



Evaluasi Kinerja Fuzzy C-Means dan K-Means dalam Menganalisis Faktor Depresi Mahasiswa

Ahmadyani¹, Muhahmmad Ariando Ferdian², Christian Mahatma Bimo Alpindo³, Muhammad Asadin Nur⁴, Christian Farrel Argya Putra⁵, Anindita Septiarini⁶, Novianti Puspitasari⁷

^{1,2,3,4,5,6,7} Informatika, Universitas Mulawarman, Indonesia

Jl. Gunung Kelua, Kec. Samarinda Utara, Kota Samarinda, Kalimantan Timur 75119

E-Mail *: novipuspitasari@unmul.ac.id

ABSTRACT

Depression among college students has become a serious global mental health concern, significantly affecting academic performance and overall well-being. Factors such as academic pressure, financial stress, sleep patterns, dietary habits, and social relationships contribute to the rising prevalence of mental health disorders in campus environments. Despite growing attention to this issue, previous studies have rarely compared clustering-based data mining approaches specifically within the context of student depression in Indonesian higher education settings. This study evaluates the performance of Fuzzy C-Means (FCM) and K-Means algorithms in clustering students based on depression-related factors, using a dataset comprising features such as age, GPA, study satisfaction, sleep duration, dietary habits, history of suicidal ideation, financial pressure, family mental health history, and depression level. Performance was evaluated using Silhouette Score, Davies-Bouldin Index (DBI), and Sum of Squared Errors (SSE). Results indicate that K-Means produced more distinct clusters with a Silhouette Score of 0.18 and lower DBI of 1.89, with a notable depression score difference between Cluster 0 (0.79) and Cluster 1 (0.23), making it effective for identifying high-risk students. Meanwhile, FCM generated overlapping clusters with a Fuzzy Partition Coefficient (FPC) of 0.50 and Silhouette Score of 0.15, which more accurately reflects the ambiguous and continuous nature of psychological symptoms. This study bridges the research gap by demonstrating the complementary strengths of both algorithms and contributes to the development of data-driven early detection frameworks for student depression.

Keywords : Fuzzy C-Means; K-Means; Depression; Clustering; Mental Health

ABSTRAK

Depresi di kalangan mahasiswa telah menjadi permasalahan kesehatan mental global yang serius dan berdampak signifikan terhadap kinerja akademik serta kesejahteraan secara keseluruhan. Berbagai faktor seperti tekanan akademik, tekanan finansial, pola tidur, kebiasaan makan, dan hubungan sosial berkontribusi terhadap meningkatnya prevalensi gangguan kesehatan mental di lingkungan kampus. Meskipun perhatian terhadap isu ini semakin berkembang, penelitian terdahulu jarang membandingkan pendekatan data mining berbasis klusterisasi secara spesifik dalam konteks depresi mahasiswa di perguruan tinggi Indonesia. Penelitian ini mengevaluasi kinerja algoritma Fuzzy C-Means (FCM) dan K-Means dalam mengelompokkan mahasiswa berdasarkan faktor-faktor yang berkaitan dengan depresi, menggunakan dataset yang mencakup fitur usia, IPK, kepuasan belajar, durasi tidur, pola makan, riwayat keinginan bunuh diri, tekanan keuangan, riwayat penyakit mental keluarga, dan tingkat depresi. Evaluasi kinerja dilakukan menggunakan Silhouette Score, Davies-Bouldin Index (DBI), dan Sum of Squared Errors (SSE). Hasil menunjukkan bahwa K-Means menghasilkan kluster lebih tegas dengan Silhouette Score 0,18 dan DBI 1,89, serta perbedaan skor depresi signifikan antara Cluster 0 (0,79) dan Cluster 1 (0,23). FCM menghasilkan kluster yang lebih tumpang tindih dengan FPC 0,50 dan Silhouette Score 0,15, yang lebih mencerminkan sifat ambigu dari gejala psikologis. Penelitian ini berkontribusi pada pengembangan kerangka deteksi dini berbasis data untuk depresi mahasiswa.

Kata Kunci : Fuzzy C-Means; K-Means; Depresi; Klusterisasi; Kesehatan Mental



1. PENDAHULUAN

Kesehatan mental mahasiswa merupakan salah satu aspek yang perlu mendapatkan perhatian karena mahasiswa berada pada tahap perkembangan menuju dewasa yang rentan terhadap berbagai permasalahan psikologis. Kondisi kesehatan mental mahasiswa dapat dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. (Pratiwi dkk., 2023) menunjukkan bahwa *self-esteem* dan keharmonisan keluarga memiliki hubungan yang sangat berpengaruh bagi tingkat depresi mahasiswa. Dari penelitian ini terlihat bahwa semakin baik *self-esteem* dan keharmonisan keluarga, sangat mempengaruhi pula kondisi tingkat depresi mahasiswa. Menurut World Health Organization, depresi memengaruhi lebih dari 280 juta orang di seluruh dunia, dengan kelompok usia 18–25 tahun sebagai salah satu kelompok paling rentan (World Health Organization, 2023).

