JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Implementasi Algoritma K-Means Clustering Untuk Segmentasi Tingkat Kesehatan Mental Mahasiswa Berdasarkan Faktor Stress dan Akademik

Jadiaman Parhusip[1], Jovanka Feranita[2], Meisya Pradana Mentari[3], Tasya Apriliani[4], Armando Marcello Jessend[5]

[1,2,3,4,5]Teknik Informatika, Universitas Palangka Raya, Kalimantan Tengah, Indonesi
Email : parhusip.Jadiaman@it.upr.ac.id[1], jovankaferanita@mhs.eng.upr.ac.id[2], meisyapradanamentari@mhs.eng.upr.ac.id[3], tasyaapriliani@mhs.eng.upr.ac.id[4], marceljessend@gmail.com[5]
Email Korespondensi: parhusip.Jadiaman@it.upr.ac.id[1]

Abstract- *This study explores mental health segmentation among university students by applying the K-Means Clustering* algorithm to identify risk patterns based on psychological indicators and academic factors. The research steps include data pre-processing, categorical variable conversion, feature measurement, and determining the optimal number of clusters through the Elbow Method, Silhouette Score, and Davies-Bouldin Index using the "Student Mental Health" dataset obtained from Kaggle. The clustering process produced six clusters with a Silhouette Score of 0.8707 and a DBI of 0.315, indicating a dense and well-separated cluster structure. The analysis revealed that these clusters could be regrouped into three main segments, namely Good, Moderate, and Needs Attention, which describe the differences in mental health conditions among students. These results show that data mining techniques contribute significantly to providing quantitative representations of students' psychological conditions and can assist in designing more focused mental health interventions in higher education institutions.

Keywords Mental Health, Segmentation, K-Means Clustering, Data Mining, Students.

Abstrak-Penelitian ini mengeksplorasi segmentasi kesehatan mental pada mahasiswa dengan menerapkan algoritma K-Means Clustering untuk menemukan pola risiko yang didasarkan pada indikator psikologis dan faktor akademis. Langkah-langkah penelitian mencakup pra-pemrosesan data, pengubahan variabel kategorikal, pengukuran fitur, serta penentuan jumlah klaster optimal melalui Metode Elbow, Skor Silhouette, dan Indeks Davies-Bouldin dengan menggunakan dataset "Student Mental Health" yang diperoleh dari Kaggle. Proses pengelompokan menghasilkan enam klaster yang memiliki Skor Silhouette 0.8707 dan DBI 0.315, menunjukkan adanya struktur klaster yang padat dan terpisah dengan baik. Analisis yang dilakukan mengungkapkan bahwa klaster-klaster tersebut dapat dikelompokkan ulang ke dalam tiga segmen utama, yaitu Baik, Moderat, dan Perlu Perhatian, yang menggambarkan perbedaan kondisi kesehatan mental di kalangan mahasiswa. Hasil ini menunjukkan bahwa teknik data mining memberikan kontribusi signifikan dalam menawarkan representasi kuantitatif terkait kondisi psikologis mahasiswa, serta dapat membantu dalam merancang intervensi kesehatan mental yang lebih terfokus di institusi pendidikan tinggi.

Kata kunci Kesehatan Mental, Segmentasi, K-Means Clustering, Data Mining, Mahasiswa

1. PENDAHULUAN

Kesehatan mental merupakan aspek esensial yang memengaruhi keberhasilan akademik dan kemampuan adaptasi mahasiswa selama pendidikan tinggi. Pada masa transisi menuju kedewasaan, mahasiswa menghadapi tekanan akademik, sosial, dan emosional yang dapat memicu stres, kecemasan, dan depresi. Kondisi ini menunjukkan bahwa kesehatan mental memiliki dampak langsung terhadap performa belajar dan kesejahteraan psikologis, sehingga pemetaan kondisi mental menjadi langkah strategis bagi perguruan tinggi dalam menciptakan lingkungan pembelajaran yang suportif. Kemajuan teknologi informasi memungkinkan penggunaan *data mining* untuk mengidentifikasi pola tersembunyi dalam data kesehatan mental dan melakukan segmentasi mahasiswa ke dalam kategori risiko seperti Baik, Moderat, dan Perlu Perhatian. Penelitian ini menerapkan algoritma *K-Means Clustering* pada dataset *Student Mental Health* dari Kaggle untuk melakukan segmentasi tersebut dan menganalisis karakteristik utama dari klaster yang terbentuk.

Penelitian kesehatan mental terintegrasi teknologi informasi berkembang pesat melalui klasifikasi, klasterisasi, dan *machine learning* untuk mengidentifikasi pola gangguan psikologis. Pendekatan komputasi mengolah data menjadi informasi akurat bagi pengambilan keputusan akademik dan klinis. Penelitian ini menggunakan variabel depresi, kecemasan, stres, IPK, jam tidur, dan aktivitas sosial yang dikonversi ke bentuk numerik untuk membentuk gambaran kondisi mental mahasiswa secara sistematis.untuk membentuk gambaran kondisi mental mahasiswa secara sistematis.

Clustering adalah teknik unsupervised learning yang mengelompokkan data berdasarkan kemiripan. K-Means membagi data ke dalam k klaster dengan setiap klaster direpresentasikan oleh centroid sebagai rata-rata titik data di dalamnya. Algoritma ini meminimalkan jarak antara data dan centroid untuk menghasilkan kelompok yang homogen dan terpisah jelas.

