Jurnal Ilmiah KURSOR Menuju Solusi Teknologi Informasi

Vol. 10, No. 3, Juli 2020

ISSN 0216 – 0544

e-ISSN 2301– 6914

IMPLEMENTATION OF K-MEDOIDS CLUSTERING FOR HIGH EDUCATION ACCREDITATION DATA

[a]Ghufron, [b]Bayu Surarso, [c]Rahmat Gernowo

[a]Magister of Information System, Post Graduate Program, Semarang, Indonesia
[b]Department of Mathematics, Diponegoro University, Semarang, Indonesia
[c]Department of Physics, Diponegoro University, Semarang, Indonesia
E-mail: ghufronasadly@gmail.com, bayusurarso@yahoo.com, gernowo@yahoo.com

Abstract

The need for data analysis in tertiary education every semester is needed, this is due to the increasingly large and uncontrolled data, on the other hand generally higher education does not yet have a data warehouse and big data analysis to maintain data quality at tertiary institutions is not easy, especially to estimate the results of university accreditation high, because the data continues to grow and is not controlled, the purpose of this study is to apply K-medoids clustering by applying the calculation of the weighting matrix of higher education accreditation with the data of the last 3 years namely length of study, average GPA, student and lecturer ratio and the number of lecturers according to the study program, so that it can predict accurate cluster results, the results of this study indicate that k-medoid clustering produces good cluster data results with an evaluation value of the Bouldin index davies cluster index of 0.407029478 and is said to be a good cluster result.

Key words: K-medoids, clustering, national higher education accreditation, davies bouldin index

INTRODUCTION

Accreditation is the determination of quality standards and assessments of an educational institution (higher education) by parties outside an independent institution in this case BAN-PT. Accreditation also means a government effort to standardize and guarantee quality at tertiary institutions so that quality among tertiary institutions is not too varied and according to the standards of the times, accreditation begins with the implementation of the relevant self-evaluation. The self-evaluation refers to the self-evaluation guidelines that have been issued by BAN-PT, however, if deemed necessary, the manager of the tertiary institution can add additional elements to be evaluated by the interests of the tertiary institution concerned. From the results of the implementation of the self-evaluation, an executive summary was made[1], which was subsequently enclosed in an executive letter of application for accreditation sent to the BAN-PT [2].

Nowadays data warehouse technology cannot handle the process of loading and analytic data into meaningful information for management. Big data technology must be implemented to expand existing data warehouse solutions[3].

This study aims to understand the assessment of basic education in the perspective of the State Reviewer as a mechanism that generates information regarding the positivity and weaknesses of a school or an educational system to provide improvements. For this reason, a Data Warehouse was created and later some analysis of the indicators were performed through clustering.The university has a large amount of data so that the academic data of the university has grown significantly and become big data. This data set is rich and continues to grow. University Management requires a tool to produce information from the information records generated [4] expected to support the top-level management decision-making process. Big data can be implemented with a data warehouse to support the decision-making process [5]

Information system data at higher educations each semester becomes very large data, data warehouse (dw) one of the containers for collecting a number of data in decision-making policies, which functions to store, analyze and visualize data effectively, extract transform load (ETL) processes) [6] which functions like data collection from several sources, cannot be updated in the data warehouse optimally, so ETL must be distributed in the data warehouse for an approach to data archive reporting and visualization [7].

K-medoids or Partitioning Around Method (PAM) is a non-hierarchical grouping method in this clustering method, which is a partition to group n sets of objects into a number of k clusters, K-medoids is applied to accurately identify candidate fire areas in the Fire Detection research for surveillance applications. video using K-medoid-based ICA color models and efficient Spatio-temporal visual features [8] and Zademehdi, 2019) and K-medoids applied according to previous studies are superior compared to K-means and FCM grouping, one of the advantages being that it can reduce noise in outliers[9].