Di lingkungan perguruan tinggi, tekanan akademik yang tinggi, tuntutan finansial, pola tidur yang tidak teratur, kebiasaan makan yang buruk, serta minimnya dukungan sosial menjadi faktor risiko utama yang memicu terjadinya depresi pada mahasiswa (Sumantiawan dkk., 2024). Sejumlah penelitian menunjukkan bahwa prevalensi depresi di kalangan mahasiswa berkisar antara 20% hingga 30%, jauh lebih tinggi dibandingkan populasi umum (Cheung dkk., 2022). Kondisi ini mendorong kebutuhan mendesak akan pendekatan berbasis data yang mampu mendeteksi dan mengidentifikasi mahasiswa berisiko secara lebih akurat dan efisien.

Semakin berkembangnya teknologi, terutama dalam bidang data mining dan machine learning, membuka peluang baru dalam analisis data, termasuk data yang berkaitan dengan kesehatan mental. Salah satu teknik yang dapat digunakan adalah klusterisasi, yaitu metode yang memungkinkan data dikelompokkan berdasarkan kesamaan atau kemiripan karakteristik tanpa memerlukan kelas atau label yang telah ditentukan sebelumnya (Simanjuntak dkk., 2023). Dalam konteks analisis depresi mahasiswa, pendekatan ini sangat relevan karena kondisi psikologis bersifat multidimensional dan tidak selalu dapat diklasifikasikan secara biner.

Dua algoritma klusterisasi yang paling banyak digunakan adalah K-Means dan Fuzzy C-Means (FCM). K-Means merupakan algoritma hard clustering yang menempatkan setiap data ke dalam satu klaster secara tegas berdasarkan jarak Euclidean terdekat (Akbar dkk., 2021). Sementara itu, FCM mengadopsi pendekatan soft clustering yang memungkinkan setiap data memiliki derajat keanggotaan parsial dalam beberapa klaster sekaligus (Uperiati dkk., 2021; Rahakbauw dkk., 2022).

Perbandingan kedua metode ini telah dilakukan pada berbagai domain, seperti pengelompokan tingkat kedisiplinan karyawan (Agustina & Prihandoko, 2018), pemetaan daerah rawan kriminalitas (Firdaus dkk., 2021), klusterisasi performa karyawan (Pramitasari & Nataliani, 2021), analisis kepuasan pelanggan jasa pengiriman

(Fakhriza, 2021), dan penentuan strategi promosi universitas (Herijanto dkk., 2024). Penerapan K-Means juga telah digunakan pada berbagai domain kesehatan, seperti pengelompokan risiko penyakit jantung (Haryadi dkk., 2021), analisis risiko penyakit stroke (Nurhakim dkk., 2024), serta klusterisasi data penggunaan obat (Manik dkk., 2023). Namun, kajian komparatif keduanya secara khusus dalam konteks faktor depresi mahasiswa di Indonesia masih sangat terbatas (Sumantiawan dkk., 2024; Ibrahim dkk., 2024).

Sejumlah penelitian terdahulu yang membandingkan K-Means dan FCM menunjukkan hasil yang beragam bergantung pada domain data. Penelitian oleh (Wulandari dkk., 2022) menemukan bahwa FCM memberikan hasil yang lebih baik pada data ekonomi-kesehatan negara, sementara Rakhmawaty dkk. (2022) menyimpulkan bahwa kedua metode memiliki kekuatan berbeda pada analisis kinerja pegawai. (Pramitasari dkk., 2021) juga membandingkan kedua algoritma untuk klusterisasi karyawan berdasarkan nilai kinerja dengan hasil yang bervariasi tergantung metrik evaluasi yang digunakan. Berdasarkan kesenjangan tersebut, studi ini bertujuan mengevaluasi dan membandingkan kinerja K-Means dan FCM menggunakan tiga metrik evaluasi kuantitatif: Silhouette Score, Davies-Bouldin Index (DBI), dan Sum of Squared Errors (SSE), dalam mengelompokkan mahasiswa berdasarkan faktor depresi guna mendukung pengembangan strategi deteksi dini kesehatan mental di perguruan tinggi.

2. METODE

2.1 Klusterisasi

Data mining merupakan proses ekstraksi pengetahuan atau pola tersembunyi dari kumpulan data berukuran besar dengan memanfaatkan berbagai teknik, seperti statistik, matematika, dan kecerdasan buatan untuk memperoleh informasi yang bermanfaat (Mardiani dkk., 2023). Klusterisasi merupakan teknik data mining yang paling relevan untuk analisis depresi karena memungkinkan pengelompokan data tanpa label yang telah ditentukan sebelumnya (unsupervised learning). Beberapa penelitian terkini menunjukkan bahwa metode klusterisasi mampu mengidentifikasi kelompok mahasiswa berisiko tinggi mengalami depresi dengan hasil yang cukup baik (Zhu dkk., 2022).

