Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 135–146
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50061

Genetic Algorithm–Optimized Clustering for University Promotion Target Recommendation

Ulla Delfana Rosiani[1*], Clauria Dwi Putri Nabillah[2], M. Hasan Basri[3], Ahmadi Yuli Ananta[4], Yushintia Pramitarini[5]

***Abstract*—Competition among higher education institutions demands promotional strategies that are more targeted and data-driven. This study proposes a clustering-based recommendation model for determining university promotion targets by integrating Genetic Algorithm (GA) optimization into three clustering methods: K-means, Fuzzy C-means (FCM), and K-medoids. The dataset consists of 925 student records (cohorts 2021–2023) from the Information Technology Department, with the selected attributes including school origin, NPSN, school location (city and province), and GPA. Clustering performance was evaluated using the Davis-Bouldin Index (DBI) and the Silhouette Coefficient as primary metrics, with intra- and inter-cluster distances as supporting indicators. The results show that GA-K-means achieves the best performance at K = 3, with a DBI of 1.2792 and a Silhouette Coefficient of 0.2876, and the improvement is statistically significant ($p < 0.05$). GA optimization also improves FCM performance but does not significantly improve K-medoids performance. Although the GA increases computational time by approximately two to three times, the improvement in clustering quality justifies its use in non-real-time decision-support scenarios. The proposed model enables universities to determine promotion targets in a more objective, adaptive, and data-driven manner, supporting strategic decision-making in higher education promotion.**

***Index Terms*—College promotion, fuzzy C-means, genetic algorithm, K-means, K-medoids.**

I. Introduction

Higher education institutions play a crucial role in shaping national human resources, prompting universities to continuously improve their promotional strategies to attract prospective students effectively [1]. XYZ University is one such institution facing a decline in new student enrollment despite implementing various promotional activities, including alumni outreach, education fairs, and social media outreach. Several challenges remain, including delays in information delivery and limited promotional coverage, which reduce the effectiveness of these efforts. Consequently, universities require a data-driven promotional strategy that can objectively and efficiently identify potential target schools.

One approach that can support this need is data mining, particularly clustering techniques, which enable the grouping of students or schools based on similar characteristics [2]. Clustering is a multivariate analysis method that aims to group objects with high intra-cluster similarity and significant inter-cluster differences [3]. Previous studies have used clustering to analyze student data, such as school of origin, Grade Point Average (GPA), and geographical location, to support promotional decision-making in higher education. Among the widely used clustering algorithms are K-means, Fuzzy C-means (FCM), and K-medoids, each with distinct characteristics and advantages depending on data distribution and application requirements [4]. K-means applies strict cluster boundaries [5], FCM allows flexible membership degrees for data points [6], and K-medoids is more robust to outliers because cluster centers are selected directly from the existing data [7].

However, conventional implementations of K-means, FCM, and K-medoids suffer from a common limitation: random initialization of cluster centers, which can lead to suboptimal clustering results and convergence to local optima [8]. Many previous studies have focused on comparing clustering algorithms using evaluation metrics without addressing this initialization problem, leading to inconsistent cluster quality and reduced reliability when applied to decision-support systems. This limitation is particularly critical in promotion target recommendation, where inaccurate clustering may lead to inefficient allocation of promotional resources.

Furthermore, although K-means clustering has been widely adopted in various applications, the algorithm has inherent weaknesses in determining the optimal number of clusters and in accurately initializing centroids [9]. These issues further undermine clustering performance and underscore the need for

Received: 13 January 2026; Revised: 27 February 2026; Accepted: 3 April 2026.
*Corresponding author
[1]Ulla Delfana Rosiani, Politeknik Negeri Malang, Indonesia (e-mail: rosiani@polinema.ac.id).
[2]Clauria Dwi Putri Nabillah, Politeknik Negeri Malang, Indonesia (e-mail: clauriaputri150303@gmail.com).
[3]M. Hasan Basri, Politeknik Negeri Malang, Indonesia (e-mail: hasanbasr054@gmail.com).
[4]Ahmadi Yuli Ananta, Politeknik Negeri Malang, Indonesia (e-mail: ahmadi@polinema.ac.id).
[5]Yushintia Pramitarini, Institute of Industrial Technology, Korea University, Sejong, Republic of Korea (e-mail: shintia88@korea.ac.kr).

more robust optimization approaches.

To overcome these issues, optimization techniques such as the Genetic Algorithm (GA) have been applied to improve clustering performance by optimizing the initialization of cluster centers. The GA is an evolutionary optimization method that simulates natural selection through selection, crossover, and mutation to search for near-optimal solutions. Prior research has demonstrated that GA-based optimization can enhance clustering stability and quality by reducing dependency on random initialization and improving convergence behavior [10], [11]. Nevertheless, studies that integrate GA with multiple clustering algorithms within a unified framework for university promotion-targeted recommendation remain limited. These findings provide a strong methodological foundation for applying GA-based optimization in clustering models for promotion target recommendation, as explored in this study.

Therefore, this study proposes a clustering-based promotion target recommendation model optimized using the GA. The proposed model applies and compares GA-optimized K-means, Fuzzy C-means, and K-medoids to analyze student data and identify priority schools for university promotion. Clustering performance is evaluated using the Davis-Bouldin Index (DBI) and Silhouette Coefficient to ensure a reliable assessment of cluster compactness and separation. By integrating GA optimization into the clustering process, this study aims to provide a more objective, adaptive, and data-driven basis for determining university promotion targets, thereby supporting strategic decision-making in higher education institutions.

II. Related Work

Several studies have explored the application of clustering techniques to support data-driven decision-making in higher education promotion and related domains. K-means and FCM are among the most widely used algorithms for grouping student data due to their simplicity and effectiveness in identifying patterns from academic and demographic attributes. Reference [12] compared K-means and FCM for determining the university promotion strategy and reported that K-means achieved a higher average Silhouette Coefficient, while FCM demonstrated faster convergence. Similar findings were reported by Kurnia [13], who showed that K-means outperformed FCM in clustering public health data, although FCM provided more flexible membership representation. In contrast, [14] found that FCM produced better results for crime-prone area mapping, indicating that algorithm performance is highly dependent on data characteristics and application context.

In the context of higher education promotion, several studies have utilized clustering to segment prospective or enrolled students. Reference [15] applied K-means clustering to PPMB data to generate targeted promotion strategies based on school origin and study program, while [16] employed FCM to identify potential promotion areas for a higher education institution. Other studies have also explored alternative approaches, such as the predictive apriori method [1] for student profile exploration, decision tree methods for determining promotion locations [17], and K-medoids for university promotion strategy analysis [18]. These studies demonstrate the potential of data mining techniques in supporting promotional decision-making; however, most focus on a single algorithm without systematic optimization or cross-method comparison.

Recent research has highlighted the importance of optimization techniques to improve clustering quality, particularly in addressing the sensitivity of clustering algorithms to random initialization. Rinaldi et al. [8] reported that the GA could reduce the likelihood of local optima in K-means-based clustering. Al Rivan et al. [10] demonstrated that integrating the GA with FCM and K-means improved cluster quality and reduced iteration counts in medical datasets, while Wahyudi et al. [11] showed that GA-optimized K-means produced superior Silhouette values in agricultural data clustering. These findings indicate that GA-based optimization effectively enhances clustering stability and performance across diverse application domains.

