Variance: Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X
Volume 7 Issue 2 | October 2025
10.30598/variancevol7iss2page231-240

COMPARATIVE STUDY OF CLUSTERING ALGORITHM TO DETERMINE STUDENT PROFILES BASED ON ACADEMIC ABILITY

Ilham Faishal Mahdy[1*], Mohammad Wildan Nurul Haqqi[2], Rafif Naufal Oktiardi[3]
[1,2,3]Statistics Study Program, Faculty of Mathematics and Natural Science, Universitas Islam Bandung
Jl. Tamansari No. 1, Bandung, 40116, West Java, Indonesia

Corresponding Author's Email: *ilham.faishal@unisba.ac.id*

Abstract: *Competition among similar study programs requires each Program to develop strategies to compete for new students, especially for private universities like Universitas Islam Bandung. One of the study programs at Universitas Islam Bandung, namely Statistics, has experienced a downward trend in new student enrollment over the last 3 years. One strategy that can be employed is to identify student profiles. Creating student profiles can serve as an evaluation tool for improving educational quality and the quality of student input. One method for finding patterns in data, in this case, student profiles, is clustering. In this research endeavor, a comparison of clustering algorithms for selecting the best model was conducted using K-Means, Complete Linkage, Average Linkage, and Ward's Method. The outcomes show that Average Linkage with two clusters was selected as the best model with the minimum DBI value. The Average Linkage clustering reveals that the academic performance of students in cluster one surpasses that of students in cluster two. The promotional strategy can be focused on cluster 1 to improve the quality of student input.*

Keywords: *academic ability, clustering, student profiles*

1. INTRODUCTION

Competition among similar study programs requires each Program to develop strategies to attract new students, especially at private universities. Higher education is the level of education that leads to a degree and can influence the development of younger generations' talents and abilities, making them more beneficial to community life. With the passage of time, it cannot be denied that universities are engaged in fierce competition to attract prospective students [1]. The competition between public and private higher education institutions is currently very intense, driven by the many methods these institutions use to attract new students [2].

Universitas Islam Bandung (UNISBA) is one of the private higher education institutions in Bandung, Indonesia. As one of the best private Islamic universities in Bandung, UNISBA has received the highest accreditation rating and enjoys a strong reputation. Despite receiving this designation, admissions of new students at UNISBA have declined across several undergraduate programs. One of the study programs at UNISBA, Statistics, has experienced a decline in new student enrollment over the past 3 years. One strategy is to identify student profiles. The creation of student profiles can serve as a basis for evaluation to improve the quality of education and student input. In addition, student profiles can serve as a foundation for determining promotional strategies for schools that produce graduates with strong academic abilities.

One method for determining the student profile is clustering. Currently, clustering algorithms continue to evolve. In this case, the algorithms to be used are K-Means and Agglomerative Hierarchical Clustering (AHC) methods, such as Complete Linkage, Average Linkage, and Ward's Method. Each of these computational methods has its advantages and disadvantages compared to the others. The merit of K-Means lies in its capacity to manage extensive datasets while maintaining a comparatively rapid computational efficiency, though it is sensitive to outliers and to the initial selection of centroids [3]. The AHC method is a hierarchical clustering technique that progressively combines the most similar clusters, based on preset metrics or distances, after merging n clusters into a single cluster [4].

In this study, we used a different approach than in the previous study. Before creating a clustering model, we first performed anomaly detection using DBSCAN, a robust algorithm for detecting noise in large datasets [5].

 Submitted:2025-08-01 *Accepted: 2025-12-08*

After that, it continues with a comparison of clustering algorithms to produce optimal clusters that achieve the research objectives. Clustering results are evaluated using DBI. Thus, the determination of the student profile will be based on the best algorithm, namely the one that produces the smallest DBI value. This research is expected to serve as a basis for determining targeted promotional activities and for producing higher-quality student input.

