International Journal of Informatics and Computation (IJICOM)
Vol.6, No.2, 2024
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Comparative Analysis of K-Means and K-Medoids Algorithms in New Student Admission

Tikaridha Hardiani[1], Esi Putri Silmina[2]

Abstract

Universitas 'Aisyiyah Yogyakarta is one of the private universities in Yogyakarta. The large number of private universities in Yogyakarta has intensified the competition for new student admissions. In this situation, every university requires the right strategy to attract prospective students. One of the strategies used by Universitas 'Aisyiyah Yogyakarta to capture the interest of potential students is by conducting direct promotions to schools in Yogyakarta, Java, and Sumatra. In the admission process for new students in the Information Technology Study Program, a common problem arises, which is the number of prospective students who do not complete re-registration each year. These students pass the selection and are declared accepted, but they do not proceed with re-registration. The school presentation strategy contributes to student admissions, making it a good strategy, but it requires significant operational costs. Promotion area segmentation is needed so that this strategy can be more targeted, resulting in more efficient spending. Segmenting or grouping promotion areas can be addressed using data mining techniques, specifically clustering. This study aims to segment promotion areas using clustering algorithms, namely K-Means and K-Medoids, along with the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The evaluation of DBI (Davies-Bouldin Index) showed that the K-Means algorithm performed better than the K-Medoids algorithm. The comparison between the K-Means and K-Medoids algorithms was assessed based on the DBI evaluation results, with the smallest DBI value observed in the K-Means algorithm. The DBI value for K-Medoids was 0.196, while for K-Means it was 0.170.

Keywords:
K-Means, K-Medoids, CRISP-DM, Data Mining, Clustering

1. Introduction

Universitas 'Aisyiyah Yogyakarta is one of the private universities in Yogyakarta. According to the LLDikti region 5 websites, there were 104 private universities in Yogyakarta in 2022. The large number of private universities in Yogyakarta has resulted in increasingly tight competition in the admission of new students. In this situation, each university requires an appropriate strategy to attract prospective students. One of the strategies used by Universitas 'Aisyiyah Yogyakarta to attract the interest of prospective students is by directly promoting to schools in Yogyakarta, Java, and Sumatra. A common problem in the admission of new students is the discrepancy in the number of students who do not re-register each year. Prospective students have passed the selection and been declared accepted but fail to complete their re-registration. The school presentation strategy contributes to the admission of new students and can be considered a good strategy, although it requires significant operational costs [1].

Promotion area segmentation is necessary so that this strategy can be more targeted and cost-efficient. The annual student admissions process generates a large amount of data in the form of new student profiles. From this large amount of student data, important information can be discovered by processing the data. The processing of student data is needed to extract new knowledge [2].

Corresponding Author: Tikaridha hardiani(tikarida8@gmail.com)
1 Tikaridha hardiani, Universitas 'Aisyiyah Yogyakarta, tikaridha@unisayogya.ac.id
2 Esi Putri Silmina, Universitas 'Aisyiyah Yogyakarta, esiputrisilmina@unisayogya.ac.id

Handling large data sets with many data points and attributes is not easy. One approach that can be used to process large data sets is data mining. The data mining technique used in this research is Clustering. Clustering is one of the data mining techniques that functions to group a set of data or objects into clusters (groups) so that each cluster contains as similar data as possible and is distinct from objects in other clusters [1]

The K-Means algorithm has advantages in that it is easy to implement and execute, relatively fast, flexible, and the most widely used in clustering[3]. This algorithm is one of the most important in data mining. The K-Medoids algorithm offers advantages in addressing the weaknesses of the K-Means algorithm, which is sensitive to noise and outliers, where objects with large values can deviate from the data distribution. Another advantage is that the clustering results are not dependent on the order of dataset entries[4]. This research aims to cluster promotion areas using clustering algorithms, namely K-Means and K-Medoids, and the Cross-Industry Standard Process for Data Mining (CRISP-DM) method[5]. The performance of both algorithms is compared in solving the new student admission data problem. The results of this analysis can assist in decision-making for determining the promotion strategy at Universitas 'Aisyiyah Yogyakarta.