Despite these advances, existing studies generally examine GA optimization applied to a single clustering method or within domains unrelated to higher education promotion. Comparative analyses that integrate GA optimization across multiple clustering algorithms, such as K-means, FCM, and K-medoids, within a unified framework for university promotion target recommendation remain limited. Moreover, few studies explicitly position clustering results within a recommendation model that directly supports strategic promotional decision-making.

This research addresses these gaps by proposing a GA-optimized clustering model for recommending university promotion targets. Unlike previous studies that focus on individual algorithms or unoptimized clustering, this study systematically compares GA-optimized K-means, FCM, and K-medoids using consistent evaluation metrics: the Davis-Bouldin Index (DBI) and the Silhouette Coefficient. By integrating clustering optimization and recommendation-oriented analysis within the context of higher education promotion, this study contributes a more objective and adaptive decision-support model for determining priority promotion targets.

III. Research Method

This study adopts a structured research methodology consisting of data collection, data processing based on the CRISP-DM framework, algorithm modeling and comparison, and result analysis. The overall research workflow is illustrated in Figure 1. The workflow is further described as follows:

A. Data Collection

Data collection in this study was conducted using three main sources. First, internal data were obtained from the Department of Information Technology at Politeknik Negeri Malang, comprising 925 student records from the 2021–2023 cohorts.

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 135–146
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50061

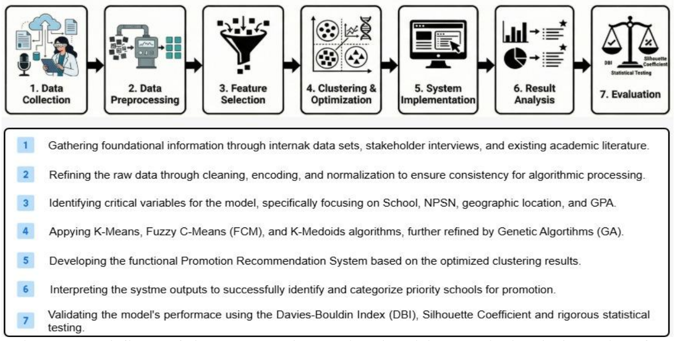

Fig. 1. Workflow of the proposed genetic-algorithm-optimized clustering for university promotion target recommendation.

The data were organized into two types of files for each cohort: student profile data (including student ID (NIM), admission pathway, class, gender, address, school of origin, NPSN, city, and province) and academic data (including GPA per semester, credits, and academic status). From these datasets, only relevant attributes were selected, namely school of origin, NPSN, city, province, and GPA.

Second, interviews were conducted with the Public Relations department to obtain insights into existing promotional strategies and challenges. Third, a literature review of related studies was carried out to support the development of data-driven clustering and optimization approaches for promotion target recommendation.

B. Data Processing

The data processing stage was conducted using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology and implemented in Python. This framework consists of six main stages:

1) *Business understanding*
This stage focuses on identifying the core problem: the decline in student enrollment and the lack of targeted promotion strategies. By analyzing historical student data, this study aims to identify patterns among schools that significantly contribute to enrollment and academic performance (GPA). The objective is to support the development of a data-driven promotion strategy that enables more effective, efficient, and targeted allocation of promotional resources.

2) *Data understanding*
In this stage, the dataset's characteristics and structure are analyzed. The dataset includes attributes such as school of origin, NPSN, city, province, and GPA. This analysis is conducted to understand the data distribution and identify key indicators that reflect the student population and academic performance. These indicators are essential for grouping schools with similar characteristics in the clustering process.

3) *Data Preparation*
In this stage, the dataset is prepared to ensure data quality and consistency before modeling. The preprocessing steps include removing missing or irrelevant data, transforming categorical variables into numerical format (encoding), normalizing data using standard scaling, and merging datasets from multiple cohorts into a unified structure. The result is a clean and structured dataset suitable for clustering analysis.

4) *Modeling*
Three clustering algorithms, namely K-means, FCM, and K-medoids, were applied and compared in this study. Each algorithm was enhanced using GA optimization to improve the initialization of cluster centers. The modeling process begins with determining the optimal number of clusters using the Silhouette method (with the best result at K = 3). The GA optimization process includes initial population generation, chromosome evaluation using the DBI as the fitness function, followed by selection, crossover, and mutation operations. The optimized clustering results are then compared across algorithms to determine the most suitable method for the dataset.

5) *Evaluation*
The performance of the clustering models was evaluated using two primary metrics: DBI and Silhouette Coefficient. A lower DBI value indicates better cluster compactness and separation [19], while a higher Silhouette value indicates more distinct and well-separated clusters [12]. In addition, statistical significance testing was conducted to validate the performance differences between models. These evaluation results were used to identify the best-performing clustering algorithm.

6) *Deployment*
In the final stage, the clustering results were used to develop a recommendation system for determining promotion targets. Schools with larger student populations and better academic performance were identified as priority targets for promotional activities. The system supports decision-making by providing data-driven insights and can be updated periodically using new student data, ensuring adaptability in dynamic institutional environments.

C. Algorithm Comparison

Genetic algorithms are used in this study to optimize centroid selection during clustering. This method overcomes the drawbacks of random initialization, which often produces suboptimal results. Based on the principles of selection and evolution, this approach helps identify more accurate and stable centroid positions [10]. In this study, the GA parameters are set as follows:

1) *Determining the number of clusters*
The optimal number of clusters is determined using the Silhouette method, which assesses clustering quality by measuring the suitability of each data point relative to other clusters. This method calculates a global Silhouette score for each configuration, and the configuration with the highest score is considered optimal. The method ensures

that data within a single cluster are highly similar to each other and sufficiently separated from data in other clusters, resulting in more accurate and relevant segmentation [20].

2) *Determination of initial population*

The population size affects solution quality. Larger population sizes result in longer processing times but do not necessarily yield more optimal results than smaller populations. In addition, a larger population leads to smaller changes in average fitness, and the resulting generations during reproduction tend to resemble their parents due to insufficient exploration capacity [21]. In this study, population sizes ranging from 10 to 40 in increments of 5 were tested, with a final population size of 30.

3) *Chromosome representation and fitness encoding*

In this study, each chromosome represents a candidate configuration of cluster centers. For K-means and FCM, chromosomes are encoded as numerical vectors containing centroid values for all clusters across all feature dimensions. For K-medoids, chromosomes represent candidate medoid indices selected from actual data points, resulting in a more discrete search space. The fitness of each chromosome is evaluated using the DBI, with lower values indicating better clustering quality. Through selection, crossover, and mutation, the population evolves toward better cluster center configurations. The fitness function evaluates each chromosome's quality to determine its viability for the next generation. In each generation, the chromosome closest to the ideal solution is selected based on its fitness score [22]. In the K-means algorithm, data points are grouped around the given centroids, and the clustering results are assessed using the DBI. The K-means method produces clustering labels for fitness evaluation. In the FCM algorithm, a soft-clustering approach is used to assess fitness: each data point has a degree of membership in each cluster. Prior to calculating the DBI value, these membership values form a fuzzy membership matrix, which is then used to determine the final label using argmax to identify the position of the largest value in the array [10].

4) *Selection process*

Each population member is represented as a sequence of fixed-length strings (chromosomes). A selection procedure is needed to determine which members of the population will proceed to the next generation, given a fixed population size N. Before selection, the population is supplemented with the results of genetic operations, such as crossover and mutation. The population and the results of the genetic operations are then selected using specific techniques to obtain the N best population members. The N individuals with the lowest fitness values are selected.