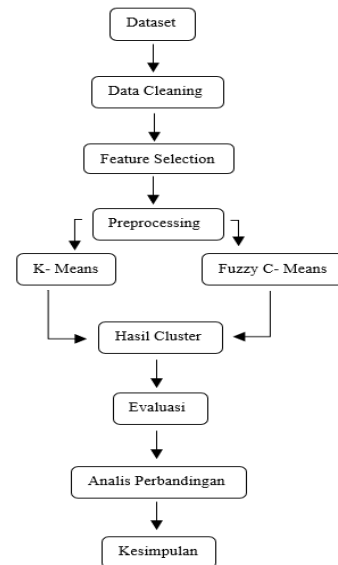
2.2 Dataset dan Pra Pemrosesan Data

Penelitian ini menggunakan dataset Student Depression yang diambil dari platform Kaggle (Opeyemi, 2024). Dataset yang berisi data mahasiswa dengan total 27.901 sampel dan sembilan fitur utama, yaitu: usia, IPK (GPA), kepuasan belajar, durasi tidur, pola makan, riwayat keinginan bunuh diri, tekanan keuangan, riwayat penyakit mental keluarga, dan tingkat depresi, sehingga mampu memberikan representasi yang lebih baik terhadap variasi karakteristik mahasiswa serta meningkatkan reliabilitas proses analisis clustering. Pemilihan

dataset ini didasarkan pada kelengkapan atribut yang merepresentasikan berbagai faktor risiko depresi sehingga sesuai digunakan untuk membandingkan performa algoritma K-Means dan Fuzzy C-Means dalam proses klasterisasi mahasiswa berdasarkan karakteristik tingkat depresinya. Dataset ini telah melalui proses pembersihan data untuk menghilangkan missing values dan outlier ekstrem. Seleksi fitur dilakukan menggunakan algoritma Decision Tree untuk mengidentifikasi fitur-fitur yang paling berkontribusi terhadap variabel tingkat depresi berdasarkan nilai feature importance. Dari proses seleksi tersebut diperoleh sembilan fitur teratas yang digunakan dalam analisis clustering: usia, IPK, kepuasan belajar, durasi tidur, pola makan, riwayat keinginan bunuh diri, tekanan keuangan, riwayat penyakit mental keluarga, dan tingkat depresi.

Tahapan pra-pemrosesan dilakukan secara berbeda untuk masing-masing algoritma sesuai dengan karakteristik komputasinya. Pada K-Means, penanganan data kosong dilakukan dengan menghapus baris yang mengandung missing values (dropna), kemudian fitur numerik dinormalisasi menggunakan StandardScaler untuk menghasilkan distribusi Z-score dengan mean 0 dan standar deviasi 1. Standardisasi ini diperlukan untuk mencegah bias skala pada perhitungan jarak Euclidean. Pada FCM, missing values ditangani melalui imputasi—nilai kolom kategorikal diisi dengan modus sedangkan kolom numerik diisi dengan mean—kemudian fitur numerik dinormalisasi menggunakan MinMaxScaler ke rentang [0,1], yang selaras dengan prinsip logika fuzzy di mana derajat keanggotaan berada pada interval yang sama.

Perbedaan teknik pra-pemrosesan pada kedua algoritma dilakukan berdasarkan karakteristik masing-masing metode. K-Means menggunakan perhitungan jarak Euclidean sehingga memerlukan standardisasi menggunakan StandardScaler agar setiap fitur memiliki skala yang sebanding. Sebaliknya, Fuzzy C-Means menghitung derajat keanggotaan fuzzy sehingga normalisasi menggunakan MinMaxScaler dipilih untuk menjaga rentang nilai antar fitur tetap konsisten. Dengan demikian, perbedaan proses pra-pemrosesan bertujuan menyesuaikan kebutuhan komputasi masing-masing algoritma. Perlu diakui bahwa perbedaan teknik ini merupakan keterbatasan dalam hal perbandingan yang sepenuhnya setara, sehingga interpretasi perbedaan nilai metrik evaluasi perlu mempertimbangkan pengaruh perbedaan skala ruang fitur tersebut. Alur penelitian yang digunakan dapat dilihat pada framework berikut:



Gambar 1. Framework penelitian

2.3 Algoritma Fuzzy C-Means

Fuzzy C-Means (FCM) adalah metode pengelompokan data di mana nilai derajat keanggotaan menentukan keberadaan setiap data dalam suatu kluster. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kluster yang terbentuk dengan derajat keanggotaan berbeda antara 0 hingga 1 (Uperiati dkk., 2021).

Pada penelitian ini, algoritma Fuzzy C-Means diimplementasikan menggunakan dua cluster ($c = 2$) dengan parameter fuzziness (m) sebesar 2, batas galat (error tolerance) sebesar 0,005, dan maksimum iterasi sebanyak 1000 iterasi. Jumlah cluster ditetapkan sama dengan algoritma K-Means agar proses perbandingan dilakukan pada konfigurasi yang identik. Inisialisasi matriks partisi dilakukan secara acak sesuai mekanisme bawaan algoritma Fuzzy C-Means hingga memenuhi kriteria konvergensi yang ditentukan. Selanjutnya algoritma FCM meminimalkan fungsi objektif berikut:

$$J_m = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^w |x_i - v_k|^2 \quad (1)$$

di mana J_m adalah fungsi objektif, μ_{ik} adalah derajat keanggotaan data x_i pada kluster ke- k , w adalah parameter pembobot fuzzy ($w = 2$), dan v_k adalah pusat kluster ke- k . Langkah-langkah algoritma FCM adalah sebagai berikut:

- Masukkan data X berupa matriks berukuran $n \times m$ (n = jumlah sampel, m = jumlah atribut). Tentukan jumlah kluster (c), parameter w , maksimum iterasi, dan nilai error terkecil (ϵ).
- Bangkitkan bilangan acak sebagai elemen matriks partisi awal U . Hitung jumlah setiap kolom.
- Hitung derajat keanggotaan μ_{ik} menggunakan persamaan 2, kemudian

hitung pusat kluster V_{kj} menggunakan persamaan 3.