2. Related Works

Previous research on the analysis of new student data includes a study conducted by Alhapizi on data mining for the registration of prospective new students using clustering techniques with the K-Means algorithm. Rapid Miner tools were used to run the algorithm. The case study was conducted at Universitas Bina Darma, Palembang, with the aim of determining new student promotion strategies[3].

A similar study focused on determining promotional areas or locations that have the potential to attract new students. The case study was conducted at Universitas Banten Jaya using the K-Means clustering method. This method was used to identify areas or locations with the potential for new student admissions by grouping research object items based on their similarities [5].

In research related to student data clustering, to facilitate the promotion of new student admissions, the author used the K-Means Algorithm for data clustering. The case study was conducted at SMK Wahidin. In this research, the researcher used two different operators for the measure types, namely Numerical Measures and Bregman Divergences, resulting in a Davies Bouldin Index [6].

Research on clustering student data to support promotion strategies by applying the K-Means algorithm was conducted at STMIK Bina Bangsa Kendari. This was done to determine promotional strategies for study programs at STMIK Bina Bangsa Kendari based on the results of clustering the most popular study programs from various schools [2].

Research on mapping toddler nutritional status by comparing the K-Means and Fuzzy C-Means algorithms was conducted at Puskesmas Pasir Jaya. From the comparison of the highest validation results and the comparison of the graphical results of the Fuzzy C-Means and K-Means algorithms, it was found that the K-Means algorithm had the highest accuracy value [7].

Related research published in 2024 by Jhiro Farana et al., titled "Comparison of K-Means and K-Medoids Algorithms in Class Clustering for New Graduate Students," concluded that K-Means produced two clusters with 40 students in Cluster 1 and 10 students in Cluster 2, and K-Medoids also produced two clusters with similar data in both Cluster 1 and Cluster 2 [8].

Research on clustering for students with intellectual disabilities to determine class placements showed that when experimenting with a small dataset, K-Means was more effective at handling small data sizes. Using the K-Means algorithm on the intellectual disability dataset resulted in a DBI of 0.161, while using K-Medoids resulted in an evaluation score of 0.281[9].

A study comparing the K-Means and K-Medoids algorithms in clustering suggested that the results could serve as a consideration for determining promotion strategies at TEDC Polytechnic Bandung in the New Student Admissions process. The K-Means algorithm performed better compared to the K-Medoids algorithm in addressing promotional infrastructure issues for new student admissions at TEDC Polytechnic Bandung[10].

Research published in 2022 by Karno Nur Cahyo et al., titled "Performance and Computational Speed Analysis of K-Means and K-Medoids Algorithms in Text Clustering," concluded that in tests conducted using the Davies Bouldin Index (DBI) calculation, the K-Means algorithm had a DBI value closer to 0, specifically -0.426, when tested with Term Weighting using Term Occurrences and NGrams set to 2. Meanwhile, the DBI range for the K-Medoids algorithm under the same conditions was quite far from 0, at -1.631 [11].

Previous research on clustering employee performance data, titled "Analysis of K-Means and K-Medoids Algorithms for Clustering Employee Performance Data at the National Housing Corporation," used 20 randomly selected data samples. Several parameters were used during the performance evaluation process, including strategy, job description, tasks, attendance, appearance, aggressiveness, problem-solving, and work results. Using the accuracy parameter, the K-Means algorithm scored 56%, while K-Medoids scored 14%. Using the recall parameter, the K-Means algorithm scored 60%, while K-Medoids scored 25%. Using the accuracy parameter again, K-Means scored 25%, and K-Medoids also scored 25%.

Clustering algorithms, particularly K-Means and K-Medoids, have been extensively studied and applied across various domains for data segmentation and analysis. K-Means is renowned for its simplicity, efficiency, and wide adoption in data mining tasks due to its fast computation and easy implementation. However, it is sensitive to noise and outliers, which can significantly impact clustering accuracy. To address this limitation, the K-Medoids algorithm was developed, offering robustness by using actual data points as cluster centers rather than centroids. The difference between this research and previous studies lies in the use of the K-Means and K-Medoids algorithms, comparing the two algorithms to assess which performs better in solving problems.

