



Penerapan Algoritma K-Means untuk Mengidentifikasi Minat dan Bakat Siswa Sekolah Dasar

Sonianto^{1*}, Hartono²

sonianto@umpri.ac.id¹, hartono@umko.ac.id²

¹Universitas Muhammadiyah Pringsewu Lampung, Indonesia

²Universitas Muhammadiyah Kotabumi, Indonesia

*Korespondensi: ✉

Abstrak

This study aims to identify the interests and talents of elementary school students using the K-Means clustering algorithm, a data mining technique used to group data based on shared characteristics. The disciplinary areas employed in this research are education and data analysis, particularly in the context of curriculum development based on students' interests. The choice of this topic is driven by the need to create a more personalized and focused learning approach that aligns with students' interests and talents, ultimately enhancing the effectiveness of the learning process. The study utilizes questionnaire data consisting of four main dimensions: Academic Interest, Artistic Interest, Sports Interest, and Social and Leadership Interest, which are then analyzed using the K-Means clustering method. The formulated hypothesis is that K-Means clustering can group students into clusters that share similar interests and talents. The main findings of this study indicate that students can be classified into three primary clusters, each characterized by distinct patterns of interests and talents: students with high academic interest, students with strong interest in arts and social activities, and students with high interest in sports and leadership. The contribution of this study lies in providing practical insights for educators in designing more personalized and effective learning strategies, grounded in students' individual interests and talents.

Status Artikel:

Diterima: 10-05-2025

Direvisi: 14-05-2025

Diterima: 03-06-3035

Kata Kunci:

K-Means Clustering;

Minat dan Bakat Siswa;

Pendekatan Pembelajaran Personal;

Data Mining dalam Pendidikan;



© 2025 Sunianto, Hartono

This work is licensed under a

[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

PENDAHULUAN

Identifikasi minat dan bakat siswa sejak dini merupakan langkah krusial dalam membentuk arah perkembangan akademik dan non-akademik mereka. Di tingkat sekolah dasar (SD), pemahaman terhadap potensi individu siswa dapat membantu pendidik dalam merancang pendekatan pembelajaran yang lebih efektif dan sesuai dengan kebutuhan masing-masing siswa. Namun, proses identifikasi ini sering kali dilakukan secara manual dan subjektif, yang dapat mengakibatkan ketidaktepatan dalam penentuan minat dan bakat siswa. Untuk itu, diperlukan pendekatan yang lebih objektif dan efisien, salah satunya dengan memanfaatkan algoritma *K-Means clustering*.

K-Means clustering adalah metode dalam unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan fitur. Algoritma ini bekerja dengan cara membagi data ke dalam K cluster, dengan setiap data akan dimasukkan ke dalam cluster yang memiliki nilai rata-rata (*centroid*) yang paling dekat. Dalam konteks pendidikan, K-Means dapat digunakan untuk mengelompokkan siswa berdasarkan berbagai atribut, seperti minat, bakat, dan prestasi akademik.

Dalam konteks pendidikan dasar, beberapa studi sebelumnya telah menunjukkan pentingnya mengetahui minat dan bakat siswa sejak dini. Sebagaimana diungkapkan oleh Taufiq et al. (2024), pemetaan minat dan bakat siswa dapat meningkatkan efektivitas pembelajaran yang lebih personal dan sesuai dengan potensi masing-masing siswa.

Beberapa penelitian telah dilakukan untuk menerapkan algoritma K-Means dalam mengidentifikasi minat dan bakat siswa. Misalnya, penelitian yang dilakukan oleh Tiana, Siswanto, dan Kanedi (2024) dalam *Jurnal Media Infotama* mengaplikasikan K-Means untuk mengklasterisasi minat siswa terhadap mata pelajaran matematika di SMP Negeri 19 Kota Bengkulu. Hasil penelitian ini menunjukkan bahwa K-Means efektif dalam mengelompokkan siswa berdasarkan minat mereka terhadap mata pelajaran tertentu, yang dapat membantu guru dalam merancang strategi pembelajaran yang lebih tepat sasaran.

