

# Analisis Efektivitas Algoritma K-Means Clustering dalam Pengelompokan Siswa Berdasarkan Kemampuan Multidimensi

*Analysis of the Effectiveness of the K-Means Clustering Algorithm in Grouping Students Based on Multidimensional Abilities*

Hanif Naufal Rafandi<sup>1</sup>, Usman Nurhasan<sup>2</sup>, Endah Septa Sintiya<sup>\*3</sup>

<sup>1,2,3</sup>Prodi Teknik Informatika, Politeknik Negeri Malang

E-mail: <sup>1</sup>2141720127@student.polinema.ac.id, <sup>2</sup>usmannurhasan@polinema.ac.id,

<sup>3</sup>\*e.septa@polinema.ac.id

\*Corresponding author

Received 10 September 2025; Revised 12 October 2025; Accepted 20 October 2025

**Abstrak** - Pengelompokan siswa berbasis data sangat penting untuk mendukung evaluasi yang adil dan menyeluruh, mengingat penilaian potensi selama ini cenderung terfokus pada aspek akademik saja. Penelitian ini mengembangkan sistem rekomendasi regu inti lomba kepramukaan menggunakan algoritma *K-Means Clustering*, dengan *dataset* berisi 120 siswa SMP yang dinilai berdasarkan parameter akademik, non-akademik, serta pencapaian SKU dan SKK. Jumlah *cluster* ditentukan sebanyak 24, sesuai dengan kategori lomba berdasarkan aturan Kwarnas mengenai lomba pramuka tingkat penggalang. Proses pengolahan data meliputi normalisasi dan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Evaluasi kualitas *clustering* dilakukan menggunakan metrik *Silhouette Score* dan *Davies–Bouldin Index* (DBI). Hasil terbaik diperoleh pada konfigurasi *random\_state* = 42, *n\_init* = 20, dan *max\_iter* = 300, dengan *Silhouette Score* sebesar 0,1238 dan DBI sebesar 1,4418. Meskipun kualitas pengelompokan tergolong rendah dengan hasil *Silhouette Score* = 0.102 dan DBI = 1.362, sistem ini tetap memberikan solusi objektif bagi pembina dalam memilih siswa berpotensi secara adil dan menyeluruh. Sistem ini juga menjawab keluhan orang tua terkait ketidakterlibatan anak dalam lomba, karena pemilihan dilakukan berdasarkan potensi keseluruhan kategori lomba, bukan hanya satu kategori untuk membentuk tim regu inti pramuka.

Kata kunci: *K-Means Clustering*, *Principal Component Analysis* (PCA), *Silhouette Score*, *Davies–Bouldin Index*, Regu Inti Pramuka.

**Abstract** - Data-based student grouping is very important to support fair and comprehensive evaluation, considering that potential assessment has tended to focus only on academic aspects. This study developed a recommendation system for scouting competition core teams using the *K-Means Clustering* algorithm, with a dataset containing 120 junior high school students assessed based on academic and non-academic parameters, as well as SKU and SKK achievements. The number of clusters was determined to be 24, in accordance with the competition categories based on Kwarnas regulations regarding scout competitions at the scout level. The data processing involved normalization and dimension reduction using *Principal Component Analysis* (PCA). The clustering quality was evaluated using the *Silhouette Score* and *Davies–Bouldin Index* (DBI) metrics. The best results were obtained with the configuration *random\_state* = 42, *n\_init* = 20, and *max\_iter* = 300, with a *Silhouette Score* of 0.1238 and a DBI of 1.4418. Although the clustering quality was relatively low, with a *Silhouette Score* of 0.102 and a DBI of 1.362, this system still provides an objective solution for coaches in selecting potential students fairly and comprehensively. This system also addresses parents'

*complaints regarding their children's lack of involvement in competitions, because the selection is based on the overall potential of the competition categories, not just one category to form a core scout team.*

*Keywords: K-Means Clustering, Principal Component Analysis (PCA), Silhouette Score, Davies–Bouldin Index, Core Scout Team.*

## 1. PENDAHULUAN

Pendidikan merupakan suatu proses yang terencana dan sistematis untuk mengembangkan potensi manusia secara menyeluruh. Pendidikan memiliki peranan yang krusial dalam menciptakan kehidupan yang efektif, meningkatkan keyakinan pada diri sendiri, serta memberikan sumber daya yang diperlukan [1]. Dalam penyelenggaraan proses belajar mengajar, tanggung jawab tidak hanya berada pada institusi sekolah semata, melainkan juga memerlukan keterlibatan aktif dari seluruh aspek agar dapat memberikan kontribusi secara seimbang terhadap kemajuan pendidikan [2]. Tidak sedikit usaha yang dilakukan oleh pihak sekolah serta lembaga pendidikan untuk memperbaiki prestasi siswa di ranah akademis yang esensial untuk mencapai norma pendidikan nasional [3]. Tingkat keberhasilan murid mencerminkan mutu dari sistem pendidikan. Selain itu, sistem informasi saat ini berfungsi sebagai sumber daya yang dapat dimanfaatkan untuk meningkatkan kualitas dan pencapaian siswa [4].

Meskipun demikian, dalam konteks pendidikan di sekolah masih terdapat kesenjangan antara kemampuan akademik dan pendidikan karakter. Pendidikan karakter kurang diperhatikan di lingkungan sekolah dibandingkan kemampuan akademik. Pendidikan karakter bisa dipahami sebagai pendidikan mengenai nilai-nilai, pendidikan tentang akhlak, pendidikan etika, dan pengembangan kepribadian [5]. Oleh karena itu, ekstrakurikuler perlu dimanfaatkan semaksimal mungkin agar dapat berperan sebagai pendukung pendidikan karakter [5][6], salah satunya melalui kegiatan non akademik kepramukaan. Aktivitas kepramukaan sangat efektif dalam mengembangkan karakter dan rasa tanggung jawab siswa [7].

Meskipun penelitian sebelumnya telah mengidentifikasi efektivitas *K-Means* dalam mengidentifikasi kemampuan akademik atau mengelompokkan siswa berprestasi, ada kekurangan penelitian yang cukup besar. Hingga saat ini, belum ada sistem rekomendasi untuk regu inti yang dengan jelas mengintegrasikan parameter multidimensi, seperti kemampuan akademik, non-akademik, bersama dengan pencapaian Syarat Kecakapan Umum (SKU) dan Syarat Kecakapan Khusus (SKK), khususnya dalam lingkungan kepramukaan. Pendekatan yang melibatkan berbagai dimensi ini sangat penting untuk memperoleh penilaian yang lebih menyeluruh mengenai potensi siswa. Penelitian ini bertujuan untuk menutupi kekurangan tersebut dengan menciptakan sistem yang mampu mengklasifikasikan siswa secara objektif berdasarkan kombinasi dari berbagai parameter tersebut.

