Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 24, No. 1, November 2024, pp. 141∼154
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v24i1.4244

Higher Education Institution Clustering Based on Key Performance Indicators using Quartile Binning Method

Virdiana Sriviana Fatmawaty , Imam Riadi , Herman
Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received July 12, 2024
Revised October 25, 2024
Accepted November 07, 2024

Keywords:

Descriptive Analytics
Higher Education Institution
Key Performance Indicators
Quartile-Based Clustering
Statistical Analysis

ABSTRACT

The Key Performance Indicators of Higher Education Institutions (KPI-HEIs) are a crucial component of the internal quality assurance system that supports the achievement of excellence status for higher education institutions. Many private higher education institutions face challenges in independently analyzing the key performance assessment indicators of Private Higher Education Institutions (PHEIs), which often require complex methodological approaches and specialized expertise. The research aims to cluster PHEIs based on achieving key performance indicators (KPIs). Research the method used descriptive statistical methods and quartile binning techniques to analyze and cluster data based on the achievement of KPI-HEIs. The research results, based on descriptive statistical analysis, identified outliers in eight KPI-HEIs, along with a dominance of zero values in KPI 1, KPI 2, KPI 6, KPI 7, and KPI 8, with the highest proportion reaching 90.91% for KPI 8. Based on these findings, clustering using the quartile binning method resulted in four clusters of PHEIs based on KPIs: Cluster 1 consists of 19 institutions with poor, Cluster 2 consists of 14 institutions with fair achievement, Cluster 3 consists of 16 institutions with good achievement, and Cluster 4 consists of 17 institutions with very good achievement, which can serve as examples for other institutions. This research concludes that the quartile binning method successfully categorized private higher education institutions based on their achievement of KPIs into four clusters: poor, fair, good, and very good. This outcome demonstrates the effectiveness of the method in understanding the performance distribution of these institutions. It provides valuable insights for stakeholders to develop data-driven strategies aimed at enhancing educational quality.

Corresponding Author:

Virdiana Sriviana Fatmawaty, +6281328000903
Master Program of Informatics,
Universitas Ahmad Dahlan, Yogyakarta, Indonesia.
Email: 2307048011@webmail.uad.ac.id

How to Cite:
V. Fatmawaty, I. Riadi, and H. Herman, "Higher Education Institution Clustering Based on Key Performance Indicators using Quartile Binning Method", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 24, No. 1, pp. 141-154, November, 2024.

1. INTRODUCTION

The quality assurance system of higher education institutions is a series of processes consisting of planning, implementing, monitoring, and improving higher education standards consistently and continuously [1]. This process is designed to ensure that all stakeholders, both internal and external, are satisfied with the performance and results achieved by higher education institutions [2]. It reflects the institution's commitment to responsibility and transparency in its educational management [3]. The Decree of the Minister of Education and Culture of the Republic of Indonesia in 2023 concerning Key Performance Indicators of Higher Education Institutions (KPI-HEIs), also known in Indonesia as Key Performance Indicators for Higher Education Institutions (IKU-PT) and Higher Education Service Institutions (HESI), referred to in Indonesia as Higher Education Service Institution (LLDikti), emphasizes that key performance indicators (KPI) are important instruments in the internal quality assurance system of higher education institutions (HEIs). They function as benchmarks used to assess the performance of higher education institutions across various aspects, including academics, research, community service, and administration.

The assessment process of KPI-HEIs involves data from various sources, including the higher education database, tracer study, Simkatmawa, sister, and partnership data of higher education institutions. This data includes critical indicators such as the percentage of graduates work readiness, the number of students engaged in the Freedom of Learning independent campus activities, the number of lecturers participating in activities outside the main campus, and the percentage of partnerships with partners [4]. Despite its importance, many private higher education institutions (PHEIs) in Indonesia called Private Higher Education Institution (PTS), particularly in HESI Region V Yogyakarta, face challenges in collecting and analyzing data from various indicators. These challenges are due to limitations in skilled human resources for data analysis [5], inadequate technology, and the complexity of data collection procedures.

Similar research that analyzes data characteristics using clustering methods and descriptive statistics includes research on clustering distributional data based on microscopic similarities using quantile values [6]. The results demonstrate that the proposed method is effective for symbolic data with complex distributions, such as histograms, and is also beneficial for simpler distributions, such as intervals. Using quantiles allows for comparing various types of symbolic objects without adding complexity. Another study [7] used a descriptive quantitative design with a cross-sectional approach to analyze the knowledge and anxiety level about covid 19 vaccination in cadres. The results showed that 6% of respondents had a high level of knowledge, 10% had a medium level of knowledge, and 84% had a low level of knowledge. In addition, 64% of respondents experienced severe anxiety, 30% moderate anxiety, and 6% low anxiety. In addition, research [8] aims to analyze the effect of service quality on inpatient patient satisfaction at Bahagia Makassar Hospital. The method used is descriptive statistical analysis. The results showed that the dimensions of service quality, namely tangible, responsiveness, assurance, and empathy, affected increasing patient satisfaction at Bahagia Hospital Makassar, as evidenced by the regression and correlation coefficient analysis results. Research [9] aims to determine students' levels and opinions regarding the e-module of vector material to be developed. This research uses a mixed method with a purposive sampling technique. Descriptive statistical results, with an average value of 23.07 in the required category, show that most samples think that e-modules are a very good idea.

Some differences between this research and previous research are that advanced analytical techniques, especially quartile binning, have not been applied to analyze the data characteristics of KPI-HEIs indicators data for higher education institutions. This data was obtained directly from a website with limited access: https://iku-pt.kemdikbud.go.id/. Furthermore, no research has yet examined the characteristics of KPI-HEIs data for higher education institutions using descriptive analysis methods with quartile binning. This research aims to address the gap by applying the quartile binning approach to gain deeper insights into the distribution and characteristics of Key Performance Indikator data for higher education institutions.

