Ranah Research
Journal of Multidisciplinary Research and Development
E-ISSN: 2655-0865
DINASTI RESEARCH
082170743613 ranahresearch@gmail.com https://jurnal.ranahresearch.com

DOI: https://doi.org/10.38035/rrj.v8i6
https://creativecommons.org/licenses/by/4.0/
Received: 14, Semptember 2026 | **Revised:** 19, September 2026 | **Accepted:** 15, Oktober 2026

Implementation of the K-Means Algorithm to Cluster Levels of Violations by Islamic Boarding School Students in the Area of Ubudiyah

Siti Hotija[1], Imam Wahyudi[2]
[1]Information Technology Program, Annuqayah University, East Java, Indonesia, sitihotija2304@gmail.com
[2]Information Technology Program, Annuqayah University, East Java, Indonesia, imamwahyudi@ua.ac.id

Corresponding Author: sitihotija2304@gmail.com[1]

Abstract: Violations in the field of ubudiyah committed by students can affect the discipline building process within the Islamic boarding school environment. The large amount of stored violation data has not been optimally utilized to identify patterns of violations, so the discipline building process is still carried out in a general manner. This study aims to implement the K-Means algorithm to cluster the severity of religious discipline violations committed by students at the Annuqayah Lubangsa Putri Islamic Boarding School. The research method used is Knowledge Discovery in Databases (KDD), which includes data cleaning, data transformation, normalization using the Min-Max method, determining the number of clusters using the Elbow method, the clustering process using the K-Means algorithm, and evaluation using the Silhouette Score. The dataset consists of 1,022 student records with 58 variables related to religious discipline violations. The results indicate that the optimal number of clusters is three (K=3). The clustering results yielded Cluster 0 with 983 students (96.18%), Cluster 1 with 14 students (1.37%), and Cluster 2 with 25 students (2.45%). The evaluation using the Silhouette Score yielded a value of 0.6491, indicating that the quality of the clustering falls into the good category. These findings demonstrate that the K-Means algorithm is capable of clustering data on student violations based on the similarity of characteristics, thereby providing a foundation for school administrators to develop more effective, objective, and data driven guidance strategies.

Keywords: K-Means, Data Mining, Clustering, Elbow Method, Student Misconduct.

INTRODUCTION

Advances in information technology over the past few years have greatly significantly contributed to improving the quality of data management in various educational institutions, including Islamic boarding schools. The use of technology is not limited to learning activities but has also expanded to data driven monitoring, evaluation, and decision making systems (Raj & Paliwal, 2022). Research indicates that data mining is an effective approach for managing

big data to identify specific patterns, regularities, or trends, thereby assisting institutions in formulating accurate policies. This technique has been widely applied in analyzing student behavior, evaluating academic performance, and mapping discipline levels within the context of Islamic education (Berkah & Zamroni, 2024).

Islamic boarding school, as religions based educational institutions, have a strict disciplinary system, particularly regarding religious practices. The performance of obligatory and recommended congregational prayers, orderly participating in religious activities, and adherence to religious practices are integral parts of the character building process for students. However, violations of religious practice rules still occur frequently occur, such as arriving late for congregational prayers, missing mandatory activities, or neglecting routine acts of worship. Recording these violations are generally done manually, making it difficult to identify patterns, analyze the severity of violations, and conduct a comprehensive evaluation of the mentoring process (Berkah & Zamroni, 2024). This situations results in the available data nott being optimaly untilized to support evidence based decision making.