Teori yang mendasari penelitian ini adalah teori pendidikan yang menekankan pada pentingnya pemahaman terhadap perbedaan individu dalam proses pembelajaran. Pandangan mutakhir dalam psikologi pendidikan menekankan pentingnya memahami keragaman kecerdasan dan potensi siswa sebagai dasar pengembangan pembelajaran yang personal. Meskipun teori kecerdasan majemuk Gardner telah lama dikenal, studi-studi terbaru tetap menguatkan relevansinya dalam konteks pendidikan abad ke-21 yang menuntut pendekatan diferensiasi dalam pembelajaran (Gardner, 2020; Chen et al., 2021). Pendekatan ini selaras dengan pemanfaatan algoritma clustering seperti K-Means, yang memungkinkan pengelompokan siswa berdasarkan kemiripan karakteristik secara objektif dan berbasis data.

Dalam ranah data mining pendidikan, algoritma K-Means telah terbukti efektif dalam mengelompokkan data siswa dan mengungkap pola-pola tersembunyi yang tidak terdeteksi oleh metode konvensional. Penelitian terkini (Romero & Ventura, 2020; Albreiki et al., 2021; Ranjan et al., 2022) menunjukkan bahwa K-Means mampu membantu pendidik dalam menganalisis minat, bakat, dan perilaku belajar siswa secara lebih efisien dan akurat. Dengan demikian, algoritma ini menjadi alat yang strategis dalam mendukung perancangan kurikulum dan strategi pembelajaran yang lebih adaptif terhadap kebutuhan masing-masing siswa..

Adapun tujuan dari penelitian ini antara lain adalah untuk mengimplementasikan algoritma K-Means clustering untuk mengidentifikasi dan mengelompokkan siswa SD berdasarkan minat dan bakat mereka, selain itu juga untuk menganalisis efektivitas algoritma K-Means dalam mengelompokkan siswa berdasarkan minat dan bakat. Harapannya hasil penelitian ini dapat memberikan rekomendasi bagi pendidik dalam merancang pendekatan pembelajaran yang sesuai dengan profil minat dan bakat siswa.

METHODS

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen yang diterapkan pada siswa sekolah dasar untuk mengidentifikasi minat dan bakat mereka

menggunakan algoritma *K-Means clustering*. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran dan analisis data secara numerik dengan cara yang sistematis, objektif, dan dapat digeneralisasi ke populasi yang lebih luas (Sugiyono, 2017). Penelitian eksperimen memberikan kesempatan untuk mengontrol variabel tertentu dan memfokuskan pada efek dari penerapan algoritma K-Means dalam pengelompokan minat dan bakat siswa (Creswell, 2014).

B. Populasi dan Sampel

Populasi dalam penelitian ini adalah seluruh siswa yang terdaftar di beberapa sekolah dasar di wilayah tertentu. Sampel yang digunakan terdiri dari 150 siswa yang dipilih menggunakan teknik *simple random sampling*. Teknik ini dipilih untuk memastikan bahwa setiap siswa dalam populasi memiliki peluang yang sama untuk terpilih sebagai sampel, sehingga hasil penelitian dapat digeneralisasi secara lebih akurat ke populasi yang lebih luas (Sugiyono, 2021). Penggunaan *random sampling* juga bertujuan untuk meminimalkan potensi bias dalam pemilihan sampel dan meningkatkan validitas eksternal penelitian (Etikan & Bala, 2020).

C. Pengumpulan Data

Data yang digunakan dalam penelitian ini dikumpulkan melalui kuesioner minat dan bakat yang terdiri dari 45 pernyataan. Kuesioner ini dirancang berdasarkan teori kecerdasan majemuk yang dikemukakan oleh Gardner (1983), yang mengidentifikasi berbagai jenis kecerdasan yang relevan dengan minat dan bakat siswa di sekolah dasar. Skala Likert 5 poin digunakan dalam pengukuran, dengan penilaian dari 1 (Sangat Tidak Setuju) hingga 5 (Sangat Setuju), yang memungkinkan siswa untuk mengekspresikan tingkat persetujuan mereka terhadap setiap pernyataan dalam kuesioner. Validitas dan reliabilitas kuesioner diuji menggunakan metode Cronbach Alpha yang diperkenalkan oleh Tavakol & Dennick (2011).