Initial data analysis of this research identified issues with outliers and high variation in the collected data. To address these problems, this research proposes a solution by grouping the data using quartile binning. The solution provides a clearer picture of the distribution and characteristics of the data The quartile binning or discretization method is used to cluster data based on quartile values, which allows the identification of distribution patterns and variations between quartile clusters, making it easier to analyze and understand the distribution of data and helping to identify anomalies and outliers that can affect the results of the analysis [10]. This research clusters PHEIs in the HESI Region V Yogyakarta, based on the achievement of KPI-HEIs data, considering data characteristics such as distribution and spread. Through a descriptive statistical approach and utilizing quartile binning techniques, the research examines data distribution and optimizes the clustering of PHEIs. The hope is that higher education institutions will be better prepared to report KPI-HEIs achievement data, make the necessary improvements, and ultimately enhance their education quality. The solutions proposed in this research are expected to significantly contribute to improving the effectiveness and efficiency of the KPI-HEIs data collection process. Additionally, the results of this research could serve as a reference for the HESI Region V Yogyakarta as a liaison between PHEIs and the Ministry of Education, Culture, Research, and Technology in developing a more effective, data-driven strategy for mapping higher education quality.

2. RESEARCH METHOD

This research employs quantitative methods using a descriptive statistical approach and quartile binning. Descriptive statistics are applied as an initial step to identify the characteristics of the data. Subsequently, quartile binning is used as a further step to classify the data based on specific performance metrics. The research focuses on KPI data from PHEIs under the jurisdiction of the HESI Region V Yogyakarta. The dataset includes 66 PHEIs, comprising universities, institutes, and sekolah tinggi (college).

2.1. Data Source

The KPI-HEIs data for higher education institutions used in this research consists of achievement assessment data obtained from the Key Performance Indicator system for Higher Education Institutions, accessed through the official portal at https://iku-pt.kemdikbud.go.id/. This portal aims to monitor and evaluate the internal quality of higher education institutions. Eight assessment indicators are used in the Key Performance Indicator system, following the Indonesian Minister of Education and Culture Regulation of 2021, updated in 2023, on Key Performance Indicators for HEIs and HESI, detailed in Table 1.

Table 1. Indicator on KPI-HEIs

Indicator Codes		Indicator	Unit
I1	KPI 1	S1 and D4 / D3 / D2 / D1 graduates who succeed in getting a job, continuing their studies, or becoming self-employed to the number of graduates	%
I2	KPI 2	Percentage of the number of students who carry out the independent campus learning independence program (MBKM), and student achievement that meets the requirements of the total number of MBKM students and student achievement	%
I3	KPI 3	Percentage of the total number of permanent lecturers working in Tridharma at other higher education institutions, practitioners in the industrial world, or supervising students engaged in activities outside their study program compared to the total number of permanent lecturers	%
I4	KPI 4	Percentage of the total number of permanent lecturers (NIDN) or non-permanent lecturers (NIDK) who have competency/professional certificates, the number of teaching staff from professional practitioners in the industry or workforce, compared to the total number of permanent lecturers (NIDN), non-permanent lecturers (NIDK), and non-lecturer staff (NUP)	%
I5	KPI 5	The percentage ratio of the number of works by permanent lecturers (NIDN) or non-permanent lecturers (NIDK) that receive international recognition or are utilized by the community/industry/government to the total number of NIDN/NIDK lecturers	%
I6	KPI 6	Percentage ratio of the number of collaborations in undergraduate (S1) or diploma (D4/D3/D2/D1) programs that meet performance standards to the total number of undergraduate (S1) and diploma (D4/D3/D2/D1) programs	%
I7	KPI 7	Percentage of the total number of courses using case method or team-based project as a teaching method and part of the evaluation weight compared to the total number of courses offered during the current year	%
I8	KPI 8	Percentage of the total number of undergraduate (S1) and diploma (D4/D3) programs that have government-recognized international accreditation or certification compared to the total number of undergraduate (S1) and diploma (D4/D3) programs that have graduated at least one batch of students	%

The eight KPI-HEIs are labeled from I1 to I8. The unit of measurement for each indicator is a percentage. These indicators are designed to provide a comprehensive picture of various aspects of institutional performance and quality, including graduate success (I1), student participation in the strategic Merdeka Belajar Kampus Merdeka (freedom of learning independent campus) programs (I2), lecturer involvement in academic and industrial activities (I3-I5), the number of collaborative partners (I6), recognition of learning methods (I7), and international recognition of study programs (I8). This research uses data from 66 PHEIs within the HESI Region V Yogyakarta, including various types of institutions such as universities, institutes, and colleges. The data obtained then went through a transformation stage to change the format of the indicator data and data cleansing [11]. After data transformation [12] and data cleansing [13], the next step is to perform data characterization analysis to identify potential outliers and anomalies that may require further attention [14]. The final step in this process is data clustering based on quartiles. The quartile method divides the data into four groups, each reflecting different characteristics. Researchers can cluster higher education institutions by dividing the data into quartiles based on their key performance indicator achievements. This clustering process is crucial because it allows for effectively identifying and analyzing variations and trends within the dataset. In this way, performance differences among the institutions can be observed more clearly and in greater depth. This provides sharper insights into areas needing improvement and enables researchers to make more targeted recommendations.

2.2. Research Steps

This aims to illustrate each phase of the analytical process to provide a clearer understanding of the stages involved in analyzing data characteristics, particularly in examining data distribution and spread. It demonstrates how data is systematically processed, analyzed, and interpreted to reveal patterns, variability, and overall trends, ensuring a comprehensive dataset evaluation. Details of these steps are visually presented in Figure 1.

