

# ANALISIS KLASTERISASI PRESTASI SISWA DENGAN METODE K-MEANS

Achmad Lutfi Fuadi<sup>1,\*</sup>, Agus Tri Haryanto<sup>2</sup>

<sup>1,2</sup>) Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang  
Jl. Surya Kencana No.1, Pamulang Bar., Kec. Pamulang, Kota Tangerang Selatan, Banten 15417  
e-mail: dosen02524@unpam.ac.id<sup>1</sup>), channelagus2020@gmail.com<sup>2</sup>)

\*corresponding author

(Naskah masuk : 14 Agustus 2025 Diterima untuk diterbitkan : 24 November 2025)

## ABSTRAK

*Evaluasi prestasi siswa di SMP Negeri 92 Jakarta sering kali terbatas pada nilai rata-rata dan perbandingan Kriteria Ketuntasan Minimal (KKM), sehingga pola capaian akademik yang lebih mendalam kurang tergali dan program pembinaan menjadi kurang spesifik. Untuk mengatasi hal ini, penelitian ini mengaplikasikan teknik data mining dengan algoritma K-Means untuk mengelompokkan 250 siswa kelas VIII Tahun Ajaran (TA) 2024/2025 secara objektif berdasarkan 30 atribut nilai dari tiga jenis asesmen pada sepuluh mata pelajaran. Proses analisis diawali dengan pra-pemrosesan data menggunakan normalisasi Min-Max untuk menyetarakan skala nilai, diikuti penentuan jumlah kluster (k) optimal menggunakan metrik Davies-Bouldin Index (DBI) untuk memastikan kualitas pemisahan kluster. Klasterisasi kemudian dilakukan menggunakan K-Means dan dianalisis berdasarkan nilai centroid-nya. Hasil analisis menunjukkan jumlah kluster paling optimal adalah k=2, yang dibuktikan dengan skor DBI terendah (-4,625). Klasterisasi ini berhasil membagi siswa menjadi dua kelompok berbeda: Cluster\_1 (133 siswa) yang diidentifikasi sebagai kelompok berprestasi tinggi, dan Cluster\_0 (117 siswa) sebagai kelompok berprestasi sedang hingga rendah. Pengelompokan ini mengubah evaluasi subjektif menjadi pemetaan berbasis data yang dapat ditindaklanjuti. Kontribusi utama penelitian ini adalah menyediakan model pemetaan prestasi siswa yang objektif dan tervalidasi, yang dapat ditindaklanjuti secara praktis oleh institusi pendidikan untuk beralih dari evaluasi subjektif ke strategi pembinaan berbasis data.*

**Kata Kunci:** Data Mining, Davies-Bouldin Index, K-Means, Klasterisasi, Prestasi Siswa.

## ABSTRACT

*Student achievement evaluation at SMP Negeri 92 Jakarta is often limited to average scores and minimum completeness criteria, failing to capture deeper academic patterns and leading to less specific, non-data-driven support programs. This study applies K-Means clustering to objectively group 250 eighth-grade students (2024/2025) based on 30 score attributes from three assessment types across ten subjects. The methodology includes Min-Max normalization to equalize data scales, optimal cluster number (k) determination using the Davies-Bouldin Index (DBI) to ensure cluster quality, and subsequent K-Means clustering with centroid analysis. The results indicate that k=2 is the optimal number of clusters, achieving the lowest DBI score (-4.625). The clustering successfully partitioned students into two distinct groups: Cluster\_1 (133 students) as the high-achieving group and Cluster\_0 (117 students) as the medium-to-low achieving group. This data-driven mapping provides an objective foundation for designing targeted interventions. The primary contribution of this research is the provision of an objective and validated student performance mapping model, enabling educational institutions to transition from subjective evaluations to data-driven support strategies.*