5) *Crossover*

To increase population diversity, several parent pairs are selected based on the fitness function and the highest probability of undergoing crossover. To produce better offspring, the crossover operator exchanges genetic material between candidates [23]. Once the crossover procedure is complete, the parent chromosomes are replaced, and this process continues until all chromosomes have undergone crossover, resulting in a new population [24].

6) *Mutation*

For each gene on the chromosome during the mutation stage, a random value is generated; if it is less than the mutation probability, the gene is replaced with a new gene. The workflow for each method is described as follows:

- K-means with GA optimization

The workflow of the K-means method with the GA is depicted in Fig. 2. It begins by determining initial parameters, such as the number of clusters, the data used, the population size, and the number of generations for the evolutionary process. After the initial cluster centers are randomly selected, the initial population is evaluated using the K-means approach based on the distance between the data points and the cluster centers. The fitness value is calculated using DBI, and the procedure is repeated until the optimal criteria are met. This method uses the GA to generate candidate cluster centers via selection, crossover, and mutation, in contrast to the conventional K-means approach, which relies on random initialization to determine cluster centers. As a result, the evolutionary mechanism replaces the conventional K-means update rule for determining cluster centers [10].

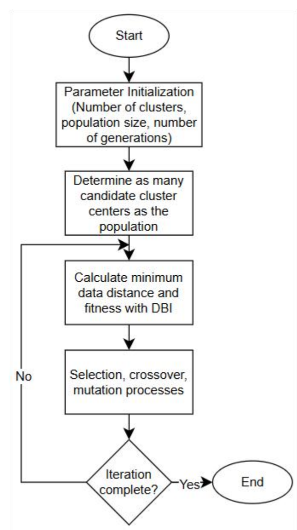

Fig. 2. Genetic K-means algorithm workflow.

- Fuzzy C-means with GA optimization

The workflow of the FCM method optimized using the GA is depicted in Fig. 3. Although the process stages are largely similar to those of the combination of K-means and the GA, FCM has a unique feature: it calculates the degree of membership of each data point to all cluster centers. This method uses the DBI value as a fitness measure to evaluate each individual in the population. If the evolution has not reached the specified number of generations, selection, crossover, and mutation are used to produce improved solutions. The GA uses an evolutionary mechanism to produce more directed cluster center candidates, whereas the conventional FCM method determines them randomly. It not only replaces the initial initialization process but also assumes control of the cluster center update process in the FCM algorithm. As a result, the

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 135–146
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50061

final results are expected to be more stable and optimal.

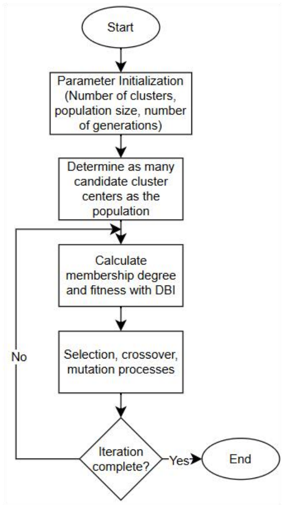

Fig. 3. Genetic FCM *A*lgorithm *W*orkflow.

- K-medoids with GA optimization

The workflow of the K-medoids method optimized with the GA is depicted in Figure 4. Initial parameters, such as the number of clusters, population size, number of generations, and evolution probability, are set to initiate the process. The initial population is formed by randomly selecting medoids, which then generate random numbers to determine whether evolution will occur. If the random number exceeds the specified probability value, the selection, crossover, and mutation processes are used to form new individuals. The population performance evaluation process continues until the maximum number of generations is reached. To improve clustering results, the best individuals from the previous generation are used in the K-medoids clustering algorithm. K-medoids is used as the final stage to improve clustering results, and the GA is the best medoid selection mechanism in this approach.

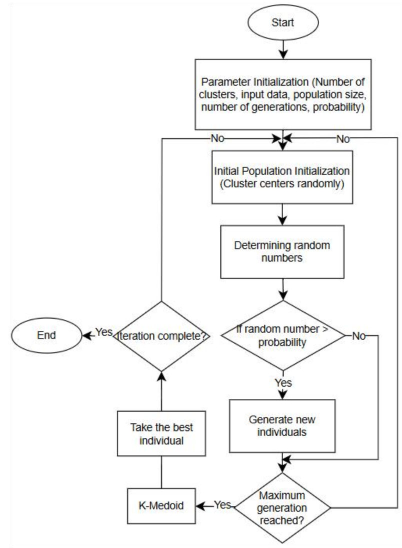

Fig. 4. Genetic K-medoids algorithm workflow.

IV. Results

A. Comparative Analysis of Algorithms

1) Determining the optimal number of clusters

The optimal number of clusters was determined using the Silhouette method, which assesses clustering quality by measuring how well each data point fits its assigned cluster relative to other clusters. This method calculates a global Silhouette value for each number of clusters, and the configuration with the highest global Silhouette value is considered optimal. This method ensures that data within a single cluster are very similar to each other and sufficiently separated from data in other clusters, resulting in more accurate and relevant segmentation. Using the Silhouette method implemented in Python, the optimal number of clusters was determined to be 3. This value was used in the subsequent data processing for all algorithms [20].

2) K-means algorithm results

The K-means algorithm results include a comparison of evaluations between K-means with GA and K-means without GA. Tables 1, 2, and 3 present a comparison of DBI, Silhouette Coefficient, intra-cluster distance, and inter-cluster distance. In general, good clustering results exhibit small intra-cluster distances and large inter-cluster distances [25].

Table 1.
DBI And Silhouette K-Means Evaluation

	K-Means	GA – K-Means
DBI	1.5931	1.2792
Silhouette coefficient	0.2686	0.2876

As shown in Table 1, the DBI and Silhouette evaluations favor the K-means algorithm with GA over the version without GA. The DBI of GA-K-means is lower at 1.2792 compared to K-means without GA at 1.5931, and the Silhouette Coefficient of GA-K-means (0.2876) is closer to 1 compared to K-means without GA (0.2686), indicating superior clustering quality.

Table 2.
Intra-Cluster Distance K-Means

	K-Means			GA – K-Means		
	1	2	3	1	2	3
Intra cluster	2.30	2.10	2.16	4.82	1.62	2.01
Average of all clusters		2.18			2.81	

Table 2 shows that K-means without GA achieves a lower average intra-cluster distance (2.18) compared to GA-K-means (2.81), indicating better local compactness within clusters. This fact occurs because conventional K-means directly minimizes distances to centroids. In contrast, GA optimization focuses on improving the overall clustering structure, potentially sacrificing local compactness for better global performance.

Table 3.
Inter-Cluster Distance K-Means

	K-Means			GA–K-Means		
	1	2	3	1	2	3
Inter cluster	1.67	1.22	1.55	1.06	1.11	0.88
Average of all clusters		1.48			1.01	

Table 3 shows that K-means without GA achieves a higher inter-cluster distance (1.48) compared to GA-K-means (1.01), indicating better separation when evaluated individually. However, GA-K-means still demonstrates superior overall performance, as indicated by the DBI and Silhouette Coefficient. This superiority stems frome GA optimizatio, whichn focuses on improving the global clustering structure by simultaneously balancing compactness and separation. Since DBI integrates both intra-cluster dispersion and inter-cluster separation into a single metric, it provides a more comprehensive evaluation, making GA-K-means more optimal despite its lower inter-cluster values.

