Jurnal Varian
Vol. 8, No. 1, October 2024, pp 11–24
e-ISSN: 2581-2017

Clustering of Study Program Using Block-Based K-Medoids

Asa Nugrahaini Itsnal Muna, Kariyam
Universitas Islam Indonesia, Yogyakarta, Indonesia

Article Info

Article history:

Received : 07-26-2023
Revised : 06-06-2024
Accepted : 07-26-2024

Keywords:

Block-Based KM;
Clustering;
DRIM;
MANOVA;
Study program.

ABSTRACT

The purpose of this research is to classify Study Programs based on eleven mixed data from Internal Quality Management System (QMS) indicators. This grouping can provide a clearer picture of how QMS affects the performance and quality of study programs. By understanding these clusters, universities can identify and design more effective strategies to improve the quality of education. The data used comes from the National Accreditation Board for Higher Education (BAN-PT) and the website database, which consists of seven numerical variables: number of lecturers, percentage of doctors, percentage of professors and associate professors, student enumeration, percentage of graduates, program experience, and availability of laboratories. Meanwhile, the categorical variable consists of four variables: National Accreditation Board of Higher Education (BAN-PT) research ranking, accreditation, international recognition, and level of community service. The clustering method used is the block-based k-medoids (block-based KM), and multivariate analysis of variance (MANOVA). We applied the Deviation Ratio Index based on K-Medoids (DRIM) to determine the number of clusters. This research results that the optimal number of groups that must be formed is three. Based on MANOVA the results showed that the group consisting of 12 study programs had better QMS outcomes than the other two groups.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i1.3181

Corresponding Author:

Kariyam,
Department of Statistics, Universitas Islam Indonesia,
Email: kariyam@uii.ac.id

How to Cite:
Muna, A. N. I., & Kariyam, K. (2024). Clustering of Study Program Using Block-Based K-Medoids. *Jurnal Varian*, *8*(1), 11-24.

A. INTRODUCTION

The development of the times provides an impetus for change in all aspects, including changes in higher education in management (Ab Wahid & Grigg, 2021; Cardoso et al., 2019). College management is an effort to keep up with the times. Standardization is used as a reference for all management carried out by universities. All universities must meet the standardization that has been set in order to maintain their existence. Higher education is the next level of education from secondary education to formal education. The higher education organizing unit is called a college. Every college is required to prepare itself to compete with others. Increasingly fierce competition will certainly spur universities to improve the quality of each of them (DeSimone & Rich, 2020; Halibas et al., 2020; Hauptman Komotar, 2020; Zuhairi et al., 2020).

The low quality of universities in Indonesia can be seen based on the results of the accreditation of universities and study programs. National Accreditation Board for Higher Education (BAN-PT, 2023) reported that 287 out of 28.540 study programs in Indonesia are not accredited. While in Indonesia, there were 2.154 or 7.55% of study programs accredited as Excellent, and the composition of the percentage value of the "A" rating in the study program reached 10.8%. This information proves that the quality of study programs at universities in Indonesia must improve. The standardization set by the National Accreditation Board of Higher Education (BAN-PT) must be implemented to improve the quality of higher education.

Higher education standardization is expected to improve the quality of higher education so that the quality assurance pattern is not only carried out externally. However, it must also be carried out autonomously by universities following the Regulation of the Minister of Research, Technology and Higher Education Number 53 of 2023 concerning the Quality Assurance of Higher Education. The quality of higher education is the level of conformity between the implementation of higher education and educational standards consisting of National Higher Education Standards and standards set by higher education institutions. While accreditation or ranking is a component used to see how far the college is recognized as a credible educational institution. Meanwhile, study program accreditation is an assessment to determine the feasibility of a study program.

Cluster analysis is a term applied to analysis that partitions a set of objects into several homogeneous groups when there is no a priori information about the data structure (Lahouaoui et al., 2022; Randriamihamison et al., 2021; Solimun & Fernades, 2022). Their research uses the clustering partitioning method to group a set of n objects into k clusters. The k-medoids algorithm uses objects in a collection of objects to represent a cluster (Sureja et al., 2022). The object chosen to represent is called a medoid. Clusters are obtained by calculating the closest distance between medoids and non-medoids. K-Medoids is an alternative algorithm to k-means that can be used for mixed data analysis. A simple and fast k-medoids is like k-means, which is an algorithm that calculates the distance matrix and uses it to get a new medoid at each iteration. The study found that the simple and fast k-medoids (SFKM) algorithm performed better than k-means and was more time efficient than the PAM algorithm. Budiaji & Leisch (2019) researched SKM (Simple K-Medoids), an algorithm that improves SFKM using Gower, Wishart, Podani, Huang, and Harikumar-PV distance calculations. Researchers used real and artificial data sets, using initial medoid initialization with the seeding parameter (s) while giving similar results to PAM. The research shows the SKM algorithm has better results than SFKM. K. Kariyam et al. (2022) proposed a new research algorithm called FKM (Flexible K-Medoids Partitioning Method). The procedure is divided into two phases: selecting the initial medoids and determining the partitioned data set. The initial medoid is selected based on the block representation of the combination of the sum of values and standard deviation. The relative positions of the objects will be separated when the sum of the variable values is different, even though the objects have the same variance. Objects are flexibly selected from each block as initial medoids to build initial groups. The FKM algorithm overcomes empty groups that appear in SFKM and overcomes identical objects in different initial groups in SKM. K-Medoids also have shortcomings in its unpredictable iteration process, so the computation takes a long time. Therefore, Schubert & Rousseeuw was developed a new method, Fast and Eager K-Medoids Clustering, to reduce the running time of algorithms. At the same time, the Block-Based K-Medoids (Block-KM) Partitioning Method was investigated as a refinement of the k-medoids algorithm, which aims to reduce the iteration process (Kariyam et al., 2022). The Block-KM algorithms update the set of medoid based on the object, which minimizes the average distance to other group members in the group. The result of the study found that the Block-KM algorithm is more efficient in reducing the number of iterations than SKM and FKM. (Kariyam et al., 2023) also developed the Deviation Ratio Index based on K-Medoids (DRIM) to determine the number of clusters in datasets. The DRIM indicator is calculated based on a distance matrix so that this index is flexible for any type of data scale.