Di sisi lain, pelaksanaan kegiatan kepramukaan di sekolah diharapkan mampu menghasilkan regu inti yang kompeten dalam berbagai jenis lomba sesuai ketentuan Kwartir Nasional (Kwarnas), yang menetapkan 24 kategori lomba. Namun, proses pembentukan regu inti tersebut menghadapi tantangan yaitu karena jumlah peserta didik yang besar serta keragaman kriteria penilaian yang mencakup aspek akademik, non-akademik, maupun pencapaian Syarat Kecakapan Umum (SKU) dan Syarat Kecakapan Khusus (SKK). Kondisi ini sering kali menyulitkan pembina dalam melakukan pengelompokan siswa secara tepat serta memicu ketidakpuasan dari orang tua terhadap keputusan keterlibatan anaknya dalam kompetisi pramuka. Salah satu pendekatan teknik yang tepat menggunakan pengelompokan *K-Means Clustering* [8].

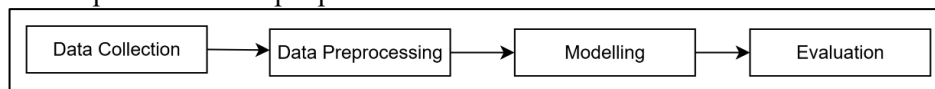
Penelitian terdahulu menunjukkan bahwa pendekatan menggunakan *K-Means* efektif dalam memetakan kemampuan akademik dalam mengoptimalkan pembelajaran [9]. Penelitian lain juga menunjukkan algoritma *K-Means* sangat baik dalam mengatasi pemetaan siswa unggulan dalam jumlah banyak [10]. Dari beberapa referensi jurnal terdapat keterbatasan masalah dan parameter yang digunakan, di beberapa referensi hanya menyertakan 1 parameter saja yaitu akademik atau non-akademik. Penelitian ini menyertakan penggabungan parameter akademik dan non akademik sebagai bahan penilaian multidimensi. Dengan demikian, sistem ini diharapkan mampu membagi siswa secara objektif ke dalam kategori lomba yang sesuai dengan kebutuhan pramuka menggunakan teknologi *data mining*.

Dengan menggunakan teknologi *data mining*, analisis rekomendasi pembentukan regu inti dapat dilakukan. Proses pengelompokan ini menggunakan metode *clustering* yang membagi sekelompok siswa menjadi subset yang dikenal sebagai *cluster* [3]. *Clustering* adalah proses mengelompokkan data dengan tingkat kemiripan yang tinggi antara satu data dengan data lainnya ke dalam kelompok atau *cluster*. Tujuan *clustering* adalah untuk memastikan bahwa data dalam satu *cluster* memiliki tingkat kemiripan yang paling tinggi dan data di antara *cluster* memiliki tingkat kemiripan yang paling rendah [11].

*K-Means* adalah salah satu metode pengelompokan data non hierarki yang dapat membagi data ke dalam dua kelompok atau lebih dengan karakteristik yang sama ke dalam satu kelompok, dan kelompok lain menggabungkan data dengan karakteristik yang berbeda [12].

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *data mining* dengan melibatkan informasi yang sudah ada dalam pusat data untuk dilakukan analisis dan menemukan solusi dari suatu permasalahan. Alur yang digunakan terdapat 4 tahapan yaitu *data collection*, *data preparation*, *modelling*, dan *evaluation*. Berikut alur penelitian penerapan algoritma *K-Means* untuk *clustering* pembentukan rekomendasi regu inti di bidang kepramukaan. Adapun alur penelitian terdapat pada Gambar 1.



**Gambar 1.** Tahapan Penyelesaian Penelitian

Tahapan penelitian pada gambar 1 dilakukan untuk menguji hasil *clustering* pembentukan rekomendasi regu inti pada algoritma *K-Means Clustering*. Pengelompokan nilai siswa dilakukan mulai dari pengumpulan data, pemrosesan data, hingga evaluasi hasil model yang digunakan. Algoritma *K-Means* membutuhkan penentuan nilai *k* yang harus ditetapkan untuk melakukan setiap jenis analisis pengelompokan [13]. Algoritma yang memisahkan individu dalam kumpulan data menjadi beberapa bagian yang umumnya tidak saling terkait disebut sebagai algoritma *clustering* [14].

### 2.1 Data Collection

Pada Data yang digunakan adalah data hasil nilai akademik, non akademik, dan nilai SKU serta SKK dengan jumlah 120 siswa SMP dalam satu sekolah. Data yang digunakan mencakup 13 fitur yang dibagi ke dalam 3 sumber data yaitu tabel nilai non akademik dan nilai akademik yang diambil dari inputan guru dan pembina, serta nilai SKU dan SKK yang diambil dari inputan pembina. Penjabaran sumber data dapat dilihat pada Tabel 1.

**Tabel 1.** Sumber Dataset

Sumber Data	Data yang Diambil	Jenis Data
Nilai Akademik	Nilai per mata pelajaran	Numerik (0-100)
Nilai Non Akademik	Nilai kategori yang dibutuhkan di kepramukaan	Numerik (0-100)
Penilaian SKU dan SKK	Status penyelesaian SKU dan SKK tingkat penggalang	Numerik (0/1)

Nilai akademik diambil dari skor tiap pelajaran (0-100) yang mencerminkan kemampuan kognitif siswa, sementara nilai non akademik mencakup kategori keterampilan kepramukaan (0-100) yang menunjukkan potensi di luar bidang akademik untuk menunjang kepramukaan. Di sisi lain, evaluasi SKU dan SKK dicatat dalam format biner (0/1) untuk menggambarkan status penyelesaian syarat kecakapan umum dan khusus pada tingkat penggalang. Ketiga elemen ini dipilih karena dapat mendefinisikan nilai multidimensi siswa.

## 2.2 Data Preprocessing

Data yang telah dikumpulkan akan melewati tahap pengolahan data agar menghasilkan informasi data yang lebih tepat [15]. Pada proses preprocessing data dilakukan serangkaian proses mulai dari pemilihan fitur, *data cleaning*, *data normalization*, dan reduksi dengan PCA untuk menghasilkan kualitas data yang akurat sesuai [16]. Proses *data preprocessing* memastikan semua datanya lengkap tanpa ada yang kurang atau *null* [17].