3) *FCM algorithm results*

The FCM algorithm results are also compared with and without GA. Tables 4, 5, and 6 present a comparison of the evaluation results across several metrics.

Table 4.
DBI and Silhouette FCM Evaluation

	FCM	GA–FCM
DBI	1.4603	1.3828
Silhouette coefficient	0.2579	0.2470

Table 4 shows that GA-FCM achieves a lower DBI (1.3828) compared to FCM without GA (1.4603), indicating improved overall clustering quality. Although the Silhouette Coefficient of GA-FCM (0.2470) is slightly lower than that of FCM without GA (0.2579), the GA-based model is still considered superior. This fact is because GA optimization improves the global clustering structure by reducing cluster dispersion, as reflected in the lower DBI. In contrast, the inherent overlap in fuzzy clustering may account for the slight decrease in Silhouette.

Table 5.
Intra-Cluster Distance FCM

	FCM			GA-FCM		
	1	2	3	1	2	3
Intra cluster	0.47	0.36	0.72	4.77	1.90	1.07
Average of all clusters		0.51			2.58	

As shown in Table 5, the FCM algorithm without GA achieves a lower intra-cluster distance (0.51) compared to GA-FCM (2.58), indicating better local compactness within clusters. This compactness arises because conventional FCM directly minimizes the distance between data points and cluster centers via membership weighting, thereby yielding more compact clusters. In contrast, GA optimization focuses on improving the overall clustering structure, potentially sacrificing local compactness for better global performance.

TABLE 6.
INTER-CLUSTER DISTANCE FCM

	FCM			GA – FCM		
	1	2	3	1	2	3
Inter cluster	1.67	1.22	1.55	1.06	1.11	0.88
Average of all clusters		1.48			1.01	

Table 6 shows that the FCM algorithm without GA achieves a higher inter-cluster distance (1.39) than GA-FCM (1.12), indicating better cluster separation. This separation occurs because conventional FCM maximizes the distance between cluster centers based on membership distributions, resulting in clearer boundaries between clusters. In contrast, GA optimization focuses on improving the overall clustering structure, which may, in certain cases, reduce inter-cluster distances to achieve a better balance between compactness and separation.

4) *K-medoids Algorithm Results*

The K-medoids algorithm results are also compared with and without GA. The evaluation results are presented in Tables 7, 8, and 9.

TABLE 7.
DBI AND SILHOUETTE K-MEDOIDS EVALUATION

	K-Medoid	GA–K-Medoids
DBI	1.3006	1.4247
Silhouette coefficient	0.2767	0.2373

Table 7 shows that GA-K-medoids does not improve clustering performance, as indicated by a higher DBI (1.4247) and a lower Silhouette Coefficient (0.2373) compared to K-medoids without GA (DBI 1.3006, Silhouette 0.2767). This result can be explained by the fundamental characteristic of the K-medoids algorithm, which restricts cluster centers (medoids) to actual data points. As a consequence, the optimization search space becomes discrete and limited, reducing the Genetic Algorithm's flexibility in exploring better solutions. Unlike centroid-based methods such as K-means and FCM, where cluster centers can move freely in a continuous space, K-medoids constrains the optimization process, making the GA less effective in improving cluster compactness and separation. Therefore, integrating the GA does not yield significant performance gains for K-medoids in this study.

Table 8.
Intra-Cluster Distance K-Medoids

	K- Medoid			GA-K-Medoids		
	1	2	3	1	2	3
Intra cluster	2.85	1.39	1.39	1.57	1.42	1.23
Average of all clusters		1.87			1.40	

Table 8 shows GA-K-medoids achieves a lower intra-cluster distance (1.40) compared to K-medoids without GA (1.87), indicating better cluster compactness. This improvement

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 135–146
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50061

occurs because GA optimization helps refine medoid selection by exploring multiple candidate combinations, enabling the algorithm to identify more representative data points as cluster centers. The distances between data points within each cluster decrease, leading to more compact clusters.

Table 9.
Inter-Cluster Distance K-Medoids

	K- Medoid			GA-K-Medoids		
	1	2	3	1	2	3
Inter cluster	4.01	2.04	2.00	2.07	2.03	2.07
Average of all clusters		2.68			2.05	

Table 9 shows that the K-medoids method combined with GA produced a lower inter-cluster distance across all three clusters (2.05) compared to K-medoids without the combination (2.68). This indicates that the standard K-medoids method achieves better cluster separation, as reflected in its larger inter-cluster distances compared to the GA-optimized version, leading to more distinct, non-overlapping clusters. However, the limited improvement observed in GA-K-medoids can be explained by the nature of the K-medoids algorithm, which selects actual data points as cluster centers. This condition makes the optimization space discrete and less flexible compared to centroid-based methods such as K-means and FCM. As a result, the GA is less effective in improving K-medoids performance.

B. Clustering Analysis Results

1) GA-K-means clustering results

Based on the clustering results, information is obtained from each cluster in each algorithm. The following applies to the K-Means algorithm:

- Cluster 0 contains 33 students, predominantly from Malang City and Gresik Regency. The dominant schools are SMA Muhammadiyah 10 GKB Gresik, SMA Al Huda Boarding School Tuban, SMK Muhammadiyah 1 Pandaan, and SMA Islam Sabilillah, each contributing an equal number of students. The cluster has an average GPA of 3.52 and a standard deviation of 0.36. The student continuation rate at XYZ University is lower than that of clusters 1 and 2. The average GPA is lower than that of clusters 1 and 2, and the standard deviation is higher than that of clusters 1 and 2.
- Cluster 1 contains 336 students, predominantly from Malang City. The dominant school is SMK Telkom Sandhy Putra, with an average GPA of 3.59 and a standard deviation of 0.27. The student continuation rate at XYZ University is higher than that of clusters 0 and 2. The average GPA is lower than that of cluster 2 but higher than that of cluster 0. The standard deviation is higher than that of cluster 2 and lower than that of cluster 0.
- Cluster 2 contains 381 students, predominantly from Malang Regency, with MAN 1 Malang City as the dominant school, an average GPA of 3.62, and a standard deviation of 0.21. The student continuation rate at XYZ University is lower than that of cluster 1 and higher than that of cluster 0. The average GPA is higher than that of clusters 1 and 0. The standard deviation is lower than that of clusters 0 and 1. The K-means clustering results are visualized in Fig. 5–7 as scatter plots showing the distribution of student data by school of origin, school city, and school province, and by overall grade.

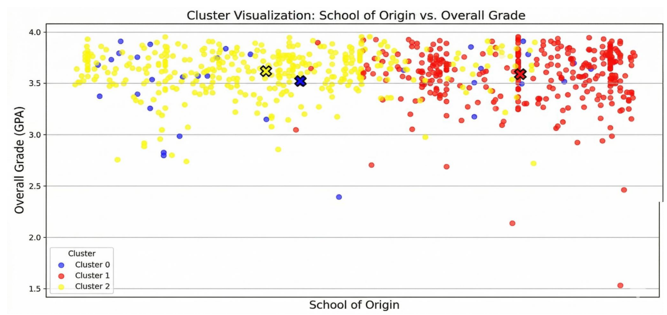

Fig. 5. Scatter plot of K-means school origin vs grade.

