

Klasterisasi Sekolah Menggunakan Algoritma K-Means dan K-Medoids Berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik

Ridwan Nillah

Universitas Buana Perjuangan Karawang
Karawang, Indonesia
If17.ridwannillah@mhs.ubpkarawang.ac.id ,
085157226489

Sutan Faisal

Universitas Buana Perjuangan Karawang
Karawang, Indonesia
sutan.faisal@ubpkarawang.ac.id

Santi Arum Puspita Lestari

Universitas Buana Perjuangan Karawang
Karawang, Indonesia
santi.arum@ubpkarawang.ac.id

Abstrak— Kualitas pendidikan dipengaruhi oleh fasilitas, pendidik, dan tenaga pendidik, yang merupakan indikator penting dalam pencapaian pembelajaran di sekolah. Penelitian ini bertujuan membantu Pemerintah Kabupaten Karawang dalam mengelompokkan sekolah untuk distribusi kebutuhan yang lebih merata menggunakan algoritma K-Means dan K-Medoids dengan dua klaster. Hasil analisis menunjukkan bahwa K-Means membagi sekolah menjadi 32 sekolah dalam klaster 1 dan 140 sekolah dalam klaster 2, sementara K-Medoids membagi menjadi 85 sekolah dalam klaster 1 dan 87 sekolah dalam klaster 2. Nilai Davies-Bouldin Index adalah -0,648 untuk K-Means dan -3,027 untuk K-Medoids. Dari analisis tersebut, 140 sekolah termasuk dalam klaster kurang menurut K-Means, dan 87 sekolah menurut K-Medoids, sehingga disarankan agar Dinas Pendidikan memberikan perhatian khusus pada sekolah-sekolah dalam klaster kurang ini.

Kata kunci — data mining, K-Means, K-Medoids, klasterisasi

I. PENDAHULUAN

Pendidikan merupakan fondasi utama dalam pembangunan suatu bangsa, di mana kualitasnya menjadi indikator kemajuan negara [6]. Di Indonesia, persoalan ketimpangan kualitas pendidikan masih menjadi tantangan besar, khususnya dalam hal pemerataan fasilitas serta distribusi pendidik dan tenaga kependidikan [8]. Ketidakseimbangan ini berdampak pada capaian pembelajaran dan kualitas sekolah secara keseluruhan.

Fasilitas pendidikan dan ketersediaan guru memainkan peran penting dalam mendukung proses pembelajaran yang efektif [5]. Di Kabupaten Karawang, berdasarkan data Dapodik, tercatat 172 sekolah menengah (SMA dan SMK) berada di bawah tanggung jawab Dinas Pendidikan setempat. Namun, distribusi sarana, prasarana, dan tenaga pendidik di wilayah ini belum merata.

Seiring dengan berkembangnya teknologi informasi, data mining telah digunakan secara luas untuk membantu pengambilan keputusan berbasis data dalam sektor pendidikan. Salah satu pendekatan yang banyak diterapkan adalah algoritma clustering, khususnya K-Means dan K-Medoids, yang mampu mengelompokkan data pendidikan berdasarkan karakteristik tertentu.

Penelitian terdahulu menunjukkan bahwa algoritma K-Means sering kali memberikan hasil clustering yang lebih baik dibandingkan K-Medoids dalam berbagai konteks, seperti prediksi pembelian kredit [15], pemetaan distribusi guru [5], dan pengelompokan sekolah berdasarkan nilai ujian nasional [10]. Meskipun demikian, efektivitas masing-masing algoritma tetap kontekstual terhadap jenis data dan tujuan analisis.

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dan K-Medoids dalam mengklasterisasi sekolah menengah di Kabupaten Karawang berdasarkan data fasilitas, pendidik, dan tenaga pendidik. Selain itu, penelitian ini membandingkan performa kedua algoritma dari segi akurasi dan efektivitas, serta memberikan rekomendasi berbasis data guna mendukung kebijakan pemerataan pendidikan.

II. DATA DAN METODE

A. Pengumpulan Data

Dataset penelitian ini diperoleh dari website resmi Kementerian Pendidikan dan Kebudayaan (Kemdikbud) melalui <https://dapo.dikdasmen.kemdikbud.go.id>. Data yang diambil merupakan data pokok pendidikan yang memiliki karakteristik serupa berdasarkan fasilitas, pendidik, dan tenaga pendidik di Kabupaten Karawang. Data yang digunakan adalah data tingkat SMA dan SMK pada tahun ajaran 2023/2024, yang akan dibagi menjadi dua klaster yaitu “baik” dan “kurang”.

