Online Version Available at: http://bsti.ubd.ac.id/e-jurnal

JOURNAL OF TECH-E

|2581-1916 (Online) | 2598-7585 (Printed) |

Article

CLUSTERING ANALYSIS OF ADMISSION OF NEW STUDENTS USING K-MEANS CLUSTERING AND K-MEDOIDS ALGORITHMS TO INCREASE CAMPUS MARKETING POTENTIAL

Hassan Amin [1], *Wiranto Herry Utomo* [2],

[1,2] *President University, Teknik Informatika, Jawa Barat, Indonesia*

SUBMISSION TRACK

Recieved: August 10, 2023
Final Revision: August 21, 2023
Available Online: August 29, 2023

KEYWORD

Data Mining, Pmb, K-Means, K-Medoid, New Student Data, Clustering

CORRESPONDENCE

E-mail: hasan0amin@gmail.com
wiranto.herry@president.ac.id

A B S T R A C T

Acceptance of new students is a very important activity for a high school or university. The admissions data has not been utilized by the campus in making strategic decisions, marketing potential, and considering invitations through academic admissions. So, to assist in processing the new student admissions data, in this study the design and analysis of new student admissions data was carried out using stages in data mining. The clustering method approach can be applied in analyzing the potential level of PMB quality produced by utilizing the PMB recording dataset for the 2023 period. 86 data records. The K-Means and K-Medoids algorithm models that are applied have results that show a new insight, namely grouping based on 2 clusters, cluster 1 (C0) is a pass category while cluster 2 (C1) has not been determined. The results of the K-Medoids algorithm which has cluster 1 (C0) 60 results, cluster 2 (C1) has 26 results is a potential pass of 60 and has not yet been determined 26 of the data tested 86 while the results of the K-Means cluster 1 algorithm (C0) 40 , cluster 2 (C1) 46 is a potential pass consisting of 40 and 46 undetermined data from the 86 datasets tested. Testing using the RapidMiner Studio application can also produce similar insights, namely each cluster has Davies Bouldin Index or DBI results from each K-Means and K-Medoids algorithm. K-Means has a Davies Bouldin Index result of -0.533 while K-Medoids has a Davies Bouldin Index result of -0.877

INTRODUCTION

The new student admission activity is one of the annual routine activities as a medium for recruiting prospective new students. The high school is one of the high schools in Bekasi that opens a pathway for new student admissions through independent exams. In practice, the high school uses a new student admission information system as an operation to record new students who register and re-register.

Admission of new students is carried out based on national exam scores and attitude scores, skills and partly carried out based on non-academic achievements. It's just that after being accepted in this way, students take a written test to determine the class based on the grades obtained when taking the written test. The written tests included Indonesian, English and Mathematics subjects. It turns out that the results obtained show that the national exam scores obtained by students do not guarantee that students with high national exam scores will also get high scores when the written test is carried out, so that students will be in classes with low grades and *high* grades . With this process running, the amount of new student data from year to year has increased, so that more and more data must be managed. The PMB data includes student origin data, school origin data, employment data, study programs taken, as well as reference lines used to enter high schools.

Currently the PMB process in the high school environment has been running using a web-based information system for each registration or re-registration as a student. The process of accepting new students is very important for the high school, especially campus management, in this case the foundation. With the increase in student admission data from year to year, it sometimes makes the data more and more and only has a transactional nature. Meanwhile, campus management only requires data in the form of reports that are fast, easy, and do not interfere with transactions with the aim of facilitating management in making strategic decisions related to campus marketing potential with data on new student admissions.

Data Mining is an important element in managing intellectual capital and the decision-making process to help leaders and managers improve university performance in accepting new students. In research conducted by Ika Kurniawati using business intelligence to determine promotion strategies for new students, the results obtained were promotional strategies for prospective students according to the study program they were interested in.

While the research by Indri Fatma, Heru Satham Tambunan, Fitri Rizki *cluster* analysis of outstanding students using data mining, namely the K-Medoid *Cluster Algorithm* . Where previously the school still used a manual method in determining students who excel at the school, so it took quite a long time and the results were not quite right. K-Medoid *Cluster* is one of the algorithms used for classification or grouping of data, the authors apply the *K-Medoid Cluster algorithm* in grouping students with achievements in order to get more accurate, fast and effective results.

Meanwhile, research by Ramdani Udiman, Rudianto. This research intends to determine which locations or areas have the potential to bring in new students at the college. By applying data mining with the clustering method grouping research object items based on their similarity, so that it will be known which areas have the potential to bring in new students. Determining the location of the promotion of new student admissions using the data mining method will have a good and targeted impact in carrying out promotions, so as to increase the number of new students each year.

Based on the description of the problem above, this research takes the Grouping Analysis of New Student Admissions Using the K-Means Clustering and K-Medoids Algorithms To Increase Campus Marketing Potential to find out how much campus marketing potential is accepted by each study program. one of the algorithms used for *clustering or grouping data is the K-Means and K-Medoid Cluster* algorithms in grouping based on a set of data from the attributes.