D. Metode Analisis Data

Untuk menganalisis data yang terkumpul, digunakan algoritma K-Means clustering, sebuah metode dalam Unsupervised Learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan dalam fitur-fitur yang ada. K-Means clustering adalah salah satu teknik yang populer dalam data mining dan telah banyak digunakan dalam berbagai bidang untuk mengidentifikasi pola dalam data (Han, Kamber, & Pei, 2011). Langkah-langkah analisis yang dilakukan adalah sebagai berikut :

1. Preprocessing Data

Sebelum dilakukan clustering, data yang diperoleh dari kuesioner akan melalui tahap preprocessing, yang meliputi pengecekan data hilang dan normalisasi data. Normalisasi dilakukan untuk memastikan bahwa semua atribut memiliki skala yang sama, yang penting untuk memaksimalkan hasil clustering (Suthaharan, 2016). Preprocessing juga meliputi penghapusan data yang tidak valid dan pembersihan data untuk memastikan kualitas data yang lebih baik (Kotsiantis, 2011).

2. Menentukan Jumlah Cluster

Penentuan jumlah cluster yang optimal sangat penting dalam penerapan K-Means. Salah satu metode yang digunakan untuk menentukan nilai K adalah Elbow Method yang mengukur variasi within-cluster sum of squares (WCSS) untuk berbagai nilai K.

Metode ini banyak digunakan dalam penelitian untuk memilih jumlah cluster yang optimal dalam algoritma K-Means (Lloyd, 1982; Wu et al., 2012).

3. Penerapan K-Means clustering

Setelah menentukan jumlah cluster yang optimal, algoritma K-Means diterapkan untuk mengelompokkan siswa berdasarkan minat dan bakat mereka. Proses ini dilakukan dengan menggunakan perangkat lunak Python dan pustaka Scikit-learn, yang telah terbukti efektif dalam implementasi algoritma machine learning (Pedregosa et al., 2011).

4. Evaluasi Hasil Clustering

Evaluasi hasil clustering dilakukan dengan menggunakan Silhouette Score, yang mengukur kualitas clustering berdasarkan kesamaan data dalam satu cluster dibandingkan dengan cluster lainnya. Nilai Silhouette yang lebih tinggi menunjukkan bahwa clustering yang dilakukan lebih baik dan memiliki pemisahan yang jelas antara cluster (Rousseeuw, 1987). Silhouette Score adalah metrik yang umum digunakan dalam banyak penelitian clustering untuk mengevaluasi ketepatan pembentukan cluster (Xia et al., 2015).

RESULTS AND DISCUSSION

Hasil

Data Hasil Kuesioner

Berikut ini adalah data hasil identifikasi minat dan bakat dari 150 responden melalui instrumen kuesioner. Setiap responden diminta untuk memberikan penilaian terhadap 45 pernyataan yang mencakup berbagai dimensi minat dan bakat, seperti Minat Akademik, Minat dalam Seni, Minat Olahraga, dan Minat Sosial dan Kepemimpinan. Penilaian diberikan menggunakan skala Likert 5 poin, yang berkisar dari 1 (Sangat Tidak Setuju) hingga 5 (Sangat Setuju).