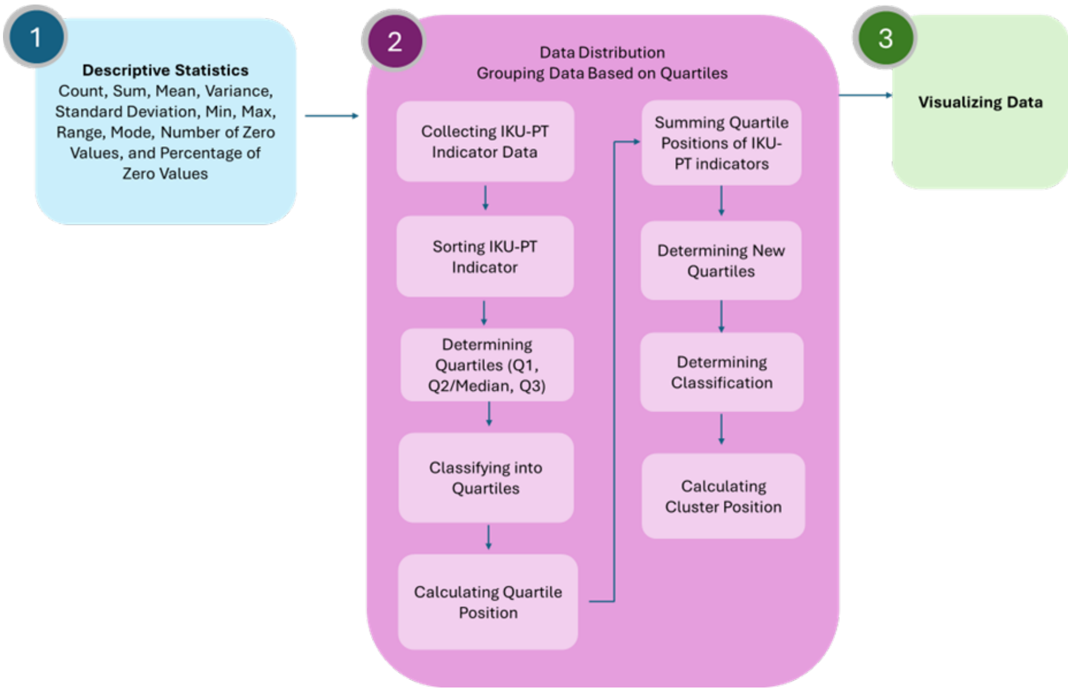

Figure 1. Data Characteristics Analysis Process

This research applied a comprehensive descriptive statistical approach to explore the distribution of data [15]. Descriptive statistical methods are techniques used to describe and summarize data [16], providing a detailed understanding of data characteristics [17]. Descriptive analysis aims to understand the distribution of data, identify patterns and trends, provide statistical summaries, and detect errors and anomalies [18]. Descriptive statistics encompass measures of central tendency, measures of dispersion (variability), assessment of distribution shape, measures of position, and data visualization. In addition, the minimum and maximum values are identified to determine the range of the lowest and highest values in the data, providing insight into the scale of the data distribution. The data range, which is the difference between the maximum and minimum values, provides additional information about the width of the data distribution. The mode, or the value that appears frequently in the dataset, is analyzed to determine the tendency of certain frequencies in the data. The number of zero values and percentage of zero values are also calculated to assess the proportion of data that has zero values, which can be an important indicator in some types of data analysis, especially in fields where the presence or frequency of zero values has special significance.

2.3. Quartile Binning

The initial descriptive analysis of key performance indicator data from 66 PHEIs identified several outliers and indications of high variance. To address this, further analysis was conducted using the quartile binning method to examine the data distribution within smaller groups. Binning itself is the process of converting continuous data into discrete data by dividing a range of values into several intervals or bins [19] to observe patterns or trends within each interval. Quartile binning is a statistical technique that divides data using three dividers (Q1, Q2, Q3). This division process results in four groups of data, namely the first quartile group, the second, the third, and the fourth [11]. Distribution analysis using the quartile binning method begins with data collection, followed by initial data analysis. Subsequently, the process involves sorting the KPI-HEIs from the lowest to the highest value for each indicator. After the data sorting, the dividing process uses three dividers. The next step is to determine the quartiles, namely Q1, Q2, and Q3, using the formulas shown in Equations 1, 2, and 3.

$$Q1 = \frac{1}{4} \times (n + 1) \tag{1}$$

$$Q2 = \frac{1}{2} \times (n + 1) \tag{2}$$

$$Q3 = \frac{3}{4} \times (n + 1) \tag{3}$$

Quartile 1 (Q1) is the value that divides the data into the lowest 25%. Quartile 2 (Q2), or the median, is the middle value that divides the data into two equal parts. Quartile 3 (Q3) is the value that divides the data into the top 25%. The variable n represents the number of data points. After the quartile values (Q1, Q2, and Q3) are determined, the key performance indicators data is divided into four clusters based on quartiles [**?**]. By categorizing into four groups, the main purpose is to gain a deeper understanding of the existing patterns and make the interpretation of the results easier [20] because each group has the same or approximately the same number [21] This allows the researcher to see variations and trends that may be hidden in the data in greater detail, facilitates a better understanding of each cluster's characteristics, and provides a solid foundation for deeper analysis and more accurate, detailed interpretation. Thus, the breakdown into four clusters provides a solid foundation for deeper analysis and more accurate interpretation. The division of the four clusters is defined in Table 2.

Table 2. Grouping Based on Quartile

Cluster	Value	Description
1	data $\leq$ Q1	Poor
2	Q1 $<$ data $\leq$ Q2	Fair
3	Q2 $<$ data $\leq$ Q3	Good
4	data $>$ Q3	Very Good