K-Means

K-means is an unsupervised machine learning algorithm used to partition a dataset into distinct, non-overlapping subgroups (clusters). Its primary goal is to group data points in such a way that points within the same cluster are more similar to each other than to points in other clusters. The algorithm iterates to assign each data point to one of \(k\) clusters based on the provided features [12]. K-Means clustering involves the following key steps:

1. **Initialization**:
   Select random data points from the dataset to act as the initial centroids (average positions of clusters). Alternatively, more advanced initialization methods, such as k-means++, can be used to increase the chances of finding better initial clustering.
2. **Assignment**:
   For each data point, calculate the distance to each centroid using a distance metric (usually Euclidean distance). Assign each data point to the nearest centroid, thus forming k clusters. Here is the Euclidean Distance formula:

$$De = \sqrt{(x_i - s_i)^2 (y_i - t_i)^2} \quad (1)$$

   Where:

- De is Euclidean Distance,
- i is the number of objects,
- (x, y) are the coordinates of the object, and
- (s, t) are the coordinates of the centroid.

3. **Update**:
   After all the points are assigned, recalculate the cluster centroids by computing the average of all data points in each cluster. The new centroid represents the updated center of the cluster.
4. **Repeat**:
   Reassign each data point to the nearest centroid and update the centroids again. Continue this process iteratively until the centroids do not change significantly (convergence) or until the predetermined number of iterations is reached.
   The value of the centroid can be obtained from the average of the respective cluster using the formula:

$$v_{ij} = \frac{1}{N} \sum_{k=0}^{N_i} x_{kj} \quad (2)$$

Where:

- v_{ij} is the centroid/mean of the i-th cluster for the j-th variable
- N_i is the number of data points that belong to the i-th cluster
- i, k are indices of the clusters
- j is the index of the variable

x_{kj} is the value of the k-th data point within that cluster for the j-th variable

5. **Termination**:

The algorithm stops when the centroids stabilize (i.e., there is minimal change in their positions between iterations) or the maximum number of iterations is reached [13].

K-Medoids

The K-Medoids algorithm is a clustering algorithm closely related to K-Means but differs in how the cluster centers (medoids) are determined. Instead of using the average (centroid) of data points within a cluster, K-Medoids selects actual data points as the cluster centers. This makes K-Medoids more robust to noise and outliers as it minimizes the sum of differences between points in the cluster and the medoid, rather than the sum of squared distances as in K-Means [14].

The steps for solving the K-Medoids algorithm are as follows [15].

1. **Initialization:**
   Select a number of k medoids randomly from the dataset as the initial cluster centers.
2. **Cluster Assignment:**

Calculate the distance between each data point and all medoids. Assign each data point to the cluster whose medoids has the shortest distance. Allocate each object to the nearest cluster using Euclidean Distance with the following formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

3. **Selection of New Medoid:**
   For each cluster, test all data points within the cluster as candidates for the new medoid. Choose the new medoid that minimizes the total distance within the cluster (sum of squared distances). The total deviation (S) is obtained by calculating the new total distance minus the old total distance. If the value of S is less than 0, swap the object with the cluster data to form a new set of cluster objects as the medoid.
4. **Iteration:**
   Repeat the steps of cluster assignment and selection of new medoid until the medoid no longer changes or convergence is achieved. Repeat steps 3 through 5 if medoid changes still occur; if no changes are observed, then the clusters and their respective members are obtained.
5. **Termination:**
   The algorithm stops if there are no changes in the medoid, or if the changes between iterations are very small.

K-Medoids is more resistant to outliers compared to K-Means but can be slower on large datasets due to the combinatorial evaluation of new medoids.
4o mini.

3. Proposed Method

The research method proposed by the researcher uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) [16], [17]

A. Business Understanding
   This stage determines the objectives of the research case predicting the spread of COVID-19. It involves translating and analyzing the goals and constraints, which become the formula for the data mining problem. Following this, strategies are prepared to achieve the objectives.
B. Data Understanding
   This stage involves data collection for analysis and investigation to understand the initial data pattern structure and obtain preliminary knowledge from data mining. The data is then evaluated for quality to eliminate missing values, duplicates, and typographical errors. If possible, a small subset of the data group containing patterns related to the problem is selected.
C. Data Preparation
   This stage involves preparing the data by selecting cases and variables to analyze that match the type of analysis to be performed. The data is then examined to determine if changes are needed for certain variables. After this stage, it is expected that the data will be ready and meet the criteria for modeling. Additional data or information may be added to facilitate the data mining process. This stage allows for optimization of attribute selection, thus obtaining significant attributes that improve the accuracy of the data mining process.
D. Modeling
   Modeling includes setting up the situation and specifications to enable data to be processed using the planned data mining methods. This stage requires tools or coding using specific programming languages so that the data mining process can be performed by a computer system. RapidMiner software is used at this stage. Modeling

involves the use of K-Means and K-Medoids algorithms. The process may return to the data preparation stage to adjust the data to better fit the specific data mining method's requirements.