I. LITERATURES REVIEW

The Promise of MOOCs Revisited? Demographics of Learners Preparing for University[28]. From an initial sample of 260,239 learners, we cluster analyze a subset of data from 29,083 participants who submitted an assignment in one of nine entry-level MOOC courses. Manhattan distance and Gower distance measures are computed based on engagement, achievement, and demographic data. To our knowledge, this marks one of the first such uses of Gower distance to cluster mixed-variable data to explore fairness and equity in the MOOC literature. The clusters are derived from CLARA and PAM algorithms, enriched by demographic data, with a particular focus on education level, as well as approximated socioeconomic status (SES) for a smaller subset of learners. Results indicate that learners without a college degree are more likely to be high-performing compared to college-educated learners. Learners from lower SES backgrounds are just as likely to be successful as learners from middle and higher SES backgrounds. While MOOCs have struggled to improve access to learning, more fair and equitable outcomes for traditionally underrepresented learners.

Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study[26]. This study identified 108 (4.63%) out of 2,334 feeder schools as the prioritized target market, contributing 51.54% of 18,537 enrolled students to the university during the analysis period. The prioritized feeder schools are in 32 cities and 23 regencies from 368 cities/regencies, with the majority being private schools. The research's findings revealed the distribution of feeder schools in regencies/cities and the trend of enrolled students from the highest value feeder schools segment, which can assist university management in selecting target feeder schools more precisely. Based on the findings, decision-makers can create a geo-marketing strategy for promotional activities and direct resources to the prioritized feeder schools. This study contributes by reinforcing a modified RFM model with spatial and temporal analysis to help university decision-makers choose feeder schools as the university's target market and develop a geo-marketing strategy.

Combination of K-Means and Simple Additive Weighting in Deciding Locations and Strategies of University Marketing[22]. t uses the K-Means method for data grouping and the Simple Additive Weighting (SAW) for ranking the results of data grouping. The result of this research suggests that location of promotion may be determined from the clustering process using the K-Means method. The silhouette coefficient test invalidates the data clustering, and the SAW method helps the ranking process to obtain a sequence of promotion locations. The ranking results reflect the predetermined decision table that directs promotion location selection according to the promotion strategy. The combination of the two methods helps to decide the location and marketing strategy to optimize time, effort, and cost. The results of this study may be used as a comparative reference for the management to decide the right promotion strategy based on the locations and student background.

II. FRAMEWORK

This research creates a framework that is useful as a research guide so that it can be carried out consistently. For this reason, the algorithm used is the K-Means and K-Medoids algorithm to solve the problem by testing the performance of the method. For the development and testing of the methods used by using the RapidMiner application. The following is a picture framework thinking:

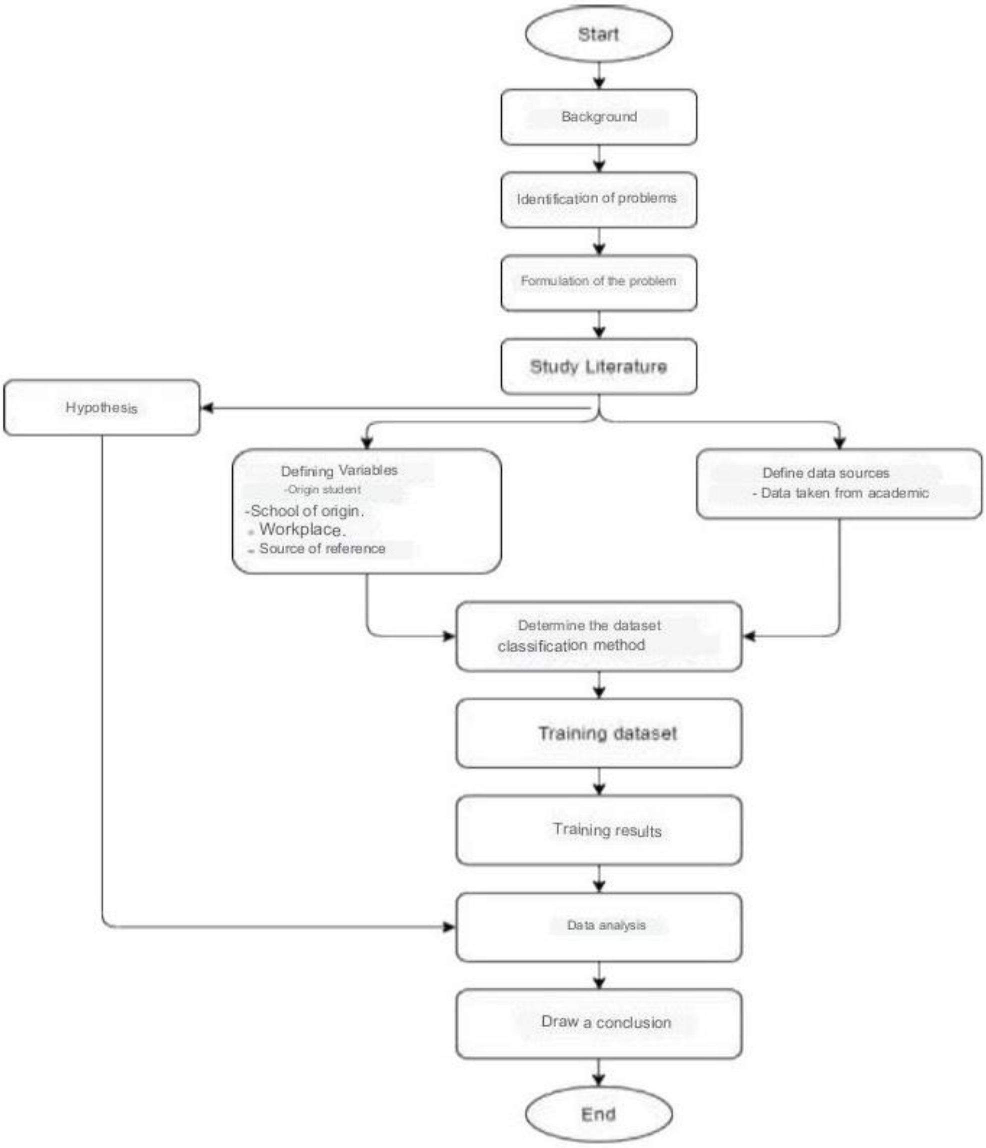

Figure 1 : Description of the framework

III. METHODS

3.1 Object of research

In this study, the stages that will be used in clustering new student admissions data, in determining attributes to facilitate research so that research can run properly and systematically, and fulfill the desired goals, to describe a situation or phenomena as they are and describe or analyze a research result but not used to make broader conclusions.

3.2 K-means Clustering

K-means clustering is a non-hierarchical data clustering method that groups data in the form of one or more clusters/groups. Data that has the same characteristics are grouped in one cluster/group and data that has different characteristics are grouped with other clusters/groups so that data in one cluster/group has a small degree of variation (Ika Kurniawati, 2019). K-means is one of the representative algorithms belonging to the partition-based clustering algorithm [10] . We need to set the number of data sample clusters, initialize them randomly to each center, which has the same length as the vector of each data point. This requires us to predict in advance the number of clusters (that is, the number of centers). Then the distance from each data point to the center point is calculated, and the data points are divided into which class clusters are closest to which center point, and the midpoint of each class is calculated as a new center point. This was repeated until each type of center did not change much after each repetition (Haonan Ma et al, 2019). The k-means algorithm is an easy and effective algorithm for finding clusters in data. The stages in the clustering process with k-means include [11] .

1. Specifies how much data to partition.
2. Randomly assign k records to be the initial cluster center locations.
3. Finds the nearest cluster center. So, in a sense, each cluster center "owns" a subset of records, thus representing a partition of the data set. We therefore have k clusters, C1, C2,. . . , ck.
4. Finding the cluster centroid and updating the location of each cluster to the new centroid value in each of the k clusters
5. Repeats steps 3 to 5 until convergence or stops. The closest criterion at step three, is usually rarely Euclidean, although other criteria can be applied. Cluster centroid in step 4 is found as follows. Suppose there are n data point values (a1, b1, c1), (a2, b2, c2) ,..., (an,bn,cn), the centroid of each of these points is the process of point gravity and is located at the point (∑ai /n, ∑bi /n, ∑ci /n).

The k-means algorithm will end when the centroid no longer changes. That is, the algorithm ends when all clusters C1, C2, C3 ….., Ck. All records owned by each main cluster center in the cluster. And also the k-means algorithm will end when some convergence criterion is met, such as there is no significant reduction in the number of squared errors [11]

$$\text{Vij} = \frac{1}{Ni}\sum_{k=0}^{Ni}\text{Xkj}$$

Information :

Vij = The average centroid of the ith cluster for the ith variable

Ni = number of clusters -i

i, k = index of the cluster

j = index of variable

Xkj = the kth data value of the jth variable in the cluster

$$D = \sqrt{(Xi - Si)^2 + (yi - ti)^2}$$

Information :

D = Euclidean Distance

i = the number of objects

(x,y) = object coordinates

(s,t) = centroid coordinates

3.3 K-Medoids Cluster Algorithm

K-Medoids Cluster model is a *clustering technique* in data mining. There is another *cluster* model , namely *the K-Means Cluster* . The most striking difference between the two *clustering methods* is at the center of each *cluster* . If *K-Medoids* uses the object as a representative (medoid) as the center *of the cluster* , *K-Means* uses the average value (mean) as the center of *the cluster* . Following are the steps for solving the *K-Medoids Cluster* :

1. Initialize k cluster centers (*number* of *clusters)* .
2. Allocate each data (object) to the nearest *cluster* using size Euclidean Distance with the equation:

$$D = \sqrt{(Xi - Si)^2 + (yi - ti)^2}$$

3. Randomly select objects in each *cluster* as candidates new medoid.
4. Calculate the distance of each object that is in each of each *cluster* by taking new medoids.
5. Calculate the total deviation (S) by calculating the new total distance value long total distance. If S<0, then replace the object with *cluster data* to get a new group of k objects as medoids *d* = (1)

3.4 Research Stage

The research phase can be defined as the decomposition of a research process that is intended to identify and evaluate the problems that occur. In the picture below is the Research stage.

