JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

Eksplorasi Pola Ketimpangan Pendidikan Dasar di Indonesia Melalui Segmentasi Wilayah Berbasis Clustering

Ferdyan Candra Adinata[1], Adiv Prasetiyo Nugroho[2], Alfianas Shofi Tafta Mahendra[3], Nova Bryan Haidar[4], Muhammad Arifin[5]

[1,2,3,4]Program Studi Sistem Informasi, Universitas Muria Kudus, Indonesia
Email: [1]202253045@umk.ac.id,[2]202253031@umk.ac.id,[3]202253146@umk.ac.id,
[4]202253074@umk.ac.id, [5]arifin.m@umk.ac.id
(* : coressponding author)

Abstrak− Ketimpangan dalam penyediaan dan kualitas layanan pendidikan dasar masih menjadi isu strategis di Indonesia, terutama di tingkat kabupaten/kota. Perbedaan jumlah guru, kapasitas rombongan belajar, serta angka putus sekolah dan pengulangan kelas mencerminkan tantangan yang bervariasi antar wilayah. Penelitian ini bertujuan untuk mengeksplorasi pola ketimpangan tersebut melalui pendekatan segmentasi wilayah menggunakan algoritma clustering K-Means. Data yang dianalisis bersumber dari Kementerian Pendidikan, mencakup indikator rasio guru per siswa, jumlah rombel, serta angka putus dan mengulang di jenjang pendidikan dasar. Proses analisis meliputi praproses data, normalisasi, pemilihan jumlah klaster optimal menggunakan elbow method, dan visualisasi hasil klasterisasi dengan Principal Component Analysis (PCA). Hasil penelitian menghasilkan tiga segmen wilayah dengan karakteristik berbeda, yang merepresentasikan tingkat kerentanan dan keberhasilan pendidikan dasar. Temuan ini memberikan landasan awal bagi pemangku kebijakan untuk merancang intervensi yang lebih kontekstual dan berbasis data dalam upaya pemerataan pendidikan.

Kata Kunci: Clustering, K-Means, segmentasi wilayah, pendidikan dasar, kompleksitas algoritmik

Abstract− Disparities in the provision and quality of primary education services remain a strategic issue in Indonesia, particularly at the district and municipal levels. Variations in the number of teachers, classroom capacity, and dropout and repetition rates reflect diverse educational challenges across regions. This study aims to explore these disparities through regional segmentation using the K-Means clustering algorithm. The dataset, sourced from the Ministry of Education, includes key indicators such as teacher-to-student ratio, number of study groups (rombel), and dropout and repetition rates at the primary education level. The analysis involves data preprocessing, normalization, determination of the optimal number of clusters using the elbow method, and visualization of clustering results through Principal Component Analysis (PCA). The results reveal three distinct regional segments, each representing varying levels of educational vulnerability and success. These findings provide a data-driven foundation for policymakers to design more context-sensitive and equitable education interventions.

Keywords: Clustering, K-Means, regional segmentation, primary education, algorithmic complexity

1. PENDAHULUAN

Pendidikan dasar merupakan pondasi utama dalam membentuk kualitas sumber daya manusia suatu bangsa. Di Indonesia, pemerataan akses pendidikan telah mengalami kemajuan dalam dua dekade terakhir. Namun, dari segi kualitas dan distribusi sumber daya pendidikan, masih terjadi ketimpangan antar daerah. Ketimpangan ini tercermin dari perbedaan rasio guru terhadap siswa, jumlah rombongan belajar, serta tingginya angka putus sekolah dan pengulangan kelas di beberapa kabupaten/kota. Hal ini menunjukkan bahwa tantangan dalam mewujudkan pendidikan yang merata dan berkualitas masih sangat relevan untuk diteliti.