In previous studies, the k-medoid algorithm was used to classify heterogeneous or different data, the k-medoid algorithm was also used for a fast way to predict big data, the K-medoids algorithm was used for clustering data in a forest fire area, the k-means algorithm comparison and k-medoid where K-medoids are superior to outliers. In this study, K-medoids clustering was applied to the college data warehouse in terms of the classification of study programs. This study aims to classify study programs to produce forecast accuracy, as in previous studies using K-medoids clustering.

Based on the explanation above the K-medoids clustering method in tertiary data for supporting accreditation is very important, while the system for processing such data does not yet exist, currently, data processing is still manual or only based on experience or not based on criteria set by tertiary institutions or competent body. This is certainly a mistake in making decisions and results in data quality at tertiary institutions which cannot be predicted early on, especially based on management information system data [10].

The aim of this research is to predict the data for accreditation purposes by taking data from existing data sources in the tertiary institution and then entering the data warehouse system, existing data in the data warehouse then calculated using the matrix of higher education accreditation calculations, namely data applicants to the number of applicants who pass the selection, the average study period of students for each program in the last 3 years, the ratio of the number of students to the number of permanent lecturers, the ratio of the number of permanent lecturers who meet the qualification requirements of lecturers for K-medoids clustering students is applied at Pancasakti University Tegal..

DATA MINING

Data mining is used to apply data mining efficiently from a collection of data, oversampling on data mining is used to avoid the negative effects of samples with lower classes on voltage prediction, data mining can also be used to predict from large data to be analyzed for example are k-nearest neighbors (KNN), Artificial Neural Network (ANN) and Random Forest (RF) [11].

INSTITUTIONAL ACCREDITATION ASSESSMENT MATRIX

The higher education accreditation assessment matrix has been prepared by BAN-PT to assess the quality of data at the tertiary institution according to the determined instruments and weights because the Higher Education and Data and Information Centers play an important role in ensuring the availability, availability, accuracy, validity, and reliability of data PD-Dikti which will be used by BAN-PT. The following is a table of accreditation assessment weights, the following is a table of matrices of institutional accreditation weights at tertiary institutions shown in tables 1 through 4

Table 1. Matrix of prospective new students

SCORE				
4	3	2	1	0
If the ratio is ≥ 3, then the score = 4.	If 1 < Ratio <3, then Score = 1 + Ratio.	If 1 < Ratio <3, then Score = 1 + Ratio.	If Ratio ≤ 1, Score = 2 x Ratio.	If Ratio ≤ 1, Score = 2 x Ratio.

Ratio = N_{Ai}/N_{Bi}

N_{Ai} = Number of prospective students taking part in the selection program. i = 1, 2, atau 7.

N_{Bi} = Number of prospective students who passed the selection in the main program. i = 1, 2, atau 7.

Tabel 2. Weighting average matrix of graduates

SCORE				
4	3	2	1	0
If P_{TWI} is 50%, then Scores = 4.	If P_{TWI} is 50%, then Scores = 4.	If P_{TWI} <50%, then $SCore_1$ = 1 + (6 x P_{TWI}).	If P_{TWI} <50%, then $SCore_1$ = 1 + (6 x P_{TWI}).	There is no score less than 1.

The percentage for the i-th education program is calculated using the following formula:

P_{TWI} = $f_i/d_i \times 100\%$

f_i = Number of students who graduate on time in the i-th educational program.

d_i = Number of students accepted in the class in the i-th education program.

The final score is calculated based on the weighted average calculation of the number of study programs in each education program.

Final score = $\sum(Score_i \times N_{pi})/\sum N_{pi}$

N_{pi} = number of study programs in the i-i education program, i = 1, 2, ..., 7

Tables 1 and 2 are the matrices of the calculation of the accreditation matrix of private tertiary institutions, namely the calculation of the weight of new students and the period of study of students. perhitungan dosen tetap dan mahasiswa pada tabel 3 dan 4 berikut

Table 3. Lecturer to Student Ratio

SCORE				
4	3	2	1	0
If 20 ≤ R_{MDT} ≤ 30, then scor e =4.	If R_{MDT} <20, then score = R_{MDT} / 5.	If 30 < R_{MDT} < 50 ,then score = 10 - (R_{MDT} / 5) .	If 30 < R_{MDT} < 50, then score = 10 - (R_{MDT} / 5) .	If R_{MDT} ≥ 50 , then score = 0.

