



Pengelompokan Gaya Belajar Mahasiswa Menggunakan Metode K-Means dan Validasi Menggunakan Davies Bouldin Index

Latifah Hanum

Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

E-Mail : latifahnm004@email.com

Article Info

Article history:

Received April 29, 2025

Revised Mei 10, 2025

Accepted Jun 15, 2025

Kata Kunci:

Gaya Belajar

Klasifikasi

K-Means

VARK

Davies-Bouldin Index

Keywords:

Learning Style

Classification

K-Means

VARK

Davies-Bouldin

ABSTRAK

Penelitian ini dilakukan untuk mengelompokkan gaya belajar mahasiswa berdasarkan model VARK (Visual, Auditory, Reading/Writing, Kinesthetic) dengan memanfaatkan algoritma K-Means. Selain itu, penelitian ini juga bertujuan mengevaluasi kualitas hasil pengelompokan tersebut menggunakan indeks Davies-Bouldin. Data diperoleh melalui penyebaran kuesioner kepada sepuluh mahasiswa dari berbagai program studi. Setiap mahasiswa memberikan penilaian terhadap empat jenis gaya belajar yang kemudian diolah menjadi data numerik. Dari hasil analisis, terbentuk empat klaster, namun hanya tiga klaster yang terisi. Klaster pertama terdiri dari tujuh mahasiswa dengan karakteristik gaya belajar yang serupa. Dua klaster lainnya masing-masing berisi satu dan dua mahasiswa yang menunjukkan kecenderungan belajar yang berbeda dari kelompok mayoritas. Nilai Davies-Bouldin sebesar 0,727 menunjukkan bahwa hasil pengelompokan tergolong baik, dengan tingkat kekompakan yang tinggi dalam klaster dan pemisahan yang jelas antar klaster. Hasil penelitian ini menunjukkan bahwa metode K-Means dapat digunakan secara efektif untuk mengenali pola gaya belajar mahasiswa secara objektif. Temuan ini diharapkan dapat menjadi acuan dalam pengembangan strategi pembelajaran yang menyesuaikan dengan karakteristik peserta didik.

ABSTRACT

This study was conducted to group students' learning styles based on the VARK model (Visual, Auditory, Reading/Writing, Kinesthetic) by utilizing the K-Means algorithm. In addition, this study also aims to evaluate the quality of the grouping results using the Davies-Bouldin index. Data were obtained by distributing questionnaires to ten students from various study programs. Each student gave an assessment of four types of learning styles which were then processed into numerical data. From the results of the analysis, four clusters were formed, but only three clusters were filled. The first cluster consisted of seven students with similar learning style characteristics. The other two clusters each contained one and two students who showed different learning tendencies from the majority group. The Davies-Bouldin value of 0.727 indicates that the grouping results are quite good, with a high level of cohesiveness in the cluster and clear separation between clusters. The results of this study indicate that the K-Means method can be used effectively to objectively identify student learning style patterns. These findings are expected to be a reference in developing learning strategies that are tailored to the characteristics of students.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Latifah Hanum,

Sistem Informasi, STIKOM Tunas Bangsa,

Jl. Jend Sudirman Blok A No. 1,2&3, Kec. Siantar Barat, Pematangsiantar, Sumatera Utara, 21127, Indonesia.

Email: latifahnm004@email.com

1. PENDAHULUAN

Pendidikan tinggi mengharuskan mahasiswa terlibat aktif dalam proses belajar, memahami materi, dan dapat mengasah keterampilan berpikir kritis secara mandiri (Yuningsih & Devi, 2024). Namun, tidak semua mahasiswa dapat menerima materi pembelajaran dengan cara yang sama. Perbedaan cara menyerap informasi ini dikenal sebagai gaya belajar. Gaya belajar merujuk pada preferensi individu dalam menerima, mengolah, dan memahami informasi, yang mencerminkan strategi kognitif paling efektif bagi seseorang dalam proses pembelajaran (Essa et al., 2023). Sebagian mahasiswa cenderung lebih mudah memahami materi melalui representasi visual seperti gambar atau diagram, sedangkan yang lain lebih mengandalkan informasi yang disampaikan secara auditori atau melalui pengalaman langsung (Nanduarti et al., 2024). Pemahaman terhadap gaya belajar mahasiswa merupakan aspek yang krusial dan tidak dapat diabaikan dalam upaya meningkatkan efektivitas proses pembelajaran. Keselarasan antara metode pembelajaran yang diterapkan oleh dosen dengan gaya belajar mahasiswa berkontribusi signifikan terhadap peningkatan efektivitas proses transfer pengetahuan. Sebaliknya, ketidaksesuaian antara metode pembelajaran dan gaya belajar mahasiswa dapat menghambat pemahaman materi, yang pada akhirnya berdampak negatif terhadap pencapaian akademik (Mindarsih & Syafitri, 2017). Dalam konteks tersebut, identifikasi gaya belajar mahasiswa menjadi elemen strategis bagi pendidik dan institusi pendidikan dalam merancang pendekatan pembelajaran yang selaras dengan karakteristik peserta didik. Selama ini, identifikasi gaya belajar mahasiswa umumnya dilakukan melalui instrumen kuesioner, salah satunya dengan menggunakan model VARK (*Visual, Auditory, Reading/Writing, dan Kinesthetic*) yang mengklasifikasikan preferensi belajar berdasarkan cara individu menerima dan mengolah informasi (Balti et al., 2023). Meskipun metode ini cukup populer, pendekatan tersebut memiliki sifat yang manual dan cenderung subjektif. Kondisi ini menjadi tantangan tersendiri, khususnya ketika pengolahan data dilakukan pada skala yang besar. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu mengelompokkan mahasiswa secara objektif dan efisien berdasarkan kemiripan karakteristik gaya belajar mereka (Kanchon et al., 2024).