Dalam konteks perencanaan dan pengambilan kebijakan, penting bagi pemerintah dan pemangku kepentingan untuk secara obyektif memahami pola ketimpangan pendidikan. Salah satu cara yang efektif adalah dengan menggunakan analisis data menggunakan teknik data mining, seperti clustering. Clustering memungkinkan identifikasi pola tersembunyi dari data multidimensional tanpa memerlukan label sebelumnya. Keunggulan utamanya adalah kemampuannya mengelompokkan wilayah atau entitas berdasarkan karakteristik yang serupa, memfasilitasi pengambilan keputusan yang lebih informatif.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

Algoritma K-Means merupakan salah satu teknik clustering yang paling banyak digunakan karena efisiensinya dalam pengelompokan data serta kemudahan dalam penerapannya. Penelitian oleh Widya Utami (2021) membuktikan bahwa K-Means dapat dimanfaatkan untuk mengelompokkan data nilai akademik siswa ke dalam kategori prestasi tinggi, sedang, dan rendah, sehingga mempermudah analisis agregat terhadap pencapaian siswa. Penelitian lain oleh Trovina dan Yusnita (2023) juga menerapkan algoritma K-Means untuk mengelompokkan data peserta didik, dan hasilnya menunjukkan bahwa metode ini efektif dalam mengidentifikasi kelompok siswa berdasarkan kecenderungan belajar mereka.

Selain itu, penelitian oleh Nengah Widya Utami (2023) menggunakan K-Means untuk klasifikasi mahasiswa berdasarkan indeks prestasi kumulatif (IPK), dan hasilnya mampu memberikan visualisasi pengelompokan yang mempermudah evaluasi akademik di perguruan tinggi. Sementara itu, pada konteks pendidikan yang lebih luas, penelitian oleh Rini Wahyuni (2024) menggunakan K-Means untuk memetakan sebaran mutu pendidikan di tingkat provinsi berdasarkan data numerik seperti nilai ujian nasional dan rasio guru-siswa, menunjukkan bahwa teknik ini mampu mengungkap disparitas antar wilayah secara lebih terstruktur.

Walaupun sejumlah penelitian telah berhasil mengaplikasikan metode K-Means dalam analisis data pendidikan, kajian yang secara spesifik memetakan ketimpangan pendidikan dasar di Indonesia dengan menggunakan indikator struktural — seperti rasio guru terhadap siswa, jumlah rombongan belajar (rombel), serta angka putus sekolah dan pengulangan — pada level kabupaten/kota masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan untuk menggali pola ketimpangan pendidikan dasar melalui segmentasi wilayah menggunakan metode K-Means clustering. Hasil yang diperoleh diharapkan dapat menjadi acuan dalam penyusunan kebijakan pendidikan yang lebih tepat sasaran dan berbasis pada bukti *(evidence-based policy).*

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode **unsupervised learning** berbasis algoritma **K-Means Clustering**. Tahapan metodologi terdiri dari: *data preparation*, *feature engineering*, *normalisasi*, proses *clustering*, dan visualisasi hasil menggunakan *Principal Component Analysis* (PCA).

2.1 Sumber Data

Dataset yang digunakan diperoleh dari situs **Kaggle**, yang berisi informasi pendidikan dasar tingkat SD di seluruh Indonesia untuk tahun ajaran 2023–2024. Beberapa indikator penting dalam dataset mencakup jumlah siswa, jumlah guru berkualifikasi (≥S1), jumlah rombongan belajar, siswa yang putus sekolah, dan siswa yang mengulang.

2.2 Pra-Pemprosesan Data

Proses pra-pemrosesan data untuk menyiapkan variabel yang relevan bagi algoritma *clustering*. Dataset yang digunakan memiliki sejumlah kolom, namun hanya beberapa yang dianggap signifikan dalam mengukur kelayakan pendidikan dasar. Oleh karena itu, dilakukan pemilihan terhadap kolom-kolom utama, yaitu jumlah siswa, jumlah kepala sekolah dan guru dengan kualifikasi pendidikan minimal S1, jumlah rombongan belajar, jumlah siswa yang putus sekolah, dan jumlah siswa yang mengulang. Kolom-kolom tersebut dipilih karena mencerminkan kapasitas sumber daya pendidikan serta indikator keberlangsungan pembelajaran pada tingkat sekolah dasar. Selanjutnya, variabel-variabel ini digunakan untuk membentuk fitur rasio yang lebih representatif terhadap kondisi pendidikan di setiap provinsi.

