

Pengelompokan Hasil Pembelajaran Mahasiswa dengan Algoritma K-Means Clustering

Sitti Rahmah¹, Muh jamil², Andi Aslindah³, Aldi bas Fawait⁴, Yudhi Fajar Saputra⁵

^{1,2,3,4,5} Universitas Widya Gama Mahakam, Samarinda

Email: sitti.rahmah@uwgm.ac.id

 <https://doi.org/10.31004/jerkin.v3i4.1217>

ARTICLE INFO

Article history

Received: 20 April 2025

Revised: 31 April 2025

Accepted: 14 Juny 2025

Kata kunci

Hasil, Pembelajaran,
Algoritma, K-Means,
Clustering

Keywords

Results, Learning,
Algorithm, K-Means,
Clustering



ABSTRACT

Pengelompokan hasil pembelajaran mahasiswa adalah proses untuk mengklasifikasikan mahasiswa berdasarkan tingkat pemahaman atau pencapaian akademik yang diperoleh selama periode pembelajaran. Tujuan pengelompokan ini, untuk mengetahui kelompok mana yang membutuhkan perhatian lebih serta strategi pengajaran yang lebih tepat. Pengelompokan hasil pembelajaran mahasiswa dalam penelitian ini menggunakan algoritma clustering K-Means yang bertujuan untuk mengelompokkan mahasiswa berdasarkan kesamaan dalam pencapaian akademik mereka, sehingga memudahkan dalam analisis lebih lanjut. Sumber data penelitian ini berasal dari data sekunder sebuah universitas swasta di Samarinda, yang merupakan nilai mahasiswa semester genap tahun 2024 sebanyak 24 Mahasiswa. Temuan penelitian ini adalah cluster 0 (baik sekali) dengan persentase 50%. Cluster 1 (baik) dengan persentase 37,5%. Cluster 2 (cukup baik) dengan persentase 12,5%. Maka disimpulkan bahwa hasil pembelajaran mahasiswa adalah baik sekali, karena 87,5% mahasiswa mendapatkan nilai yang baik.

Grouping of student learning outcomes is a process of classifying students based on the level of understanding or academic achievement obtained during the learning period. The purpose of this grouping is to find out which groups need more attention and more appropriate teaching strategies. Grouping of student learning outcomes in this study uses the K-Means clustering algorithm which aims to group students based on similarities in their academic achievement, thus facilitating further analysis. The data source for this study comes from secondary data from a private university in Samarinda, which is the value of 24 students in the even semester of 2024. The findings of this study are cluster 0 (very good) with a percentage of 50%. Cluster 1 (good) with a percentage of 37.5%. Cluster 2 (quite good) with a percentage of 12.5%. So it is concluded that student learning outcomes are very good, because 87.5% of students get good grades.



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

How to Cite: Sitti Rahmah, et al (2025) Pengelompokan Hasil Pembelajaran Mahasiswa dengan Algoritma K-Means Clustering, 3(4). 4215 -4221 <https://doi.org/10.31004/jerkin.v3i4.1217>

PENDAHULUAN

Pengelompokan hasil pembelajaran mahasiswa merupakan upaya untuk menganalisis data hasil belajar. Hal ini bertujuan untuk mengidentifikasi pola dan mengelompokkan mahasiswa berdasarkan karakteristik tertentu, seperti nilai akademik, partisipasi kelas, ataupun tingkat kehadiran. Proses ini penting karena memberikan wawasan kepada pendidik dan institusi pendidikan tentang kinerja mahasiswa secara keseluruhan, sehingga dapat digunakan untuk mendesain strategi pembelajaran yang lebih efektif. Dengan mengelompokkan mahasiswa, institusi dapat mengidentifikasi kelompok mahasiswa yang membutuhkan intervensi khusus, ataupun merancang program pembelajaran tambahan