Data Mining ialah metode yang sering digunakan dalam analisis data berskala besar, termasuk di sektor pendidikan. Dengan cara ini, data pendidikan dalam kuantitas besar bisa diolah dengan cara teratur untuk mengidentifikasi pola-pola tersembunyi yang relevan, sehingga dapat memberikan pemahaman yang berguna dalam mendukung proses pengambilan keputusan yang objektif dan berbasis data (Ray & Saeed, 2018). Dalam bidang *data mining*, *clustering* ialah salah satu teknik yang efisien untuk mengelompokkan data berdasarkan kesamaan atribut, tanpa membutuhkan informasi label sebelumnya. *Clustering* merupakan metode pengelompokan data yang bertujuan untuk membagi data ke dalam beberapa grup yang seragam berdasarkan kesamaan karakteristik, sehingga setiap grup mencerminkan tingkat kesamaan internal yang tinggi di antara anggotanya (Ran et al., 2023). Dalam bidang pendidikan, metode *clustering* bisa digunakan untuk mengategorikan mahasiswa ke dalam berbagai tipe gaya belajar, berdasarkan pola data yang diperoleh dari kegiatan atau tanggapan mereka selama proses belajar. Algoritma *clustering* yang banyak digunakan adalah *K-Means*. Algoritma *K-Means* membagi data ke dalam sejumlah kluster yang telah ditentukan sebelumnya dengan menghitung jarak antara setiap titik data dan pusat kluster (*centroid*), kemudian mengalokasikan data tersebut ke kluster yang memiliki jarak terdekat (Liu et al., 2024). Metode ini memiliki kelebihan utama yaitu struktur algoritma yang sederhana serta kemampuan untuk menangani data dalam skala besar dengan waktu komputasi yang cukup cepat (Wang et al., 2022). Dalam penelitian ini, metode *K-Means* diterapkan untuk mengelompokkan mahasiswa berdasar pada data gaya belajar yang diperoleh dari hasil penyebaran kuesioner gaya belajar berbasis model VARK kepada mahasiswa. Algoritma *K-Means* dapat menghasilkan pengelompokan data, namun kualitas hasil tersebut tidak dapat dikatakan optimal tanpa evaluasi yang menyeluruh (Hendrastuty, 2024). Agar hasil pengelompokan yang diperoleh berkualitas baik, penerapan metode validasi yang sesuai sangatlah diperlukan. *Davies-Bouldin Index (DBI)* adalah salah satu metode validasi yang sering digunakan untuk menilai seberapa baik hasil pengelompokan data yang diperoleh dari algoritma *clustering* (Kristanto et al., 2023). Semakin rendah nilai *Davies-Bouldin Index (DBI)*, semakin bagus kualitas pengelompokan, karena ini menunjukkan bahwa kluster yang terbentuk memiliki tingkat kekompakan yang tinggi dan jelas terpisah satu sama lain (Rahmawati et al., 2025).

Beberapa penelitian sebelumnya menunjukkan bahwa teknik *K-Means* sangat efektif untuk mengelompokkan data mahasiswa, terutama dalam konteks analisis ciri-ciri dan kebiasaan belajar. (Fish, 2020), Artikel ini membahas penggunaan algoritma *K-Means* untuk mendeteksi gaya belajar siswa dalam lingkungan pembelajaran daring. (Muttaqin & Defriani, 2020) Penelitian ini menggunakan metode *K-Means* untuk mengelompokkan nilai mahasiswa setiap semester. Hasil pengelompokan ini memudahkan dosen dalam mengevaluasi dan menyesuaikan materi ajar sesuai dengan kelompok hasil belajar mahasiswa. Namun, sampai saat ini jumlah penelitian yang secara khusus menggabungkan metode *K-Means* dengan evaluasi menggunakan *Davies-Bouldin Index (DBI)* dalam konteks gaya belajar mahasiswa masih terbatas, padahal validasi seperti ini sangat penting untuk memastikan akurasi dan keandalan hasil pengelompokan.

Penelitian ini bertujuan untuk mengklasifikasikan gaya belajar mahasiswa menggunakan algoritma *K-Means* dan mengevaluasi kualitas hasil pengelompokan menggunakan indeks *Davies-Bouldin* sebagai ukuran validitas pengelompokan. Data dari penelitian ini berasal dari hasil penyebaran kuesioner model VARK yang diisi oleh mahasiswa dari berbagai program studi. Pendekatan ini diharapkan dapat menawarkan alternatif untuk memetakan gaya belajar siswa dengan objektivitas yang lebih besar, serta membantu lembaga pendidikan dalam merumuskan kebijakan pengajaran adaptif yang responsif terhadap kebutuhan para peserta didik. Hasil penelitian ini memberikan sumbangan secara langsung terhadap peningkatan efektivitas dan usaha pembelajaran yang dilakukan, maupun dari pembukuan kajian yang lebih luas dalam pengembangan ilmu pendidikan berbasis teknologi. Melalui penggunaan metode *data mining* dan validasi *clustering* yang sesuai, proses untuk mengidentifikasi gaya belajar dapat dilaksanakan dengan lebih efisien, akurat, dan dapat diperluas pada populasi mahasiswa dalam jumlah yang besar.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen semu yang bersifat eksploratif, bertujuan mengelompokkan mahasiswa berdasarkan gaya belajar menggunakan algoritma *K-Means Clustering* dan memvalidasi hasilnya dengan *Davies-Bouldin Index* (DBI). Penelitian dilakukan di lingkungan perguruan tinggi dengan asumsi bahwa mahasiswa memiliki kecenderungan belajar yang berbeda, sehingga pendekatan berbasis data dianggap mampu memetakan gaya belajar secara objektif dan adaptif. Populasi terdiri dari mahasiswa berbagai program studi, dengan sampel sebanyak 80 orang yang dipilih secara *purposive*, yaitu mereka yang mengisi kuesioner VARK secara lengkap dan valid. Data dikumpulkan melalui kuesioner daring dengan pertanyaan tertutup berdasarkan model VARK (Visual, Auditory, Reading/Writing, Kinesthetic). Setelah data terkumpul, dilakukan pra-pemrosesan berupa pemeriksaan duplikasi, identifikasi data kosong, konversi ke format numerik, dan normalisasi untuk memastikan kontribusi yang seimbang dari setiap atribut dalam proses pengelompokan. Data diklasifikasikan menggunakan algoritma *K-Means* untuk menentukan kluster-kluster yang mengelompokkan gaya belajar mahasiswa dengan menggunakan model VARK dapat dilihat pada tabel 1.