This quartile classification will cluster the data from each KPI-HEIs from all data points, consisting of 66 PHEIs, into four quartiles. This process involves dividing the indicator data into four equal parts based on their quartile values, each representing 25% of the data. The classification position of each indicator is calculated based on the assessment of 8 indicators, each with a different formula. For each data point from each PHEIs, the indicator value will is computed by summing the values corresponding to the determined quartile data. In other words, each PHEIs is evaluated based on its position within the distribution of its key performance indicator data, where the value is measured against quartile boundaries namely the first cluster (1) includes data whose value is less than or equal to the first quartile (Q1) with a description of poor. the second cluster (2) includes data whose value is greater than the first quartile (Q1) and less than or equal to the second quartile (Q2) with a description of fair. Similarly, the third cluster (3) includes data whose value is greater than the second quartile (Q2) and less than or equal to the third quartile (Q3) with a good description. The fourth cluster (4) includes data whose value exceeds the third quartile (Q3) with a very good description. After determining the quartile group, the next step is to calculate the position for each indicator based on the classification in Table 2. The quartile position data for each data point from the higher education institutions are summed across indicators 1 to 8. The results of this summation are then used to calculate new quartiles using the formula provided in Formula 1. Subsequently, the total number of quartile position data points is grouped based on the new quartile values, resulting in the final cluster value for each higher education institution. Data visualization is an effective way to summarize the results of data distribution and clustering [22]. According to John W. Tukey in his book "Exploratory Data Analysis" (1977), data visualization plays an important role in conveying data analysis in a clear and informative manner [23]. Data visualization can also be in the form of various forms of diagrams such as histograms, scatter, and heatmaps, each of which helps in understanding and interpreting the data better [24].

3. RESULT AND ANALYSIS

This research will focus on data obtained from 66 PHEIs within the HESI Region V Yogyakarta, which includes universities, institutes, and colleges. Initial data plays a crucial role in providing an early overview of how data is distributed across these institutions [25]. This data enables researchers to assess each institution's various aspects and key performance indicator values before conducting a more in-depth analysis. By examining the initial data, researchers can understand the data distribution among the institutions, including how each institution's performance compares to others. Moreover, the initial data is important for evaluating relevant key performance indicator values, which will assist in identifying critical aspects that need attention. This preliminary understanding provides a solid foundation before conducting further analysis to identify potential patterns, trends, and variations within the data. The initial overview derived from this data will serve as the basis for preliminary descriptive statistics analysis, as presented in Table 3.

Table 3. Sample Data of KPI-HEIs

PT Code	I1	I2	I3	I4	I5	I6	I7	I8
	%	%	%	%	%	%	%	%
PT1	14.64	0	22.29	36.52	230.6	0	3.97	28.57
PT2	9.64	2.23	21.03	35.17	379.66	5.88	45.31	0
...	...	...	...	...	...	...	...	...
PT65	0	0	57.89	31.58	89.47	0	0	0

This dataset presents the values of 8 indicators measured at 66 universities, institutes, and colleges. As an illustration, higher education is represented by the code PT1 (representing higher education); in indicator I1, which measures the percentage of graduate success, 14.64% of graduates successfully get jobs, are self-employed, and continue their studies. Another example, indicator I3 at PT65, which measures the involvement of lecturers in conducting teaching, research, and community service outside their homes as practitioners or guiding students in the Merdeka Belajar Kampus Merdeka program, shows a value of 57.89%. The initial data shows indications of outliers, as seen in the scatter diagram in Figure 2.

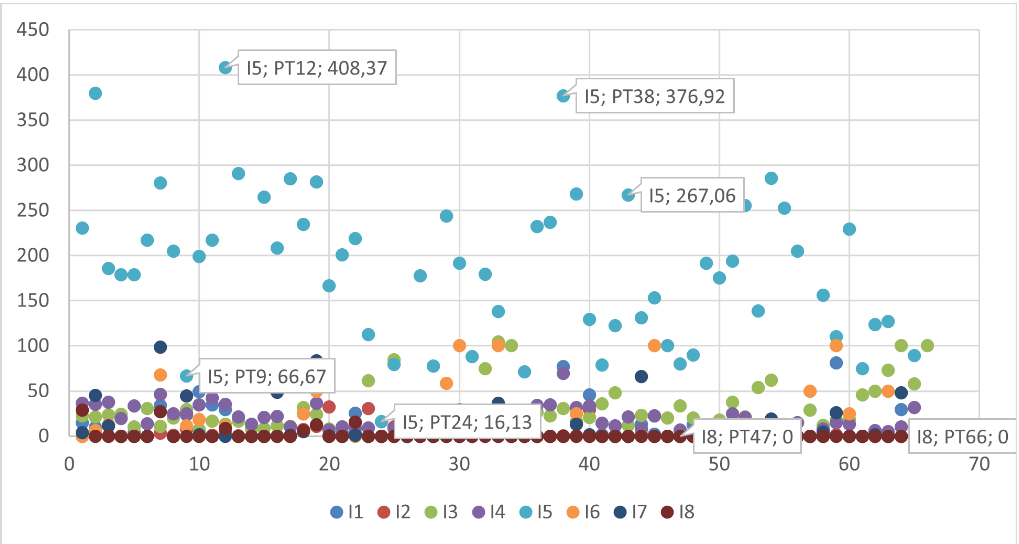

Figure 2. Initial Data Diagram of KPI-HEIs

In this diagram, the X-axis represents the 66 private universities identified by numerical codes from PT1 to PT66. Meanwhile, the Y-axis depicts the percentage of KPI values from I1 to I8. For example, at the PHEIs with code PT12, the percentage value for the 5th KPI (I5) is 408.37 percent. In contrast, the percentage value for the PHEIs with code PT24 for the 5th KPI (I5) is 16.13 percent. Other examples in the diagram show zero values for some indicators, as seen in the 8th KPI (I8) for PHEIs with codes PT47 and PT66. These examples highlight significant differences or high variance in KPI values among PHEIs. This is particularly evident in the 5th KPI (I5), where some institutions, such as PT12, display very high values, while others, such as PT24, show very low values. Additionally, the I8 indicator shows many zero values, reflecting a lack of achievement or reporting on this indicator at some institutions, such as PT47 and PT66. This research utilized the data described in Table 1, Table 3, and Figure 2. The preliminary analysis presents initial results on the characteristics, which is a crucial step in the research process to provide an overview of the fundamental attributes of the collected data. At this stage, the primary focus is to identify and describe the fundamental characteristics of the data, including the mean, median, standard deviation, and range of the variables studied. This preliminary analysis also helps to identify potential patterns or anomalies that may affect further analysis. The results of the preliminary analysis are shown in Table 4.