This research proposes a new approach to clustering Statistics study programs by utilizing output indicators from implementing the Quality Management System (QMS). To the researcher's knowledge, this topic has never been studied. Meanwhile, in the era of big data, statistical methods are increasingly widespread. Therefore, researchers are interested in discussing the profile of Statistics study programs based on the outcomes of QMS implementation. The novelty of this research also lies in using mixed data (categorical and numerical) from academic and non-academic performance indicators as a basis for evaluating and grouping study programs. This research also uses a relatively new method: the object block-based k-medoids partitioning method (Block-based KM). The Block-KM algorithm is a development of the k-medoids method, which uses representations of object blocks with the same standard deviation and the same amount of data on variables. This method guarantees that if the initial medoid is an identical object, then the block identical objects will only be represented by one object and the identical object will occupy the same initial and final group. Another difference with existing research is using the new DRIM method to determine the optimal number of groups. The DRIM method is derived based on homogeneity within groups and heterogeneity between groups, which is calculated based on the distance of the object to the group medoid containing the object. The DRIM method, known for overcoming mixed data problems and increasing estimation accuracy, is applied to group Statistics study programs into clusters based on QMS output indicators. The use of the DRIM method in this context is an innovation that is expected to provide more reliable and precise results. This study also aims to profile groups using multivariate analysis of variance. The benefit that prospective students obtain from the results of this research is that it can be used as a consideration in selecting a study program. Meanwhile, the benefit for study program managers is knowing the relative position of the study program and getting information about universities that might be used as a reference in administering QMS.

B. RESEARCH METHOD

The flow diagram of this research is shown in Figure 1. Based on the study literature, we collecting multivariate data about the characteristics of the Statistics Study Program in Indonesia based on the Quality Management System outputs. We combined cluster analysis and Multivariate Analysis of Variance (MANOVA) to analyze the profile of the Statistics Study Program group. Both methods are part of multivariate analysis, a type of statistical analysis used to analyze data consisting of two or more variables. Multivariate analysis is used because, in reality, the problems that occur cannot be solved by simply connecting two variables or looking at the effect of one variable on another, including in this case study. Before classifying the Statistics Study Program, we conduct a pre-processing data cover check of missing values and outliers and choose the suitable transformation method. Next, we determine the number of clusters using a new approach, namely the medoid-based deviation ratio index. According to the optimal number of clusters, we classify the dataset using block-based k-medoids and profiling each group based on MANOVA. Finally, we cross the results with other clustering methods and use an appropriate way to analyze the data and interpret the results.

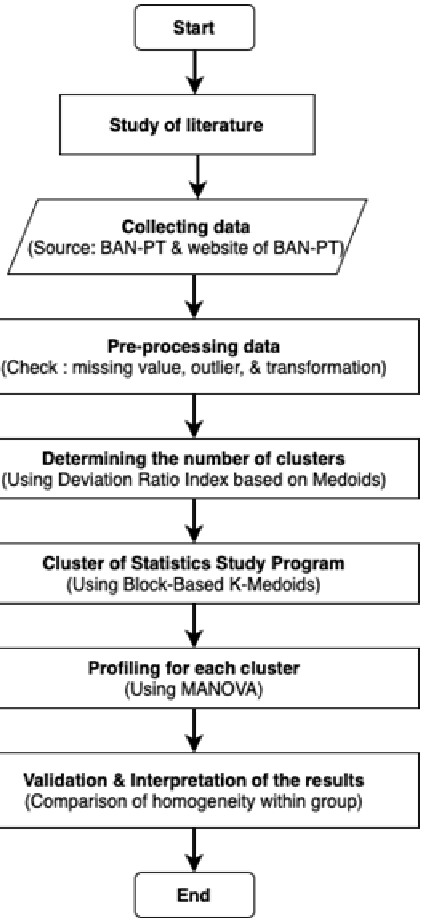


Figure 1. The flow diagram of research

1. Real Datasets

The population used in this study is represented by universities in Indonesia that organize undergraduate statistics education. The sample used is data from 39 universities in Indonesia that organize Undergraduate Statistics education. The data is secondary data taken from the BAN-PT website via the address https://banpt.or.id/direktori/data_borang/data_borang.php and the PD Dikti page, with the address https://pddikti.kemdikbud.go.id. Data was collected from December 1, 2022, to January 31, 2023. The study used 11 variables, consisting of 7 numerical type variables and 4 categorical type variables A more complete explanation of the research variables is contained in Table 1. The reason for selecting these variables is to determine the reputation of the study program based on outcomes of the Quality Management System, especially in the academic and non-academic fields. The academic areas in question are education, research, and community service. Apart from that, non-academic fields include human resources, finance, facilities, management, etc.