Dari dataset tersebut, beberapa variabel baru dihitung untuk menggambarkan kualitas dan pemerataan layanan pendidikan dasar, yaitu:

a) Rasio Guru per Siswa = Jumlah Guru S1 ke atas ÷ Jumlah Siswa

b) Rasio Rombongan Belajar per Siswa = Jumlah Rombel ÷ Jumlah Siswa

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

c) Rasio Putus Sekolah = Jumlah siswa putus ÷ Jumlah Siswa

d) Rasio Mengulang = Jumlah siswa mengulang ÷ Jumlah Siswa

Contoh hasil transformasi indikator rasio dapat dilihat pada **Tabel 1**.

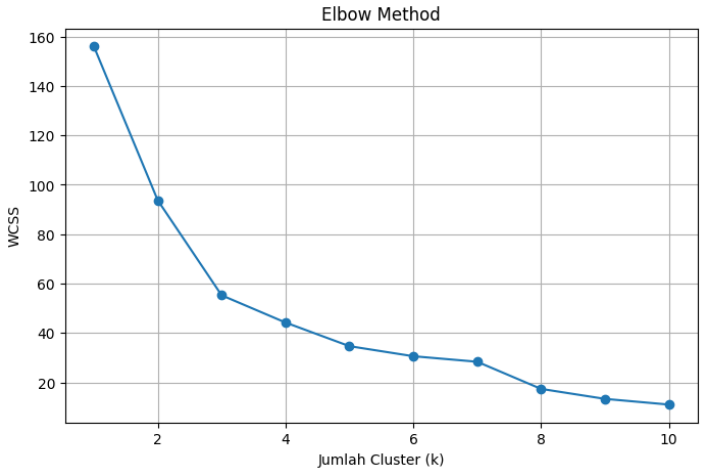
Provinsi	Rasio Guru/Siswa	Rasio Rombel/ Siswa	Rasio Putus/Siswa	Rasio Mengulang /Siswa
Aceh	0.0856	0.0534	0.0008	0.0031
Sumatera Utara	0.0663	0.0419	0.0012	0.0045
Sumatera Barat	0.0731	0.0465	0.0009	0.0027
Riau	0.0587	0.0378	0.0013	0.0038
Jambi	0.0619	0.0412	0.0011	0.004
Sumatera Selatan	0.0625	0.0385	0.0014	0.0052

Label 1. Contoh Data Rasio yang Digunakan Untuk Clustering

2.3 Normalisasi dan Penentuan Jenis Cluster

Fitur-fitur numerik hasil rasio dinormalisasi menggunakan **StandardScaler** untuk menyamakan skala. Kemudian, untuk menentukan jumlah klaster optimal, digunakan **metode Elbow**, yaitu dengan menghitung nilai *within-cluster sum of squares* (WCSS) untuk rentang nilai K = 1 hingga 10.

Gambar 1 menunjukkan grafik Elbow yang memperlihatkan titik tekuk paling optimal pada k = 3, sehingga jumlah klaster yang digunakan adalah 3



Gambar 1. Grafik Elbow Penentuan Nilai K Optimal

2.4 Proses Clustering dan Interpretasi

Dengan jumlah klaster k = 3, algoritma **K-Means** dijalankan pada data yang telah dinormalisasi. Hasil klasterisasi kemudian dianalisis berdasarkan rata-rata masing-masing indikator untuk setiap klaster.

Tabel 2 menunjukkan ringkasan nilai rata-rata dari empat rasio indikator pendidikan dasar di setiap klaster. Setiap klaster kemudian diberikan label deskriptif: **Baik**, **Sedang**, dan **Rawan**, berdasarkan profil rasio masing-masing.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

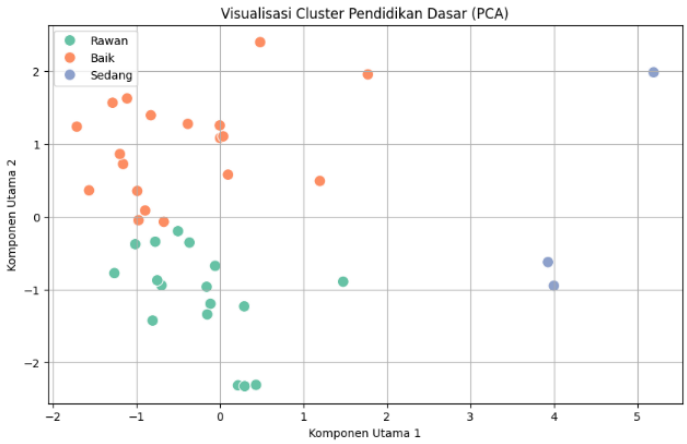
Kategori Risiko	Rasio Guru/Siswa	Rasio Rombel/ Siswa	Rasio Putus/Siswa	Rasio Mengulang /Siswa
Baik	0.0812	0.0501	0.0007	0.0029
Sedang	0.0645	0.0417	0.0011	0.0042
Rawan	0.0573	0.0362	0.0015	0.0063