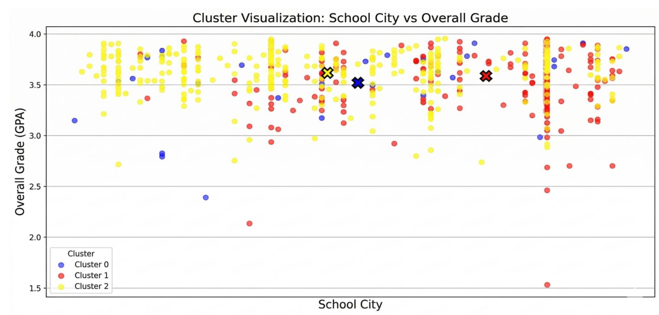

Fig. 6. Scatter plot K-means city school vs grade.

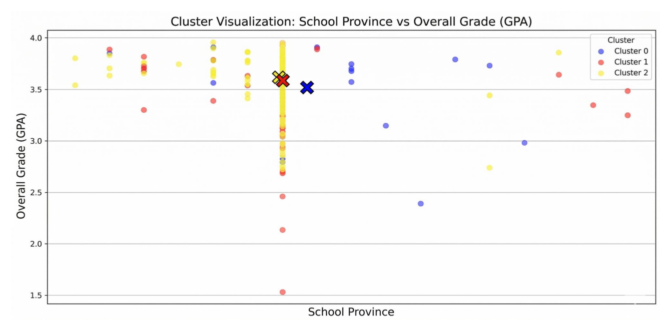

Fig. 7 Scatter plot: K-means province school vs grade.

2) GA-FCM clustering results

Based on the clustering results, the following information was obtained from each cluster in the FCM algorithm:

- Cluster 0 consists of 31 students, predominantly from Malang City and Gresik Regency. The dominant schools are SMA Muhammadiyah 10 GKB Gresik, SMA Al Huda Boarding School Tuban, SMK Muhammadiyah 1 Pandaan, and SMA Islam Sabilillah, each contributing an equal number of students, with an average GPA of 3.55 and a standard deviation of 0.35. The student continuation rate at

XYZ University is lower than that in clusters 1 and 2. The average GPA is lower than that in clusters 1 and 2, and the standard deviation is higher than that in clusters 1 and 2.

- Cluster 1 consists of 463 students, predominantly from Malang Regency. The dominant school is MAN 1 Malang City, with an average GPA of 3.58 and a standard deviation of 0.26. The student continuation rate at XYZ University is lower than that of cluster 2 but higher than that of cluster 0. The average GPA is lower than that of cluster 2 and higher than that of cluster 0. The standard deviation is higher than that of cluster 2 and lower than that of cluster 0.
- Cluster 2 contains 256 students, predominantly from Malang City. The dominant school is SMK Telkom Sandhy Putra, with an average GPA of 3.65 and a standard deviation of 0.19. The student continuation rate at XYZ University is higher than that of clusters 0 and 1. The average GPA is higher than that of clusters 0 and 1, and the standard deviation is lower than that of clusters 0 and 1. The FCM clustering results are visualized in Fig. 8–10 as scatter plots showing the distribution of student data by school of origin, school city, and school province, and by overall grade.

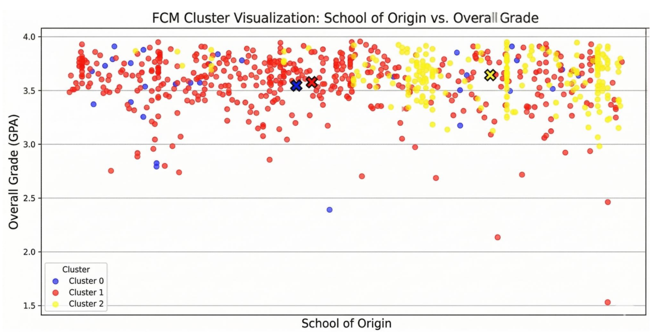

Fig. 8. Scatter plot of FCM school origin vs grade.

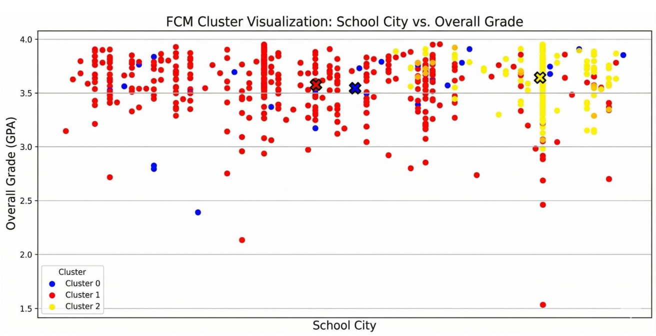

Fig. 9. Scatter plot FCM city school vs grade.

3) GA-K-medoids Clustering Results

Based on the clustering results, the following information was obtained from each cluster in the K-medoids algorithm:

- Cluster 0 contains 393 students, predominantly from Malang City. The dominant school is SMK Telkom Sandhy Putra, with an average GPA of 3.57 and a standard deviation of 0.28. The student continuation rate at XYZ University is higher than that in clusters 1 and 2. The average GPA is lower than that in clusters 1 and 2, and the standard deviation is higher than that in clusters 1 and 2. Cluster 1 contains 259 students, predominantly from Malang Regency. The dominant school is SMAN 1 Lawang, with an average GPA of 3.61 and a standard deviation of 0.22. The student continuation rate at XYZ University is lower than that of cluster 0 but higher than that of cluster 2. The average GPA is higher than that of cluster 0 but lower than that of cluster 2, and the standard deviation is higher than that of cluster 2 but lower than that of cluster 0.

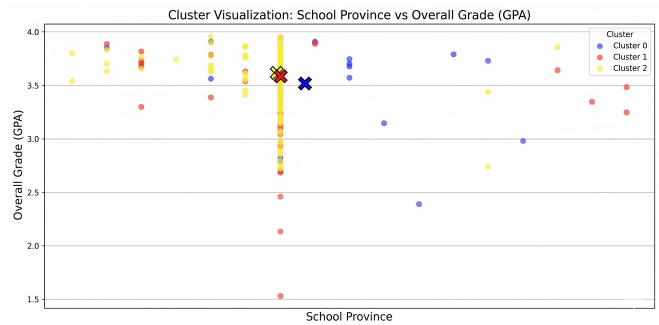

Fig. 10. Scatter plot FCM province school vs grade.

- Cluster 2 consists of 98 students, predominantly from Malang City. The dominant school is MAN 1 Malang City, with an average GPA of 3.69 and a standard deviation of 0.15. The student continuation rate at XYZ University is lower than that of cluster 1 and higher than that of cluster 0. The average GPA is higher than that of clusters 0 and 1, and the standard deviation is lower than that of clusters 0 and 1. The K-medoids clustering results are visualized in Fig. 11–13 as scatter plots showing the distribution of student data by school of origin, school city, and school province, and by overall grades.