Table 4. Distribution of KPI-HEIs Data

index	I1	I2	I3	I4	I5	I6	I7	I8
count of data	66.00	66.00	66.00	66.00	66.00	66.00	66.00	66.00
sum of data	589.82	169.99	1974.74	1167.50	11222.98	948.84	567.20	98.19
mean	8.94	2.58	29.92	17.69	170.05	14.38	8.59	1.49
variance	314.36	44.18	634.04	197.05	8670.21	798.55	417.63	28.79
standard deviation	17.73	6.65	25.18	14.04	93.11	28.26	20.44	5.37
minimum	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
maximum	80.82	32.35	104.55	69.23	408.37	100.00	98.48	28.57
range	80.82	32.35	104.55	69.23	408.37	100.00	98.48	28.57
number of zero values	35.00	40.00	1.00	9.00	4.00	46.00	42.00	60.00
Percentage of zero values	53.03	60.61	1.52	13.64	6.06	69.70	63.64	90.91
mode	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00

Initial descriptive analysis of KPI-HEIs data from these 66 data points showed high variance, as indicated by the total number of values showing significant variation between indicators. The largest variance is found in indicator I5 (8670.21) with a standard deviation of 93.11, indicating a very wide data spread. Indicator I8 has the smallest variance and standard deviation (28.79 and 5.37), indicating data that is very consistent and does not vary much. The minimum value of all indicators is 0, indicating the presence of zero data in each indicator. Maximum values range from 28.57 (I8) to 408.37 (I5), indicating outliers or extreme values in these indicators. The range shows the difference between the maximum and minimum values; for example, indicator I5 has the largest range (408.37). Some indicators have a high percentage of zero values, such as I8 with 90.91%, indicating most of the data in this indicator is zero. Overall, the dataset shows significant variation among the various indicators, with some indicators showing highly scattered data and others being highly consistent. Indicators I6, I7, and I8 present a high percentage of zero values. For some indicators that show outliers, the data distribution is significantly around the median value, with some values significantly higher or lower than the median. Seeing this condition, further analysis is needed in addition to the descriptive statistical approach to see the data distribution. This study tries to group universities based on indicators using binning techniques with quartiles. Following the preliminary analysis, the initial step in the quartile binning process is to determine the quartile values for each KPI-HEIs, namely I1 through I8. At this stage, data points for each indicator are sorted from the lowest to the highest value. Subsequently, the first quartile (Q1), second quartile (Q2), and third quartile (Q3) values are calculated for each of these indicators using Formulas 1, 2, and 3. The results of these quartile calculations are then presented in Table 5, which provides details of the quartile values for each KPI-HEIs.

Table 5. Quartile Values of KPI-HEIs

Quartile	I1	I2	I3	I4	I5	I6	I7	I8
Q1	0.00	0.00	13.12	6.99	92.50	0.00	0.00	0.00
Q2	0.00	0.00	20.93	14.52	178.81	0.00	0.00	0.00
Q3	7.43	1.20	32.93	26.25	231.71	12.15	1.64	0.00

Quartile calculations based on 8 indicators from I1 to I8 at 66 PHEIs were performed to analyze data distribution before data clustering. Quartile 1 is represented by (Q1), Quartile 2 is represented by (Q2), and quartile 3 is represented by (Q3). For example, the first quartile (Q1) in indicator I3 is generated from calculating the first quartile (25%) of the data contained in the range of data points in I3, with a Q1 value of 13.12. Similarly, the second quartile (Q2), which is the median quartile (50%) in the range of data points in I2, results in a Q2 value of 20.93, and the third quartile (Q3) is calculated as the third quartile (75%) of the range of data points in I3 with a Q3 value of 32.93. Determination of quartile limits by separating the data into four quartiles, as presented in Table 2. This quartile limit calculation results are then used to group the data into four groups. Determination of higher education groups with the aim of more in-depth analysis of data distribution and facilitate identification of data occupancy quality categories based on quartile values. The distribution of data grouping based on quartiles is shown in Table 6.

Table 6. Distribution of Quartile Clusters Based on KPI-HEIs

PT Code	I1	I2	I3	I4	I5	I6	I7	I8
PT1	4	1	3	4	3	1	4	4
PT2	4	4	3	4	4	3	4	1
PT3	3	3	3	4	3	4	4	1
PT4	1	1	3	3	3	1	1	1
PT5	3	1	1	4	2	1	1	1
PT6	1	1	3	2	3	1	1	1
PT7	4	4	1	4	4	4	4	4
...	...	...	...	...	...	...	...	...
PT63	1	1	4	1	2	4	1	1
PT64	4	1	4	2	1	1	4	1
PT65	1	1	4	4	1	1	1	1
PT66	1	1	4	1	1	1	1	1

The distribution of quartile groups is represented by different colors: quartile group 1 is shown in red, quartile group 2 in brown, quartile group 3 in yellow, and quartile group 4 in green. For instance, this distribution can be observed with the institution code "PT1". Regarding the key performance indicator I1 indicator, PT1 falls into quartile group 4, indicating that the data achievement for this indicator at PT1 is considered very good. Conversely, for the I1 indicator, PT65 is in quartile group 1, indicating that the graduate success rate at PT65 is below expectations. Thus, the quality of its data achievement is deemed poor. After determining the group position of each data point at PHEIs based on the values of the 8 KPI-HEIs indicators, it is not immediately clear which group each institution falls into. To address this, the research developed a method by creating a new quartile from the sum of the quartile positions for each data point at each institution. This step is followed by a new grouping to determine the final group for each institution. Before determining the new quartile, the step taken is to sum the values for each data point at the institutions across the eight indicators. The results of this summation are shown in Figure 3.