B. Fasilitas

Fasilitas dalam konteks pendidikan merujuk pada semua sarana dan prasarana yang tersedia di lingkungan sekolah untuk mendukung proses pembelajaran dan aktivitas siswa. Ini meliputi ruang kelas yang nyaman dan dilengkapi dengan teknologi pendidikan modern, laboratorium untuk eksperimen ilmiah, perpustakaan dengan koleksi buku dan sumber belajar yang beragam, fasilitas olahraga seperti lapangan dan gymnasium, serta ruang komputer untuk mengembangkan keterampilan teknologi. Fasilitas

ini tidak hanya mendukung aspek akademis, tetapi juga memfasilitasi pengembangan keterampilan sosial, seni, dan olahraga siswa, menciptakan lingkungan belajar yang lengkap dan mendukung pertumbuhan mereka secara menyeluruh.

C. Pendidik

Pendidik adalah sosok kunci dalam dunia pendidikan yang bertanggung jawab tidak hanya untuk mentransfer pengetahuan, tetapi juga membimbing dan menginspirasi siswa untuk mengembangkan potensi mereka secara maksimal. Mereka memiliki peran penting dalam menciptakan lingkungan belajar yang mendukung dan merangsang minat siswa terhadap berbagai bidang studi. Sebagai pemimpin intelektual di dalam kelas, pendidik tidak hanya mengajar materi kurikuler, tetapi juga mendukung perkembangan keterampilan sosial, emosional, dan kritis siswa, serta membantu mereka mengatasi tantangan akademis dan nonakademis. Pendidik yang efektif adalah mereka yang tidak hanya memiliki pengetahuan yang mendalam dalam bidang akademisnya, tetapi juga memahami kebutuhan individu siswa dan mampu mengadaptasi metode pengajaran sesuai dengan gaya belajar mereka. Mereka berperan sebagai panutan positif, membantu siswa membangun nilai-nilai moral dan etika, serta mengajarkan keterampilan penting seperti pemecahan masalah, kerja sama tim, dan komunikasi efektif. Melalui interaksi harian di dalam dan luar kelas, pendidik memainkan peran yang tak tergantikan dalam membentuk masa depan generasi muda, mempersiapkan mereka untuk berkontribusi secara positif dalam masyarakat global yang kompleks.

D. Tenaga Pendidik

Tenaga pendidik merupakan kelompok profesional yang mendukung operasional dan pengelolaan sekolah serta memberikan layanan pendidikan kepada siswa di luar lingkungan kelas. Mereka terdiri atas berbagai peran, termasuk staf administrasi, konselor, psikolog sekolah, dan spesialis pendidikan khusus. Staf administrasi sekolah, seperti kepala sekolah dan manajer administrasi, bertanggung jawab atas pengelolaan harian, keuangan, dan infrastruktur sekolah untuk memastikan operasional yang lancar. Sementara itu, konselor dan psikolog sekolah membantu siswa dalam pengembangan pribadi, sosial, dan akademik dengan memberikan layanan konseling, penilaian psikologis, dan dukungan yang dibutuhkan. Tenaga pendidik juga mencakup spesialis pendidikan khusus yang membantu siswa dengan kebutuhan khusus, seperti disabilitas atau kebutuhan belajar lainnya, untuk mencapai potensi mereka secara penuh. Mereka bekerja sama dengan guru dan orang tua untuk menyusun program pendidikan yang sesuai dan memberikan dukungan tambahan yang diperlukan. Secara keseluruhan, peran tenaga pendidik di sekolah sangat penting karena mereka tidak hanya mendukung pengelolaan sekolah yang efisien, tetapi juga memberikan layanan khusus yang mendukung perkembangan holistik siswa di semua tingkatan.

E. Klasterisasi

Klasterisasi merupakan salah satu metode dalam data mining yang bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan di antara objek-objeknya. Objek yang berada dalam satu klaster memiliki karakteristik yang serupa, sedangkan antar-klaster berbeda secara signifikan [1]. Proses ini memerlukan pengukuran kedekatan (proximity measure), yang umumnya dihitung menggunakan jarak Euclidean, Manhattan, atau Minkowski.