Table 1. Research Variables

Variable	Information	Types	Category/Unit
X_1	Lecturer enumeration	Numerical	People
X_2	Percentage of S3 Lecturers	Numerical	Percent
X_3	Percentage of professors and associate professors	Numerical	Percent
X_4	Student enumeration	Numerical	People
X_5	Percentage of graduation	Numerical	Percent
X_6	Higher Education Research Cluster	Categorical (Ordinal)	1: Built; 2: Middle 3: Main; 4: Independent
X_7	Study Program Accreditation	Categorical (Ordinal)	1: Good; 2: B; 3: Very good; 4: A; 5: Excellent
X_8	International recognition	Binary	0: No; 1: Yes
X_9	Experience of Program	Numerical	Year
X_10	Laboratory availability	Numerical	Space
X_11	Community service cluster	Categorical (Ordinal)	1: Very good; 2: Excellent

2. Cluster Analysis

Cluster analysis is an unsupervised learning technique that does not make assumptions about the number of groups or class structure. The primary purpose of clustering is to find cluster objects that show a high similarity while objects belonging to different groups have significant differences.

a. Data Transformation

One of the critical stages in data analysis is preprocessing. Data preprocessing is cleaning data from noise or converting data into a more straightforward format. This stage aims to improve the results of a method. One of the data preprocessing techniques is transformation to standardize the data. Transformation steps for ordinal, interval, and ratio scaled data are shown below.

(i) Ranking n objects in the th variable l (two same values have the same order), i.e. $x_{1l} \leq x_{2l} \leq \ ... \leq x_{nl}$ becomes $r_{1l} \leq r_{2l} \leq \ ... \leq r_{nl}$.

(ii) Transformation to interval $[0, f]$ such in the Equation 1,

$$z_{li} = f \ . \left(\frac{r_{li} - r_{l1}}{r_{lm} - r - l1} \right) ; \ i = 1, 2, \ldots, n. \quad (1)$$

The value of r_{li} is ranking data for variable number l on the th object i. Meanwhile, the r_{l1} is the smallest rank for the variable l, and conversely, the r_{lm} is the highest rank for the variable l. The value of f shows the standardization weight for transformation.

b. Simple Matching Coefficient

Closeness measures on both binary and categorical variables use a simple matching coefficient. Suppose two objects i and j are observed in p discrete random variables that have binary and categorical data types, respectively, the variables are categorized by 0 (zero) and 1 (one). The values of a and d denote the frequency of the same data (matches), i.e. objects i and j have data categorized as 0 (zero) as much as a and data categorized as 1 (one) as many as d. Meanwhile, the values of b and c indicate the frequency of mismatches. In simple terms, if the frequencies a and d are summed up, the result obtained is close to the number of variables, then the objects of i and j are more similar. If the value of $a + d = p$, then the object i and j are identical (Yuan et al., 2020).

c. Manhattan Distance

Manhattan distance is used because, overall, the data used in this research is mixed data consisting of categorical (binary and ordinal) and numerical (interval and ratio). The distance measure using Manhattan as in Equation 2.

$$d_{ij} = \sum_{l=1}^{p} |x_{il} - x_{jl}| \tag{2}$$

with x_{il} is the value of the object i in the numerical variable lth, and x_{jl} is the value of the object j in numeric variable lth.

3. Block-Based K-Medoids Partitioning Method

Block-based K-Medoids Partitioning Method (Block-KM) was developed to improve the SFKM (Simple and Fast K-Medoids) algorithm that allows empty clusters. Meanwhile, the SKM (Simple K-Medoids) algorithm simplifies the SFKM algorithm by selecting an initial medoid on the object located at the center. In contrast, the initial medoid members are determined randomly based on iteration. This process causes the similarity of objects in different groups. Another problem with the k-medoids algorithm is that the initial medoid selection is done randomly. In addition, the computational process takes a long time for large data because the iterations performed cannot be predicted. Block-KM analysis consists of two stages: selecting initial representative objects and obtaining data set partitions.

Stage 1: Selecting an initial representative object

1-1 For each object, i, $(i = 1, 2, \ldots, n)$, two parameters are calculated according to the standard deviation as in Equation 3, and sum values as in Equation 4,

$$u_i = \sqrt{\frac{\sum_{l=1}^{p} (x_{il} - \underline{x}_i)^2}{p-1}} \tag{3}$$

where $\underline{x}_i = w_i/p$ with w_i, is the sum of p-variables as Equation 4,

$$w_i = \sum_{l=1}^{p} x_{il} \tag{4}$$

with $i = 1, 2, \ldots, n$ and $l = 1, 2, \ldots, p$. These parameters are used as a reference to select the initial medoid.

1-2 Arrange all objects based on Equation 3 in ascending order, then for each block with the same standard deviation (if any), the objects are re-sorted based on Equation 4 also in ascending order.

1-3 For k the first block of the combination u_i and w_i (or maybe just the u_i). Select the first object from each block as the initial medoid.

1-4 Determine the members of k initial group based on the object's distance to the nearest medoid.

Stage 2: Obtain data set partitions

2-1 Update the medoid in each group formed based on the object that minimizes the average distance to other group members in the group. The calculation of the average distance is by Equation 5.

$$\underline{D}_i = \frac{1}{n_k} \sum_{j=1}^{p} d_{ij} \tag{5}$$

2-2 Determine the cluster by marking each object to the nearest medoid and calculate the sum deviation value within the group, $SDW(k)$, such as Equation 6 with m_i is the medoid of the object group x_i,

$$SDW\,(k) = \sum_{i=1}^{n} d_{(x_i, m_i)} \tag{6}$$

2-3 Repeat steps 2-1 and 2-2 until the value $TD(k)$ is equal to one of the previous steps, or the pre-defined number of iterations has been reached, or the set of medoids does not change.

4. Deviation Ratio Index based on K-Medoids

The deviation ratio index based on K-Medoids is a method used to determine the best number of clusters in a data set (Kariyam et al., 2023). Based on the final medoids for k a particular cluster, the deviation ratio or $DR(k)$ is defined by Equation 7.

$$DR\,(k) = \frac{SDW(k)/(n-k)}{SDB(k)/(k-1)} \tag{7}$$

where $SDW\,(k)$ such as in Equation 6, and $SDB\,(k)$ is the sum of distances of all objects to other medoids (between groups), formulated in Equation 8.

$$SDB\,(k) = \sum_{i=1}^{n}\sum_{k=1}^{g-1} d_{(x_i, m_k)} \tag{8}$$

with m_k shows the medoids of the other groups. Deviation ratio index is defined as the ratio of the deviation ratio of a group against k group to $(k-1)$ group. Deviation ratio index for k group is formulated as Equation 9.

$$DRI\,(k) = \frac{DR\,(k)}{DR\,(k+1)} \tag{9}$$

The best group is determined based on k the first smallest group with a value of $DRI\,(k) < 1$. Determination $DR\,(k)$ starts from $k = 2$ and increase the number of k groups until the value $DR\,(k)$ obtained is less than $DR\,(k+1)$ or $DR\,(k) < DR\,(k+1)$.