2. METHODOLOGY

2.1. Data Source

The data utilized in the present research were procured from a repository that is maintained by the Information and Technology System Development Section (PSITEK) of UNISBA. The collected data consists of 144 students who are at least in their 6th semester in the Statistics Study Program at UNISBA. The variables used in this study are the final grades of several core courses present in the curriculum, such as Statistical Methods, Introduction to Probability, Data Management, Exploratory Data Analysis, Mathematical Statistics, Regression Analysis, Time Series Analysis, Analysis & Design of Surveys, and Multivariate Statistics.

2.2. Detection of Multicollinearity

Variance Inflation Factor (VIF) analysis is conducted to identify potential multicollinearity. A VIF value which surpasses the threshold of 10 signifies that the associated predictor demonstrates significant multicollinearity with the other variables. Nevertheless, if the resulting value is less than 10, then the variable can be considered free from multicollinearity. VIF can be calculated using the following equation [6]:

$$VIF_i = \frac{1}{1 - R_i^2} \tag{1}$$

in which context R_i^2 represents the proportion of variation in the i-th predictor that can be explained by all the other predictors. Meanwhile, VIF_i is the VIF value of the i-th variable. The multicollinearity problem can be addressed using Principal Component Analysis (PCA).

2.3. Principal Component Analysis

Principal Component Analysis (PCA) constitutes a statistical technique that condenses a group of interrelated original variables to reduced quantity of newly generated variables that exhibit no correlation with one another [7]. The steps taken to determine the principal components are as follows [8]:

1) Standardizing data so that it has equivalent value scales.
2) Calculate the covariance matrix to determine the correlation of each variable.
3) Calculate the eigenvalue using the equation:

   $$\det(A - \lambda I) = 0 \tag{2}$$

   where A is the covariance matrix and λ is the eigenvalue. This value reflects the magnitude of data variance in a specific direction.
4) Calculating eigenvectors using the equation:

   $$(A - \lambda I)v = 0 \tag{3}$$

5) Calculating the proportion of variance for each component.

   $$P_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \tag{4}$$

   P_i indicates the proportion of variance from component i, and λ_i is the corresponding eigenvalue.
6) Calculate the cumulative proportion of variance.

   $$CP_r = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{j=1}^{n} \lambda_j} \tag{5}$$

with $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. The eigenvalues are sorted from the largest to the smallest to facilitate the identification of the requisite number of components necessary for accurately depicting the data.

7) Final transformation into the principal component space

$$\boldsymbol{X} = \boldsymbol{ZW} \tag{6}$$

where Z is the standardized data and W is the eigenvector matrix.

2.4. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) represents a partition methodology that emphasizes the compactness of point distributions. The spatial aspect of this algorithm concerns the position and proximity of points in the data when forming a cluster, which depends on density. In DBSCAN, density refers to the formation of clusters: areas with high density are considered clusters. Conversely, areas that are sparse or have low density are considered as noise [9].

DBSCAN operates based on two essential parameters: the neighbourhood radius ε (epsilon) and the minimum number of points, minPts, which together determine whether a region can be considered dense. With these settings, the algorithm categorizes data into core points, border points, and noise. A point is labelled as a core point when it has at least minPts other points located within its ε-distance, indicating that it lies in the dense center of a cluster. In contrast, a border point is situated within the ε-range of a core point but lacks a sufficient number of a nearby points to be classified as a core point itself. Noise refers to the atypical data points that do not qualify as either core points or boundary points. Additionally, DBSCAN does not need prior assumptions about how many clusters exist [10]. The value of ε is determined using the k-distance graph obtained from the K-Nearest Neighbour algorithm. Then, the value of minPts is obtained from the calculation of $(2 \times \text{dimension of data}) - 1$ [11].

In this study, the DBSCAN method is used for anomaly detection. When DBSCAN is used for anomaly detection, the primary interest shifts from identifying clusters to identifying noise points. Observations allocated to a particular cluster exhibit analogous characteristics with a minimum of minPts neighboring points, thereby rendering them improbable anomalies. The absence of their capacity to either establish or integrate into a cluster signifies a substantial divergence in their attributes in comparison to those represented by the clustered data [12]. Observations within a cluster will be further analyzed using additional clustering techniques to identify groups with similar characteristics.

