International Journal of Computer Sciences and Mathematics Engineering Vol. 1 No. 1 (2021)
Published by: CV. HAWARI

International Journal of Computer Sciences and Mathematics Engineering
Journal homepage: www.ijecom.org

Analysis Of Increasing Student Service Satisfaction Using K-Means Clustering Algorithm and Gaussian Mixture Models (GMM)

Adek Maulidya[1], Khairul[2], Zulham Sitorus[3] , Andysah Putera Utama Siahaan[4], Muhammad Iqbal[5]
[1,2,3,4,5]Department of Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

Article Info

Article history:

Received May 3, 2024
Revised May , 2024
Accepted May , 2024

Keywords:

Student
Service
K-Means Clustering
Gaussian Mixture Model

ABSTRACT

This research analyzes the comparison between two cluster analysis algorithms, namely K-Means Clustering and Gaussian Mixture Model (GMM), to gain a deeper understanding of data structure and model suitability. The results of the analysis show that the silhouette score value from using the K-Means algorithm is 0.44528, indicating relatively good cluster grouping, while the use of the Gaussian Mixture Model produces a silhouette score value of -0.500119, indicating a mismatch between the data points in the cluster and the probability overlap between clusters. Therefore, the conclusion states that based on the silhouette score value, using the K-Means Clustering algorithm is better because it produces better and more cohesive cluster grouping. The results of this analysis are that campuses can use this information to understand student needs more effectively and take appropriate corrective steps.

Corresponding Author:

Adek Maulidya
Master Of Information Technology
Universitas Pembangunan Panca Budi
Jl.Gatot Subroto Km.4,5 Sei Sikambing, Kota Medan, Indonesia.
Email: adek.maulidya@gmail.com

1. INTRODUCTION

An educational institution confirms its commitment to providing the best service to students, realizing that the quality of service affects the learning environment and the interests of prospective students. However, despite regularly collecting input through evaluation questionnaires, institutions often fail to take concrete action based on the results of these questionnaires. This has resulted in stagnation in the quality of services and facilities, as well as a decrease in the number of active students each semester. To overcome this problem, research was conducted to increase student service

E-ISSN 2962-4274

satisfaction by applying the K-Means Clustering algorithm and Gaussian Mixture Models. It is hoped that this analysis will help institutions respond to student needs more quickly and effectively, as well as determine improvement priorities and design more targeted strategies.

Through the use of the K-Means Clustering algorithm and Gaussian Mixture Models, this research aims to provide in-depth insight into the patterns and structure of student responses to institutional services. K-Means Clustering is expected to identify groups of students with similar levels of satisfaction, while Gaussian Mixture Models will provide an understanding of the probability distribution of the data. Thus, it is hoped that the results of this research will help educational institutions to more effectively adapt their services to student needs, increase student retention, and attract the interest of prospective new students, which in turn will strengthen the institution's reputation as a quality education provider.

2. RESEARCH METHOD

This study employs a quantitative approach to analyze student satisfaction with campus services. Utilizing a combination of the K-Means Clustering and Gaussian Mixture Model algorithms, the research focuses on identifying patterns and clusters of student satisfaction.

a) Sample and Research Method :

This study employs a quantitative research method titled "Analysis of Student Satisfaction Using K-Means Clustering and Gaussian Mixture Model Algorithms." The research sample consists of 1000 student data.

b) Data Collection Method :

Data were collected through questionnaires completed by students at the end of the even semester of the Academic Year 2022/2023. The questionnaires were made available through the university's official academic administration portal at www.sia.kaputama.ac.id.

c) Variables and Questions :

The study considers 17 attributes or variables, each with specific questions answered by students. These questions are designed to evaluate various aspects of campus services and student experiences.

d) Measurement Method :

The analysis of student satisfaction levels utilizes the Likert scale as a measurement method. This scale is used to assess the level of student satisfaction with campus services across predetermined aspects. The following are the Likert scale values from the students that will be used in this research :

Table. 1 Linkert Scale

Kriteria Jawaban Kuesioner Mahasiswa	Skor
Sangat Puas	5
Puas	4
Kurang Puas	3
Tidak Puas	2
Sangat Tidak Puas	1

e) Data Analysis :

This data analysis involves assessing students' evaluation of the services and facilities provided within the campus environment. The university requires students to provide their perspectives and feedback at the end of each semester regarding their experiences with various aspects of campus life. These questions are carefully crafted to obtain comprehensive insights into students' needs and preferences, as well as to contribute to the enhancement of the quality of services and facilities provided by the institution.