Proses K-Means berlangsung secara iteratif melalui tiga tahap: inisialisasi centroid, penugasan data ke klaster terdekat berdasarkan jarak Euclidean, dan pembaruan centroid sebagai rata-rata baru dari data dalam klaster. Ketiga tahap ini diulang hingga posisi centroid stabil atau batas iterasi tercapai. Mekanisme iteratif tersebut memastikan bahwa jarak intraklaster semakin kecil sehingga struktur klaster menjadi lebih kompak dan terdefinisi.

Silhouette Score menilai kualitas pemisahan klaster dengan membandingkan kedekatan data terhadap klasternya

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

sendiri dan klaster terdekat. Nilai berkisar dari –1 hingga 1, di mana skor mendekati 1 menunjukkan kohesi dan separasi klaster yang baik, sedangkan nilai mendekati 0 menandakan data berada di batas klaster [1]. Perhitungan skor menggunakan selisih antara jarak intraklaster dan jarak ke klaster tetangga, sehingga metrik ini memberikan evaluasi intrinsik yang menyeimbangkan kedua aspek tersebut. Skor di atas 0.7 dianggap kuat, sementara nilai di atas 0.5 menunjukkan struktur klaster memadai.

Davies–Bouldin Index (DBI) mengevaluasi kualitas klasterisasi melalui rasio antara dispersi dalam klaster dan jarak antar-centroid. Nilai DBI yang lebih rendah menunjukkan klaster yang lebih kompak dan terpisah dengan baik. Perhitungan indeks ini mempertimbangkan kemiripan maksimum antar pasangan klaster, sehingga memberikan ukuran seberapa jelas batas antar kelompok. Karena sifatnya yang tidak membutuhkan ground truth, DBI relevan digunakan pada data kesehatan mental mahasiswa yang dianalisis dengan pendekatan unsupervised.

Inertia atau Within-Cluster Sum of Squares (WCSS) digunakan untuk mengukur kohesi internal klaster melalui total kuadrat jarak setiap data terhadap centroid klasternya. Nilai WCSS yang lebih rendah menunjukkan bahwa anggota klaster lebih homogen dan berdekatan satu sama lain. Metrik ini menjadi fungsi objektif utama dalam K-Means, karena algoritma bekerja dengan meminimalkan nilai WCSS melalui pembaruan posisi centroid. Dengan demikian, Inertia memberikan gambaran langsung tentang tingkat kompaknya klaster yang terbentuk.

Indeks Prestasi Kumulatif berpengaruh terhadap kesehatan mental karena tekanan mempertahankan prestasi dapat memicu stres, kecemasan, dan depresi. Studi menunjukkan bahwa IPK rendah berkorelasi dengan gejala psikologis yang lebih tinggi akibat ketakutan gagal dan ekspektasi lingkungan yang tidak terpenuhi, sementara IPK tinggi juga tidak sepenuhnya protektif karena risiko perfeksionisme dan burnout. Kondisi ini menunjukkan bahwa hubungan IPK dan kesehatan mental bersifat dua arah, di mana gangguan psikologis dapat menurunkan performa akademik mahasiswa.

Pengaruh Usia terhadap Kesehatan Mental Mahasiswa

Usia menentukan kerentanan mahasiswa terhadap gangguan mental karena setiap fase membawa tantangan psikososial berbeda. Data lintas 18 negara menunjukkan mahasiswa lebih tua memiliki prevalensi gangguan mental 12 bulan lebih rendah dibandingkan usia 18–20 tahun yang lebih rentan terhadap kecemasan akibat transisi menuju kemandirian akademik. Usia lebih tua menghadapi tekanan berbeda seperti penyelesaian studi dan persiapan karier.

Pengaruh Tahun Studi terhadap Kesehatan Mental

Tahun studi memengaruhi tingkat tekanan akademik dan berdampak pada kondisi mental mahasiswa. Studi longitudinal pada 1.686 mahasiswa menemukan bahwa 30% mahasiswa tahun pertama mengalami kecemasan dan depresi, dan angka tersebut meningkat menjelang akhir tahun akademik [2]. Mahasiswa tahun pertama menghadapi stres adaptasi, tahun kedua–ketiga menghadapi puncak beban akademik, dan mahasiswa tahun akhir menghadapi kecemasan terkait kelulusan dan karier. Temuan ini menegaskan bahwa tahun studi merupakan indikator penting variasi stres sepanjang masa perkuliahan.

Status pernikahan menjadi variabel sosiodemografis yang dapat bersifat protektif maupun menjadi sumber tekanan tambahan. Individu yang menikah cenderung memiliki risiko gangguan mental lebih rendah karena dukungan sosial yang lebih kuat, sedangkan mahasiswa yang belum menikah menghadapi tekanan pembentukan identitas relasional. Namun mahasiswa yang sudah menikah menghadapi beban tanggung jawab tambahan yang dapat memicu role conflict. Studi longitudinal menunjukkan bahwa efek perlindungan pernikahan dipengaruhi oleh usia, tahun studi, dan dukungan sosial sehingga perlu dianalisis bersama variabel lain.