$R_{MDT} = N_M / N_{DT}$

N_M = Number of students (regular and transfer) in the main program at the time of TS.

N_{DT} = Number of permanent lecturers.

Table 4. The number of lecturers qualified

SCORE				
4	3	2	1	0
If $R_{DPS} \geq 12$, then score = 4	If $6 \leq R_{DPS} < 12$,then Score = R_{DPS} / 3	If $6 \leq R_{DPS} < 12$,then Score = R_{DPS} / 3	If $R_{DPS} <$ 6,hence tertiary institutions are not accredited.	If $R_{DPS} <$ 6,hence tertiary institutions are not accredited.

Note: Lecturer data remains listed on the PD DIKTI page. If there is a study program that does not meet the minimum number of lecturers (the number of lecturers is less than 6), the tertiary institution is not accredited.

$R_{DPS} = N_{DT} / N_{PS}$

N_{DT} = Number of permanent lecturers.

N_{PS} = Number of study programs.

In table 3 above is the weight calculation of the ratio of students and lecturers while in table 4 is the calculation of the number of lecturers who meet the qualifications where the minimum number of lecturers in the study program is at least 6 lecturers.

K-MEDOIDS CLUSTERING

The problem of grouping, according to attributes, has been widely studied because of its application in various fields such as machine learning, data mining and discovery of pattern recognition knowledge and pattern classification. homogeneous groups so that the patterns in each group are similar. Several unsupervised learning algorithms have been proposed to partition a group of objects into a number of groups according to the optimization criteria. One of the most popular and widely studied grouping methods is k-means algorithmically simple, relatively powerful, and gives "good enough" answers to various data sets. [12].

Clustering Analysis is the process of dividing a set of objects into a set of non-overlapping parts. Each subset is a cluster, such that the objects in the cluster are similar to each other and are different from the objects in the other cluster. Most of the algorithms in the grouping partitioning approach suffer from trapped in optimal locality and sensitivity for initialization and outliers [13]similarity or object equation by means of distance calculation [14] The steps of the K-medoids algorithm is as follows[15][16]

1. Initialize the cluster center by k (number of clusters)
2. Allocate each data (object) to the nearest cluster using Euclidian Distance equation with the equation below :

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

Annotation:

d = the distance between x and y
x = cluster data center
y = data on the attribute
i = every data
n = amount of data,
x_i = data in the center of cluster i
y_i = data in each data to i

3. Assign each instance to the nearest medoid x
4. Calculate the objective function
5. The sum of dissimilarities of all instances to their nearest medoids
6. Randomly select an instance y
7. Swap x by y if the swap reduces the objective function
8. Repeat (3-6) until no change whereas the k-medoid algorithm flow chart process is shown in Figure 1 the following flow chart.

To evaluate the clustering algorithm process, especially in this study using the K-medoids algorithm, clustering evaluation in this study uses Davies Bouldin index for the davies Bouldin index formula and silhoutee index coeficient explained in the following sub