5. MANOVA

MANOVA, or Multivariate Analysis of Variance, is a statistical technique used to measure the effect of independent variables with categorical scales on several dependent variables with quantitative data scales. MANOVA can be continued with paired multivariate. The Confidence Interval (CI) in paired multivariate testing is formulated in Equation 10.

$$\left(\underline{x}_{gl} - \underline{x}_{hl}\right) \pm t_{(n-k),\left(\frac{\propto}{pk(k-1)}\right)} \cdot \sqrt{\frac{w_{ii}}{n-k}\left(\frac{1}{n_g} + \frac{1}{n_h}\right)} \tag{10}$$

for all variables $l = 1,\ 2,\ \cdots,\ p$ and all differences group $g < h = 1,\ 2,\ \cdots,\ k$. Here w_{ii} is the ith diagonal element of matrix of sum of squares and cross products $W(k)$ such as Equation 11.

$$W(k) = \sum_{g=1}^{k}\sum_{i=1}^{n_g}\left(x_{gi} - \underline{x}_g\right)\left(x_{gi} - \underline{x}_g\right)' \tag{11}$$

where n_g is the number of group members, x_{gi} is observation matrix of gth group with p variables, and $\underline{x}_g$ is mean vector of p variables in g th group.

C. RESULT AND DISCUSSION

1. Data Description

Descriptive statistics describe research subjects from samples or populations to provide useful information. In this study, descriptive statistics are used to understand 11 variable characteristics. Table 2 shows the descriptive statistics of numerical variables. Furthermore, the descriptive statistics of categorical variables are presented in the form of a pie chart visualization as shown in Figure 2.

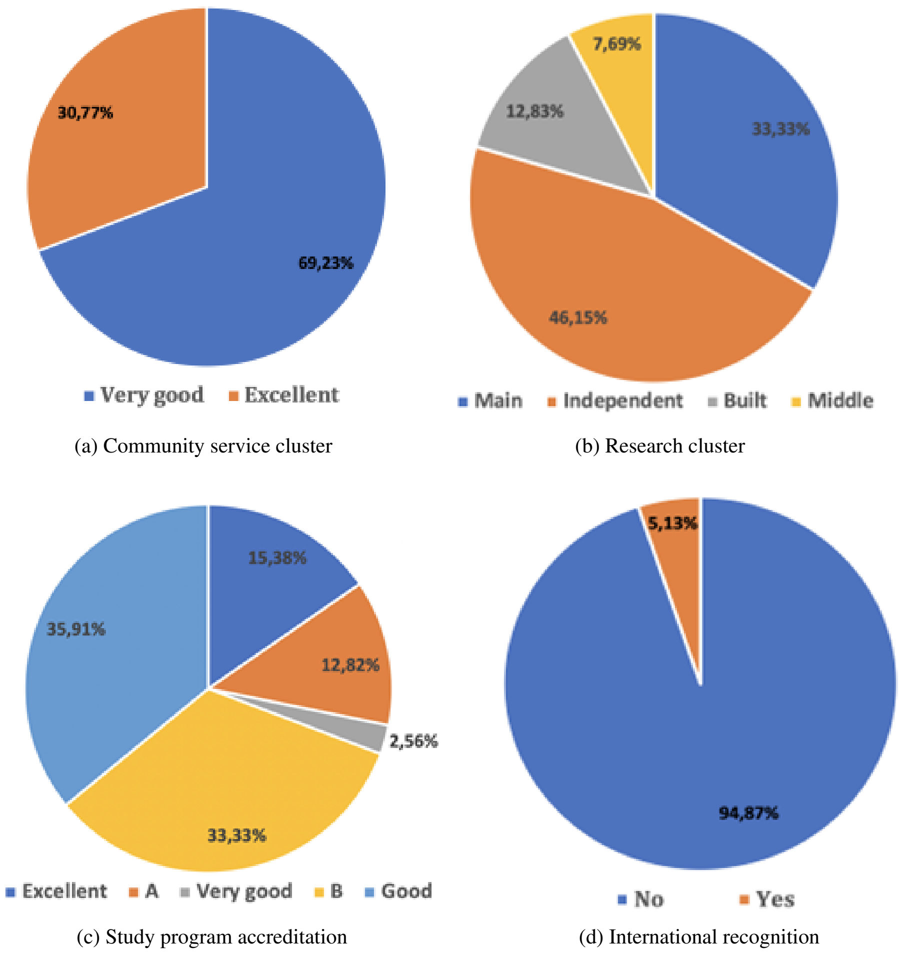
Table 2. Descriptive Statistics of Numerical Variables

Variable	Mean
X_1 (Lecturer enumeration)	10
X_2 (Percentage of S3 Lecturers)	19%
X_3 (Percentage of professors and associate professors)	17%
X_4 (Student enumeration)	509

Variable	Mean
X_5 (Percentage of graduation)	26%
X_9 (Experience of Program)	15&
X_{10} (Laboratory availability)	2

Based on Table 2, the average number of lecturers in the Statistics Undergraduate Program at 39 universities in Indonesia is 10 people. This number can be considered adequate for effective and personalized teaching. However, increasing the number may be needed to ensure each student receives optimal attention and guidance. The average percentage of lecturers with a doctoral degree is 19 percent, while the percentage of professors and head lecturers is 17 percent. This indicates room for improvement in terms of the lecturers' academic qualifications. Increasing the number of doctoral degree holders can strengthen the quality of research and teaching. Meanwhile, the percentage of professors and associate professors reflects higher experience and expertise in teaching and research. Each university with an Undergraduate Statistics Study Program in Indonesia has a total of 509 students. This number indicates a significant interest in the program. However, it is essential to maintain the student-to-lecturer ratio to ensure quality learning.