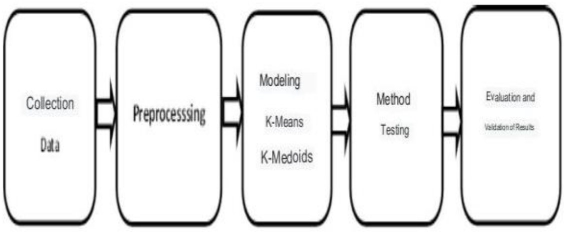

Figure 2 : Research Stage

Data Processing

Data processing in this study uses data *clustering techniques* , data mining for data *clustering* of new student admissions or bad product results using the *k-means algorithm, K-Medoids* . Data can be processed and used as a dataset in this study, this data will be divided into training data, testing and separating data as needed in this study so that the model can be obtained, has good generalization abilities in clustering data, training data or data testing is part of from powder coating data that is trained to create data clustering or run the function of an algorithm according to its purpose in this study so that it can see the accuracy or performance of a data.

3.2.3 Data Preprocessing

Data Preprocessing is the process of selecting data from a set of existing operational data before entering the data and information mining stage. Data that has gone through the data *selection process* will become data that is ready to be processed and stored in excel type files. At this stage data selection will also be carried out by selecting the required attributes for the next process, so that attributes that have no connection with the stages of model implementation will later be reduced so that the data dimensions become more concise. In the initial dataset there are 7 attributes as shown in Table 1 below.

Table 1 Dataset Initial Attribute

NO	Attribute Name	type
1	**Name**	date
2	**Study Program Choices**	date
3	**Test Pass**	date
4	**Finalized**	date
5	**File Pass**	date
6	**Study Program Accepted**	date
7	**Already a Student**	date

In the attribute selection process, several attributes that are unrelated and related to the clustering modeling stage and the *K-Means* and *K-Medoids algorithms* will not be used in the next process. Attributes that are used are only attributes **of passing tests** and **study programs being accepted** both of which have a numeric data type.

Table 2 Attribute Selection Results

No	Attribute Name	type
1	**Test Pass**	date
2	**Study Program Accepted**	date

The process of converting the initial data format into a standard data format for the process of reading data with the algorithm in the application used, then the process of testing the data using the programs and tools used . The following are the results of the initial data processing after going through the above stages to be used as a dataset at a later stage.

Table 3 Dataset Initial Attributes

Registration_No	Test Pass	Study Program Accepted
0432902210428	1	3
0432902210429	2	2
0432902210430	1	4

IV. RESULT

4.1 Application of K-Means and K-Medoids Algorithms

4.1.1 Data analysis

The initial stage carried out in this study was to prepare data where the data to be processed is a collection of new student admissions datasets. As a test of the K-

Means and K-Medoids algorithms with the clustering method the author takes the data to be modeled as many as 86 data records, where the data is an aggregation of raw data with a total of 200 rows of data records that have passed pre-processing data as shown in the table below this :.

Table 4 Datasets

No	Name	Study Program Accepted	Test Pass
1	Irvandita Pratama Alamsyah	3	1
2	Nur Sadiyah	2	1
3	Muhammad Yusuf	2	1
4	Mochammad Rizky Purnomo	1	1
5	Kusumawati Rohmah	1	1
6	Ratno Wijaya Kusuma	3	1
7	Diamond Wulan Sari	2	1
8	Isnanto	2	2
9	Rachel Sanosa	3	2
10	Yohana Sari Sitompul	3	2
11	Charisma Indra Saputra	1	2
12	Muhammad Barokah Tri Setianto	2	1
...	...	...	...
...	...	...	...
...	...	...	...
83	Erlang Tri Hardana	1	1
84	Ribie Ibn Hasan	1	1
85	Iqbal Febriansyah	2	2
86	Widiya Mufitalia	1	2

Evaluation of Tests on the RapidMiner Studio Application

In this process the clustering method with the K-Means algorithm is applied to form cluster groups with precise accuracy. In this study the authors used a calculation test with the help of the RapidMiner Studio application. Testing the model obtained by using the RapidMiner Studio application is carried out with the following steps:

1. Import the data needed for processing on the Rapid Miner tools. In the RapidMiner Studio application select and click **Import Data** , then select the data to be used and then determine the attributes and labels to be used.
2. Click the **Design menu** , in the process view, add the dataset in the folder to the process view screen. On **the Names & Roles menu, look for the** *Set Role* function which will later be used to set the attribute roles, then drag it to the process display screen.
3. Then on the **Normalization menu** select *Normalize* and drag it to the process display screen, through this function you can set the data normalization to be carried out from the dataset used in this process.
4. Then on the **Modeling menu** , in the **Segmentation submenu, select** the K-Means and K-Medoids function , to apply the K-Means and K-Medoids algorithms on the clustering process to be carried out. In this function we can determine the number (k) of clusters to be used in data modeling. The following are the parameters used to calculate the clustering process on the PMB recording data .

