Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Implementation of the K-Medoids Algorithm in a Clustering System for Determining Promotional Strategies at SMK Bina Satria

Rizky Wira Nanda Pasaribu[1]*, Marah Doly Nasution[2]

[1]Universitas Muhammadiyah Sumatera Utara
[2]Universitas Muhammadiyah Sumatera Utara
rizkywirananda14@gmail.com[1*]

Abstract

The competition between public and private vocational high schools (SMK) requires private schools, such as SMK Bina Satria, to adopt more effective and data-driven promotional strategies. This study aims to implement the K-Medoids algorithm in the clustering process of student origin school data to group the contribution potential of partner schools toward new student enrollment. The variables used in the analysis include the total number of student contributions and the participation status in the Member Get Member (MGM) program.
This research employs a quantitative approach using data mining methods. The stages include data collection, preprocessing, determining the optimal number of clusters using the Elbow Method and Silhouette Score, implementation of the K-Medoids algorithm, and development of a web-based system for visualizing the results. The analysis results show that origin schools can be grouped into three clusters: High Potential Loyal Partners, Potential Future Partners, and Incidental Contributors. The developed system assists the school in determining more targeted and efficient promotional strategies. With this implementation, SMK Bina Satria can formulate promotion policies based on objective partner segmentation and enhance the effectiveness of its promotional programs to attract new students.

Keywords*: K-Medoids; Clustering; Promotional Strategy; Member Get Member; Data Mining.*

1. Introduction

In the era of globalization and rapid technological advancement, vocational education plays a crucial role in preparing a workforce that is ready to compete in the industrial world. Vocational High Schools (SMK) hold an essential position in producing graduates equipped with skills aligned with industry needs. However, with the increasing number of SMKs in Indonesia, competition among schools has become more intense, particularly between public and private SMKs. According to data from the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) as of December 31, 2023, the number of SMKs in Indonesia reached 14,445 units, consisting of 3,752 public SMKs and 10,693 private SMKs. Despite the higher number of private SMKs, public SMKs still dominate in terms of student enrollment. This relatively small difference highlights the need for private SMKs to offer competitive advantages in order to attract prospective students [1].

One of the main differences between public and private SMKs lies in tuition fees and available facilities. According to [2], public SMKs are generally more attractive to the community because they provide lower tuition fees due to government subsidies, as well as more comprehensive facilities compared to private SMKs, which rely heavily on funding from schools or foundations. This condition leads many prospective students to prefer public SMKs as their primary choice, while private SMKs must work harder to promote their strengths to remain attractive. In such a situation, effective promotional strategies become a key factor in increasing the appeal and enrollment numbers of private SMKs.

Traditional promotional efforts, such as distributing brochures, installing banners, and visiting junior high schools, are often insufficient in reaching potential students. Non–data-driven promotional methods frequently result in less optimal decisions when determining marketing strategies. Therefore, a data-driven approach is required to identify student segmentation more accurately. Through comprehensive data analysis, schools can better understand student preferences and develop targeted promotional strategies that align with their needs [3]. One such technique is clustering, a method of grouping data based on fundamental similarities and differences within a dataset. The main goal of clustering is to divide datasets into groups that share similar characteristics while remaining distinct from other groups [4].

This study applies one of the data mining methods, the K-Medoids algorithm. This algorithm is advantageous in handling outliers and producing more stable clusters compared to the K-Means algorithm. K-Medoids selects actual data points from the dataset as cluster centers (medoids), making the clustering results more representative and easier to interpret [3]. The method has been widely used in various domains, such as customer segmentation, consumer behavior analysis, and data-driven marketing strategies. For instance, at STIKES Perintis Padang, this algorithm was used to group students based on their school of origin and region, enabling the institution to focus promotional efforts on areas with higher success rates [5]. Hence, implementing the K-Medoids algorithm for school promotional strategy can provide valuable insights in formulating more effective marketing approaches.

In the context of SMK Bina Satria, the application of the K-Medoids algorithm can help classify feeder schools based on their contribution potential in student admissions, considering factors such as the number of students enrolled from each school, the implementation of the Member Get Member (MGM) program, and annual contribution intensity. By identifying schools with high, medium, and low potential, SMK Bina Satria can design more effective and targeted promotional strategies, focusing efforts on schools with the greatest potential to contribute to future student intake.

2. Literature Review

2.1. Data Mining

Data mining is the process of analyzing large datasets to discover hidden patterns, correlations, and useful information that can support decision-making. It integrates methods from machine learning, statistics, pattern recognition, and database systems to extract knowledge from complex data sources [6]. Often referred to as Knowledge Discovery in Databases (KDD), data mining transforms raw data into valuable insights by identifying trends and relationships within historical datasets [7].