E. Evaluation

Evaluating one or more models used in modeling to determine their quality and effectiveness before use. It involves determining if any model meets the objectives set in the initial phase. Evaluation and validation of performance (accuracy and processing time) of the three algorithms are conducted using confusion matrices and ROC curves. The evaluation and validation results will indicate the best algorithm for solving the problem.

F. Deployment

Deployment here refers to the use of the resulting model, which includes a series of methods and representative data that have been processed to provide optimal information during the data mining process. In a simple context, deployment involves using the final results of data mining, for example: reporting the results of the process using data mining to predict the spread of COVID-19. This research uses the data mining model based on the Cross-Industry Standard Process for Data Mining (CRISP-DM), developed in 1996 by analysts from various industries such as Daimler Chrysler, SPSS, and NCR. CRISP-DM provides a standard for the data mining process as a problem-solving strategy commonly used in both business and research. In the CRISP-DM standard, data mining has a lifecycle divided into 6 phases. The subsequent phase in the sequence depends on the output of the previous phase.

4. Experimental Setup

CRISP-DM Process

a. Business Understanding The business understanding of this research involves clustering and analyzing the cluster results of student data who have applied to Universitas ‘Aisyiyah Yogyakarta from 2017 to 2022. The results of this clustering can provide recommendations to decision-makers.
b. Data Understanding The student dataset consists of 12,750 records with 20 attributes. These attributes include: year of registration, registration number, student ID, full name, gender, religion, high school name, choice 1, choice 2, choice 3, accepted program, admission path, examination path, accepted, registration, province of origin, parent income, occupation, and knowledge of UNISA.
c. Data Preparation The dataset has 4,455 missing values, resulting in 8,295 records that can be processed.
d. Modeling is performed using the K-Means algorithm with RapidMiner software, as shown in Figure 1 and Figure 2.

K-Means

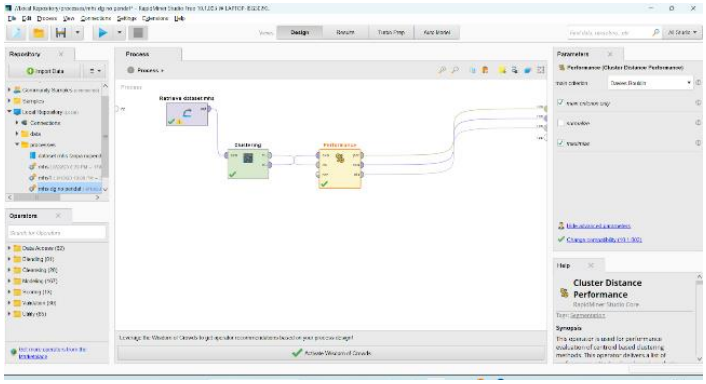

Fig 1. Modeling the K-Means Algorithm in RapidMiner

Figure 1 shows the RapidMiner application interface for modeling the K-Means algorithm, where the three interconnected relations are connected according to the defined algorithm flow.

K-medoids

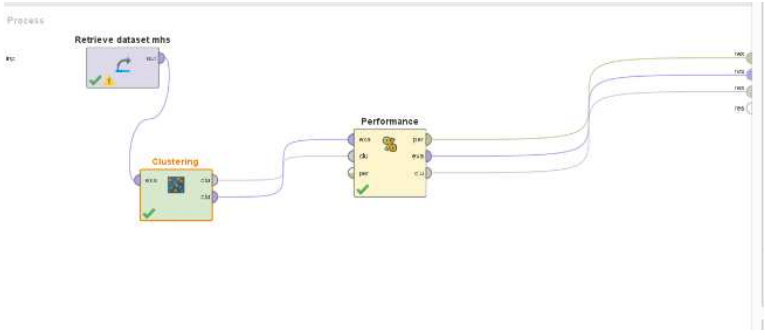

Fig 2. K-Medoids Modeling in RapidMiner

Figure 2 shows the RapidMiner application displaying the K-Medoids modeling results from the algorithm according to the rules.

a. Evaluation

Clustering evaluation uses the Davies-Bouldin Index (DBI). The DBI evaluation results for K-Means are shown in Table 1.