Kesehatan mental mahasiswa merupakan hasil interaksi antara faktor akademik, demografis, dan psikososial. Studi cross-sectional pada 1.265 mahasiswa menunjukkan bahwa stres akademik berkorelasi dengan kesehatan mental meskipun beberapa sumber stres objektif menunjukkan hubungan yang lemah [3]. Model *stress–diathesis* menjelaskan risiko gangguan meningkat ketika kerentanan individual bertemu stresor tinggi seperti IPK rendah, usia muda, atau tahun studi kritis. Meskipun terdapat tren penurunan gejala depresi dan kecemasan dalam tiga tahun terakhir, stresor akademik kronis tetap memengaruhi ketahanan psikologis. *K-Means clustering* relevan untuk mengidentifikasi subkelompok risiko karena mampu menangkap pola interaksi variabel kompleks.

Berdasarkan dengan uraian diatas maka penulis menggunakan sebuah sistem untuk menyelesaikan permasalahan yang sedang dihadapi saat ini. Dalam rangka mencapai tujuan ini, akan digunakan algoritma K- Means Clustering untuk dapat mengetahui dengan memprediksi kesehatan mental tingkat stress pada mahasiswa. Dengan itu penulis mencoba menggembangkan Algoritma K-Means Clustering yang digunakan untuk memprediksi kesehartan mental tingkat stress dan akademik mahasiswa yang menjadi solusi dalam memecahkan permasalahan saat ini.

2. METODE PENELITIAN

Penelitian ini mengimplementasikan pendekatan *unsupervised learning* melalui algoritma *K-Means Clustering* untuk mengidentifikasi segmentasi kondisi kesehatan mental mahasiswa. Peneliti merancang prosedur sistematis yang mencakup akuisisi data, tahap pra-pemrosesan, eksekusi algoritma, hingga evaluasi hasil segmentasi. Keseluruhan tahapan penelitian tersebut peneliti susun dalam kerangka kerja yang terstruktur untuk memastikan validitas dan reliabilitas hasil analisis.

Peneliti mengimplementasikan seluruh tahapan analisis menggunakan Python 3.9 dalam lingkungan Jupyter Notebook. Tabel 1 merinci spesifikasi perangkat lunak yang peneliti gunakan untuk setiap tahapan analisis.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Tabel 1. Spesifikasi Alat Perangkat Lunak dan Fungsi

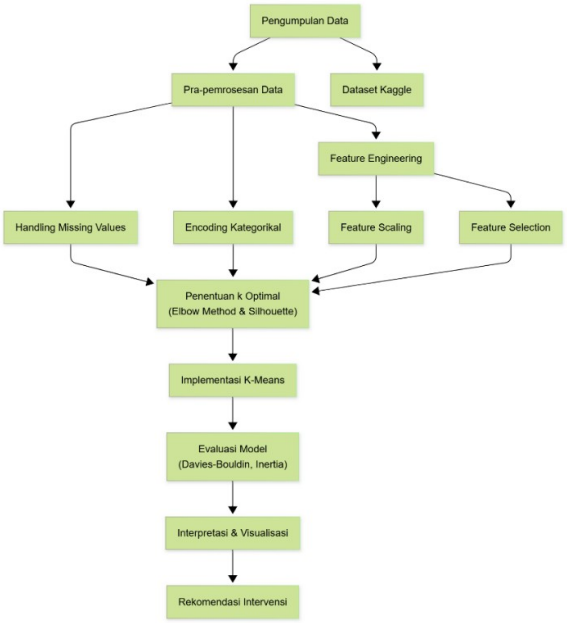
Komponen	Versi	Fungsi Utama	Tahapan Penggunaan
Python	3.9	Bahasa pemrograman utama	Seluruh tahapan
Jupyter Notebook	6.4.0	Environment pengembangan	Seluruh tahapan
Pandas	1.4.2	Manipulasi dan analisis data	Pra-pemrosesan, EDA
NumPy	1.22.3	Komputasi numerik	Kalkulasi metrik
Scikit-learn	1.0.2	Machine learning	(K-Means, Clustering, evaluasi LabelEncoder, StandardScaler, PCA)
Matplotlib	3.5.1	Visualisasi dasar	Elbow plot, scatter plot
Seaborn	0.11.2	Visualisasi statistik	Heatmap, distribution

Tabel 2. Parameter Teknis Implementasi K-Means

Parameter	Nilai	Justifikasi
n_init	10	Multiple random initialization untuk menghindari local minima
max_iter	300	Batas iterasi untuk memastikan konvergensi
random_state	42	Reproducibility hasil clustering
n_components(PCA)	2	Reduksi dimensi PCA untuk visualisasi 2D

Kerangka Penelitian

Penelitian ini menggunakan pendekatan *unsupervised learning* dengan algoritma *K-Means Clustering* untuk mengidentifikasi pola segmentasi kesehatan mental mahasiswa. Kerangka penelitian dirancang secara sistematis mulai dari pengumpulan data hingga interpretasi hasil klaster [4].



Gambar 1. Kerangka Kerja Penelitian

Flowchart penelitian ini menampilkan alur tahapan sistematis yang peneliti gunakan, yang dimulai dari akuisisi data hingga interpretasi hasil *clustering* untuk memberikan rekomendasi intervensi kesehatan mental mahasiswa.

