

IMPLEMENTASI CLUSTERING DATA NILAI SISWA MENGGUNAKAN ALGORITMA K-MEANS: SEBUAH STUDI KASUS DI SMK NASIONAL NAMOTERASI

Arya Wira Permana¹, Rusmin Saragih², I Gusti Prahmana³
Teknik Informatika, STMIK Kaputama, Binjai

E-mail: *aryawira2810@gmail.com¹, evitha1204@gmail.com², igustiprahmana4@gmail.com³

ABSTRAK

Pemanfaatan teknologi informasi dalam pendidikan saat ini menjadi strategi penting, khususnya dalam mengelompokkan siswa berdasarkan kemampuan akademik. Penelitian ini bertujuan untuk menerapkan algoritma K-Means *clustering* pada data nilai siswa SMK Nasional Namoterasi untuk menentukan kelompok pembinaan akademik. Metodologi penelitian mengikuti tahapan *Knowledge Discovery in Database* (KDD) yang mencakup seleksi data, *preprocessing*, transformasi data ke bentuk skoring, penentuan jumlah kluster, dan proses *clustering* dengan K-Means, serta evaluasi hasil. Dataset yang digunakan adalah nilai mata pelajaran umum dan kejuruan siswa dari tahun ajaran 2022/2023 hingga 2024/2025. Hasil pengelompokan berhasil membagi 300 data siswa menjadi tiga kluster, yaitu siswa dengan capaian akademik tinggi, sedang, dan rendah. Hasil ini memungkinkan pihak sekolah menyusun strategi pembinaan yang lebih tepat sasaran, seperti program remedial untuk siswa berprestasi rendah, pembinaan reguler untuk siswa berprestasi sedang, dan program akselerasi bagi siswa berprestasi tinggi. Dengan demikian, penerapan algoritma K-Means terbukti efektif dalam mendukung pengambilan keputusan berbasis data untuk meningkatkan kualitas pembelajaran.

Kata kunci

Clustering, K-Means, Data Mining, Pembinaan Akademik, MATLAB.

ABSTRACT

This study aims to cluster student academic scores using the K-Means algorithm to determine academic coaching groups at SMK Nasional Namoterasi. The research method follows the stages of Knowledge Discovery in Database (KDD), including data selection, preprocessing, transformation into scoring format, determining the optimal number of clusters using the Elbow Method, clustering with the K-Means algorithm, and evaluating the results with MATLAB. The dataset consists of students' scores from both general and vocational subjects, which were successfully grouped into three clusters: high, medium, and low academic achievement. The clustering results enable the school to design more targeted academic coaching strategies, such as remedial programs for low-achieving students, regular coaching for medium-level students, and enrichment or acceleration programs for high-achieving students. Thus, the implementation of the K-Means algorithm has proven effective in supporting data-driven decision-making to improve the quality of learning.

Keywords

Clustering, K-Means, Data Mining, Academic Coaching, MATLAB

1. PENDAHULUAN

Pengambilan keputusan di dunia pendidikan kini semakin mengandalkan data, dan salah satu pendekatannya adalah dengan mengelompokkan siswa berdasarkan nilai akademik menggunakan algoritma *clustering*. Metode K-Means adalah salah satu yang populer dan efisien karena kemampuannya dalam mengidentifikasi kelompok-kelompok homogen dari data yang besar (Oktarian et al., 2020). Pengelompokan ini sangat bermanfaat bagi sekolah untuk merancang program pembinaan akademik yang spesifik dan terarah, sesuai dengan kebutuhan tiap kelompok siswa.

Beberapa penelitian terdahulu telah membuktikan efektivitas metode K-Means dalam konteks pendidikan. Studi yang dilakukan di SMK Negeri 1 Gowa menunjukkan

bahwa *clustering* dapat membagi siswa menjadi tiga kelompok, yang membantu guru dalam menentukan strategi pembelajaran yang sesuai (Nur Risal et al., 2024). Demikian pula, penelitian di SMK PGRI 2 Karawang membandingkan K-Means dengan Fuzzy C-Means dan menemukan bahwa K-Means memberikan hasil pengelompokan yang lebih optimal (Kurniawan et al., 2023).