2.5. K-Means Clustering

K-Means method works by separating objects into different groups, placing objects with identical characteristics into one group and those with different characteristics into another. Objects within the same cluster differ from those belonging to other clusters [13]. The primary procedures inherent in the K-Means algorithm are delineated as follows:

1) Begin by randomly choosing k points to serve as the initial centroids.
2) Allocate every observation to the cluster whose centroid exhibits the least distance from it. The proximity between two objects is calculated using the Euclidean distance, is given by Equation (7).

$$d(x, y) = \sqrt{\sum_{j=1}^{p} \left(x_j - y_j\right)^2} \tag{7}$$

where x_j is the value of the j-th variable on object x, y_j is the value of the j-th variable on object y, p denotes the quantity of variables employed in the analysis.

3) Recalculate the centroid of each group by calculating the average of all data points included in that group.
4) The processes outlined in Steps 2 and 3 are iteratively executed until there is an insignificant alteration in the centroid's location or the data points cease to transition into alternative clusters.

5) The process stops when the stopping criteria are met, for example, when there are no further changes in the centroid's position.

K-Means performs clustering by partitioning observations into a predetermined number of groups, denoted as K. For example, when K = 2, the algorithm assigns the data points into two separate clusters.

2.6. Complete Linkage

The Complete Linkage algorithm merges clusters A and B using a distance metric defined as the maximum pairwise distance separating any point in cluster A from any point in cluster B [14]. The Complete Linkage distance is expressed mathematically as:

$$d(A,B) = max\{d(y_i, y_j)\}; A \epsilon y_i; B \epsilon y_j \tag{8}$$

the term $d(y_i, y_j)$ denotes the Euclidean distance or any other chosen metric between the vectors y_i and y_j. During each iteration, distance $d(A,B)$ is computed for all possible cluster pairs, and the algorithm then fuses the pair of clusters whose maximum inter-point distance is greatest.

2.7. Average Linkage

The Average Linkage algorithm operates by clustering data based on the average distance among the entire dataset. In the Average Linkage approach, the inter-cluster distance between clusters A and B is ascertained s by calculating the mean of all pairwise distances that are present between the points in cluster A and those in cluster B [14]. This metric can be computed utilizing the subsequent formula:

$$d(A,B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j) \tag{9}$$

At each step of Average Linkage, we use $d(A,B)$ to identify the amalgamation of the two clusters characterized by the minimal distance.

2.8. Ward's Method

Ward's method is a widely adopted for linkage strategy in hierarchical agglomerative clustering that aims to ensure clusters remain internally homogeneous. The underlying principle of this method is to identify two clusters whose combination results in the smallest additional squared deviation between the observations and the centroid of the merged cluster [15]. The expression presented in equation (10) quantifies this by measuring how much the sum of squared errors (SSE) increases when the clusters are merged [16].

$$\Delta SSE = SSE(AB) - [SSE(A) + SSE(B)] \tag{10}$$

where SSE(.) takes the form:

$$SSE(A) = \sum_{i=1}^{n_A} (x_i - \bar{x}_A)' (x_i - \bar{x}_A) \tag{11}$$

$$SSE(B) = \sum_{i=1}^{n_B} (x_i - \bar{x}_B)' (x_i - \bar{x}_B) \tag{12}$$

$$SSE(AB) = \sum_{i=1}^{n_{AB}} (x_i - \bar{x}_{AB})' (x_i - \bar{x}_{AB}) \tag{13}$$

In these expressions, $n_{(.)}$ represents the quantity of observations encompassed within a cluster, x_i represents an individual data point within that cluster, and $\bar{x}_{(.)}$ refers to centroid of the cluster. Ward's method proceeds by selecting the sequence of merges that yields the smallest possible increase in ΔSSE at each iteration.