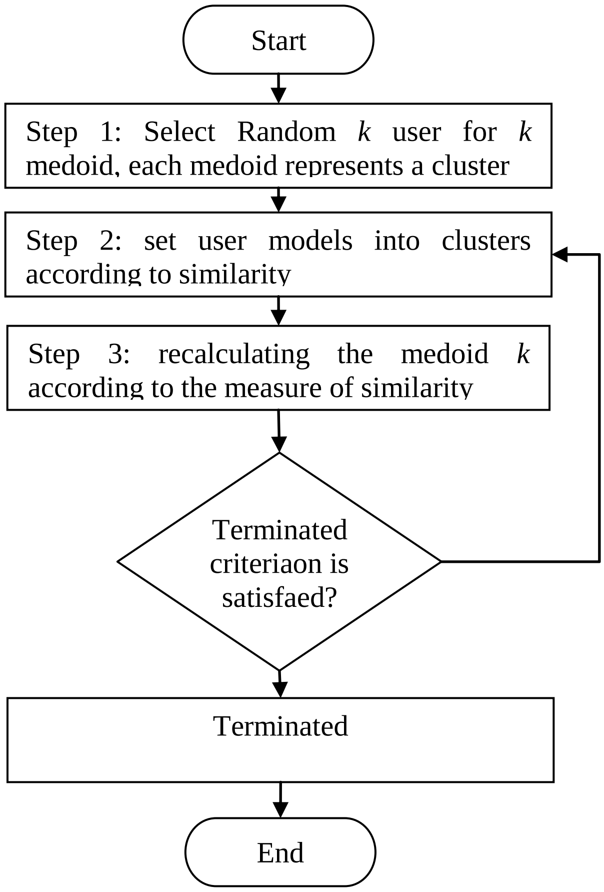

Fig 1. Flowchart K-medoids process

DAVIES BOULDIN INDEX

The evaluation of clustering in this study will use clustering evaluation To evaluate the quality of grouping the authors consider the Davies Bouldin Index (DBI). Cluster evaluation using the boudin index davies uses an internal evaluation scheme in which the cluster results can be seen whether the quantity and proximity of the cluster data result [17]. Davies Bouldin's criteria are based on the ratio in clusters and the distance between clusters. In the K-medoid formulation, the cohesiveness of the corresponding clusters and the separation between them is the main parameter that distinguishes one cluster from another. The Davies-Bouldin Index is one measure in cluster evaluations [18].

$$\mathrm{var}(x) = \frac{1}{N-1}\sum_{i-1}^{N}(x_i - \bar{x})^2 \quad (2)$$

Where $\bar{x}$: the average of the clusters x and N is the number of cluster members

$$DBI = \frac{1}{k} \bullet \sum_{i=1}^{k} R_i \quad (3)$$

From this equation, k is the number of clusters, the smaller the davies Bouldin index value obtained (non-negative> = 0), the better the clusters obtained from clustering using the clustering algorithm.

$$R_{i=j=1\ldots k,1\neq j}^{\max R_{ij}} \quad (4)$$

$$\begin{matrix} R_{i,j} \\ i \neq j \end{matrix} = \frac{\mathrm{var}(c_i) + \mathrm{var}(c_j)}{\left\| c_i - c_j \right\|} \quad (5)$$

Research Methods

In this study it is proposed by the K-medoids clustering method, previously the data was taken from the information system data in the tertiary institutions and then the data was extracted and entered into a data warehouse for analysis [19].

SILHOUETTE COEFFICIENT

The silhouette coefficient is an evaluation method to test the optimal or accuracy of a cluster that has been formed from the clustering process[20]. The silhouette coefficient tests the quality of groupings based on dispersion in groups and the distance between them, but also provides the best information, from estimates between samples from different groups that are not taken into account. Thus, the Silhouette coefficient is also used to measure how close each point is in a cluster to the points in a neighboring group and measure the results of cluster group validation [21], [22]

$$a(i) = \frac{1}{|A|-1}\sum j \in_{A,j\neq 1} d(i,j) \quad (6)$$

Where :

$a(i)$ = Difference in average of object (i) to all other objects in A

$d(i, j)$ = distance between data i and j

Then look for differences in the average of objects with equation 7

$$d(i,C) = \frac{1}{|A|}\sum j \in cd(i,j) \quad (7)$$

Explanation:

$d(i,C)$ = difference in average of objects (i) to all other objects in C
C = cluster other than cluster A or cluster C is not the same as cluster A.
Then find the minimum value of b (i) in equation 8 below:

$$b(i) = \min_{c \neq A} d(i, j) \quad (8)$$

Cluster B that reaches a minimum d(i,B) = $b(i)$ is called the neighbor of the object (i). This is the second-best cluster for the object (i).
Then calculate the silhouette index with the following 9 equation

$$S(i) = \frac{(b(i) - a(i))}{\max a(i), b(i)} \quad (9)$$

RESULT AND DISCUSSION

College data are taken from the existing information system at the university that is at the University Pancasakti Tegal then the data is extracted data into the data warehouse after inside the data warehouse the data is calculated according to the college accreditation matrix 9 criteria using the calculation matrix weighting of accreditation Higher education in the accreditation matrix table tables 1 to 4, this study uses the CakePHP framework and SQL server enterprise as a database and data warehouse, while data from the results of the data mining process from the college data warehouse are presented in tables 5 to 9 below:

Table 5. Number of Prospective Student

Data	Study Program Code	Amount
1	21201	310
2	21401	22
3	22201	251
4	26201	290
5	26401	15
6	54243	80
7	54246	206
8	61201	1545
9	61403	60
10	62201	856
11	65201	300
12	70201	145
13	74101	132
14	74201	832
15	84202	181
16	84206	91
17	86201	248
18	87203	144
19	87205	63
20	88201	254
21	88203	206

Table 6. Number of Students Table

Data	Study Program Code	Amount
1	21201	300
2	21401	22
3	22201	250
4	26201	110
5	26401	4
6	54243	62
7	54246	55
8	61201	1148
9	61403	57
10	62201	849
11	65201	280
12	70201	126
13	74101	122
14	74201	810
15	84202	181
16	84206	91
17	86201	248
18	87203	144
19	87205	62
20	88201	248
21	88203	203

Table 7. Average students' study period

#	Study Program Code	Amount
1	21201	4
2	21401	3
3	22201	4
4	26201	4
5	26401	3
6	54243	4
7	54246	4

8	61201	4
9	61403	3
10	62201	4
11	65201	4
12	70201	3
13	74101	2
14	74201	3
15	84202	4
16	84206	4
17	86201	4
18	87203	4
19	87205	4
20	88201	4
21	88203	4

Table 8. Number of lecturers in the study program

#	Study Program Code	Amount
1	21201	11
2	21401	4
3	22201	6
4	26201	8
5	26401	4
6	54243	10
7	54246	8
8	61201	6
9	61403	8
10	62201	17
11	65201	9
12	70201	7
13	74101	6
14	74201	24
15	84202	18
16	84206	8
17	86201	21
18	87203	6
19	87205	12
20	88201	15
21	88203	21

Table 9. Number of lecturer qualifications

#	Study Program Code	Amount
1	21201	10
2	21401	4
3	22201	5
4	26201	8
5	26401	4
6	54243	7
7	54246	6
8	61201	14
9	61403	8
10	62201	13
11	65201	7
12	70201	7
13	74101	3
14	74201	22
15	84202	14
16	84206	6
17	86201	14
18	87203	9
19	87205	6
20	88201	12
21	88203	18

Tables 5 and 9 above are the results of the data extraction process from existing academic and academic information systems in colleges, Data that has been mined are then entered into the data warehouse, after that the data is weighted with the university accreditation matrix after the data has a value according to the weight of the accreditation matrix then the college data cluster is performed using the k-medoid clustering algorithm, then to test the data clusters used davies Bouldin index. In the graph below, there is a significant difference between the number of study program students and other study programs.

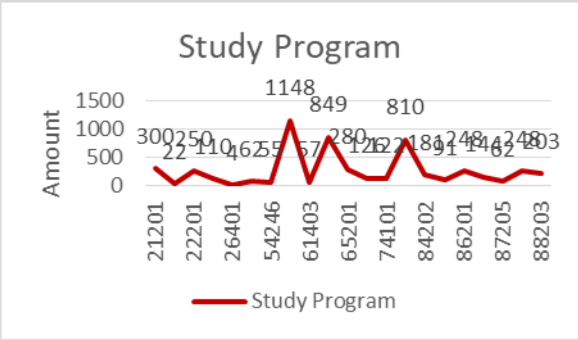

Fig 2. The difference in the number of students

The following table describes the results of the clustering process with the K-medoids algorithm. In the initial table using the calculation process in the formula and the first step number the results of the first iteration calculation are shown in the following table 10