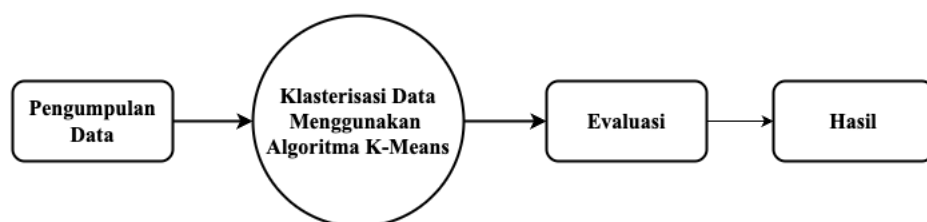
yang sesuai (Br Sembiring et al., 2022). Salah satu teknik yang populer digunakan dalam kasus pengelompokan adalah teknik Clustering. Clustering adalah suatu teknik dalam data mining yang bersifat unsupervised, yang berarti proses ini tidak memerlukan label (Harahap & Rambe, 2021). Dalam banyak kasus banyak peneliti yang memanfaatkan algoritma yang sudah ada seperti algoritma K-Means. Algoritma ini merupakan teknik yang digunakan untuk mengelompokkan suatu kumpulan data berdasarkan kesamaan atribut atau karakteristiknya ke dalam sejumlah kelompok data tertentu (Hidayat, 2022). Algoritma K-Means Clustering beroperasi dengan membagi data ke dalam beberapa kelompok atau cluster berdasarkan tingkat kesamaan data. Sehingga data dalam satu cluster akan memiliki kesamaan yang sangat tinggi, sedangkan data di luar cluster cenderung memiliki tingkat kesamaan yang jauh lebih rendah (Matdoan et al., 2023).

Penggunaan algoritma K-Means sebenarnya bukanlah hal yang baru di kalangan beberapa peneliti. Terbukti dari beberapa penelitian sebelumnya yang berkaitan dengan klasterisasi data banyak yang memanfaatkan algoritma K-Means, seperti penelitian yang dilakukan oleh Putra dan Yuniarti pada tahun 2020, klasifikasi hasil belajar mahasiswa dengan memanfaatkan variabel IPK (Putra & Yuniarti, 2020). Ada juga penelitian menggunakan algoritma K-Means untuk membantu dalam mengklasifikasikan kelas unggulan di sebuah sekolah (Pratiwi & Wahyu Wibowo, 2022). Salah satu faktor yang berpengaruh pada hasil klasterisasi data dengan algoritma K-Means adalah penentuan nilai K, nilai ini berkaitan dengan jumlah cluster optimum terhadap sebuah data semisal penelitian Anwar dan Syaifulloh dengan kasus pengelompokan hasil belajar siswa dengan nilai $K = 4$ (Anwar et al., 2022) (Syaefudulloh et al., 2022).

Tujuan penelitian yang dikerjakan adalah untuk mengelompokkan hasil pembelajaran mahasiswa untuk mengidentifikasi pola kinerja mahasiswa yang dapat membantu dalam merancang metode pembelajaran yang lebih efektif dan sesuai dengan kebutuhan mereka. Selain itu, pengelompokan ini dapat memberikan wawasan mengenai kelompok mahasiswa yang memerlukan dukungan tambahan, sehingga intervensi pendidikan dapat dilakukan secara lebih tepat sasaran.

METODE

Penelitian tentang pengelompokkan hasil pembelajaran mahasiswa dengan algoritma clustering K-Means bertujuan untuk mengidentifikasi pola atau kelompok mahasiswa berdasarkan hasil pembelajaran yang serupa. Tahapan dalam penelitian ini dapat dilihat pada gambar diagram berikut.



Gambar 1. Diagram tahapan penelitian

Desain penelitian dan Metode Pengumpulan Dataset

Penelitian ini termasuk dalam kategori penelitian deskriptif kuantitatif. Penelitian deskriptif kuantitatif bertujuan untuk memberikan gambaran mengenai suatu fenomena melalui data yang berbentuk angka, tanpa bermaksud menguji hipotesis tertentu (Anugrah & Dewi, 2024). Dataset yang diambil adalah nilai mahasiswa semester 2 Genap 2024. Pengumpulan dataset dilakukan di Universitas Widya Gama Mahakam Samarinda. Pada tahap ini dipilih sampel data dari 1 kelas pada Program Studi Ilmu Komputer, data yang digunakan adalah data nilai di 8 mata kuliah selama satu semester.