The average graduation rate of 26 percent shows that only about a quarter of students graduate on time. This may indicate challenges in the educational process that need to be addressed, such as curriculum, teaching methods, or academic support for students. The average age of the Bachelor of Statistics program is 15 years, indicating that these programs have substantial experience in delivering education. Older programs usually correlate with stability and maturity in the education process. Each Bachelor of Statistics program has 2 laboratory rooms. Higher Education in Indonesia obtains facilities in the form of a laboratory with as many as two rooms. In this era of widespread statistical applications, these facilities still need to be increased. The availability of adequate laboratories can improve the quality of education by providing students with better practical experience.



(a) Community service cluster (b) Research cluster

(c) Study program accreditation (d) International recognition

Figure 2. Descriptive Statistics of Categorical Variables

Furthermore, based on Figure 2, part (a), only about 30% of universities have achieved very good for the community service category. This indicates that institutions must prioritize activities that address community needs, foster partnerships with local organizations, and implement sustainable community service projects. Effective community engagement enhances the institution's societal impact and aligns with its mission to contribute to social development. Figure 2, part (b), shows that the majority of universities fall into independent cluster research. In the QMS, research collaboration and interaction between institutions are critical aspects that can enhance the quality and relevance of research. Increasing cooperation and integration in research can strengthen institutions' capacity to produce high-impact and globally relevant research. Figure 2, part (c), indicates that the achievement of Excellent and A accreditation rankings for study programs is no more than 30%. This means that only a small portion of study programs can achieve the highest rankings in the national accreditation system. This signals the need to strengthen the implementation of internal QMS so that more departments can meet the criteria for excellent accreditation. Meanwhile based on Figure 2, part (d), shows that the international recognition of study programs is only 5.13%, which is very low. In the quality assurance framework, international recognition is evidence that the study programs are globally acknowledged for their quality. The low percentage indicates the need for improvements in aspects of internationalization of education, such as collaboration with foreign universities, increasing faculty and student mobility, and adopting curricula that meet international standards.

2. Pre-processing and estimation the number of clusters

This research used 39 complete Statistics Study Program data (non-missing). Each Study Program is observed on seven numerical variables and four categorical variables. Some variables contain outliers, namely student enumeration, experience of the Study Program, and availability of laboratory. Therefore, before determining the number of clusters and classification of the object, a transformation is carried out using Equation 1. Based on the final medoids produced by the object block-based k-medoids partitioning method, the Deviation Ratio Index (DRI) value is obtained as in Figure 3.

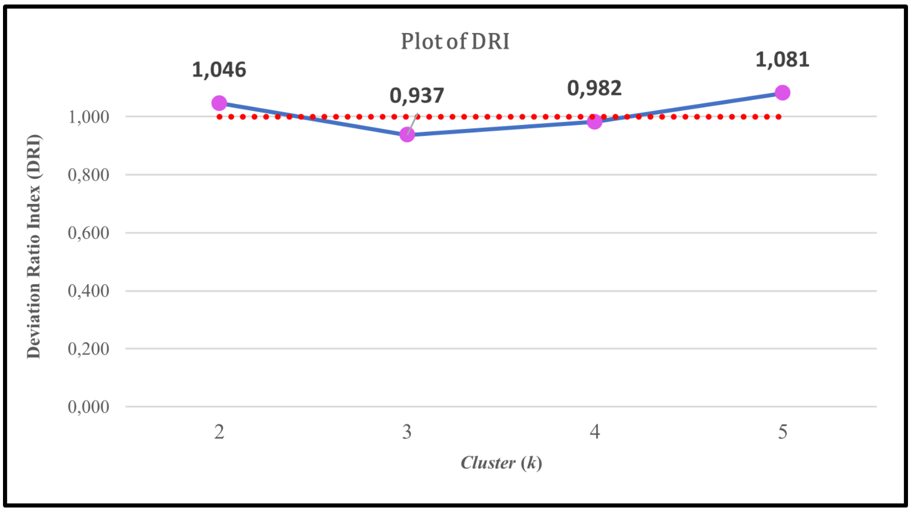


Figure 3. Graph of Deviation Ratio Index Value of Each Cluster

Figure 3 shows the Deviation Ratio Index (DRI) values for groups of sizes two to five. Determination of the best group size in Block K-Medoids analysis is determined using the Deviation Ratio Index based on K-Medoids (DRIM) method. If the (DRI) value is less than 1, it can be said that the resulting cluster is good, and vice versa. The Deviation Ratio (DR) value can also determine the best cluster. This study will divide the data clustering experiment into several k clusters, namely 2 to 5. Figure 3 shows the first k smallest so that Deviation Ratio Index (DRI) is less than one in three clusters. So, it can be said that cluster number three is the optimal cluster in the experiment.

