

Penerapan K-Means Clustering pada Absensi Mahasiswa Semester Ganjil untuk Mengelompokkan Kehadiran Mahasiswa

¹Rezti Deawinda Parinduri

¹Universitas Muhammadiyah Tapanuli Selatan; Jalan St Mohd Arief No 32 Padangsidimpuan,
(0634)21696
e-mail: ¹rezti@um-tapsel.ac.id

Abstrak

Kedisiplinan mahasiswa dalam hal kehadiran menjadi indikator penting dalam menjamin mutu pembelajaran di perguruan tinggi. Penelitian ini bertujuan untuk menganalisis pola kehadiran mahasiswa pada mata kuliah semester ganjil dengan menerapkan algoritma K-Means Clustering sebagai bagian dari proses Knowledge Discovery in Database (KDD). Hipotesis penelitian menyatakan bahwa mahasiswa dapat dikelompokkan ke dalam kategori tingkat kehadiran yang berbeda secara signifikan. Data yang digunakan berupa absensi 35 mahasiswa selama 16 pertemuan. Proses penelitian mencakup tahapan KDD: pengumpulan data, pra-pemrosesan, penerapan algoritma K-Means, visualisasi dengan PCA, serta evaluasi menggunakan Silhouette Score. Hasil menunjukkan terbentuknya tiga cluster: kehadiran tinggi (90%), sedang (70%), dan rendah (50%), dengan skor siluet rata-rata sebesar 0.72 yang menunjukkan pemisahan cluster yang cukup baik. Kesimpulannya, algoritma K-Means efektif dalam mengelompokkan pola kehadiran mahasiswa secara objektif. Saran diberikan agar pihak universitas menggunakan hasil ini sebagai dasar perumusan kebijakan peningkatan kedisiplinan. Penelitian ini penting sebagai langkah awal dalam memahami perilaku mahasiswa berbasis data dan dapat dikembangkan lebih lanjut dengan menambahkan variabel serta mengeksplorasi algoritma lain.

Kata Kunci : K-Means Clustering, Kehadiran Mahasiswa, Absensi

Abstract

Student discipline, particularly in terms of attendance, is a crucial indicator in ensuring the quality of learning in higher education. This study aims to analyze student attendance patterns in odd semester courses by applying the K-Means Clustering algorithm as part of the Knowledge Discovery in Database (KDD) process. The research hypothesis states that students can be grouped into significantly different categories based on their attendance levels. The data used consists of attendance records from 35 students over 16 sessions. The research process follows the KDD stages: data collection, preprocessing, implementation of the K-Means algorithm, visualization using PCA, and evaluation using the Silhouette Score. The results indicate the formation of three clusters: high attendance (90%), moderate attendance (70%), and low attendance (50%), with an average Silhouette Score of 0.72, indicating a fairly good cluster separation. The study concludes that the K-Means algorithm is effective in objectively clustering student attendance patterns. It is recommended that universities use these findings as a basis for formulating policies to improve student discipline. This research is important as an initial step in understanding student behavior through data-driven approaches and can be further developed by adding more variables and exploring other clustering algorithms.

Keywords : K-Means Clustering, Student Attendance, Absence

PENDAHULUAN

Pertumbuhan teknologi informasi dimana saat ini berkembang secara cepat menuntut kita untuk mampu memanfaatkan teknologi informasi secara maksimal, memanfaatkan teknologi informasi digunakan dalam berbagai aspek kehidupan manusia untuk menciptakan alur kerja yang mudah digunakan dan menyediakan informasi untuk kebutuhan spesifik yang lebih rinci, mudah mengikuti, dan efisiensi. Pemanfaatan teknologi juga akan meningkatkan keakuratan data dan informasi [1]. Pembangunan pendidikan sangat diperlukan untuk membangun karakter Sumber Daya Manusia (SDM) yang berkualitas. Universitas sebagai lembaga pendidikan

memiliki tanggung jawab penting dalam pembentukan karakter, terutama karakter disiplin mahasiswa. Absensi adalah kegiatan pengambilan data untuk mengetahui jumlah mahasiswa yang hadir [2]. Data mining adalah metode untuk mengekstraksi informasi dari database besar dengan menggabungkan berbagai bidang ilmiah seperti statistika, matematika, kecerdasan buatan, dan machine learning. Tujuan utamanya adalah untuk menemukan pola serta mengenali data yang bermanfaat. Salah satu metode yang digunakan dalam penelitian ini adalah clustering, yang merupakan proses pengelompokan data secara otomatis tanpa menggunakan label kelas. Algoritma K-Means, sebagai salah satu algoritma clustering yang paling populer, dipilih karena keunggulannya yang mudah diimplementasikan dan efektif dalam mengelompokkan data berdasarkan kemiripan tertentu.

