JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALYSING STUDENT MENTAL HEALTH THROUGH K-MEANS CLUSTERING AND MULTI-STAGE SAMPLING METHODS

Rahmat Hidayat[1*], Dede Pratama[2]
[1]TVRI Stasiun Sumatera Barat, Kementrian Komunikasi Digital
[2]Master of informatics engineering, UPI YPTK Padang
email: 048409024@ecampus.ut.ac.id

Abstract: Mental health is an essential aspect of overall well-being, particularly for university students vulnerable to emotional strain. This study aims to identify clusters of student mental health trends using the K-Means clustering technique. The research involved 60 students from four academic programs at the Faculty of Science and Technology, selected using stratified and cluster sampling techniques. Data were collected using a modified Mental Health Inventory (MHI). The results revealed distinct commonalities among majors: the Statistics program was predominantly defined by the depressed cluster at 53.3%, while Mathematics followed at 40% within the same cluster. In contrast, Biology students predominantly fell under the neutral/stable cluster (66.7%), whilst Information Systems students exhibited an even distribution (33.3% per cluster) without a dominant trend. The clustering quality was evaluated using the Silhouette Coefficient, yielding a range of 0.39 to 0.60. Biology (0.60) and Statistics (0.54) exhibited a reasonable structure, but Information Systems (0.39) and Mathematics (0.34) demonstrated a deficient structure. In conclusion, K-Means effectively discerns mental health patterns, providing a data-driven basis for targeted psychological interventions in educational settings.

Keywords: biology; information systems; k-means; mental health; silhouette coefficient.

Abstrak: Kesehatan mental merupakan komponen vital dari kesejahteraan total, terutama bagi maha-siswa yang rentan terhadap stres emosional. Penelitian ini bertujuan untuk mengidentifikasi kelompok tren kesehatan mental mahasiswa melalui penerapan metode pengelompokan K-Means. Studi ini mencakup 60 mahasiswa dari empat program studi di Fakultas Sains dan Teknologi, yang dipilih melalui metode pengambilan sampel bertingkat dan kelompok. Data dikumpulkan dengan menggunakan Inventaris Kesehatan Mental (MHI) yang dimodifikasi. Temuan menunjukkan kesamaan yang jelas di antara jurusan: program studi Statistika terutama ditandai oleh kelompok depresi (53,3%), diikuti oleh Matematika dengan 40% dalam kelompok depresi. Sebaliknya, mahasiswa Biologi terutama termasuk dalam kelompok netral/stabil (66,7%), sedangkan mahasiswa Sistem Informasi memiliki distribusi yang merata (33,3% per kelompok) tanpa pola yang dominan. Kualitas pengelompokan dinilai dengan Koefisien Sil-houette, menghasilkan rentang 0,39 hingga 0,60. Biologi (0,60) dan Statistika (0,54) memiliki struktur sedang, sedangkan Sistem Informasi (0,39) dan Matematika (0,34) menunjukkan struktur yang buruk. Kesimpulannya, K-Means secara akurat mengidentifikasi tren kesehatan mental, menawarkan landasan berbasis data untuk terapi psikologis yang ditargetkan di ling-kungan pendidikan.

Kata kunci: biologi; kesehatan mental; K-Means; silhouette coefficient; sistem informasi.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Mental health is a crucial component of overall health, contributing to phys-ical well-being. It is defined as a situation that promotes individuals' physical, mental, spiritual, and social growth, allowing them to achieve their potential, contribute to so-ciety, and manage environmental challeng-es [1]. College students, transitioning from late adolescence to early adulthood, are particularly vulnerable to emotional stress arising from social, intellectual, and per-sonal challenges. If unaddressed, this stress often manifests as Mental Emotional Disorders (MDDs), including depression, anxi-ety, and different psychiatric comorbidities.

The significance of this issue is underscored by World Health Organisa-tion (WHO) statistics, which reveal that 3.8% of the global population—approximately 280 million individuals—suffer from depression [2]. Although these statistics reflect the broader popu-lation, they are a vital metric for higher education institutions, as the academic milieu frequently intensifies these stressors.