2.9. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) constitutes a quantitative measure employed to assess the efficacy of clustering outcomes. It was selected in this study because it can assess whether a chosen number of clusters is appropriately separated and compact given the clustering algorithm used [17]. As an internal validation measure, DBI evaluates cluster quality based on two key aspects: cohesion (how compact each cluster is) and separation (how distinct different clusters are). The formal definition of DBI is provided in equation (14) [18].

$$DBI = \frac{1}{K}\sum_{i=1}^{K} max_{j\neq i}\left(\frac{\sigma_i + \sigma_m}{d(c_i, c_m)}\right) \tag{14}$$

where σ_i denotes the mean distance from every point within cluster i to its corresponding centroid c_i, σ_j repre denotes the mean distance of points in cluster j to centroid c_j, $d(c_i, c_j)$ refers to the distance between the respective centroids of c_i and c_j, and K specifies the total number of clusters.

3. RESULTS AND DISCUSSION

3.1. Data Description

The exposition of data delineates a comprehensive outline of the dataset employed within the scope of the research. The findings derived from the data exposition are illustrated in Table 1.

Table 1. Summary Statistics

Variable	Minimum	Mean	Maximum
Statistical Methods (X_1)	50.50	81.03	105.85
Introduction to Probability (X_2)	17.71	85.95	104.30
Data Management (X_3)	19.60	86.18	99.30
Exploratory Data Analysis (X_4)	55.50	76.26	92.42
Mathematical Statistics (X_5)	36.90	65.55	85.93
Regression Analysis (X_6)	47	72.85	94
Time Series Analysis (X_7)	49	75.13	97
Analysis & Design of Surveys (X_8)	55.70	84.86	96.85
Multivariate Statistics (X_9)	31.85	74.35	87.08

Upon reviewing the analytical findings presented in Table 1, it is evident that Data Management is the subject with the highest average score compared to other subjects. Conversely, Mathematical Statistics has the lowest average score among the subjects. The average scores for each subject will be used to determine student profiles based on academic ability.

3.2. Detection of Multicollinearity

Detection of multicollinearity issues is performed using VIF calculation. The computational outcome illustrated in Table 2.

Table 2. VIF Calculation

Variable	VIF
Statistical Methods (X_1)	59.10
Introduction to Probability (X_2)	80.37
Data Management (X_3)	91.13
Exploratory Data Analysis (X_4)	72.74
Mathematical Statistics (X_5)	78.85
Regression Analysis (X_6)	210.12
Time Series Analysis (X_7)	127.29
Analysis & Design of Surveys (X_8)	123.28
Multivariate Statistics (X_9)	195.42

Based on the results of the VIF calculation, it was found that all variables have a VIF value greater than 10, thus concluding that multicollinearity issues exist. Therefore, the assumption of non-multicollinearity is not met and will be addressed using the PCA method.

3.3. Principal Component Analysis

This study will use PCA on the original dataset, and variables exhibiting multicollinearity will be transformed into new, lower-dimensional variables that are independent of each other. Below are the results of the cumulative explained variance ratio calculations based on PCA.

Table 3. Cumulative Explained Variance Ratio

PC Component	Cumulative Variance
PC1	47.45%
PC2	62.16%
PC3	71.11%
PC4	78.58%
PC5	84.22%

The PCA results show that the use of three principal components can explain more than 70% of the variance. Accordingly, the three extracted principal components (PC1, PC2, and PC3) will serve as the basis for the next stage of analysis. Here are the three principal components that are formed:

$$PC1 = 0.290\ Z_1 + 0.372\ Z_2 + 0.376\ Z_3 + 0.280\ Z_4 + 0.184\ Z_5 + 0.422\ Z_6 + 0.360\ Z_7 + 0.269\ Z_8 + 0.380\ Z_9$$
$$PC2 = -0.216\ Z_1 - 0.368\ Z_2 - 0.024\ Z_3 + 0.349\ Z_4 + 0.673\ Z_5 + 0.181\ Z_6 + 0.166\ Z_7 - 0.415\ Z_8 - 0.099\ Z_9$$
$$PC3 = 0.715\ Z_1 - 0.122\ Z_2 + 0.165\ Z_3 - 0.223\ Z_4 - 0.096\ Z_5 + 0.036\ Z_6 - 0.014\ Z_7 - 0.623\ Z_8 + 0.026\ Z_9$$

3.4. Anomaly Detection using DBSCAN

DBSCAN is used to separate observations that are considered as noise or anomalies from the clusters that will be formed. The initial step taken is the determination of the DBSCAN parameters, namely ε and minPts.