**Keywords:** Clustering, Data Mining, Davies-Bouldin Index, K-Means, Student Achievement.

## I. PENDAHULUAN

Evaluasi hasil belajar di jenjang sekolah menengah sering kali membandingkan nilai siswa terhadap Kriteria Ketuntasan Minimal (KKM), namun pendekatan ini belum mampu menggambarkan pola capaian akademik secara menyeluruh [1]. Akibatnya, program pembinaan siswa menjadi kurang spesifik dan tidak sepenuhnya berbasis data. Keterbatasan ini menciptakan kesenjangan (gap) di mana evaluasi masih bersifat subjektif dan reaktif, padahal tuntutan pendidikan modern memerlukan

pendekatan yang lebih proaktif dan personal. Terdapat urgensi untuk menyediakan alat bantu (tools) berbasis data bagi institusi pendidikan agar dapat mengidentifikasi kebutuhan siswa secara akurat dan merancang program pembinaan yang lebih efektif dan tepat sasaran.

Untuk menjawab tantangan ini, pemanfaatan teknik data mining dapat menjadi solusi, di mana metode klusterisasi mampu mengelompokkan siswa secara objektif ke dalam grup-grup berdasarkan kemiripan pola prestasi mereka tanpa memerlukan label awal [2]. Di antara berbagai algoritma, K-Means sering menjadi pilihan utama karena terbukti efisien dalam mengolah data berukuran besar dan mudah diimplementasikan [2], [15]. Beberapa penelitian di Indonesia juga telah menunjukkan keberhasilan K-Means untuk melakukan pemetaan prestasi siswa, penentuan kelas unggulan, serta perancangan intervensi pendidikan yang lebih terarah [3]–[5], [14]. Kualitas dari klaster yang dihasilkan dapat dievaluasi secara kuantitatif menggunakan metrik seperti Davies–Bouldin Index (DBI), di mana nilai DBI yang lebih rendah mengindikasikan partisi klaster yang lebih baik dan lebih terpisah [6]. Perkembangan terkini bahkan menunjukkan adanya upaya integrasi klusterisasi dengan sistem rekomendasi atau model prediksi untuk mendukung pembelajaran yang lebih personal [7], [8].

Penelitian ini bertujuan untuk menerapkan algoritma K-Means untuk mengelompokkan prestasi siswa kelas VIII di SMP Negeri 92 Jakarta berdasarkan 30 atribut nilai. Dengan menentukan jumlah klaster optimal menggunakan DBI, penelitian ini diharapkan dapat menyajikan implikasi praktis bagi sekolah dalam merancang program pengayaan dan remedial yang berbasis data.

## II. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik *data mining* untuk klusterisasi yang sistematis [2]. Alur penelitian dirancang secara sistematis yang dimulai dari pengumpulan dan deskripsi data, diikuti oleh tahap pra-pemrosesan data, penentuan jumlah klaster ( $k$ ) optimal, proses klusterisasi menggunakan algoritma K-Means, dan diakhiri dengan evaluasi hasil klaster untuk memastikan validitas pengelompokan. Seluruh proses analisis diimplementasikan menggunakan perangkat lunak RapidMiner.

### A. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder berupa nilai akademik dari 250 siswa kelas VIII SMP Negeri 92 Jakarta untuk Tahun Ajaran 2024/2025. Dataset terdiri dari 30 atribut numerik yang merepresentasikan nilai dari 10 mata pelajaran utama (Bahasa Indonesia, Matematika, IPA, IPS, Bahasa Inggris, Pendidikan Agama & Budi Pekerti, Pendidikan Pancasila, PJOK, TIK, serta Seni Budaya & Prakarya). Setiap mata pelajaran memiliki tiga jenis nilai asesmen: Tugas Harian (TH), Asesmen Tengah Semester (ATS), dan Asesmen Akhir Semester (AAS). Untuk menjaga etika penelitian, seluruh data identitas siswa telah dianonimkan.