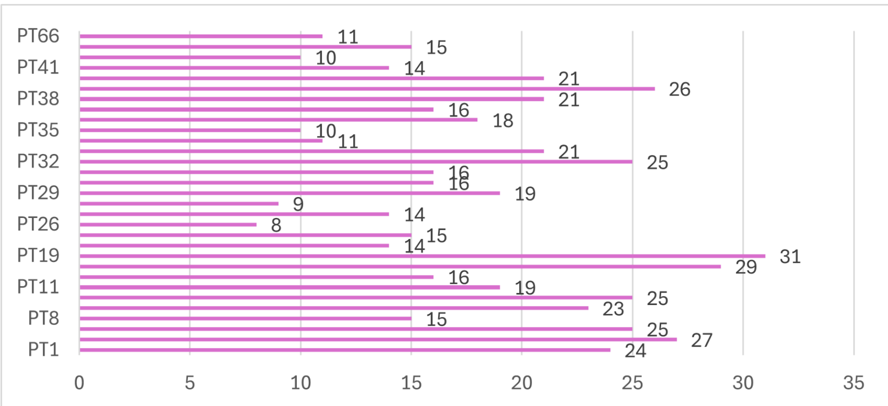

Figure 3. The Summation Results of the KPI-HEIs Group Values

In this diagram, the X axis represents the total number of data point position values for the KPI-HEIs, while the Y axis represents 66 PHEIs. For example, at PT1, the number of data point positions from I1 to I8 is 24. Another example is PT 19, which has several data point positions of 31. After getting the total number of indicators for each PHEIs, then determine the new quartile by using the data from the sum of all indicators at each PHEIs, the results of the new quartile are shown in Table 7. From 66 PHEIs, the following quartile values were obtained: the first quartile (Q1) is 13, the second quartile (Q2) is 15.5, and the third quartile (Q3) is 20.75. These quartile values serve as the basis for clustering PHEIs into four clusters in accordance with the provisions outlined in Table 2. A recapitulation of the clustering results based on KPI-HEIs achievement values is presented in Table 8, with visualization provided in the form of a diagram shown in Figure 4.

Table 7. Distribution of Quartile Clusters Based on KPI-HEIs

Quartile	Score
Q1	13
Q2	15.5
Q3	20.75

Table 8. Recap of PHEIs Clustering Final Results Based on Quartiles

PHEIs Cluster	Number of PHEIs	Description
1	19	Poor
2	14	Fair
3	16	Good
4	17	Very Good

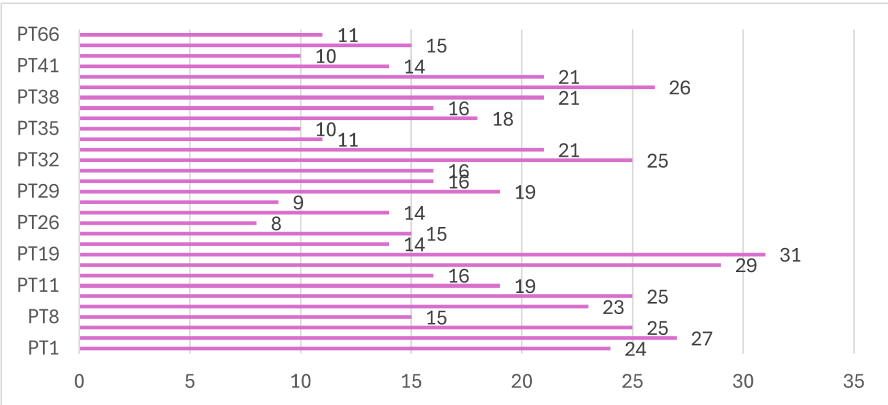

Figure 4. Summary Diagram of Final Clustering Results of PHEIs Based on Group of Quartiles

Cluster 1 consists of 19 PHEIs with the lowest quality among all assessed indicators. Institutions in this cluster show significant weaknesses in achieving KPI-HEIs data. Contributing factors to this low performance may include limited resources, lack of information, ineffective management, or insufficient support. Cluster 2 includes 14 PHEIs that have achieved the KPI-HEIs data milestone with a fair score. Although these institutions have met key performance indicator data requirements, they still need further improvement to reach the good category by enhancing the quality of education and data reporting in higher education databases to meet higher standards. Cluster 3 comprises 16 PHEIs. This cluster shows good achievement in key performance indicator data. Institutions in this cluster demonstrate effective management and adequate support but still require increased innovation in teaching and research collaboration with domestic and international partners to achieve higher categories. Cluster 4 consists of 17 PHEIs that show the best performance in meeting all KPI-HEIs. Institutions in this cluster have adequate resources, highly qualified faculty, complete facilities, and strong participation in research and community service. These institutions can serve as good practice examples for other PHEIs and need special attention from the government and other stakeholders to strengthen capacity and resources to improve overall performance, based on the obtained results. Table 9 shows several comparisons between the findings of this research and previous studies as references.

Table 9. Pembagian data untuk Training dan Testing

Research Tittle	Author	Method	Research Findings
Hierarchical conceptual clustering based on quantile method for identifying microscopic details in distributional data	K.Umbleja, et al, 2020	clustering distributional data based on microscopic similarities using quantile values	Effective Clustering: The proposed algorithm was found to produce more accurate and adequate conceptual clusters compared to traditional methods. Versatility: It effectively handles both complex symbolic data forms like histograms and simpler distributions such as intervals. Comparison Capability: The use of quantiles allows for easy comparison between different types of symbolic data without adding complexity to the method.
Analisis Tingkat Pengetahuan Dan Ansietas Tentang Vaksinasi Covid_19 Pada Kader Kesehatan	S.Nurtanti, et al, 2022	Descriptive quantitative design with a cross-sectional approach	The results showed that 6% of respondents had a high level of knowledge, 10% had a medium level of knowledge, and 84% had a low level of knowledge. Additionally, 64% of respondents experienced severe anxiety, 30% moderate anxiety, and 6% low anxiety.
Analisis Pengaruh Kualitas Pelayanan terhadap Kepuasan Pasien Rawat Inap pada RS. Bahagia Makassar	R.R.S. Baan, 2020	Descriptive statistical analysis	The results showed that the dimensions of service quality, namely tangible, responsiveness, assurance, and empathy, had an effect on increasing patient satisfaction at Bahagia Makassar Hospital, as evidenced by the results of regression and correlation coefficient analysis
Analisis Kebutuhan Mahasiswa terhadap Penggunaan E-Modul pada Perkuliahan Fisika Matematika I Materi Vektor	Auliya Ramadhanti, et al, 2022	Mixed-method with purposive sampling technique	Descriptive statistical results with an average value of 23.07 in the required category show that most samples think that e-modules are a very good idea.
Higher Education Institution Clustering Based on Key Performance Indicators using Quartile Binning Method	V.S. Fatmawaty, et al	Data characterization analysis using the quartile method	This research clusters 66 PHEIs into 4 groups based on Key Performance Indicators data achievement. Cluster 1 consists of 19 PHEIs with the lowest quality, Cluster 2 includes 14 PHEIs that have met milestones but need further improvement, Cluster 3 consists of 16 PHEIs with good performance but require further innovation, and Cluster 4 includes 17 PHEIs with the best performance, which can serve as good practice examples