### 2.2.1 Feature Selection

Proses pemilihan fitur dilakukan untuk menentukan variabel yang digunakan dalam penelitian [18]. Pemilihan fitur dengan 13 fitur yang terdiri dari 3 jenis data. Pemilihan parameter ini dilakukan dengan pertimbangan bahwa pembentukan regu inti pramuka tidak hanya dipengaruhi oleh prestasi akademik, tetapi juga oleh kemampuan non-akademik yang berkaitan dengan aktivitas kepramukaan, serta tingkat partisipasi siswa dalam memenuhi syarat keterampilan. Penjabaran variabel ditampilkan pada Tabel 2.

**Tabel 2.** Variabel Clustering

Jenis Data	Justifikasi	Tipe Data
Akademik	Matematika	Numerik
	Olahraga	Numerik
	IPA	Numerik
	IPS	Numerik
	Bahasa Indonesia	Numerik
	Bahasa Inggris	Numerik
Non Akademik	Skor Tes Bahasa	Numerik
	Skor Tes TIK	Numerik
	Kehadiran	Numerik
	Skor Penerapan	Numerik
	Hasta Karya	Numerik
Nilai SKU dan SKK	Status SKU	Numerik
	Status SKK	Numerik

Pemilihan fitur yang digunakan untuk proses *clustering* yang mencakup 13 fitur pembentukan regu inti yang kemudian akan dilakukan proses *modelling* dari penentuan *cluster* setiap kategori lomba dengan 3 sumber data yaitu tabel nilai akademik, nilai non akademik, dan nilai SKU dan SKK dengan tipe data numerik.

### 2.2.2 Data Cleaning

Proses *data cleaning* dilakukan untuk mengatasi data yang hilang / tidak lengkap yang menyebabkan data menjadi tidak akurat. *Data cleaning* dilakukan dengan cara

menghilangkan data yang tidak sesuai [19] atau mengisi data yang hilang dengan rata-rata [20] sebelum masuk ke proses *modelling*.

2.2.3 Data Normalization

Proses normalisasi data dilakukan dengan teknik *Standard Scaler* untuk memastikan bahwa setiap fitur memiliki kontribusi yang seimbang dalam proses *clustering*. Metode *Standard Scaler* dipilih karena metode ini efektif dalam mengatasi perbedaan skala antar variabel dalam penelitian ini. Metode ini mengubah data ke dalam distribusi dengan rata-rata 0 dan standar deviasi 1. Hasil normalisasi data dijabarkan pada Gambar 2.

	Bahasa Indonesia	Bahasa Inggris	Hasta Karya	IPA	IPS	Kehadiran	Matematika	Olahraga	Skor Penerapan	Skor Tes Bahasa	Skor Tes TIK	Status SKU	Status TIK
0	-0.151485	0.132046	1.319739	-0.365584	1.203250	1.183442	0.913023	-0.750906	1.582488	1.338617	-0.625842	-0.229416	6.244998
1	-0.221401	-1.199514	-1.197298	0.299114	-1.426927	-0.235337	0.235036	-0.003398	1.649002	-0.281222	-1.210538	4.358899	-0.160128
2	0.407843	0.132046	-0.071794	-0.365584	0.866048	-0.843386	-0.714147	1.083887	0.318715	-0.753675	1.583011	-0.229416	6.244998
3	0.757423	0.664670	0.214698	-1.325704	-1.494368	0.102467	1.184218	-0.886817	0.850830	0.258724	0.348652	4.358899	6.244998
4	-0.990477	-0.001110	0.337481	1.480801	-1.359487	-1.316312	-1.188738	1.695485	1.715516	-1.091142	0.868382	4.358899	-0.160128

Gambar 2. Hasil Normalisasi Data

Metode yang digunakan pada normalisasi data menggunakan *Z-Score Normalization* dengan mengubah standar deviasi untuk masing-masing atribut fitur untuk mengubah skala nilai data [21]. Didapatkan hasil normalisasi dengan nilai skala yang tidak terlalu besar antara setiap variabelnya.

2.2.4 Reduksi PCA

Tahapan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) dilakukan untuk menyederhanakan kompleksitas dari 13 fitur variabel menjadi dua komponen utama. Meskipun data sudah terstruktur dan bersih, penggunaan PCA bertujuan utama untuk mempermudah visualisasi hasil pengelompokan. Dengan melakukan reduksi dimensi data, data siswa dapat diproyeksikan ke dalam *scatter plot* dua dimensi yang mudah diinterpretasi, sehingga pola pengelompokan dapat terlihat lebih jelas dan akurat. Selain itu, PCA membantu menghilangkan korelasi antar fitur, memastikan bahwa setiap dimensi yang digunakan dalam visualisasi bersifat independen.

2.3 Modelling

Pada proses *modelling* menggunakan algoritma *K-Means* dengan jumlah *cluster* sebanyak 24 cluster sesuai dengan jumlah kategori lomba yang ditetapkan oleh Kwartir Nasional untuk penggalang. Dalam penelitian ini, penentuan jumlah *cluster* (k) ditetapkan secara spesifik sebanyak 24, tidak melalui metode *Elbow Method* yang umum digunakan dikarenakan tujuannya untuk memetakan nilai berdasarkan lombanya secara langsung. Justifikasi utama untuk pendekatan ini adalah untuk memetakan setiap siswa secara langsung ke dalam 24 kategori lomba yang telah ditetapkan oleh Kwartir Nasional (Kwarnas) untuk tingkat penggalang. Tujuan kami bukanlah untuk menemukan jumlah *cluster* optimal secara matematis, melainkan untuk menciptakan sebuah model fungsional yang secara langsung mendukung kebutuhan operasional pembentukan regu inti kepramukaan.

Dengan memilih k=24, setiap *cluster* secara langsung merepresentasikan satu jenis lomba, yang memungkinkan sistem untuk memberikan rekomendasi yang relevan dan praktis bagi pembina. Pemilihan jumlah *cluster* secara langsung ini dilakukan dengan tujuan memetakan siswa secara langsung ke dalam setiap kategori lomba. Dengan menentukan jumlah *cluster* ini, setiap kategori kompetisi dapat dipetakan secara langsung ke dalam satu kelompok yang mewakili karakteristik siswa. *Clustering* dilakukan secara iteratif hingga jumlah kebutuhan lomba tercapai dan distribusi siswa stabil di setiap *cluster*.