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{|x_i - v_k|}{|x_i - v_j|} \right)^{\frac{2}{w-1}}} \quad (2)$$

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w \cdot x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

- d) Hitung fungsi objektif pada iterasi ke- t menggunakan persamaan 1, kemudian perbarui matriks partisi menggunakan persamaan 5.
- e) Cek kondisi berhenti: jika $|P_t - P_{t-1}| < \epsilon$ atau $t > \text{MaxIter}$, maka berhenti; jika tidak, ulangi langkah c.

2.4 Algoritma K-Means

K-Means merupakan salah satu metode klusterisasi yang berusaha membagi objek ke dalam satu atau beberapa kluster berdasarkan karakteristiknya (Akbar dkk., 2021). Proses K-Means dimulai dengan menentukan jumlah kluster (K) dan memilih secara acak K titik sebagai pusat kluster (centroid). Setiap data kemudian dihitung jaraknya ke semua centroid menggunakan Euclidean Distance:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (4)$$

Setelah penetapan kluster awal, centroid diperbarui berdasarkan rata-rata posisi data dalam masing-masing kluster. Langkah ini diulang hingga posisi centroid stabil (konvergen) atau jumlah iterasi maksimum tercapai. Penentuan jumlah kluster optimal dilakukan menggunakan Elbow Method dengan menganalisis nilai Within-Cluster Sum of Squares (WCSS). Berdasarkan hasil Elbow Method, jumlah kluster optimal yang diperoleh adalah dua kluster ($K = 2$). Nilai tersebut kemudian digunakan sebagai jumlah kluster pada algoritma Fuzzy C-Means agar kedua algoritma dibandingkan menggunakan konfigurasi yang sama. Pendekatan ini bertujuan mengurangi bias akibat perbedaan jumlah kluster sehingga evaluasi kinerja dapat dilakukan secara lebih objektif.

2.5 Pengujian Hasil Clustering

Evaluasi kinerja kedua algoritma dilakukan menggunakan tiga metrik pengujian hasil clustering yang umum digunakan dalam penelitian komparatif (Puspitasari dkk., 2023):

12.5.1 Silhouette Score (SS)

Silhouette Score digunakan untuk mengukur kualitas dan kekuatan hasil klusterisasi, yaitu seberapa baik suatu data ditempatkan dalam kluster dibandingkan kluster lain (Kamila dkk., 2019). Nilai silhouette yang lebih dari 0 atau positif mengindikasikan kluster yang kuat, sedangkan nilai

negatif menunjukkan struktur kluster yang lemah. Nilai SS mendekati 1 mengindikasikan klusterisasi yang sangat baik, sedangkan nilai mendekati -1 mengindikasikan penempatan yang tidak tepat. Rumus perhitungan:

$$s(i) = \frac{b(i) - a(i)}{(a(i), b(i))} \quad (5)$$

di mana $a(i)$ adalah rata-rata jarak data i ke semua data lain dalam kluster yang sama, dan $b(i)$ adalah rata-rata jarak data i ke semua data pada kluster terdekat yang berbeda.

12.5.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) merupakan metrik evaluasi internal yang mengukur kualitas klusterisasi berdasarkan rasio antara sebaran intra-kluster dan jarak antar-kluster (Wahyudi dkk., 2022). Semakin kecil nilai DBI yang dihasilkan, semakin baik kualitas kluster yang terbentuk, karena menunjukkan bahwa kluster bersifat kompak dan terpisah dengan baik satu sama lain. Rumus DBI:

$$DBI = \frac{1}{K} \sum_{i=1}^K \left(\frac{S_i + S_j}{d(c_i, c_j)} \right) \quad (6)$$

di mana K adalah jumlah kluster, S_i adalah rata-rata jarak data dalam kluster i ke centroid-nya, dan $d(c_i, c_j)$ adalah jarak antara centroid kluster i dan j .

12.5.3 Sum of Squared Errors (SSE)

Sum of Squared Errors (SSE) atau Within-Cluster Sum of Squares (WCSS) digunakan untuk mengukur total kesalahan kuadrat antara setiap data dan centroid kluster (Refialy dkk., 2021). Nilai SSE yang semakin kecil atau mendekati 0 menandakan hasil clustering yang lebih baik karena menunjukkan data lebih padat di sekitar centroid masing-masing kluster. Rumus SSE:

$$SSE = \sum_{k=1}^K \sum_{i \in C_k} |x_i - c_k|^2 \quad (7)$$

3. PEMBAHASAN

3.1 Prapemrosesan dan Normalisasi Data

Penanganan missing values dan normalisasi fitur numerik diterapkan sesuai dengan karakteristik komputasi masing-masing algoritma. Pada K-Means, missing values dihapus menggunakan dropna dan fitur numerik distandarisasi dengan StandardScaler. Pada FCM, missing values diimputasi dengan modus (kategorikal) dan mean (numerik), kemudian fitur dinormalisasi menggunakan MinMaxScaler ke rentang $[0,1]$. Seleksi fitur dilakukan menggunakan Decision Tree untuk mengidentifikasi sembilan fitur paling relevan: usia, IPK, kepuasan belajar, durasi tidur, pola makan, riwayat keinginan bunuh diri, tekanan keuangan, riwayat penyakit mental keluarga, dan tingkat depresi.