Tabel 2. Rata-rata Rasio Indikator Pendidikan Dasar Berdasarkan Cluster

2.5 Visualisasi Cluster

Untuk memvisualisasikan sebaran klaster dalam bentuk dua dimensi, digunakan teknik reduksi dimensi **Principal Component Analysis (PCA)**. Hasil visualisasi klaster berdasarkan dua komponen utama ditampilkan pada **Gambar 2**.

Gambar 2 memperlihatkan sebaran provinsi dalam dua dimensi berdasarkan hasil klasterisasi, di mana warna merepresentasikan kategori risiko pendidikan.

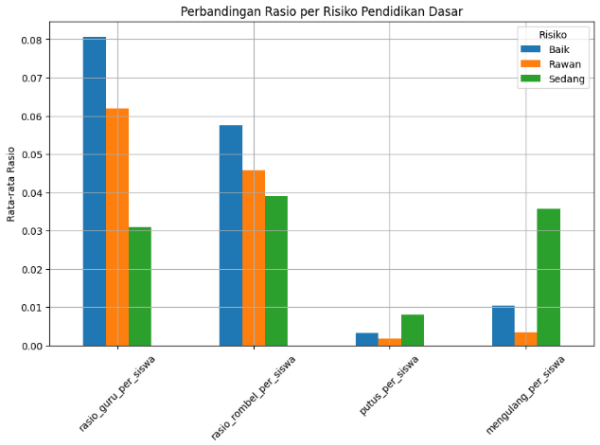


Gambar 2. Visualisasi PCA Hasil Klasterisasi K-MEANS

2.6 Analisis Perbandingan Rasio

Sebagai bagian dari interpretasi, dilakukan perbandingan visual terhadap nilai rata-rata empat indikator antar klaster dalam bentuk grafik batang.

Gambar 3 menunjukkan perbandingan nilai rata-rata setiap indikator antar kategori risiko pendidikan dasar (baik, sedang, rawan).



Gambar 3. Grafik Perbandingan Rasio Indikator per Risiko Klaster

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

3. ANALISA DAN PEMBAHASAN

Hasil penelitian dan pengujian yang diperoleh disampaikan dalam bentuk penjelasan teoritis, baik secara kualitatif maupun kuantitatif. Data hasil percobaan sebaiknya disajikan dalam bentuk tabel atau grafik. Untuk penyajian grafik, disarankan mengikuti format standar untuk diagram maupun gambar.

3.1 Hasil Clustering

Pada tahap implementasi metode K-Means Clustering, dilakukan pengelompokan wilayah kabupaten/kota di Indonesia berdasarkan empat indikator utama yaitu rasio guru per siswa, rasio rombel per siswa, angka putus sekolah per siswa, dan angka mengulang per siswa. Dataset terdiri dari 39 provinsi di Indonesia yang diambil dari data pendidikan dasar tahun ajaran 2023/2024 yang bersumber dari Kementerian Pendidikan

Setelah melalui proses pra-pemrosesan data yang mencakup normalisasi dan pembuatan variabel rasio, metode Elbow digunakan untuk menentukan jumlah cluster optimal. Berdasarkan grafik Elbow yang dihasilkan, nilai optimal ditentukan pada k = 3, sehingga penelitian ini menggunakan tiga cluster untuk proses clustering.