2.2. Clustering

Clustering methods are commonly divided into partitioning methods (e.g., K-Means, K-Medoids), hierarchical methods, and density-based methods. Among these, K-Means and K-Medoids are the most widely used due to their efficiency and interpretability. K-Means uses centroids that are not necessarily actual data points, while K-Medoids selects representative data objects (medoids) as cluster centers, making it more robust against outliers [3].

2.3. K-Medoids

The Partitioning Around Medoids (PAM) algorithm, commonly known as K-Medoids, was introduced by Leonard Kaufman and Peter J. Rousseeuw in 1987. It is a partitioning clustering method that groups a set of n objects into k clusters. Unlike K-Means, where clusters are represented by centroids, K-Medoids uses actual data objects, called medoids, as cluster representatives, making it more robust to outliers and noise.[8]

3. Research Method

The research process followed the Knowledge Discovery in Databases (KDD) framework, which outlines the steps required to extract meaningful knowledge from raw data.

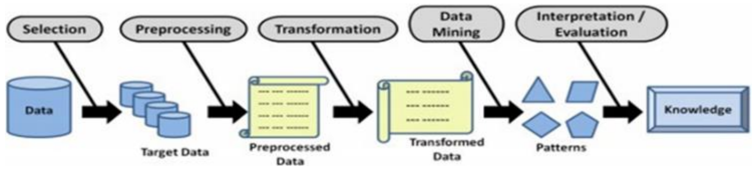


Fig. 1: Research Method

Description of Research Methods Using KDD:

1. Selection: Identifying and selecting relevant data sources, in this case enrollment and feeder school contribution data from SMK Bina Satria.
2. Preprocessing: Cleaning and integrating raw data to remove inconsistencies, missing values, and errors, ensuring data quality.
3. Transformation: Converting data into an appropriate format for analysis, including normalization of numeric attributes and encoding categorical variables such as MGM status.
4. Data Mining: Applying the K-Medoids clustering algorithm to segment feeder schools into groups based on contribution potential.
5. Interpretation/Evaluation: Interpreting cluster results and evaluating their effectiveness using validation techniques (Elbow Method and Silhouette Score).
6. Knowledge: Generating insights that support SMK Bina Satria in formulating targeted and data-driven promotional strategies.

3.1. Data Selection

Relevant data were collected from SMK Bina Satria covering the period of 2020–2025. The dataset consisted of feeder school contributions, including the number of students enrolled from each school, the number of years of contribution, and the status of participation in the Member Get Member (MGM) program.

3.2. Data Preprocessing

Data preprocessing was conducted to ensure accuracy and consistency, including:
- Data cleaning: handling missing values and removing duplicates.
- Data integration: combining enrollment data with MGM records.
- Data reduction: focusing only on attributes relevant to clustering (student contributions, years of contribution, MGM status).

3.3. Data Transformation

The dataset was normalized using the Standard Scaler method to ensure that all attributes contributed proportionally to distance calculations. For example, jumlah_siswa ranged from 1–402, while tahun_berkontribusi ranged only from 1–5. Normalization prevented the larger-scale attribute from dominating the clustering process.

3.4. Data Analysis

The implementation of the K-Medoids algorithm in this study involved several mathematical formulations, as described below:

3.4.1. Data Normalization (Standard Scaler)

To ensure that each variable contributed equally to the clustering process, the dataset was normalized using the Standard Scaler method:

$$x_{new} = \frac{x_{old} - \mu}{\delta} \quad (1)$$

Where:

x_{old}	= Original data value
μ	= Mean of the variable
δ	= Standard deviation of the variable

3.4.2. Euclidean Distance

The dissimilarity between an observation and a cluster center was calculated using the Euclidean Distance:

$$d_{euc}(x_{ij}, c_{kj}) = \sqrt{\sum_{j=1}^{p}\sum_{i=1}^{n}\left(x_{ij} - c_{kj}\right)^2} \quad (2)$$

Where:

$d_{euc}(x_{ij}, c_{kj})$	= Euclidean Distance between the i-th observation on variable j and the k-th cluster center on variable j
x_{ij}	= Value of the i-th observation on variable j
c_{kj}	= Value of the k-th cluster center on variable j
p	= Number of observed variables
n	= Number of total observations

3.5. Interpretation and Evaluation

The resulting clusters were analyzed and labeled according to their characteristics:
1. Cluster 1: High Contribution Loyal Partners
2. Cluster 2: Potential Future Partners
3. Cluster 3: Incidental Contributors

Interpretation focused on how each cluster could guide promotional strategies. Evaluation ensured that the clustering provided meaningful and actionable insights.