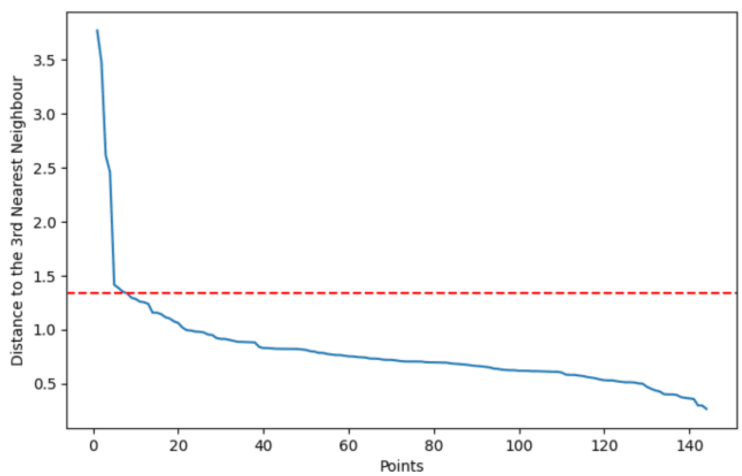

Figure 1. K-Distance Graph to Determine ε Parameter

Based on Figure 1, the value of ε obtained from the K-Distance Graph is 1.336. By using this ε value and the minPts value of 5, DBSCAN modeling is used to detect noise. The results of DBSCAN show that the 5th, 69th, 82nd, and 101st observations are detected as noise. Therefore, this data is separated and not included in any group formed based on the clustering algorithm.

3.5. Comparison of Clustering Algorithms

Data that has been freed from noise is then modeled using several clustering algorithms. The best algorithm is determined by comparing DBI values. Table 4 compares DBI values across clustering algorithms.

Table 4. Comparison of Clustering Algorithms

Algorithm	Number of Clusters	DBI
K-Means	5	0.935
Complete Linkage	2	0.938
Average Linkage	2	0.704
Ward's Method	2	0.795

Based on Table 4, the average linkage algorithm produces the smallest DBI value compared to other algorithms. Therefore, the determination of student profiles will focus on the clustering results obtained from the average linkage algorithm.

3.6. Average Linkage

The initial phase of the average linkage algorithm involves the computation of the distance between various observations. Next, clusters are formed hierarchically using equation (9). The clustering results are visualized in the dendrogram in Figure 2.