The clustering of higher education institutions into these clusters reveals a relatively even distribution. This even distribution indicates that the applied assessment and clustering system effectively identifies varying performance levels among institutions. It also suggests that, while many institutions are at very high levels or require significant improvement, some are moving towards better quality. The results of this study show that the quartile method effectively clusters 66 PHEIs based on KPI-HEIs data achievement. The research divides PHEIs into four clusters: Cluster 1, with the lowest quality; Cluster 2, requiring further improvement; Cluster 3, with good performance but needs innovation; and Cluster 4, with the best performance that can serve as a model for best practices. These findings align with previous research that utilized descriptive statistical and quartile methods in data analysis [26] . However, compared to earlier research, this study emphasizes the specific advantages of using quartile binning for clustering in the analysis of educational quality achievement, offering a more detailed perspective on quality distribution and identifying potential areas for improvement in higher education institutions. These findings are consistent with previous studies that used quartile data analysis methods. However, compared to earlier research, this study emphasizes the specific advantages of using quartile binning for clustering in the analysis of educational quality achievement, offering a more detailed perspective on quality distribution and identifying potential areas for improvement in higher education institutions.

4. CONCLUSION

The research results show that data clustering using the quartile binning method, analyzing distribution characteristics and KPI-HEIs data distribution, produced evenly distributed clusters of PHEIs in the HESI Region V Yogyakarta. Cluster 1 includes 19 PHEIs with poor KPI-HEIs data achievement due to low indicator values and many zero values, particularly in indicators I6, I7, and I8. Cluster 2 consists of 14 PHEIs with fair KPI-HEIs data achievement, although there are still significant variations and zero values. Cluster 3 comprises 14 PHEIs with good KPI-HEIs data achievement, featuring higher indicator values and fewer zero values, although outliers and variations still need improvement. Cluster 4 includes 17 PHEIs with very good KPI-HEIs data achievement and can serve as examples for other institutions. Clustering data based on the analysis of distribution characteristics and the distribution of KPI-HEIs data provides a clear view of how data is distributed and how higher education institutions perform in achieving KPI-HEIs targets. Further evaluation and corrective actions can be focused on institutions in clusters with low KPI-HEIs achievement to improve overall quality and consistency. By understanding the position of each institution within these clusters, the HESI Region V Yogyakarta and PHEIs can design coaching and support programs tailored to the specific needs of each group. Moreover, these results promote continuous improvement across the higher education sector, thereby enhancing the quality of education and services offered by institutions in the future.

5. ACKNOWLEDGEMENTS

This study was supported by the Ministry of Education, Culture, Research, and Technology (Ministry of Education, Culture, Research, and Technology) through an excellent scholarship with number 0379/J5.2.2/BP/PKS/X/2023 and LLDIKTI Region V, and the Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta.

6. DECLARATIONS

AUTHOR CONTIBUTION

The first author contributed to designing the overall outline of the study and seeking more knowledge about descriptive analysis methods using a descriptive statistical approach. The second author provided feedback on the journal content, and the third author provided feedback on the journal writing.

FUNDING STATEMENT

-

COMPETING INTEREST

The author declares that there are no conflicts or interests between the editors and reviewers regarding the publication of this article.

REFERENCES

[1] N. Nugraha, Y. T. Prasetyo, H. Sugiharti, I. Lhutfi, A. Widyaningsih, A. Triantoro, A. K. S. Ong, M. N. Young, S. F. Persada, R. A. Kristamtomo Putra, and R. Nadlifatin, "Quality Assurance in Higher Educational Institutions: Empirical Evidence in Indonesia," *SAGE Open*, vol. 13, no. 4, pp. 1–12, 2023, https://doi.org/10.1177/21582440231203060.

[2] M. Fadhli, "Sistem Penjaminan Mutu Internal Dan Ekstenal Pada Lembaga Pendidikan Tinggi," *AL-TANZIM: Jurnal Manajemen Pendidikan Islam*, vol. 4, no. 2, pp. 53–65, sep 2020, https://doi.org/10.33650/al-tanzim.v4i2.1148.

[3] A. Taryana, A. Fadli, and S. R. Nurshiami, "Merancang Perangkat Lunak Sistem Penjaminan Mutu Internal (SPMI) Perguruan Tinggi yang Memiliki Daya Adaptasi Terhadap Perubahan Kebutuhan Pengguna secara Cepat dan Sering," *Jurnal Al-Azhar Indonesia Seri Sains Dan Teknologi*, vol. 5, no. 3, April, pp. 121–122, 2020, https://doi.org/10.36722/sst.v5i3.372.

[4] V. S. Fatmawaty, I. Riadi, and H. Herman, "Klasterisasi Perguruan Tinggi LLDIKTI V Berdasarkan Indikator Kinerja Utama dan PDDIKTI Menggunakan K-Means Clustering," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 878–888, 2024, https://doi.org/10.30865/mib.v8i2.7497.