Selain nilai akademik, faktor lain seperti kehadiran juga dapat dipertimbangkan untuk membentuk klaster yang lebih komprehensif, mencerminkan disiplin dan prestasi siswa secara bersamaan (Mukhsyi et al., 2024). Berdasarkan penelitian-penelitian tersebut, K-Means dapat memberikan dasar yang kuat bagi sekolah untuk merancang intervensi pendidikan berbasis data (Siregar et al., 2024). Dengan melihat urgensi dan manfaat dari pendekatan ini, penelitian ini bertujuan untuk mengimplementasikan algoritma K-Means pada data nilai siswa SMK Nasional Namoterasi. Hasil dari penelitian ini diharapkan dapat membantu pihak sekolah dalam mengidentifikasi kelompok pembinaan akademik, yang pada akhirnya dapat meningkatkan mutu pendidikan dan prestasi siswa secara keseluruhan.

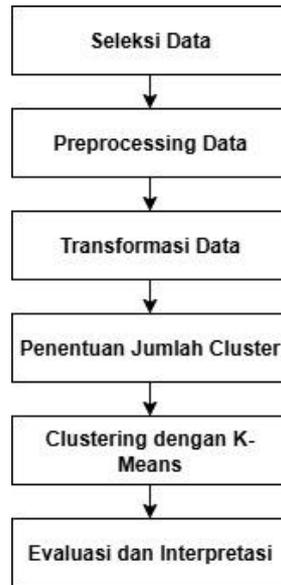
2. METODE PENELITIAN

Penelitian ini menggunakan metodologi *Knowledge Discovery in Database* (KDD), sebuah proses yang sistematis dan interaktif untuk menemukan pola dan pengetahuan dari data (Kesuma et al., 2020). Tahapan KDD dalam penelitian ini adalah sebagai berikut:

- a. Seleksi Data: Mengumpulkan data nilai mata pelajaran umum dan kejuruan dari 300 siswa SMK Nasional Namoterasi.
- b. *Preprocessing*: Melakukan pembersihan data untuk mengatasi data yang hilang atau tidak lengkap.
- c. Transformasi Data: Mengubah data nilai numerik menjadi bentuk skoring (1-5) seperti pada Tabel 2 untuk memudahkan analisis.
- d. Penentuan Jumlah Klaster: Menggunakan metode *Elbow Method* untuk menentukan jumlah klaster (k) yang optimal.
- e. *Clustering* dengan K-Means: Mengelompokkan data siswa berdasarkan kemiripan nilai.
- f. Evaluasi dan Interpretasi: Menganalisis hasil klasterisasi dan memberikan rekomendasi untuk pembinaan akademik.

Metodologi ini digambarkan dalam bentuk *flowchart* yang komprehensif (lihat Gambar 1). Perhitungan jarak antar data dan pusat klaster menggunakan rumus *Euclidean Distance* yang didefinisikan sebagai berikut:

$$D(ij) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + (X3i - X3j)^2 + (X4i - X4j)^2}$$