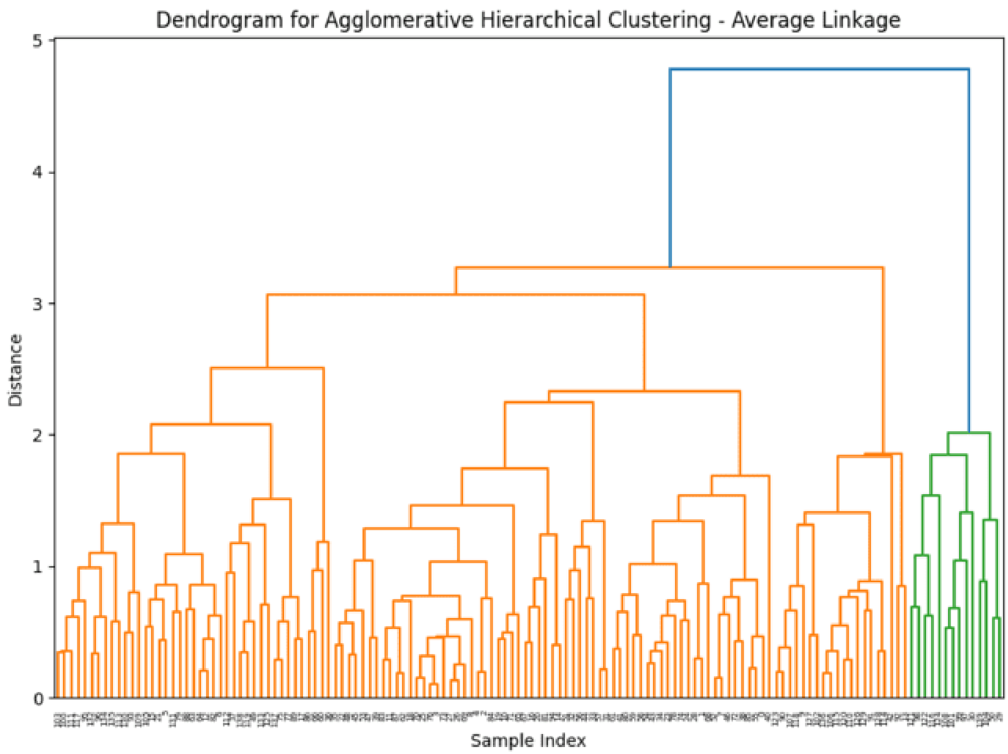

Figure 2. Dendrogram

The dendrogram illustrates the hierarchical merging of clusters based on their distance. The height of each merge on the vertical axis indicates the dissimilarity level at which two clusters are joined. To objectively determine the best-fitting number of clusters, a trial-and-error methodology was employed, which involved comparing DBI values for cluster counts ranging from 2 to 5. The results show that a cluster count of 2 yields the minimum DBI value, as explained in Table 4. In the next step, student profiles will be created based on the formed clusters.

Table 5. Student Profiles Based on Clustering Results

Cluster	Cluster Size	Characteristics Based on Mean								
		Statistical Methods	Introduction to Probability	Data Management	Exploratory Data Analysis	Mathematical Statistics	Regression Analysis	Time Series Analysis	Analysis & Design of Surveys	Multivariate Statistics
1	126	83.59	89.18	89.33	77.57	66.28	74.94	76.55	86.46	75.98
2	14	57.83	63.80	65.75	69.21	64.21	58.50	65.36	76.89	64.11

The clustering results show that 126 observations are in cluster 1 and 14 are in cluster 2. The characteristics of cluster 1, observed based on the average scores of several core subjects, indicate better academic performance compared to cluster 2. The average scores obtained from the members of the first cluster also show higher values compared to the overall average scores, which can be seen in Table 1.

4. CONCLUSION

Based on the research findings, the best clustering algorithm to determine the profile of students in the Statistics Study Program at UNISBA is the Average Linkage, which yields the minimum DBI value. The clustering results using Average Linkage produced 2 clusters with different academic abilities. The characteristics of each cluster indicate that the first cluster has better academic performance compared to the second cluster. This suggests that the profiles of students in the first cluster can serve as a reference for promotional efforts to improve the quality of student input. Promotional activities can be carried out by visiting the schools of the members of the first cluster, and other activities can also be conducted, focusing on the information from the students' profiles in the first cluster.

ACKNOWLEDGMENT

The authors express their gratitude to Universitas Islam Bandung for providing financial support for the dissemination of this research through the Research and Community Services Grants, as well as to all individuals who contributed to the completion of this study.

REFERENCES

[1] Garaika and W. Feriyan, "Promosi dan Pengaruhnya terhadap Animo Calon Mahasiswa Baru dalam Memilih Perguruan Tinggi Swasta," *Jurnal Aktual STIE Trisna Negara*, vol. 16, no. 1, pp. 21–27, 2018.

[2] R. Kanada, "Trend Promosi Perguruan Tinggi yang Ampuh dalam Menarik Minat Mahasiswa Baru (Studi Kasus Perguruan Tinggi di Kota Palembang)," *Journal of Islamic Education Management*, vol. 5, no. 1, pp. 81–92, 2019, [Online]. Available: http://jurnal.radenfatah.ac.id/index.php/El-idare

[3] R. Aulia, N. Julianti, S. F. Putri, L. Efrizoni, and Rahmaddeni, "Optimalisasi Pengelompokan Gangguan Kecemasan dalam Mendukung Tujuan Pembangunan Berkelanjutan Menggunakan Algoritma K-Means dan K-Medoids," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 12, no. 2, pp. 348–360, 2025.

