Sains Data Jurnal Studi Matematika dan Teknologi
ISSN 2986-903X

v4i1(2026) pp. 34-41
https://pub.nuris.ac.id/sainsdata

Analisis Model Algoritma K-Means Clustering untuk Pengelompokan Prestasi berdasarkan Nilai Pengetahuan Siswa (Studi Kasus : SMP Negeri 2 Mekar Baru)

Fajar Widianto[1], Alfaiko Janita[2]

[1] Universitas Pamulang
Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
[2]Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) Banten
Jl Raya. Ciruas - Walantaka KM. 1 Kota Serang, Banten
Email : [1]fajarwidianto95@gmail.com [2]alfaikojanita@gmail.com

DOI : https://doi.org/10.52620/sainsdata.v4i1.347

Abstrak

Perkembangan teknologi memberikan dampak nyata hari ini. Khususnya dunia pendidikan dalam melakukan penilaian siswa, mengevaluasi keberhasilan proses pembelajaran di sekolah. Namun, pengelompokan prestasi berdasarkan nilai pengetahuan seringkali masih dilakukan secara manual sehingga kurang efektif dan kurang mampu memberikan gambaran pola akademik secara menyeluruh. Tujuan dari penelitian ini untuk menganalisis penerapan model algoritma K-Means clustering dalam mengelompokkan prestasi siswa berdasarkan nilai pengetahuan di SMP Negeri 2 Mekar Baru. Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan teknik data mining. Data nilai pengetahuan siswa diolah melalui tahapan preprocessing, penentuan jumlah cluster (k), proses iterasi K-Means, serta mengevaluasi hasil pengelompokan. Hasil yang didapat menunjukkan bahwa algoritma K-Means mampu melakukan pengelompokan siswa ke dalam beberapa kategori prestasi, seperti tinggi, sedang maupun rendah, secara objektif berdasarkan kedekatan nilai centroid. Kesimpulannya, penerapan K-Means efektif membantu pihak sekolah dalam memetakan prestasi siswa sebagai dasar pengambilan keputusan adapun kebaruan dalam penelitian ini terletak pada penerapan model clustering sebagai sistem pendukung evaluasi prestasi berbasis data di tingkat SMP.
Kata Kunci: K-Means Clustering, Data Mining, Prestasi Siswa, Nilai Pengetahuan

Abstract
Student Technological developments have a real impact today. Especially in the world of education in conducting student assessments, evaluating the success of the learning process in schools. However, grouping achievements based on knowledge scores is often still done manually so it is less effective and less able to provide a comprehensive picture of academic patterns. The purpose of this study is to analyze the application of the K-Means clustering algorithm model in grouping student achievements based on knowledge scores at SMP Negeri 2 Mekar Baru. The research method used is a quantitative approach with data mining techniques. Student knowledge score data is processed through preprocessing stages, determining the number of clusters (k), the K-Means iteration process, and evaluating the clustering results. The results obtained show that the K-Means algorithm is able to group students into several achievement categories, such as high, medium, and low, objectively based on the proximity of the centroid value. In conclusion, the application of K-Means effectively helps schools in mapping student achievement as a basis for decision making. The novelty in this study lies in the application of the clustering model as a data-based achievement evaluation support system at the junior high school level.
Keywords*: K-Means Clustering, Data Mining, Student Achievement, Knowledge Score*

v4i1(2026) pp. 34-41
https://pub.nuris.ac.id/sainsdata

PENDAHULUAN

Perkembangnya transformasi digital dalam bidang pendidikan, maka muncul kebutuhan untuk menggunakan data akademik sebagai landasan pengambilan keputusan berbasis bukti (evidence-based decision making). Dalam konteks di sekolah, data nilai pengetahuan siswa merupakan sumber informasi yang sangat berharga dan dapat dianalisis untuk memahami pola prestasi dan distribusi kemampuan akademik. Namun, dalam kenyataannya, pengolahan nilai masih cenderung bersifat deskriptif dan administratif, sehingga belum sepenuhnya dimanfaatkan untuk analisis pola berbasis data. Kondisi ini juga terlihat dalam pengelolaan data akademik di SMP Negeri 2 Mekar Baru, dimana pengelompokan prestasi siswa belum menggunakan pendekatan komputasional yang sistematis.