Metode Klasterisasi Data dengan K-Means

Tahap implementasi algoritma K-Means dilakukan dengan beberapa tahap sebagai berikut.

A. Menentukan Nilai K Awal dan Centroid Awal

Tahap awal adalah memilih nilai K secara acak pada data. Karena algoritma K-Means bekerja dengan cara membagi data berdasarkan nilai K yang telah ditentukan dan nilai K adalah jumlah kluster yang diinginkan (Hendrastuty, 2024). Kemudian dilanjutkan dengan menentukan titik centroid awal yang diperoleh secara acak (Muhammad Zulfadhilah et al., 2022). Penentuan jarak terdekat bisa

menggunakan metode Euclidean Distance dengan rumus yang ditunjukkan pada persamaan 1 berikut (Mangku Negara et al., 2021).

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_m - q_m)^2} \quad (1)$$

B. Menghitung Ulang Posisi Centroid

Tahap dilakukan dengan cara menentukan nilai centroid rata-rata dari setiap kelompok yang terbentuk pada tahap kedua. Proses penghitungan centroid baru ini dilakukan berdasarkan data yang telah dikelompokkan ke dalam masing-masing grup (Gustian & Al-Farits, 2023). Rumus yang digunakan ditunjukkan pada persamaan 2.

$$c = \frac{\sum m}{n} \quad (2)$$

C. Iterasi Hingga Optimal (Convergen)

Tahap terakhir adalah mengulangi iterasi pada proses sebelumnya hingga hasilnya convergen (Gustian & Al-Farits, 2023). Artinya tidak ada lagi perubahan yang signifikan pada posisi centroid atau kluster sudah stabil.

D. Metode Evaluasi Kinerja Klasterisasi

Metode evaluasi kinerja algoritma klasiterisasi yang digunakan adalah Davies-Bouldin index (DBI). Metode ini sudah cukup umum digunakan untuk menilai kinerja algoritma klastering seperti K-Means. DBI akan melakukan proses pengukuran hasil pengelompokan dengan parameter separasi dan kohesi. Parameter kohesi akan menghitung kedekatan data dengan titik pusat atau centroid (Suraya et al., 2023). Berdasarkan metrik Davies-Bouldin index (DBI) bahwa jika score DBI hasil pemisahan < 1,0 maka pemisahan kluster cukup baik, jika score 1,0-2,0 maka kluster tumpang tindih dan agak kabur, sementara jika > 2,0 pemisahan kluster buruk atau tidak terpisah dengan baik (Henderi et al., 2024).

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Dataset

Dataset yang telah berhasil dikumpulkan terdiri dari 24 instance dan 9 kolom. Kemudian 8 kolom digunakan sebagai kolom fitur (nilai mahasiswa pada 8 mata kuliah) dan 1 kolom yaitu identitas mahasiswa akan diabaikan. Dataset yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Dataset yang digunakan.

NO	Kode	AOK	BI	DB	KAL	PAK	PKN	PSD	TSD
1	M1	84	85	90	70	78	85	91	86
2	M2	87	85	85	72	80	81	75	94
3	M3	84,5	80	95	70	78	85	93	91
4	M4	87,5	85	95	80	80	85	96	95
5	M5	88,5	85	98	80	85	85	98	96
6	M6	89	80	85	74	75	81	94	70
7	M7	87	80	60	70	70	70	75	69
8	M8	87,5	80	75	72	70	85	96	79
9	M9	87,5	85	98	90	85	85	99	92
10	M10	87,5	85	98	90	78	85	99	96
11	M11	84	80	80	70	78	85	89	80
12	M12	85	80	85	72	70	81	91	72
13	M13	84	80	80	72	70	80	80	71
14	M14	84,5	80	90	72	85	85	89	70
15	M15	86,5	80	80	72	78	85	93	94
16	M16	82,5	80	60	70	78	80	75	70
17	M17	87,5	80	90	70	75	82	91	87
18	M18	88	85	98	70	76	82	98	95
19	M19	86,5	80	75	70	73	70	78	60
20	M20	87,5	85	90	80	80	82	99	92
21	M21	87,5	85	98	85	70	81	95	84
22	M22	87	80	85	70	70	81	82	73