Tabel 1. Dataset Gaya Belajar Mahasiswa

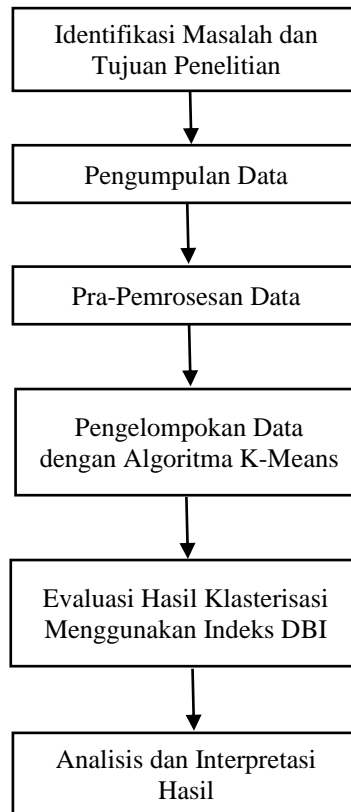
No	Nama	Program Studi	Visual	Auditory	Reading/Writing	Kinesthetic
1	Fani	Manajemen Informatika	4	4	4	5
2	Azhari	Sistem Informasi	4	4	3	4
3	Rahmat	Sistem Informasi	5	3	4	5
4	Fadiah	Sistem Informasi	5	5	5	5
5	Saudi	Sistem Informasi	5	3	4	5
6	Arsyinta	Akuntansi	4	5	4	4
7	Vira	Akuntansi	4	4	3	5
8	Dilla	Manajemen	5	5	5	4
9	Femi	Sistem Informasi	4	4	4	4
10	Adinda	Sistem Informasi	3	4	4	4
....
80	Asri	Psikolog	4	4	3	4

Tabel 1 menyajikan sebagian data mentah yang diperoleh dari hasil pengisian kuesioner VARK oleh mahasiswa responden dalam penelitian ini. Data ini merepresentasikan skor preferensi masing-masing individu terhadap empat kategori utama gaya belajar, yaitu visual, auditori, membaca/menulis (*reading/writing*), dan kinestetik. Setiap baris di tabel merepresentasikan satu responden, yang dikenali melalui nomor urut dan nama, serta disertai informasi asal program studi. Di sisi lain, empat kolom terakhir mencerminkan skor atau tingkat preferensi mahasiswa terhadap setiap gaya belajar sesuai dengan jawaban mereka dalam kuesioner. Nilai tersebut diberikan dalam skala ordinal, dengan rentang dari 1 (sangat rendah) sampai 5 (sangat tinggi), yang mencerminkan seberapa menonjol suatu gaya belajar pada setiap individu. Misalnya, responden pertama, Fani Safira dari program studi Manajemen Informatika, mendapatkan skor tinggi pada aspek kinestetik 5, sedangkan skor untuk tiga gaya belajar lainnya memiliki nilai yang sama 4. Ini menunjukkan bahwa individu tersebut cenderung paling dominan dalam mencerna informasi melalui pengalaman langsung atau kegiatan fisik. Responden keempat, Fadiah Rizky, mencetak skor tertinggi 5 pada keempat kategori, yang menunjukkan bahwa ia adalah mahasiswa dengan kemampuan adaptasi yang sangat baik terhadap beragam metode pembelajaran. Data yang ditunjukkan pada tabel ini berfungsi sebagai input awal dalam tahap pra-pemrosesan data, sebelum dilakukan pengelompokan menggunakan algoritma *K-Means*. Skor dari setiap dimensi gaya

belajar digunakan sebagai atribut numerik untuk mengukur kedekatan atau kesamaan antar responden, sehingga memungkinkan pembentukan kluster mahasiswa dengan preferensi belajar yang cukup seragam.

2.1 Rancangan Alur Penelitian

Rancangan alur penelitian tersebut disajikan pada Gambar 1, guna memastikan setiap tahapan berjalan secara sistematis dan terstruktur.



Gambar 1. Rancangan Penelitian

Tahapan awal dalam penelitian ini adalah mengidentifikasi permasalahan yang relevan dengan konteks gaya belajar mahasiswa. Permasalahan dirumuskan berdasarkan kenyataan bahwa setiap mahasiswa memiliki kecenderungan belajar yang berbeda, yang apabila tidak diidentifikasi dengan tepat dapat menghambat efektivitas pembelajaran. Setelah permasalahan dirumuskan, peneliti menyusun tujuan yang jelas, yakni untuk mengelompokkan mahasiswa berdasarkan preferensi gaya belajar menggunakan pendekatan ilmiah yang berbasis data. Tujuan ini menjadi dasar pijakan dalam seluruh rangkaian proses penelitian. Data dikumpulkan melalui kuesioner model VARK yang disebarluaskan secara daring. Responden mengisi pertanyaan untuk mengidentifikasi preferensi belajar mereka (visual, auditori, membaca/menulis, dan kinestetik). Data yang diperoleh divalidasi untuk menghilangkan duplikasi dan kesalahan pengisian, kemudian dinormalisasi agar semua atribut memiliki skala yang seimbang sebelum dianalisis. Tahapan inti dari penelitian ini adalah proses pengelompokan data menggunakan algoritma *K-Means* dengan menggunakan rumus (1) dan (2). Algoritma ini bekerja dengan cara membagi data ke dalam sejumlah kluster berdasarkan tingkat kemiripan karakteristik.