3.2 Hasil Clustering K-Means

Penentuan jumlah klaster optimal dilakukan menggunakan Elbow Method, yang menunjukkan K=2 sebagai titik siku yang paling signifikan. K-Means mengelompokkan data menjadi dua klaster dengan hasil sebagai berikut:

Tabel 1. Hasil K-Means

Fitur	Claster 0	Claster 1
Umur	25.40	26.55
GPA	7.67	7.64
Kepuasan Belajar	2.86	3.09
Waktu Tidur	1.49	1.54
Pola Makan	1.57	1.28
Pernah Punya Keinginan Bunuh Diri	1.00	0.0004
Tekanan Keuangan	3.37	2.74
Riwayat Penyakit Mental Keluarga	0.49	0.46
Depresi	0.79	0.23

Dari Tabel 1 terlihat bahwa Cluster 0 memiliki nilai tekanan finansial, keinginan bunuh diri, dan tingkat depresi yang lebih tinggi dibandingkan Cluster 1. Dengan perbedaan skor depresi yang sangat signifikan (0,79 vs 0,23), Cluster 0 dapat diinterpretasikan sebagai kelompok mahasiswa dengan risiko depresi tinggi, sementara Cluster 1 merepresentasikan mahasiswa yang lebih stabil secara mental.

3.3 Hasil Clustering Fuzzy C-Means

FCM membentuk dua klaster dengan mempertimbangkan derajat keanggotaan tiap data. Evaluasi menggunakan Fuzzy Partition Coefficient (FPC) menghasilkan nilai 0,50, yang menunjukkan pembagian klaster yang seimbang namun tidak sejelas K-Means. Jumlah anggota dominan: Cluster 0 sebanyak 14.449 data dan Cluster 1 sebanyak 13.452 data.

Tabel 2. Hasil Fuzzy C-Means

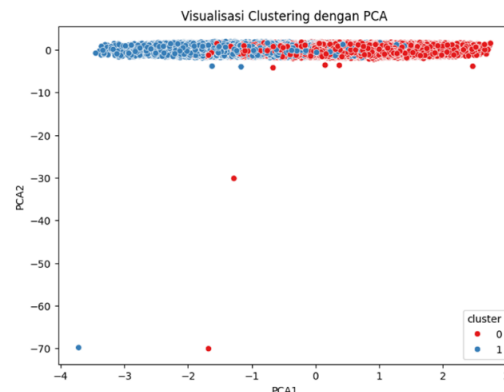
Fitur	Claster 0	Claster 1
Umur	25.76	25.89
GPA	7.71	7.60
Kepuasan Belajar	2.89	3.00
Waktu Tidur	1.51	1.50
Pola Makan	1.63	1.28
Pernah Punya Keinginan Bunuh Diri	0.68	0.58
Tekanan Keuangan	3.25	3.02
Riwayat Penyakit Mental Keluarga	0.48	0.49
Depresi	0.63	0.54

Dari Tabel 2 terlihat bahwa klaster dari FCM menunjukkan perbedaan tingkat depresi yang tidak

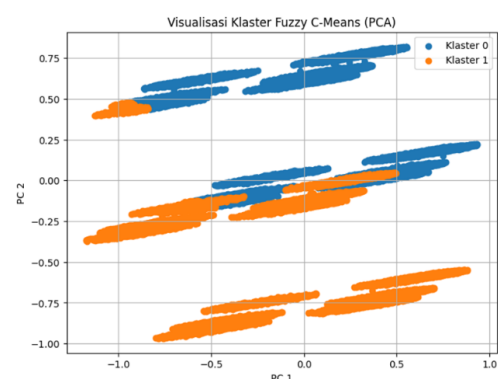
setajam K-Means (0,63 vs 0,54). Namun, karakteristik ini justru lebih mencerminkan sifat ambiguitas alami dari gejala psikologis, mengingat depresi tidak selalu dapat diklasifikasikan secara biner.

3.4 Hasil Clustering dengan PCA

Visualisasi hasil clustering menggunakan Principal Component Analysis (PCA) dilakukan untuk mereduksi dimensi data ke dua komponen utama (PC1 dan PC2) agar dapat ditampilkan secara grafis. PC1 dan PC2 secara kumulatif menjelaskan sekitar 35–40% variansi total data, yang merupakan karakteristik umum pada dataset berdimensi tinggi dengan atribut yang saling berkorelasi rendah. Pada visualisasi K-Means (Gambar 3), klaster berdistribusi cukup jelas meskipun terdapat sedikit tumpang tindih di area pusat, dan beberapa outlier terlihat jauh dari pusat data. Pada visualisasi FCM (Gambar 4), klaster tampak lebih menyebar dan tumpang tindih, sesuai dengan sifat fuzzy dari metode ini yang lebih mencerminkan realitas kompleks kondisi mental mahasiswa.



Gambar 3. Visualisasi Hasil Clustering



Gambar 4. Visualisasi Hasil Clustering dengan C-Means

3.5 Evaluasi Kinerja Clustering

Pengujian kinerja kedua algoritma dilakukan menggunakan tiga metrik evaluasi: Silhouette Score (SS), Davies-Bouldin Index (DBI), dan Sum of Squared Errors (SSE). Hasil pengujian ditampilkan pada Tabel 3.