23	M23	84,5	80	98	73	70	80	75	65
24	M24	87	80	80	70	75	82	96	78

Hasil Penentuan Nilai K dan Centroid Awal

Hasil penentuan nilai K dan Centroid awal dilakukan dengan menggunakan nilai K = 3 dengan hasil ditunjukkan pada Tabel 2.

Tabel 2. Centroid awal setiap cluster.

Centroid	AOK	BI	DB	KAL	PAK	PKN	PSD	TSD
0	88,5	85	98	80	85	85	98	96
1	86,5	80	75	70	73	70	78	60
2	82,5	80	60	70	78	80	75	70

Selanjutnya dilakukan perhitungan jarak antara data dengan centroid menggunakan persamaan 1 yang disajikan pada Tabel 3.

Tabel 3. Jarak terdekat dengan cluster pada iterasi 1.

Kode	C0	C1	C2	Jarak Terdekat	Klaster
M1	19,55121	36,75935	38,26552	19,55121	0
M2	28,44732	38,26552	35,42951	28,44732	0
M3	15,77973	42,89522	44,93328	15,77973	0
...	...	...	...	...	...
M7	57,26474	18,00694	13,61066	13,61066	2
M8	33,71943	30,39737	29,41088	29,41088	2
...	...	...	...	...	...
M22	36,36276	20,37768	27,71732	20,37768	1
M23	42,7785	26,1725	39,31921	26,1725	1
M24	29,80352	28,65746	30,63087	28,65746	1

Hasil Perhitungan Ulang Posisi Centroid

Proses selanjutnya dilanjutkan pada iterasi ke 2 dan 3 menggunakan persamaan 2. berdasarkan rata-rata anggota dalam klaster sebelumnya. Hasil iterasi ke 3 ditampilkan pada Tabel 4 dan Tabel 5.

Tabel 4. Centroid 3 setiap cluster.

Centroid	AOK	BI	DB	KAL	PAK	PKN	PSD	TSD
0	86,95833	83,75	92,9166667	77,416667	78,5833333	83,583	93,917	91,83
1	85,9375	80	84,75	71,625	73,5	80	85,625	69,88
2	85,25	80	68,75	70,5	74	80	83,75	74,5

Tabel 5. Jarak terdekat dengan cluster pada iterasi 3.

Kode	C0	C1	C2	Jarak Terdekat	Klaster
M1	10,895	18,69	31,38	10,895	0
M2	21,559	26,72	31,52	21,559	0
M3	9,1288	24,59	37,84	9,1288	0
...	...	...	...	...	...
M7	47,854	29,03	12,89	12,893	2
M8	24,686	16,98	19,82	16,979	1
...	...	...	...	...	...
M22	26,623	6,366	20,41	6,3657	1
M23	35,069	19,16	34,83	19,159	1
M24	21,141	13,18	21,69	13,177	1

Tabel 5 menunjukkan terdapat perubahan data salah satunya adalah pada data Mahasiswa ke 8. Pada iterasi pertama posisi Mahasiswa 8 berada pada kluster 2, namun pada iterasi ke 2 posisi mahasiswa berpindah ke kluster 1. Sehingga iterasi terus dilanjutkan.

Hasil Iterasi Optimal

Setelah dilakukan proses yang sama sampai dengan iterasi ke 4 maka nilai centroid dan cluster sudah tidak mengalami perubahan yang signifikan. Sehingga data sudah mencapai titik convergen, hasil iterasi ke 4 ditampilkan pada Tabel 6 dan Tabel 7.