Sereval previous studies have applied data ming techniques to analyze studens violations. A study on patterns of student violations in the area of religious practice the Apriori algorithm was able to identify correlations between types of violatons but did not produce clustering results (Ridla & Munazilin, 2024). Another study oclustered student violations based on sans sans tions using clustering techniques but did not specifically focus on religions practice violations (Sari et al., 2024). Furthermore, research on similar violatons did not place my emphasis on aspects of religious practice (Hakim et al., 2024; Wahyuni & Arianto, 2024). Thus, specialized studies on the clustering of religious practace violations within Islamic boarding schools remain limited. In line with this, the K-Means algorithm is widely used in the fields of education and Islamic boarding schools. The study (Dila et al., 2025) applies K-Means to cluster student academic performance, the study (Lubis & Tamam, 2022) uses it to determine scholarship eligibility for students; and the study employs the K-Means algorithm to analyze students Qur'an memorization scores based on their fluency and tajwid scores (Najib & Rasiban, 2023). The results of these studies indicate that K-Mens is effective in forming clusters of data with similar characteristics, and the novelty of this study lies in its specific focus on analyzing students violations of religious practices using the K-Mens algorithm. In addition, this study combines data preprocessing using the Min-Max method with cluster number optimization using the Elbow method based on the Sum of Squared Error (SSE) value. This approach has not been widely applied in studies of students' violations of religious practices and is expected to produce a more objective, measurable, and practical mapping of violation levels for the development of Islamic boarding schools.

Meanwhile, the K-means algorithm has become one of the most widely used clustering methods in educational research for grouping student behavior (Chi, 2021; Mawarni & Budi, 2022), and mapping specific academic achievements (Dila et al., 2025; Lubis & Tamam, 2022). K-mens has the advantage of producing clear and easily interpretable group divisions, making it highly suitable for categorizing violation levels into categories such as minor, moderate, and severe. Applying this method to data on disciplinary violations is expected to provide an objective picture of the students disciplinary status and assist the administration in developing more targeted guidance strategies.

At the Annuqayah Lubangsa Putri Islamic boarding school, the recording of violations of religious obligations is still done manually. Ideally, this record-keeping should be done digitally and analyzed to generate information about trends in violations, groups of students requiring special attention or rehabilitation, and the effectiveness of the disciplinary policies in place. However the available data has not yet been used for further analysis. This gap between the ideal and the actual situation highlights the need for a more comprehensive, data-driven system for analyzing violations (Raj & Paliwal, 2022).

In light of the above discussion, the objective of this study is to apply the K-means algorithm to categorize the levels of violations of religious practices (ubudiyah) committed by students at the Anuqayah Lubangsa putri Islamic Boarding School. The findings of this study are expected to contribute to a more systematic mapping of violations of religious practices, support the disciplinary evaluation process, and serve as a basis for the development of more effective and measurable guidance policies.

METHODS

This study descriptive analytical research aimed at categorizing the levels of violations committed by Islamic boarding school students in the area of ubudiyah. The research approach used is *data mining* with the K-Means Clustering algorithm, which is widely applied in the categorization of educational data (Chi, 2021; Mawarni & Budi, 2022). Data collection was conducted through documentary and literature reviews. The data analysis method began with a data preprocessing stage using Min-Max normalization to standardize the scale across attributes (Maharana et al., 2022), followed by optimization of the number of clusters using the Elbow Method (Almohammedi & Abido, 2020; Ishak & Bengnga, 2022) based on the Sum of Squared Error (SSE) value (Wahyudi et al., 2021; Rodi et al., 2024; Hartanti et al., 2024; Irawan et al., 2025). The research process is presented in the form of a research flowchart.

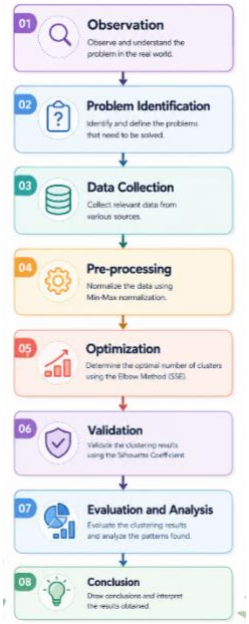


Figure 1. Research Flowchart

Based on the resech flowchart, the data analyzed consists of violations of religious obligations by female students at the Annuqayah Lubangsa Islamic Boarding School, comprising approximately 1.022 studenst. The variables used include the students age, educational level, and type religious violation, when were then normanized and clustered using K-mens alghoritm.

Data Mining

Data mining is the process of processing and analyzing large amounts of data to discover hidden patterns, relationships, or trends that can be used as a basis for decisionmaking. In an educational context, data mining is used to analyze student behavior and discipline, as well as for data driven evaluation (Raj & Paliwal, 2022). The data mining approach transforms previously unstructured data into meaningful information that can be objectively used in the institutional evaluation process.