Tabel 3. Perbandingan Evaluasi Kinerja *K-Means* dan *Fuzzy C-Means*

Matrix Evaluasi	K-Means	Fuzzy C-Means	Keeterangan
Silhouette Score	0.18	0.15	Lebih tinggi lebih baik
Davies-Bouldin Index	1.89	2.14	Lebih rendah lebih baik
SSE (WCSS)	12.543	14.872	Lebih rendah lebih baik
FPC	-	0.50	Khusus FCM

Berdasarkan Tabel 3, K-Means unggul pada ketiga metrik evaluasi kuantitatif. Nilai Silhouette Score K-Means (0,18) lebih tinggi dari FCM (0,15), mengindikasikan bahwa K-Means menghasilkan kluster dengan separabilitas yang lebih baik. Meskipun demikian, kedua nilai Silhouette Score tergolong rendah secara absolut, yang mengindikasikan bahwa batas pemisahan antar kluster tidak tajam pada dataset ini. Hal ini konsisten dengan sifat data psikologis yang bersifat kontinu dan multidimensional, di mana pemisahan kluster yang tegas memang tidak selalu dapat dicapai. Nilai DBI K-Means (1,89) lebih rendah dari FCM (2,14), membuktikan bahwa kluster yang dihasilkan K-Means lebih kompak dan terpisah dengan baik. Nilai SSE K-Means (12,543) juga lebih kecil dibandingkan FCM (14,872), menunjukkan bahwa data lebih padat di sekitar centroid kluster K-Means. Perlu dicatat bahwa SSE kedua algoritma dihitung dari ruang fitur yang berbeda (StandardScaler untuk K-Means dan MinMaxScaler untuk FCM), sehingga perbandingan nilai SSE bersifat indikatif dan harus diinterpretasikan dengan mempertimbangkan perbedaan skala tersebut.

Untuk mengevaluasi konsistensi hasil clustering, penelitian ini juga melakukan uji stabilitas terhadap kedua algoritma melalui 10 kali pengulangan dengan variasi inisialisasi, yaitu parameter *seed* pada Fuzzy C-Means dan *random_state* pada K-Means. Hasil pengujian menunjukkan bahwa Fuzzy C-Means menghasilkan nilai FPC yang konsisten dengan rata-rata sebesar 0,5000 dan standar deviasi sebesar $8,81 \times 10^{-10}$, sedangkan nilai SSE memiliki rata-rata sebesar 40.552,27 dengan standar deviasi sebesar 0,326. Pada K-Means, rata-rata nilai SSE sebesar 219.415,51 dengan standar deviasi sebesar 0,0177, sementara nilai Silhouette Score dan Davies-Bouldin Index juga menunjukkan variasi yang sangat kecil pada setiap pengujian. Hasil tersebut menunjukkan bahwa kedua algoritma memiliki stabilitas yang baik karena perubahan inisialisasi tidak memberikan pengaruh yang signifikan terhadap kualitas clustering.

Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan kualitas kluster yang lebih baik dibandingkan Fuzzy C-Means, ditunjukkan oleh

nilai Silhouette Score yang lebih tinggi serta nilai Davies-Bouldin Index (DBI) dan Sum of Squared Errors (SSE) yang lebih rendah. Hasil tersebut menunjukkan bahwa kluster yang dihasilkan K-Means memiliki tingkat kekompakan (cohesion) yang lebih baik dan pemisahan antarkluster (separation) yang lebih jelas pada dataset yang digunakan. Temuan ini sejalan dengan penelitian Herijanto dan Riti (2024) yang membandingkan K-Means dan Fuzzy C-Means pada data promosi universitas dan melaporkan bahwa K-Means mampu menghasilkan kualitas kluster yang lebih baik pada data dengan karakteristik batas pemisahan yang relatif jelas. Hasil ini juga konsisten dengan temuan Pramitasari dan Nataliani (2021) serta Rakhmawaty dkk. (2022) yang menunjukkan bahwa kinerja relatif kedua algoritma bervariasi tergantung pada karakteristik domain data dan metrik evaluasi yang digunakan. Pada domain psikologis seperti penelitian ini, K-Means menghasilkan segmentasi yang lebih kontras, sementara FCM lebih mampu merepresentasikan kondisi yang bersifat gradual. Di sisi lain, Fuzzy C-Means tetap memiliki keunggulan karena setiap data dapat memiliki derajat keanggotaan pada lebih dari satu kluster. Pendekatan ini lebih mampu merepresentasikan kondisi yang bersifat tidak pasti (uncertainty), seperti karakteristik kesehatan mental mahasiswa yang tidak selalu dapat dipisahkan secara tegas. Oleh karena itu, meskipun pada penelitian ini K-Means memberikan hasil evaluasi yang lebih baik berdasarkan metrik internal, Fuzzy C-Means tetap relevan digunakan apabila tujuan penelitian lebih menekankan pada analisis tingkat keanggotaan setiap objek terhadap masing-masing kluster.