### B. Tahapan Analisis Data

Proses analisis data dilakukan secara bertahap untuk mengolah informasi secara efektif. Tahapan-tahapan inti dalam analisis ini mencakup beberapa langkah kunci sebagai berikut:

1. *Pra-pemrosesan Data*: Sebelum dianalisis, data mentah melewati tahap pra-pemrosesan untuk memastikan kualitas dan kesetaraan skala. Proses ini mencakup dua langkah utama: (1) Validasi Data, di mana data diperiksa untuk menangani nilai yang hilang (missing values) atau inkonsistensi, meskipun pada dataset ini data yang diterima sudah lengkap. (2) Normalisasi Data. Atribut non-numerik (seperti ID Siswa) dikecualikan dari perhitungan, dan untuk menangani perbedaan rentang nilai antar asesmen, dilakukan normalisasi data menggunakan metode Min-Max Scaling. Metode ini mentransformasikan setiap nilai ke dalam rentang (0, 1) [2], sesuai dengan rumus persamaan (1) sebagai berikut:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

di mana  $x'$  adalah nilai setelah normalisasi,  $x$  adalah nilai asli, sedangkan  $x_{min}$  dan  $x_{max}$  adalah nilai minimum dan maksimum pada atribut tersebut.

2. *Proses Klusterisasi*: Pengelompokan siswa dilakukan menggunakan algoritma K-Means, yang mempartisi data ke dalam  $k$  klaster berdasarkan kedekatan jarak Euclidean setiap titik data ke

pusat kluster (*centroid*). Penentuan jumlah kluster optimal (nilai  $k$ ) dilakukan dengan menguji beberapa nilai  $k$  (misalnya,  $k=2, 3, 4, 5$ ) dan memilih nilai  $k$  yang menghasilkan skor Davies-Bouldin Index (DBI) terendah [6].

3. *Evaluasi Hasil Kluster*: Kualitas hasil pengelompokan dievaluasi menggunakan dua metrik komplementer [2], [6]. Pertama, Davies-Bouldin Index (DBI) digunakan untuk mengukur rasio antara kekompakan internal kluster (intra-kluster) dan tingkat pemisahan antar kluster (inter-kluster). Nilai DBI yang lebih rendah mengindikasikan partisi kluster yang lebih baik. Kedua, **Average** Within-Cluster Distance (AWCD) digunakan untuk mengukur tingkat kohesi atau kepadatan setiap kluster. Nilai AWCD yang rendah menunjukkan bahwa anggota dalam satu kluster memiliki tingkat kesamaan yang tinggi.

### C. Perangkat yang Digunakan

Penelitian ini memanfaatkan dua perangkat lunak utama. Microsoft Excel digunakan untuk pengumpulan, penyimpanan, dan validasi awal data nilai siswa. Sementara itu, seluruh proses inti *data mining* mulai dari pra-pemrosesan (normalisasi), implementasi algoritma K-Means, hingga evaluasi performa kluster menggunakan DBI dan AWCD dilakukan dengan bantuan perangkat lunak RapidMiner Studio.

## III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari proses klusterisasi menggunakan algoritma K-Means, mulai dari penentuan jumlah kluster optimal hingga analisis karakteristik setiap kluster yang terbentuk.

### A. Penentuan Jumlah Kluster Optimal

Untuk menentukan jumlah kluster ( $k$ ) yang paling sesuai untuk dataset, dilakukan pengujian dengan rentang  $k$  dari 2 hingga 5. Kualitas hasil klusterisasi untuk setiap partisi diukur menggunakan *Davies-Bouldin Index* (DBI), hasil yang lebih optimal ditandai dengan perolehan skor DBI yang cenderung lebih rendah. Hasil evaluasi menunjukkan bahwa  $k=2$  menghasilkan skor DBI terendah (-4.625), sehingga ditetapkan sebagai jumlah kluster yang paling optimal untuk penelitian ini.