K-Means Algorithm

The K-Means algorithm is a non hierarchical clustering method that aims to divide data into *k* clusters based on the degree of similarity among the data points. Each cluster is represented by a centroid, and the clustering process is performed by minimizing the distance of each data point from its nearest centroid (Oti et al., 2021).

K-mens Algorithm Formula

a) Euclidean Distance Formula

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - c_{jk})^2}$$

Description:
d(x,μi) = distance of data poin *x* to the i-th centroid
xj = the j-th attribute value of the data point
μij = the j-th attribute value of the i-th centroid
n = number of attribute

b) Centroid Update Formula

$$\mu_i = \frac{1}{|C_i|}\sum_{x \in C_i} x$$

Description:
μi = the i-th centroid
Ci = the set of data in the i-th cluster
|*Ci*| = number of data points in the i-th centroid

Min-Max Normalization

Min-Max normalization is used to equalize the range of values across attributes so that no single variable dominates the clustering process. This technique transforms data values into a range of 0 to 1 (Maharana et al., 2022). This normalization is important because the K-Means algorithm uses distance calculations as the basis for data clustering.

Min-Max Normalization Formula

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

The normalization is important in K-Means because the algorithm uses distance calculations as the basis for data clustering, where *x′* is the normalized value, *x min* is the origin*al data value, x max is the minimum value, and is the maximum value of that attribute.*

Elbow

The Elbow Method is used to determine the optimal number of clusters by analyzing changes in the elbow value (SSE) for each number of clusters. SSE measures the total squared distance between the data and their cluster centroids (Wahyudi et al., 2021; Rodi et al., 2024; Hartanti et al., 2024; Irawan et al., 2025).

Sum Of Squared Eror (SSE) Formula

$$SSE = \sum_{j=1}^{k}\sum_{x_i \in C_j}(x_i - c_j)^2$$

A smaller SSE value indicates that the data within a cluster are more homogeneous. In this study, SSE was used both as the basis for determining the number of clusters and for evaluating the clustering results; thus, the *Elbow* method and SSE were applied in an integrated and consistent manner.

RESULTS AND DISCUSSION

Data Selection

The first stage in the *Knowledge Discovery in Databases* (KDD) process is data selection. This stage aims to select and prepare the data to be used in the clustering process using the K-Means algorithm. The dataset used in this study consists of data on religious discipline violations by Islamic boarding school students, comprising 1,022 records with 58 variables, including student identity, age, education level, and various types of violations. The dataset was then imported into Google Colab using the Pandas library to ensure the data was successfully loaded and ready for processing in the next stage.

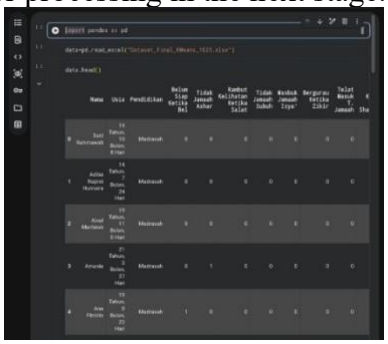


Figure 2. Importing the Dataset and *data.head*

The import result show that the dataset successfully loaded into the system. The fist five record were displayed using head() function to verivy that the data structure matched the research dataset before proceeding to the *preprocessing* stage.

Preprocessing / Data Cleaning

After the data *selection* process was completed, teh next step was *preprocessing* or data clening. This step aimed to asure that the data to be used in the clustering process had the appropriate structure so that is cloud be processed by K-mens alghorithm. In the step, the dataset attributes were checked, unused attributes were removed, and the data format was adjusted to prepare it for the trasformations stage.