Selanjutnya, untuk menunjukkan proporsi pemetaan ke masing-masing kategori lomba, hasil pemodelan ditampilkan dalam bentuk jumlah siswa setiap *cluster*. Grafik hasil *clustering* digunakan untuk visualisasi. Pola pengelompokan siswa digambarkan lebih jelas dengan menggunakan plot dispersi dua dimensi berbasis PCA. Oleh karena itu, proses *modelling* tidak hanya menghasilkan pembagian *cluster*, tetapi juga menunjukkan distribusi potensial siswa sesuai bidang lomba. Algoritma pengelompokan menggunakan fondasi distribusi data dan menetapkan kriteria untuk mengatur data berdasarkan sifat yang sebanding [13]. Algoritma *K-Means* merupakan teknik pengelompokan yang berbasis partisi dan banyak dipakai. Dari sisi teori, metode ini memiliki keandalan yang tinggi, mudah diterapkan, cepat dalam konvergensi, serta mampu mengelola data dalam ukuran kecil hingga menengah dengan baik, sehingga sering digunakan [22].

## 2.4 Evaluation

Kualitas hasil pengelompokan dinilai melalui dua metrik internal, yaitu nilai *Silhouette Score* dan Indeks *Davies-Bouldin Index* (DBI) [17][23]. Nilai *Silhouette Score* berfungsi untuk menilai seberapa efektif suatu objek ditempatkan dalam *cluster* yang benar dibandingkan dengan *cluster* terdekat lainnya [20][24], sedangkan DBI bertujuan untuk mengevaluasi rata-rata kesamaan antara *cluster* dengan memperhatikan jarak pusat *cluster* [25][26]. Nilai *Silhouette Score* yang mendekati angka 1 menunjukkan pemisahan *cluster* yang baik, sedangkan nilai DBI yang semakin rendah mendekati 0 menunjukkan kualitas *cluster* yang lebih baik [15][27]. Metrik ini dipilih karena dapat memberi gambaran objektif tentang kualitas pembentukan *cluster* tanpa perlu mengandalkan label kelas sebagai referensi.

## 3. HASIL DAN PEMBAHASAN

Proses *clustering* dalam penelitian ini dilakukan melalui beberapa tahapan, dimulai dari penentuan jumlah *cluster* sesuai jumlah kategori lomba hingga evaluasi kualitas hasil pengelompokan. Evaluasi dilakukan menggunakan metrik *Silhouette Score* dan *Davies Bouldin Index* (DBI) untuk memastikan kualitas pemisahan *cluster*, serta validasi jumlah siswa pada setiap kategori lomba agar sesuai dengan kebutuhan pembentukan regu inti pramuka. Alur lengkap proses *modelling* dapat dilihat pada Gambar 3.



**Gambar 3.** Langkah *Clustering* dengan *K-Means*

Proses dimulai dengan menghitung jumlah *cluster*, menggunakan *Silhouette Score* dan *Davies-Bouldin Index* sebagai metrik evaluasi, dan memetakan *cluster* ke dalam kategori lomba sesuai jumlahnya. Selanjutnya, iterasi pengelompokan dilakukan hingga hasil sesuai dengan kebutuhan distribusi siswa. Hasil akhir ditampilkan dalam bentuk jumlah siswa setiap kelompok dan grafik yang menunjukkan pola pengelompokan dapat dilihat dalam bentuk *scatter plot*. *K-Means* bisa jadi adalah teknik pengelompokan yang paling umum diterapkan, terutama yang paling terkenal di antara teknik pengelompokan yang berbasis pemisahan yang memanfaatkan pusat *cluster* sebagai representasi *cluster* [28].

### 3.1 Penentuan Jumlah Cluster

Menentukan total kluster berdasarkan jenis kompetisi yang diambil dari *database* yang mengacu pada ketentuan Kwarnas mengenai pramuka penggalang menghasilkan 24 jenis lomba. Dari total kategori lomba tersebut, akan dilakukan pengelompokan untuk masing-masing kategori lomba sehingga diperoleh  $k = 24$ . Hasil jumlah data siswa di setiap cluster disesuaikan dengan kebutuhan setiap lomba yang kebutuhannya masing-masing lomba berbeda-beda, dikarenakan dari setiap kebutuhan lomba terdapat jumlah tetapnya berdasarkan aturan dari Kwarnas sehingga hasil akhir jumlah setiap *cluster* tidak bisa beragam atau hanya sesuai dengan kebutuhan siswa di setiap lomba yang telah ditetapkan bersamaan dengan jumlah *cluster* regu inti.

### 3.2 Menghitung Nilai Silhouette Score dan DBI

Menilai kualitas hasil pengelompokan, penelitian ini menggunakan dua metrik evaluasi, yaitu *Silhouette Score* dan *Davies-Bouldin Index* (DBI). *Silhouette Score* digunakan karena mampu mengukur tingkat kedekatan objek dengan *cluster* yang ditempatinya sekaligus keterpisahannya dari *cluster* lain, sedangkan DBI digunakan untuk menilai tingkat kompak dan terpisahnya antar *cluster*. Dalam penelitian ini jumlah *cluster* ditetapkan sebanyak 24 sesuai dengan kategori lomba pada tingkat penggalang yang ditetapkan oleh Kwartir Nasional. Interpretasi rentang nilai *Silhouette Score* yang digunakan sebagai acuan evaluasi diambil dari referensi penelitian [11][20] yang ditunjukkan pada Tabel 2.

**Tabel 2.** Rentang Nilai *Silhouette Score*

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur yang kuat
0.51 – 0.70	Yang beralasan
0.26 – 0.50	lemah
<0.25	Tidak ditemukan struktur yang substansial

Nilai *Silhouette Score* yang tinggi (lebih dari 0.71) menunjukkan struktur *cluster* yang kuat, sedangkan nilai yang rendah (kurang dari 0.25) menunjukkan bahwa tidak ada struktur yang terbentuk. Oleh karena itu, skala nilai ini berfungsi sebagai dasar untuk menilai seberapa efektif hasil pengelompokan dapat membentuk pengelompokan yang signifikan. Dengan mengintegrasikan hasil *Silhouette Score* dengan DBI, evaluasi yang lebih akurat. Hasil pengujian nilai *Silhouette Score* dan DBI bisa dilihat pada Tabel 3.