Gambar 1. Tahapan Clustering

2.1 Tahapan Penelitian

- a. Seleksi Data
Mengambil data nilai siswa SMK Nasional Namoterasi pada mata pelajaran umum dan kejuruan.
- b. Preprocessing Data
Melakukan pembersihan data (cleaning), memastikan tidak ada data kosong atau outlier yang bisa mengganggu proses analisis.
- c. Transformasi Data
Mengubah data nilai ke dalam bentuk skoring (rentang 1-5) agar lebih mudah diproses dan diinterpretasikan dalam clustering.
- d. Penentuan Jumlah Cluster
Menggunakan Elbow Method untuk menentukan nilai k (jumlah cluster) yang paling optimal.
- e. Clustering dengan Algoritma K-Means
 - 1) Melakukan pengelompokan data nilai siswa menggunakan algoritma K-Means berdasarkan variabel yang telah ditransformasi.
 - 2) Proses ini dilakukan melalui perhitungan jarak (Euclidean Distance), pembentukan centroid, dan iterasi hingga hasil stabil.
- f. Evaluasi dan Interpretasi Hasil
 - 1) Mengevaluasi kualitas cluster yang terbentuk.
 - 2) Menafsirkan hasil clustering dengan membagi siswa ke dalam kelompok berprestasi tinggi, sedang, dan rendah, lalu menghubungkannya dengan strategi pembinaan akademik yang sesuai (remedial, reguler, pengayaan).

3. HASIL DAN PEMBAHASAN

3.1 Transformasi dan Pengujian

Data nilai siswa dari SMK Nasional Namoterasi ditransformasi menjadi skala 1-5 berdasarkan kategori nilai yang telah ditetapkan, seperti yang ditunjukkan pada Tabel 1.

Tabel I. Transformasi Data Mapel C3 dan Mapel Umum

No	Nilai Rata-Rata	Nilai Transformasi
1	51 – 60	1
2	61 – 70	2
3	71 – 80	3
4	81 – 90	4
5	91 – 100	5

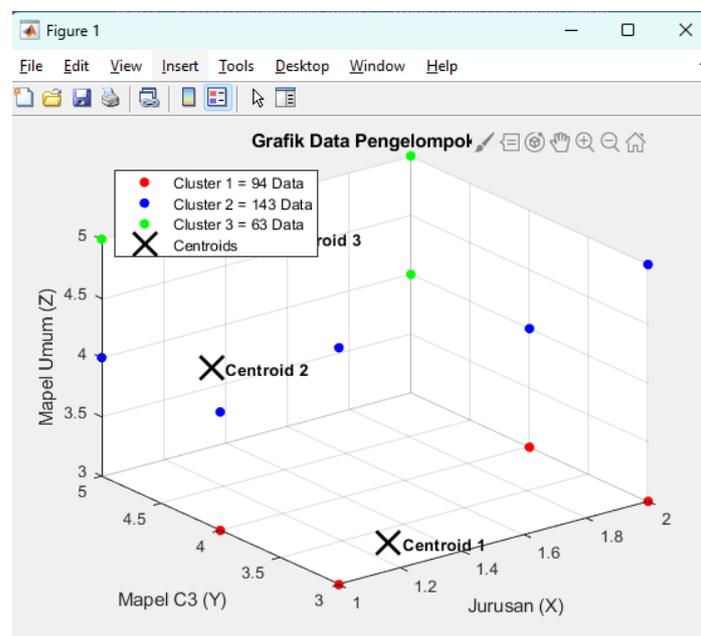
Untuk pengujian awal, dipilih 30 data siswa dan dihitung jaraknya ke tiga pusat *cluster* (centroid) yang diinisialisasi secara acak. Proses iterasi terus dilakukan hingga tidak ada perubahan pada keanggotaan klaster, yang menandakan bahwa algoritma telah konvergen. Perhitungan iterasi I dan II menunjukkan adanya perubahan keanggotaan klaster, sehingga proses dilanjutkan ke iterasi III. Pada iterasi III, keanggotaan klaster sudah stabil, dan perhitungan berhenti.