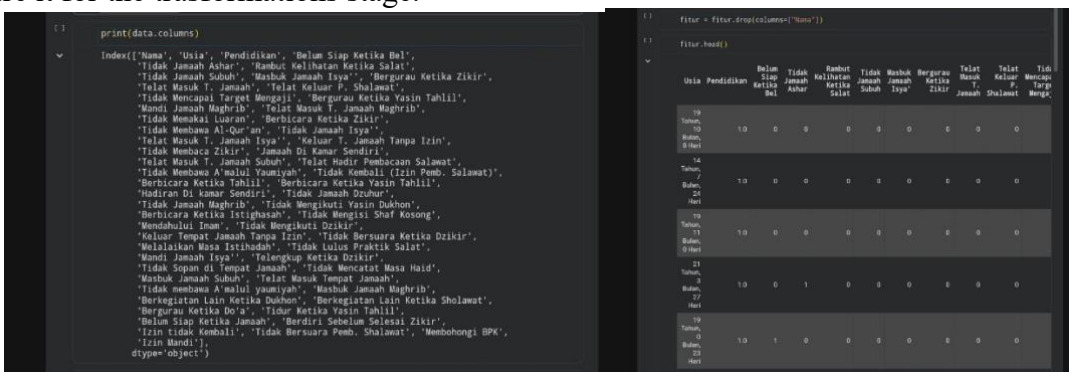


Figure 3. Data Preprocessing Process

Based on the *preprocessing* result, all attributes were checked to ensure that each variable had been read correctly. Next, the Name attribute was removed because it served only as the student's identifier and had no impact on the clustering process. Thus, the dataset used consisted only of variables relevant to the analysis using the K-Means algorithm.

Data Normalization Result

Before the clustering process using the K-Means algorithm was carried out, the data first underwent a preprocessing stage that included data cleaning, data transformation, and normalization using the Min-Max method. Normalization aims to standardize the range of values for each attribute to 0 to 1 so that all variables have the same scale and do not dominate the Euclidean distance calculation process in the K-Means algorithm. The dataset used in this study consists of 1,022 student records with 58 variables, including age, education level, and type of religious conduct violation. The results of the normalization of a portion of the data are shown in Table 1.

Table 1. Data Normalization Result (Fist 5 Data Points)

No	Age	Education	Not Ready When the bel	Dit Not Atted Ashar	Hair Visible During Prayer (Shalat)
1	0,122951	0,250000	0,000000	0,000000	0,000000
2	0,081967	0,250000	0,000000	0,000000	0,000000
3	0,122951	0,250000	0,000000	0,000000	0,000000
4	0,139344	0,250000	0,000000	0,111111	0,000000
5	0,122951	0,250000	0,500000	0,000000	0,000000

Source: Data processing result using Python (2026)

Based on Table 1, the normalization results shown represent only the first 5 data points as a sample of the entire dataset. The normalization process was actually performed on 1,022 student data points and the 58 variables used in the study. All attributes were successfully transformed into a range of values from 0 to 1 using the Min-Max method. The Age and Education variables were successfully normalized so that they have a uniform scale, while violation variables such as Not Ready When the Bell Rings, Did Not Attend Ashar Prayer, and Hair Visible During Prayer are also within the same range of values. Presenting a portion of the data in the table aims to illustrate the normalization results, while the entire set of normalized data is used as input for the clustering process using the K-Means algorithm so that the clustering process can be carried out more objectively and accurately.

Determining the Number of Clusters Using the Elbow Method

After all the data has been normalized, the next step is to determine the optimal number of clusters using the Elbow method. This method involves calculating the Sum of Squared Error (SSE) for various numbers of clusters (K), ranging from K = 1 to K = 10. The SSE value represents the total squared distance between each data point and the centroid of its respective cluster. The smaller the SSE value, the better the level of data homogeneity within a cluster. However, increasing the number of clusters excessively does not always result in a significant improvement in clustering quality. Therefore, the Elbow method is used to determine the most optimal number of clusters Table 2. Nilai Sum of Squared Error (SSE).

Table 2. Sum of Squared Error (SSE) Values

Jumlah Cluster (K)	Nilai SSE
1	315,698
2	285,912
3	265,400
4	256,314
5	252,499
6	246,549
7	236,427
8	199,581
9	192,395
10	189,767