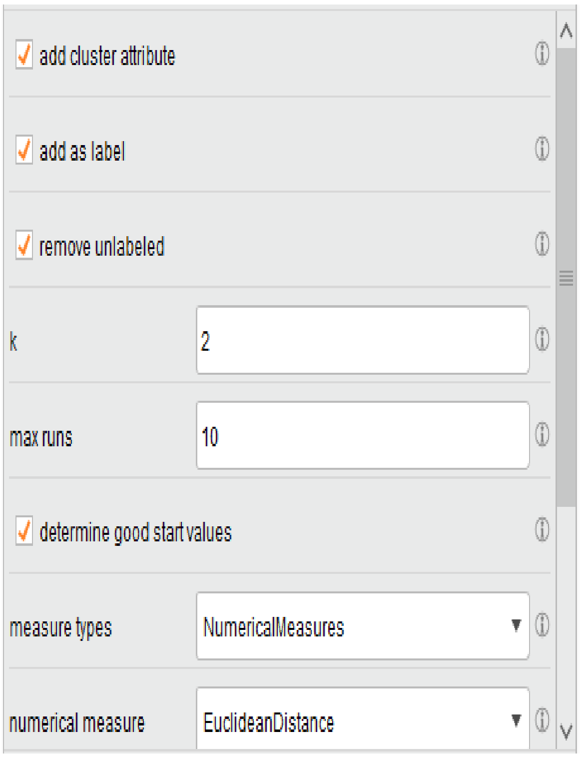

Figure 3 Parameters in *Function Clustering*

5. **Performance** function to display the David-Bouldin Index (DBI) value obtained from the clustering process of the PMB recording data used.
6. Connect all these commands so that the process display screen shows the following flow:

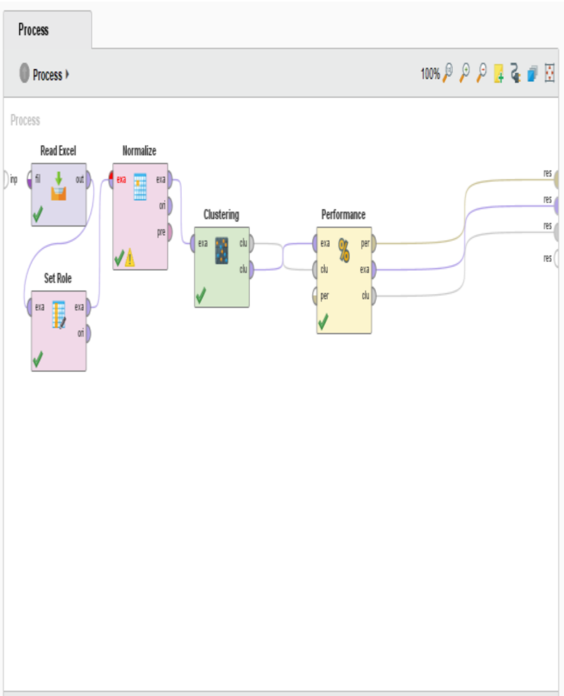

Figure 4 A series of *clustering* processes in the RapidMiner Studio application

7. Do ***the Running Process*** to get the clustering results from 86 *record* dataset that is used.

4.2 Analysis of Test Results

After carrying out the steps in finding cluster groups through the clustering method, the use of the K-Means and K-Medoids algorithms used produces a cluster grouping of each data. The PMB recording dataset used is 86 data records which will be tested in the process of forming cluster groups with the K-Means and K-Medoids algorithms. The results of the K-Medoids cluster model in the RapidMiner Studio application test show that of the 86 data, 60 data entered the first cluster group (C0) and 26 data entered the second cluster group (C1), which can be seen in the following figure.

Cluster Model

```
Cluster 0: 60 items
Cluster 1: 26 items
Total number of items: 86
```

Figure 6 : Results of the K-Medoids Cluster Model

The results of the K-Means cluster model in the RapidMiner Studio application test show that of the 86 data, there are 46 data entered in the first cluster group (C0) and 40 data entered in the second cluster group (C1), which can be seen in the following figure.

Cluster Model

```
Cluster 0: 46 items
Cluster 1: 40 items
Total number of items: 86
```

Figure 5: K-Means Cluster Model Results

The general description of cluster groups from each cluster 0, 1 and 2 can be seen in the clustering tree as follows.