Peneliti terlebih dahulu menentukan jumlah kluster (k), kemudian algoritma akan secara iteratif mengelompokkan data dengan menghitung jarak antar titik data dan titik pusat kluster (*centroid*). Proses ini akan terus berlangsung hingga komposisi kluster stabil dan tidak lagi mengalami perubahan. Hasil dari tahap ini adalah pembagian mahasiswa ke dalam kelompok-kelompok gaya belajar yang seragam. Setelah proses pengelompokan selesai, penting untuk mengevaluasi seberapa baik kluster yang terbentuk. Dalam penelitian ini, digunakan metode validasi *Davies-Bouldin Index* (DBI) untuk mengukur kualitas klusterisasi dengan menggunakan rumus (3). DBI memberikan gambaran mengenai tingkat kekompakan tiap kluster serta seberapa jauh pemisahan antar kluster. Semakin kecil nilai DBI, semakin baik hasil pengelompokan karena menunjukkan bahwa anggota dalam satu kluster memiliki kemiripan tinggi, sementara antar kluster saling berbeda secara signifikan. Evaluasi ini menjadi penentu utama dalam menilai keberhasilan penggunaan algoritma *K-Means* dalam konteks data yang digunakan. Hasil pengelompokan dianalisis untuk memahami

karakteristik tiap klaster. Temuan ini digunakan untuk merekomendasikan strategi pembelajaran yang sesuai dengan gaya belajar mahasiswa.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

$$C_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (2)$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Data yang digunakan adalah hasil pengumpulan melalui *google* formulir yang menghasilkan 10 data untuk dimodelkan menggunakan *K-means Clustering*. Tabel 2 di bawah ini menunjukkan hasil kuesioner mengenai gaya belajar mahasiswa :

Tabel 2. Hasil Kuisisioner Gaya Belajar Mahasiswa

No	Nama	Program Studi	Visual	Auditory	Reading/ Writing	Kinesthetic
1	Fani	Manajemen Informatika	4	4	4	5
2	Azhari	Sistem Informasi	4	4	3	4
3	Rahmat	Sistem Informasi	5	3	4	5
4	Fadiah	Sistem Informasi	5	5	5	5
5	Saudi	Sistem Informasi	5	3	4	5
6	Arsyinta	Akuntansi	4	5	4	4
7	Vira	Akuntansi	4	4	3	5
8	Dilla	Manajemen	5	5	5	4
9	Femi	Sistem Informasi	4	4	4	4
10	Adinda	Sistem Informasi	3	4	4	4

Di bawah ini terdapat contoh kasus dan perhitungan dengan menggunakan algoritma *K-Means Clustering* :

1. Menetapkan jumlah cluster data dalam penelitian ini adalah sebanyak 4 *cluster*. Dalam tabel 2, nilai *visual*, *auditory*, *reading/writing*, dan *kinesthetic* yang diperoleh mahasiswa Fani adalah 4, 4, 3, dan 5, berdasarkan hasil pengisian kuesioner gaya belajar yang dilakukan oleh mahasiswa tersebut.

Tabel 3. Titik Pusat Awal Tiap Cluster

Titik Pusat Awal	Visual	Auditory	Reading/Writing	Kinesthetic
<i>Cluster 1</i>	4	4	4	5
<i>Cluster 2</i>	4	5	4	4
<i>Cluster 3</i>	3	5	3	4
<i>Cluster 4</i>	5	4	5	5

2. Menghitung jarak setiap data ke pusat *cluster* antara objek ke *centroid* dengan perhitungan jarak *euclidean*. $d(x_1, y_1) = \sqrt{(x_{1a} - \mu_{ja})^2 + (x_{1n} - \mu_{jn})^2}$ (4)

$$d_{11} = \sqrt{(4^2 - 4^2) + (4^2 - 4^2) + (4^2 - 4^2) + (5^2 - 5^2)} = 0.00$$

$$d_{12} = \sqrt{(4^2 - 4^2) + (4^2 - 5^2) + (4^2 - 4^2) + (5^2 - 4^2)} = 1.41$$

$$d_{13} = \sqrt{(4^2 - 3^2) + (4^2 - 5^2) + (4^2 - 3^2) + (5^2 - 4^2)} = 2.00$$

$$d_{14} = \sqrt{(4^2 - 5^2) + (4^2 - 4^2) + (4^2 - 5^2) + (5^2 - 5^2)} = 1.41$$

3. Kelompokan data ke dalam *cluster* dengan jarak minimal

Tabel 4. Perhitungan Setiap Data Ke Setiap *Cluster* Iterasi 1

Nama	Visual	Auditory	Reading /Writing	Kinesthetic	Cluster Center 1				Jarak Record ke CC-1
					V1	A1	R1	K1	
Fani	4	4	4	5	4	4	4	5	0.00
Azhari	4	4	3	4	4	4	4	5	1.41
Rahmat	5	3	4	5	4	4	4	5	1.41
Fadiyah	5	5	5	5	4	4	4	5	1.73
Saudi	5	3	4	5	4	4	4	5	1.41
Arsyinta	4	5	4	4	4	4	4	5	1.41
Vira	4	4	3	5	4	4	4	5	1.00
Dilla	5	5	5	4	4	4	4	5	2.00
Femi	4	4	4	4	4	4	4	5	1.00
Adinda	3	4	4	4	4	4	4	5	1.41

Tabel 5. Lanjutan Tabel 4. Tabel Perhitungan Setiap Data Ke Setiap *Cluster* Iterasi 1

Nama	Cluster Center 2				Jarak Record Ke CC-2	Cluster Center 3				Jarak Record Ke CC-3
	V2	A2	R2	K2		V3	A3	R3	K3	
Fani	4	5	4	4	1.41	3	5	3	4	2.00
Azhari	4	5	4	4	1.41	3	5	3	4	1.73
Rahmat	4	5	4	4	2.45	3	5	3	4	3.32
Fadiyah	4	5	4	4	1.73	3	5	3	4	3.16
Saudi	4	5	4	4	2.45	3	5	3	4	3.32
Arsyinta	4	5	4	4	0.00	3	5	3	4	1.41
Vira	4	5	4	4	1.73	3	5	3	4	1.73
Dilla	4	5	4	4	1.41	3	5	3	4	3.00
Femi	4	5	4	4	1.41	3	5	3	4	2.00
Adinda	4	5	4	4	1.41	3	5	3	4	1.73

Tabel 6. Lanjutan Tabel 5. Tabel Perhitungan Setiap Data Ke Setiap *Cluster* Iterasi 1