### B. Karakteristik Hasil Klusterisasi

Dengan menerapkan K-Means untuk  $k=2$ , sebanyak 250 siswa berhasil dikelompokkan ke dalam dua kluster:

- 1) Cluster\_0: Terdiri dari 117 siswa.
- 2) Cluster\_1: Terdiri dari 133 siswa.

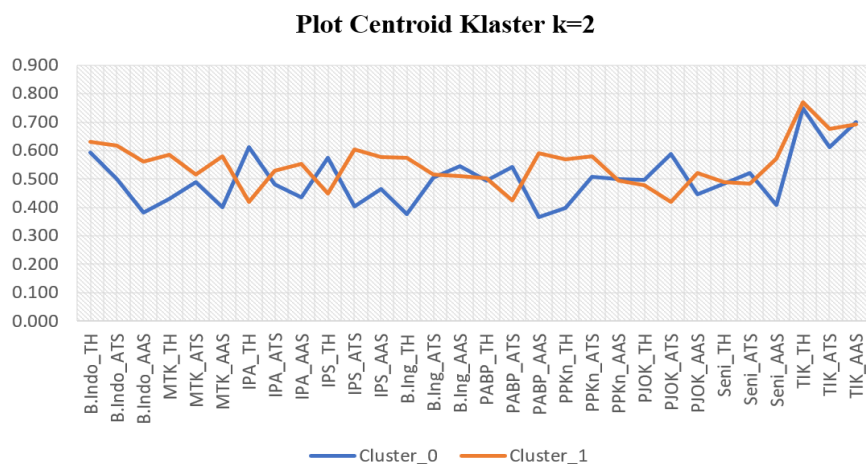
Untuk memahami profil dari masing-masing kluster, nilai *centroid* (rata-rata nilai untuk setiap atribut) dianalisis seperti yang disajikan pada Tabel 1.

Tabel 1. Nilai Centroid untuk Setiap Kluster ( $k=2$ )

Attribute	Cluster_0	Cluster_1
B.Indo_TH	0,594	0,631
B.Indo_ATS	0,500	0,616
B.Indo_AAS	0,382	0,561
MTK_TH	0,429	0,584
MTK_ATS	0,489	0,515
MTK_AAS	0,401	0,581
IPA_TH	0,611	0,419
IPA_ATS	0,480	0,529
IPA_AAS	0,434	0,554

IPS_TH	0,575	0,447
IPS_ATS	0,403	0,605
IPS_AAS	0,466	0,576
B.Ing_TH	0,375	0,574
B.Ing_ATS	0,504	0,515
B.Ing_AAS	0,545	0,511
PABP_TH	0,493	0,502
PABP_ATS	0,541	0,425
PABP_AAS	0,365	0,589
PPKn_TH	0,399	0,568
PPKn_ATS	0,506	0,581
PPKn_AAS	0,500	0,493
PJOK_TH	0,496	0,479
PJOK_ATS	0,587	0,420
PJOK_AAS	0,446	0,521
Seni_TH	0,485	0,489
Seni_ATS	0,520	0,484
Seni_AAS	0,407	0,572
TIK_TH	0,749	0,769
TIK_ATS	0,612	0,675
TIK_AAS	0,701	0,693

Visualisasi hasil klusterisasi juga menunjukkan adanya pemisahan yang jelas antara kedua kelompok, memperkuat temuan bahwa K-Means berhasil membentuk dua grup yang berbeda secara visual.



**Gambar 1.** Grafik Visualisasi Centroid Kluster

### C. Pembahasan Hasil

Analisis terhadap nilai centroid pada Tabel 1 menunjukkan adanya perbedaan pola prestasi yang signifikan antara kedua klaster. Cluster\_1 secara konsisten menunjukkan nilai centroid yang lebih unggul pada sebagian besar atribut akademik, terutama pada mata pelajaran inti seperti Bahasa Indonesia, Matematika, dan IPS. Hal ini mengindikasikan bahwa Cluster\_1 dapat diinterpretasikan sebagai kelompok siswa berprestasi tinggi.