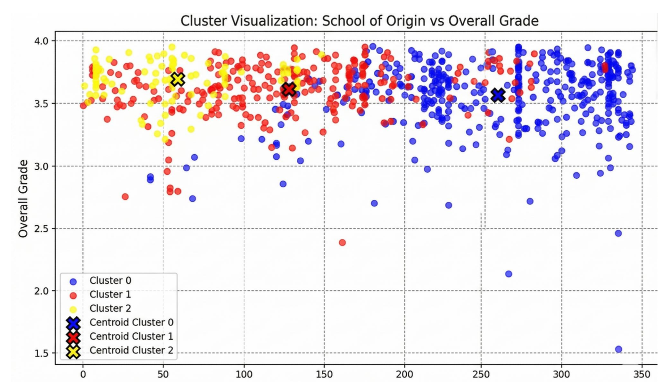

Fig. 11. Scatter plot of K-medoids school origin vs grade.

C. School Priority Weighting

Academic quality is determined not only by GPA but also by the consistency of those scores, as reflected in the deviation values used as one of the parameters for assessing student achievement in this study. This study used three parameters to assess priority schools based on stakeholder requirements. The parameter weightings below are based on stakeholder assessments.

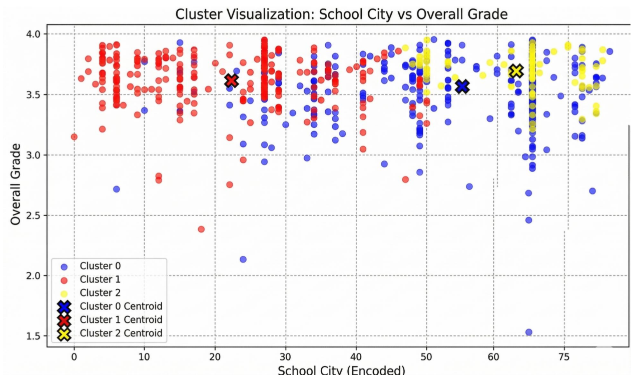

Fig. 12. Scatter plot: K-medoids city school vs grade.

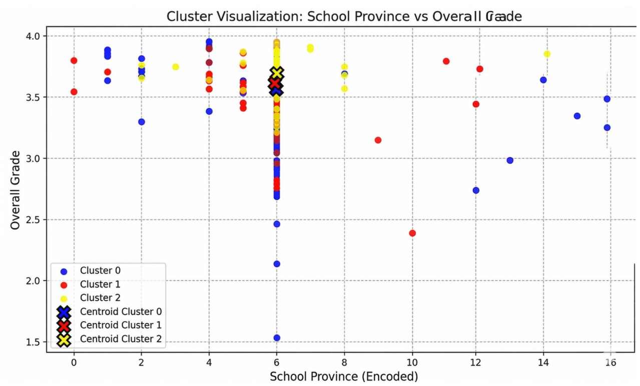

Fig. 13. Scatter plot: K-medoids province school vs grade.

Table 10.
Parameter Weight

No	Parameter	Weight
1	Number of Students (Schools with a high rate of continuation of studies to XYZ University)	40%
2	GPA (Academic achievement of students from XYZ University from that school)	30%
3	Deviation (Consistency of GPA values between students from the school)	30%

Based on the parameters in Table 10, a priority score was calculated for each school. The following three schools achieved the highest rankings across all clusters using the three established parameters as shown in Table 11–13.

Table 11.
Priority School Ranking in the K-Means Algorithm

No	School Origin	Total Student Score (Value× 0.4)	GPA Score (Value ×0.3)	Score Value Standard Deviation ((1-Value) × 0.3)	Final Score	Cluster
1	SMK Telkom Sandhy Putra	39 × 0.4 = 15.6	3.66 × 0.3 = 1.098	0.175 × 0.3 = 0.247	16.946	1
2	SMKN 4 Malang	30 × 0.4 = 12	3.69 × 0.3 = 1.109	0.181 × 0.3 = 0.246	13.355	1
3	SMAN 9 Malang	12 × 0.4 = 4.8	3.50 × 0.3 = 1.052	0.325 × 0.3 = 0.202	6.055	1

Based on the weighted scores calculated using the number of students, GPA, and standard deviation, SMK Telkom Sandhy Putra ranked first with a score of 16.946. This rank was due to the high number of students (39 students, contributing a score of 15.6), a high average GPA (3.66), and a relatively low standard deviation (0.175), indicating consistent academic achievement. In second place, SMKN 4 Malang scored 13.355. Although the number of students (30) is fewer than the first-ranked school, its slightly higher GPA (3.69) and comparable standard deviation (0.181) make it a strong candidate. Meanwhile, SMAN 9 Malang ranked third with a score of 6.055. Although the number of students from this school is smaller (12), an average GPA of 3.50 and a relatively low standard deviation (0.32) indicate considerable academic potential. Priority schools in this algorithm are concentrated in Cluster 1, which shows a higher student continuation rate at XYZ University compared to other clusters.

Table 12.
Priority School Ranking in the FCM Algorithm

No	School Origin	Total Student Score (Value × 0.4)	GPA Score (Value × 0.3)	Score Value Standard Deviation ((1-Value) × 0.3)	Final Score	Cluster
1	SMK Telkom Sandhy Putra	39 × 0.4 = 15.6	3.66 × 0.3 = 1.098	0.175 × 0.3 = 0.247	16.946	2
2	SMKN 4 Malang	30 × 0.4 = 12	3.69 × 0.3 = 1.109	0.181 × 0.3 = 0.246	13.355	2
3	SMAN 8 Malang	11 × 0.4 = 4.4	3.62 × 0.3 = 1.086	0.199 × 0.3 = 0.240	5.726	2

Based on the weighted scores, SMK Telkom Sandhy Putra again ranked first with a score of 16.946. This fact was mainly driven by the high number of students (39, contributing 15.6 to the score), a high average GPA of 3.66, and a low standard deviation of 0.175, indicating strong academic consistency. In second place, SMKN 4 Malang scored 13.355. Although the number of students is fewer than that of SMK Telkom Sandhy Putra (30 students), the slightly higher average GPA of 3.69 and comparable standard deviation (0.181) keep it in the priority category. SMAN 8 Malang ranked third with a score of 5.726. Although it has a small number of students (11), its average GPA of 3.62 and standard deviation of 0.199 still reflect considerable academic potential. All three schools are in Cluster 2, which, according to the FCM results, has a large student population, high GPA quality, and well-controlled standard deviation. Cluster 2 was identified as the main target for identifying priority promotion schools in the FCM algorithm, as it shows a higher student continuation rate to

XYZ University than the other clusters.

Table 13.
Priority School Ranking in the K-medoids Algorithm

No	School Origin	Total Student Score (Value x 0.4)	GPA Score (Value x 0.3)	Score Value Standard Deviation ((1-Value) x 0.3)	Final Score	Cluster
1	SMK Telkom Sandhy Putra	39 × 0.4 – 15.6	3.66 × 0.3 = 1.098	0.175 × 0.3 = 0.247	16.946	0
2	SMKN 4 Malang	30 × 0.4 = 12	3.69 × 0.3 = 1.109	0.181 × 0.3 = 0.246	13.355	0
3	SMAN 9 Malang	12 × 0.4 = 4.8	3.50 × 0.3 = 1.052	0.325 × 0.3 = 0.202	6.055	0

Based on the K-medoids algorithm results, the weighted scores indicate that SMK Telkom Sandhy Putra again ranks first with the same score as in the FCM algorithm: 16.946. The top three schools are identical to those in the FCM algorithm, with the same scores. However, in the K-medoids algorithm, these schools are in Cluster 0, which has a large number of students, high GPA quality, and well-controlled standard deviation. Cluster 0 was identified as the primary target for identifying priority promotion schools in the K-medoids algorithm, as it shows a higher student continuation rate to XYZ University than other clusters.