F. Algoritma K-Means

Algoritma K-Means merupakan salah satu metode klasterisasi yang paling umum digunakan karena kesederhanaannya dalam membagi data ke dalam kelompok. Teknik ini bekerja dengan cara membagi dataset menjadi sejumlah klaster yang telah ditentukan sebelumnya [14]. Setiap klaster berisi data yang memiliki kemiripan tinggi satu sama lain. Salah satu keunggulan algoritma ini adalah kemampuannya dalam meminimalkan jarak antara data dan pusat klasternya.

Prosedur algoritma K-Means terdiri dari beberapa tahapan sebagai berikut:

- 1) Tentukan jumlah klaster yang diinginkan (c).
- 2) Pilih secara acak titik awal sebagai pusat klaster (centroid).
- 3) Hitung jarak antara setiap data dan pusat klaster menggunakan rumus Euclidean Distance, seperti terlihat pada Persamaan (1):

$$D_{i,j} = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ip} - X_{jp})^2} \quad (1)$$

Keterangan:

$D_{i,j}$ = jarak *Euclidean* antara titik

X_i = nilai koordinat titik pertama

X_j = nilai koordinat titik kedua

- 4) Kelompokkan data ke dalam klaster berdasarkan jarak terdekat ke pusat klaster.

- 5) Hitung ulang posisi pusat klaster dengan menghitung rata-rata dari semua data dalam klaster menggunakan rumus (2):

$$v = \frac{\sum_{p=1}^n x_i}{n}; p = 1, 2, 3, \dots, n \quad (2)$$

- 6) Ulangi proses dari langkah ketiga hingga posisi klaster tidak lagi berubah, menandakan bahwa proses iterasi telah mencapai konvergensi.

G. Algoritma K-Medoids

K-medoids adalah salah satu teknik klasterisasi yang digunakan untuk menentukan pusat klaster berdasarkan medoid, yaitu data aktual yang paling representatif dalam suatu kelompok [11]. Metode ini dianggap cukup sederhana dan mudah dipahami, serta banyak diterapkan dalam menyelesaikan permasalahan klasifikasi atau pengelompokan data [2].

Tahapan pelaksanaan algoritma K-Medoids adalah sebagai berikut:

- 1) Tentukan jumlah klaster k dan pilih secara acak titik awal sebagai medoid.
- 2) Hitung jarak antara setiap data dengan semua medoid menggunakan rumus Manhattan Distance, sebagaimana terlihat dalam Persamaan (3):

$$d_{\text{man}}(x, y) = \sum_{i=1}^d |x_i - y_i| \quad (3)$$
- 3) Secara acak, pilih kandidat baru sebagai medoid dari data yang ada dalam setiap klaster.
- 4) Hitung kembali jarak dari semua data ke medoid yang baru dipilih.
- 5) Evaluasi perubahan total jarak (cost). Jika total jarak dari medoid baru lebih kecil dibanding sebelumnya (yakni jika $S = \text{jarak baru} - \text{jarak lama} < 0$), maka medoid akan diperbarui dengan data tersebut.
- 6) Ulangi proses dari langkah ketiga hingga tidak ada lagi perubahan medoid, menandakan proses telah mencapai kondisi stabil.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data merupakan tahap krusial dalam penelitian klasterisasi sekolah yang menggunakan algoritma K-Means dan K-Medoids. Data yang dikumpulkan meliputi informasi tentang fasilitas, pendidik, dan tenaga pendidik di berbagai sekolah. Fasilitas mencakup keberadaan laboratorium, perpustakaan, dan ruang kelas yang mendukung proses pembelajaran. Data tentang pendidik mencakup jumlah guru yang mengajar mata pelajaran masing-masing. Sementara itu, informasi mengenai tenaga pendidik mencakup jumlah staf administratif dan teknis yang mendukung kegiatan sekolah.