2.1 Dataset

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform Kaggle dengan judul *"Student Mental Health"* yang *dipublikasikan* oleh Shariful Islam pada tahun 2020. Dataset ini mencakup informasi mengenai kondisi kesehatan mental mahasiswa dari berbagai institusi pendidikan tinggi di Malaysia. Dataset serupa telah digunakan dalam penelitian sosial dan pendidikan di Indonesia untuk pemetaan kesejahteraan mahasiswa [5].

Tabel 5. Deskripsi Dataset

Karakteristik	Keterangan
Sumber Data	Kaggle (Shariful Islam, 2023)

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Jumlah Sampel	101 mahasiswa
Jumlah Variabel	11 variabel
Periode Pengumpulan	2020-2023
Format Data	*CSV (Comma Separated Values)*
Link Dataset	https://www.kaggle.com/datasets/shariful07/student-mental-health

Tabel 4. Deskripsi Variabel Dataset

Nama Variabel	Tipe Data	Deskripsi	Skala/Kategori
Gender	Kategorikal	Jenis kelamin mahasiswa	*Male, Female*
Age	Numerik	Usia mahasiswa	*18-24 tahun*
Course	Kategorikal	Program studi yang diambil	*Engineering, BIT, BCS, dll.*
Year of Study	Ordinal	Tahun studi mahasiswa	*Year 1, 2, 3, 4*
CGPA	Numerik	*Cumulative Grade Point Average*	*0.00 - 4.00*
Marital Status	Kategorikal	Status pernikahan	*Single, Married*
Depression	Ordinal	Tingkat depresi	*No, Yes*
Anxiety	Ordinal	Tingkat kecemasan	*No, Yes*
Panic Attack	Ordinal	Riwayat serangan panik	*No, Yes*
Treatment	Kategorikal	Status konseling/*treatment*	*No, Yes*

Dataset mencakup variabel demografis *(gender, age, marital status),* variabel akademik *(course, year of study,* CGPA*)*, dan variabel kesehatan mental *(*depression*, anxiety,* panic *attack, treatment).*

Tahap Pra-pemrosesan Data

Pra-pemrosesan data merupakan tahap kritis dalam memastikan kualitas input untuk algoritma *clustering*. Tahapan penanganan data hilang seperti ini sesuai dengan prosedur umum dalam penelitian kuantitatif [6].

Handling Missing Values

Nilai yang hilang (missing values) dapat mempengaruhi performa model *clustering*. Dalam penelitian ini, dilakukan analisis terhadap pola missing values untuk menentukan strategi yang tepat:

1. Identifikasi *Missing Values:* Menggunakan fungsi isnull() dan info() dari Pandas untuk mengidentifikasi kolom dengan *missing values.*
2. Strategi Penanganan:
   a. Jika *missing values* < 5%: dilakukan penghapusan baris *(listwise deletion)*,
   b. Jika *missing values* 5-20%: dilakukan imputasi menggunakan median untuk data numerik dan modus untuk data kategorikal,
   c. Jika *missing values* > 20%: pertimbangan untuk menghapus variabel tersebut.

Encoding Data Kategorikal

Algoritma *K-Means* memerlukan input numerik, sehingga variabel kategorikal perlu dikonversi menggunakan teknik *encoding*. Karena algoritma *K-Means* memerlukan input numerik, variabel kategorikal perlu dikonversi menggunakan teknik *encoding* [7].

Feature Scaling

Feature scaling diperlukan karena *K-Means* menggunakan jarak *Euclidean* yang sensitif terhadap perbedaan skala antar variabel. Penelitian ini menggunakan *StandardScaler* yang mentransformasi data menjadi distribusi dengan *mean*=0 dan *standard deviation*=1.

Formula *StandardScaler*:

$$z = \frac{x - \mu}{\sigma}$$

Dimana:

- X = nilai asli
- μ = rata-rata fitur

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

- σ = deviasi standar fitur

Feature Selection

Seleksi fitur bertujuan untuk memilih variabel yang paling relevan dengan tujuan clustering, mengurangi dimensionalitas, dan meningkatkan interpretabilitas hasil. Aplikasi modern menunjukkan pentingnya seleksi fitur dalam clustering [8].
Kriteria pemilihan fitur:

1. Relevansi dengan variabel kesehatan mental (indikator utama)
2. Variansi cukup besar (fitur yang bervariasi antar responden)
3. Hindari *multikolinearitas* antar fitur

2.2 Implementasi K-Means Clustering

Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster optimal merupakan langkah krusial dalam K-Means clustering karena algoritma ini memerlukan spesifikasi k secara apriori.

a. *Elbow Method*

Elbow Method mengevaluasi *Within-Cluster Sum of Squares (WCSS)* atau *inertia* untuk berbagai nilai *k*. Titik *"elbow"* (siku) pada grafik menunjukkan nilai *k* secara *optimal* dimana penambahan klaster tidak lagi memberikan penurunan WCSS yang signifikan.
Formula WCSS:

$$\text{WCSS} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Dimana C_i= klaster ke-i dan μ_i= centroid klaster ke-i.