Additionally, in Indonesia, the Ministry of Health said that 6.1% of individuals aged 15 and above suffered depression. In East Java, Basic Health Research (Riskesdas) recognised the province as having the twelfth highest frequency of severe mental problems. Furthermore, data from the Indonesian Psychiatric Association (PDSKJI) indi-cates a rising trend in mental health dis-orders, with screening rates increasing by 11.8% from 2020 to 2022 [3]. While these values reflect the broader popu-lace, they strongly indicate an increasing necessity to assess mental health among students, a demographic that belongs to this high-risk age group yet frequently lacks targeted, representative data.

The initial assessment of psychological tests is essential for identifying behavioural patterns, cognitive abilities, and problem-solving characteristics in students [4]. These assessments serve as a preventive measure against severe mental disorders. Conventional screen-ing procedures sometimes struggle to effectively manage large, diverse student populations. As a result, data mining techniques, particularly clustering, have shown to be powerful instruments. Clustering organises data with similar characteristics into distinct categories without predetermined criteria [5], [10], [14].

A multitude of research have employed this methodology in analo-gous fields. Timothy et al. examined student mental health from 2017 to 2020 via K-Means Clustering, resulting in a robust structure with an average silhou-ette score ranging from 0.49 to 0.63 [7]. Likewise, Angelina et al. examined the anxiety levels of academics during the Covid-19 outbreak utilising K-Means, attaining 99% accuracy [8]. Although these studies illustrate the method's efficacy, many depended on rudimentary sampling techniques or constrained psychological variables. A more stringent methodology is required to guarantee data representativeness in a diverse stu-dent group [15] [16].

This study seeks to fill these gaps by employing the K-Means clustering tech-nique to assess mental health trends among students in the Faculty of Science and Technology. This research introduces a du-al sampling methodology that merges strat-ified and one-stage cluster sampling with K-Means clustering analysis applied to psychological evaluation data.

This strate-gy gives a more equal data representation compared to the basic sample techniques employed in previous

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

studies and offerK-Means Clusterings a comprehensive per-spective on student mental health trends in the post-pandemic context. The resultant clusters are further assessed by the Silhou-ette Coefficient to ascertain the precision and legitimacy of the results, offering a sol-id data-driven foundation for educational institutions to develop targeted psychologi-cal interven-tion programs.

METHOD

Identifying Mental Health Problems

This study employed a validated 10-item Mental Health Inventory (MHI) on a five-point Likert scale to assess both psychological well-being and distress, ensuring reliable measurement of positive and negative mental health perspectives.

Table 1. Psychological well-being questions

No.	Psychological Wellbeing (x_1)
1	My life right now is full of interesting things.
2	I feel comfortable communicating with the people around me.
3	I feel valued because I am treated well by my friends.
4	I feel happy living my life.
5	I enjoy what is happening in my life at this moment.

Table 2. Psychological stress questions

No.	Psychological Distress (x_2)
1	I am currently in a phase of confusion or frustration.
2	I currently feel exhausted or helpless.
3	I am currently at my lowest point.
4	I am currently losing control over my thoughts, feelings, and behavior.
5	I feel as though I have nothing to look forward to in the future.

Stratified Sampling Technique

The study was conducted at the Faculty of Science and Technology, Universitas Terbuka, involving 60 students selected through stratified sampling from four study programs, with 15 participants from each cohort.

$$n=\frac{\sum_{i=1}^{L} N_i p_i q_i}{ND+\frac{1}{N}\sum_{i=1}^{L} N_i p_i q_i} \quad (1)$$

Where,

$$D=\frac{B^2}{4} \quad (2)$$

$$n_i=n\left(\frac{N_i}{\sum_{k=1}^{L} N_k}\right)=n\left(\frac{N_i}{N}\right) \quad (3)$$

Description:
n : the population.
B : the estimation error,
P : the estimated proportion.
q : the subtraction of 1 from the *p*-value.

Table 3. Results of stratified Sampling calculations

No.	Year	Calculation (n=4)	Sample Result (ni)
1	2020	n=4	1
2	2021	n=4	1
3	2022	n=4	1
4	2023	n=4	1
Total		4	

One-Stage Cluster Sampling Technique

One study program sample was obtained from each stratum, so random sampling was conducted using a random table for the seven study programs, resulting in the sample as shown in Table 4.