[4] W. Usna and R. Aprilia, "Comparison of agglomerative hierarchical clustering (AHC) algorithm and k-means algorithm in poverty data clustering in north sumatra," *Desimal: Jurnal Matematika*, vol. 7, no. 3, pp. 489–500, 2024, doi: 10.24042/djm.

[5] H. Lu, "Evaluating the Performance of SVM, Isolation Forest, and DBSCAN for Anomaly Detection," in *ITM Web of Conferences*, EDP Sciences, 2025, p. 04012. doi: 10.1051/itmconf/20257004012.

[6] S. Maharani and R. Yotenka, "Pengelompokan Kecamatan di Daerah Istimewa Yogyakarta berdasarkan Jumlah Pengguna Alat Kontrasepsi Tahun 2022 dengan K-Medoids Cluster," *Emerging Statistics and Data Science Journal*, vol. 2, no. 2, pp. 222–237, 2024.

[7] I. A. Rosyada and D. T. Utari, "Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering," *Jambura J. Probab. Stat*, vol. 5, no. 1, pp. 6–13, 2024, doi: 10.34312/jjps.v5i1.18733.

[8] M. D. Salman *et al.*, "Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Province Based on Availability of Facilities and Infrastructure," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 797–806, 2025, doi: 10.57152/malcom.v5i3.1950.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques 3rd Edition*. Waltham: Elsevier, 2012.

[10] M. I. Khelil, M. Ladjal, M. A. Ouali, and H. Bennacer, "A Clustering-Driven Strategy Utilizing DBSCAN for Detecting Outliers in Water Quality Data," *Communication Science et Technologie*, vol. 23, no. 1, pp. 01–12, 2025.

[11] N. Afriyal and I. F. Mahdy, "Pengelompokan Daerah Rawan Gempa Bumi di Pulau Jawa Menggunakan Metode Density Based Spatial Clustering of Application with Noise (DBSCAN)," in *Bandung Conference Series: Statistics*, 2024, pp. 490–499. doi: 10.29313/bcss.v4i2.13873.

[12] Y. Kim and M. Vasarhelyi, "Anomaly detection with the density based spatial clustering of applications with noise (DBSCAN) to detect potentially fraudulent wire transfers," *International Journal of Digital Accounting Research*, vol. 24, pp. 57–91, 2024, doi: 10.4192/1577-8517-v24_3.

[13] M. Suyal and S. Sharma, "A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning," *Journal of Artificial Intelligence and Systems*, vol. 6, no. 1, pp. 85–95, 2024, doi: 10.33969/ais.2024060106.

[14] M. E. Mustafidah and M. D. Purnama, "Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kasus DBD Menggunakan Complete Linkage dan Average Linkage," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 12, no. 2, pp. 337–343, 2024.

[15] M. Urdabayev, A. Kireyeva, L. Vasa, I. Digel, K. Nurgaliyeva, and A. Nurbatsin, "Discovering smart cities' potential in Kazakhstan: A cluster analysis," *PLoS One*, vol. 19, no. 3, Mar. 2024, doi: 10.1371/journal.pone.0296765.

[16] B. Salsabilla, A. Arif, A. Rusgiyono, and A. Hoyyi, "Pengelompokan Provinsi-Provinsi di Indonesia Menggunakan Metode Ward (Studi Kasus: Produksi Tanaman Pangan di Indonesia Tahun 2018)," *JURNAL GAUSSIAN*, vol. 9, no. 1, pp. 112–121, 2020, [Online]. Available: http://ejournal3.undip.ac.id/index.php/gaussian

[17] K. A. Abbas *et al.*, "Unsupervised machine learning technique for classifying production zones in unconventional reservoirs," *International Journal of Intelligent Networks*, vol. 4, pp. 29–37, Jan. 2023, doi: 10.1016/j.ijin.2022.11.007.

[18] F. Ros, R. Riad, and S. Guillaume, "PDBI: a Partitioning Davies-Bouldin Index for Clustering Evaluation," *Neurocomputing*, vol. 528, pp. 178–199, 2023.