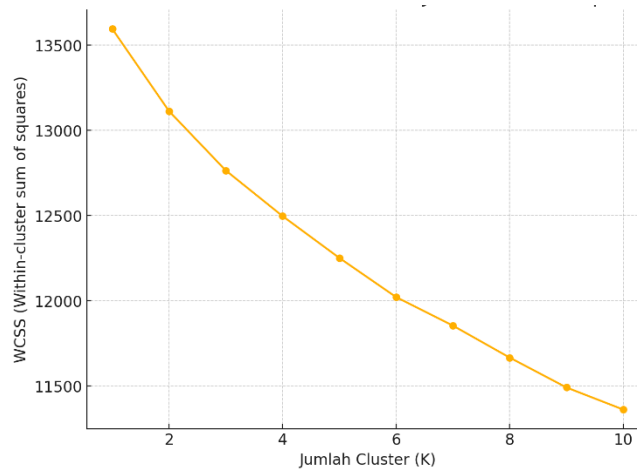
Tabel 1. Data Hasil Kuesioner

Responden	Matematika	IPA	Baca_Buku	Pecahkan_Masalah_Matematika	Sains_Experiment	Menggambar	Dengarkan_Musik	Bernyanyi	Main_Alat_Musik	Menari	Sepak_Bola	Bulu_Tangkis	Lomba_Lari	Percaya_Diri_Situasi_Sosial	Kegiatan_Luar_Sekolah	Eksplorasi_Baru
R1	4	5	3	5	5	2	3	3	3	5	4	3	5	...	4	2	2
R2	1	2	5	2	4	4	4	4	5	3	1	4	2	...	1	4	3
R3	3	1	3	3	1	3	5	2	2	1	4	1	4	...	4	5	3
R4	3	4	2	2	5	1	5	4	4	4	4	4	3	...	1	5	1
R5	1	3	1	2	2	4	5	1	1	3	2	5	4	...	3	4	1
R6	1	2	2	3	4	2	1	4	4	1	2	1	4	...	1	5	4
R7	3	1	1	4	3	3	5	3	3	3	2	5	1	...	2	1	2

....	
R148	5	2	2	2	3	4	3	5	1	4	1	5	5	...	4	1	4
R149	5	5	2	1	5	3	1	1	5	3	2	5	3	...	4	3	1
R150	5	2	5	3	1	4	5	4	1	5	4	4	5	...	5	1	2

Menentukan Jumlah Cluster

Langkah berikutnya adalah menentukan jumlah cluster. Pada penelitian ini, penentuan jumlah cluster menggunakan metode elbow, yakni teknik yang umum digunakan untuk memilih jumlah cluster yang optimal berdasarkan Within-Cluster Sum of Squares (WCSS). WCSS mengukur seberapa baik data dalam setiap cluster dikelompokkan, semakin rendah nilai WCSS, semakin baik pengelompokan data tersebut. Pada metode Elbow, nilai WCSS dihitung untuk berbagai nilai K (jumlah cluster), dan grafik WCSS terhadap K digunakan untuk mengidentifikasi "elbow point" — titik di mana penurunan WCSS mulai melambat. Titik ini menunjukkan jumlah cluster yang optimal, karena penurunan lebih lanjut tidak memberikan pengurangan WCSS yang signifikan.



Gambar 1. Jumlah Cluster menggunakan Metode Elbow

Grafik Elbow di atas menunjukkan hubungan antara jumlah cluster (K) dan Within-cluster Sum of Squares (WCSS). Berdasarkan grafik tersebut, kita dapat mengamati bahwa K = 1 hingga K = 3 terjadi penurunan WCSS cukup tajam, menunjukkan bahwa pengelompokan data semakin baik seiring bertambahnya jumlah cluster. Kemudian pada K = 4 terjadi penurunan yang lebih lambat dalam WCSS, yang menandakan bahwa penambahan jumlah cluster lebih lanjut tidak memberikan peningkatan yang signifikan dalam kualitas clustering. Oleh karena itu dapat disimpulkan bahwa K = 3 adalah jumlah cluster yang paling efektif untuk digunakan dalam analisis ini, karena memberi keseimbangan terbaik antara kompleksitas model dan kualitas clustering.

Inisialisasi Centroid

Setelah menentukan jumlah cluster yang optimal menggunakan metode Elbow (K = 3), langkah selanjutnya dalam proses K-Means clustering adalah centralisasi centroid, yaitu menghitung posisi centroid untuk masing-masing cluster. Proses ini adalah bagian penting dari algoritma K-Means untuk memastikan bahwa setiap cluster memiliki titik pusat yang representatif dari data di dalamnya.

Berikut ini adalah centroid untuk masing-masing cluster dengan menggunakan empat dimensi minat dan bakat.

Tabel 2. Centroid Masing-Masing Cluster

Cluster	Minat Akademik	Minat dalam Seni	Minat Olahraga	Minat Sosial dan Kepemimpinan
Cluster 1	2.64	2.67	3.08	2.91
Cluster 2	2.88	3.44	2.84	3.11
Cluster 3	3.50	2.96	3.17	3.07

Dari tabel 2 dapat dijelaskan sebagai berikut :

Cluster 1. Siswa dalam cluster ini memiliki minat rendah pada seni, tetapi minat yang cukup tinggi pada olahraga dan sosial/kepemimpinan. Mereka mungkin lebih aktif dalam kegiatan sosial dan kepemimpinan.