E-ISSN 2962-4274

Table. 2 Student Questionnaire Table

Aspek ke	**Keterangan**
Aspek 1	Kenyamanan ruang kuliah teori
Aspek 2	Kenyamanan dan ketersediaan Laboratorium-1
Aspek 3	Kenyamanan dan ketersediaan Laboratorium-2
Aspek 4	Kenyamanan dan ketersediaan Laboratorium-3
Aspek 5	Kelengkapan alat praktik untuk praktikum
Aspek 6	Pelayanan bagian laboratorium (Asisten LAB dan Kepala
Aspek 7	Pelayanan bagian perpustakaan
Aspek 8	Jumlah dan mutu buku di perpustakaan
Aspek 9	Pelayanan bagian keuangan
Aspek 10	Pelayanan administrasi akademik
Aspek 11	Pelayanan bimbingan akademik (Dosen penasehat
Aspek 12	Pelayanan keamanan kampus
Aspek 13	Kebersihan lingkungan kampus
Aspek 14	Kenyamanan dan ketersediaan sarana ibadah
Aspek 15	Ketersediaan sarana olahraga
Aspek 16	Cakupan dan bandwidth hotspot (wifi) kampus
Aspek 17	Pelayanan sistem informasi

e) Dataset :
The dataset utilized comprises student assessments, as presented in Table 3 below :

Table. 3 Dataset Student Assessment Questionnaire

MHS	Aspek 1	Aspek 2	Aspek 3	Aspek 4	Aspek 5	Aspek 6	Aspek 7	Aspek 8	Aspek 9	Aspek 10	Aspek 11	Aspek 12	Aspek 13	Aspek 14	Aspek 51	Aspek 16	Aspek 17
1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3	2	2	3	3	3	5	5	5	2	3	2	2	3	4	2	1	4
4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6	4	1	3	3	4	4	4	4	5	4	4	4	4	4	4	4	4
7	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
8	4	1	1	1	1	4	5	2	2	4	5	5	2	3	1	2	4
9	4	3	3	3	2	3	3	3	4	4	3	4	3	3	1	4	4
10	2	1	1	1	1	1	2	2	2	2	3	3	3	3	1	2	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
999	3	1	1	1	1	3	4	4	4	4	4	2	4	2	2	2	4
1000	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

E-ISSN 2962-4274

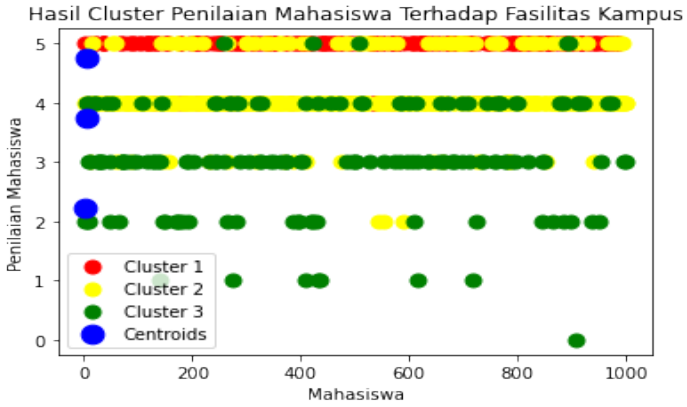
3. RESULTS AND DISCUSSIONS

In this study, an analysis was conducted on the level of service satisfaction among students using the K-Means Clustering and Gaussian Mixture Model (GMM) algorithms. The analysis results indicated that the service satisfaction data were successfully grouped into several clusters through K-Means Clustering, with priority-1, priority-2, and priority-3 clusters. Additionally, specific characteristics of each cluster could be identified, such as the responsiveness of students to services or the need for special attention in clusters with low satisfaction.

3.1. Testing the K-Means Clustering Algorithm

After the 1000 datasets above are processed using the Python programming language and processed using predefined source code, the clustering results are divided into three clusters based on the closest distance from the centroid points. This data includes Student (X) and Student Assessment (Y), as seen in the following image **:**

Figure 1. Cluster Results from the K-Means Clustering Algorithm



The explanation of the cluster results above is :

1. The number obtained for cluster 1 is 156 students. This is the cluster with the smallest number, thus it can be concluded that cluster 1 is **Priority 3 (P3)**, meaning that improvements to campus facilities can be postponed in this cluster. The aspects to be improved are: Aspects 5, 2, 16, 15, and Aspect 4.
2. The number obtained for cluster 2 is 573 students. This is the cluster with the highest number, thus it can be concluded that cluster 2 is **Priority 1 (P1)**, meaning that improvements to campus facilities must be made immediately in this cluster. The aspects that need immediate improvement are: Aspects 5, 15, 2, 16, and Aspect 4.
3. The number obtained for cluster 3 is 271 students. This is a cluster with a moderate number, thus it can be concluded that cluster 3 is **Priority 2 (P2)**, meaning that improvements to campus facilities can be postponed until next month in this cluster. The aspects to be improved are: Aspects 5, 2, 15, 16, and Aspect 3.