3.6. Knowledge Extraction

The final step was transforming the clustering results into practical knowledge. A web-based decision support system was developed using PHP and MySQL to automate clustering, visualize cluster profiles, and generate strategic promotional recommendations for SMK Bina Satria.

4. Result and Discussion

The following figures illustrate the implementation of the web-based system, including its main dashboard, data management, and input-output functionalities.

Fig. 2: Display Implementation Dashboard

The dashboard represents the central interface of the system, designed to provide users with streamlined access to its functionalities. It presents a comprehensive overview of the dataset, facilitates initiation of clustering processes, and displays essential performance indicators. This centralized interface enhances user efficiency in system navigation and monitoring.

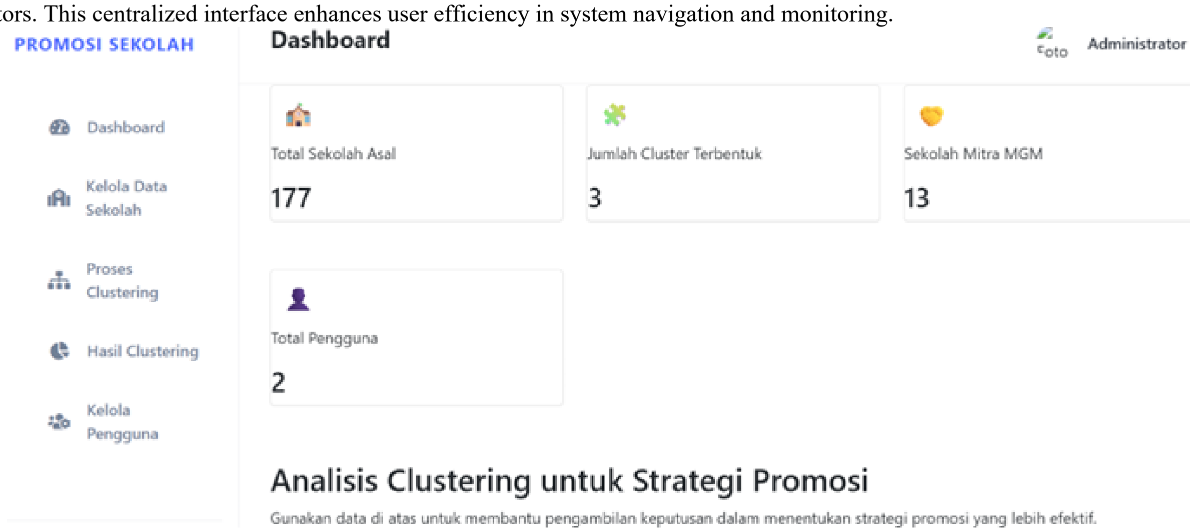


Fig. 2: Display Implementation Dashboard

Fig. 3: Implementation of the School Data Management View
This view enables comprehensive management of feeder school records. Administrators can access the full list of schools, track student contribution figures, and verify participation in the Member Get Member (MGM) program. The structured presentation ensures clarity, accuracy, and accessibility of data.