Table 10. Distance (3,2,2,2) and (3,2,4,4)

K	A	B	C	D	C 1	C 2
Engineering	3	2	3	2	V	
D3 Engineering	3	4	4	4		V
Civil Engineering	3	2	2	2	V	
Industrial Engineering	3	2	4	4		V
D3 Industrial Engineering	3	4	4	4		V
Aquaculture	3	2	4	4		V
Fisheries Resources	3	2	4	4		V
Management	3	2	1	1	V	
D3 Tax Management	3	4	4	4		V
Accounting	3	2	2	1	V	
Public administration	3	2	2	2	V	
Communication Studies	3	4	4	4		V
Master in Law	3	4	3	2	V	
Legal studies	3	4	2	2	V	
Mathematics education	3	2	4	4		V
Ipa Education	3	2	4	4		V
Guidance and counseling	3	2	4	4		V
Economic Education	3	2	3	4		V
Pancasila Education	3	2	4	4		V
Indonesian Education	3	2	4	3		V
English language	3	2	4	4		V

Then in the second initialization, the next stage is calculated by randomly selecting distance, for the results of calculations in the second iteration shown in table 11 below:

Table 11. Distance (3,2,2,2) and (3,4,2,2)

K	A	B	C	D	C 1	C 2
Engineering	3	2	3	2	V	
D3 Engineering	3	4	4	4		V
Civil Engineering	3	2	2	2	V	
Industrial Engineering	3	2	4	4	V	
D3 Industrial Engineering	3	4	4	4		V
Aquaculture	3	2	4	4	V	
Fisheries Resources	3	2	4	4	V	
Management	3	2	1	1	V	
D3 Tax Management	3	4	4	4		V
Accounting	3	2	2	1	V	
Public administration	3	2	2	2	V	
Communication Studies	3	4	4	4		V
Master in Law	3	4	3	2		V
Legal studies	3	4	2	2		V
Mathematics education	3	2	4	4	V	
Ipa Education	3	2	4	4	V	
Guidance and counseling	3	2	4	4	V	
Economic Education	3	2	3	4	V	
Pancasila Education	3	2	4	4	V	
Indonesian Education	3	2	4	3	V	
English language	3	2	4	4	V	

Because the first and second iteration processes have exceeded one, the clustering process is stopped. The author does not write in length the results of the medoids and non-medoids data clusters for the results of the cluster testing can be seen in table 12. To prove that the cluster is a good cluster, the researchers conducted clustering using the Bouldin index davies, in the first iteration table the number of clusters 1 was 7 data while the second cluster is 14 clusters. To calculate DBI, you must first calculate the ratio to find out the ratio of each of the previous cluster variants following the DBI calculation sample:

Number of clusters 1 = 7
The amount of data in cluster 1 = 9
Number of clusters 2 = 14
The amount of data in cluster 2 = 10
Cluster average 1 = 9/7 = 1.285714286
Cluster average 2 =10/14= 0.714285714

Cluster variant 1 = 1.06122449
Cluster variant 2 =0.775510204
Ratio = (1.06122449 + 0.775510204) / 9 = 0.459183673
DBI = 1/2 (0.459183673 + 0.459183673) = 0.459183673

Above is a sample calculation test of the Bouldin index davies, for the test results table with DBI and silhouette index coefficient, it can be shown by the following table 11.