**Tabel 3.** Perbandingan Nilai Algoritma *K-Means* dan *K-Medoids*

Model	Cluster (k)	Max_iter	Nilai <i>Silhouette Score</i>	Nilai <i>Davies-Bouldin Index</i>
<i>K-Means</i>	24	300	0.102	1.362
<i>K-Medoids</i>	24	300	0.056	1.388

Evaluasi kualitas pengelompokan menunjukkan bahwa kedua model, baik *K-Means* maupun *K-Medoids*, menghasilkan nilai *Silhouette Score* dan *Davies-Bouldin Index* (DBI) yang tergolong rendah. Pada model *K-Means*, nilai *Silhouette Score* adalah 0.102 dan DBI sebesar 1.362, sedangkan *K-Medoids* memiliki nilai *Silhouette Score* 0.056 dan DBI 1.388. Rendahnya nilai ini menunjukkan kualitas pengelompokan yang kurang optimal dan keterpisahan antar *cluster* yang rendah.

Kualitas yang menurun ini disebabkan oleh parameter yang kompleks di dalam *dataset*. Mengingat setiap kategori lomba memiliki variabel penilaian yang berbeda, algoritma *clustering* menghadapi tantangan dalam membentuk kelompok yang padat dan terpisah secara optimal. Selain itu, penetapan jumlah *cluster* yang sangat spesifik ( $k=24$ ) yang didasarkan pada aturan lomba, bukan pada struktur data itu sendiri, memperparah kesulitan algoritma dalam menemukan pemisahan yang ideal. Meskipun demikian, dari hasil pengujian, *K-Means*

tetap menunjukkan performa yang sedikit lebih baik dibandingkan *K-Medoids* dalam kasus ini, terlihat dari nilai *Silhouette Score* yang lebih tinggi dan DBI yang lebih rendah, yang mengindikasikan *cluster* yang lebih kompak.

### 3.3 Pemetaan Cluster ke Setiap Lomba

Pemetaan *cluster* ke setiap lomba ditentukan dalam aturan Kwarnas mengenai lomba pramuka tingkat penggalang yang didapatkan jumlah lomba sebanyak 24 jenis lomba yang di *clusterkan* ke dalam 24 jenis *cluster*. Penentuan ini didasarkan setiap *cluster* lomba memiliki parameter variabel yang berbeda-beda sehingga tidak bisa digabungkan untuk setiap jenis lombanya. Hasil pemetaan lengkap bisa dilihat pada Gambar 4.

Pemetaan Cluster ke Lomba:	
Cluster 0	-> Pionering
Cluster 1	-> Administrasi Regu
Cluster 2	-> Packing Pengembaraan
Cluster 3	-> Semboyan dan Isyarat
Cluster 4	-> Sketsa Panorama
Cluster 5	-> Peta Pita dan Peta Perjalanan
Cluster 6	-> Menaksir
Cluster 7	-> Pertolongan Pertama
Cluster 8	-> Masak Rimba
Cluster 9	-> Bivak
Cluster 10	-> Obat dan Ramuan Tradisional
Cluster 11	-> Baris Berbaris Tongkat
Cluster 12	-> Senam Pramuka
Cluster 13	-> E-Sport
Cluster 14	-> Robotik
Cluster 15	-> Coding
Cluster 16	-> Panjat Dinding
Cluster 17	-> Memanah
Cluster 18	-> Renang
Cluster 19	-> Halang Rintang
Cluster 20	-> Pidato
Cluster 21	-> Melukis Poster
Cluster 22	-> Hasta Karya
Cluster 23	-> Reportase

**Gambar 4.** Hasil Pemetaan *Cluster* Lomba

Pemetaan *cluster* ditentukan oleh perbedaan variabel parameter yang digunakan pada setiap lomba berbeda-beda, sehingga penentuan *cluster* harus di pisah antara *cluster* 1 dengan *cluster* lain. Jika variabel parameter setiap lomba sama jumlah dan kategori parameternya maka dari penentuan *cluster* bisa digabung menjadi beberapa *cluster* kecil yang bisa digunakan bersama.

### 3.4 Melakukan Iterasi Clustering

Proses iterasi dilakukan selama 2 kali dengan melihat pemenuhan kebutuhan siswa dari setiap kategori lomba. Iterasi ke-1 bertujuan mencari nilai siswa dengan kecocokan parameter variabel yang ada di setiap kategori lomba. Pada proses iterasi ke-2 bertujuan mengisi kebutuhan siswa di setiap lomba yang belum terpenuhi dengan mengambil sisa siswa yang tidak tergabung dalam iterasi ke-1. Penjabaran proses iterasi dapat dilihat di Gambar 5.



```

Memulai generasi rekomendasi siswa (Fase 2: Mengisi Kebutuhan Lomba yang Kurang):

--- Iterasi Fase 2: 1 ---
Lomba 'Semboyan dan Isyarat' masih membutuhkan 2 siswa.
Menambahkan 2 siswa ke 'Semboyan dan Isyarat'. Lomba ini sekarang memiliki 7/7 siswa.
Lomba 'Peta Pita dan Peta Perjalanan' masih membutuhkan 3 siswa.
Menambahkan 3 siswa ke 'Peta Pita dan Peta Perjalanan'. Lomba ini sekarang memiliki 7/8 siswa.
Lomba 'Pertolongan Pertama' masih membutuhkan 6 siswa.
Menambahkan 6 siswa ke 'Pertolongan Pertama'. Lomba ini sekarang memiliki 7/8 siswa.
Lomba 'Obat dan Ramuan Tradisional' masih membutuhkan 1 siswa.
Menambahkan 1 siswa ke 'Obat dan Ramuan Tradisional'. Lomba ini sekarang memiliki 2/2 siswa.
Lomba 'Baris Berbaris Tongkat' masih membutuhkan 7 siswa.
Menambahkan 7 siswa ke 'Baris Berbaris Tongkat'. Lomba ini sekarang memiliki 4/8 siswa.
Lomba 'E-Sport' masih membutuhkan 4 siswa.
Menambahkan 4 siswa ke 'E-Sport'. Lomba ini sekarang memiliki 5/5 siswa.

--- Iterasi Fase 2: 2 ---
Lomba 'Peta Pita dan Peta Perjalanan' masih membutuhkan 1 siswa.
Menambahkan 1 siswa ke 'Peta Pita dan Peta Perjalanan'. Lomba ini sekarang memiliki 8/8 siswa.
Lomba 'Pertolongan Pertama' masih membutuhkan 1 siswa.
Menambahkan 1 siswa ke 'Pertolongan Pertama'. Lomba ini sekarang memiliki 8/8 siswa.
Lomba 'Baris Berbaris Tongkat' masih membutuhkan 4 siswa.
Menambahkan 4 siswa ke 'Baris Berbaris Tongkat'. Lomba ini sekarang memiliki 8/8 siswa.

--- Iterasi Fase 2: 3 ---
Semua kebutuhan lomba sudah terpenuhi. Menghentikan iterasi Fase 2.