Pada tahap pertama ini, penguji memulai pengumpulan data sekolah di website <https://dapo.dikdasmen.kemdikbud.go.id> dibantu oleh kenalan operator di salah satu sekolah SMK di Karawang. Data yang didapatkan merupakan data sekolah jenjang SMA/SMK yang berada di wilayah Kabupaten Karawang dengan jumlah 172 record dan 17 atribut. Data yang didapat berupa file berekstensi .xslm. Data yang dikumpulkan kemudian diproses (selections dan transformations) untuk digunakan sebagai input untuk algoritma klasterisasi. Data selengkapnya bisa dilihat pada Tabel 4.1 berikut:

Tabel 4. 1 Dataset Sebelum Tahap *Selection*

Nama Sekolah	Jenis Sekolah	Status	Peserta Didik	Rombel	Perpustakaan
SMKN 1 KARAWANG	SMK	Negeri	2,589	73	1
SMKN 2 KARAWANG	SMK	Negeri	2,109	60	1
SMKN 1 CIKAMPEK	SMK	Negeri	1,904	54	3
SMKN 3 KARAWANG	SMK	Negeri	1,799	51	1
SMA BAITUL ULYA BOARDING SCHOOL	SMA	Swasta	0	0	0

Tabel ini berfungsi sebagai dasar untuk klasterisasi menggunakan algoritma K-Means dan K-Medoids. Algoritma ini akan mengelompokkan sekolah-sekolah berdasarkan kesamaan dalam jumlah fasilitas, pendidik, dan tenaga pendidik. Klaster-klaster yang terbentuk kemudian dapat dianalisis untuk mengidentifikasi pola-pola dalam distribusi sumber daya pendidikan dan untuk merancang kebijakan yang lebih baik dalam meningkatkan kualitas pendidikan.

B. Selection

Selection data yaitu tahap di mana pemilihan atribut dilakukan untuk memfokuskan data hanya pada atribut-atribut yang diperlukan saja. Adapun atribut yang akan digunakan penulis yaitu sebagai berikut:

1. Nama Sekolah
2. Jenis Sekolah
3. Status Sekolah
4. Jumlah Guru Pengajar Matematika
5. Jumlah Guru Pengajar Bahasa Indonesia
6. Jumlah Guru Pengajar Bahasa Inggris

7. Jumlah Guru Pengajar PJOK
8. Jumlah Guru Pengajar Pendidikan Agama
9. Jumlah Guru Pengajar Pendidikan Pancasila dan Kewarganegaraan
10. Jumlah Guru Pengajar Ilmu Pengetahuan Alam
11. Jumlah Guru Pengajar Ilmu Pengetahuan Sosial
12. Jumlah Tenaga Pendidik (Tendik)
13. Jumlah Fasilitas Kelas
14. Jumlah Fasilitas Laboratorium
15. Jumlah Fasilitas Perpustakaan

Atribut tersebut selanjutnya akan diproses ke tahap *transformation* yang nantinya akan merubah data menjadi data yang siap untuk di proses.

Tabel 4. 2 Data Sebelum Tahap Transformation

Nama Sekolah	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
SMKN 1 KARAWANG	SMK	Negeri	16	15	1
SMKN 2 KARAWANG	SMK	Negeri	15	14	1
SMKN 1 CIKAMPEK	SMK	Negeri	11	11	3
SMKN 3 KARAWANG	SMK	Negeri	10	10	1
SMA BAITUL ULYA BOARDING SCHOOL	SMA	Swasta	0	0	0

Tabel ini memberikan gambaran dasar tentang data yang dipilih untuk klusterisasi dan membantu dalam memahami bagaimana variabel yang berbeda (fasilitas, pendidik, tenaga pendidik) diintegrasikan dalam analisis. Dengan data yang dipilih secara hati-hati, analisis klusterisasi dapat menghasilkan wawasan yang berharga tentang bagaimana sekolah dikelompokkan berdasarkan sumber daya mereka, yang dapat mendukung perencanaan dan pengambilan keputusan dalam kebijakan pendidikan.

C. Transformation

Tahapan transformation data merupakan tahapan untuk mengubah nilai dari bentuk polinomial menjadi integer, dan tahapan ini merupakan akhir dari proses persiapan data. Data yang ditransformasikan yaitu data Jenis Sekolah yang nilainya berupa data polinomial SMA/SMK menjadi 1 dan 2, serta data Status yang nilainya berupa data polinomial Negeri/Swasta menjadi 1 dan 2.