Nama	Cluster Center 4				Jarak Record Ke CC-4	Cluster
	V4	A4	R4	K4		
Fani	5	4	5	5	1.41	1
Azhari	5	4	5	5	2.45	1
Rahmat	5	4	5	5	1.41	1
Fadiyah	5	4	5	5	1.00	4
Saudi	5	4	5	5	1.41	1
Arsyinta	5	4	5	5	2.00	2
Vira	5	4	5	5	1.73	1
Dilla	5	4	5	5	1.00	4
Femi	5	4	5	5	1.73	1
Adinda	5	4	5	5	2.45	1

Berdasarkan tabel 4 tersebut didapatkan hasil perhitungan jarak euclidean sebagai berikut:

Cluster 1 = {Fani, Azhari, Rahmat, Saudi, Vira, Femi, Adinda}

Cluster 2 = {Arsyinta}

Cluster 3 = Tidak Ada

Cluster 4 = {Fadiyah, Dilla}

4. Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut. Setelah menentukan titik pusat baru untuk setiap *cluster*, ulangi langkah ketiga sampai titik pusat setiap *cluster* stabil dan tidak ada lagi data yang bergerak antar *cluster*.

Tabel 7. Pengelompokan Data dari Iterasi Pertama

Nama	Cluster 1				Cluster 2			
	Visual	Auditory	Reading /Writing	Kinesthetic	Visual	Auditory	Reading /Writing	Kinesthetic
Fani	4	4	4	5	0	0	0	0
Azhari	4	4	3	4	0	0	0	0
Rahmat	5	3	4	5	0	0	0	0
Fadiyah	0	0	0	0	0	0	0	0
Saudi	5	3	4	5	0	0	0	0
Arsyinta	0	0	0	0	4	5	4	4

Nama	Cluster 1				Cluster 2			
	Visual	Auditory	Reading /Writing	Kinesthetic	Visual	Auditory	Reading /Writing	Kinesthetic
Vira	4	4	3	5	0	0	0	0
Dilla	0	0	0	0	0	0	0	0
Femi	4	4	4	4	0	0	0	0
Adinda	3	4	4	4	0	0	0	0
Jumlah	29	26	26	32	4	5	4	4
Jumlah data	7	7	7	7	1	1	1	1
<i>Centroid baru</i>	4.14	3.71	3.71	4.57	4.00	5.00	4.00	4.00

Tabel 8. Lanjutan Tabel 7. Pengelompokan Data dari Iterasi Pertama

Nama	Cluster 3				Cluster 4			
	Visual	Auditory	Reading /Writing	Kinesthetic	Visual	Auditory	Reading /Writing	Kinesthetic
Fani	0	0	0	0	0	0	0	0
Azhari	0	0	0	0	0	0	0	0
Rahmat	0	0	0	0	0	0	0	0
Fadiyah	0	0	0	0	5	5	5	5
Saudi	0	0	0	0	0	0	0	0
Arsyinta	0	0	0	0	0	0	0	0
Vira	0	0	0	0	0	0	0	0
Dilla	0	0	0	0	5	5	5	4
Femi	0	0	0	0	0	0	0	0
Adinda	0	0	0	0	0	0	0	0
Jumlah	0	0	0	0	10	10	10	9
Jumlah data	0	0	0	0	2	2	2	2
<i>Centroid baru</i>	0.00	0.00	0.00	0.00	5.00	5.00	5.00	4.50

Berdasarkan tabel 7 dan tabel 8 di atas, diperoleh hasil berikut: *Cluster 1* terdiri atas 7 data (Fani, Azhari, Rahmat, Saudi, Vira, Femi, Adinda), *Cluster 2* memiliki 1 data (Arsyinta), *Cluster 3* tidak memiliki anggota, dan *Cluster 4* berisi 2 data (Fadiyah dan Dilla). Setelah memperoleh hasil dari perhitungan iterasi pertama, Langkah selanjutnya adalah menghitung iterasi kedua. Pada kasus ini, nilai *centroid* yang diterapkan berbeda; perlu menghitung *centroid* baru dengan menambahkan nilai *Cluster* di kolom C1, C2, dan C3, kemudian membaginya dengan total data itu sendiri.

Tabel 9. Titik Pusat Setiap Cluster 2

Titik Pusat Awal	Visual	Auditory	Reading/Writing	Kinesthetic
<i>Cluster 1</i>	4.14	3.71	3.71	4.57
<i>Cluster 2</i>	4	5	4	4
<i>Cluster 3</i>	0	0	0	0
<i>Cluster 4</i>	5	5	5	4.50

Tabel 10. Perhitungan Setiap Data terhadap Setiap Cluster pada Iterasi Kedua

Nama	Visual	Auditory	Reading /Writing	Kinesthetic	Cluster Center 1				Jarak Record ke CC-1
					V1	A1	R1	K1	
Fani	4	4	4	5	4.14	3.71	3.71	4.57	0.61
Azhari	4	4	3	4	4.14	3.71	3.71	4.57	0.96
Rahmat	5	3	4	5	4.14	3.71	3.71	4.57	1.23
Fadiyah	5	5	5	5	4.14	3.71	3.71	4.57	2.06
Saudi	5	3	4	5	4.14	3.71	3.71	4.57	1.23
Arsyinta	4	5	4	4	4.14	3.71	3.71	4.57	1.44
Vira	4	4	3	5	4.14	3.71	3.71	4.57	0.89
Dilla	5	5	5	4	4.14	3.71	3.71	4.57	2.09
Femi	4	4	4	4	4.14	3.71	3.71	4.57	0.71
Adinda	3	4	4	4	4.14	3.71	3.71	4.57	1.33

Tabel 11. Lanjutan Tabel 10. Perhitungan Setiap Data terhadap Setiap *Cluster* pada Iterasi Kedua