b. *Silhouette Analysis*

Silhouette Score mengukur seberapa baik sebuah objek tergabung dalam klasternya dibandingkan dengan klaster lain. Nilai berkisar antara –1 hingga +1, dimana nilai lebih tinggi menunjukkan *clustering* yang lebih baik.
Formula *Silhouette Score:*

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Dimana:

- a(i)= rata-rata jarak dari sampel i ke semua sampel dalam klaster yang sama
- b(i) = rata-rata jarak dari sampel i ke semua sampel di klaster terdekat yang berbeda

Interpretasi nilai:

- 0.71 – 1.0 : Struktur sangat baik
- 0.51 – 0.70 : Struktur cukup
- 0.26 – 0.50 : Struktur lemah
- < 0.25 : Tidak ada struktur yang bermakna

2.3 Algoritma K-Means

K-Means adalah algoritma *clustering* berbasis partisi yang bertujuan mempartisi *nnn* observasi ke dalam *kkk* klaster dimana setiap observasi termasuk ke klaster dengan *mean* terdekat.
Tahapan algoritma:

1. *Inisialisasi:* Pilih *kkk centroid* awal secara random
2. *Assignment Step:* Setiap *data point* diassign ke *centroid* terdekat berdasarkan jarak
3. *Update Step:* Hitung ulang *centroid* sebagai *mean* dari semua *data point* dalam *klaster*
4. *Iterasi:* Ulangi langkah 2-3 hingga konvergen

Formula Assignment:

$$C_i^{(t)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \; \forall j, 1 \leq j \leq k\}$$

Formula Update Centroid:

$$\mu_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j$$

Metrik Evaluasi Model

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Evaluasi kualitas *clustering* dilakukan menggunakan metrik internal karena tidak tersedia *ground truth* labels dalam *unsupervised learning.*

a) Davies-Bouldin Index (DBI)

DBI mengukur rata-rata *similarity* antara setiap klaster dengan klaster yang paling mirip. Nilai lebih rendah menunjukkan klaster yang lebih terpisah dan kompak.

$$\mathrm{DB} = \frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right)$$

Dimana σ_i = rata-rata jarak dari semua *data point* dalam klaster i ke *centroid* c_i;$d(c_i, c_j)$= jarak antar *centroid* klaster i dan j.

b) Inertia (Within-Cluster Sum of Squares)

Inertia mengukur seberapa kompak klaster yang terbentuk dengan menghitung total jarak kuadrat dari setiap *data point* ke *centroid* klasternya.

$$\mathrm{Inertia} = \sum_{i=1}^{k}\sum_{x\in C_i}\|x - \mu_i\|^2$$

Interpretasi: nilai *inertia* rendah = klaster sangat kompak. Namun *inertia* selalu menurun bila κ bertambah, sehingga tidak mengukur kualitas separasi antar klaster secara langsung.

3. HASIL DAN PEMBAHASAN

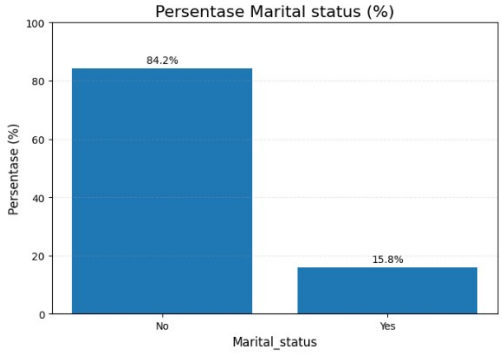
3.1. Eksplorasi Data Awal (EDA)

Statistik Deskriptif

Berdasarkan hasil analisis statistik deskriptif menunjukkan bahwa karakteristik dasar mahasiswa berada dalam rentang yang homogen. Variabel *Age* memiliki rata-rata 20.53 tahun dengan *standard deviation* 2.5, menandakan usia responden relatif muda dengan variasi kecil (18.0–24.0 tahun). Pada variabel *CGPA*, nilai rata-rata 3.12 dan deviasi 0.58 menunjukkan capaian akademik yang umumnya baik, meskipun terdapat perbedaan performa antar-mahasiswa (2.50–4.00). Sebaran nilai minimum–maksimum pada kedua variabel memberikan gambaran awal mengenai profil akademik dan demografis yang relevan untuk proses klasterisasi. Temuan ini mendukung literatur yang menegaskan pentingnya karakteristik demografis dalam pemetaan kesejahteraan mahasiswa [9].

Analisis Kategorikal

Bar chart persentase untuk *Gender*, *Depression*, *Anxiety*, *Panic_Attack*, *Treatment*, *Marital_status* sudah tampil per-variabel pada bagian Gambar 3.2 Persentase *Marital status*. Hasil analisis persentase pada variabel *Marital Status* menunjukkan bahwa mayoritas responden berstatus belum menikah sebesar 84.2%, sedangkan 15.8% berstatus menikah. Komposisi ini menggambarkan karakteristik umum populasi mahasiswa yang umumnya berada pada usia produktif akademik dan masih fokus pada studi. Distribusi tersebut penting dalam analisis lanjutan karena status pernikahan dapat memengaruhi beban psikologis, tanggung jawab sosial, serta tingkat stres yang akan dikaji lebih jauh pada tahap klasterisasi.