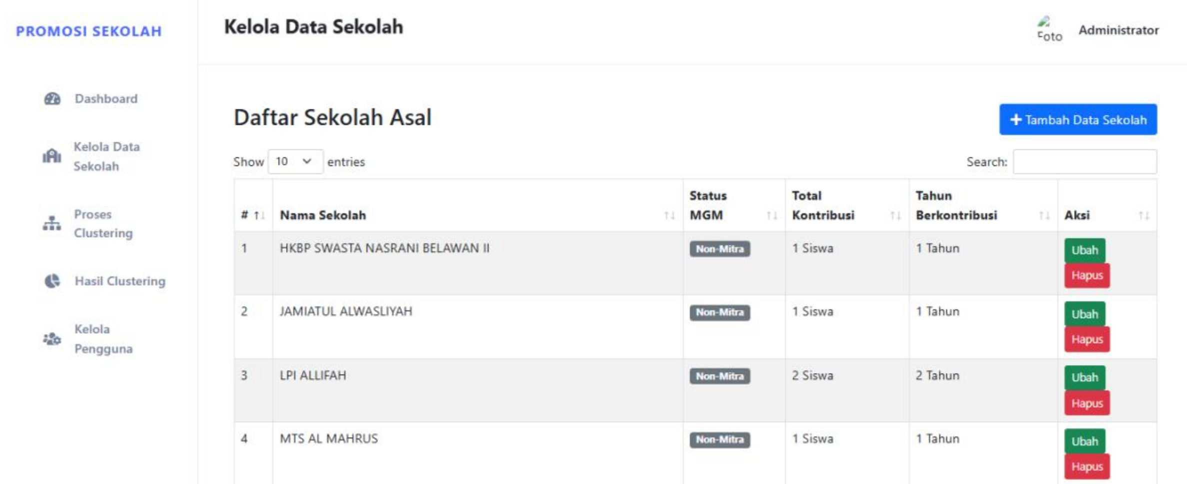


Fig. 3: Implementation of the School Data Management View

Fig. 4: Implementation of School Data Add View
The “Add Data” form supports the systematic entry of new feeder school information, including school identity, contribution size, contribution history, and MGM status. This feature improves data acquisition by ensuring accuracy, consistency, and efficiency in record integration.

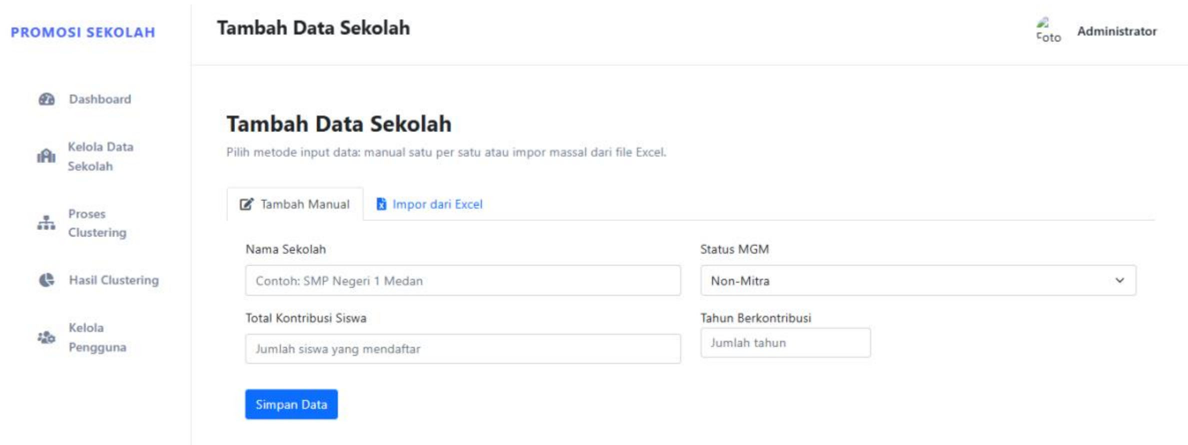


Fig. 4: Implementation of School Data Add View

Fig. 5: School Data Edit Implementation
This interface allows administrators to modify existing feeder school records. Revisions may include updated student contribution numbers or changes to MGM participation status. Such functionality ensures that the dataset remains current and valid for subsequent clustering analysis.