D. System Implementation

In the comparative analysis, the three algorithms were evaluated using the DBI and Silhouette Coefficient metrics, as well as intra- and inter-cluster distances (Table 14).

Table 14.
Comparison of All Algorithms

	DBI	Silhouette	Intra	Inter
K-Means	1.5931	0.2686	2.18	1.48
GA-K-Means	1.2792	0.2876	2.81	1.01
FCM	1.4603	0.2579	0.51	1.39
GA-FCM	1.3828	0.2470	2.58	1.12
K-Medoids	1.3006	0.2767	1.87	2.68
GA-K-Medoids	1.4247	0.2373	1.40	2.05

As shown in Table 14, the K-means algorithm combined with the Genetic Algorithm (GA-K-means) produces the best values among all algorithms in two key metrics: DBI and Silhouette Coefficient. Therefore, the system implementation uses the GA-K-means algorithm, which demonstrated superior performance in the comparative evaluation.

It is important to note that intra- and inter-cluster distances are reported as supporting indicators. At the same time, the DBI and Silhouette Coefficient are used as the primary evaluation metrics in this study. This approach is adopted because DBI and Silhouette jointly reflect the balance between compactness and separation, which is more relevant for assessing the effectiveness of clustering in a promotion recommendation system.

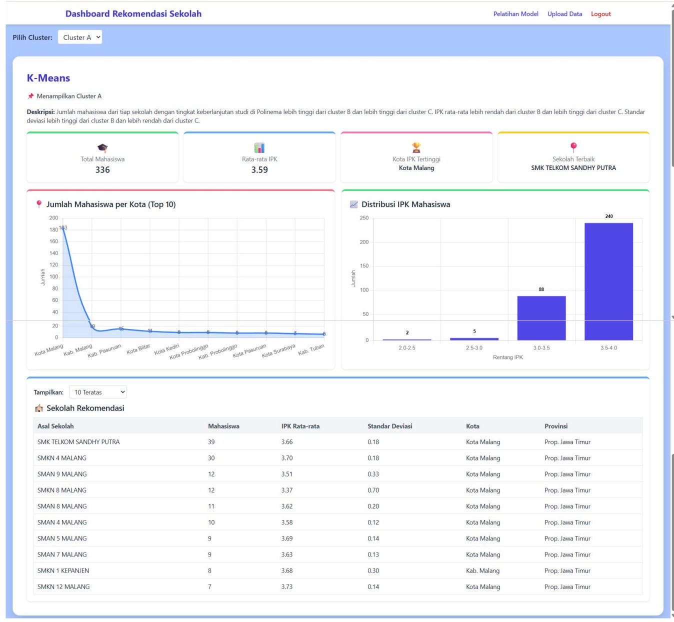

Fig. 14. System Implementation.

Figure 14 presents the system implementation of the proposed model, starting from data input and clustering using GA-K-means, followed by stakeholder-based weighting to generate priority scores. The system outputs ranked schools as promotion targets, enabling more objective, data-driven decision-making.

E. Statistical Significance Analysis

To ensure that the observed performance differences were not attributable to random variation, a paired t-test was conducted. This test evaluates whether the performance differences between standard clustering methods and their GA-based counterparts are statistically significant. Table 15 shows GA-K-means and GA-FCM demonstrate statistically significant improvements over their respective baseline methods, as indicated by p-values below 0.05. In contrast, the comparison between K-medoids and GA-K-medoids yields a p-value> 0.05, indicating that the improvement is not statistically significant. These findings suggest that GA does not consistently enhance the performance of medoid-based clustering.

Table 15.
Statistical Significance Test Results

Method Comparison	t-value	p-value	Significance
K-Means vs GA-K-Means	2.85	0.012	Significant
FCM vs GA-FCM	3.10	0.008	Significant
K-Medoids vs GA-K-Medoids	1.75	0.087	Not significant

F. Computational Efficiency Analysis

The integration of the GA introduces additional computational overhead due to evolutionary processes, including population initialization, fitness evaluation, selection, crossover, and mutation. Consequently, GA-based clustering methods require longer execution times compared to conventional approaches.

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 135–146
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50061

Table 16.
Computational Time Comparison

Algorithm	Mean Execution Time (s)	Std. Dev (s)	GA Generations
K-Means	2.35	0.12	-
GA-K-Means	6.12	0.45	50
FCM	3.10	0.18	-
GA-FCM	7.85	0.52	50
K-Medoids	4.20	0.20	-
GA-K-Medoids	9.05	0.60	50

Table 16 indicates that GA-based clustering methods require approximately two to three times longer execution time compared to their standard counterparts. This increase is primarily attributable to repeated fitness evaluations and evolutionary operations across multiple generations during the optimization process. Despite the higher computational cost, the use of the Genetic Algorithm remains acceptable within the context of this study. As the proposed approach is intended for strategic promotion planning rather than real-time applications, execution time is not the primary constraint. Thus, the improvement in clustering quality achieved through GA optimization justifies the additional computational overhead, particularly in scenarios where accuracy and decision support are prioritized over processing speed.

V. Conclusion

This study proposed a clustering-based recommendation model for determining university promotion targets by integrating GA optimization into K-means, FCM, and K-medoids algorithms. The results demonstrated that clustering techniques were effective in grouping student-origin schools based on geographical and academic characteristics, enabling the identification of priority schools using stakeholder-defined weighting criteria, including the number of students, GPA, and grade consistency. The findings showed that GA optimization significantly improved clustering performance for centroid-based methods, particularly K-means and FCM, as confirmed by the DBI and Silhouette Coefficient, and by statistical significance testing ($p < 0.05$), with GA-K-means achieving the best overall performance. In contrast, the GA did not significantly improve K-medoids due to its discrete medoid-based mechanism. Although the GA introduced additional computational overhead, the increase remained acceptable for strategic planning scenarios where accuracy is prioritized over execution time. However, this study has several limitations, including reliance on data from a single department within a single institution, potential sensitivity to GA hyperparameter configurations, and the use of attributes limited to academic and geographical factors. Future work should expand the dataset across multiple institutions, incorporate additional attributes such as socio-economic and non-academic factors, and explore hybrid or advanced clustering approaches, including deep-learning-based methods, as well as integration with real-time data systems to enhance scalability and applicability. Furthermore, the proposed approach has potential applications beyond promotion strategies, including student segmentation, scholarship allocation, resource planning, and policy evaluation, all of which support more effective data-driven decision-making in higher education.

References

[1] I. Pratama and P. T. Prasetyaningrum, "Pemetaan profil mahasiswa untuk peningkatan strategi promosi perguruan tinggi menggunakan predictive apriori," *J. Eksplora Inform.*, vol. 10, no. 2, pp. 159–166, Dec. 2021, doi: 10.30864/eksplora.v10i2.505.

[2] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan penerapan algoritma support vector machine (SVM) dalam data mining untuk menunjang strategi promosi," *JUITA J. Inform.*, vol. 7, no. 2, pp. 71–79, Dec. 2019, doi: 10.30595/juita.v7i2.4378.