Dalam konteks ini, K-Means digunakan untuk menganalisis dan mengelompokkan tingkat kehadiran mahasiswa berdasarkan data absensi pada mata kuliah semester ganjil [3]. Algoritma K-Means merupakan salah satu algoritma clustering yang termasuk dalam kelompok unsupervised learning. Algoritma ini digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi. K-Means menerima masukan berupa data tanpa label kelas, sehingga komputer dapat mengelompokkan data berdasarkan kesamaan pola tanpa mengetahui terlebih dahulu target kelasnya. Dalam konteks penelitian ini, data yang diterima adalah data absensi mahasiswa, dan algoritma ini digunakan untuk mengelompokkan mahasiswa ke dalam beberapa cluster berdasarkan tingkat kehadiran mereka pada mata kuliah semester ganjil [4]. Ketidakhadiran atau membolos dapat terjadi akibat ketidakminatan mahasiswa terhadap mata kuliah, yang dapat berdampak negatif pada prestasi akademik mereka. Oleh karena itu, pemantauan kehadiran mahasiswa menggunakan metode K-Means Clustering diharapkan dapat memberikan gambaran yang lebih jelas mengenai pola kehadiran mahasiswa dan membantu pihak universitas mengambil tindakan yang lebih tepat untuk meningkatkan disiplin dan kualitas pembelajaran [5]. Dengan demikian, inovasi ini tidak hanya akan meningkatkan efisiensi pemantauan kehadiran, tetapi juga berkontribusi pada peningkatan kualitas pendidikan secara keseluruhan di universitas tersebut [6]

METODE PENELITIAN

Metode penelitian adalah langkah-langkah sistematis yang diterapkan untuk mengumpulkan, mengolah, dan menganalisis data guna menjawab rumusan masalah atau menguji hipotesis dalam suatu penelitian. Dalam penelitian ini, metode yang digunakan adalah proses Knowledge Discovery in Database (KDD) dengan pendekatan kuantitatif. Pendekatan kuantitatif digunakan karena penelitian ini berfokus pada data numerik yang dapat diukur dan dianalisis secara statistik, dalam hal ini adalah data kehadiran mahasiswa [7]. Proses KDD merupakan kerangka kerja yang umum digunakan dalam bidang data mining untuk menemukan pengetahuan yang bermanfaat dari sekumpulan data yang besar. Tahapan KDD dalam penelitian ini terdiri dari beberapa langkah utama, antara lain:

I. Pengumpulan Data

Pada tahap ini, data absensi mahasiswa pada mata kuliah semester ganjil dikumpulkan. Data yang diperoleh mencakup informasi tentang kehadiran mahasiswa dalam setiap pertemuan perkuliahan selama semester ganjil. Pengumpulan data dilakukan melalui sistem absensi yang tersedia di universitas, baik melalui sistem elektronik (misalnya sistem informasi akademik atau aplikasi absensi digital) maupun manual jika diperlukan. Data yang dikumpulkan meliputi:

- a. Nama mahasiswa
- b. Tanggal dan waktu pertemuan kuliah
- c. Status kehadiran (Hadir/Tidak Hadir/Izin/Sakit)
- d. Jumlah pertemuan yang dihadiri oleh mahasiswa
- e. Jumlah pertemuan total dalam mata kuliah tersebut

Pengumpulan data ini bertujuan untuk memperoleh informasi yang valid dan representatif tentang kehadiran mahasiswa pada mata kuliah semester ganjil.