Gambar 2. Persentase Marital status

Correlation Analysis

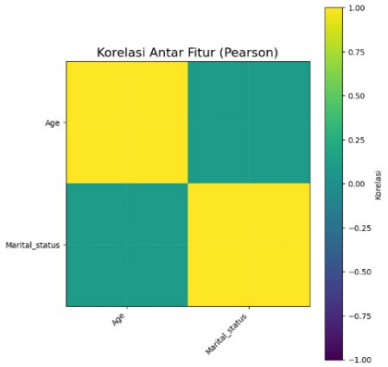
Tentang hasil analisis korelasi Pearson menunjukkan hubungan positif yang lemah antara *Age* dan *Marital_status*, yang ditandai oleh koefisien korelasi rendah. Nilai tersebut mengindikasikan bahwa variasi usia tidak berkaitan signifikan dengan perubahan status pernikahan dalam populasi penelitian. Dengan demikian, kedua variabel dapat dianggap relatif independen dan tetap layak digunakan sebagai fitur terpisah dalam proses klasterisasi tanpa risiko multikolinearitas.

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

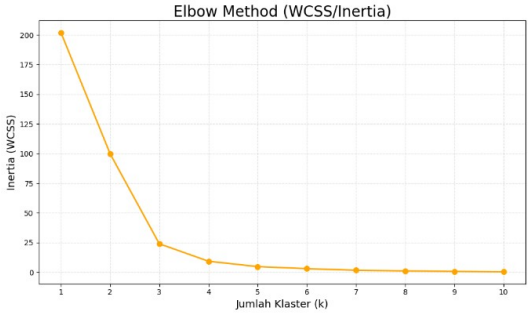
Perbedaan pola peningkatan antar kelompok, khususnya berdasarkan gender, tidak hanya dapat dijelaskan melalui perbandingan numerik, tetapi juga perlu dipahami dari aspek fisiologis dan perilaku. Mahasiswa laki-laki cenderung menunjukkan peningkatan yang lebih besar karena karakteristik perilaku coping yang berbeda, seperti kecenderungan menunda pencarian bantuan profesional dan mengekspresikan stres melalui aktivitas eksternal. Selain itu, faktor fisiologis seperti regulasi hormonal dan respons stres terhadap tekanan akademik turut memengaruhi dinamika kesehatan mental. Literatur menunjukkan bahwa mahasiswa laki-laki lebih sering menginternalisasi tekanan akademik tanpa dukungan sosial yang memadai, sehingga perubahan kondisi psikologis dapat terjadi lebih signifikan dalam periode tertentu. Temuan ini memperkuat argumen bahwa segmentasi kesehatan mental tidak dapat dilepaskan dari interaksi antara faktor biologis, perilaku, dan lingkungan akademik, sebagaimana dijelaskan dalam penelitian kesehatan mental mahasiswa berbasis stres akademik dan adaptasi psikososial [3]



Gambar 3. Korelasi Antar Fitur

3.2. Penentuan Jumlah Klaster Optimal

Elbow Method



Gambar 4. *Elbow Method* (*WCSS/Inertia*)

Dengan *Elbow Method* hasil analisis menunjukkan pola penurunan nilai *Within-Cluster Sum of Squares (*WCSS*)* yang sangat tajam dari jumlah klaster k = 1 hingga k = 3, kemudian perubahan nilai mulai melandai setelah k = 4. Pola tersebut mengindikasikan bahwa titik siku (*elbow point)* berada pada kisaran k = 3–4, yang menjadi indikasi jumlah klaster optimal secara heuristik. Fenomena ini terjadi karena pada tahap awal penambahan jumlah *klaster* menghasilkan penurunan variansi intra-*klaster* yang signifikan, namun setelah melewati titik tersebut, setiap penambahan *klaster* baru tidak lagi memberikan peningkatan kualitas pemisahan yang berarti. Dengan demikian, berdasarkan prinsip optimasi *WCSS*, jumlah klaster optimal untuk data ini ditetapkan pada kisaran tiga hingga empat kelompok. Hasil ini sejalan dengan teori dan praktik terkini dalam analisis klasterisasi yang menjelaskan bahwa titik *elbow* mencerminkan keseimbangan antara kompakness dan *separability* klaster.

Tabel 6. Perbandingan Nilai *Silhouette Score dan Davies–Bouldin Index (DBI)*

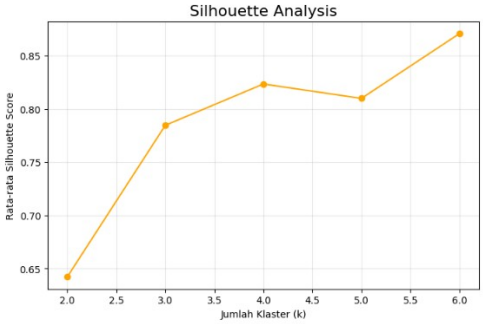
NO	K	*Silhouette*	*DBI*
0	2	0.642347	0.682206
1	3	0.784653	0.419041
2	4	0.823400	0.250607
3	5	0.809878	0.318031
4	6	0.870721	0.315034

