly, this study offers pathways for future research through the integration of classification methods for predictive and forecasting purposes, thereby further advancing learning analytics and reinforcing evidence based decision making within higher education settings.

It is suggested that future studies explore alternative clustering algorithms, increase the scope of datasets to improve generalizability, and employ classification techniques to enable prediction and forecasting of learning patterns. Furthermore, the inclusion of additional variables could offer deeper insights, while empirical testing in real-world settings is necessary to assess the effectiveness of the proposed learning strategies.

4. Conclusion

This research was conducted to determine the appropriate learning method strategy using the k-means clustering algorithm. The method applied was CRISP-DM. The results and visualizations from the k-means process show that the highest total data frequency was found in offline learning, the medium frequency was in hybrid learning, and the lowest frequency was in online learning. Meanwhile, the testing results using the Silhouette Coefficient method showed the lowest (minimum) score in C1 of 0.47, medium in C3 of 0.56, and the highest (maximum) in C2 of 0.65. This category falls into the medium structure cluster, which means it is a good cluster and aligns with the research objectives, thus the results of this study can be implemented. Based on the total value of each cluster, the recommended learning methods are: first (1) offline learning, second (2) hybrid learning, and third (3) online learning. Under normal conditions, offline learning is most suitable, while during pandemic conditions, hybrid learning is preferable compared to online learning. Nevertheless, all methods can be chosen and implemented by adjusting to the recommendations and regulations set by both the central and local governments. The clustering method employed in this study can be further extended by conducting research that applies classification techniques such as Regression, Decision Trees, Naive Bayes, and other methods commonly used in data mining, which can subsequently be utilized for prediction, forecasting, and other analytical purposes.

5. References

Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2021). The application of K-means Clustering for Province Clustering in Indonesia of The Risk of The COVID-19 Pandemic Based on COVID-19 Data. *Springer*. https://doi.org/10.1007/s11135-021-01176-w

Alasali, T., & Ortakci, Y. (2024). Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications. *Journal of Computer Science*, *9*(1). https://doi.org/10.53070/bbd.1421527

Ambarita, J., Jarwati, & Restanti, D. K. (2020). *Pembelajaran Luring*. Adanu Abimata.

Andiopenta, & Aripudin. (2021). Pengembangan Model Pembelajaran Sosiolinguistik Berbasis Hybrid Learning Melalui BORG and GALL Model Pada Mahasiswa Prodi Pendididkan Bahasa dan Sastra Indonesia FKIP Universitas JAMBI 2019/2020. *Jurnal Inovasi Penelitian*, *1*(9), 2011–2017. https://doi.org/10.47492/jip.v1i9.395

Aryanto, P., Indrayana, I., Imanuddin, B., Satriyanto, P., & Yamin, M. (2024). Minat Berwirausaha Mahasiswa Kewirausahaan Universitas Yatsi Madani Tahun 2023. *Journal of Entrepreneurial Behaviour and Research (JUBIR)*, *1*(1), 1–10.

Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi (JATI)*, *13*(2), 122–137. https://doi.org/10.34010/jati.v13i2.10090

Awaludin, M., & Gani, A. G. (2024). Pemanfaatan Kecerdasan Buatan Pada Algoritma K-Means Klastering Dan Sentiment Analysis Terhadap Strategi Promosi Yang Sukses Untuk Penerimaan Mahasiswa Baru. *Jurnal Sistem Informasi (JSI)*, *11*(1). https://doi.org/10.35968/jsi.v11i1.1120

Fahmi, R. N., Jajuli, M., & Sulistiyowati, N. (2021). Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma K-Means. *Journal of Information Technology and Computer Science (INTECOMS)*, *4*(1), 67–79. https://doi.org/10.31539/intecoms.v4i1.2413

Fimawahib, L., & Rouza, E. (2021). Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian. *Jurnal INOVTEK POLBENG - Seri Informatika*, *6*(2), 234–247. https://doi.org/10.35314/isi.v6i2.2096

Gerawati, A. P., Wardani, K., Yusuf, P. S. N., Elmenes, F. A., Rustam, N., & Khairunnisa, A. (2025). Efektivitas Model Hybrid Learning Dalam Meningkatkan Kualitas Pembelajaran. *Jurnal Riset Ilmiah (SINERGI)*, *2*(7), 3019–3030. https://doi.org/10.62335/sinergi.v2i7.1474