II. Penerapan Algoritma K-Means

Setelah data siap, tahap selanjutnya adalah penerapan algoritma K-Means untuk mengelompokkan mahasiswa berdasarkan tingkat kehadiran mereka. Algoritma ini akan mengelompokkan mahasiswa ke dalam beberapa cluster sesuai dengan kemiripan pola kehadiran. Proses ini terdiri dari:

- a. Inisialisasi Cluster: Tentukan jumlah cluster (k) yang diinginkan. Jumlah cluster ini dapat ditentukan berdasarkan tujuan penelitian atau menggunakan metode seperti Elbow Method untuk menemukan jumlah cluster yang optimal.
- b. Penentuan Titik Pusat Cluster (Centroid): Algoritma K-Means akan menginisialisasi titik pusat cluster secara acak.
- c. Penugasan Data ke Cluster: Setiap data (mahasiswa) akan dikelompokkan ke cluster berdasarkan kedekatannya dengan titik pusat cluster. Mahasiswa dengan tingkat kehadiran yang mirip akan dikelompokkan bersama.
- d. Perhitungan Ulang Centroid: Setelah data dikelompokkan, posisi centroid dihitung ulang berdasarkan rata-rata posisi data dalam cluster tersebut.
- e. Iterasi: Proses penugasan data dan perhitungan ulang centroid dilakukan berulang kali sampai posisi centroid tidak berubah atau mencapai konvergensi.

HASIL DAN PEMBAHASAN

Pada bagian ini, kami akan membahas secara rinci hasil penerapan algoritma K-Means Clustering pada data absensi mahasiswa. Proses ini bertujuan untuk mengelompokkan mahasiswa berdasarkan tingkat kehadiran mereka selama satu semester, yang memberikan wawasan mengenai pola kedisiplinan dan komitmen mahasiswa terhadap perkuliahan.

I. Deskripsi Data

Sebelum menganalisis hasil clustering, penting untuk memahami deskripsi data absensi yang digunakan dalam penelitian ini:

- a. Jumlah Mahasiswa: 35 mahasiswa
- b. Jumlah Pertemuan: 16 pertemuan kuliah
- c. Status Kehadiran: Hadir, Tidak Hadir, Izin, Sakit

Data absensi yang digunakan mencatat kehadiran mahasiswa pada setiap pertemuan selama semester ganjil. Data ini mencakup informasi tentang apakah mahasiswa hadir, tidak hadir, izin, atau sakit pada masing-masing pertemuan. Proses pra-pemrosesan yang dilakukan meliputi pembersihan data dan normalisasi agar data yang digunakan sesuai dengan standar untuk diterapkan pada algoritma K-Means. Tahap normalisasi penting dilakukan agar variasi dalam data kehadiran, yang beragam antara mahasiswa satu dengan yang lainnya, tidak mempengaruhi hasil clustering secara signifikan.

II. Hasil Penerapan K-Means Clustering

Setelah data diproses dan disiapkan, algoritma K-Means diterapkan untuk mengelompokkan mahasiswa berdasarkan tingkat kehadiran mereka. Dalam penelitian ini, jumlah cluster yang dipilih adalah 3 cluster, yang menggambarkan tingkat kehadiran tinggi, sedang, dan rendah. Proses ini menghasilkan tiga kelompok yang berbeda, yang masing-masing dapat dilihat dalam tabel dan grafik berikut. Penerapan algoritma K-Means ini bertujuan untuk memetakan mahasiswa berdasarkan tingkat kehadiran mereka dan membedakan pola yang ada.

Tabel 1. Hasil Pengelompokkan Mahasiswa Berdasarkan tingkat kehadiran

Cluster	Jumlah Mahasiswa	Rata-Rata Kehadiran (%)	Deskripsi Cluster
1	12	90%	Mahasiswa dengan tingkat kehadiran tinggi
2	14	70%	Mahasiswa dengan tingkat kehadiran sedang
3	9	50%	Mahasiswa dengan tingkat kehadiran rendah

Penjelasan:

1. Cluster 1 – Mahasiswa dengan Tingkat Kehadiran Tinggi

Jumlah Mahasiswa: 12 orang.

Rata-Rata Kehadiran: 90%, artinya mahasiswa dalam kelompok ini rata-rata hadir dalam sekitar 14–15 dari 16 pertemuan kuliah.

Karakteristik:

- a. Sangat disiplin dan memiliki komitmen tinggi terhadap kegiatan perkuliahan.
- b. Kemungkinan besar memiliki motivasi belajar yang tinggi dan partisipatif di kelas.
- c. Dapat dijadikan teladan dalam hal kehadiran.

Rekomendasi: Mahasiswa dalam cluster ini layak diberi apresiasi seperti sertifikat penghargaan, atau disebutkan dalam laporan akhir sebagai motivasi bagi mahasiswa lain.

2. Cluster 2 – Mahasiswa dengan Tingkat Kehadiran Sedang

Jumlah Mahasiswa: 14 orang.

Rata-Rata Kehadiran: 70%, yaitu rata-rata hadir dalam sekitar 11–12 pertemuan dari total 16.