Table 12. Results of DBI and SC

Data	Distance	Distance	DBI	SC
1	Distance (3,2-2,2)	Distance (3,2,4,4)	0,459183673	0,579614
2	Distance (3,2,2,2)	Distance (3,4,2,2)	2,549444444	0,443476
3	Distance (3,2-4,4)	Distance (3,4,2,2)	0,407029478	0,393103
4	Distance (3,4,4,4)	Distance (3,4,2,2)	0,756405817	0,105573

On the test results in Table 12 shows the results of distance calculation and evaluation using the DBI best score 0,407029478 and silhouette index coefficients best score 0,579614 showed the best results with the same cluster results at a distance of 3,2-2,2 and distance 3,2-4,4 with the results of the first cluster namely Engineering, Civil Engineering, Management, Accounting, Public administration, Masters in Law And Legal Studies and the results of the second cluster namely: D3 Engineering, Industrial Engineering, D3 Industrial Engineering, Aquaculture, Fisheries Resources, D3 Tax Management, Communication Studies, Mathematics Education, IPA Education, Guidance and Counseling, Economics Education, Pancasila Education, Indonesian Education, English Education while the evaluation graph of clustering davies Bouldin index and silhouette index coefficient are indicated by the following graphic figure 3:

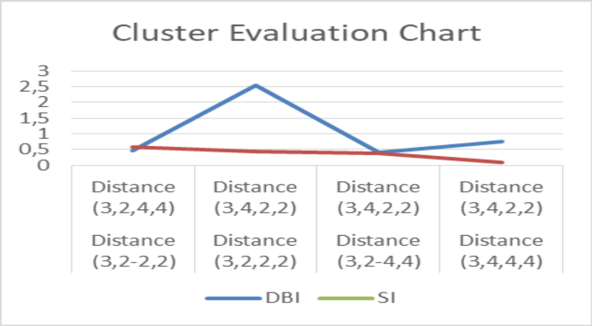

Fig 3. Clustering evaluation chart

CONCLUSION

In this research, the K-medoids algorithm method can not only be applied in testing large datasets for fire detection in forests, but this study applies the K-medoids algorithm and the combination of the weight calculation of university accreditation matrices from the college data mining process for the classification of study programs with the K-medoids clustering algorithm, based on the results of research with the K-medoids algorithm regarding data classification with the K-medoids method, it can be concluded that the K-medoids classification implemented is included in a good cluster category, namely by using a bouldin index davies test with a cluster evaluation value of 0, 459183673 and also testing with Silhouette index Coefficient 0.579614389 and said to be a good cluster result.

ACKNOWLEDGMENT

The authors thank the University of Pancasakti Tegal for the data used for the authors to complete this scientific article.

REFERENCES

[1] P. A. Dover, “Creating learning solutions for executive education programs,” Int. J. Manag. Educ., p. 12, 2018.
[2] BAN-PT, “Peraturan BAN-PT No.59Th 2018 Instrumen APT,” Oct. 2018.
[3] T. G. Ramos, J. C. F. Machado, and B. P. V. Cordeiro, “Primary Education Evaluation in Brazil Using Big Data and Cluster Analysis,” Procedia Comput. Sci., vol. 55, pp. 1031–1039, 2015, doi: 10.1016/j.procs.2015.07.061.
[4] W. Jemaa, P. J. Carew, and L. Stapleton, “Implementing Effective Information Control Systems for Primary Education Support in Developing Countries: An Investigative Field Study of a Voluntary Non-Profit Organisation in Tunisia,” IFAC-Pap.,

vol. 51, no. 30, pp. 171–176, 2018, doi: 10.1016/j.ifacol.2018.11.281.

[5] L. W. Santoso and Yulia, “Data Warehouse with Big Data Technology for Higher Education,” Procedia Comput. Sci., vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.

[6] Z. Ouaret, D. Boukraa, O. Boussaid, and R. Chalal, “AuMixDw: Towards an automated hybrid approach for building XML data warehouses,” Data Knowl. Eng., vol. 120, no. January, pp. 60–82, 2019, doi: 10.1016/j.datak.2019.01.004.

[7] B. Boulekrouche, N. Jabeur, and Z. Alimazighi, “An Intelligent ETL Grid-Based Solution to Enable Spatial Data Warehouse Deployment in Cyber Physical System Context,” Procedia - Procedia Comput. Sci., vol. 56, no. MobiSPC, pp. 111–118, 2015, doi: 10.1016/j.procs.2015.07.176.