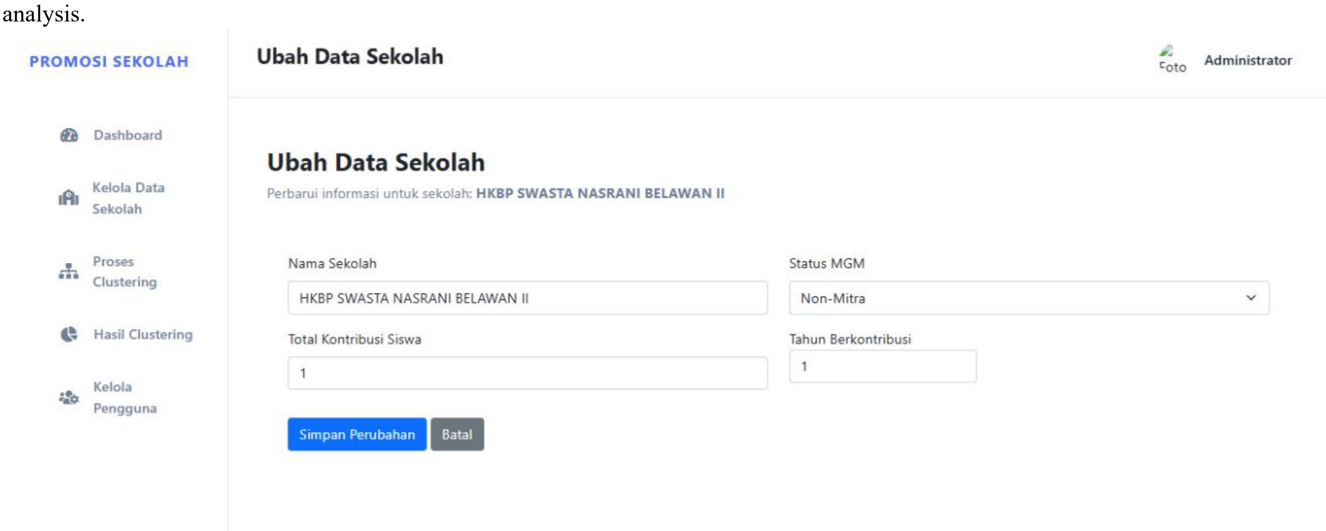


Fig. 5: School Data Edit Implementation

Fig. 6: Process of Running Clustering
The figure illustrates the automated execution of the K-Medoids clustering process. The system processes normalized data, applies distance computations, and assigns feeder schools to clusters. This mechanism minimizes manual intervention, reduces error potential, and ensures the consistency of results.

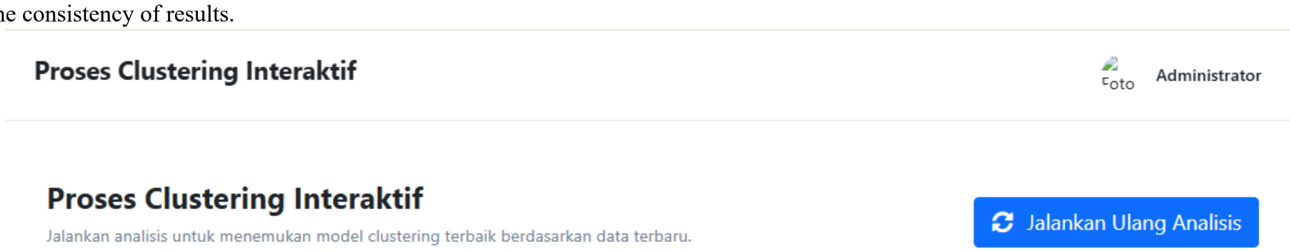
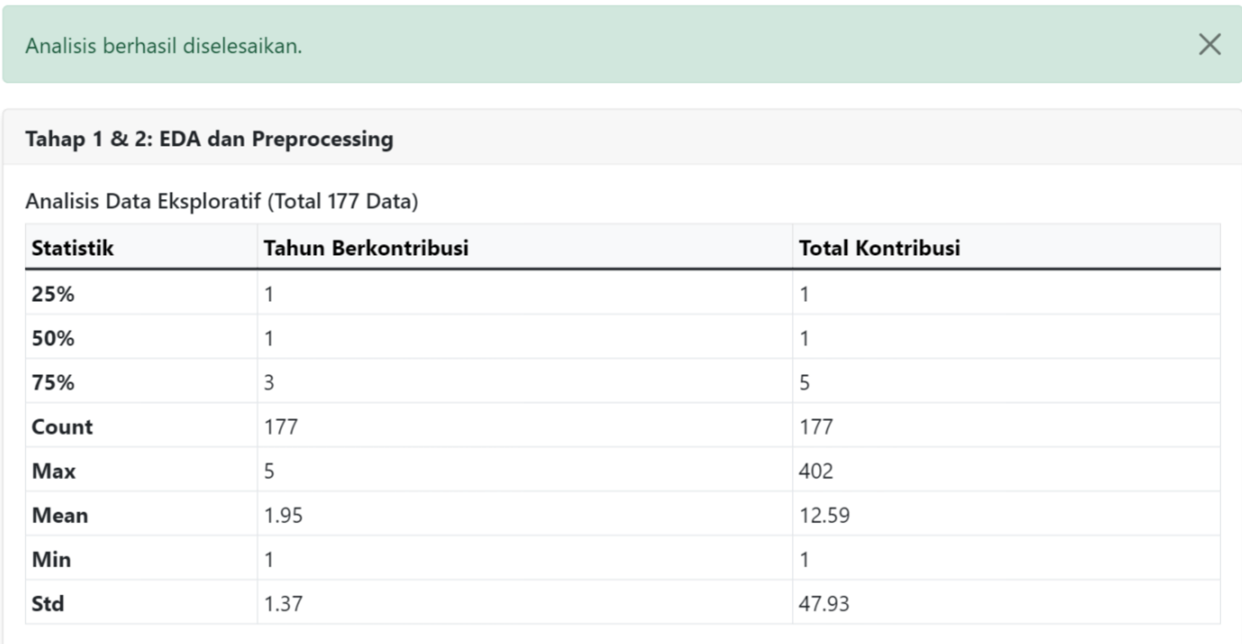


Fig. 6: Process of Running Clustering

Fig. 7: Stages of Data Analysis
The workflow demonstrates the sequential stages of analysis, beginning with preprocessing and normalization, followed by clustering, and concluding with evaluation using the Elbow Method and Silhouette Score. This structured methodology ensures the statistical validity and interpretability of the derived clusters.