Tabel 4. 3 Dataset Setelah Transformations

Nama Sekolah	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
SMKN 1 KARAWANG	2	1	16	15	1
SMKN 2 KARAWANG	2	1	15	14	1
SMKN 1 CIKAMPEK	2	1	11	11	3
SMKN 3 KARAWANG	2	1	10	10	1
SMA BAITUL ULYA B S	1	2	0	0	0

D. Data Mining

Pada tahap ini, dataset yang telah melalui proses transformasi data akan diklusterisasi menggunakan algoritma K-Means dan K-Medoids.

a) Inisialisasi Pusat Klaster

Tahap pertama dalam perhitungan algoritma K-Means adalah menentukan pusat klaster secara acak untuk menghitung jarak data point selama proses klusterisasi. Pada penelitian ini, jumlah klaster ditentukan menjadi 2, yaitu C1 dan C2, dengan masing-masing data yang dipilih secara acak sebagai pusat klaster. Penentuan pusat klaster ini dilakukan dengan mengambil data SMKN 3 Karawang untuk klaster 1 (C1), dan data SMAN 3 Cikampek untuk klaster 2 (C2). Pusat klaster dapat diketahui pada Tabel 4.4.

Tabel 4. 4 Pusat Klaster Awal

	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
C1	2	1	10	10	1
C2	1	1	8	9	1

Setelah ditentukan pusat klaster, maka langkah selanjutnya adalah menghitung jarak setiap data ke pusat klaster menggunakan rumus Euclidean Distance.

b) Hitung Jarak Pusat Klaster

Pada tahap kedua ini, dilakukan proses perhitungan jarak antara data point dan pusat klaster yang akan menjadi kunci untuk menentukan klaster yang tepat untuk setiap data point. Jarak ini dihitung menggunakan rumus Euclidean Distance pada setiap data point untuk dimasukkan ke klaster yang paling dekat dengan pusat klaster, yaitu:

Jarak data ke pusat klaster C1:

$$(1,1) = \sqrt{(2-2)^2 + (1-1)^2 + (16-10)^2 + (15-10)^2 + \dots + (1-1)^2} \\ = 60,324$$

$$(3,1) = \sqrt{(2-2)^2 + (1-1)^2 + (11-10)^2 + (11-10)^2 + \dots + (3-1)^2} \\ = 21,748$$

$$(172,1) = \sqrt{(1-2)^2 + (2-1)^2 + (0-10)^2 + (0-10)^2 + \dots} \\ + (0-1)^2 = 54,891$$

Jarak data ke pusat C2:

$$(1,2) = \sqrt{(2-1)^2 + (1-1)^2 + (16-8)^2 + (15-9)^2 + \dots + (1-1)^2} \\ = 87,960$$

$$(3,2) = \sqrt{(2-1)^2 + (1-1)^2 + (11-8)^2 + (11-9)^2 + \dots + (1-1)^2} \\ = 49,487$$

$$(172,2) = \sqrt{(1-2)^2 + (2-1)^2 + (0-8)^2 + (0-9)^2 + \dots} \\ + (0-1)^2 = 27,221$$

Hasil dari perhitungan jarak antara pusat klaster dengan tiap titik objek dapat dilihat pada Tabel 4.5.

Tabel 4. 5 Menghitung Jarak dari Pusat Klaster Awal

Nama Sekolah	Jenis Sekolah	Status	C1	C2	Klaster
SMKN 1 KARAWANG	2	1	60,324	87,960	1
SMKN 2 KARAWANG	2	1	26,381	52,287	1
SMKN 1 CIKAMPEK	2	1	21,748	49,487	1
SMKN 3 KARAWANG	2	1	0	30,951	1
SMA BAITUL ULYA BOARDING SCHOOL	1	2	54,891	27,221	2

Setelah berhasil menghitung dan memasukkan data point ke pusat klaster terdekat, maka langkah selanjutnya adalah membuat pusat klaster baru dengan mencari rata-rata dari data point yang berada dalam klaster tersebut. Perhitungan untuk mencari pusat klaster baru pada iterasi ke-1 dapat dilihat berikut.