Nama	<i>Cluster Center 2</i>				Jarak Record Ke CC-2	<i>Cluster Center 3</i>				Jarak Record Ke CC-3
	V2	A2	R2	K2		V3	A3	R3	K3	
Fani	4	5	4	4	1.41	0	0	0	0	0
Azhari	4	5	4	4	1.41	0	0	0	0	0
Rahmat	4	5	4	4	2.44	0	0	0	0	0
Fadiah	4	5	4	4	1.73	0	0	0	0	0
Saudi	4	5	4	4	2.44	0	0	0	0	0
Arsyinta	4	5	4	4	0.00	0	0	0	0	0
Vira	4	5	4	4	1.73	0	0	0	0	0
Dilla	4	5	4	4	1.41	0	0	0	0	0
Femi	4	5	4	4	1.00	0	0	0	0	0
Adinda	4	5	4	4	1.41	0	0	0	0	0

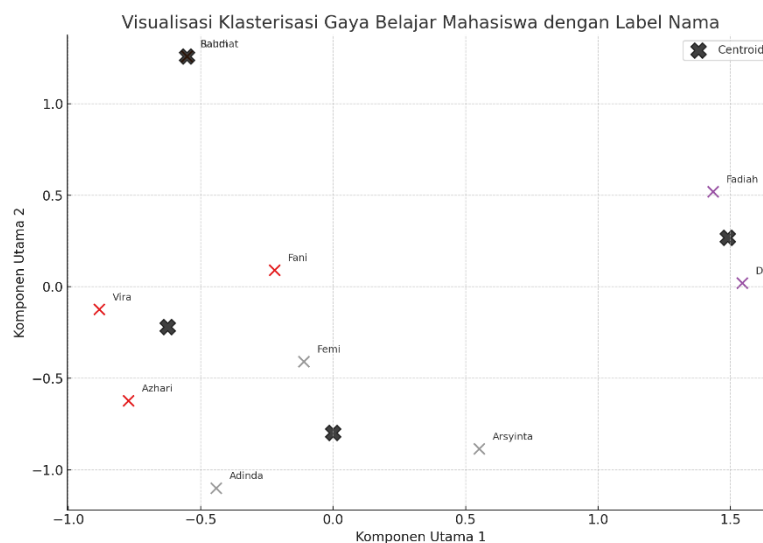
Tabel 12. Lanjutan Tabel 11. Tabel Perhitungan Setiap Data terhadap Setiap *Cluster* pada Iterasi Kedua

Nama	<i>Cluster Center 4</i>				Jarak Record Ke CC-4	<i>Cluster</i>
	V4	A4	R4	K4		
Fani	5	5	5	4.50	1.80	1
Azhari	5	5	5	4.50	2.50	1
Rahmat	5	5	5	4.50	2.29	1
Fadiah	5	5	5	4.50	0.50	4
Saudi	5	5	5	4.50	2.29	1
Arsyinta	5	5	5	4.50	1.50	2
Vira	5	5	5	4.50	2.50	1
Dilla	5	5	5	4.50	0.50	4
Femi	5	5	5	4.50	1.80	1
Adinda	5	5	5	4.50	2.50	1

Berdasarkan tabel 11 dan tabel 12 di atas, diperoleh hasil perhitungan jarak *euclidean* sebagai berikut:

1. *Cluster 1* = {Fani, Azhari, Rahmat, Saudi, Vira, Femi, Adinda}
2. *Cluster 2* = {Arsyinta}
3. *Cluster 3* = {Tidak ada}
4. *Cluster 4* = {Fadiah, Dilla}

Berdasarkan tabel 7, tabel 8, tabel 9, tabel 10, tabel 11, dan tabel 12 tampak bahwa posisi *cluster* menunjukkan nilai *cluster* yang serupa dan tidak ada pergeseran. Proses perhitungan k-means berakhir pada iterasi kedua, karena hasil iterasi kedua sama dengan hasil dari iterasi sebelumnya. Hasil yang didapat menunjukkan bahwa *cluster 1* memiliki 7 data, *cluster 2* memiliki 1 data, *cluster 3* tidak memiliki anggota, dan *cluster 4* memiliki 2 data. Visualisasi dua dimensi dari hasil *clustering* ditunjukkan pada Gambar 2 berikut ini :

Gambar 2. Visualisasi hasil *clustering* gaya belajar mahasiswa berdasarkan PCA

Gambar ini memperlihatkan hasil pengelompokan gaya belajar mahasiswa menggunakan metode *K-Means* yang divisualisasikan dalam ruang dua dimensi. Proyeksi data dilakukan dengan pendekatan *Principal*

Component Analysis (PCA) agar data yang semula terdiri dari empat variabel (visual, auditori, membaca/menulis, dan kinestetik) dapat ditampilkan secara lebih sederhana tanpa menghilangkan makna pentingnya. Setiap titik dalam gambar mewakili satu mahasiswa, dengan warna yang berbeda menunjukkan kelompok atau klaster tempat mahasiswa tersebut tergolong. Penambahan nama pada setiap titik berfungsi untuk mengidentifikasi siapa saja yang termasuk dalam setiap klaster. Selain itu, tanda silang berwarna hitam menunjukkan posisi pusat kelompok (*centroid*) dari masing-masing klaster hasil pengelompokan.

Dari hasil visualisasi tersebut, dapat terlihat bahwa sebagian besar mahasiswa tergolong dalam satu kelompok besar, menandakan adanya kesamaan dalam preferensi gaya belajar mereka. Sebaliknya, terdapat satu atau dua mahasiswa yang membentuk kelompok kecil tersendiri, yang menandakan adanya perbedaan signifikan dalam cara mereka menyerap informasi dibandingkan dengan mayoritas. Hasil ini selaras dengan proses perhitungan sebelumnya, yang menunjukkan bahwa hanya tiga dari empat klaster memiliki anggota. Visualisasi ini memberikan gambaran yang lebih jelas mengenai sebaran kecenderungan belajar mahasiswa, serta mendukung pemahaman yang lebih mendalam terhadap pola pembelajaran yang berbeda-beda di antara mereka. Dengan informasi ini, dosen dan pihak institusi dapat mempertimbangkan penyesuaian metode pembelajaran agar lebih sesuai dengan karakteristik mahasiswa dalam setiap kelompok.