Sebaliknya, Cluster\_0 memiliki nilai centroid yang cenderung lebih rendah pada mayoritas atribut, yang secara umum menunjukkan performa akademik di tingkat sedang hingga rendah. Oleh karena itu, Cluster\_0 diidentifikasi sebagai kelompok siswa yang memerlukan perhatian dan pembinaan lebih lanjut. Meskipun demikian, Cluster\_0 justru unggul pada beberapa atribut spesifik seperti IPA\_TH dan PJOK\_ATS. Hal ini mengindikasikan bahwa kelompok siswa Cluster\_0 cenderung merespon baik pada metode pembelajaran berbasis Tugas Harian yang bersifat formatif, sementara nilai Asesmen Akhir Semester (AAS) yang bersifat sumatif mereka cenderung lebih rendah.

Temuan ini tidak hanya sejalan, namun juga memperkuat hasil penelitian sebelumnya [3]–[5], [9]–[13] yang menunjukkan efektivitas K-Means dalam pemetaan prestasi. Perbedaan penelitian ini terletak pada penggunaan 30 atribut multi-asesmen yang komprehensif, memberikan gambaran yang lebih granular (rinci) dibandingkan studi yang hanya berfokus pada nilai akhir. Pemisahan biner ( $k=2$ ) ini mengkonfirmasi temuan oleh [4] dan [5] bahwa partisi menjadi kelompok prestasi tinggi dan rendah seringkali merupakan langkah intervensi paling praktis bagi sekolah.

Untuk lebih memperkuat validitas temuan, sebuah analisis komparatif diimplementasikan dengan mengevaluasi hasil dari metode normalisasi Min-Max dan Z-Score. Analisis ini menunjukkan bahwa kedua metode secara konsisten menunjuk  $k=2$  sebagai jumlah klaster optimal, yang membuktikan kekokohan (*robustness*) model penelitian. Hasil perbandingan skor DBI antara kedua metode normalisasi disajikan pada Tabel 2.

Tabel 2. Perbandingan Nilai DBI antara Normalisasi Min-Max dan Z-Score

Jumlah Klaster (k)	Nilai DBI Min-Max	Nilai DBI Z-Score
2	-4.625	-4.745
3	-4.066	-4.184
4	-3.838	-3.975
5	-3.608	-3.739

Hasil pengelompokan yang objektif ini memberikan landasan berbasis data bagi pihak sekolah. Berdasarkan umpan balik yang diterima, pihak SMP Negeri 92 Jakarta mengonfirmasi bahwa temuan ini sangat relevan dan akan digunakan untuk merancang strategi pembinaan yang lebih tepat sasaran yaitu program pengayaan materi menysasar siswa Cluster\_1 dan program remedial dikhususkan bagi siswa Cluster\_0. Dengan demikian, penelitian ini berhasil menjembatani analisis data akademik dengan pengambilan keputusan strategis di lingkungan sekolah.

## IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma klasterisasi K-Means untuk mengelompokkan prestasi siswa di SMP Negeri 92 Jakarta secara objektif berdasarkan 30 atribut nilai akademik. Hasil analisis menunjukkan bahwa jumlah klaster optimal adalah dua ( $k=2$ ), yang secara efektif membagi siswa ke dalam dua kelompok dengan karakteristik berbeda: Cluster\_1 sebagai kelompok berprestasi tinggi dan Cluster\_0 sebagai kelompok dengan prestasi sedang hingga rendah. Pengelompokan ini memberikan landasan berbasis data yang kuat bagi pihak sekolah untuk merancang strategi pembinaan yang lebih personal dan tepat sasaran, seperti program pengayaan bagi kelompok berprestasi tinggi dan program remedial bagi kelompok yang membutuhkan perhatian lebih.