Dengan berkembangnya teknik data mining, maka metode clustering menjadi salah satu pendekatan yang efektif dalam mengelompokkan data tanpa label (unsupervised learning). Penelitian terbaru menunjukkan bahwa algoritma K-Means masih menjadi metode yang banyak digunakan dalam analisis data pendidikan karena efisiensi komputasi dan kemampuannya mengelompokkan data numerik secara objektif (Zhang & Wu, 2023; Kumar et al., 2024). Dalam konteks pendidikan menengah, maka penerapan clustering telah terbukti membantu sekolah dalam mengidentifikasi kelompok siswa berdasarkan capaian akademik untuk mendukung intervensi pembelajaran yang lebih tepat sasaran (Rahman & Pratama, 2023)

Namun demikian, sebagian besar penelitian lebih terkonsentrasi pada tingkat pendidikan tingkat tinggi atau menerapkan analisis berbasis platform *e-learning*. Penelitian ini menerapkan K-Means pada pengelompokan prestasi siswa berdasarkan nilai pengetahuan siswa tingkat SMP di sekolah daerah masih tergolong terbatas. Penelitian ini memiliki urgensi karena dapat mengisi kekosongan tersebut dengan menerapkan model analisis clustering sebagai sistem pendukung dalam evaluasi akademik di SMP Negeri 2 Mekar Baru.

Penelitian ini bertujuan menganalisis penerapan algoritma K-Means Clustering dalam mengelompokkan prestasi siswa di sekolah berdasarkan nilai pengetahuan siswa, menganalisis pola prestasi yang terbentuk, serta memberikan rekomendasi strategis bagi sekolah dalam menyusun kebijakan pembinaan akademik. Kebaruan ilmiah pada penelitian ini terletak pada implementasi model clustering sebagai pendekatan analitik dalam bentuk praktis di tingkat SMP, dengan fokus pada optimalisasi pengelompokan siswa.

TINJAUAN PUSTAKA

Prestasi Siswa

Prestasi siswa dalam konteks penelitian berbasis data ini tidak hanya dimaknai sebagai hasil belajar secara umum, tetapi lebih spesifik sebagai capaian akademik yang terukur melalui indikator kuantitatif, seperti nilai pengetahuan pada mata pelajaran tertentu. Menurut Rahman dan Pratama (2023), hasil belajar siswa dapat diartikan sebagai angka-angka yang menggambarkan tingkat ketercapaian kompetensi kognitif yang telah diperoleh melalui proses evaluasi pembelajaran yang dapat dianalisis .

Sedangkan Zhang dan Wu (2023) mengatakan bahwa dalam educational data mining, nilai akademik siswa dipandang sebagai atribut numerik yang dapat digunakan untuk mengidentifikasi pola performa belajar melalui teknik clustering. Dalam konteks ini, prestasi siswa tidak hanya dilihat dari rata-rata nilai, tetapi dari pola distribusi dan kemiripan karakteristik capaian akademik antar siswa.

Jadi, definisi dari prestasi siswa dalam penelitian ini adalah capaian kompetensi kognitif siswa yang dinyatakan melalui nilai pengetahuan dan dianalisis menggunakan model algoritma K-Means Clustering

v4i1(2026) pp. 34-41
https://pub.nuris.ac.id/sainsdata

dalam mengidentifikasi pola untuk mengklasterisasikan prestasi siswa secara objektif dan berbasis data. Definisi ini juga menegaskan hubungan antara konsep prestasi belajar.

Algoritma K-Means Clustering

Algoritma K-Means Clastering adalah suatu metode yang termasuk dalam data mining yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kesamaan karakteristiknya atau kedekatan. (Yuni Franata Sinurat, Masrizal, 2024)

Sedangkan Metode K-Means Claustering merupakan metode yang banyak digunakan untuk mengklasterisasi data Dimana dibagi ke dalam sejumlah klaster yang telah ditetapkan sebelumnya. Setiap titik data ditetapkan ke klaster dengan centroid (titik tengah) terdekat. Centroid diupdate hingga tujuannya tercapai.