3.1.1 Evaluasi Hasil Klasterisasi Menggunakan Indeks DBI

Evaluasi kualitas klasterisasi dilakukan menggunakan *Davies-Bouldin Index* (DBI). Indeks ini mengukur seberapa baik klaster terbentuk berdasarkan tingkat kekompakan dalam klaster dan pemisahan antar klaster. Semakin kecil nilai DBI, semakin baik hasil klasterisasi. Berdasarkan hasil perhitungan, diperoleh nilai DBI sebesar 0,727. Nilai ini menunjukkan bahwa hasil pengelompokan mahasiswa cukup baik, dengan klaster yang relatif terpisah dan homogen di dalamnya. Berikut adalah visualisasi hasil nilai DBI yang diperoleh :

Davies Bouldin

Davies Bouldin: 0.727

Gambar 3. Hasil Evaluasi DBI

3.2 Pembahasan

Pengelompokan gaya belajar mahasiswa menggunakan algoritma *K-Means* menghasilkan empat klaster, namun hanya tiga klaster yang terisi, yaitu Klaster 1 dengan tujuh anggota, Klaster 2 dengan satu anggota, dan Klaster 4 dengan dua anggota. Klaster 3 tidak memiliki anggota karena tidak ada data mahasiswa yang cukup dekat jaraknya dengan pusat klaster tersebut. Proses iterasi berhenti pada iterasi kedua karena tidak terjadi perubahan pada pusat klaster, yang menandakan bahwa hasil klasterisasi telah stabil. Pola ini menunjukkan bahwa sebagian besar mahasiswa memiliki preferensi belajar yang mirip dan dapat dikelompokkan bersama, sementara sebagian kecil memiliki karakteristik yang lebih unik. Evaluasi terhadap hasil klasterisasi dilakukan dengan menggunakan indeks *Davies-Bouldin* (DBI), yang merupakan salah satu metode validasi internal dalam algoritma clustering. Hasil perhitungan menunjukkan nilai DBI sebesar 0,727. Nilai ini tergolong rendah dan mengindikasikan bahwa klaster yang terbentuk memiliki kekompakan internal yang baik dan jarak antar klaster yang cukup jelas. Semakin kecil nilai DBI, semakin baik kualitas klasterisasi. Oleh karena itu, angka tersebut menunjukkan bahwa metode *K-Means* amat efisien dalam mengelompokkan mahasiswa sesuai dengan preferensi gaya belajarnya yang diperoleh dari kuesioner VARK.

Jika dibandingkan dengan penelitian terdahulu, seperti studi oleh (Fish, 2020) dan (Muttakin & Defriani, 2020), hasil penelitian ini selaras dalam hal efektivitas algoritma *K-Means* untuk klasifikasi karakteristik mahasiswa. Penelitian-penelitian sebelumnya umumnya menggunakan *K-Means* untuk mengelompokkan data prestasi atau aktivitas pembelajaran, sedangkan penelitian ini secara spesifik mengelompokkan gaya belajar. Perbedaan ini menegaskan bahwa metode *K-Means* fleksibel dan dapat diterapkan untuk berbagai konteks data pendidikan. Namun, penelitian ini menambahkan validasi menggunakan indeks DBI, yang tidak banyak ditemukan pada studi sejenis sebelumnya. Proses pengolahan data juga menjadi faktor penting yang memengaruhi keberhasilan klasterisasi. Normalisasi data dilakukan agar semua atribut (*Visual, Auditory, Reading/Writing, Kinesthetic*) memiliki skala yang setara, sehingga tidak ada satu atribut pun yang mendominasi hasil pengelompokan. Selain itu, pemilihan jumlah klaster ($k = 4$) didasarkan pada pertimbangan jumlah gaya belajar utama yang terdapat dalam model VARK. Walaupun terdapat satu klaster yang hanya berisi satu individu (Arsyinta), hal ini bukan merupakan kesalahan, melainkan

mencerminkan bahwa terdapat gaya belajar yang cukup berbeda dari mayoritas responden. Keberadaan *outlier* semacam ini juga penting untuk diketahui agar institusi dapat merancang pendekatan pembelajaran yang lebih personal jika diperlukan.

Temuan ini memiliki implikasi praktis yang cukup kuat, khususnya bagi institusi pendidikan tinggi. Dengan mengetahui kecenderungan gaya belajar mahasiswa secara lebih objektif, pendidik dapat merancang metode pembelajaran yang lebih adaptif, sesuai dengan karakteristik mahasiswa di setiap klaster. Misalnya, mahasiswa yang tergolong dalam klaster dengan dominansi visual dapat diberikan materi berbasis grafis, infografis, atau video. Sebaliknya, mahasiswa kinestetik dapat lebih diuntungkan dengan pendekatan pembelajaran berbasis praktik dan simulasi. Meskipun penelitian ini menggunakan sampel terbatas, hasilnya menunjukkan potensi besar penerapan metode data mining dalam mendukung keputusan di bidang pendidikan. Ke depan, penelitian lanjutan disarankan menggunakan jumlah responden yang lebih besar dan mempertimbangkan perbandingan dengan algoritma clustering lain, seperti DBSCAN atau *Agglomerative*, untuk menguji konsistensi dan kualitas hasil pengelompokan.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengkategorikan gaya belajar mahasiswa menurut model VARK dengan memanfaatkan algoritma *K-Means* dan menilai hasil pengelompokan tersebut menggunakan indeks *Davies-Bouldin* (DBI). Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa metode *K-Means* dapat mengelompokkan mahasiswa ke dalam klaster yang mencerminkan kesamaan dalam preferensi belajar. Dari empat klaster yang ditentukan, tiga di antaranya berisi anggota, sedangkan satu klaster tidak terisi karena tidak ada data yang cukup dekat dengan pusat klaster tersebut. Hasil pengelompokan menunjukkan bahwa sebagian besar mahasiswa cenderung memiliki kesamaan dalam gaya belajar, terutama pada aspek visual dan kinestetik. Terdapat pula beberapa mahasiswa yang memiliki pola belajar yang lebih unik dan tergolong dalam klaster terpisah. Evaluasi menggunakan indeks DBI menghasilkan nilai sebesar 0,727 yang menunjukkan bahwa kualitas klasterisasi cukup baik, dengan tingkat kekompakan antaranggota klaster yang tinggi dan pemisahan yang jelas antar klaster.