[5] I. Riadi, R. Umar, and R. Anggara, "Comparative Analysis of Naive Bayes and K-NN Approaches to Predict Timely Graduation using Academic History," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1163–1174, 2024, https://doi.org/10.12785/ijcds/160185.

[6] K. Umbleja, M. Ichino, and H. Yaguchi, "Hierarchical conceptual clustering based on quantile method for identifying microscopic details in distributional data," *Springer : Advances in Data Analysis and Classification*, vol. 15, no. 2, pp. 407–436, 2020, https://doi.org/10.1007/s11634-020-00411-w.

[7] S. Nurtanti and P. H. Husna, "Analisis Tingkat Pengetahuan Dan Ansietas Tentang Vaksinasi Covid_19 Pada Kader Kesehatan," *Jurnal Ilmu Keperawatan Jiwa*, vol. 53, no. 9, pp. 1689–1699, 2022, https://doi.org/10.32584/jikj.v5i1.

[8] R. R. S. Baan, "Analisis Pengaruh Kualitas Pelayanan terhadap Kepuasan Pasien Rawat Inap pada RS. Bahagia Makassar," *Ekombis Sains: Jurnal Ekonomi, Keuangan dan Bisnis*, vol. 5, no. 1, pp. 45–52, 2020, https://doi.org/10.24967/ekombis.v5i1.591.

[9] Auliya Ramadhanti, A. Astalini, and D. Darmaji, "Analisis Kebutuhan Mahasiswa terhadap Penggunaan E-Modul pada Perkuliahan Fisika Matematika I Materi Vektor," *Jurnal Pendidikan Mipa*, vol. 12, no. 1, pp. 13–22, 2022, https://doi.org/10.37630/jpm.v12i1.528.

[10] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," *Decision Analytics Journal*, vol. 6, no. 6, pp. 10–22, 2023, https://doi.org/10.1016/j.dajour.2023.100164.

[11] E. D. B. Sudewo, M. K. Biddinika, and A. Fadlil, "DenseNet Architecture for Efficient and Accurate Recognition of Javanese Script Hanacaraka Character," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 453–464, 2024, https://doi.org/10.30812/matrik.v23i2.3855.

[12] I. Riadi, A. Yudhana, and M. R. Djou, "Optimization of Population Document Services in Villages using Naive Bayes and k-NN Method," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 127–138, 2024, https://doi.org/10.12785/ijcds/150111.

[13] D. Faroek, Rusydi Umar, and Imam Riadi, "Classification Based on Machine Learning Methods for Identification of Image Matching Achievements," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 198–206, 2022, https://doi.org/10.29207/resti.v6i2.3826.

[14] V. Magar, D. Ruikar, S. Bhoite, and R. Mente, "Innovative Inter Quartile Range-based Outlier Detection and Removal Technique for Teaching Staff Performance Feedback Analysis," *Journal of Engineering Education Transformations*, vol. 37, no. 3, pp. 176–184, 2024, https://doi.org/10.16920/jeet/2024/v37i3/24013.

[15] T. Z. Dessiaming, S. Anraeni, and S. Pomalingo, "College Academic Data Analysis Using Data Visualization," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, pp. 1203–1212, 2022, https://doi.org/10.20884/1.jutif.2022.3.5.310.

[16] M. Wahyuni, *Statistik Deskriptif Untuk Penelitian Olah Data Manual dan SPSS versi 25*. Bintang Pustaka Madani Yogyakarta, 2020, no. 1.

[17] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using VGG19: Caltech-101 image data set," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3609–3620, 2023, https://doi.org/10.1007/s12652-021-03488-z.

[18] I. Riadi, H. Herman, F. Fitriah, and S. Suprihatin, "Optimizing Inventory with Frequent Pattern Growth Algorithm for Small and Medium Enterprises," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 1, pp. 169–182, 2023, https://doi.org/10.30812/matrik.v23i1.3363.

[19] G. Navas-Palencia, "Optimal binning: mathematical programming formulation," *arvix*, vol. 3, no. 8, pp. 1–22, 2022, https://doi.org/10.48550/arXiv.2001.08025.

[20] O. R. Y. Pradana, "Menentukan Letak Kuartil Pada Data Tunggal Menggunakan Metode Penjumlahan Satu," *Indonesia Journal of Education and Learning Mathematics*, vol. 1, no. 1, pp. 6–16, 2020.

[21] R. Prieto Curiel, C. Cabrera Arnau, M. Torres Pinedo, H. González Ramírez, and S. R. Bishop, "A continuous binning for discrete, sparse and concentrated observations," *MethodsX*, vol. 7, no. 7, pp. 1–12, 2020, https://doi.org/10.1016/j.mex.2019.10.020.

[22] D. Vinsensia, S. Amri, J. Sihotang, and H. T. Sihotang, "New Method for Identification and Response to Infectious Disease Patterns Based on Comprehensive Health Service Data," vol. 23, no. 3, pp. 583–592, 2024, https://doi.org/10.30812/matrik.v23i3.4000.

[23] M. Waskom, "Seaborn: Statistical Data Visualization," *Journal of Open Source Software*, vol. 6, no. 6, p. 3021, 2021, https://doi.org/10.21105/joss.03021.

[24] S. Rawat, A. Rawat, D. Kumar, and A. S. Sabitha, "Application of machine learning and data visualization techniques for decision support in the insurance sector," *International Journal of Information Management Data Insights*, vol. 1, no. 2, pp. 1–12, 2021, https://doi.org/10.1016/j.jjimei.2021.100012.

[25] A. Fadlil, I. Riadi, and F. Andrianto, "Improving Sentiment Analysis in Digital Marketplaces through SVM Kernel Fine-Tuning," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 159–171, 2024, https://doi.org/10.12785/ijcds/160113.

[26] T. Jambudi and S. Gandhi, "An Effective Initialization Method Based on Quartiles for the K-means Algorithm," *Indian Journal Of Science And Technology*, vol. 15, no. 35, pp. 1712–1721, 2022, https://doi.org/10.17485/ijst/v15i35.714.

[This page intentionally left blank.]