3.6 Perbandingan Karakteristik Algoritma

Tabel 4. Perbandingan Evaluasi Kinerja *K-Means* dan *Fuzzy C-Means*

Aspek	K-Means	Fuzzy C-Means
Teknik Keanggotaan	Tegas (hard clustering)	Fuzzy (soft clustering)
Pemisahan Kluster	Lebih tajam dan kontras	Lebih tumpang tindih
Depresi Antar Kluster	Jauh berbeda (0.79 vs 0.23)	Lebih dekat (0.63 vs 0.54)
Silhouette Score	0.18 (lebih baik)	0.15
Davies-Bouldin Index	1.89 (lebih baik)	2.14
SSE	12.543 (lebih baik)	14.872
Interpretasi Psikologis	Jelas membedakan tingkat depresi-	Mencerminkan spektrum gejala

Secara keseluruhan, K-Means lebih efektif dalam memisahkan mahasiswa dengan tingkat depresi

tinggi dan rendah secara kuantitatif. Hal ini ditunjukkan oleh performa yang lebih baik pada ketiga metrik evaluasi dan perbedaan signifikan pada fitur tekanan keuangan, keinginan bunuh diri, dan tingkat depresi antar klaster. Sementara itu, FCM memberikan hasil yang lebih realistis untuk kasus psikologis yang bersifat kontinu dan ambigu, meskipun kurang tegas dalam pemisahan klaster.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja algoritma Fuzzy C-Means (FCM) dan K-Means dalam mengelompokkan mahasiswa berdasarkan faktor-faktor yang berkaitan dengan depresi. Berdasarkan pengujian menggunakan tiga metrik evaluasi—Silhouette Score, Davies-Bouldin Index (DBI), dan Sum of Squared Errors (SSE)—K-Means terbukti lebih unggul secara kuantitatif dengan nilai SS 0,18, DBI 1,89, dan SSE 12.543, dibandingkan FCM dengan nilai SS 0,15, DBI 2,14, dan SSE 14.872.

Dari sisi interpretasi klaster, K-Means menghasilkan pengelompokan yang lebih tegas dengan perbedaan skor depresi yang signifikan antara Cluster 0 (0,79) dan Cluster 1 (0,23), sehingga lebih efektif untuk mengidentifikasi mahasiswa berisiko tinggi. FCM menghasilkan klaster yang lebih tumpang tindih dengan nilai FPC 0,50, yang lebih mencerminkan sifat ambigu dan kontinu dari gejala psikologis. Dengan demikian, K-Means lebih sesuai untuk segmentasi risiko yang jelas, berdasarkan hasil penelitian ini, FCM lebih relevan untuk analisis spektrum psikologis yang memerlukan tingkat keanggotaan parsial. Selain itu, hasil uji stabilitas menunjukkan bahwa K-Means maupun Fuzzy C-Means memiliki konsistensi hasil clustering yang baik terhadap variasi inisialisasi, sehingga kedua algoritma dapat dianggap cukup stabil untuk digunakan pada dataset penelitian ini.

Penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari satu sumber, yaitu Student Depression Dataset, sehingga hasil clustering belum tentu dapat digeneralisasikan pada populasi atau dataset lain. Selain itu, penelitian ini belum mengevaluasi pengaruh variasi parameter algoritma maupun jumlah cluster terhadap kualitas hasil clustering. Penelitian selanjutnya dapat menggunakan dataset dari berbagai sumber atau institusi pendidikan untuk meningkatkan kemampuan generalisasi hasil penelitian. Selain itu, evaluasi dapat diperluas dengan membandingkan algoritma clustering lainnya, seperti DBSCAN, Agglomerative Clustering, atau Gaussian Mixture Model, serta melakukan analisis sensitivitas terhadap parameter dan jumlah cluster guna memperoleh hasil clustering yang lebih komprehensif.

PUSTAKA

Agustina, N., & Prihandoko, P. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem*

Dan Teknologi Informasi), 2(3), 621–626.
<https://doi.org/10.29207/resti.v2i3.492>

Akbar, M. F., Mahmudi, A., & Zahro, H. Z. (2021). K-Means Clustering untuk Pengelompokan Tingkat Resiko Ibu Hamil. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 581–588. <https://doi.org/10.36040/jati.v5i2.3772>

Cheung, T., Fong, T. K. H., & Bressington, D. (2022). COVID-19 Under the SARS Cloud: Mental Health Nursing During the Pandemic in Hong Kong. *Journal of Psychiatric and Mental Health Nursing*, 28(2), 115–117. <https://doi.org/10.1111/jpm.12639>

Dewa, C. S., Trojanowski, L., Joosen, M. C. W., & Bonato, S. (2022). Interventions to Improve Presenteeism in the Workplace: A Systematic Review. *Journal of Occupational and Environmental Medicine*, 59(10), 677–691. <https://doi.org/10.1097/JOM.0000000000001106>

Fakhriza, M. H. (2021). Comparison Of K-Means Method And Fuzzy Clustering Algorithm In Determining Customer Satisfaction Test In Delivery Services. *JIKA (Jurnal Informatika)*, 5(2), 152. <https://doi.org/10.31000/jika.v5i2.4511>

Firdaus, H. S., Nugraha, A. L., Sasmito, B., & Awaluddin, M. (2021). Perbandingan Metode Fuzzy C-Means dan K-Means untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang. *Elipsoida: Jurnal Geodesi dan Geomatika*, 4(01). <https://doi.org/10.14710/elipsoida.2021.9219>

Haryadi, D., & Atmaja, D. M. U. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung. *DES 2021 Journal of Informatics and Communications Technology*, 3(2), 51–066. https://doi.org/10.52661/j_ict.v3i2

Herijanto, C. K., & Riti, Y. F. (2024). Perbandingan Algoritma K-Means Clustering dan Fuzzy C-Means untuk Penentuan Strategi Promosi Universitas Katolik Darma Cendika. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3931–3937. <https://doi.org/10.36040/jati.v8i3.9660>

Ibrahim, A. K., Kelly, S. J., Adams, C. E., & Glazebrook, C. (2024). A Systematic Review of Studies of Depression Prevalence in University Students. *Journal of Psychiatric Research*, 47(3), 391–400. <https://doi.org/10.1016/j.jpsychires.2012.11.015>

Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 5(1), 119–125.