Table 1. Results of DBI Calculation

Cluster	DBI Value
2	0.478
3	0.495
4	0.389
5	0.277
6	**0.170**
7	0.249
8	0,317
9	0,378

10	0,412

The most optimal cluster is with 6 clusters, with a DBI value of 0.170. The results of the 6-cluster grouping are shown in Table 2. Cluster 0: 594 items, Cluster 1: 1044 items, Cluster 2: 1012 items, Cluster 3: 998 items, Cluster 4: 646 items and Cluster 5: 505 items. Total number of items: 4799

K-Medoids

Table 2. Results of DBI Calculation

Cluster	DBI Value
2	0.447
3	0.803
4	0,582
5	0,411
6	**0,196**
7	0,249
8	0,405
9	0,263
10	0,264

The most optimal cluster is with 6 clusters, with a DBI value of 0.196 for K-Medoids. The results of the 6-cluster grouping are shown in Table 2. Cluster 0: 505 items, Cluster 1: 1044 items, Cluster 2: 594 items, Cluster 3: 1012 items, Cluster 4: 998 items and Cluster 5: 646 items. Total number of items: 4799

5. Result and Analysis

In this study, a comparison of the K-Means and K-Medoids algorithms was conducted on new student admissions data, including the province of residence, parents' occupations, and obtained information. From the dataset of 8,295 records, several clusters with varying results were generated. The researcher explained that the most optimal cluster is with 6 clusters and a DBI value of 0.196 for K-Medoids. The DBI evaluation results show a comparison between K-Means and K-Medoids with the DBI value of K-Medoids at 0.196 and the DBI value of K-Means at 0.170, with the smallest DBI value found in the K-Means algorithm.

Cluster Analysis Results:

1. **Cluster 1:** Consists of 594 new student admissions from various provinces in Indonesia. The majority of new students come from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu, with most parents working as civil servants, private employees, or entrepreneurs/traders. Most of these new students learned about UNISA by visiting the university directly, with a smaller proportion learning through the UNISA website or from their parents/family.
2. **Cluster 2:** Consists of 1,044 new student admissions from various provinces in Indonesia. The majority of these students are from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu, with a diverse range of parental occupations including farmers, retirees, entrepreneurs/traders, private employees, laborers, civil servants, teachers, state-

owned enterprises, fishermen, military personnel, and undefined occupations. Most students learned about UNISA from UNISA alumni, friends/classmates, social media (Instagram, Facebook, Twitter), parents/family, the website, counseling teachers, and electronic brochures.

3. **Cluster 3:** Consists of 1,012 new student admissions from various provinces in Indonesia. The majority of these students come from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu, with diverse parental occupations such as civil servants, farmers, entrepreneurs/traders, laborers, private employees, teachers, state-owned enterprises, fishermen, military personnel, retirees, and undefined occupations. Most students learned about UNISA from social media (Instagram, Facebook, Twitter), friends/classmates, parents/family, the website, counseling teachers, high school promotions, UNISA alumni, and electronic brochures.
4. **Cluster 4:** Consists of 998 new student admissions from various provinces in Indonesia. The majority come from Aceh, Bali, Bangka Belitung, and Banten, with diverse parental occupations including civil servants, entrepreneurs/traders, farmers, laborers, fishermen, private employees, military personnel, and undefined occupations. Most students learned about UNISA from social media (Instagram, Facebook, Twitter), friends/classmates, parents/family, the website, UNISA research/community service/PKM, and counseling teachers.
5. **Cluster 5:** Consists of 646 new student admissions from various provinces in Indonesia, with some students coming from abroad (Timor Leste). The majority come from Aceh, Bangka Belitung, Banten, and Bengkulu, with diverse parental occupations including civil servants, police, entrepreneurs/traders, farmers, laborers, teachers, state-owned enterprises, retirees, and undefined occupations. Most students learned about UNISA by visiting the university directly, from friends/classmates, parents/family, the website, social media (Instagram, Facebook, Twitter), and other media.
6. **Cluster 6:** Consists of 505 new student admissions from various provinces in Indonesia. The majority come from Aceh, Bali, Bangka Belitung, Banten, Bengkulu, and Yogyakarta, with diverse parental occupations including teachers, entrepreneurs/traders, private employees, civil servants, laborers, teachers, military personnel, and retirees. Most students learned about UNISA by visiting the university directly, from the web.