Dari sisi teoritis, penelitian ini memperkuat pemahaman bahwa penerapan algoritma *clustering* seperti *K-Means* dapat menjadi pendekatan objektif dalam mengidentifikasi pola belajar mahasiswa secara kelompok. Secara praktis, hasil penelitian ini dapat menjadi dasar bagi lembaga pendidikan untuk merancang strategi pembelajaran yang lebih tepat sasaran dan adaptif, disesuaikan dengan karakteristik gaya belajar dari masing-masing klaster mahasiswa. Namun demikian, penelitian ini memiliki batasan pada jumlah responden yang tergolong sedikit, yaitu hanya 10 mahasiswa. Jumlah sampel yang terbatas dapat memengaruhi generalisasi temuan ke populasi yang lebih luas. Selain itu, pemilihan jumlah klaster ditentukan secara manual tanpa eksplorasi komparatif terhadap nilai k lainnya.

Sebagai tindak lanjut, disarankan agar penelitian serupa dilakukan dengan jumlah data yang lebih besar dan cakupan program studi yang lebih beragam. Selain itu, penggunaan algoritma clustering lainnya seperti DBSCAN atau *Agglomerative Hierarchical Clustering* dapat menjadi alternatif pembanding untuk menguji stabilitas dan akurasi hasil pengelompokan. Penelitian mendatang juga dapat menambahkan aspek visualisasi lanjutan atau pengayaan data untuk memperoleh pemahaman yang lebih dalam mengenai perilaku belajar mahasiswa.

ACKNOWLEDGEMENTS

Penulis mengucapkan terima kasih yang mendalam kepada semua pihak yang telah memberikan bantuan selama pelaksanaan penelitian ini. Ucapan terima kasih yang istimewa ditujukan kepada dosen pengampu yang telah memberikan petunjuk, saran, dan koreksi yang sangat berharga dalam penyusunan artikel ini. Penulis mengucapkan terima kasih kepada mahasiswa responden dari beragam program studi yang telah meluangkan waktu untuk mengisi kuesioner dengan lengkap dan jujur, sehingga data yang terkumpul dapat diolah dan dianalisis secara maksimal. Harapan kami, temuan dari penelitian ini bisa memberikan sumbangan yang berguna, baik untuk pengembangan ilmu pengetahuan di sektor pendidikan dan teknologi informasi, maupun untuk penerapan kebijakan pembelajaran yang lebih responsif di lingkungan pendidikan tinggi.

REFERENCES

Balti, R., Hedhili, A., & Chaari, W. L. (2023). Hybrid analysis of the learner's online behavior based on

- learning style. *Educ Inf Technol*, 28, 12465–12504. <https://doi.org/10.1007/s10639-023-11595-x>
- Essa, S. G., Celik, T., & Human-Hendricks, N. E. (2023). Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review. *IEEE Access*, 11(May), 48392–48409. <https://doi.org/10.1109/ACCESS.2023.3276439>
- Fish, B. (2020). *No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析* Title. 2507(February), 1–9.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Kanchon, M. K. H., Sadman, M., Nabila, K. F., Tarannum, R., & Khan, R. (2024). Enhancing personalized learning: AI-driven identification of learning styles and content modification strategies. *International Journal of Cognitive Computing in Engineering*, 5(July), 269–278. <https://doi.org/10.1016/j.ijcce.2024.06.002>
- Kristanto, B., Turmudi Zy, A., & M. Fatchan. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, 4(1), 112–120. <https://doi.org/10.47065/bit.v4i1.521>
- Liu, Y., Ma, S., & Du, X. (2024). A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Initial Cluster Centroids of K-means. *IEEE Access*, 1–1. <https://doi.org/10.1109/ACCESS.2020.3044069>
- Mindarsih, E., & Syafitri, T. (2017). Hubungan Antara Gaya Belajar Dan Hasil Belajar Mahasiswa Diii Kebidanan. *Edudikara: Jurnal Pendidikan Dan Pembelajaran*, 2(4), 340–347. <https://doi.org/10.32585/edudikara.v2i4.67>
- Muttaqin, M. R., & Defriani, M. (2020). Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa. *ILKOM Jurnal Ilmiah*, 12(2), 121–129. <https://doi.org/10.33096/ilkom.v12i2.542.121-129>
- Nanduarti, N. F., Abida, A., Azzahra, H., Pratama, D. A., Arumsari, F. Q., Saputro, R. A., Nugroho, J. T., & Tidar, U. (2024). *ANALISIS GAYA KOGNITIF BELAJAR PADA MAHASISWA*. 1(4), 227–236. <https://doi.org/10.70248/jolale.v1i4.1512>
- Rahmawati, R., Prihartono, W., & Cirebon, K. (2025). *OPTIMASI STOK DENGAN CLUSTERING DATA MENGGUNAKAN*. 13(2).
- Ran, X., Xi, Y., Lu, Y., Wang, X., & Lu, Z. (2023). *Comprehensive survey on hierarchical clustering algorithms and the recent developments*. 56(August 2023), 8219–8264. <https://doi.org/10.1007/s10462-022-10366-3>
- Ray, S., & Saeed, M. (2018). Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education. *Applications of Big Data Analytics*, 135–160. https://doi.org/10.1007/978-3-319-76472-6_7
- Wang, M., Fu, W., He, X., Hao, S., & Wu, X. (2022). A Survey on Large-Scale Machine Learning. *IEEE*, 34(6), 2574–2594. <https://doi.org/10.1109/TKDE.2020.3015777>
- Yuningsih, T., & Devi, W. S. (2024). Dinamika pembelajaran retorika dan berpikir kritis pada mahasiswa pendidikan bahasa dan sastra Indonesia Universitas Muhammadiyah Jakarta. *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, 10(2), 152. <https://doi.org/10.29210/020243706>