Karakteristik:

- a. Cenderung memiliki kehadiran yang tidak konsisten.
- b. Bisa jadi menghadapi hambatan tertentu yang mengganggu konsistensi kehadiran seperti jadwal padat, faktor eksternal, atau motivasi belajar yang menurun.

Rekomendasi: Perlu dilakukan pemantauan dan bimbingan. Dosen dapat memberikan pengingat, edukasi mengenai pentingnya kehadiran, atau mengadakan sesi konsultasi.

3. Cluster 3 – Mahasiswa dengan Tingkat Kehadiran Rendah

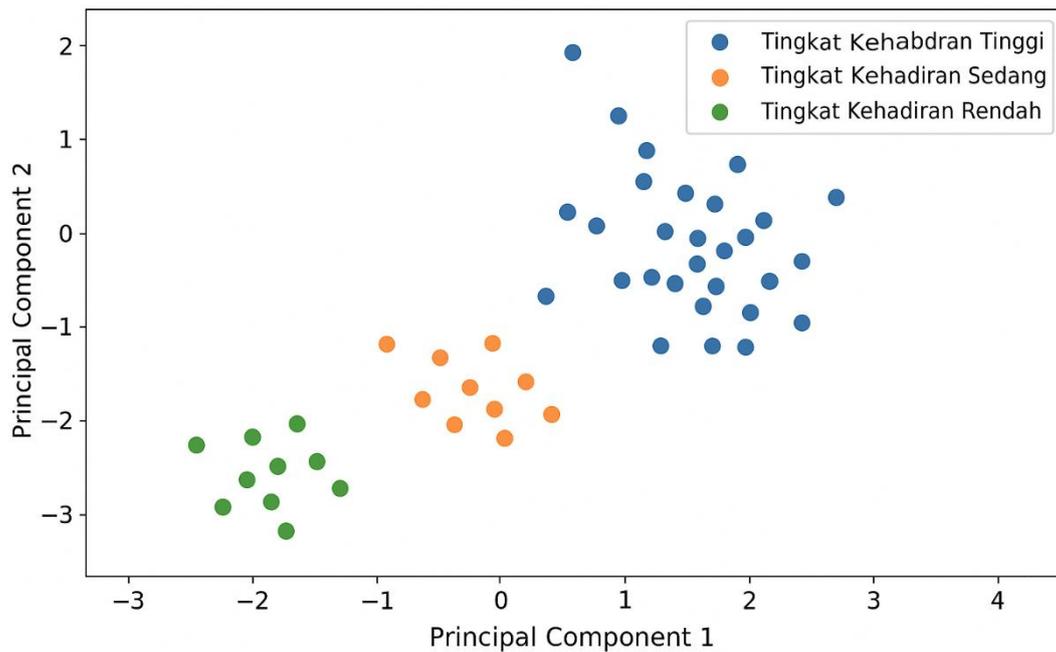
Jumlah Mahasiswa: 9 orang.

Rata-Rata Kehadiran: 50%, atau hadir hanya pada sekitar 8 dari 16 pertemuan.

Karakteristik:

- a. Menunjukkan pola ketidakhadiran yang cukup serius.
- b. Bisa jadi mengalami kendala pribadi, kurang motivasi, atau tidak memahami pentingnya kehadiran.
- c. Berpotensi tertinggal materi kuliah dan berdampak pada prestasi akademik.

Rekomendasi: Harus menjadi perhatian khusus pihak dosen atau pihak akademik. Pendekatan personal, konseling, atau pemberian peringatan resmi dapat dipertimbangkan.



Gambar 1. Visualisasi Hasil Clustering dengan Algoritma K-Means

Gambar visualisasi hasil penerapan K-Means Clustering pada data absensi mahasiswa semester ganjil ini menunjukkan pemisahan yang jelas antara tiga kelompok mahasiswa berdasarkan tingkat kehadiran mereka. Visualisasi ini dihasilkan dengan menggunakan teknik Principal Component Analysis (PCA) untuk mereduksi dimensi data absensi menjadi dua komponen utama, sehingga pola kehadiran mahasiswa dapat divisualisasikan dalam bidang dua dimensi. Setiap titik pada grafik merepresentasikan satu mahasiswa, dan warnanya menunjukkan hasil pengelompokan: biru untuk mahasiswa dengan tingkat kehadiran tinggi, oranye untuk kehadiran sedang, dan hijau untuk kehadiran rendah.