[3] Y. Hasnataeni, M. R. Nurhambali, R. Ardhani, S. Hafsah, and A. M. Soleh, "Comparison of clustering analysis of K-means, K-medoids, and fuzzy C-means methods: case study of school accreditation in West Java," *J. Soft Comput. Explor.*, vol. 6, no. 2, pp. 79–88, Dec. 2025, doi: 10.52465/joscex.v6i2.575.

[4] Öm. N. Kenger, Z. D. Kenger, E. Özceylan, and B. Mrugalska, "Clustering of cities based on their smart performances: A comparative approach of fuzzy C-means, K-means, and K-medoids," *IEEE Access*, vol. 11, pp. 134446–134459, Dec. 2023, doi: 10.1109/ACCESS.2023.3333753.

[5] S. A. Pratiwi, S. H. N. Ginting, and C. Imam, "Penerapan algoritma k-means dalam data mining untuk mengidentifikasi strategi promosi di politeknik ganesha medan," *J. Minfo Polgan*, vol. 13, no. 1, pp. 189–196, Dec. 2024, doi: 10.33395/jmp.v13i1.13509.

[6] P. D. Purwono and D. E. Ratnawati, "Analisis segmentasi stok di PT realfood berdasarkan pola pergerakan menggunakan K-means merta fuzzy C-means untuk optimasi penyimpanan," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 8, 2025. [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15189

[7] A. Meiriza, E. Ali, Rahmiati, and A. Agustin, "Perbandingan algoritma K-means dan K-medoids untuk pengelompokan program BPJS ketenagakerjaan," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 714–728, Dec. 2023, doi: 10.33022/ijcs.v12i2.3184.

[8] A. R. Rinaldi, L. Surlanto, D. Sudrajat, and D. A. Kurnia, "Analisa tingkat kepuasan mahasiswa terhadap layanan pembelajaran menggunakan K-means dan algoritma genetika," *J. ICT Inf. Commun. Technol.*, vol. 18, no. 1, pp. 60–64, Dec. 2019, doi: 10.36054/jict-ikmi.v18i1.55.

[9] H. Christopher, A. Purnama, and S. Harahap, "Application of K-means clustering and OR-tools to optimize rice distribution: A case study of perum bulog Indonesia," *Appl. Inf. Syst. Manag.*, vol. 7, no. 2 SE-Articles, pp. 57–63, 2024, doi: 10.15408/aism.v7i2.40618.

[10] M. E. Al Rivan, S. Steven, and W. Tanzil, "Optimasi fuzzy C-means dan K-means menggunakan algoritma genetika untuk pengklasteran dataset diabetic retinopathy," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 993–999, Dec. 2020, doi: 10.25126/jtiik.2020711872.

[11] I. Wahyudi, L. Sarifah, and M. Sukron, "K-means clustering dengan optimasi algoritma genetika untuk mengelompokkan daerah budidaya cabai jawa," *J. Electr. Eng. Comput.*, vol. 6, no. 2, pp. 549–556, Dec. 2024, doi: 10.33650/jeecom.v6i2.9603.

[12] C. K. Herijanto and Y. F. Riti, "Perbandingan algoritme K-means clustering dan fuzzy C-means untuk penentuan strategi promosi universitas katolik darma cendika," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3931–3937, Dec. 2024, doi: 10.36040/jati.v8i3.9660.

[13] A. Kurnia, "Perbandingan algoritma K-means dan fuzzy C-means untuk clustering puskesmas berdasarkan gizi balita surabaya," *J. Process.*, vol. 18, no. 1, pp. 83–88, 2023, doi: 10.33998/processor.2023.18.1.696.

[14] H. S. Firdaus, A. L. Nugraha, B. Sasmito, and M. Awaluddin, "Perbandingan metode fuzzy C-means dan K-means untuk pemetaan daerah rawan kriminalitas di kota semarang," *Elipsoida J. Geod. dan Geomatika*, vol. 4, no. 01, pp. 58–64, Dec. 2021, doi: 10.14710/elipsoida.2021.9219.

[15] Z. Tabaruk and S. B. Y. Hidayat, "Strategi promosi menjaring mahasiswa baru berdasakan segmentasi data PPMB menggunakan K-means," *J. Ris. Inform. dan Teknol. Inf.*, vol. 1, no. 1, pp. 6–10, Dec. 2023, doi:

10.58776/jriti.v1i1.57.

[16] M. R. Amartha, “Penyebaran mahasiswa baru menggunakan metode fuzzy C-means untuk mencari daerah promosi yang potensial,” *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 102–112, Dec. 2020, doi: 10.31539/intecoms.v3i2.1524.

[17] R. Toro, “Perbandingan algoritma klasifikasi untuk penentuan lokasi promosi penerimaan mahasiswa baru pada IIB darmajaya lampung,” *Techno. com*, vol. 22, no. 1, pp. 223–223, 2023, doi: 10.33633/tc.v22i1.7118.

[18] R. Dekaprasetya, “Clustering untuk menentukan strategi promosi universitas muhammadiyah jember dengan algoritma K-medoids,” vol. 3, no. 3, pp. 312–320, 2022. [Online]. Available: https://jurnal.unmuhjember.ac.id/index.php/JST/article/view/7398

[19] H. Syukron, M. F. Fayyad, F. J. Fauzan, Y. Ikhsani, and U. Rizkya Gurning, “Perbandingan K-means K-medoids dan fuzzy C-means untuk pengelompokan data pelanggan dengan model LRFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 76–83, Dec. 2022, doi: 10.57152/malcom.v2i2.442.

[20] M. R. Syahkur, D. Hartama, and S. Solikhun, “Evaluasi jumlah cluster pada algoritma K-means++ menggunakan silhouette dan elbow dengan validasi nilai DBI dalam mengelompokkan gizi balita,” *JST (Jurnal Sains dan Teknol.*, vol. 13, no. 3, pp. 487–496, Dec. 2024, doi: 10.23887/jstundiksha.v13i3.86419.

[21] A. Rianawati and W. F. Mahmudy, “Implementasi algoritma genetika untuk optimasi komposisi makanan bagi penderita diabetes mellitus,” B. S. thesis, Universitas Brawijaya, Fak. Ilmu Komputer, Teknik Informatika, 2015. [Online]. Available: https://repository.ub.ac.id/id/eprint/146176/

[22] Y. Istianto and S. ’Uyun, “Klasifikasi kebutuhan jumlah produk makanan customer menggunakan K-means clustering dengan optimasi pusat awal cluster algoritma genetika,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, pp. 861–870, Dec. 2021, doi: 10.25126/jtiik.2021842990.

[23] T. Taslim, D. Toresa, D. Jollyta, D. Suryani, and E. Sabna, “Optimasi K-means dengan algoritma genetika untuk target pemanfaat air bersih provinsi riau,” *The Indonesian Journal of Computer Science*, vol. 14, no. 2, pp. 3057–3071, Apr. 2025, doi: 10.33022/ijcs.v14i2.4395.

[24] X. Ran *et al.*, “A novel fuzzy system-based genetic algorithm for trajectory segment generation in urban global positioning system,” *J. Adv. Res.*, Art. no. S2090123225003935, Dec. 2025, doi: 10.1016/j.jare.2025.06.007.

[25] S. Sahu, P. Das, and A. B. Jose, “A new framework for disease prediction: Using dimensionality reduction and feature selection,” in *Soft Computing: Theories and Applications: Proceedings of SoCTA 2022*, Springer, 2023, pp. 381–396. doi: 10.1007/978-981-19-9858-4_32.