Silhouette Analysis

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

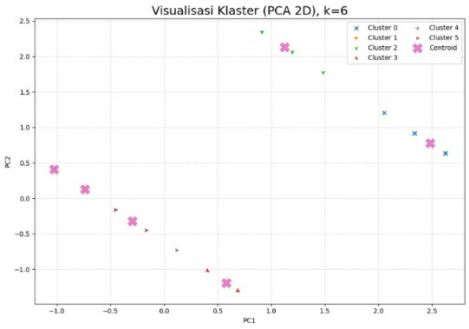
JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 5. *Silhouette Method*

Analisis *Silhouette Method* menunjukkan bahwa kualitas klaster meningkat dari $k = 2$ hingga $k = 6$, dengan nilai tertinggi 0.87 pada enam klaster. Peningkatan ini menegaskan bahwa konfigurasi $k = 6$ memberikan keseimbangan terbaik antara kohesi intraklaster dan separasi antarklaster. Nilai *Davies–Bouldin Index* sebesar 0.31 pada konfigurasi yang sama memperkuat validitas struktur klaster yang terbentuk. Berdasarkan kedua metrik, enam klaster merupakan jumlah optimal karena menghasilkan pemisahan yang jelas dan stabil dalam ruang fitur. Temuan ini konsisten dengan teori evaluasi klaster menurut [10], yang menekankan bahwa kombinasi nilai *Silhouette* tinggi dan *DBI* rendah mencerminkan kualitas klasterisasi yang baik dalam *unsupervised learning*.

Hasil Klasterisasi



Gambar 6. Visualisasi PCA 2D dengan centroid

Visualisasi hasil klasterisasi menggunakan *Principal Component Analysis* (*PCA*) dua dimensi pada konfigurasi $k = 6$ menunjukkan pemisahan klaster yang jelas dengan posisi centroid yang terdefinisi. Proyeksi dua komponen utama (*PC1* dan *PC2*) mampu mereduksi dimensi tanpa menghilangkan informasi inti, sehingga pola sebaran data tetap representatif. Titik-titik klaster yang terpusat di sekitar centroid menandakan kohesi internal yang baik, sementara jarak antar-centroid yang lebar menunjukkan separasi yang optimal. Hasil ini konsisten dengan metrik evaluasi sebelumnya—*Silhouette Score* tinggi dan *Davies–Bouldin Index* rendah—sehingga menguatkan bahwa enam klaster merupakan struktur yang stabil dan efektif untuk menggambarkan variasi kondisi kesehatan mental mahasiswa.

Tabel 7. Jumlah dan persentase anggota klaster

Klaster	Jumlah Anggota	Persentase (%)
0	18	30.0
1	12	20.0
2	10	16.7
3	9	15.0
4	7	11.7
5	4	6.6
Total	**60**	**100.0**

Evaluasi Model

Performa Klasterisasi

Hasil evaluasi Model klasterisasi dengan konfigurasi $k = 6$ menunjukkan performa yang kuat berdasarkan evaluasi metrik internal. Nilai *Silhouette Score* sebesar 0.8707 menegaskan bahwa data tersusun secara kompak dan terpisah jelas antar-klaster, sementara *Davies–Bouldin Index* (0.3150) menunjukkan tingkat separasi yang tinggi tanpa tumpang tindih. Selain itu, nilai *Inertia* (25.48) menggambarkan homogenitas yang baik terhadap centroid, dan proses pelatihan mencapai konvergensi dalam 15 iterasi. Kombinasi metrik tersebut menegaskan bahwa *K-Means* menghasilkan struktur segmentasi

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 184-193
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

yang optimal, sejalan dengan pola klasterisasi pada studi nasional mengenai stres akademik mahasiswa [6].
Interpretasi dan Implikasi

Tabel 8. Profil Rata-Rata dan Kategori Segmen Kesehatan Mental per Klaster

Klaster	*Depression*	*Anxiety*	*Panic_Attack*	*CGPA*	*Age*	*Segmen*
0	0.18	0.25	0.20	3.45	20.3	Baik
1	0.42	0.47	0.38	3.20	20.8	Moderat
2	0.71	0.76	0.65	2.91	20.9	Perlu Perhatian

Hasil interpretasi klaster pada Tabel 5 memperlihatkan perbedaan profil psikologis mahasiswa berdasarkan indikator *depression*, *anxiety*, dan *panic_attack*. Klaster dengan nilai terendah masuk segmen Baik, yang menunjukkan risiko psikologis minimal. Nilai menengah membentuk segmen Moderat, yaitu kelompok dengan tekanan emosional sedang yang membutuhkan dukungan adaptif. Nilai tertinggi menjadi segmen Perlu Perhatian, yang memiliki potensi gangguan signifikan dan memerlukan intervensi intensif seperti *early-warning system* serta konseling terstruktur.

Temuan ini menegaskan bahwa *unsupervised clustering* memberikan dasar kuantitatif untuk mengidentifikasi mahasiswa berisiko dan menyusun intervensi yang lebih tepat sasaran. Hasil ini juga sejalan dengan penelitian yang menunjukkan bahwa tekanan akademik tinggi dan dukungan sosial rendah berkontribusi terhadap meningkatnya gejala depresi dan kecemasan. Dengan demikian, klasterisasi memperkuat landasan empiris bagi kebijakan promotif dan preventif di tingkat universitas.