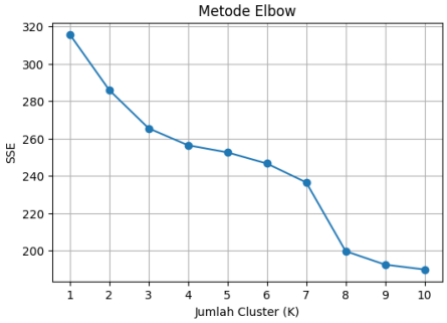


Figure 4. Elbow Method Graph

Based on Table 2 and Figure 4, the Sum of Squared Error (SSE) value decreases as the number of clusters increases. A significant decrease in the SSE value is observed from K = 1 to K = 3, after which the rate of decrease begins to level off at K = 3. This condition indicates the formation of an elbow point, so three clusters were selected as the optimal number of clusters in this study. Selecting K = 3 is considered capable of producing good data clustering without unduly increasing the model's complexity. Therefore, the clustering process in this study was conducted using 3 clusters: Cluster 0, Cluster 1, and Cluster 2.

Clustering Results Using the K-Means Algorithm

Based on the results of data processing using the K-Means algorithm with an optimal number of 3 clusters, all 1,022 student data points were successfully grouped into three clusters based on the similarity of their religious violation characteristics. The clustering results are presented in Table 3.

Table 3. Clustering Result Using the K-mens Alghorithm

Cluster	Number of Data Points	Percentage (%)
Cluster 0	983	96,18
Cluster 1	14	1,37
Cluster 2	25	2,45
Total	1.022	100,00

Source: Data processing result using Python (2026)

Based on Table 3, the K-Means algorithm successfully grouped 1,022 student data points into three clusters. Cluster 0 is the group with the largest number of members, namely 983 students (96.18%), so it can be said that the majority of the data share relatively similar characteristics. Cluster 2 consists of 25 students (2.45%), while Cluster 1 has the fewest members, namely 14 students (1.37%). The difference in the number of members across each cluster indicates that the data distribution is uneven. This is a common characteristic of the K-Means algorithm because the clustering process is based on the degree of similarity in data characteristics, not on the even distribution of members across each cluster. These clustering results are then used as the basis for analyzing the characteristics of each cluster based on centroid values, thereby identifying the patterns of religious practice violations that dominate in each group.

Characteristics of Each Cluster

The characteristics of each cluster were obtained based on the average value (centroid) of each variable resulting from the clustering process using the K-Means algorithm. The centroid value indicates the tendency of violation characteristics in each cluster. The higher the average value of a variable, the higher the tendency for that violation to occur in the cluster in question. A summary of the characteristics of each cluster is presented in Table 4.

Table 4. General Characteristics of Each Cluster

Cluster	General Characteristics	Nomber of Data Point
Cluster 0	Has a low average number of violations across nearly all religious observance indicators, indicating students with a high level of discipline.	983
Cluster 1	Has a low average of violatons Has a higher average number of violations on several indicators compared to Cluster 0, particularly regarding certain violations, thus requiring special attention.	14
Cluster 2	Has the highest average violation scores on several religious practice indicators, indicating a group of students with a relatively higher level of violations.	25

Source: Data analysis results using Python (2026)

Based on Table 4, the clustering results show that each cluster has distinct characteristics. Cluster 0 is the largest group with 983 students and is characterized by low average violation scores across nearly all religious observance indicators. Cluster 1 consists of 14 students who tend to have higher average violation scores on certain indicators compared to Cluster 0. Meanwhile, Cluster 2, comprising 25 students, exhibits the highest average scores for several types of violations, such as failing to attend the Ashar congregational prayer, failing to attend the Subuh congregational prayer, and failing to meet the Quran recitation target; thus, this group requires greater attention in the discipline-building process within the pesantren environment.

Evaluation of Clustering Result Using the Silhouette Score

To determine the quality of the clustering results, an evaluation was conducted using the Silhouette Score. This method is used to measure the degree of similarity between data points and the cluster to which they belong, compared to other clusters. The Silhouette Score ranges from -1 to 1, where a value closer to 1 indicates better clustering quality.

Table 5. Evaluation Clustering Results Using the Silhouette Score

Evaluation Method	Value
Silhouette Score	0,6491

Source: Data proccessing result using Python (2026)

Based on Table 5, the evaluation results using the Silhouette Score yielded a value of 0.6491. This value indicates that the clustering results using the K-Means algorithm are of good quality, as they fall within the range of 0.50-0.70. This suggests that members within each cluster share a sufficiently high level of similarity, while the differences between clusters are also quite distinct. Thus, the K-Means algorithm is capable of effectively clustering data on students violations in the ubudiyah domain, so that the clustering results can be used as a basis for identifying the characteristics of violations in each group of students.