Results of Analysis per Cluster

a. The analysis results for Cluster 1 include 505 new student admission data from various provinces in Indonesia, namely Aceh, Bali, Bangka Belitung, Banten, Bengkulu, and DI Yogyakarta. The majority of their parents work as private employees and entrepreneurs/traders. Among these 505 data entries, most new students learned about UNISA by visiting the university directly, while a smaller portion found out through UNISA's website or through their parents/family.
b. The analysis results for Cluster 2 include 1,044 new student admission data from various provinces in Indonesia, mostly from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu. The diverse parental occupations include farmers, laborers, fishermen, retirees, BUMN/BHMN employees, teachers, civil servants, private employees, and entrepreneurs/traders. Among these 1,044 data entries, most new students learned about UNISA from their parents/family, UNISA alumni, and through social media (Instagram, Facebook, Twitter), with a smaller portion learning from friends/classmates, guidance counselors, and UNISA's website.
c. The analysis results for Cluster 3 include 594 new student admission data from various provinces in Indonesia, with the majority coming from Aceh, Bali, Banten, and Bengkulu. Parental occupations vary, with the majority being civil servants. Among these 594 data entries, most new students learned about UNISA from their parents/family, UNISA alumni, and through social media (Instagram, Facebook,

Twitter), with a smaller portion learning from friends/classmates, guidance counselors, and UNISA's website.

d. The analysis results for Cluster 4 include 1,012 new student admission data from various provinces in Indonesia, mostly from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu. The various parental occupations include civil servants, private employees, entrepreneurs/traders, laborers, military personnel, and others. Among these 1,012 data entries, most new students learned about UNISA from friends/classmates, parents/family, and through social media (Instagram, Facebook, Twitter), with a smaller portion learning from UNISA's website, guidance counselors, and electronic brochures.
e. The analysis results for Cluster 5 include 998 new student admission data from various provinces in Indonesia, with the majority of new students coming from Aceh, Bali, Bangka Belitung, Banten, and Bengkulu. Parental occupations are diverse, including farmers, laborers, fishermen, entrepreneurs/traders, as well as some military personnel, civil servants, private employees, and some unknown. Among these 998 data entries, most new students learned about UNISA through UNISA's website, parents/family, and friends/classmates, with a smaller portion learning from UNISA alumni, electronic brochures, and social media (Instagram, Facebook, Twitter).
f. The analysis results for Cluster 6 include 646 new student admission data from abroad, specifically Timor Leste, and various provinces in Indonesia such as Aceh, Bangka Belitung, Banten, and Bengkulu. Most new students' parents work as civil servants and entrepreneurs/traders. Among these 646 data entries, most new students learned about UNISA by visiting the university directly and through the website, while a smaller portion learned through parents/family and friends/classmates.

   The comparison of K-Means and K-Medoids algorithms based on DBI evaluation results shows that K-Means has the smallest DBI. The DBI value for K-Medoids is 0.196, whereas for K-Means it is 0.170.
g. Deployment The final stage of the CRISP-DM method is reporting creation. The final report includes the knowledge gained or pattern recognition from the data during the data mining process.