Manik, T. G., Rahayu, W. I., & Fathonah, Rd. N. S. (2023). Perbandingan Metode Fuzzy C-Means

- dan K-Means Clustering pada Data Penggunaan Obat di R.S National Hospital Surabaya. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 591–597.
<https://doi.org/10.36040/jati.v7i1.6430>
- Mardiani, E., Rahmansyah, N., Kurniati, I., Matondang, N., Tesalonika, Z. A., & Romzy, I. (2023). Membandingkan algoritma data mining dengan tools Orange untuk social economy. Digital Transformation Technology, 3(2).
- Nurhakim, B., Septiani, I., Anam, K., & Pratama, D. (2024). Penerapan Algoritma K-Means Clustering dalam Menganalisis Resiko Penyakit Stroke. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 318–322.
<https://doi.org/10.36040/jati.v8i1.8452>
- Opeiyemi, S. (2024). Student Depression Dataset. Kaggle.
<https://www.kaggle.com/datasets/hopesb/student-depression-dataset>
- Opeiyemi, S. (2024). Student Depression Dataset. Kaggle.
<https://www.kaggle.com/datasets/hopesb/student-depression-dataset>
- Pramitasari, A. E., & Nataliani, Y. (2021). Perbandingan Clustering Karyawan Berdasarkan Nilai Kinerja dengan Algoritma K-Means dan Fuzzy C-Means. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 8(3), 1119–1132.
<https://doi.org/10.35957/jatisi.v8i3.957>
- Puspitasari, N., Lempas, G., Hamdani, Haviluddin, & Septiari, A. (2023). Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia. Building of Informatics, Technology and Science (BITS), 4(4), 2015–2027.
<https://doi.org/10.47065/bits.v4i4.2994>
- Rahakbauw, D., Huwae, V., & Talakua, M. (2022). Implementasi Metode Fuzzy C-Means (FCM) dalam Pengelompokan Data Klaim Asuransi Kendaraan Bermotor. Jurnal Ilmu Matematika dan Terapan, 16(1), 71–82.
- Rakhmawaty, N., Nasution, Y. N., & Amijaya, F. D. T. (2022). Perbandingan Metode K-Means Dan Metode Fuzzy C-Means (FCM) Pada Analisis Kinerja Pegawai PT. Cemara Khatulistiwa Persada Bontang. EKSPONENSIAL, 13(1), 63.
<https://doi.org/10.30872/eksponsional.v13i1.886>
- Refialy, L. P., Maitimu, H., & Pesulima, M. S. (2021). Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster. Techno.Com, 20(2), 321–329.
<https://doi.org/10.33633/tc.v20i2.4572>
- Sumantiawan, D. I., Amaliyah, S., Narulita, S., Kholilurrahman, M., & Suharmanto, A. Y. (2024). Analisis Faktor Status Depresi Kehidupan Mahasiswa Menggunakan Machine Learning. Digital Transformation Technology, 4(1), 705–713.
<https://doi.org/10.47709/digitech.v4i1.4577>
- Simanjuntak, H. T. A., Silaban, P. E. P., Manurung, J. K. S., & Sormin, V. H. (2023). Klasterisasi berita Bahasa Indonesia dengan menggunakan K-Means dan Word Embedding. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(3), 641–652.
<https://doi.org/10.25126/jtiik.2023106468>
- Uperiati, A., Bettiza, M., & Puspasari, A. (2021). Perbandingan Metode Fuzzy C-Means dan K-Means dalam Klasifikasi Kelulusan Mahasiswa. Jurnal Sustainable: Jurnal Hasil Penelitian dan Industri Terapan, 9(2), 75–81.
<https://doi.org/10.31629/sustainable.v9i2.1409>
- Wahyudi, M., Solikhun, S., & Pujiastuti, L. (2022). Komparasi K-Means Clustering dan K-Medoids Clustering dalam Mengelompokkan Produksi Susu Segar di Indonesia Berdasarkan Nilai DBI. Jurnal Bumigora Information Technology, 4(2), 243–254.
- World Health Organization. (2023). Depressive disorder (depression). WHO Fact Sheet. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/depression>
- Wulandari, L., & Yogantara, B. O. (2022). Algorithm Analysis of K-Means and Fuzzy C-Means for Clustering Countries Based on Economy and Health. Faktor Exacta, 15(2).
<https://doi.org/10.30998/faktorexacta.v15i2.12106>
- Zhu, J., Xie, Q., Luo, J., & Chen, Z. (2022). Identifying Depression Based on Clustering and Machine Learning Methods. Frontiers in Psychiatry, 13, 877195.
<https://doi.org/10.3389/fpsy.2022.877195>