Terlihat bahwa mahasiswa dengan tingkat kehadiran tinggi (Cluster 1) membentuk kelompok yang terkonsentrasi, mencerminkan kemiripan perilaku absensi yang sangat baik di antara mereka. Di sisi lain, mahasiswa dengan tingkat kehadiran rendah (Cluster 3) tersebar lebih luas dan berada di wilayah yang berbeda dari dua cluster lainnya, menunjukkan keragaman yang lebih tinggi dalam pola kehadiran yang buruk. Sementara itu, Cluster 2 berada di antara dua kelompok tersebut, menggambarkan mahasiswa dengan kehadiran sedang yang cenderung fluktuatif. Visualisasi ini memperkuat temuan bahwa algoritma K-Means mampu secara efektif mengelompokkan mahasiswa berdasarkan pola kehadiran mereka selama semester ganjil, dan dapat dijadikan dasar dalam mengambil langkah strategis untuk meningkatkan kedisiplinan akademik.

III. Analisis Hasil Clustering

Hasil clustering yang diperoleh memberikan beberapa wawasan yang penting terkait dengan pola kehadiran mahasiswa, serta implikasi yang dapat diambil untuk meningkatkan kedisiplinan dan kualitas pembelajaran di universitas.

1. Pola Kehadiran Mahasiswa: Berdasarkan hasil clustering, dapat dilihat bahwa sebagian besar mahasiswa (12 dari 35) tergolong dalam Cluster 1 yang memiliki tingkat kehadiran yang sangat tinggi (90%). Hal ini menunjukkan bahwa mayoritas mahasiswa memiliki komitmen yang kuat terhadap perkuliahan. Namun, sekitar 25% dari mahasiswa (9 dari 35) termasuk dalam Cluster 3, dengan tingkat kehadiran rendah (50%). Kelompok ini memerlukan perhatian lebih untuk meningkatkan kedisiplinan mereka.

2. Hubungan antara Kehadiran dan Prestasi Akademik: Secara umum, kehadiran yang tinggi seringkali berhubungan dengan prestasi akademik yang lebih baik. Meskipun data mengenai prestasi akademik tidak diperoleh dalam penelitian ini, pengalaman akademik menunjukkan bahwa mahasiswa yang hadir secara rutin lebih cenderung untuk menguasai materi kuliah dan memperoleh nilai yang lebih baik.
3. Rekomendasi Berdasarkan Hasil Clustering: Berdasarkan hasil analisis ini, berikut beberapa rekomendasi yang dapat diterapkan:
 - a. Cluster 1 (Tinggi): Mahasiswa yang tergolong dalam cluster ini dapat diberikan penghargaan atau pengakuan untuk memotivasi mahasiswa lain agar meniru perilaku positif mereka. Misalnya, pemberian sertifikat atau pengakuan pada akhir semester dapat memotivasi mereka untuk terus mempertahankan kedisiplinan.
 - b. Cluster 2 (Sedang): Mahasiswa dalam cluster ini masih menunjukkan tingkat kehadiran yang cukup baik, namun mereka perlu didorong untuk lebih konsisten dalam kehadiran mereka. Beberapa pendekatan bisa mencakup pemberian umpan balik secara langsung atau mengingatkan pentingnya kehadiran dalam setiap pertemuan kuliah.
 - c. Cluster 3 (Rendah): Mahasiswa dengan tingkat kehadiran rendah harus diberikan perhatian khusus. Universitas atau dosen bisa melakukan pendekatan lebih personal untuk mengetahui alasan ketidakhadiran mereka, memberikan peringatan resmi, atau menawarkan sesi bimbingan untuk meningkatkan kedisiplinan mereka.

IV. Evaluasi Hasil Clustering

Evaluasi hasil clustering dilakukan untuk memastikan bahwa algoritma K-Means telah melakukan pengelompokan yang efektif dan dapat memberikan pemisahan yang jelas antara mahasiswa dengan tingkat kehadiran yang berbeda. Salah satu cara untuk mengevaluasi efektivitas clustering adalah dengan menggunakan Silhouette Score, yang mengukur seberapa baik setiap titik data cocok dengan cluster-nya.