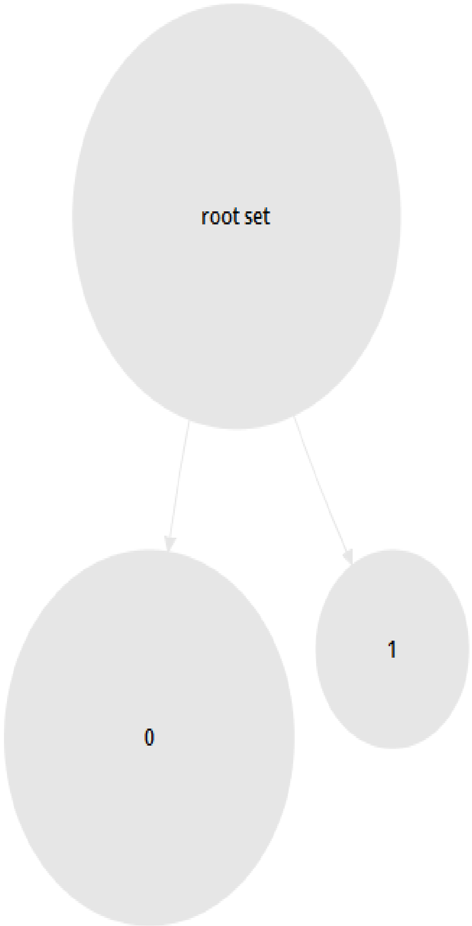

Figure 7: Cluster Tree formed on K-Medoids

The general description of cluster groups from each cluster 0, 1 and 2 can be seen in the clustering tree as follows.

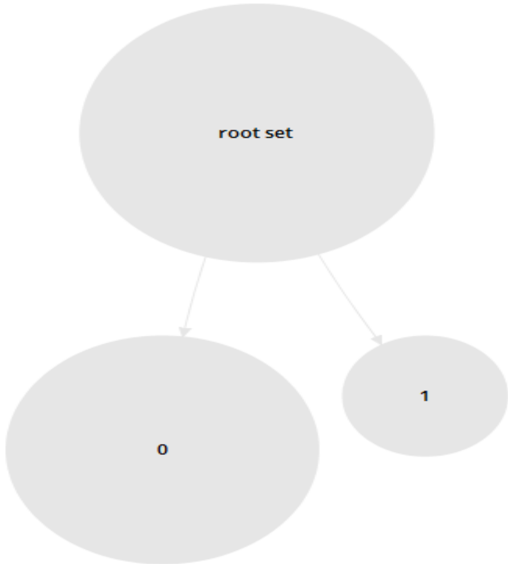

Figure 8: Cluster Tree formed on K-Means

From the results above it can be seen that the formation of cluster members obtained through testing with the RapidMiner application is relevant to the K-Medoids model. It's just that in the process of using the RapidMiner Studio application.
The most optimal cluster point for each variable for Cluster 0 (C0) is 1 & 3 and for Cluster 1 (C1) is 2 & 2 as can be seen in the image below.

Table 5 Optimal Cluster in the RapidMiner Studio application

attributes	Cluster 0	Cluster 1
Study Program Accepted	1	3
Test Pass	2	2

From the results above it can be seen that the formation of cluster members obtained through testing with the RapidMiner application is relevant to the K-Means model. The most optimal cluster point for each variable for Cluster 0 (C0) is 2,565 & 1 and for Cluster 1 (C1) is 1,304 & 1,100 as can be seen in the image below.
Table 6 Optimal Cluster in the RapidMiner Studio application

attributes	Cluster 0	Cluster 1
Study Program Accepted	2.565	1
Test Pass	1.304	1.100

Performance testing of models and algorithms is carried out with the intention of knowing the results of the calculations being analyzed and measuring the K-Medoids methods and algorithms used whether they function properly or not.

PerformanceVector
PerformanceVector:
Avg. within centroid distance: -1.023
Avg. within centroid distance_cluster_0: -1.150
Avg. within centroid distance_cluster_1: -0.731
Davies Bouldin: -0.877

Figure 9: Performance of the K-Medoids Algorithm Clustering

Performance testing of models and algorithms is carried out with the intention of knowing the results of the calculations being analyzed and measuring the K-Means methods and algorithms used whether they function properly or not.

PerformanceVector
PerformanceVector:
Avg. within centroid distance: -0.287
Avg. within centroid distance_cluster_0: -0.457
Avg. within centroid distance_cluster_1: -0.090
Davies Bouldin: -0.533

Figure 10: Clustering Performance of the K-Means Algorithm

The results of the evaluation value with the Davies Bouldin Index or DBI based on the RapidMiner Studio application obtained from tests obtained from the test results on the RapidMiner Studio application show the number 0.877 as shown in Figure 7.