Data Mining

Konsep data mining dapat diartikan sebuah proses pencarian informasi dan pola yang bermanfaat dari suatu data dalam jumlah besar. Proses data mining yaitu dilakukan dari pengumpulan data, ekstraksi, analisa data, dan juga statistik data. Data mining juga merupakan metode yang banyak digunakan dalam mencari pola penggambaran dan kecenderungan pada sejumlah data. Deskripsi dari pola dan kecenderungan data.(Wahidin & Syukrilla, 2023)

METODE

Metode penelitian ini melibatkan beberapa proses tahapan dan pendekatan kuantitatif dalam menganalisis data, K-means *clustering* sala satu algoritma yang digunakan dalam data mining. Metode ini bertujuan untuk mengeklasterisasi data nilai siswa SMP berdasarkan tingkat kemiripan karakteristik akademik tanpa label kelas sebelumnya (*unsupervised learning*).

Jenis dan Desain Penelitian

Jenis penelitian ini yaitu penelitian kuantitatif deskriptif-analitik. Penelitian berfokus pada analisis pola pengelompokan prestasi siswa berdasarkan data numerik berupa nilai pengetahuan. Desain penelitian mengikuti tahapan proses data mining yang meliputi: pengumpulan data, preprocessing, pemodelan (*modeling*), evaluasi, dan interpretasi hasil. Sumber nya di dapat data sekunder, yaitu data nilai pengetahuan siswa yang diperoleh dari pihak sekolah.

Tahap Analisis data

1. Data Preprocessing
   Tahap ini meliputi:
   - Proses seleksi data nilai yang relevan
   - Pembersihan data (*data cleaning*) dari nilai kosong atau duplikasi
   - Normalisasi data (jika diperlukan) untuk menyamakan skala nilai

2. Penentuan Jumlah Cluster (k)
   Jumlah cluster ditentukan berdasarkan kebutuhan analisis, misalnya 3 cluster (tinggi, sedang, rendah). Penentuan nilai k dapat menggunakan metode *Elbow Method* untuk memperoleh jumlah cluster optimal.

3. Proses Algoritma K-Means
   Adapun Langkah-langkah algoritma K-Means adalah sebagai berikut:

v4i1(2026) pp. 34-41
https://pub.nuris.ac.id/sainsdata

- Menentukan jumlah cluster (k).
- Menetapkan sejumlah centroid awal secara acak.
- Memperhitungkan sejumlah jarak masing-masing centroid serta merumuskan *euclidean distance* :

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

- Mengelompokkan sejumlah data dalam ke cluster dengan jarak terdekat.
- Menghitung ulang centroid kedalam rata-rata anggota cluster :

$$C_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i$$

- Mengulangi tahapan hingga centroid tidak berubah (konvergen)

4. Evaluasi Model
   Evaluasi dilakukan untuk melihat stabilitas cluster dan interpretasi karakteristik masing-masing kelompok berdasarkan nilai rata-rata tiap cluster.

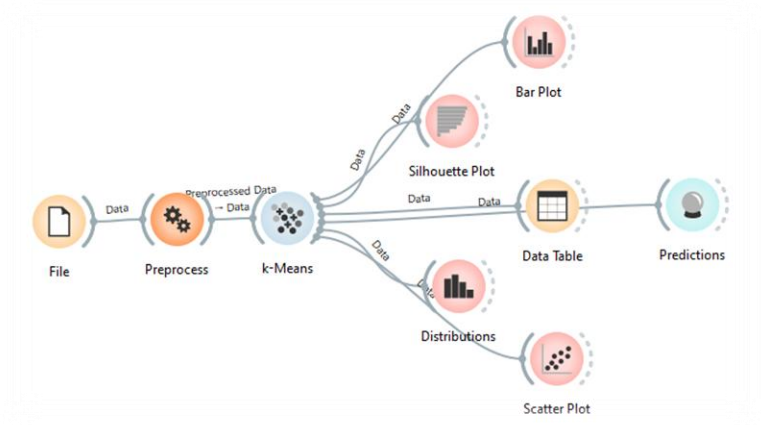
Interpretasi Hasil

Hasil clustering diinterpretasikan menjadi kategori prestasi siswa (tinggi, sedang, rendah). Interpretasi ini digunakan sebagai dasar rekomendasi akademik bagi sekolah, seperti program remedial atau pengayaan.