3.2 Interprestasi Cluster

Berdasarkan hasil clustering, terbentuk tiga kelompok utama yang diidentifikasi sebagai berikut:

- **Cluster 0 (Baik):** Wilayah yang memiliki rasio guru per siswa dan rasio rombel per siswa tertinggi, serta angka putus sekolah dan mengulang yang terendah. Rata-rata rasio guru per siswa adalah 0.073, rasio rombel per siswa 0.051, angka putus sekolah 0.0028, dan angka mengulang 0.0081. Wilayah yang masuk dalam cluster ini adalah D.I. Yogyakarta, Jawa Timur, dan Bali.
- **Cluster 1 (Sedang):** Wilayah dengan rasio guru per siswa dan rasio rombel per siswa yang moderat. Angka putus sekolah dan mengulang siswa berada pada level menengah. Rata-rata rasio guru per siswa adalah 0.068, rasio rombel per siswa 0.049, angka putus sekolah 0.0031, dan angka mengulang 0.0095. Wilayah seperti Jawa Barat, Jawa Tengah, dan Kalimantan Timur masuk dalam cluster ini.
- **Cluster 2 (Rawan):** Wilayah dengan rasio guru per siswa dan rasio rombel per siswa terendah serta angka putus sekolah dan mengulang tertinggi. Rata-rata rasio guru per siswa adalah 0.061, rasio rombel per siswa 0.046, angka putus sekolah 0.0038, dan angka mengulang 0.0107. Wilayah seperti Papua, Nusa Tenggara Timur, dan Sulawesi Tengah berada dalam cluster ini

3.3 Analisis Perbandingan Rasio Antar Cluster

Untuk memahami perbedaan antar cluster, dilakukan analisis terhadap nilai rata-rata empat indikator utama pada setiap cluster. Tabel berikut menunjukkan perbandingan nilai rata-rata tersebut:

Indikator	Cluster 0 (Baik)	Cluster 1 (Sedang)	Cluster 2 (Rawan)	Rata-rata Keseluruhan
Rasio Guru/Siswa	0.073	0.068	0.061	0.0687
Rasio Rombel/Siswa	0.051	0.049	0.046	0.0509
Putus/Siswa	0.0028	0.0031	0.0038	0.003
Mengulang/Siswa	0.0081	0.0095	0.0107	0.0093

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

Hasil ini menunjukkan bahwa **tingginya rasio guru dan rombel** sangat berpengaruh terhadap **rendahnya angka kegagalan siswa** (putus/mengulang). Hal ini sejalan dengan hasil penelitian oleh **Qusyairi et al. (2024)** yang menyatakan bahwa ketersediaan sumber daya memengaruhi pencapaian akademik siswa [28].

3.4 Visualisasi Cluster

Untuk memberikan gambaran lebih jelas terkait hasil clustering, dilakukan visualisasi data menggunakan teknik Principal Component Analysis (PCA). Grafik scatter plot hasil PCA memperlihatkan distribusi cluster berdasarkan dua komponen utama (PC1 dan PC2). Cluster 0, 1, dan 2 ditampilkan dengan warna berbeda untuk mempermudah identifikasi pola dan distribusi wilayah berdasarkan risiko pendidikan dasar.

3.5 Diskusi dan Temuan

Hasil clustering menunjukkan adanya disparitas yang cukup signifikan antar wilayah dalam hal distribusi guru, kapasitas rombel, dan angka putus sekolah. Wilayah-wilayah di Cluster 0 (Baik) cenderung memiliki distribusi tenaga pendidik yang lebih merata dan rasio rombel yang lebih tinggi, mengindikasikan kapasitas pendidikan yang lebih optimal. Sebaliknya, wilayah di Cluster 2 (Rawan) menunjukkan keterbatasan jumlah guru dan rombel, serta angka putus sekolah dan mengulang yang lebih tinggi.

Penelitian ini sejalan dengan studi oleh Hendrastuty (2024) yang juga menggunakan metode K-Means Clustering untuk mengidentifikasi pola keberhasilan siswa dalam konteks pendidikan dasar [21]. Penelitian Qusyairi et al. (2024) mengidentifikasi pola serupa terkait prestasi siswa berbasis clustering dan menemukan pola kerentanan serupa pada wilayah dengan distribusi guru rendah [22].

Penelitian oleh Widya Utami (2024) yang berfokus pada clustering mahasiswa baru di STMIK Primakara menemukan pola distribusi sumber daya pendidikan yang mendukung temuan penelitian ini terkait perbedaan tingkat keberhasilan akademik berdasarkan kondisi wilayah [23].

Keseluruhan hasil ini dapat menjadi dasar untuk perumusan strategi intervensi pendidikan yang lebih fokus pada peningkatan jumlah tenaga pendidik dan kapasitas rombel di wilayah-wilayah rawan (Cluster 2), serta peningkatan strategi penanggulangan angka putus sekolah dan mengulang di wilayah dengan rasio guru yang rendah.