Davies Bouldin

```
Davies Bouldin: -0.877
```

Figure 11 Results of the Davies-Bouldin Index K-Medoids Algorithm

The results of the evaluation value with the Davies Bouldin Index or DBI based on the RapidMiner Studio application obtained from tests obtained from the test results on the RapidMiner Studio application show the number 0.533 as shown in Figure 8

Davies Bouldin

```
Davies Bouldin: -0.533
```

Figure 12 Results of the Davies-Bouldin Index K-Means Algorithm

V. DISCUSSION

The value of novelty in this study is the existence of data attributes and comparison of algorithms as variables and data processing that will be used as research. Using 2 attributes and two algorithms as a comparison, so that the research that I am doing needs to be updated in this research or needs to be improved in terms of the algorithm and data regarding the attributes used and the comparison of an algorithm so that it can produce a novelty value in the research that I am doing. The choice of method for this study was due to the relatively high percentage. Testing using the RapidMiner Studio application can also produce similar insights, namely that each cluster has Davies Bouldin Index or DBI results from each K-Means and K-Medoids algorithm. K-Means has the results of the Davies Bouldin Index giving a better value.

VI. CONCLUSION

The clustering method approach can be applied in analyzing the potential level of PMB quality produced by utilizing the PMB recording dataset for the 2023 period, there are 86 data records. The K-Means and K-Medoids algorithm models that are applied have results that show a new insight, namely grouping based on 2 clusters, cluster 1 (C0) is a pass category while cluster 2 (C1) has not been determined. The results of the K-Medoids algorithm which has a cluster 1 (C0) result of 60, cluster 2 (C1) has a result of 26 which is a potential pass of 60 and has not been determined 26 from the data tested 86 while the results of the K-Means algorithm testing cluster 1 (C0) 40 , cluster 2 (C1) 46 is a potential pass consisting of 40 and 46 undetermined data from the 86 datasets tested. Testing using the RapidMiner Studio application can also produce similar insights, namely each cluster has Davies Bouldin Index or DBI results from each K-Means and K-Medoids algorithm. K-Means has a Davies Bouldin Index result of -0.533 while K-Medoids has a Davies Bouldin Index result of -0.877.

VII. ACKNOWLEDGEMENT

Praise and Gratitude the writer goes to Allah Subhanallahu Wata'ala who has given His mercy and guidance so that the writer can finish writing this thesis. On this occasion the author expresses his deepest gratitude to:

1. Mr Prof. Dr. Ir. Wiranto Herry Utomo, M.Kom. Head of the S-2 Information Technology Study Program
2. Mr Prof. Dr. Ir. Wiranto Herry Utomo, M.Kom. as supervising lecturer
3. All lecturers, family, friends, who have helped the writer in the preparation of this THESIS to completion which the writer cannot mention one by one.

The writer realizes that this writing is far from perfect, for that the writer really hopes for the support from the readers to provide constructive criticism and suggestions for future improvement.

REFERENCES

[1] Alharbi, Z., Cornford, J., Dolder, L., & De La Iglesia, B. (2016). Using data mining techniques to predict students at risk of poor performance. *Proceedings of 2016 SAI Computing Conference, SAI 2016*, 523–531. https://doi.org/10.1109/SAI.2016.7556030

[2] Ardista, N., Purbandini, P., & Taufik, T. (2017). Rancang Bangun Data Warehouse Untuk Pembuatan Laporan dan Analisis pada Data Kunjungan Pasien Rawat Jalan Rumah Sakit Universitas Airlangga Berbasis Online Analytical Processing (OLAP). *Journal of Information Systems Engineering and Business Intelligence*, *3*(1), 40. https://doi.org/10.20473/jisebi.3.1.40-51

[3] Fan, Z., Sun, Y., & Luo, H. (2017). Clustering of college students based on improved K-means algorithm. *Journal of Computers (Taiwan)*, *28*(6), 195–203. https://doi.org/10.3966/199115992017122806017

[4] Gustian, D., Rahmawati, A. F., Titin, Putra, R. R., & Anisa, P. (2019). Comparison of classification data mining in process majors students. *Proceedings - 2018 4th International Conference on Computing, Engineering, and Design, ICCED 2018*, 125–130. https://doi.org/10.1109/ICCED.2018.00033

[5] Huang, S. Y., & Zhang, B. (2019). Research on improved k-means clustering algorithm based on hadoop platform. *Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2019*, 301–303. https://doi.org/10.1109/MLBDBI48998.2019.00067

[6] Jiao, Y., Shen, Y., & Fei, S. (2016). A modified K-means algorithm based RBF neural network and its application in time series modelling. *Proceedings - 14th International Symposium on Distributed Computing and Applications for Business, Engineering and Science, DCABES 2015*, 481–484. https://doi.org/10.1109/DCABES.2015.126

[7] Kurniawati, I. (2019). Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru Peran Bussines Intelligence Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru, *1*(November 2017), 70–79

[8] Ma, H., Chen, H. Y., Meng, H., Yana, W., Yuhan, C., & Yifei, J. (2019). Application of K-means clustering algorithms in optimizing logistics distribution routes. *2019 6th International Conference on Systems and Informatics, ICSAI 2019*, (Icsai), 1466–1470. https://doi.org/10.1109/ICSAI48974.2019.9010223

[9] Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462–55470. https://doi.org/10.1109/ACCESS.2020.2981905