Table 3. Summary of DRIM Based on Block-KM Partition Method ($k = 3$)

	College High	The process of getting medoid				Distance of object to medoid			$DR(k)$	
		u_i	w_i	IG	I3=I4	G1	G2	G3	$min\ d_i$	b_i
1	UNUGIRI	0,168	1,12	G1*	G1	0,79	1,86	7,11	0,79	8,97
2	UNU LAMPUNG	0,172	1,17	G2*	G1	1,22	1,81	7,07	1,22	8,87
3	MATANA	0,183	1,21	G3*	G1	1,27	1,81	7,02	1,27	8,83
4	UNHAZ	0,192	1,78	G1	G1	0,66	1,30	6,45	0,66	7,75
5	UNIBA	0,194	1,12	G1	G1*	0	1,96	7,11	0	9,07
6	UNIB	0,194	2,62	G2	G2*	1,96	0	5,61	0	7,57
7	UNSULBAR	0,212	1,28	G1	G1	0,51	1,97	6,95	0,51	8,92
8	ULM	0,231	2,77	G3	G2	2,78	1,98	5,46	1,98	8,24
9	UNPATTI	0,236	2,71	G2	G2	1,81	1,75	5,52	1,75	7,32
10	UNUGO	0,263	1,53	G3	G1	1,42	2,18	6,86	1,42	9,04
11	UGM	0,282	8,56	G3	G3	8,47	6,51	2,98	2,98	14,98
12	UNRI	0,288	4,32	G3	G2	3,41	1,91	4,46	1,91	7,87
13	UB	0,295	8,56	G3	G3	7,44	5,94	1,45	1,45	13,38
14	UNG	0,296	3,60	G3	G2	2,94	0,98	4,82	0,98	7,76
15	UII	0,297	8,23	G3	G3*	7,11	5,61	0	0	12,72
16	UNS	0,298	7,20	G3	G3	6,08	4,58	2,13	2,13	10,66
17	UNM	0,299	6,80	G3	G3	5,68	4,18	2,13	2,13	9,86
18	UNMUL	0,302	5,67	G3	G2	4,55	3,05	3,32	3,05	7,86
19	UNCEN	0,303	2,24	G2	G1	2,29	2,38	6,68	2,29	9,06
20	UNDIP	0,305	8,46	G3	G3	7,33	5,84	1,10	1,10	13,17
21	UNAIR	0,315	6,85	G3	G3	5,73	4,23	2,15	2,15	9,96
22	AKPRIND	0,316	4,80	G3	G2	4,47	2,85	4,53	2,85	9,00
23	BINUS	0,325	4,03	G2	G2	3,37	1,78	4,20	1,78	7,57
24	UNJ	0,327	4,22	G1	G2	3,89	2,58	4,33	2,58	8,22
25	UNTAD	0,328	3,79	G3	G2	2,89	2,07	4,71	2,07	7,60
26	UNIMUS	0,329	4,48	G3	G2	3,82	1,86	4,19	1,86	8,01
27	ITK	0,331	1,73	G1	G1	1,01	2,69	6,84	1,01	9,53
28	UNIPA SBY	0,336	4,01	G3	G2	2,88	1,81	4,22	1,81	7,11
29	UNSYIAH	0,338	6,57	G3	G3	5,45	3,95	2,16	2,16	9,40
30	UNY	0,339	4,40	G3	G2	4,07	3,38	3,83	3,38	7,90
31	UHO	0,344	4,92	G2	G2	4,01	2,84	4,02	2,84	8,03
32	ITS	0,350	7,91	G3	G3	6,79	5,29	1,22	1,22	12,08
33	UNHAS	0,352	7,06	G3	G3	6,73	4,77	2,43	2,43	11,50
34	UI	0,360	7,03	G3	G3	6,70	4,74	2,63	2,63	11,45
35	UNPAD	0,378	6,85	G3	G3	5,72	4,23	2,32	2,32	9,95
36	UNISBA	0,380	6,25	G3	G2	5,13	3,63	3,81	3,63	8,94
37	UT	0,380	5,13	G2	G2	4,67	3,50	4,24	3,50	8,90
38	UNP	0,390	4,64	G3	G2	4,56	3,76	4,14	3,76	8,70
39	IPB	0,431	4,86	G2	G2	4,31	2,92	3,75	2,92	8,06
	Total Deviation TD(k)			**145,9**	**74,5**	$SDW(a)$ & $SDB(b)$ $DR(k)$			**74,5** (a) **0,011**	**363,8** (b)

Based on Table 3, it was found that the medoid in cluster 1 experienced an update in the first iteration, namely UNUGIRI changed to UNIBA. From the first iteration to the fourth iteration, it did not change. Likewise, with cluster 2, UNU LAMPUNG became UNIB. In contrast, cluster 3 continued to update the medoid until the third iteration and experienced medoid stability in the fourth iteration.

Table 4. Membership of Three Clusters

Cluster	Cluster Member	Total
1	UNUGIRI, UNU LAMPUNG, MATANA, UNHAZ, UNIBA, UNSULBAR, UNUGO, UNCEN, ITK	9
2	UNIB, ULM, UNPATTI, UNRI, UNG, UNMUL, AKPRIND, BINUS, UNJ, UNTAD, UNIMUS, UNIPA SBY, UNY, UHO, UNISBA, UT, UNP, IPB	18
3	UGM, UB, UII, UNS, UNM, UNDIP, UNAIR, UNSYIAH, ITS, UNHAS, UI, UNPAD	12

Based on Table 4, the clustering results obtained by cluster 1 consisted of 9 study programs, cluster 2 of 18 study programs, and cluster 3 of 12 study programs. Statistics study programs from private universities dominate the first group. Group two members are quite balanced between state and private universities. Meanwhile, for group three, almost all study programs are held by state universities, except for one statistics department from the Universitas Islam Indonesia (UII), which is a private university.

3. MANOVA

Profilization was performed using the Multivariate Analysis of Variance (MANOVA) method for numerical data and percentages for categorical data. From the results of MANOVA testing with four test statistics, it was found that the $p - value < 0.05$, so reject H0 which means that there is one different group average vector. From this test, further testing is continued, namely paired multivariate testing. A summary of paired multivariate tests for numerical data is presented in Table 5.