6. Conclusion

Based on the research that has been conducted, clustering methods using K-Means and K-Medoids algorithms were obtained. The results of both algorithms were compared based on the DBI (Davies-Bouldin Index) evaluation. The values obtained were a DBI of 0.196 for K-Medoids and a DBI of 0.170 for K-Means. The smallest DBI value was found with K-Means. This difference indicates that K-Means generates more compact and well-separated clusters. The clustering results provide valuable insights for decision-makers in identifying more strategic and efficient promotional areas, reducing operational costs, and increasing the number of applicants that align with the university's target demographics. Additionally, the CRISP-DM methodology applied in this study proves its flexibility in handling large-scale data, encompassing phases from business understanding to the implementation of results. This study is expected to serve as a reference for developing data-driven promotional strategies in other educational institutions. Future research may explore the use of other algorithms, such as Fuzzy C-Means or hybrid models, to compare results and further enhance clustering accuracy.

References

[1] H. Hairani, D. Susilowati, I. Puji Lestari, K. Marzuki, and L. Z. A. Mardedi, "New Student Admission Promotion Location Segmentation Using RFM and K-Means Clustering Methods," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 275–282, Mar. 2022, doi: 10.30812/matrik.v21i2.1542.

[2] W. Lestari, S. Bina, and B. Kendari, "Clustering Student Data Using K-Means Algorithm to Support Promotion Strategy (Case Study: STMIK Bina Bangsa Kendari)," *SIMKOM*, vol. 4, no. 2, pp. 2715–906, Jul. 2019.

[3] R. Alhapizi, M. Nasir, and I. Effendy, "Application of Data Mining Using K-Means Clustering Algorithm to Determine New Student Promotion Strategy at Bina Darma University Palembang," *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 2775–2488, 2020, [Online]. Available: https://journal-computing.org/index.php/journal-sea/index

[4] B. Riyanto, "Application of K-Medoids Clustering Algorithm for Grouping Diarrhea Distribution in Medan City (Case Study: Medan City Health Office)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Dec. 2019, doi: 10.30865/komik.v3i1.1659.

[5] R. Budiman, "Application of Data Mining to Determine the Location of New Student Admission Promotion at Banten Jaya University (K-Means Clustering Method)," *Jurnal ProTekInfo*, vol. 6, no. 1, pp. 2406–7741, 2019.

[6] Y. Arie Wijaya, "Clustering of New Student Prospective Data Using the K-Means Method at Wahidin Vocational High School, Cirebon City," *Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, 2022.

[7] N. S. Fatonah and T. K. Pancarani, "Comparative Analysis of Clustering Algorithms for Mapping Toddler Nutritional Status at Pasir Jaya Health Center," *KONVERGENSI*, vol. 18, Jan. 2022.

[8] J. Faran and R. T. Aldisa, "Perbandingan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kelas Untuk Mahasiswa Baru Program Magister," *Journal of Information System Research*, vol. 5, no. 2, 2024.

[9] F. Harahap, "TIN: Nusantara Informatics Application Comparison of K Means and K Medoids Algorithms for Clustering Classes of Mentally Disabled Students," vol. 2, no. 4, 2021, Accessed: Dec. 06, 2024.

[10] N. Lestari Anggreini and S. Tresnawati, "Comparison of K-Means and K-Medoids Algorithms to Handle Promotion Strategy at TEDC Polytechnic Bandung," *Shandy Tresnawati TEDC*, vol. 14, no. 2, 2020.

[11] K. N. Cahyo, A. Subekti, and M. Haris, "Performance Analysis and Computational Speed of K-Means and K-Medoids Algorithms in Text Clustering," *Universitas Nusa Mandiri Jl. Raya Jatiwaringin*, vol. 15, no. 2, p. 13620, 2022.

[12] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.

[13] S. G. M. Al-Kababchee, Z. Y. Algamal, and O. S. Qasim, "Enhancement of K-means clustering in big data based on equilibrium optimizer algorithm," *Journal of Intelligent Systems*, vol. 32, no. 1, Jan. 2023, doi: 10.1515/jisys-2022-0230.

[14] E. Schubert and L. Lenssen, "Fast k-medoids Clustering in Rust and Python," *J Open Source Softw*, vol. 7, no. 75, p. 4183, Jul. 2022, doi: 10.21105/joss.04183.

[15] H. Chenan and N. Tsutsumida, "A Scalable k-Medoids Clustering via Whale Optimization Algorithm," Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.16993

[16] J. Han, M. Kamber, and J. Pei, *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, 2011.

[17] T. Hardiani, "Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta," *International Journal of Health Science and Technology*, vol. 2, no. 1, pp. 75–85, Jan. 2021.