Discussion

Based on the clustering results using the K-Means algorithm, three clusters with distinct characteristics of violations in the ubudiyah domain were identified. The number of clusters was determined using the Elbow method, which indicated that the optimal number of clusters is three (K = 3). Furthermore, the evaluation using the Silhouette Score yielded a value of 0.6491, indicating that the quality of the clustering results falls into the good category. This value indicates that the data within each cluster exhibit a suffciently high level of similarity and are clearly distinct from the other clusters, allowing the clustering results to serve as a basis for decision making. The clusrering results show that Cluster 0 has the largest number of members, namely 983 students (96.18%), indicating that the majority of students exhibit relatively low levels of violations in the ubudiyah domain. Meanwhile, Cluster 1, consisting of 14 students (1.37%), and Cluster 2, consisting of 25 students (2.45%), exhibit distinct disciplinary violation characteristics, thus requiring more intensive attention and guidance from the boarding school's administrative staff.

These findings align with the research by Mawarni and Budi (2022), which demonstrated that the K-Means algorithm can cluster student disciplinary data based on the similarity of characteristics, thereby facilitating the evaluation and decision-making processes. The similarity between that study and this one lies in the use of the K-Means algorithm to cluster disciplinary behavior data, while the difference lies in the research subject: this study focuses on violations related to students' religious practices within the boarding school environment. Furthermore, the results of this study are supported by research by Lubis and Tamam (2022), which concluded that the K-Means algorithm is effective for clustering student data based on specific characteristics, thereby supporting an objective decision-making (Mawarni & Budi, 2022; Lubis & Tamam, 2022).

The results of this study indicate that the application of the K-Means algorithm was able to cluster data on religious discipline violations committed by students at the Annuqayah Lubangsa Putri Islamic Boarding School into three clusters based on the similarity of violation characteristics. These clustering results provide more systematic information regarding the students' disciplinary status, enabling the administrative staff to identify groups of students requiring more intensive guidance. Students in Clusters 1 and 2 can be prioritized in guidance and evaluation activities, while students in Cluster 0 still require routine guidance to maintain their level of discipline. Thus, the clustering results not only provide information on patterns of violations in the area of ubudiyah but can also serve as a basis for developing more effective, objective, and data-driven guidance policies at the Annuqayah Lubangsa Putri Islamic Boarding School.

CONCLUSIONS AND RECOMMENDATIONS

Penelitian This study successfully applied the K-Means algorithm to cluster the levels of students' violations in the area of ubudiyah at the Annuqayah Lubangsa Putri Islamic Boarding School. Before the clustering process was carried out, the data underwent a preprocessing stage consisting of data cleaning, data transformation, and normalization using the Min-Max method. Determining the number of clusters using the Elbow method indicated that the optimal number of clusters was three (K = 3). The clustering results grouped 1,022 student data points into Cluster 0 (983 students, 96.18%), Cluster 1 (14 students, 1.37%), and Cluster 2 (25 students, 2.45%). Evaluation using the Silhouette Score yielded a value of 0.6491, indicating that the quality of the clustering falls into the good category. The results of this study demonstrate that the K-Means algorithm is capable of grouping student disciplinary violation data based on the similarity of characteristics, thereby serving as a foundation for school administrators to develop more effective, objective, and data-driven guidance strategies.

REFERENCES

Almohammedi, A., & Abido, M. (2020). RES on power operation: K-means clustering over elbow approach. *WSEAS Transactions on Power Systems, 15*, 214–221. https://doi.org/10.37394/232016.2020.15.25

Berkah, D., & Zamroni, M. A. (2024). Management of Islamic boarding school shapes the character of santri discipline. *CJoTL, 3*(2), 147–159. https://doi.org/10.31538/cjotl.v3i2.1109

Chi, D. (2021). Research on the application of K-means clustering algorithm in student achievement. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 435–438). IEEE. https://doi.org/10.1109/ICCECE51280.2021.9342164

Dila, R., Defit, S., & Arlis, S. (2025). Analisis algoritma K-means clustering dalam pengelompokan prestasi belajar siswa menengah atas (SMA). *Vol. 5*(5).