HASIL DAN PEMBAHASAN

Implementasi K-Means Orange Data Mining

Alur ini dirancang untuk melakukan pengelompokan data (clustering) menggunakan algoritma k-Means. Berikut adalah penjelasan langkah demi langkah dari setiap komponen (widget) yang ada :

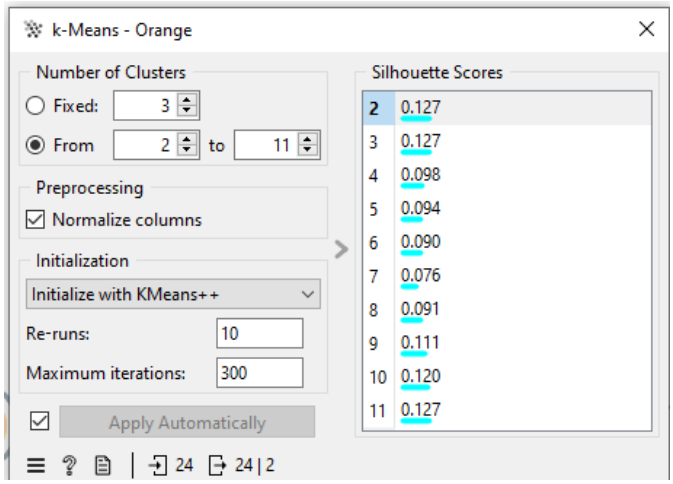


Gambar 1. Alur kerja (workflow)

1. File : merupakan titik awal di mana data mentah (seperti file .csv atau .xlsx) dimasukkan ke dalam sistem.
2. Preprocess : Data mentah jarang sekali langsung siap pakai. Widget ini biasanya digunakan untuk membersihkan data, seperti menangani nilai yang hilang (missing values), melakukan normalisasi (penskalaan angka agar setara), atau memilih fitur tertentu agar hasil clustering lebih akurat.
3. K-Means: merupakan "otak" dari alur ini. Algoritma ini memilah data nilai ke sejumlah klaster (k) berdasarkan kesamaan karakteristiknya. Data yang berada di dalam satu kelompok akan memiliki kesamaan yang tinggi satu sama lain.

v4i1(2026) pp. 34-41
https://pub.nuris.ac.id/sainsdata

4. Bar Plot & Distributions: Digunakan untuk melihat perbandingan frekuensi atau distribusi data di setiap cluster yang terbentuk.
5. Silhouette Plot: Ini adalah alat evaluasi yang sangat penting. Widget ini menunjukkan seberapa baik sebuah data "pas" berada di clusternya dibandingkan dengan cluster lain. Semakin tinggi nilainya, semakin baik pengelompokannya.
6. Scatter Plot: Memvisualisasikan data dalam bentuk titik-titik pada koordinat 2D untuk melihat secara langsung bagaimana pengelompokan terjadi secara spasial
7. Data Table: Menampilkan data dalam bentuk tabel baris dan kolom, biasanya sekarang sudah menyertakan label "Cluster" yang dihasilkan oleh k-Means.
8. Predictions: Berdasarkan tabel data tersebut, widget ini digunakan untuk melihat hasil klasifikasi atau memprediksi kategori data baru berdasarkan model clustering yang sudah dibuat.