Table 4. Results of study program sampling

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

No.	Batch	Major
1	2020	Information System
2	2021	Statistics
3	2022	Mathematics
4	2023	Biology

Two-Stage Cluster Sampling Technique

Based on the results of the first-stage cluster, we will continue with sampling of students in each study program selected using equations (4) and (5).

$$n=\frac{Npq}{(N-1)D+pq} \quad (4)$$

Where,

$$D=\frac{B^2}{4} \quad (5)$$

Table 5. Results of student sampling

Stratum	Calculation (n=(N−1)D+pqNpq)
1	(66−1)(0.013)+(0.5×0.5) 66(0.5×0.5) = 15.06
2	(62−1)(0.013)+(0.5×0.5) 62(0.5×0.5) = 14.86
3	(69−1)(0.013)+(0.5×0.5) 69(0.5×0.5) = 15.30
4	(66−1)(0.013)+(0.5×0.5) 66(0.5×0.5) = 15.06
Total	60.18

The sample size chart for each stratum is shown in Image 1.

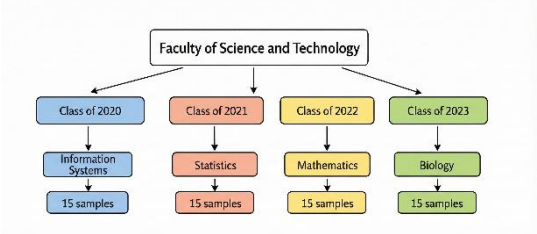

Image 1. Sample Distribution Chart

K-Means Clustering

The K-Means Clustering method is one type of clustering method used to separate or group data into several groups using a partition system [9]. The K-Means Clustering method divides data into similar characteristics or details into one cluster, while data with different specifications or characteristics are placed in another cluster [10].

Silhouette Coefficient

The Silhouette coefficient is one method for evaluating or validating cluster results. Validation will yield the partition that best fits the data [11]. The Silhouette coefficient aims to evaluate or retest the placement of each object within a cluster by comparing the average distance between centroids within one cluster and the distance between centroids within other clusters [12]. The following is the formula for the Silhouette coefficient [13]:

$$si=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{k(i)-l(i)}{\max\{k(i),l(i)\}}\right) \quad (6)$$

where, $k(i)$ is the average distance between sample i and other samples in the cluster, $l(i)$ is the minimum distance between samples i and other clusters, and n is the number of data points. The following is a Kaufman table for measuring accuracy [17].

Table 6. Silhouette coefficient accuracy scale

Silhouette Index (si) Range	Interpretation / Description
0.700<si≤1	Strong structure
0.500<si≤0.700	Reasonable (Medium) structure
0.250<si≤0.500	Weak structure
si≤0.250	No substantial structure

Analysis Method

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

The research begins with scenario creation and survey data collection, followed by descriptive analysis. The data is then grouped using K-Means Clustering and evaluated via the Silhouette Coefficient, concluding with a summary of the results as shown in the flowchart.

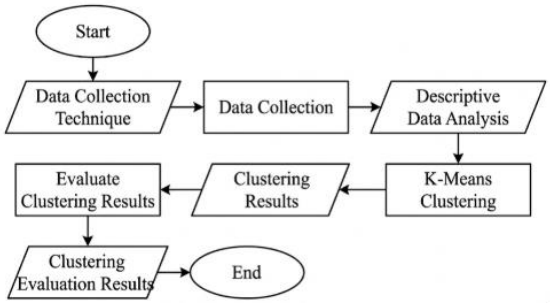

Image 2. Research Flowchar

RESULTS AND DISCUSSION

K-Means Clustering Results

This study utilizes K-Means Clustering to categorize data into three groups: depression, neutral, and happiness (*k*=3). The algorithm iteratively calculates Euclidean distances [23] to assign data to the nearest randomly initialized centroids, then updates these centers [24] until the cluster membership stabilizes and remains unchanged.

Information Systems Study Program

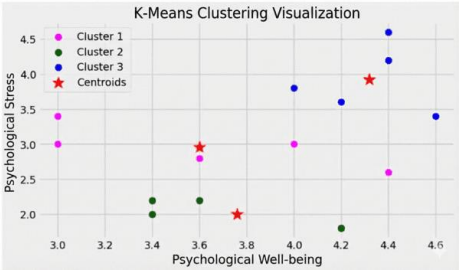

Image 3. Clustering Visualization in the Information Systems Study Program

Image 3 shows the 2020 Information Systems cohort (n=15) evenly distributed, with five students each in the depressed, neutral, and happy clusters, indicating no dominant pattern.