From the results of the K-Means Clustering algorithm above, we will now calculate the silhouette score. The silhouette score is an evaluation metric used in cluster analysis to assess how well a clustering or cluster has been formed. Its function is to measure how close each data point in a cluster is to other data points in the same cluster, compared to data points in other clusters. With the formula as follows:

$$s_i = \frac{b_i - a_i}{max(b_i, a_i)}$$

E-ISSN 2962-4274

From the usage of the K-Means Clustering algorithm, it is found that the silhouette score value of 0,44528 indicates that the clustering of clusters is relatively good, with data points close to the cluster they belong to and sufficiently separated from other clusters.

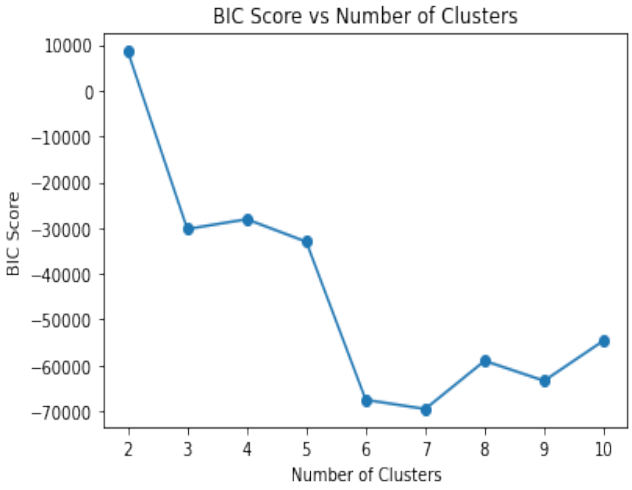
3.2. Testing the Gaussian Mixture Model Algorithm

At this stage, the results of applying the Gaussian Mixture Model (GMM) algorithm to the dataset used will be discussed. By using the GMM algorithm, you can group data into several clusters that represent different patterns in the dataset. This discussion will provide insight into how the data is structured and how the GMM algorithm interprets these patterns.

The initial steps that must be taken to calculate the Gaussian mixture model algorithm are :

1. Calculate the number of Gaussian components, namely by determining the number of Gaussian components, k, that will be used in the model. Determining the Bayesian Information Criterion (BIC) is used to determine the most appropriate number of clusters. BIC is often used to select the optimal number of clusters in grouping models such as GMM (Gaussian Mixture Model). BIC selects the model with the lowest value. Therefore, the model with the lowest BIC value is considered the model that best fits the data. With the Python programming language, the Bayesian Information Criterion (BIC) is obtained as follows :

Figure 2. Bayesian Information Criterion (BIC) values



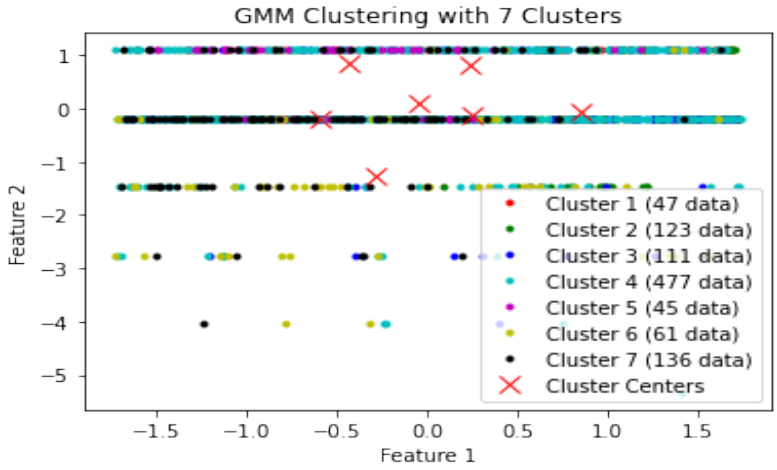
This graph illustrates the results of Bayesian Information Criterion (BIC) scores for various numbers of clusters (k) in cluster analysis. In this graph, the BIC value significantly decreases as the number of clusters increases from k=2 to k=7, and then stabilizes or starts to increase slightly afterward.

At the point k=7, the graph shows the lowest BIC value, indicating that the model with seven clusters provides optimal explanation of the data structure. This is supported by the understanding that the decrease in BIC reflects an improvement in the quality of the model with a higher number of clusters.

2. The Gaussian process aims to generate the best GMM model according to the observed data. This model will provide an accurate representation of the patterns present in the dataset, and the resulting clustering can be used to gain a deeper understanding of its structure. The output of the Gaussian Mixture Model algorithm using the Python programming language is as follows :

E-ISSN 2962-4274

Figure 3. Cluster Results from the Gaussian Mixture Model Algorithm



From the clustering results obtained using the Gaussian Mixture Model algorithm, it was found that the data can be divided into seven clusters. Each cluster has a different number of students with varying data distributions, as follows :
Cluster 1 : Consists of 47 students.
Cluster 2 : Consists of 123 students.
Cluster 3 : Consists of 111 students.
Cluster 4 : Consists of 477 students.
Cluster 5 : Consists of 45 students.
Cluster 6 : Consists of 61 students.
Cluster 7 : Consists of 136 students.
With these clustering results, we can observe variations in the number of students in each cluster, indicating that each cluster may represent groups of students with different characteristics.