Grahani, F. O., & Priambudi, S. (2024). Model Pembelajaran Hybrid Menggunakan Aplikasi Virtual Class Guna Meningkatkan Literasi Digital Mahasiswa. *Jurnal Pendidikan Informatika Dan Sains*, *13*(2), 154–164. https://doi.org/10.31571/saintek.v13i2.8174

Gunawan, A. R., Sudarmin, S., & Rais, Z. (2024). Applied the Self Organizing Maps (SOM) Method for Clustering Educational Equity in South Sulawesi. *Journal of Mathematics and Applied Science (JARRUS)*, *4*(1), 6–19. https://doi.org/10.35877/mathscience2607

Hasdyna, N., Dinata, R. K., & Yafis, B. (2025). Optimizing K-Means Algorithm Using the Purity Method for Clustering Oil Palm Producing Regions. *Jurnal Informatika Sunan Kalijaga (JISKA)*, *10*(1). https://doi.org/10.14421/jiska.2025.10.1.1-15

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Infromatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Hermawati, F. A. (2013). *Data Mining*. CV Andi Offset,.

Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Penerapan K-means Clustering Berbasis Machine Learning untuk Deteksi Penipuan Keuangan. *Academic Journal of Science and Technology*, *10*(1), 33–39. https://doi.org/10.54097/74414c90

Hutagalung, J., Ginantra, N. L. W. S. R., Bhawika, G. W., Parwita, W. G. S., Wanto, A., & Panjaitan, P. D. (2021). COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm. *Journal of Physics: Conference Series*, *1783*(012027). https://doi.org/10.1088/1742-6596/1783/1/012027

Ikhsan, E. (2021). Penerapan K-Means Clustering dari Log Data Moodle untuk Menentukan Perilaku Peserta pada Pembelajaran Daring. *Jurnal Sistem Informasi (SISTEMASI)*, *10*(2). https://doi.org/10.32520/stmsi.v10i2.1285

Khalif, A., Hasanah, A. N., Ridwan, M. H., & Sari, B. N. (2024). Clustering Poverty Levels in Indonesia Using the K-Means Algorithm. *Generation Journal*, *8*(1), 54–62. https://doi.org/10.29407/gj.v8i1.21470

Lestari, W. A., Kartika, K. P., & Budiman, S. N. (2022). Klasterisasi Siswa Berdasarkan Hasil Belajar Menggunakan K-Means Berbasis Web (Studi Kasus: TK. Prima Insan Sholeh Talun). *Jurnal Mahasiswa Teknik Informatika (JATI)*, *6*(1). https://doi.org/10.36040/jati.v6i1.4261

Lubis, D. J., & Tamam, M. B. (2022). Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor. *Jurnal Ilmiah Teknologi-Informasi & Sains (TEKNOIS)*, *12*(1), 7–20. https://doi.org/10.36350/jbs.v12i1

Nartin, N., Faturrahman, F., Deni, A., Santoso, Y. H., Paharuddin, P., Suacana, I. W. G., Indrayani, E., Utama, F. Y., Tarigan, W. J., & Eliyah, E. (2024). *Metode Penelitian Kualitatif*. Cendikia Mulia Mandiri.

Nazla, N., Sai'dah, Z., Lestari, R. A., & Husain, M. (2025). Strategi Hybrid Learning dalam Pengembangan Pembelajaran Pendidikan Islam di Era Society 5.0 Studi Kasus di SMA Darussalam Blokagung. *Journal of Islamic Education Management (MANAGIERE)*, *4*(2), 113–134. https://doi.org/10.35719/managiere.v3i2.2348

Negara, G. A. J. (2025). Hybrid Learning Strategy (Offline-Online) for Hindu Religion Subjects at Dwijendra Senior High School, Denpasar, in the Post-Pandemic Era. *Jurnal Penelitian Agama Hindu*, *9*(3), 282–298. https://doi.org/10.37329/jpah.v9i3.4548

Paramartha, G. N. W., Ratnawati, D. E., & Widodo, A. W. (2017). Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(9), 813–824.

Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, *139*. https://doi.org/https://doi.org/10.1016/j.datak.2022.102013

Pramesti, D. F., Lahan, Tanzil, F. M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(9), 723–732. https://doi.org/10.1109/EUMC.200 8.4751704

Pratiwi, N. K. T. Y., Wasundhari, P. A. E. D., Nikova, K., & Mahendra, G. S. (2024). Rekomendasi Hotel Di Kawasan Lovina Menggunakan Sistem Pendukung Keputusan Dengan Metode WASPAS. *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, *5*(1), 30–40. https://doi.org/10.55122/junsibi.v5i1.1146

Pratiwi, S. A., & Ginting, S. H. N. (2024). Penerapan Algoritma k-means dalam Data Mining untuk Mengidentifikasi Strategi Promosi di Politeknik Ganesha Medan. *Jurnal & Penelitian Manajemen Informatika Polgan (MINFO POLGAN)*, *13*(1), 189–196. https://doi.org/10.33395/jmp.v13i1.13509

Prayoga, Y., Mahmudi, A., & Zahro, H. Z. (2021). Penerapan Metode K-Means Pada Sistem Informasi Akademik Untuk Pengelompokkan Siswa Berprestasi Di UPT SMA NEGERI 3 Kota Pasuruan Berbasis Web. *Jurnal Mahasiswa Teknik Informatika (JATI)*, *5*(2), 822–828.

Pujiono, S., Astuti, R., & Basysyar, F. M. (2024). Implementation of Data Mining to Identify Product Sales Patterns Using K-Means Clustering. *Jurnal Mahasiswa Teknik Informatika (JATI)*, *8*(1), 615–620. https://doi.org/10.36040/jati.v8i1.8360

Rachmatin, D. (2014). Aplikasi Metode-Metode Agglomerative Dalam Analisis Klaster Pada Data Tingkat Polusi Udara. *Infinity Journal*, *3*(2). https://doi.org/10.22460/infinity.v3i2.p133-149

Ridzki, M. M., Hadijah, I., Mukidin, M., Azzahra, A., & Nurjanah, A. (2023). K-Means Algorithm Method for Clustering Best-Selling Product Data at XYZ Grocery Stores. *International Journal of Social Service and Research (IJSSR)*, *3*(8), 3354–3367. https://doi.org/10.46799/ijssr.v3i12.652

Setianingsih, N. A. (2021). *Manifestasi Hybrid Learning di Masa Pandemi*. Sabda Cinta: Persembahan DPD ADRI Jawa Timur untuk Khasanah Tridharma.

Shedriko, & Firdaus, M. (2022). Penentuan Klasifikasi Dengan CRISP-DM dalam Memprediksi Kelulusan Mahasiswa pada Suatu Mata Kuliah. *Seminar Nasional Riset Dan Teknologi (SEMNAS RISTEK)*. https://doi.org/10.30998/semnasristek.v6i1.5814

Tarigan, N. L. L., Wijaya, P. S. M., Wahyuni, Y., & Sulistyowati, E. (2024). Analisis Tingkat Loyalitas Konsumen Generasi Z terhadap Marketplace di Indonesia Menggunakan Metode NPS (Net Promoter Score). *Jurnal Manajemen Strategis (MANTRA)*, *1*(1), 21–34. https://doi.org/10.30588/jmt.v1i01.1222

TIM DPD ADRI Jawa Timur. (2021). *Persembahan DPD ADRI Jawa Timur untuk Khasanah Tridharma di Masa Pandemi*.

Verawati, V., & Desprayoga, D. (2019). Solusi Pembelajaran 4.0: Hybrid Learning. *Prosiding Seminar Nasional Pendidikan Program Pascasarjana UNIVERSITAS PGRI PALEMBANG*. https://jurnal.univpgri-palembang.ac.id/index.php/Prosidingpps/article/view/2739

Widyastuti, A. (2021). *Optimalisasi Pembelajaran Jarak Jauh (PJJ), Daring Luring, BDR*. Elex Media Komputindo.

Wiemer, H., Drowatzky, L., & Ihlenfeldt, S. (2018). Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. In *Science Direct, (12th CIRP Conference on Intelligent Computation in Manufacturing Engineering*. Gulf of Naples, Italy. https://doi.org/https://doi.org/10.3390/app9122407