Untuk pengembangan di masa depan, disarankan agar penelitian selanjutnya dapat menyertakan **variabel non-akademik** (seperti absensi atau keaktifan) untuk mendapatkan profil siswa yang lebih

---

holistik. Selain itu, melakukan studi komparatif dengan **algoritma klasterisasi lain** (misalnya, Hierarchical Clustering atau DBSCAN) dapat memberikan wawasan lebih lanjut mengenai metode yang paling efektif untuk konteks data prestasi siswa.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak SMP Negeri 92 Jakarta yang telah memberikan dukungan data dan masukan berharga selama proses penelitian.

#### DAFTAR PUSTAKA

- [1] Y. Elda, S. Defit, Y. Yunus, dan R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan Metode K-Means," *Jurnal Pendidikan dan Informatika*, vol. 8, no. 2, hal. 97–105, 2021.
- [2] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, ed. ke-3. Waltham, MA, USA: Morgan Kaufmann, 2012.
- [3] J. Hutagalung, Y. H. Syahputra, dan Z. P. Tanjung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, vol. 9, no. 1, hal. 606–620, 2022.
- [4] H. Winata, S. N. Br Sembiring, dan S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *Jurnal Pendidikan dan Teknologi*, vol. 12, no. 1, hal. 34–44, 2022.
- [5] A. S. T. Suprpti, G. Dwilestari, dan I. Ali, "Pengelompokan Hasil Belajar Siswa dengan Metode Clustering K-Means," *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, vol. 4, no. 2, hal. 60–72, 2022.
- [6] D. L. Davies dan D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, hal. 224–227, Apr. 1979.
- [7] M. Alshaigi, G. Qaterji, M. Nazzal, dan S. Haj Ismail, "Leveraging Clustering and Recommendations for Personalized Learning," *Education Sciences*, vol. 15, no. 8, 2025.
- [8] M. N. Sutoyo dan A. Paliling, "Optimizing Student Performance Prediction via K-Means and k-NN Integration," *Int. J. Educ. Manage. Eng.*, vol. 15, no. 4, hal. 12–22, Agu. 2025.
- [9] A. R. Antoni, D. Sartika, dan D. A. Trianggana, "Penerapan Data Mining dalam Pengelompokan Data Nilai Siswa SMA Negeri 7 Kota Bengkulu Menggunakan Metode K-Means Clustering," *Jurnal Data Mining dan Pendidikan*, vol. 12, no. 3, hal. 122–133, 2025.
- [10] I. Fitriani dan D. Astuti, "Data mining untuk analisis nilai ujian dan tingkat kehadiran siswa: Pendekatan K-Means," *Jurnal Ilmiah Teknologi dan Pendidikan*, vol. 17, no. 3, hal. 47–55, 2021.
- [11] Herlina dan Z. Ernaningsih, "Implementasi K-Means Clustering untuk Analisis Tingkat Pemahaman Computational Thinking Siswa Sekolah Dasar," *Jurnal Teknologi Pendidikan*, vol. 16, no. 2, hal. 88–94, 2023.
- [12] M. Qusyairi, Z. Hidayatullah, dan A. Sandi, "Penerapan K-Means Clustering dalam Pengelompokan Prestasi Siswa dengan Optimasi Metode Elbow," *Jurnal Teknologi Pendidikan*, vol. 45, no. 3, hal. 123–135, 2024.
- [13] J. S. Sibatuara, "Klasterisasi Tingkat Pemahaman Siswa dalam Sistem Pembelajaran Online dengan Metode K-Means Clustering," *Jurnal Pembelajaran Digital*, vol. 4, no. 1, hal. 56–65, 2022.
- [14] R. Sari dan S. Wahyuni, "Pemanfaatan data mining untuk analisis prestasi siswa menggunakan algoritma K-Means," *Jurnal Pendidikan dan Teknologi*, vol. 22, no. 4, hal. 134–141, 2020.
- [15] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, ed. ke-4. Cambridge, MA, USA: Elsevier, 2017.