```

Gambar 5. Iterasi *Clustering*

Iterasi dilakukan dua kali untuk memenuhi kebutuhan rekomendasi regu inti terpenuhi di semua kategori lomba, untuk melihat siswa berpotensi di lebih dari 1 kategori lomba untuk direkomendasikan ke tim regu inti pramuka. Dari hasil iterasi *clustering* didapatkan hasil status pemenuhan kebutuhan yang terpenuhi untuk semua lomba.

### 3.5 Menampilkan Jumlah Siswa Setiap Clusternya

Setelah dilakukan proses iterasi yang kedua akan didapatkan hasil status pengisian lomba yang sudah terpenuhi di semua lomba. Proses ini dilakukan untuk mengecek kebutuhan jumlah siswa dengan jumlah siswa dari hasil iterasi *clustering*. Tujuan penyajian ini adalah untuk mengevaluasi secara menyeluruh tingkat terpenuhinya kebutuhan pada masing-masing lomba. Jumlah terpenuhinya kebutuhan lomba dijabarkan pada Gambar 6.

--- Status Akhir Pengisian Lomba ---				
	Lomba	Kebutuhan	Terisi	Status
0	Pionering	6	6	Terpenuhi
1	Administrasi Regu	2	2	Terpenuhi
2	Packing Pengembaraan	1	1	Terpenuhi
3	Semboyan dan Isyarat	7	7	Terpenuhi
4	Sketsa Panorama	2	2	Terpenuhi
5	Peta Pita dan Peta Perjalanan	8	8	Terpenuhi
6	Menaksir	8	8	Terpenuhi
7	Pertolongan Pertama	8	8	Terpenuhi
8	Masak Rimba	3	3	Terpenuhi
9	Bivak	3	3	Terpenuhi
10	Obat dan Ramuan Tradisional	2	2	Terpenuhi
11	Baris Berbaris Tongkat	8	8	Terpenuhi
12	Senam Pramuka	8	8	Terpenuhi
13	E-Sport	5	5	Terpenuhi
14	Robotik	2	2	Terpenuhi
15	Coding	2	2	Terpenuhi
16	Panjat Dinding	1	1	Terpenuhi
17	Memarah	1	1	Terpenuhi
18	Renang	1	1	Terpenuhi
19	Halang Rintang	1	1	Terpenuhi
20	Pidato	1	1	Terpenuhi
21	Melukis Poster	1	1	Terpenuhi
22	Hasta Karya	4	4	Terpenuhi
23	Reportase	2	2	Terpenuhi

Gambar 6. Jumlah Kebutuhan Lomba

Setiap kategori lomba memiliki status terpenuhi, yang menunjukkan bahwa pemetaan dan *clustering* siswa berhasil membagi peserta sesuai kebutuhan lomba. Hasil ini menunjukkan bahwa metode *K-Means* efektif dalam mendukung proses pengelompokan

dan seleksi siswa pada setiap lomba. Oleh karena itu, sistem yang dibuat dapat memastikan bahwa kapasitas siswa sesuai dengan kebutuhan regu dan memastikan bahwa tidak ada kategori kompetisi yang kekurangan peserta. Setelah semua lomba terpenuhi maka akan dicari 8 siswa dengan potensi lomba di keseluruhan lomba untuk direkomendasikan ke dalam tim regu inti pramuka yang dijabarkan dalam Gambar 7.

--- Daftar 8 Siswa Paling Serbaguna ---

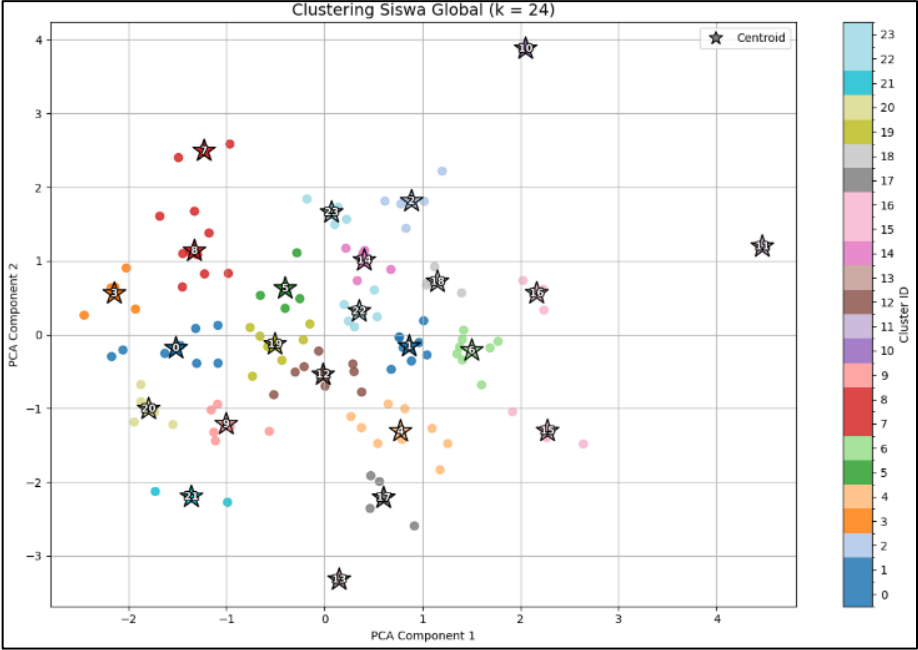
ID Siswa	Nama Siswa	Jumlah Lomba Potensial
46	Ilham Wahyudi	24
115	Maulidini Khasanah	24
62	Uswatun Hasanah	24
106	Aisyah Nursuhayli	24
53	Moh Ilham Alfarizi Ramadhani	24
69	Ainur Rizky Dafa Maulidi	24
117	Moh Alviannur	24
114	Maulidina Khasanah	24

Gambar 7. Jumlah Regu Inti

Siswa dengan rekomendasi tertinggi ditandai dengan seberapa banyak jumlah lomba yang berpotensi pada setiap siswa yang direkomendasikan di dalam tim regu inti kepramukaan. Semakin banyak lomba yang berpotensi, semakin direkomendasikan ke dalam tim regu inti. Keberadaan siswa yang berpotensi ini menunjukkan betapa efektifnya metode *clustering* dalam menemukan siswa yang unggul memiliki keterampilan multidimensi yang membantu komposisi tim menjadi lebih fleksibel dan stabil dalam lomba kepramukaan.