Table 7. Values at each cluster center

cluster	Pusat *Cluster*	
	x_1	x_2
1	3,60	2,96
2	3,76	2,00
3	4,32	3,92

Statistics Study Program

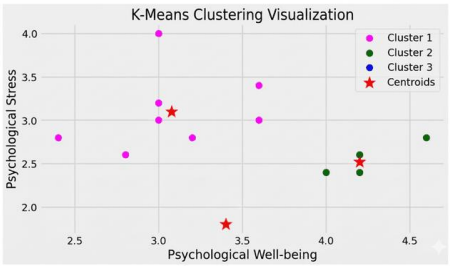

Image 4. Clustering Visualization in the Statistics Study Program

Image 4 displays K-Means results for the 2021 Statistics cohort (n=15). With 8 students clustering into depression, 5 neutral, and 2 happy, the data indicates a dominance of depressive symptoms in this group.

Table 8. Values at each cluster center

cluster	Pusat *Cluster*	
	x_1	x_2
1	3,07	3,10
2	4,20	2,52
3	3,40	1,80

Mathematics Study Program

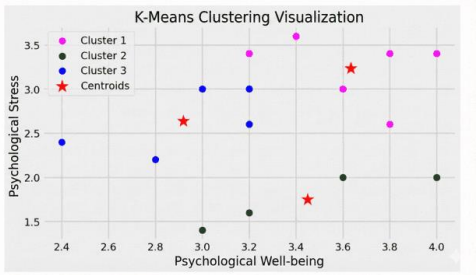

Image 5. Clustering Visualization in the Mathematics Study Program

Image 5 shows the results of clustering visualization using K-Means. The third grouping is the Mathematics Study

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 145 – 152 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v12i1.4359
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

Program, intake of 2022, with a sample size of 15 data points. The K-Means clustering results show that 6 data points fall into the depression category, 4 data points fall into the neutral or stable cluster, and 5 data points fall into the happy cluster. Based on this, it can be concluded that, from the sample taken, mathematics students predominantly have depressed mental health.

Table 9. Values at each cluster center

cluster	Pusat *Cluster*	
	x_1	x_2
1	3,63	3,23
2	3,45	1,75
3	2,92	2,64

Biology Study Program

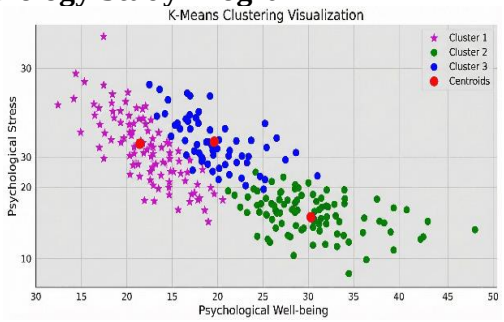

Image 6. Clustering Visualization in the Biology Study Program

Image 6 shows the results of clustering visualization using K-Means. In the fourth grouping, the class of 2023, with the Biology study program selected, the sample consisted of 15 data points. The K-Means clustering results show that 3 data points fall into the depression cluster, 10 into the neutral or stable cluster, and 2 into the happiness cluster. Based on this, it can be concluded that biology majors have a higher proportion of neutral or stable mental health among the samples taken.

Table 10. Values at each cluster center

cluster	Pusat *Cluster*	
	x_1	x_2
1	3,13	3,33
2	3,66	2,12
3	4,00	4,30

Cluster Evaluation

Table 11. Silhouette coefficient results

Study Program	Silhouette Coefficient Value	Description
Information Systems	0.39	Weak structure
Statistics	0.54	Medium structure
Mathematics	0.34	Weak structure
Biology	0.6	Medium structure

Based on Table 11, the results of the cluster evaluation using the silhouette coefficient, it can be seen that the biology and Statistics study programs have silhouette coefficient values categorized as moderately structured. The information systems and mathematics study programs have silhouette coefficient values categorized as weakly structured. The comparison of the results of this study with previous research is shown in Table 12.