Tabel 6. Centroid 4 setiap cluster.

Centroid	AOK	BI	DB	KAL	PAK	PKN	PSD	TSD
0	87	83,75	92,92	77,42	78,58	83,6	93,92	91,83
1	85,8	80	84,22	71,67	73,67	82,2	88	73,11
2	85,3	80	65	70	73,67	73,3	76	66,33

Tabel 7. Jarak terdekat dengan cluster pada iterasi 4.

Kode	C0	C1	C2	Jarak Terdekat	Kluster
M1	10,895	16,31	37,66	10,895	0
M2	21,559	25,96	36,02	21,559	0
M3	9,1288	22,19	44,19	9,1288	0
M4	5,5903	28,36	49,27	5,5903	0
M5	10,662	32,43	53,41	10,662	0
M6	24,265	8,064	28,77	8,064	1
M7	47,854	30,65	7,775	7,7746	2
M8	24,686	14,41	28,61	14,414	1
M9	15,971	34,17	54,51	15,971	0
M10	15,218	35	55,52	15,218	0
M11	20,278	9,946	27,16	9,9457	1
M12	24,334	5,156	27,08	5,1559	1
M13	30,548	10,4	18,07	10,399	1
M14	24,536	13,49	32,81	13,495	1
M15	14,776	22,5	37,94	14,776	0
M16	44,946	28,34	10,52	10,521	2
M17	11,184	15,58	36,86	11,184	0
M18	10,926	28,41	50,07	10,926	0
M19	43,274	22,58	12,53	12,531	2
M20	6,8861	25,47	45,36	6,8861	0
M21	15,102	24,04	45,71	15,102	0
M22	26,623	7,463	23,57	7,4629	1
M23	35,069	21,13	34,05	21,132	1
M24	21,141	10,57	29	10,569	1

Hasil Evaluasi Kinerja Klusterisasi

Proses akhir dari proses klusterisasi adalah melakukan evaluasi kinerja algoritma K-Means yang digunakan. Dengan menggunakan metode Davies-Bouldin index (DBI) maka diperoleh skor sebesar 0.9595. Berdasarkan rujukan pada penelitian [15] menyatakan bahwa nilai DBI < 1 masih bisa dianggap cukup baik dan bisa diterima.

SIMPULAN

Berdasarkan analisa yang dilakukan dari algoritma dan dataset yang digunakan maka dapat ditarik beberapa kesimpulan sebagai berikut.

1. Penelitian yang dikerjakan menghasilkan 3 klaster yang mencapai titik convergen pada iterasi ke 4 dengan score DBI sebesar 0.9595.
2. Terdapat 12 mahasiswa dengan predikat baik sekali dan tergolong ke dalam klaster pertama yaitu M1-M5, M9, M10, M15, M17, M18, M20, M21.
3. Terdapat 9 mahasiswa dengan predikat baik dan tergolong dalam klaster ke dua yaitu M6, M8, M11-M14, M22-M24.
4. Mahasiswa yang tergolong ke dalam predikat cukup baik dan masuk ke dalam klaster 3 yaitu M7, M16 dan M19.
5. Secara persentase pembagian jumlah mahasiswa setelah dilakukan proses klasterisasi adalah klaster 1 = 50%, klaster 2 = 37,5% dan klaster 3 = 12,5%. Sehingga dapat dikatakan bahwa hanya ada 3 mahasiswa yang membutuhkan stimulus atau rangsangan agar dapat meningkatkan hasil belajarnya