1. Perhitungan pusat klaster C1 baru:

$$\text{Jenis sekolah} = \frac{2+2+2+\dots+1}{12} = 1,5$$

$$\text{Status} = \frac{1+1+1+\dots+2}{12} = 1,08$$

$$\text{Guru Matematika} = \frac{16+15+11+10+\dots+6}{12} = 10,84$$

$$\text{Guru Bahasa Indonesia} = \frac{15+14+11+10+\dots+6}{12} = 11$$

$$\text{Perpustakaan} = \frac{1+1+3+1+\dots+1}{12} = 1$$

2. Perhitungan pusat kluster C2 baru:

$$\text{Jenis sekolah} = \frac{2+2+2+\dots+1}{160} = 1,7$$

$$\text{Status} = \frac{1+1+1+\dots+2}{160} = 1,76$$

$$\text{Guru Matematika} = \frac{8+6+8+8+\dots+0}{160} = 3,35$$

$$\text{Guru Bahasa Indonesia} = \frac{8+12+7+7+\dots+0}{160} = 3,70$$

$$\text{Perpustakaan} = \frac{2+1+2+1+\dots+0}{160} = 0,49$$

Setelah nilai rata-rata dari data dalam masing-masing kluster dihitung, maka pusat kluster yang baru dapat diidentifikasi sebagaimana ditampilkan pada tabel 4.6.

Tabel 4. 6 Pusat Kluster Baru Iterasi Ke-1

	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
C1	1,5	1,08	10,84	11	1
C2	1,7	1,76	3,35	3,70	0,49

Dari hasil pusat kluster yang didapatkan pada tabel 4.6, kemudian dilakukan kembali perhitungan tiap-tiap objek menggunakan pusat kluster baru pada iterasi ke-1 yang telah ditentukan sebelumnya. Hasil perhitungan jarak pada pusat kluster baru dapat diidentifikasi sebagaimana ditampilkan pada tabel 4.23.

Tabel 4. 7 Data Kluster Iterasi Ke-5

Nama Sekolah	Jenis Sekolah	Status	C1	C2	Kluster
SMKN 1 KARAWANG	2	1	70,12	103,9	1
SMKN 2 KARAWANG	2	1	38,04	66,42	1
SMKN 1 CIKAMPEK	2	1	32,35	65	1
SMKN 3 KARAWANG	2	1	19,7	45,78	1
SMA BAITUL ULYA BOARDING SCHOOL	1	2	43,6	9,73	2

Pada perhitungan iterasi ke-6 ini, tidak ada lagi data yang berpindah dari satu kluster ke kluster yang lain, maka tidak perlu dilakukan iterasi lagi dan proses data mining selesai.

Hasil dari total 172 data yang ada, dengan perhitungan manual algoritma K-Means, terdapat 32 data sekolah yang masuk dalam kluster baik (C1) dan 140 data yang masuk dalam kluster kurang (C2). Hasil ini didapatkan setelah menjalankan sebanyak 6 pengulangan hingga tidak ada perubahan pada kluster di setiap data.

c) Perhitungan Manual Algoritma K-Medoids

1. Inisialisasi Medoid

Tahap pertama dalam perhitungan algoritma K-Medoids adalah menentukan medoid secara acak untuk menghitung jarak data poin selama proses klusterisasi. Pada penelitian ini jumlah kluster ditentukan menjadi 2, yaitu M1 dan M2, dengan masing-masing data yang dipilih secara acak sebagai medoid. Penentuan pusat kluster ini dilakukan dengan mengambil data SMA Raudhatul Irfan untuk kluster 1 (M1), dan data SMAS Bhinneka Karawang untuk kluster 2 (M2). Medoid dapat diketahui pada tabel 4.24.

Tabel 4. 8 Medoid Awal

	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
M1	1	2	0	1	0
M2	1	2	1	1	1

Setelah ditentukan medoid, maka langkah selanjutnya adalah menghitung jarak setiap data ke medoid menggunakan rumus Manhattan Distance.

2. Hitung Jarak Medoid

Pada tahap kedua ini dilakukan proses perhitungan jarak antara data poin dan medoid, yang akan menjadi kunci untuk menentukan kluster yang tepat untuk setiap data poin. Jarak ini dihitung menggunakan rumus Manhattan Distance pada setiap data poin untuk dimasukkan ke kluster yang paling dekat dengan medoid.

Jarak data ke medoid M1:

$$(1,1) = \sqrt{|2-1| + |1-2| + |16-0| + |15-1| + \dots + |1-0|} = 264$$

$$(3,1) = \sqrt{|2-1| + |1-2| + |11-0| + |11-1| + \dots + |3-0|} = 182$$

$$(172,1) = \sqrt{|1-1| + |2-2| + |0-0| + |0-1| + \dots + |0-0|} = 2$$

Jarak data ke medoid M2:

$$(1,1) = \sqrt{|2-1| + |1-2| + |16-1| + |15-1| + \dots + |1-1|} = 250$$

$$(3,1) = \sqrt{|2-1| + |1-2| + |11-1| + |11-1| + \dots + |3-1|} = 168$$

$$(172,1) = \sqrt{|1-1| + |2-2| + |0-1| + |0-1| + \dots + |0-1|} = 14$$

Hasil dari perhitungan jarak antara medoid dengan tiap titik objek dapat dilihat pada Tabel 4.25.