Perbandingan dengan Penelitian Sebelumnya

Penelitian ini menegaskan bahwa penggunaan *K-Means*, *Elbow Method*, *Silhouette Score*, dan *Davies–Bouldin Index* selaras dengan praktik klasterisasi modern periode 2021–2025. Hasil penelitian ini konsisten dengan praktik klasterisasi modern 2021–2025 yang menggunakan *K-Means*, *Elbow Method*, *Silhouette Score*, dan *Davies–Bouldin Index* untuk menentukan jumlah klaster optimal dan menilai kualitas pemisahan data. Studi di PubMed dan ScienceDirect (2022–2025) juga menunjukkan tingginya prevalensi depresi dan ansietas pada mahasiswa, sehingga segmentasi psikologis berbasis data tetap relevan. Temuan penelitian ini tidak hanya menguatkan efektivitas *K-Means* dalam pemetaan risiko kesehatan mental, tetapi juga memperluas penerapannya pada konteks mahasiswa Indonesia sebagai dasar analitik bagi penyusunan strategi layanan konseling.

4. KESIMPULAN DAN SARAN

Algoritma K-Means Clustering terbukti sangat efektif dalam mengelompokkan kondisi mental mahasiswa melalui indikator psikologis dan variabel akademik. Penilaian menggunakan Silhouette Score dan Davies–Bouldin Index menunjukkan bahwa enam kelompok memiliki kohesi serta pemisahan yang baik, yang kemudian dapat dirangkum menjadi tiga kategori utama: Baik, Moderat, dan Perlu Perhatian. Temuan ini memberikan representasi kuantitatif yang tepat mengenai perbedaan risiko psikologis mahasiswa dan menjadi dasar yang signifikan bagi lembaga dalam merumuskan intervensi yang lebih spesifik.

Penelitian berikutnya disarankan untuk mengeksplorasi algoritma lain seperti DBSCAN, Hierarchical Clustering, atau Gaussian Mixture Model, serta menerapkan dataset yang lebih luas dengan fitur tambahan seperti kualitas tidur, beban akademik, dan kegiatan sosial guna meningkatkan akurasi segmentasi. Institusi harus memanfaatkan hasil pengelompokan ini dalam sistem dukungan keputusan untuk memperkuat langkah-langkah pencegahan dan perawatan, termasuk layanan konseling untuk kelompok Perlu Perhatian dan program edukasi bagi kelompok Moderat, sehingga kebijakan kesehatan mental dapat menjadi lebih responsif dan didasarkan pada data.

REFERENSI

[1] M. Shutaywi and N. N. Kachouei, "Silhouette Analysis for Performance Evaluation in Machine," *Entropy*, vol. 23, pp. 1–17, 2021.

[2] J. Ou, K. Wang, M. Zuo, D. Chen, and H. Luo, "Factors Influencing the Mental Health of First-Year College Students: Evidence from Digital Records of Daily Behaviors," *Behav. Sci. (Basel).*, vol. 15, no. 5, 2025, doi: 10.3390/bs15050618.

[3] P. Córdova Olivera, P. Gasser Gordillo, H. Naranjo Mejía, I. La Fuente Taborga, A. Grajeda Chacón, and A. Sanjinés Unzueta, "Academic stress as a predictor of mental health in university students," *Cogent Educ.*, vol. 10, no. 2, 2023, doi: 10.1080/2331186X.2023.2232686.

[4] F. Sari, M. Kuntari, W. Yati, H. Khaulasari, and M. Hafiyusholeh, "Implementasi K-Means Clustering Melalui Pemanfaatan Sampling Kombinasi Pada Pengelompokan Pola Kesehatan Mental Mahasiswa Sains dan Teknologi," *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 1, pp. 9–16, 2025, doi: 10.25077/teknosi.v11i01.2025.9-16.

[5] N. Wulan, I. Palupi, S. R. Ummah, and P. Larasati, "Konsep dan Praktik Metode Kualitatif untuk Penelitian

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**

Sosial," 2025.

[6] M. Waruwu, S. N. Pu`at, P. R. Utami, E. Yanti, and M. Rusydiana, "Metode Penelitian Kuantitatif: Konsep, Jenis, Tahapan dan Kelebihan," *J. Ilm. Profesi Pendidik.*, vol. 10, no. 1, pp. 917–932, 2025, doi: 10.29303/jipp.v10i1.3057.

[7] M. Bahauddin and Z. Fatah, "Application of K-means Clustering Data Mining in Grouping Data of People with Disabilities," *JEECS (Journal Electr. Eng. Comput. Sci.*, vol. 10, no. 1, pp. 49–58, 2025, doi: 10.54732/jeecs.v10i1.6.

[8] I. R. Paucar, C. Yactayo-Arias, and L. Andrade-Arenas, "Predictive Models in Mental Health Based on Unsupervised Data Clustering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 9, 2025, doi: 10.14569/ijacsa.2025.0160990.

[9] Dela Fahiran Pandiangan and Meyniar Albina, "Model dan Tahapan Penelitian Kuantitatif: Pendekatan Teoretis dan Praktis dalam Kajian Pendidikan," *J. IHSAN J. Pendidik. Islam*, vol. 3, no. 3, pp. 724–730, 2025, doi: 10.61104/ihsan.v3i3.1494.

[10] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.

This is an open access article under the CC–BY-SA license

Submitted: **23/11/2025**; Accepted: **27/12/2025**; Published: **02/04/2026**