[10] Mustafid, M. (2013). Sistem business intelligence untuk mendukung perguruan tinggi yang kompetitif. *Jurnal Sistem Informasi Bisnis*, *3*(1), 18–24. https://doi.org/10.21456/vol3iss1pp18-24

[11] Nugroho, K. (2016). Model Analisis Prediksi Menggunakan Metode Fuzzy Time Series. *Infokam*, *1*, 46–50

[12] Pratama, D. (2013). Implementasi Data Warehouse dan Business Intelligence Untuk Memonitor Penjualan Sales Group : PT XYZ. *Implementasi Data Warehouse Dan Business Intelligence Untuk Memonitor Penjualan Sales Group : PT XYZ*, (Mdm)

[13] Qi, J., Yu, Y., Wang, L., & Liu, J. (2016). K∗-means: An effective and efficient k-means clustering algorithm. *Proceedings - 2016 IEEE International Conferences on Big Data and*

Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016, 242–249. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.46

[14] Singh, I., Sabitha, A. S. A. I., Bansal, A., & Cse, A. (2016). Student Performance Analysis Using Clustering Algorithm. *IEEE Access, 8*, 294–299. https://doi.org/10.1109/CONFLUENCE.2016.7508131

[15] Stephanie Pamela Adithama, Irya wisnubhadra, B. L. S. (2013). Analisis Dan Desain Real-Time Business Intelligence Menggunakan Change Data Capture. *Seminar Nasional Teknologi Informasi Dan Komunikasi 2013 (SENTIKA 2013)*, *2013*(Bisnis Intelligent), 1–9. https://doi.org/ISSN: 2089-9815

[16] Tibay, J. P., Ambat, S. C., & Lagman, A. C. (2019). School-based management performance efficiency modeling and profiling using data envelopment analysis and K-means clustering algorithm. *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*, 149–153. https://doi.org/10.1109/CCOMS.2019.8821732

[17] Widiyaningtyas, T., Prabowo, M. I. W., & Pratama, M. A. M. (2017). Implementation of k-means clustering method to distribution of high school teachers. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, *2017-December*(September), 19–21. https://doi.org/10.1109/EECSI.2017.8239083

[18] Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru. *Sistemasi*, *7*(3), 238. https://doi.org/10.32520/stmsi.v7i3.388

[19] A. Wibisono, M. Rofik, and E. Purwanto, "Penerapan Analisis Regresi Linier Berganda dalam Penyelesaian Skripsi Mahasiswa," *J. ABDINUS J. Pengabdi. Nusant.*, vol. 3, no. 1, p. 30, 2019, doi: 10.29407/ja.v3i1.13512.

[20] A. S. Nawangsih, Ismasari, "Implementasi Data Mining Untuk Memprediksi Barang Rusak Dengan Menggunakan Algoritma C4.5 Pada Perusahaan Pt Home Center Indonesia," *Progr. Stud. Tek. Inform. Sekol. Tinggi Teknol. Pelita Bangsa*, vol. 9, pp. 34–42, 2019.

[21] P. S. Ramadhan and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *Sains dan Komput.*, vol. 18, no. 1, pp. 55–61, 2019.

[22] M. A. Kasri and H. Jati, "Combination of K-Means and Simple Additive Weighting in Deciding Locations and Strategies of University Marketing," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 2, pp. 132–141, 2020, doi: 10.23917/khif.v6i2.11281.

[23] Suyanto, *Data Mining*. Yogyakarta: Informatika, 2017.

[24] P. Studi, S. Informasi, J. Sarjana, and A. T. Nopriansyah, "Fakultas ilmu komputer universitas sriwijaya mei 2018," 2018.

[25] Sugiyono, "Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016," *Fik*, vol. 1, no. 1, pp. 1–2, 2016.

[26] Ernawati, S. S. K. Baharin, F. Kasmin, and H. A. Purwanugraha, "Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study," *J. Syst. Manag. Sci.*, vol. 12, no. 3, pp. 156–180, 2022, doi: 10.33168/JSMS.2022.0309.

[27] F. Fengki, "Implementasi Regresi Linear Untuk Memprediksi Lama Waktu Pengiriman Catering Kepada Konsumen Studi Kasus Home Catering Malang," *Cent. Libr. Maulana Malik Ibrahim State Islam. Univ. Malang*, 2018.

[28] M. J. Meaney and T. Fikes, "The Promise of MOOCs Revisited? Demographics of Learners Preparing for University," *J. Learn. Anal.*, vol. 10, no. 1, pp. 113–132, 2023, doi: 10.18608/jla.2023.7807.

[29] M. Miftakhul and S. Prihandoko, “Penerapan Algoritma K-Means dan Cure Dalam Menganalisa Pola Perubahan Belanja Dari Retail ke E-Commerce,” vol. 7, no. 2, pp. 44–49, 2017.
[30] O. Villacampa, “(Weka - Thesis) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis,” *ProQuest Diss. Theses*, no. 63, p. 188, 2015.
[31] S. Haryati, A. Sudarsono, and E. (2015) Suryana, “Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5,” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.