From the results of the Gaussian Mixture Model above, we will now calculate the silhouette score. The silhouette score is an evaluation metric used in cluster analysis to assess how well a clustering or cluster has been formed. Its function is to measure how close each data point in a cluster is to other data points in the same cluster, compared to data points in other clusters. With the formula as follows:

$$s_i = \frac{b_i - a_i}{max(b_i, a_i)}$$

From the use of the Gaussian Mixture Model (GMM) algorithm, it was found that the silhouette score value was -0,500119. This negative value indicates a mismatch between data points in a cluster and other clusters, and the possibility of overlap between clusters.

4. CONCLUSION

The cluster analysis results from both K-Means Clustering and Gaussian Mixture Model (GMM) provide insights into the distribution of students' preferences regarding campus facility improvements. Cluster 1, with 156 students, indicates a lower priority for enhancements, suggesting that improvements can be postponed. Conversely, Cluster 2, comprising 573 students, emphasizes an urgent need for immediate enhancements. Lastly, Cluster 3, encompassing 271 students, suggests a

E-ISSN 2962-4274

moderate priority for improvements. Identified aspects for enhancement across clusters include various dimensions of campus facilities.

Additionally, the silhouette score evaluation reveals interesting findings. While the silhouette score of 0.44528 from K-Means Clustering suggests satisfactory clustering performance, the negative silhouette score of -0.500119 from the Gaussian Mixture Model indicates potential discrepancies and overlap between clusters. This highlights the need for further refinement of clustering techniques to enhance accuracy and address clustering discrepancies identified by the Gaussian Mixture Model.

In conclusion, the clustering analysis offers valuable insights into students' preferences for campus facility improvements. The findings underscore the importance of prioritizing enhancements based on student preferences to effectively allocate resources. Furthermore, the evaluation of silhouette scores highlights the need for continued refinement and exploration of clustering techniques to ensure accurate and meaningful cluster assignments. Future research endeavors may focus on refining clustering algorithms and exploring additional factors to enhance the accuracy and robustness of clustering analyses in the context of campus facility management and student satisfaction.

REFERENCES

[1] Ahmad, I. A., Al-Nayar, M. M. J., & Mahmood, A. M. (2023). A comparative study of Gaussian mixture algorithm and K-means algorithm for efficient energy clustering in MWSN. Bulletin of Electrical Engineering and Informatics, 12(6), 3727–3735.

[2] Ajimotokan, H. A. (n.d.). Research Techniques Qualitative, Quantitative and Mixed Methods Approaches for Engineers.

[3] Banachewicz, K., & Massaron, L. (n.d.). The Kaggle book : data analysis and machine learning for competitive data science.

[4] Bahri, S. (2018). Metodologi Penelitian Bisnis Lengkap dengan Teknik Pengolahan Data SPSS. Yogyakarta : CV ANDI OFFSET.

[5] Borishade, T. T., Ogunnaike, O. O., Salau, O., Motilewa, B. D., & Dirisu, J. I. (2021). Assessing the relationship among service quality, student satisfaction and loyalty: the NIGERIAN higher education experience. In Heliyon (Vol. 7, Issue 7). Elsevier Ltd.

[6] Burk, S., & Miner, G. D. (n.d.). It's All Analytics!: The Foundations of AI, Big Data, and Data Science Landscape for Professionals in Healthcare, Business, and Government.

[7] Chhabra, A., Masalkovaite, K., & Mohapatra, P. (2021). An Overview of Fairness in Clustering. IEEE Access, 9, 130698–130720.

[8] Clustering Harga Rumah: Perbandingan Model K-Means dan Gaussian Mixture Model. (n.d.).

[9] Parlambang, B., & Fauziah. (2020). IMPLEMENTASI ALGORITMA K-MEANS DALAM PROSES PENILAIAN KUESIONER KEPADA DOSEN GUNA MENDUKUNG KEPUASAN MAHASISWA TERHADAP DOSEN. Jurnal Ilmiah Teknologi Dan Rekayasa, 25(2), 161–173.

[10] Rohman, A., & Rochcham, ; Muhammad. (2020). Implementasi Algoritma K-Means Untuk Clustering Kepuasan Mahasiswa Terhadap Pelayanan Akademik (Implementation of the K-Means Algorithm for Clustering Student Satisfaction on Academic Services) (Vol. 6, Issue 2).