Analisis berhasil diselesaikan.

Tahap 1 & 2: EDA dan Preprocessing

Analisis Data Eksploratif (Total 177 Data)

Statistik	Tahun Berkontribusi	Total Kontribusi
25%	1	1
50%	1	1
75%	3	5
Count	177	177
Max	5	402
Mean	1.95	12.59
Min	1	1
Std	1.37	47.93

Fig. 7: Stages of Data Analysis

Fig. 8: Sample Data Display Scaling

This illustration highlights the role of normalization in balancing variable scales. Attributes with wide-ranging values, such as student numbers and years of contribution, are transformed into standardized forms. This process ensures that no single attribute dominates the distance calculations, thereby supporting equitable clustering outcomes.

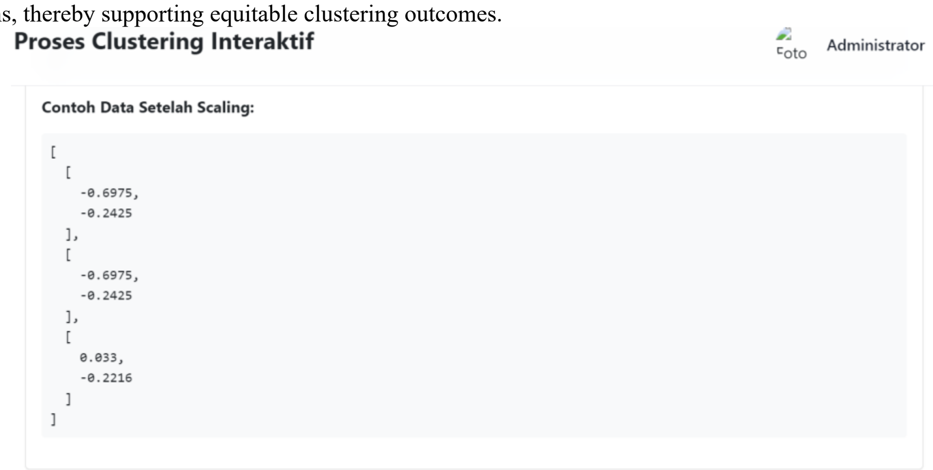


Fig. 8: Sample Data Display Scaling

Fig. 9: Model Training and Evaluation Stage

The figure demonstrates the training and evaluation of the K-Medoids model. Validation techniques are employed to confirm the optimal number of clusters and assess their stability. These procedures reinforce the robustness and reliability of the clustering results.

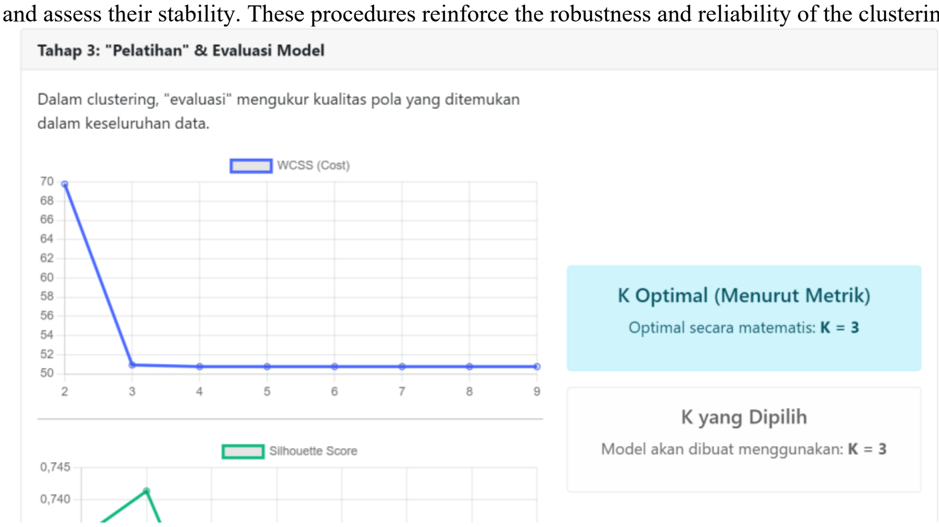


Fig. 9: Model Training and Evaluation Stage

Fig. 10: Example of Grouping Details Cluster

This figure presents the detailed assignment of feeder schools to their respective clusters. The categorization distinguishes high-potential, moderate-potential, and low-potential contributors. Such segmentation provides the foundation for designing targeted promotional strategies.