Cluster 2. Siswa dalam cluster ini menunjukkan minat tinggi pada seni, dengan minat moderat pada akademik dan olahraga. Mereka mungkin lebih tertarik pada kegiatan kreatif dan ekstrakurikuler.

Cluster 3. Siswa dalam cluster ini memiliki minat tinggi pada akademik, khususnya Matematika dan IPA, dengan minat yang seimbang pada seni, olahraga, dan sosial.

Hasil Pengelompokkan

Setelah melalui tahapan K-Means clustering dengan jumlah cluster yang ditentukan sebanyak 3 ($K = 3$), siswa dalam penelitian ini dikelompokkan berdasarkan minat dan bakat mereka. Proses pengelompokan ini bertujuan untuk mengidentifikasi pola dalam data siswa berdasarkan empat dimensi utama, yaitu Minat Akademik, Minat dalam Seni, Minat Olahraga, dan Minat Sosial dan Kepemimpinan.

Pengelompokan ini dihasilkan melalui analisis K-Means clustering, yang memungkinkan pemisahan siswa ke dalam kelompok-kelompok yang memiliki kesamaan karakteristik dalam hal minat dan bakat. Setiap responden atau siswa akan ditempatkan dalam cluster yang paling sesuai dengan data mereka berdasarkan kedekatan dengan centroid masing-masing cluster.

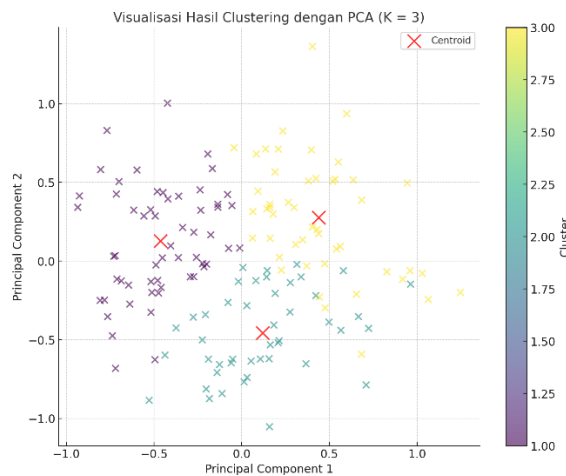
Hasil pengelompokan ini memberikan wawasan yang lebih dalam mengenai kelompok-kelompok siswa yang memiliki minat dan bakat yang serupa, yang selanjutnya dapat digunakan untuk merancang pendekatan pembelajaran yang lebih personal dan terfokus. Pembagian ini juga memberikan informasi penting tentang bagaimana masing-masing siswa cenderung berkembang dalam bidang-bidang tertentu.

Tabel di bawah ini menampilkan hasil pengelompokan setiap responden berdasarkan cluster yang telah ditentukan, yang mencakup nilai-nilai untuk Minat Akademik, Minat dalam Seni, Minat Olahraga, dan Minat Sosial dan Kepemimpinan, beserta label Cluster yang menunjukkan keanggotaan siswa dalam salah satu dari tiga cluster.

Tabel 3. Hasil Clustering

Responden	Minat Akademik	Minat dalam Seni	Minat Olahraga	Minat Sosial dan Kepemimpinan	Cluster
R1	3.5	3.2	3.3	3.0	3
R2	3.0	3.8	2.8	2.8	2
R3	2.4	2.6	2.7	4.07	1
R4	3.0	3.1	3.1	2.07	1
R5	2.0	3.3	3.2	2.67	1
R6	2.7	2.5	3.0	2.73	1
R7	2.5	3.5	3.2	3.27	2
...
R148	3.1	3.2	2.6	2.8	2
R149	3.7	2.6	2.9	2.87	3
R150	2.9	2.9	3.7	2.93	1

Hasil clustering juga dapat divisualisasikan dalam bentuk grafik dengan model *Principal Component Analysis* (PCA) untuk mereduksi dimensi menjadi dua komponen utama. Setiap titik mewakili responden (siswa), dan warna yang berbeda menunjukkan cluster yang telah ditentukan.