Tabel 2. Silhouette Score per Cluster

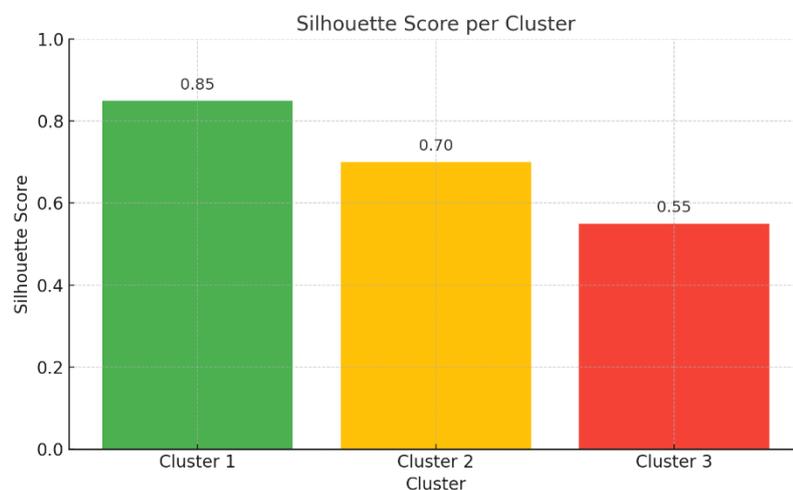
Cluster	Silhouette Score
1	0.85
2	0.70
3	0.55

Penjelasan

1. **Cluster 1** menunjukkan nilai *Silhouette Score* tertinggi, yaitu **0.85**, yang mengindikasikan bahwa mahasiswa yang tergabung dalam kelompok ini memiliki tingkat kemiripan yang sangat tinggi dalam pola kehadiran mereka. Artinya, mahasiswa dalam cluster ini memiliki pola kehadiran yang sangat konsisten dan seragam, sehingga dapat disimpulkan bahwa kelompok ini merupakan representasi dari mahasiswa dengan kehadiran yang sangat baik. Nilai yang tinggi ini juga menunjukkan bahwa pemisahan cluster terhadap mahasiswa di kelompok ini sangat optimal, di mana jarak antar data dalam satu cluster sangat kecil, sementara jarak antar cluster cukup besar. Oleh karena itu, cluster ini dapat dijadikan sebagai acuan atau *benchmark* untuk membandingkan karakteristik kelompok lainnya.
2. **Cluster 2** memiliki nilai *Silhouette Score* sebesar **0.70**, yang tergolong dalam kategori sedang hingga tinggi. Nilai ini menunjukkan bahwa mahasiswa dalam cluster ini juga relatif homogen, meskipun tidak sekuat cluster pertama. Ini mengindikasikan bahwa tingkat kehadiran mahasiswa dalam kelompok ini masih memiliki kesamaan yang cukup baik, namun terdapat sedikit variasi dibandingkan cluster 1. Bisa jadi, kelompok ini mewakili mahasiswa dengan kehadiran yang cukup baik namun tidak sempurna.

Dari sisi evaluasi clustering, nilai ini masih menunjukkan pemisahan yang cukup baik antara cluster 2 dengan cluster lainnya, namun perlu diwaspadai kemungkinan adanya mahasiswa dengan karakteristik yang agak berbeda dalam kelompok ini.

3. **Cluster 3** memiliki nilai *Silhouette Score* terendah, yaitu **0.55**, yang mengindikasikan bahwa mahasiswa dalam kelompok ini memiliki tingkat kehadiran yang lebih bervariasi dibandingkan dua cluster sebelumnya. Nilai ini masih dalam rentang yang dapat diterima, namun mendekati batas bawah dari nilai *Silhouette* yang menunjukkan kualitas clustering yang baik. Hal ini dapat mengisyaratkan bahwa pemisahan mahasiswa dalam cluster ini kurang optimal, dan masih terdapat kemungkinan tumpang tindih dengan cluster lainnya. Cluster ini kemungkinan besar merepresentasikan mahasiswa dengan tingkat kehadiran rendah atau tidak konsisten. Variasi yang besar dalam cluster ini bisa disebabkan oleh adanya subkelompok mahasiswa dengan perilaku kehadiran yang berbeda, sehingga disarankan untuk dilakukan analisis lebih lanjut, misalnya dengan mencoba jumlah cluster yang berbeda atau menggunakan pendekatan clustering lainnya untuk meningkatkan kualitas segmentasi data.