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Anugrah, P., & Dewi, C. (2024). *ANALISIS KARAKTER ENTREPRENEUR BERBASIS MATEMATIKA MELALUI MODEL PROJECT BASED LEARNING ANALYSIS OF ENTREPRENEUR CHARACTER BASED MATHEMATICS THROUGH PROJECT BASED LEARNING MODEL*. 10(2), 144–150.
- Anwar, S., Suprpti, T., Dwilestari, G., & Ali, I. (2022). Pengelompokan Hasil Belajar Siswa dengan Metode Clustering K-Means. *JURSISTEKNI (Jurnal Sistem Informasi Dan Teknologi Informasi)*, 4(2), 60–72. <https://doi.org/https://doi.org/10.52005/jursistekni.v4i2.99>
- Br Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31. <https://doi.org/10.53513/jursi.v1i1.4784>
- Gustian, D., & Al-Farits, M. S. (2023). Data Mining Untuk Melihat Minat Belajar Siswa Menerapkan Metode K-Means. *Journal of Information System Research (JOSH)*, 4(3), 775–784. <https://doi.org/10.47065/josh.v4i3.3218>
- Harahap, B., & Rambe, A. (2021). Implementasi K-Means Clustering Terhadap Mahasiswa yang Menerima Beasiswa Yayasan Pendidikan Battuta di Universitas Battuta Tahun 2020/2021 Studi Kasus Prodi Informatika. *Informatika*, 9(3), 90–97. <https://doi.org/10.36987/informatika.v9i3.2185>
- Henderi, H., Fitriana, L., Iskandar, I., Astuti, R., Arifandy, M. I., Hayadi, B. H., Mesran, M., Chin, J., & Kurniawan, A. (2024). Optimization of Davies-Bouldin Index with k-medoids algorithm. *Science and Technology Research Symposium 2022*, 3065(September), 030002. <https://doi.org/10.1063/5.0225220>
- 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/https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Hidayat, R. (2022). Pemanfaatan Data Mining untuk Melihat Minat Siswa Setelah Menyelesaikan Pendidikan Sekolah Menengah Atas (SMA) dengan Algoritma K-Means Clustering. *Technology and Informatics Insight Journal*, Volume 1(2). <https://doi.org/https://doi.org/10.32639/tij.v1i2.220>
- Mangku Negara, I. S., Purwono, P., & Ashari, I. A. (2021). Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means. *JOINTECS (Journal of Information Technology and Computer Science)*, 6(3), 153. <https://doi.org/10.31328/jointecs.v6i3.2693>
- Matdoan, M. Y., Purnamasari, N. A., & Laamena, N. S. (2023). Application of the K-Means Algorithm for Clustering Production of Capture Fisheries in Maluku Province. *Pattimura International Journal of Mathematics (PIJMath)*, 2(2), 63–70. <https://doi.org/10.30598/pijmathvol2iss2pp63-70>
- Muhammad Zulfadhilah, Mambang, & Septyan Eka Prastya. (2022). Implementasi Metode K-Means

- Clustering untuk Meningkatkan Penjaringan Mahasiswa. *Tematik*, 9(2), 152–160. <https://doi.org/10.38204/tematik.v9i2.1053>
- Pratiwi, H., & Wahyu Wibowo, A. P. (2022). Implementasi Algoritma K-Means Untuk Mengklaster Kelompok Sektor Perkebunan Di Indonesia. *JOISIE Journal Of Information System And Informatics Engineering*, 6(1), 1–10. <https://doi.org/https://doi.org/10.35145/joisie.v6i1.2252>
- Putra, B. J. M., & Yuniarti, D. A. F. (2020). Analisis Hasil Belajar Mahasiswa Dengan Clustering Menggunakan Metode K-Means. *Poros Teknik*, 12(2), 49–58. <https://doi.org/https://doi.org/10.25077/TEKNOSI.v2i3.2016.31-40>
- Suraya, S., Sholeh, M., & Lestari, U. (2023). Evaluation of Data Clustering Accuracy using K-Means Algorithm. *International Journal of Multidisciplinary Approach Research and Science*, 2(01), 385–396. <https://doi.org/10.59653/ijmars.v2i01.504>
- Syaefudulloh, M., Faqih, A., & Basysyar, F. M. (2022). Clustering Kelompok Belajar Siswa Berdasarkan Hasil Ujian Sekolah Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Dan Manajemen*, Volume 10(1), 195–199. <https://doi.org/https://doi.org/10.47024/js.v10i1.397>