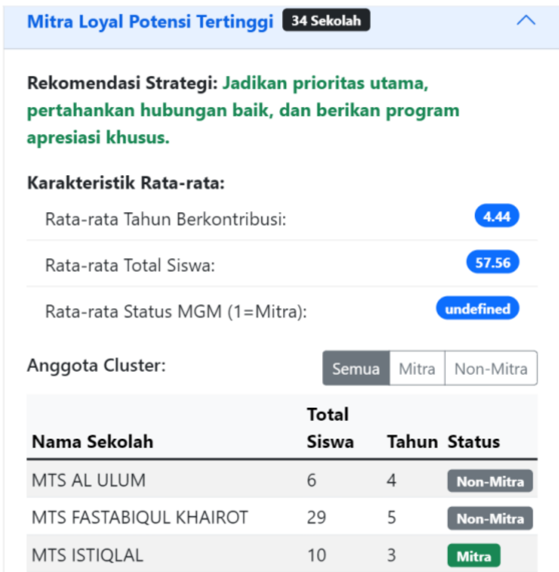


Fig. 10: Example of Grouping Details Cluster

Fig. 11: Graphics Cluster
The graphical representation illustrates the distribution of feeder schools across three clusters. The visualization facilitates intuitive interpretation by clearly distinguishing variations in contribution levels and MGM participation. This strengthens the communicability of the clustering outcomes.

#	Cluster	Jumlah Sekolah	% dari Total
1	Calon Mitra Potensial	39	22.0%
2	Mitra Loyal Potensi Tertinggi	34	19.2%
3	Kontributor Insidental	104	58.8%
	Total	**177**	**100%**

Fig. 11: Graphics Cluster

Fig. 12: Clustering Results View Implementation
This interface displays the system's clustering output, including cluster assignments, summaries, and overall grouping performance. It enhances transparency of the analytical process and delivers actionable insights for promotional planning.

Fig. 12: Clustering Results View Implementation

Fig. 13: Display After Process Completion
Following the clustering execution, this view provides confirmation of successful completion. It presents users with a concise summary and direct access to the detailed results, thereby ensuring clarity and continuity in the analysis workflow.

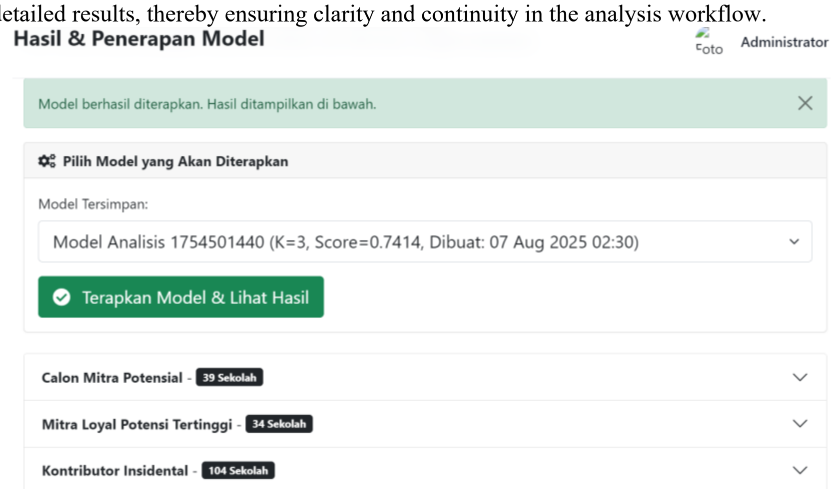


Fig. 13: Display After Process Completion

Fig. 14: Manage Users View
The user management interface supports administrative control by enabling the addition, modification, and removal of system users. This feature safeguards system security, enforces access control, and supports collaborative utilization of the platform.