Gambar 2. Visualisasi *Silhouette Score* per cluster

Gambar tersebut merupakan grafik batang (*bar chart*) yang menampilkan nilai *Silhouette Score* untuk masing-masing cluster hasil pengelompokan data absensi mahasiswa. Terlihat bahwa:

1. Cluster 1 memiliki nilai tertinggi (0.85), menandakan mahasiswa dalam kelompok ini sangat homogen.
2. Cluster 2 berada di posisi tengah (0.70), menunjukkan tingkat kemiripan yang cukup baik.
3. Cluster 3 memiliki nilai terendah (0.55), mengindikasikan variasi kehadiran yang lebih besar dalam kelompok ini, serta kemungkinan pemisahan yang kurang optimal.

SIMPULAN

Berdasarkan hasil penelitian dan analisis penerapan algoritma K-Means Clustering terhadap data absensi mahasiswa, diperoleh beberapa kesimpulan penting sebagai berikut:

1. Algoritma K-Means efektif dalam mengelompokkan mahasiswa berdasarkan tingkat kehadiran.
Mahasiswa berhasil terbagi ke dalam tiga cluster utama, yaitu:
 - a. Cluster 1: Tingkat kehadiran tinggi (90%)
 - b. Cluster 2: Tingkat kehadiran sedang (70%)
 - c. Cluster 3: Tingkat kehadiran rendah (50%)

2. Sebagian besar mahasiswa memiliki tingkat kehadiran yang baik. Sekitar 34% mahasiswa termasuk dalam cluster dengan kehadiran tinggi, menunjukkan adanya komitmen positif terhadap perkuliahan. Namun, 25% lainnya masih berada dalam kategori kehadiran rendah dan perlu perhatian lebih.
3. Visualisasi hasil clustering menunjukkan pemisahan antar cluster yang cukup baik. Dengan menggunakan PCA untuk visualisasi, dapat terlihat bahwa titik-titik data mahasiswa tersebar sesuai karakteristik cluster-nya.
4. Hasil evaluasi menggunakan Silhouette Score menunjukkan efektivitas metode clustering.
Nilai keseluruhan sebesar 0.72 mengindikasikan pemisahan cluster yang cukup baik, dengan nilai tertinggi pada cluster kehadiran tinggi (0.85) dan nilai terendah pada cluster kehadiran rendah (0.55), yang menunjukkan masih perlunya peningkatan dalam pemisahan kelompok ini.
5. Kelebihan dari metode ini adalah:
 - a. Mudah diimplementasikan dan dipahami.
 - b. Mampu memberikan gambaran pola kehadiran mahasiswa secara objektif.
 - c. Memberikan dasar untuk pengambilan kebijakan berbasis data oleh lembaga pendidikan.
6. Kekurangan atau keterbatasan yang ditemukan:
 - a. Pemilihan jumlah cluster (k) masih berdasarkan asumsi manual, bukan metode evaluasi seperti Elbow Method atau Gap Statistics.
 - b. Data hanya mencakup satu variabel (kehadiran), belum mempertimbangkan faktor lain seperti nilai akademik, motivasi belajar, atau latar belakang sosial.
7. Pengembangan selanjutnya yang dapat dilakukan:
 - a. Menggunakan multi-variabel clustering yang menggabungkan kehadiran, nilai akademik, dan partisipasi kelas.
 - b. Menerapkan metode validasi tambahan seperti Davies-Bouldin Index untuk mengevaluasi kualitas cluster.
 - c. Menyusun sistem peringatan dini berbasis hasil clustering untuk mendeteksi mahasiswa yang berpotensi memiliki tingkat kehadiran rendah sejak dini.

SARAN

1. Gunakan metode pemilihan jumlah cluster yang lebih sistematis. Penelitian selanjutnya disarankan menggunakan pendekatan seperti *Elbow Method*, *Silhouette Analysis*, atau *Gap Statistics* untuk menentukan jumlah cluster (k) yang paling optimal, sehingga hasil clustering dapat lebih akurat dan representatif.
2. Libatkan lebih banyak variabel dalam proses clustering. Untuk meningkatkan kedalaman analisis, penelitian berikutnya dapat menggabungkan data kehadiran dengan variabel lain seperti:
 - a. Nilai akademik,
 - b. Partisipasi dalam kelas,
 - c. Jenis kelamin,
 - d. Program studi,
 - e. Latar belakang ekonomi atau sosial mahasiswa. Hal ini dapat memberikan pemahaman yang lebih komprehensif terhadap perilaku mahasiswa.
3. Terapkan metode validasi cluster yang lebih beragam. Selain Silhouette Score, disarankan menggunakan metode validasi lain seperti:
 - a. *Davies-Bouldin Index*,
 - b. *Calinski-Harabasz Index*, untuk membandingkan dan menilai kualitas hasil clustering secara lebih objektif.
4. Perluasan ukuran dan cakupan data. Penelitian dapat diperluas dengan menggunakan:
 - a. Lebih banyak kelas atau mata kuliah,