Table 5. Summary of Paired Multivariate Test

Variable	Conclusion
X_1 (Number of lecturer)	$\mu_1 = \mu_2 < \mu_3$
X_2 (Percentage of S3 lecturer)	$\mu_1 < \mu_2 < \mu_3$
X_3 (Percentage of professors & associate professors)	$\mu_1 < \mu_2 = \mu_3$
X_4 (Number of student)	$\mu_1 = \mu_3 < \mu_2$
X_5 (Graduation percentage)	$\mu_1 < \mu_2 = \mu_3$
X_9 (Experience of Study Program)	$\mu_1 < \mu_2 = \mu_3$
X_{10} (Laboratory availability)	$\mu_2 < \mu_1 = \mu_3$

Based on Table 5, the results show that the number of lecturers and experience of the study program for group one is the same as that of group Two, and the average is below Group Three. The lowest achievement for the percentage of lecturers with doctoral degrees was group one, whereas the highest was in group three. Meanwhile, the graduation percentage and percentage of lecturers with the titles of professor and associate professor for group two are the same as those in group three, and both are higher than those in the first group. The number of students in group one is the same as group three, and both are lower than group two. The laboratory facilities available in group one is the same as group three, and both are higher than group one. Overall, group three has the highest outcomes compared to other groups. The average profile for numerical variables is as shown in Table 6.

Table 6. Profiling of Cluster Members Based on Numerical Variables

Cluster	X_1	X_2	X_3	X_4	X_5	X_9	X_{10}
1	7,78	6,67	0,00	84,67	10,57	5,33	2,56
2	9,28	16,67	16,70	781,60	27,26	14,83	1,56
3	12,83	31,44	29,40	418,80	38,57	24,42	2,17

When viewed from Table 6, the dark red rows are clusters with numerical variable values in the excellent category. Meanwhile, the colourless ones are clusters with numerical variable values in the fair category, and the dark green rows are in the low category. Cluster characteristics in categorical data are presented as a percentage of one category that dominates certain variables in a cluster. The percentage is presented in Table 7.

Table 7. Profiling of Cluster Members Based on Numerical Variables

Clusters	X_6 (Research cluster)				X_7 (Accreditation)				
	1	2	3	4	1	2	3	4	5
1	**55,56%**	33,33%	11,11%	-	100%	-	-	-	-
2	-	-	**61,11%**	38,89%	27,78%	**61,11%**	5,56%	-	5.56%
3	-	-	-	**100%**	-	16,67%	-	**41,61%**	**41,67%**

Clusters	X_8 (International recognition)		X_{11} (Community service cluster)	
	0	1	1	2
1	**100%**	-	100%	-
2	**100%**	-	**88,89%**	11,11%
3	**100%**	8,33%	16,67%	**83,33%**

Based on Table 7 members of cluster 1 are dominated by universities, with research clusters categorized as Binaan and service and cooperation clusters categorized as ”Very Good ”. The majority of universities in cluster 1 are accredited ”Good” and not yet internationally recognized. Cluster 2 members are dominated by universities with a category ”Main” research cluster and category ”Very Good” service and cooperation cluster. Universities in cluster 2 are mostly accredited ”B” and not yet internationally recognized. Cluster 3 members are dominated by universities with ”Independent” research clusters and Excellent (Unggul) service and cooperation clusters. Some universities are accredited as Excellent and A, and there are universities that have been internationally recognized.

The findings from this research are that there is a group of study programs with an average ratio of students to lecturers above 80, namely group two. In other sections, the average was found to be below 12. This fact is clearly far from the ideal provisions for assessing study program accreditation, namely between 15 and 25. This information also shows the inequality in the number of students and lecturers in a number of statistics study program providers. The consequences of this mismatch can certainly result in a learning process that is not optimal. Meanwhile, nine study programs do not yet have lecturers with the academic positions of associate professor and professor, while on average less than 10% of lecturers have doctoral degrees. This fact clearly reflects that the academic quality of lecturers still really needs to be improved. The highest average percentage of new associate professors and professors reached 29.4%, which also shows that achievements still do not meet the excellent accreditation ranking.

Another finding from this research is that the average passing percentage is very low in all groups, with the highest only reaching 38.57%. This achievement also shows that the learning process is not in accordance with the undergraduate program curriculum design with a normal study period of eight semesters. In other words, the majority of students experience delays in their studies. This problem may be caused by the ratio of students to lecturers not complying with the provisions or the academic quality of the lecturers, which needs to be improved. These results should be taken into consideration for evaluating the curriculum and lecturers.

A good result from this research is that one statistics study program has been recognized internationally in terms of accreditation, and on average, more than 40% of group members have an excellent accreditation rating. Of course, this study program can be used as a reference for other comparative study programs to achieve international accreditation or at least an excellent ranking. The results of this research indirectly show that the output and outcome of the Quality Management System in academic and non-academic areas still really needs to be improved. Commitment to implementing continuous quality improvement is an important part that must be considered by all parties involved in the higher education quality management system. Noting that the grouping of Statistics Study Programs has never been researched, it is hoped that these results can be used as a reference for further research, especially regarding the outcomes of implementing internal QMS.

D. CONCLUSION AND SUGGESTION

Applying the Deviation Ratio Index based on K-Medoids (DRIM) concluded that the optimal number of clusters was three groups of Statistical Study Programs. The final medoids used in this study are the results of the using Block-based K-Medoids Partitioning Method (Block-KM) with Manhattan distance and are stable at the fourth iteration. Cluster 1 consists of 9 study programs, cluster 2 has 18 study programs, and cluster 3 consists of 12. From the clustering results, characteristics are formed in each cluster. Based on numerical data, cluster 1 is a cluster with low education quality standard indicators; cluster 2 is a cluster with fairly good education quality standard indicators, and cluster 3 is a cluster with excellent education quality standard indicators. Cluster characteristics based on categorical data found that cluster 1 is dominated by universities with research clusters in the built category and community service and cooperation clusters in the very good category, accredited good and not yet internationally recognized. Cluster 2 is dominated by universities with research clusters in the primary category, community service and cooperation clusters in the very good category, accredited B, and not yet internationally recognized. Cluster 3 is dominated by universities with independent category research clusters and community service and cooperation clusters categorized as excellent, accredited as excellent, and A and some study programs that have been internationally recognized.