Tabel 4. 9 Menghitung Jarak Ke Medoid Iterasi Awal

Nama Sekolah	Jenis Sekolah	Status	M1	M2	Kluster
SMKN 1 KARAWANG	2	1	264	250	2
SMKN 2 KARAWANG	2	1	183	169	2
SMKN 1 CIKAMPEK	2	1	182	168	2
SMKN 3 KARAWANG	2	1	147	133	2
SMA BAITUL ULYA BOARDING SCHOOL	1	2	2	14	1

Setelah setiap data poin dialokasikan ke medoid terdekat, tahap berikutnya adalah menghitung nilai cost untuk masing-masing kluster berdasarkan jarak yang telah diperoleh dari medoid awal. Sebagai ilustrasi, perhitungan cost pada sekolah SMKN 1 Karawang dapat dijelaskan sebagai berikut:

$$\text{Cost (SMKN 1 Karawang)} = \min\{264, 250\} = 250$$

Hasil perhitungan cost minimal data SMKN 1 Karawang adalah 250, dan perhitungan dilanjutkan pada semua data yang ada. Hasil perhitungan untuk semua data pada medoid awal dapat dilihat pada Tabel 4.26.

Tabel 4. 10 Menghitung Cost Medoid Awal

Nama Sekolah	Jenis Sekolah	M1	M2	Cost	Kluster
SMKN 1 KARAWANG	2	264	250	250	2
SMKN 2 KARAWANG	2	183	169	169	2
SMKN 1 CIKAMPEK	2	182	168	168	2
SMKN 3 KARAWANG	2	147	133	133	2
TOTAL COST				6646	

Langkah selanjutnya mengganti medoid, dengan medoid baru yang digunakan adalah salah satu data non-medoid yang berada pada kluster yang sama dengan medoid awal. Hasil dari pengganti medoid awal dapat dilihat pada Tabel 4.27.

Tabel 4. 11 Medoid Baru Iterasi ke-1

	Jenis Sekolah	Status	Guru Matematika	Guru Bahasa Indonesia	Perpustakaan
M1	1	2	1	1	1
M2	1	2	0	1	0

Medoid yang diganti pada Tabel 4.27 adalah M1. Setelah mengganti medoid, kemudian dilakukan kembali perhitungan yang sama sampai total cost medoid lama tidak lebih besar dari total cost medoid baru. Hasil perhitungan cost medoid yang baru dapat dilihat pada Tabel 4.32.

Tabel 4. 12 Menghitung Cost Iterasi ke-2

Nama Sekolah	Jenis Sekolah	M1	M2	Cost	Klaster
SMKN 1 KARAWANG	2	246	264	246	1
SMKN 2 KARAWANG	2	165	183	165	1
SMKN 1 CIKAMPEK	2	164	182	164	1
SMKN 3 KARAWANG	2	129	147	129	1
SMA BAITUL ULYA BOARDING SCHOOL	1	18	2	2	2
TOTAL COST					6446

Setelah diperoleh hasil perhitungan jarak Manhattan dan total cost pada iterasi kedua, langkah selanjutnya adalah menghitung selisih antara total cost terbaru dengan cost pada iterasi sebelumnya. Dalam hal ini, nilai selisih S dihitung sebagai berikut:

$$S = 6446 - 6430 = 16$$

Apabila nilai S bernilai negatif (kurang dari 0), maka proses dilanjutkan dengan mengganti medoid lama menggunakan data lain dari klaster tersebut untuk membentuk komposisi medoid yang baru. Namun, karena nilai simpangan pada iterasi ini lebih besar dari nol, maka tidak dilakukan penggantian medoid, dan proses iterasi dinyatakan selesai.