3.2 Analisis dan Kategori Klaster

Berdasarkan hasil pengujian dengan perangkat lunak MATLAB, 300 data siswa berhasil dikelompokkan ke dalam tiga klaster dengan karakteristik sebagai berikut:

Tabel 2. Pusat Cluster (Centroid) dan Jumlah Data

No	Pusat Cluster	X (Jurusan)	Y (Mapel C3)	Z (Mapel Umum)	Jumlah Data
1	Centroid 1	1.2872	3.3298	3.0000	94
2	Centroid 2	1.1469	4.4545	4.0559	143
3	Centroid 3	1.5238	5.0000	4.6349	63



Gambar 2. Grafik Hasil Proses Pengelompokan

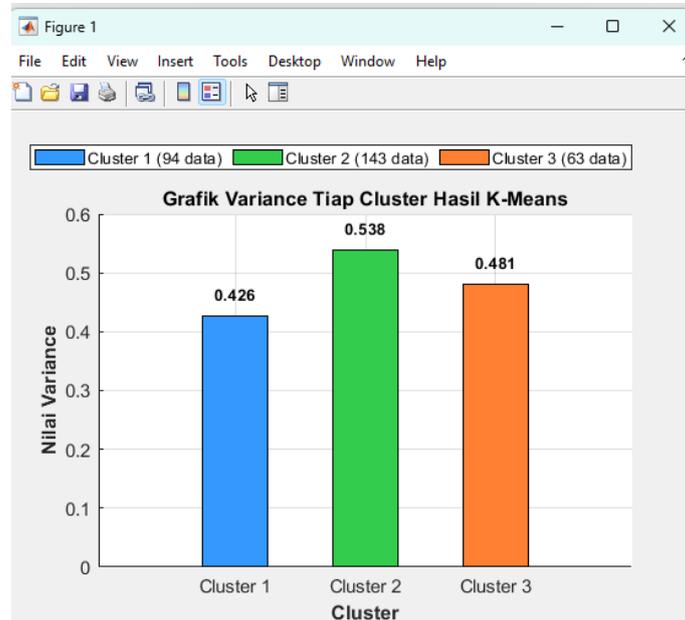
Berdasarkan nilai centroid dan sebaran data, setiap klaster dapat diinterpretasikan menjadi kategori pembinaan akademik yang spesifik.

- Klaster 1 (Kelompok Rendah): Terdiri dari 94 siswa, dengan nilai rata-rata yang mengindikasikan capaian akademik rendah (nilai transformasi sekitar 3.00). Kelompok ini memerlukan program remedial dan bimbingan belajar intensif.
- Klaster 2 (Kelompok Sedang): Terdiri dari 143 siswa, dengan nilai rata-rata sedang (nilai transformasi sekitar 4.00). Kelompok ini membutuhkan pembinaan reguler

untuk mempertahankan dan meningkatkan performa.

- c. Klaster 3 (Kelompok Tinggi): Terdiri dari 63 siswa, dengan nilai rata-rata tinggi (nilai transformasi di atas 4.50). Kelompok ini dapat mengikuti program pengayaan atau akselerasi.

Selain itu, perhitungan *cluster variance* juga dilakukan untuk mengukur tingkat homogenitas data dalam setiap klaster, seperti yang ditunjukkan pada Gambar 3.



Gambar 3. Grafik Cluster Variance

Hasil perhitungan menunjukkan bahwa Klaster 1 memiliki varian terkecil (0.426), yang berarti anggotanya paling homogen. Sebaliknya, Klaster 2 memiliki varian terbesar (0.538), menunjukkan sebaran data yang lebih heterogen.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means untuk mengelompokkan data nilai siswa SMK Nasional Namoterasi. Hasilnya menunjukkan bahwa data siswa dapat dikelompokkan secara efektif ke dalam tiga klaster yang merepresentasikan tingkat capaian akademik rendah, sedang, dan tinggi.

- a. Algoritma K-Means terbukti efektif dalam mengelompokkan siswa secara otomatis dan objektif berdasarkan kemiripan nilai akademik mereka.
- b. Hasil *clustering* menghasilkan tiga kelompok yang berbeda, yang masing-masing dapat dijadikan dasar untuk merancang program pembinaan akademik yang terarah, seperti remedial, pembinaan reguler, dan pengayaan.
- c. Penggunaan aplikasi MATLAB mempermudah proses analisis dan visualisasi data, sehingga hasil pengelompokan dapat diinterpretasikan dengan lebih mudah oleh pihak sekolah.