[8] M. Hashemzadeh and A. Zademehdi, “Fire detection for video surveillance applications using ICA K-medoids-based color model and efficient spatio-temporal visual features,” Expert Syst. Appl., vol. 130, pp. 60–78, Sep. 2019, doi: 10.1016/j.eswa.2019.04.019.A.

[9] Khatami, S. Mirghasemi, A. Khosravi, C. P. Lim, and S. Nahavandi, “A new PSO-based approach to fire flame detection using K-Medoids clustering,” Expert Syst. Appl., vol. 68, pp. 69–80, Feb. 2017, doi: 10.1016/j.eswa.2016.09.021.

[10] E. Turban, J. E. Aronson, and T.-P. Liang, “Decision Support Systems and Business Intelligence,” Decis. Support Bus. Intell. Syst. 7E, 2007, doi: 10.1017/CBO9781107415324.004.

[11] W. Hadi, N. El-Khalili, M. AlNashashibi, G. Issa, and A. A. AlBanna, “Application of data mining algorithms for improving stress prediction of automobile drivers: A case study in Jordan,” Comput. Biol. Med., vol. 114, no. July, p. 103474, 2019, doi: 10.1016/j.compbiomed.2019.103474.

[12] P. Prihandoko and B. Bertalya, “a Data Analysis of the Impact of Natural Disaster Using K-Means Clustering Algorithm,” Kursor, vol. 8, no. 4, p. 169, 2017, doi: 10.28961/kursor.v8i4.109.

[13] S. M. Razavi Zadegan, M. Mirzaie, and F. Sadoughi, “Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets,” Knowl.-Based Syst., vol. 39, pp. 133–143, 2013, doi: 10.1016/j.knosys.2012.10.012.

[14] D. Sun, H. Fei, and Q. Li, “A Bisecting K-Medoids clustering Algorithm Based on Cloud Model,” IFAC-Pap., vol. 51, no. 11, pp. 308–315, 2018, doi: 10.1016/j.ifacol.2018.08.301.

[15] J. Han, M. Kamber, and Jian Pei, Data Mining Concepts and Techniques, 2nd ed. 2012.

[16] I. C. Dewi, B. Y. Gautama, and P. A. Mertasana, “Analysis of Clustering for Grouping of Productive Industry by K-Medoid Method,” Int. J. Eng. Emerg. Technol., vol. 2, no. 1, p. 26, 2017, doi: 10.24843/ijeet.2017.v02.i01.p06.

[17] A. Bates and J. Kalita, “Counting clusters in twitter posts,” ACM Int. Conf. Proceeding Ser., vol. 04-05-Marc, 2016, doi: 10.1145/2905055.2905295.

[18] S. Harikumar and P. V. Surya, “K-Medoid Clustering for Heterogeneous DataSets,” Procedia Comput. Sci., vol. 70, pp. 226–237, 2015, doi: 10.1016/j.procs.2015.10.077.

[19] S. O. Salinas, “DATA WAREHOUSE AND BIG DATA INTEGRATION Sonia,” Int. J. Comput. Sci. Inf. Technol., vol. 9, no. 2, pp. 1–17, 2017, doi: 10.5121/ijcsit.2017.9201.

[20] D. F. Pramesti, M. T. Furqon, and C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot),” p. 10.

[21] C. M. R. do Carmo and T. H. Christensen, “Cluster analysis of residential heat load profiles and the role of technical and household characteristics,” Energy Build., vol. 125, pp. 171–180, Aug. 2016, doi: 10.1016/j.enbuild.2016.04.079.

[22] Y. Feng, R. C. Affonso, and M. Zolghadri, “Analysis of bike sharing system by clustering: the Vélib’ case,” IFAC-Pap., vol. 50, no. 1, pp. 12422–12427, Jul. 2017, doi: 10.1016/j.ifacol.2017.08.2430.