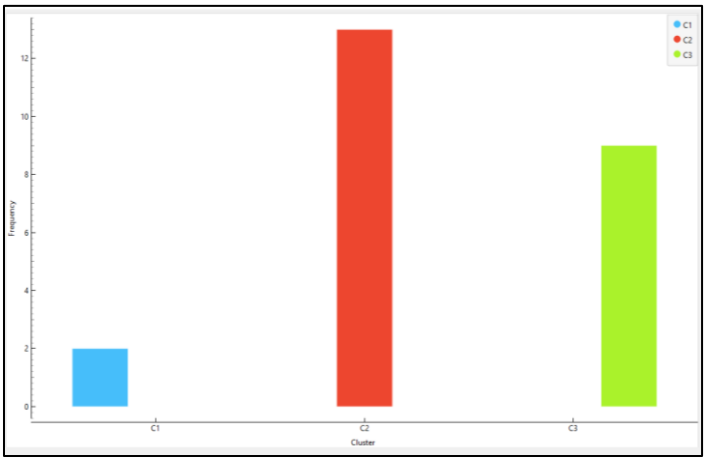


Gambar 2. K-Means Orange

Berdasarkan tampilan parameter K-Means pada software Orange, jumlah cluster diuji dari K = 2 hingga K = 11 dengan evaluasi menggunakan Silhouette Score. Kemudian mengelompokkan data ke dalam kategori tinggi, sedang, dan rendah. Proses clustering juga telah menggunakan normalisasi data serta metode inisialisasi K-Means++ untuk meningkatkan stabilitas hasil.

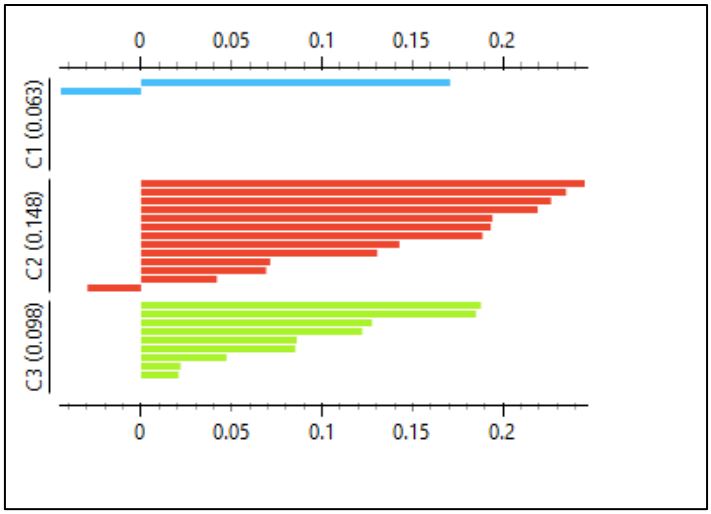
Analisis Sebaran Data

Berdasarkan hasil proses clustering di orange, diperoleh distribusi siswa sebagai berikut : Grafik Distributions menunjukkan bagaimana 24 data siswa tersebut terbagi ke dalam 3 kelompok (K=3):



Gambar 3. Grafik Distribusi

- Cluster C2 (Merah) : Merupakan kelompok mayoritas dengan jumlah sekitar 13 siswa.
- Cluster C3 (Hijau) : Berisi sekitar 9 siswa.
- Cluster C1 (Biru) : Merupakan kelompok terkecil yang hanya berisi 2 siswa



Gambar 4. Silhouette Plot

Interpretasi :

1. Cluster C2 (0,148) memiliki nilai silhouette tertinggi, sehingga dapat dikatakan sebagai cluster dengan kualitas pemisahan terbaik dibandingkan cluster lainnya. data pada cluster ini relatif lebih kompak dan cukup terpisah dari cluster lain.
2. Cluster C3 (0,098) memiliki kualitas sedang. Beberapa data sudah cukup baik dalam pengelompokan, namun masih ada kemungkinan kedekatan dengan cluster lain.
3. Cluster C1 (0,063) memiliki nilai paling rendah, menunjukkan bahwa pemisahan cluster ini kurang kuat dan kemungkinan terdapat data yang cukup dekat dengan cluster lain.

Maka, seluruh nilai silhouette masih berada di bawah 0,5 dan bahkan mendekati 0, sehingga dapat disimpulkan bahwa struktur cluster dalam data belum terlalu kuat atau masih terdapat tumpang tindih antar kelompok.