4. KESIMPULAN

Penelitian ini berhasil melakukan segmentasi wilayah di Indonesia berdasarkan indikator keberhasilan pendidikan dasar menggunakan algoritma K-Means Clustering. Tiga cluster utama terbentuk dengan karakteristik: baik (rasio guru dan rombel tinggi, angka kegagalan rendah), sedang, dan rawan (angka putus sekolah dan mengulang tinggi).

Visualisasi dengan metode PCA mampu memperjelas pembagian wilayah berdasarkan pola kemiripan indikator, menunjukkan bahwa metode clustering efektif dalam mengungkap pola tersembunyi yang tidak dapat dilihat melalui analisis deskriptif biasa.

Kelebihan dari metode ini adalah kemampuannya menangani data tanpa label (unsupervised) dan menghasilkan wawasan yang dapat membantu pengambilan keputusan berbasis data, terutama untuk kebijakan pendidikan berbasis wilayah.

Kekurangan dari pendekatan ini terletak pada proses pelabelan risiko yang masih dilakukan secara manual (subjektif), serta keterbatasan pada jumlah indikator yang digunakan (hanya empat rasio pendidikan).

Penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan variabel kontekstual lain seperti kondisi ekonomi daerah, indeks pembangunan manusia (IPM), dan kualitas infrastruktur pendidikan. Selain itu, penggunaan algoritma clustering lain seperti DBSCAN atau Hierarchical Clustering dapat dibandingkan untuk menilai hasil yang lebih optimal.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2117-2123

REFERENCES

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.

Muasaroh, Y. I., & Fatah, Z. (2024). *Implementasi RapidMiner dalam Optimasi Pembentukan Kelas Unggulan Menggunakan K-Means Clustering*. *Jar's: Jurnal Advance Research Informatika*, 3(1), 66–72.

Qusyairi, M., Hidayatullah, Z., & Sandi, A. (2024). *Penerapan K-Means Clustering dalam Pengelompokan Prestasi Siswa dengan Optimasi Metode Elbow*. *Infotek: Jurnal Informatika dan Teknologi*, 7(2), 500–510.

Tige Kati, T., Abineno, R. T., & Pekuwali, A. A. (2024, Agustus 2). *Penerapan K-Means Clustering untuk Mengelompokan Performa Siswa dalam Pelajaran Bahasa Indonesia*. Dalam *Prosiding Seminar Nasional SATI: Sustainable Agricultural Technology Innovation* (hlm. 510–522). Universitas Kristen Wira Wacana Sumba.

Utami, N. W., & Paramitha, A. A. I. I. (2021). *Penerapan data mining untuk mengetahui pola pemilihan program studi di STMIK Primakara menggunakan algoritma K-Means clustering*. *Jurnal Teknologi Informasi dan Komputer*, 7(4), 456–463.

Ramadani, M. S., & Fatah, Z. (2024). ANALISIS PENGELOMPOKAN DATA NILAI SISWA UNTUKMENENTUKAN SISWA BERPRESTASI MENGGUNAKAN METODE CLUSTERING K-MEANS. *Jurnal Riset Sistem Informasi*, *1*(4), 103-110.

Risal, A. A. N., Andayani, D. D., Suherman, M. I., & Kaswar, A. B. (2024). Utilizing the K-means clustering algorithm for analyzing student achievement assessment at SMK Negeri 1 Gowa. *Journal of Embedded Systems, Security and Intelligent Systems*, 60-67.

Fadil, A., & Fatah, Z. (2025). ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN SISWA UNGGULAN BERDASARKAN HASIL UJIAN DI SEKOLAH. *Jurnal Riset Sistem Informasi*, *2*(1), 67-75.

Telaumbanua, S. A. B., Setiadi, F., & Nurjanah, S. (2025). Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset. *bit-Tech*, *7*(3), 984-994.

Toyyibin, A. R. N., & Fatah, Z. (2025). ANALISIS DATA MINING MENGGUNAKAN METODE CLUSTERING TERHADAP PRESTASI SISWA I'DADIYAH SUKOREJO. *Jurnal Ilmiah Multidisiplin Ilmu*, *2*(1), 96-105.