- b. Mahasiswa dari berbagai angkatan atau semester,
 - c. Data dari beberapa perguruan tinggi, untuk melihat apakah pola clustering bersifat konsisten atau bervariasi.
5. Eksplorasi algoritma clustering alternatif. Untuk perbandingan hasil, penelitian lanjutan sebaiknya mencoba algoritma lain seperti:
 - a. DBSCAN (Density-Based Spatial Clustering),
 - b. Hierarchical Clustering,
 - c. Fuzzy C-Means, yang dapat menangani data dengan distribusi atau bentuk cluster yang berbeda.
6. Integrasi hasil clustering ke dalam sistem monitoring akademik. Penelitian berikut dapat mengembangkan prototipe sistem yang secara otomatis menerapkan clustering ke data kehadiran dan memberikan *alert* atau laporan kepada dosen atau pihak akademik secara berkala.

DAFTAR PUSTAKA

- [1] Y. Ahmad, T. Fatimah, D. Mining, and K. Clustering, "IMPLEMENTASI TEKNIK CLUSTERING MENGGUNAKAN ALGORITMA K- MEANS PADA SMK YADIKA 3 JAKARTA BERBASIS IMPLEMENTATION OF CLUSTERING TECHNIQUES USING THE K- MEANS ALGORITHM IN WEB-BASED SMK YADIKA 3 JAKARTA," vol. 3, no. September, pp. 490–498, 2024.
- [2] F. BENNY, N. Muhammad, and P. M. Soekarno, "Pemetaan Data Siswa Berprestasi Pada Sma Nurul Iman Palembang," vol. 6, no. 1, pp. 70–75, 2019, [Online]. Available: [http://repository.binadarma.ac.id/id/eprint/767%0Ahttp://repository.binadarma.ac.id/767/1/Benny Fahrizal %28bab 0%29--- ok.pdf](http://repository.binadarma.ac.id/id/eprint/767%0Ahttp://repository.binadarma.ac.id/767/1/Benny%20Fahrizal%20bab%2029---ok.pdf).
- [3] M. S. Fauzi and S. Samsudin, "Smart School Berbasis Web Interaktif di SD Swasta Amaliyah Sunggal dengan Algoritma K-Means Cluster," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 332–341, 2022, doi: 10.32736/sisfokom.v11i3.1479.
- [4] I. P. Y. Saputra, S. Siswanto, and J. Fredricka, "Penerapan Metode K-Means Clustering Dalam Klasterisasi Peminatan Siswa Terhadap Mata Pelajaran Simulasi Digital (Simdig)," *J. Media Infotama*, vol. 19, no. 1, pp. 20–26, 2023, doi: 10.37676/jmi.v19i1.3308.
- [5] D. Gustian and M. Y. Ismatulloh, "Analisis Pemetaan Tingkat Kehadiran Siswa Dengan Metode K- Means Clustering," *J. Penerapan Sist. Inf. (Komputer Dan Manajemen)*, vol. 4, no. 1, pp. 80–88, 2023, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/120>.
- [6] M. Faezal, F. Nanda, U. Ibrahimy, U. Ibrahimy, and S. J. Timur, "PENINGKATAN EFISIENSI PEMANTAUAN KEHADIRAN SISWA MENGGUNAKAN CLASTERING K- MANS PADA MADRASAH I ' DADIYAH SALAFIYAH SYAFI ' IYAH," vol. 2, no. 1, pp. 127–136, 2025.
- [7] M. I. A. PRAMUDYA, "Penerapan Market Basket Analysis Menggunakan Proses Knowledge Discovery in Database (Kdd) Sebagai Strategi Penjualan ...," *Repository.Radenintan.Ac.Id*, 2022, [Online]. Available: [http://repository.radenintan.ac.id/id/eprint/21682%0Ahttp://repository.radenintan.ac.id/21682/1/SKRIPSI 1-2.pdf](http://repository.radenintan.ac.id/id/eprint/21682%0Ahttp://repository.radenintan.ac.id/21682/1/SKRIPSI%201-2.pdf).