Predictions dan visualisasi

Restore Origin

NAMA	Cluster	Silhouette	N0	MTK	BINDO	PAI	IPA	IPS	BING	SBK	PJOK	UTS	UAS
ARIF	C2	0.533744	1	90	85	85	85	85	80	85	85	70	60
ASMARIYAH	C3	0.514935	2	90	85	90	90	85	90	90	85	85	60
DEDEN	C2	0.565262	3	80	80	85	80	90	80	80	80	70	60
FAISAL FARIZI	C2	0.516631	4	85	80	90	85	90	80	85	85	80	60
HAMBALI	C3	0.528033	5	85	80	90	80	85	80	90	85	85	60
KHILYATUNNU...	C2	0.573596	6	80	80	90	90	90	80	85	80	80	55
M RIZKI JANUAR	C2	0.557332	7	85	80	90	80	80	80	80	80	75	55
MAESAROH	C3	0.53321	8	90	90	85	80	95	80	85	85	80	60
MAHMOOD M...	C2	0.524718	9	90	90	85	85	85	80	85	85	75	50
MARINAH	C2	0.576298	10	85	80	85	90	80	85	85	80	75	50
MOHAMAD SU...	C2	0.547197	11	80	80	80	90	80	80	75	80	80	50
MUHAMMAD ...	C2	0.568172	12	75	80	90	85	90	80	85	85	80	50
MUSLIK	C1	0.53543	13	75	80	70	90	70	80	75	70	70	65
NASRUL	C3	0.558853	14	90	80	85	80	85	80	90	85	85	60
RIDHO SAPUTRA	C2	0.580018	15	80	80	90	90	90	80	80	80	70	60
RIYAN	C1	0.462266	16	85	80	70	85	85	80	90	70	80	55
ROMLAH	C2	0.567185	17	70	80	80	85	95	80	85	80	70	55
SAHRONI	C2	0.540957	18	90	80	85	85	90	80	85	90	75	50
SHADAH	C3	0.542999	19	75	90	80	80	80	80	90	90	80	60
SITI NURHALIZA	C3	0.540191	20	80	90	90	85	80	80	85	80	85	60
SOVI ARYANTI	C3	0.508836	21	90	80	80	85	90	80	85	90	80	60
SRI AMILAWATI	C3	0.553435	22	90	90	80	80	80	80	80	80	85	60
SYAHRIFA DEWI	C3	0.534095	23	90	90	90	85	90	90	85	80	80	60
TUHRI	C2	0.498858	24	80	85	85	85	80	90	85	80	75	55

Gambar 5. Predictions dan visualisasi

Dari tabel data siswa, kita dapat melihat pola yang membedakan ketiga kelompok ini:

- Cluster C2 (Merah - Kelompok Terbesar): Kelompok ini umumnya memiliki nilai yang sangat stabil dan tinggi di mata pelajaran eksakta dan bahasa. Sebagai contoh, Arif dan Deden memiliki nilai MTK (90 & 80), IPA (85 & 80), serta BING (80). Kelompok ini mewakili siswa dengan performa akademik yang konsisten di atas rata-rata.
- Cluster C3 (Hijau - Kelompok Menengah): Siswa di kelompok ini, seperti Asmariyah dan Hambali, menonjol di mata pelajaran seperti PAI (90) dan SBK (90). Nilai mereka cukup bersaing dengan C2, namun sering kali memiliki sedikit penurunan di nilai UAS (rata-rata di angka 60) dibandingkan nilai harian mereka.
- Cluster C1 (Biru - Kelompok Unik): Kelompok ini hanya berisi 2 siswa, yaitu Muslik dan Riyan. Perbedaan mencolok mereka terlihat pada nilai PAI (70) dan SBK (75/70) yang lebih rendah secara signifikan dibanding kelompok lain, serta nilai IPS (70) yang berada di bawah rata-rata kelas. Inilah yang menyebabkan mereka terpisah menjadi kelompok kecil sendiri.

Secara keseluruhan, pembeda utama di siswa adalah konsistensi nilai mata pelajaran pendukung (PAI, SBK, IPS). Kelompok C2 adalah kelompok unggul di hampir semua bidang, C3 unggul di bidang keagamaan dan seni, sementara C1 adalah kelompok yang membutuhkan perhatian khusus karena memiliki nilai yang paling rendah di beberapa mata pelajaran kunci tersebut.