Overall, the research findings indicate that the implementation of the internal quality management system in Indonesia still requires significant improvements. There are several areas that require more attention, such as increasing lecturer qualifications, especially in terms of lecturers with doctoral and professor degrees, as well as increasing the percentage of graduating students. Focusing on improving the quality of teaching, academic support, and educational facilities can help achieve better and higher quality education delivery standards. While some study programs have achieved excellent accreditation, the majority lag in various assessed aspects. Increasing research collaboration, strengthening quality management, and improving international recognition are some steps that can be taken to enhance the overall quality of higher education in Indonesia.

ACKNOWLEDGEMENT

The title for the thank you to the institution or the person who has contributed during the research and references is not numbered.The Acknowledgments section is optional. Research sources can be included in this section.

DECLARATIONS

AUTHOR CONTIBUTION

Asa Nugrahaini Itsnal Muna: data curation, coding, writing-original, and editing of draft preparation. Kariyam: conceptualization, supervision, validation, reviewing, editing for revision.

FUNDING STATEMENT

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

REFERENCES

Ab Wahid, R., & Grigg, N. P. (2021). A draft framework for quality management system auditor education: Findings from the initial stage of a Delphi study. *The TQM Journal*, *33*(6), 1373–1394. https://doi.org/10.1108/TQM-08-2020-0193

Budiaji, W., & Leisch, F. (2019). Simple K-Medoids Partitioning Algorithm for Mixed Variable Data. *Algorithms*, *12*(9), 177. https://doi.org/10.3390/a12090177

Cardoso, S., Rosa, M. J., Videira, P., & Amaral, A. (2019). Internal quality assurance: A new culture or added bureaucracy? *Assessment & Evaluation in Higher Education*, *44*(2), 249–262. https://doi.org/10.1080/02602938.2018.1494818

DeSimone, S., & Rich, K. (2020). Determinants and consequences of internal audit functions within colleges and universities. *Managerial Auditing Journal*, *35*(8), 1143–1166. https://doi.org/10.1108/MAJ-10-2019-2444

Halibas, A. S., Mehtab, S., Al-Attili, A., Alo, B., Cordova, R., & Cruz, M. E. L. T. (2020). A thematic analysis of the quality audit reports in developing a framework for assessing the achievement of the graduate attributes. *International Journal of Educational Management*, *34*(5), 917–935. https://doi.org/10.1108/IJEM-07-2019-0251

Hauptman Komotar, M. (2020). Discourses on quality and quality assurance in higher education from the perspective of global university rankings. *Quality Assurance in Education*, *28*(1), 78–88. https://doi.org/10.1108/QAE-05-2019-0055

Kariyam, Abdurakhman, & Effendie, A. R. (2023). A medoid-based deviation ratio index to determine the number of clusters in a dataset. *MethodsX*, *10*, 102084. https://doi.org/10.1016/j.mex.2023.102084

Kariyam, Abdurakhman, Subanar, Utami, H., & Effendie, A. R. (2022). Block-Based K-Medoids Partitioning Method with Standardized Data to Improve Clustering Accuracy. *Mathematical Modelling of Engineering Problems*, *9*(6), 1613–1621. https://doi.org/10.18280/mmep.090622

Kariyam, K., Abdurakhman, A., Subanar, S., & Utami, H. (2022). The Initialization of Flexible K-Medoids Partitioning Method Using a Combination of Deviation and Sum of Variable Values. *Mathematics and Statistics*, *10*(5), 895–908. https://doi.org/10.13189/ms.2022.100501

Lahouaoui, L., Abdelhak, D., Abderrahmane, B., & Toufik, M. (2022). Image Classification Using a Fully Convolutional Neural Network CNN. *Mathematical Modelling of Engineering Problems*, *9*(3), 771–778. https://doi.org/10.18280/mmep.090325

Randriamihamison, N., Vialaneix, N., & Neuvial, P. (2021). Applicability and Interpretability of Ward's Hierarchical Agglomerative Clustering With or Without Contiguity Constraints. *Journal of Classification*, *38*(2), 363–389. https://doi.org/10.1007/s00357-020-09377-y

Solimun, S., & Fernades, A. A. R. (2022). Cluster Analysis on Various Cluster Validity Indexes with Average Linkage Method and Euclidean Distance (Study on Compliant Paying Behavior of Bank X Customers in Indonesia 2021). *Mathematics and Statistics*, *10*(4), 747–753. https://doi.org/10.13189/ms.2022.100405

Sureja, N., Chawda, B., & Vasant, A. (2022). An improved K-medoids clustering approach based on the crow search algorithm. *Journal of Computational Mathematics and Data Science*, *3*, 100034. https://doi.org/10.1016/j.jcmds.2022.100034

Yuan, F., Yang, Y., & Yuan, T. (2020). A dissimilarity measure for mixed nominal and ordinal attribute data in k-Modes algorithm. *Applied Intelligence*, *50*(5), 1498–1509. https://doi.org/10.1007/s10489-019-01583-5

Zuhairi, A., Raymundo, M. R. D. R., & Mir, K. (2020). Implementing quality assurance system for open and distance learning in three Asian open universities: Philippines, Indonesia and Pakistan. *Asian Association of Open Universities Journal*, *15*(3), 297–320. https://doi.org/10.1108/AAOUJ-05-2020-0034

[This page intentionally left blank.]