3.6 Menampilkan Hasil Grafik Clustering

Setelah mendapatkan hasil rekomendasi regu inti dalam satu tim sebanyak 8 siswa, untuk melihat lebih detail persebaran hasil clustering disajikan grafik berbentuk *scatter plot* dari 24 *cluster* yang sebelumnya dilakukan reduksi PCA untuk menunjukkan distribusi pemenuhan siswa dalam ruang dua dimensi, memungkinkan interpretasi pola pengelompokan secara gambar. Setiap titik mewakili satu siswa, dan warna yang berbeda menunjukkan hubungan mereka dengan setiap *cluster* kategori lomba. Untuk masing-masing *cluster*, simbol bintang menunjukkan posisi *centroid*, yang merupakan titik pusat karakteristik siswa dalam setiap *cluster*. Hasil grafik *scatter plot* disajikan pada Gambar 8.



**Gambar 8.** Grafik Hasil Rekomendasi Regu Inti

Keberadaan *centroid* yang tersebar luas di antara *cluster* warna menunjukkan bahwa proses *clustering* berhasil menemukan pola kemiripan karakteristik siswa. Hasil ini mendukung pemetaan siswa ke dalam kategori lomba yang sesuai. Hal ini dapat digunakan sebagai dasar untuk saran tentang pembentukan regu inti yang terstruktur dan objektif. Visualisasi ini tidak hanya memperkuat validitas metode *K-Means* untuk mengelompokkan siswa berdasarkan fitur multidimensi, tetapi juga memberikan gambaran praktis tentang kemungkinan penyebaran peserta pada setiap jenis kategori lomba. Pengujian parameter model berpengaruh di dalam proses *clustering* yang ditampilkan pada Tabel 4.

**Tabel 4.** Pengujian Parameter *Clustering*

No	<i>random_state</i>	<i>n_init</i>	<i>max_iter</i>	<i>Silhouette Score</i>	<i>Davies Bouldin Index</i>
1	42	20	300	0,1238	1,4418
2	42	20	500	0,1238	1,4418
3	42	10	300	0,1197	1,4236
4	42	10	500	0,1197	1,4236
5	99	10	300	0,1191	1,5121
6	99	10	500	0,1191	1,5121
7	99	20	300	0,1191	1,5121
8	99	20	500	0,1191	1,5121
9	0	20	300	0,1153	1,4970
10	0	20	500	0,1153	1,4970

Berdasarkan tabel yang menampilkan uji parameter, terlihat bahwa variasi nilai pada parameter *random\_state*, *n\_init*, dan *max\_iter* berdampak pada *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Konfigurasi yang memberikan hasil optimal adalah ketika *random\_state* = 42, *n\_init* = 20, dan *max\_iter* = 300, yang menghasilkan *Silhouette Score* 0.1238 dan DBI 1.4418. Sementara itu, pengaturan dengan *max\_iter* = 500, *random\_state* = 42 dan *n\_init* = 20 menunjukkan hasil yang sama, yang berarti model telah berhasil mencapai konvergensi pada jumlah iterasi yang lebih sedikit. Di sisi lain, pengaturan dengan *random\_state* sama dengan 99 dan 0 menghasilkan *Silhouette Score* dan DBI yang lebih rendah meskipun *n\_init* dan *max\_iter* bervariasi. Hal ini mengindikasikan bahwa *random\_state* memiliki peran penting dalam memengaruhi hasil akhir dari pengelompokan.

Pengaturan parameter memainkan peran penting dalam teknik pengelompokan, terutama *K-Means*. Parameter *random\_state* berfungsi untuk menentukan titik pusat (*centroid*) awal. Karena titik awal yang berbeda dapat memunculkan kelompok yang bervariasi, penetapan *random\_state* yang benar bisa mendukung pencapaian hasil yang lebih konsisten dan efisien. Parameter *n\_init* mengatur berapa kali algoritma *K-Means* harus dijalankan dengan *centroid* awal yang berbeda, dan model akan memilih hasil terbaik dari seluruh percobaan tersebut. Nilai *n\_init* yang lebih tinggi meningkatkan peluang untuk menghasilkan kelompok yang lebih baik. Di sisi lain, *max\_iter* menetapkan batas maksimum jumlah iterasi yang diizinkan bagi algoritma. Walaupun hasil uji menunjukkan bahwa konvergensi dapat tercapai pada jumlah iterasi yang lebih rendah, sangat penting untuk memastikan bahwa nilai ini cukup tinggi agar algoritma dapat menemukan solusi yang stabil. Oleh karena itu, hasil pengujian ini menekankan pentingnya pengoptimalan parameter-parameter tersebut untuk mendapatkan hasil pengelompokan yang ideal, tidak hanya dengan berfokus pada nilai *k* (jumlah *cluster*).

#### 4. KESIMPULAN

Berdasarkan analisis pengelompokan dengan metode *K-Means* yang dilakukan terhadap

120 siswa di tingkat SMP, penerapan metode ini menghasilkan 24 cluster dengan nilai *Silhouette Score* mencapai 0.102 dan *Davies-Bouldin Index* (DBI) sebesar 1.362. Angka-angka ini mencerminkan kualitas pengelompokan yang tergolong rendah serta tingkat pemisahan antara cluster yang tidak ideal.

Rendahnya skor dari ukuran evaluasi ini disebabkan oleh kerumitan dan variasi parameter yang diterapkan dalam contoh kasus ini. Masing-masing kategori kompetisi memiliki variasi jumlah dan jenis variabel penilaian yang beragam, mencakup aspek akademis, non-akademis, serta status pencapaian SKU dan SKK. Karena adanya variasi ini, algoritma *K-Means* mengalami tantangan dalam mengorganisir kelompok yang padat dan terpisah secara ideal.