KESIMPULAN

Kesimpulan dari hasil penelitian ini yang menggunakan perangkat lunak aplikasi *Orange* Data Mining dengan alur kerja yaitu dimulai dari pengambilan data mentah (*File*), proses pembersihan (*Preprocess*), hingga tahap inti yaitu algoritma k-Means untuk mengelompokkan siswa berdasarkan kemiripan nilai pengetahuannya. Berdasarkan pengujian otomatis terhadap 2 hingga 11 kelompok, ditemukan bahwa jumlah pengelompokan yang paling optimal secara statistik adalah 3 cluster dengan skor *Silhouette* sebesar 0,127. Hasil pemrosesan terhadap 24 data siswa menunjukkan sebaran prestasi yang tidak merata. Cluster C2 (Merah) menjadi kelompok prestasi dominan yang mencakup 13 siswa dengan karakteristik nilai akademik yang cenderung stabil di atas rata-rata. Di sisi lain, Cluster C3 (Hijau) terdiri dari 9 siswa dengan performa menengah, sementara Cluster C1 (Biru) hanya terdiri dari 2 siswa. Dengan demikian, model ini memberikan landasan bagi SMP Negeri 2 Mekar Baru untuk melakukan pemetaan prestasi guna memberikan bimbingan belajar yang lebih terarah sesuai dengan kekuatan dan kelemahan masing-masing kelompok siswa.

DAFTAR PUSTAKA

Adzra, S. N. N., Hasan, F. N., & Kuntoro, A. Y. (2024). *Penerapan data mining dalam penilaian kinerja akademik siswa/i SMP YPI Pulogadung dengan metode K-Means clustering*. *Jurnal Ilmiah Informatika*, 13(02), Article 10396.

Alalawi, S. J. S., Mohd Shaharanee, I. N., & Jamil, J. (2023). *Clustering student performance data using K-Means algorithms*. *Journal of Computational Innovation and Analytics*, 2(1), 41–55

Herulambang, W., Hidayat, M. M., & Putri, N. A. (2023). *Educational data mining for mapping the comprehension of personality subject using K-Means algorithm (case study of SD Insan Mulya)*. *Jurnal Mandiri IT*, 11(4), 152–158

Kumar, R., Singh, P., & Verma, A. (2024). Student performance grouping using unsupervised machine learning techniques. *International Journal of Data Science and Analytics*, 18(1), 55–67.

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)

Rahman, A., & Pratama, D. (2023). Implementation of K-Means clustering in academic performance analysis at secondary education level. *Journal of Educational Data Science*, 5(2), 112–120.

Santos, M., & Lee, J. (2025). Data-driven decision making in secondary education: Applications of clustering models. *Education and Information Technologies*, 30(3), 2145–2162

Siti Maghfiroh & Zaehol Fatah. (2024). *Analisis Data Mining dengan Algoritma K-Means Clustering untuk Menentukan Siswa Berprestasi di MTs Miftahul Ulum Bengkak*. *Jurnal Mahasiswa Teknik Informatika (JAMASTIKA)*.

Srianthaworn, D. A., et al. (2025). *Student performance analysis using the K-Means clustering method*. *Inovasi Pembangunan: Jurnal Kelitbangan*, 13(03), Article 1386. https://doi.org/10.35450/jip.v13i03.1386

Srirahmawati, E. O., Purnamasari, A. I., Bahtiar, A., & Tohidi, E. (2025). *Pengelompokan prestasi akademik siswa SD menggunakan algoritma K-Means*. *Jurnal Informatika dan Rekayasa Elektronik*, 8(1), Article 1358.

Sujana, T. S., Astuti, R., Prihartono, W., & Hamonangan, R. (2024). *Implementation of data mining using the K-Means algorithm to group students based on academic performance*. *Journal of Artificial Intelligence and Engineering Applications*.

Wahidin, A. J., & Syukrilla, W. A. (2023). *Data mining* (A. Y. Ediana, Dina (Ed.); Cetakan Pe). PT Global Eksekutif Teknologi.

WIDIANTO, F. F. (2025). Analisis Orange Data Mining Untuk Klasifikasi Penyakit Diabetes Menggunakan Model Decision Tree. *Cipasung Techno Pesantren: Scientific Journal*, *19*(2).

Yuni Franata Sinurat, Masrizal, I. (2024). *Data Mining Pengelompokan Siswa Berprestasi Menggunakan Metode Clustering* (Cetakan ke). Penerbit NEM.

Zhang, L., & Wu, X. (2023). Recent advances in clustering algorithms for educational data mining: A review. *IEEE Access*, 11, 45678–45690.