Gambar 2. Visualisasi Hasil Clustering

Berdasarkan hasil clustering yang disajikan pada tabel 3 maupun pada gambar 2, dapat diinterpretasikan sebagai berikut :

Cluster 1. Siswa dalam cluster ini cenderung lebih aktif dalam kegiatan sosial dan kepemimpinan, mungkin berperan dalam organisasi atau kegiatan ekstrakurikuler yang berfokus pada interaksi sosial.

Cluster 2. Siswa dalam cluster ini lebih seimbang dalam minat antara akademik dan seni, cenderung mengikuti kegiatan yang mencakup keduanya.

Cluster 3. Siswa di cluster ini lebih terfokus pada akademik, khususnya pada matematika dan IPA.

SIMPULAN

Berdasarkan hasil penerapan algoritma K-Means clustering pada data minat dan bakat siswa sekolah dasar, penelitian ini berhasil mengelompokkan siswa ke dalam tiga cluster yang mencerminkan karakteristik minat dan bakat yang berbeda. Cluster pertama terdiri dari siswa yang lebih fokus pada kegiatan sosial dan kepemimpinan, dengan minat moderat pada akademik dan seni. Cluster kedua mencakup siswa yang lebih tertarik pada seni, dengan keterlibatan sedang dalam bidang akademik dan sosial. Sedangkan, cluster ketiga berisi siswa yang memiliki minat tinggi pada akademik, khususnya dalam Matematika dan IPA, meskipun tetap menunjukkan ketertarikan pada kegiatan non-akademik seperti seni dan olahraga. Penerapan K-Means clustering dengan $K = 3$ memberikan gambaran yang jelas mengenai perbedaan minat dan bakat siswa, yang selanjutnya dapat digunakan untuk merancang pendekatan pembelajaran yang lebih personal dan terfokus. Berdasarkan hasil clustering ini, direkomendasikan agar Cluster 1 diberikan lebih banyak kesempatan dalam kegiatan sosial dan kepemimpinan, Cluster 2 difokuskan pada pengembangan minat seni dan kreativitas, sementara Cluster 3 bisa diberikan tantangan akademik lebih lanjut, khususnya dalam Matematika dan IPA. Dengan demikian, hasil penelitian ini menunjukkan pentingnya pendekatan pembelajaran yang disesuaikan dengan minat dan bakat siswa untuk meningkatkan hasil pembelajaran dan pengembangan potensi individu.

REFERENSI

- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student' performance prediction using machine learning techniques. *Education and Information Technologies*.
- Arikunto, S. (2013). *Prosedur Penelitian: Suatu Pendekatan Praktek* (12th ed.). Rineka Cipta.
- Chen, J.-Q., Moran, S., & Gardner, H. (2021). *Multiple Intelligences around the World*.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.). Sage Publications.
- Etikan, I., & Bala, K. (2020). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 8(1), 001–002.
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gardner, H. (2020). *Multiple Intelligences: New Horizons in Theory and Practice*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Kotsiantis, S. B. (2011). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31(3), 249–268.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Ranjan, R., Kumari, V. A., & Kumar, A. (2022). Student Performance Analysis using K-Means Clustering in Educational Data Mining. *Materials Today: Proceedings*.
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Graphical Statistics*, 1(1), 53–65.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

- Sugiyono. (2021). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- Taufiq, R., Kartika, V. D., & Muslimah. (2024). Penerapan Metode K-Means untuk Klasterisasi Minat dan Bakat Siswa terhadap Ekstrakurikuler Sekolah. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, 8(3).
<https://doi.org/10.33395/remik.v8i3.14008>
- Tiana, M., Siswanto, S., & Kanedi, I. (2024). Penerapan Metode K-Means Clustering Dalam Mengetahui Minat Siswa Terhadap Mata Pelajaran Matematika Di SMP Negeri 19 Kota Bengkulu. *Jurnal Media Infotama*, 20(2). <https://doi.org/10.37676/jmi.v20i2.6567>
- Wu, X., et al. (2012). A Study on K-Means Clustering Algorithm: A Survey. *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, 1131–1136.