Hakim, L., Ridla, M. A., & Munazilin, A. (2024). Penentuan pola pelanggaran santri pada bidang keamanan dan ketertiban dengan menggunakan algoritma Apriori. *Vol. 1*(2).

Hartanti, N. T., Seniwati, E., Pramitasari, R., & Wulandari, I. R. (2024). Jumlah cluster optimal dalam pengelompokan siswa SMK dengan metode Elbow K-means clustering. *JAIS, 4*(2), 67–73. https://doi.org/10.31294/jais.v4i2.7531

Irawan, H., Fawzi, M. I., & Hidayat, S. (2025). Integration of K-means clustering and Elbow method for mapping Baccaurea spp. distribution to support agroindustrial development in West Sulawesi. *J. Tek. Inform. (JUTIF), 6*(2), 1057–1068. https://doi.org/10.52436/1.jutif.2025.6.2.4671

Ishak, R., & Bengnga, A. (2022). Clustering tingkat pemahaman mahasiswa pada perkuliahan probabilitas statistika dengan metode K-means. *JJEEE, 4*(1), 65–69. https://doi.org/10.37905/jjeee.v4i1.11997

Lubis, D. J., & Tamam, M. B. (2022). Penerapan K-means untuk pengelompokkan beasiswa santri di Pondok Pesantren Miftahul Huda Bogor. *Jurnal Teknologi Informasi, 12*(1).

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings, 3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Mawarni, Q. I., & Budi, E. S. (2022). Implementasi algoritma K-means clustering dalam penilaian kedisiplinan siswa. *Journal of Software Engineering, Computer Science and Information Technology (JSON), 3*(4), 522. https://doi.org/10.30865/json.v3i4.4242

Mawarni, Q. I., & Budi, E. S. (2022). Implementasi algoritma K-means clustering dalam penilaian kedisiplinan siswa. *Journal of Software Engineering, Computer Science and Information Technology (JSON), 3*(4), 5221. https://doi.org/10.30865/json.v3i4.4242

Najib, M. A., & Rasiban. (2023). Penerapan algoritma K-means untuk analisis nilai tahfidz santri berdasarkan nilai fashohah dan tajwid (studi kasus: Pesantren Modern At-Taqwa Bogor). *JIMIK, 4*(3), 1670–1679. https://doi.org/10.35870/jimik.v4i3.392

Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive review of K-means clustering algorithms. *IJASRE, 07*(08), 64–69. https://doi.org/10.31695/IJASRE.2021.34050

Raj, S., & Paliwal, M. (2022). Higher education dashboard implementation using data mining and data warehouse: A review paper. *IJIRCST*, 107–111. https://doi.org/10.55524/ijircst.2022.10.1.19

Ridla, M. A., & Munazilin, A. (2024). Penentuan pola pelanggaran santri pada bagian ubudiyah menggunakan algoritma Apriori. *Vol. 1*(2).

Rodi, M., Hendrik, Amir Bagja, M. Nurul Wathani, & Zaenul Amri. (2024). Perbandingan metode optimasi penentuan sentroid awal pada algoritma K-means menggunakan Elbow PSO dan SSE. *JINTEKS, 6*(4), 794–801. https://doi.org/10.51401/jinteks.v6i4.4803

Sari, T. P., Simanjuntak, M., & Prahmana, I. G. (2024). Implementasi data mining untuk mengelompokkan data tingkat pelanggaran santri berdasarkan sanksi yang diberikan.

Wahyudi, I., Sulthan, M. B., & Suhartini, L. (2021). Analisa penentuan cluster terbaik pada metode K-means menggunakan Elbow terhadap sentra industri produksi di Pamekasan. *Jatim, 2*(2), 72–81. https://doi.org/10.31102/jatim.v2i2.1274

Wahyuni, D., & Arianto, A. (2024). Penerapan data mining pada data pelanggaran lalu lintas menggunakan metode K-means clustering (studi kasus: Pengadilan Negeri Dumai).