Meskipun begitu, sistem rekomendasi ini masih memberikan keuntungan nyata bagi para pembina. Sistem ini berperan sebagai sarana yang adil dan objektif untuk membentuk tim inti dengan cara mengenali siswa-siswa yang memiliki potensi terbaik di semua jenis lomba. Ini juga merespons keluhan yang datang dari orang tua karena pemilihan siswa didasarkan pada potensi keseluruhan, bukan hanya di satu jenis lomba saja, sehingga dapat meningkatkan kepercayaan terhadap proses pemilihan dan mengurangi kemungkinan terjadinya konflik.

Hasil penelitian ini menunjukkan bahwa pengujian parameter dalam proses pengelompokan memiliki dampak pada nilai *Silhouette Score* dan DBI yang diperoleh. Konfigurasi yang menggunakan *random\_state* = 42, *n\_init* = 20, dan *max\_iter* = 300 menghasilkan performa terbaik, dengan *Silhouette Score* mencapai 0.1238 dan DBI sebesar 1.4418. Dengan demikian, sangat penting untuk mengoptimalkan parameter seperti *random\_state* dan *n\_init* guna memperoleh hasil pengelompokan yang lebih baik pada penerapan *K-Means*.

Untuk penelitian selanjutnya, dengan data parameter yang rumit dan tidak konsisten, disarankan untuk tetap memanfaatkan algoritma *K-Means*, tetapi dengan beberapa modifikasi. Melihat hasil perbandingan, *K-Means* menunjukkan kinerja yang lebih unggul dibandingkan *K-Medoids* dalam situasi ini, dengan nilai *Silhouette Score* yang sedikit lebih baik dan DBI yang sedikit lebih rendah. Kinerja yang lebih baik ini kemungkinan dikarenakan sifat *K-Means* yang peka terhadap perubahan kecil dalam data, yang sesuai dengan struktur multidimensi di penelitian ini.

## DAFTAR PUSTAKA

- [1] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, 2024, doi: 10.1155/2024/4067721.
- [2] I. Purwaningsih, O. Oktariani, L. Hernawati, R. Wardarita, and P. I. Utami, "Pendidikan Sebagai Suatu Sistem," *J. Vision. Penelit. dan Pengemb. dibidang Adm. Pendidik.*, vol. 10, no. 1, p. 21, 2022, doi: 10.33394/vis.v10i1.5113.
- [3] D. O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [4] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 71–83, 2024, doi: 10.53842/juki.v6i1.474.
- [5] C. Budiyanto, "Manajemen pendidikan kepramukaan dalam pembentukan karakter," *AL-IDRAK J. Pendidik. Islam dan Budaya*, vol. 1, no. 1, pp. 27–45, 2021.
- [6] O. Arifudin, "Optimalisasi Kegiatan Ekstrakurikuler dalam Membina Karakter Peserta Didik," *JiIP - J. Ilm. Ilmu Pendidik.*, vol. 5, no. 3, pp. 829–837, 2022, doi: 10.54371/jiip.v5i3.492.
- [7] M. S. Ummah, "Jurnal Metode dan Teknik kepramukaan," *Sustain.*, vol. 11, no. 1, pp. 1–

- 14, 2019, [Online]. Available: [http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484\\_SISTEM\\_PEMBETU\\_NGAN\\_TERPUSAT\\_STRATEGI\\_MELESTARI](http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETU_NGAN_TERPUSAT_STRATEGI_MELESTARI)
- [8] M. P. Utami, G. M. Rahma, and F. Suroso, "Implementasi K-Means Clustering untuk Analisis Non-Numerik Dataset Spare Part Mobil," *Techno.Com*, vol. 23, no. 2, pp. 306–318, 2024, doi: 10.62411/tc.v23i2.9446.
- [9] I. Maulana and U. Rosalina, "Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means," *PERISKOP J. Sains dan Ilmu Pendidik.*, vol. 1, no. 2, pp. 76–85, 2020, doi: 10.58660/periskop.v1i2.10.
- [10] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [11] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.
- [12] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [13] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [14] M. Zhang, "Use Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm to Identify Galaxy Cluster Members," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 252, no. 4, 2019, doi: 10.1088/1755-1315/252/4/042033.
- [15] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [16] I. Amirulloh, "Pemetaan Kelompok Kerja Siswa Dengan Metode Clustering K-Means Dan Algoritma Greedy," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 94–98, 2019, doi: 10.36499/jinrpl.v1i2.2953.
- [17] A. A. D. Sulistiyawati and M. Sadikin, "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan," *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: 10.32520/stmsi.v10i3.1332.
- [18] S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.
- [19] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.
- [20] B. N. Sari and A. Primajaya, "Penerapan Clustering DbSCAN Untuk Pertanian Padi Di Kabupaten Karawang," *J. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2019, [Online]. Available: [www.mapcoordinates.net/en](http://www.mapcoordinates.net/en).
- [21] "View of Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation.pdf."
- [22] Y. Zhao and X. Zhou, "K-means Clustering Algorithm and Its Improvement Research," *J. Phys. Conf. Ser.*, vol. 1873, no. 1, 2021, doi: 10.1088/1742-6596/1873/1/012074.
- [23] E. Supriyati *et al.*, "Peningkatan Kinerja K-Means Clustering pada Data Penggunaan ChatGPT oleh UMKM," *Techno COM*, vol. 24, no. 1, pp. 81–90, 2025, [Online].

- Available: <https://publikasi.dinus.ac.id/index.php/technoc/article/view/11838>
- [24] B. J. M. Putra and D. A. F. Yuniarti, "Analisis Gaya Belajar terhadap Nilai Mahasiswa dengan Menggunakan Metode k-Means," *Techno.Com*, vol. 21, no. 2, pp. 343–354, 2022, doi: 10.33633/tc.v21i2.5837.
  - [25] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020, doi: 10.1088/1757-899X/725/1/012128.
  - [26] A. Lesmana and W. Gunawan, "Perbandingan Algoritma K-Means dan K-Medoids Dalam Penclusteran Data Penjualan PT.," *Techno.COM*, vol. 21, no. 3, pp. 455–461, 2022.
  - [27] Y. Jumaryadi, A. Wibowo, and M. Ilmu Komputer, "Analisis Pengelompokan Gangguan TIK Pada Sistem Pencatatan Layanan Menggunakan Algoritma K-Means dan Metode Elbow Analysis of Incidents Segmentation Based on Service Recording System Data Using K-Means Algorithm and Elbow Method," *Techno.COM*, vol. 22, no. 2, pp. 348–357, 2023.
  - [28] E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *Int. J. Adv. Sci. Res. Eng.*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.