5. DAFTAR PUSTAKA

Apriandi, D., Sari, R. M., & Sarif, M. I. (2024). Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means. *Jurnal Minfo Polgan*, 13(1), 1117-1129.

- <https://doi.org/10.33395/jmp.v13i1.13959>
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 82–89. <https://doi.org/10.25077/teknosi.v6i2.2020.82-89>
- Kesuma, P. W., Risalah, A., & Purba, B. P. (2020). Penerapan Data Mining Dalam Proses Pengelompokan Data Masyarakat Kurang Mampu di Kota Deli Serdang Menggunakan Metode Clustering. *Jurnal Masyarakat Informatika Sumatera Utara*, 10(5). <https://doi.org/10.14710/jmasif.v7i1.10794>
- Kurniawan, S., Siregar, A. M., & Novita, H. Y. (2023). Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik. *Scientific Student Journal for Information, Technology and Science*, IV(1), 73–81.
- Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences (Switzerland)*, 12(19). <https://doi.org/10.3390/app12199467>
- Muhammad Hilman, Martanto, Dikananda, A. R., & Rifai, A. (2025). K-Means Algorithm for Clustering High-Achieving Student at Madrasah Tsanawiyah Yami Waled. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), 1538–1548. <https://doi.org/10.59934/jaiea.v4i3.771>
- Mukhsyi, A., Syukri, M., & Yanti, H. (2024). Analisis Kelompok Siswa Berdasarkan Nilai Akademik dan Non-Akademik dengan K-Means. *Jurnal Riset Pendidikan dan Sains*, 5(2), 211–220.
- Oktarian, S., Defit, S., & Sumijan. (2020). Klusterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma K-Means Clustering. *Jurnal Informasi Dan Teknologi*, 2(3). <https://doi.org/http://www.jidt.org>
- Parasian D P, S., & Irene Sri, M. (2017). Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering. *Jurnal TIMES*, VI(Vol 6, No 2 (2017)), 22–25.
- Purba, C., Sonata, F., Info, A., & Mining, D. (2020). Penerapan Data Mining Dalam Menganalisa Pola. *CyberTech*, 3(7), 1–8.
- Qusyairi, M., Zul Hidayatullah, & Arnila Sandi. (2024). Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 500–510. <https://doi.org/10.29408/jit.v7i2.26375>
- Rahayu, N. D., Anshor, A. H., & Afriantoro, I. (2024). Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means. *JUKI: Jurnal Komputer Dan Informatika*, 6(1), 71–83. <https://doi.org/10.53842/juki.v6i1.474>
- Sinaga, A. V., Saifullah, & Hardinata, J. T. (2021). Penerapan Algoritma Backpropagation dalam Memprediksi Kebutuhan Blangko Sertipikat Tanah pada Kantor BPN Kota Pematangsiantar. *TIN: Terapan Informatika Nusantara*, 1(4). <https://doi.org/Doi : https://ejurnal.seminar-id.com/index.php/tin>
- Siregar, A. M., Kurniawan, S., & Al Fariz, A. (2024). Pemanfaatan K-Means untuk Pengelompokan Kelas Unggulan Berbasis Nilai Akademik. *Jurnal Penelitian dan Pengembangan Pendidikan*, 8(1), 88–97.
- Suarna, N., Rahaningsih, N., & Suarna, A. A. (2025). Optimalisasi Prestasi Akademik Siswa

Melalui Pengelompokan Indeks Prestasi Dengan K-Means Clustering. Jurnal Kecerdasan Buatan Dan Teknologi Informasi, 4(2), 198-207.
<https://doi.org/10.69916/jkbt.v4i2.321>