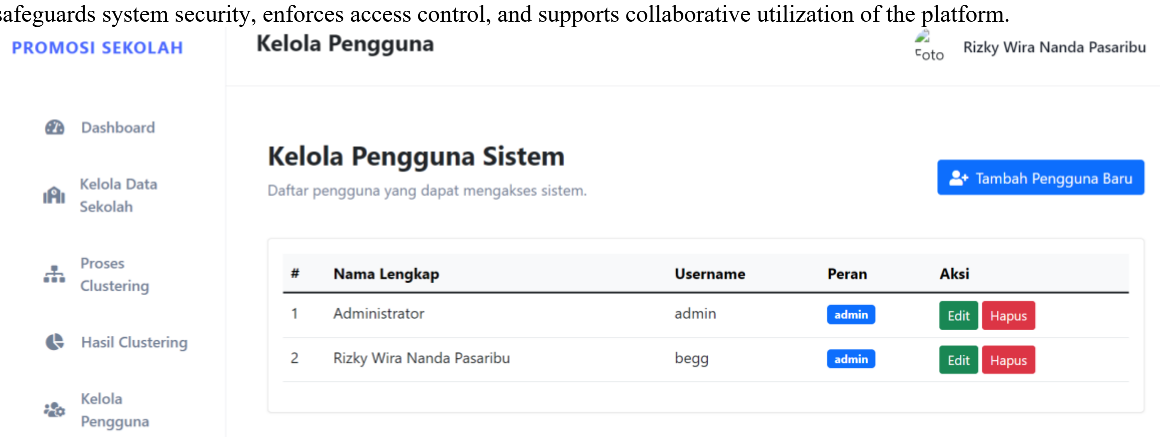


Fig. 14: Manage Users View

5. Conclusion

The application of the K-Medoids algorithm successfully grouped feeder schools of SMK Bina Satria into three clusters based on student contribution and MGM partnership status, providing a clear basis for targeted promotional strategies. Cluster 1 consisted of high-contributing and active MGM schools, Cluster 2 represented moderate contributors with strong future potential, and Cluster 3 included schools with low or inconsistent contributions. The developed web-based system further supported this process by enabling efficient data management, automated clustering, and clear visualization, making it a practical decision-support tool for data-driven promotional planning.

Acknowledgement

The author sincerely expresses gratitude to Universitas Muhammadiyah Sumatera Utara for providing the academic support and resources essential for the completion of this research. Special appreciation is extended to the management and staff of SMK Bina Satria for granting access to enrollment data and for their assistance during the data collection process. The author also wishes to acknowledge colleagues and peers for their valuable insights and constructive feedback, which significantly contributed to the development of this study. Deepest gratitude is dedicated to the author’s parents for their unwavering encouragement and support throughout the entire research process. Finally, the author conveys profound appreciation to the supervisor, Dr. Marah Doly Nasution, S.Pd., M.Si., for the patient guidance, and wholehearted supervision that made the successful completion of this study possible.

References

[1] E. A. Adi, A. A. Anwar, M. T. Lestari, A. Djuwita, R. R. W. Giri, and H. Millanyani, “Motion Graphic Sebagai Media Promosi SMK Telkom 1 Medan Sumatera Utara,” *Jurnal Seni, Desain dan Budaya*, vol. 9, pp. 249–262, Sep. 2024, doi: 10.36982/jsdb.v8i1.

[2] R. Siregar, “STRATEGI PROMOSI SEKOLAH MENENGAH KEJURUAN TARUNA MASMUR PEKANBARU DALAM MENARIK MINAT CALON SISWA BARU,” 2019.

[3] S. Darma and G. W. Nurcahyo, “Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids,” *Jurnal Informatika Ekonomi Bisnis*, vol. 3, pp. 89–94, Sep. 2021, doi: 10.37034/infeb.v3i3.87.

[4] Al-Khowarizmi, M. D. Nasution, Y. Sary, and Bela, “Clustering of uninhabitable houses using the optimized apriori algorithm,” *Computer Science and Information Technologies*, vol. 5, no. 2, pp. 150–159, Jul. 2024, doi: 10.11591/csit.v5i2.pp150-159.

[5] E. Rahmah, E. Haerani, A. Nazir, and S. Ramadhani, “Penerapan Algoritma K-Medoids Clustering Untuk Menentukan Srategi Promosi Pada Data Mahasiswa (Studi Kasus : Stikes Perintis Padang),” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 3, pp. 556–564, Jun. 2022.

[6] Yulianti, D. Y. Utami, and N. Hikmah, “Implementasi Data Mining Menentukan Game Android Paling Diminati Dengan Algoritma Apriori,” *Jurnal Komputer dan Informatika Universitas Bina Sarana Informatika*, vol. 21, no. 1, pp. 29–34, Mar. 2019, doi: 10.31294/p.v20i2.

[7] N. L. Anggreini, “TEKNIK CLUSTERING DENGAN ALGORITMA K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG,” *Jurnal Teknologi Informasi dan Pendidikan*, vol. 12, pp. 1–7, Dec. 2019, [Online]. Available: http://tip.ppj.unp.ac.id

[8] F. Hardiyanti, H. S. Tambunan, and I. S. Saragih, “PENERAPAN METODE K-MEDOIDS CLUSTERING PADA PENANGANAN KASUS DIARE DI INDONESIA,” *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1666.