Table 12. Comparison of Clustering Results Performance

References	Approach	Methode	Silhouette Coefficient
[16]	MHI	K-Means Clustering	0,13 - 0,19
[29]	non MHI	K-Means & K-Modes Clustering	0,05 - 0,15
This research	MHI	K-Means Clustering & Teknik Sampling Kombinasi	0,39 - 0,60

Achieving a Silhouette Coefficient

of 0.39–0.60, this study outperforms prior benchmarks (0.05–0.19) by integrating MHI with combined sampling. This approach ensures data representativeness, establishing a robust foundation for targeted mental health interventions.

The discussion is analyzing 60 students across four majors, K-Means clustering revealed that Statistics and Mathematics students tended toward depression, while Biology students showed stability; Information Systems showed no dominant pattern. Silhouette analysis indicated moderate structure for Statistics and Biology, but weak for the others. Future research should expand to variables like social support and focus on developing practical interventions.

CONCLUSION

This research successfully classified the mental health patterns of 60 students at the Faculty of Science and Technology, UT, using K-Means clustering and combined sampling. The findings revealed that Statistics and Mathematics students were predominantly in the depressed cluster, whereas Biology students were largely neutral/stable, and Information Systems students showed no dominant pattern. The success rate of the the K-Means was demonstrated by a Silhouette Coefficient ranging from 0.39 to 0.60, where Biology and Statistics achieved a "Reasonable Structure," proving the method's effectiveness in mapping student mental health trends for future targeted interventions.

BIBLIOGRAPHY

[1] D. Wahyuni and D. Winarso, "Penerapan metode rule based reasoning dalam sistem pakar deteksi dini gangguan kesehatan mental pada mahasiswa," *Journal of Software Engineering and Information Systems*, vol. 2, no. 2, pp. 1–10, 2021.

[2] World Health Organization, "Depressive disorder (depression)," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[3] Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI), "Masalah Kesehatan Jiwa di Indonesia," 2022. [Online]. Available: [tautan mencurigakan telah dihapus].

[4] W. A. Radiani, "Asesmen psikologis dan nilai budaya sebagai landasan konseling dalam pengembangan diri siswa," *Jurnal Nasional Bimbingan dan Konseling*, pp. 66–79, 2022.

[5] H. Prastiwi, J. Pricilia, and E. Rasywir, "Implementasi data mining untuk menentukan persediaan stok barang menggunakan metode K-Means clustering," *Jurnal Informatika dan Rekayasa Komputer*, vol. 2, no. 1, pp. 141–148, 2022.

[6] E. Prayitno, N. Tarigan, W. Sukmawaty, and U. Mauidzoh, "Gangguan mental emosional dan depresi pada remaja," *Kebangkitan UMKM Pascapandemi COVID-19*, vol. 2, no. 4, pp. 4787–4794, 2022.

[7] T. Solang and A. Nugroho, "Analisis kesehatan mental mahasiswa menggunakan algoritma K-Means," *Jurnal TEKINKOM*, vol. 6, no. 1, pp. 8–15, 2023.

[8] A. P. Thenata and M. Suryadi,

"Machine Learning Prediction of Anxiety Levels in the Society of Academicians During the Covid-19 Pandemic," *Jurnal Varian*, vol. 6, no. 1, pp. 81–88, 2022.

[9] D. Praseptian M., A. Fadlil, and H. Herman, "Penerapan clustering K-Means untuk pengelompokan tingkat kepuasan pengguna," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, p. 1693, 2022.

[10] A. Septianingsih, "Analisis K-Means clustering pada pemetaan provinsi Indonesia berdasarkan indikator rumah layak huni," *Jurnal Lebesgue*, vol. 3, no. 1, 2022.

[11] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 1990.

[12] D. Firmansyah and Dede, "Teknik pengambilan sampel umum dalam metodologi penelitian: Literature review," *Jurnal Ilmiah Pendidikan Holistik*, vol. 1, no. 2, pp. 85–114, 2022.

[13] N. Suriani, Risnita, and M. S. Jailani, "Konsep populasi dan sampling serta pemilihan partisipan ditinjau dari penelitian ilmiah pendidikan," *Jurnal IH-SAN: Jurnal Pendidikan Islam*, vol. 1, no. 2, pp. 24–36, 2023.

[14] Y. A. Rozali, N. W. Sitasari, and A. Lenggogeni, "Meningkatkan kesehatan mental di masa pandemi," *Jurnal AbdiMas*, vol. 7, no. 2, 2021.