E. Pengujian Dengan Rapid Miner

Pengujian algoritma K-Means dan algoritma K-Medoids dilakukan menggunakan RapidMiner untuk mengevaluasi efektivitas klasterisasi dalam mengelompokkan data sekolah. Proses ini melibatkan beberapa langkah kunci yang memanfaatkan kemampuan analisis dan visualisasi RapidMiner untuk mengukur dan membandingkan hasil klasterisasi.

a) Hasil Pengujian Algoritma K-Means

Hasil pengujian dengan menggunakan tool RapidMiner algoritma K-Means pada tampilan Gambar 4.3 akan menampilkan klaster model, di mana pada tahapan ini ditampilkan hasil pembagian data terhadap tiap klaster. Cluster_0 memiliki 32 anggota dan Cluster_1 memiliki 140 anggota dari total semua data berjumlah 172.

b) Hasil Pengujian Algoritma K-Medoid

Hasil pengujian dengan menggunakan tool RapidMiner algoritma K-Medoids pada tampilan Gambar 4.12 akan menampilkan klaster model, di mana pada tahapan ini ditampilkan hasil pembagian data terhadap tiap klaster. Cluster_0 memiliki 85 anggota dan Cluster_1 memiliki 87 anggota dari total semua data berjumlah 172.

IV. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan algoritma K-Means dan K-Medoids pada klasterisasi data sekolah di Kabupaten Karawang menghasilkan performa yang berbeda. K-Means menghasilkan klaster dengan kualitas lebih baik berdasarkan nilai Davies Bouldin Index (-0,648), meskipun distribusinya tidak seimbang (32 dan 140 sekolah). Sebaliknya, K-Medoids menghasilkan distribusi klaster yang lebih seimbang (85 dan 87 sekolah), tetapi memiliki kualitas klasterisasi yang lebih rendah (-3,027). Dengan demikian, algoritma K-Means lebih unggul dalam hal efektivitas klasterisasi pada dataset ini.

Penelitian selanjutnya disarankan untuk melibatkan jumlah data yang lebih besar guna meningkatkan representativitas hasil klasterisasi. Selain itu, eksplorasi menggunakan berbagai platform dan perangkat data mining seperti Python atau Weka dapat memperkaya analisis dan memungkinkan evaluasi performa algoritma secara lebih menyeluruh.

PENGAKUAN

Naskah ilmiah ini merupakan bagian dari riset tugas akhir milik Ridwan Nillah dengan judul “Klasterisasi Sekolah Menggunakan Algoritma K-Means dan K-Medoids Berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik” yang dibimbing oleh Sutan Faisal dan Santi Arum Puspita Lestari.

DAFTAR PUSTAKA

- [1] A. Aditya, I. Jovian, dan B. N. Sari, "Implementasi K-Means clustering ujian nasional sekolah menengah pertama di Indonesia tahun 2018/2019," *mib*, vol. 4, p. 51, 2020. [Online]. Tersedia: <https://doi.org/10.30865/mib.v4i1.1784>
- [2] F. Harahap, "Perbandingan algoritma K Means dan K Medoids untuk clustering kelas siswa tunagrahita 2," 2021.
- [3] R. Indriyana, S. Faisal, dan B. Priyatna, "Data mining implementasi algoritma K-Means untuk mengetahui prestasi siswa berdasarkan data nilai (studi kasus: SMA Negeri 1 Telukjambe Barat)," 2019.
- [4] A. B. Jiwandono, Okfalisa, dan Dr. M. Sc., "Analisis pengelompokan kelas BPJS Kesehatan dengan menggunakan metode K-Means studi kasus: Badan Penyelenggara Jaminan Sosial (BPJS) Kota Pekanbaru," Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau Pekanbaru, 2021.
- [5] L. Kartika, A. M. Siregar, dan D. S. Kusumaningrum, "Penerapan algoritma K-Medoids dan K-Means untuk pemetaan penyebaran guru tingkat SMP seluruh kabupaten/kota di Indonesia," 2020.
- [6] F. N. A. Kurniawati, "Meninjau permasalahan rendahnya kualitas pendidikan di Indonesia dan solusi," *AoEJ*, vol. 13, pp. 1–13, 2022. [Online]. Tersedia: <https://doi.org/10.47200/aoej.v13i1.765>
- [7] R. Maulidia, A. M. Siregar, dan S. A. P. Lestari, "Pengelompokan stok barang menggunakan algoritma K-Means," 2019.
- [8] E. Meravigliosi, "Kurangnya pemerataan pendidikan di Indonesia: Mengatasi hambatan menuju kesetaraan," 2023.
- [9] N. Nurahman, A. Purwanto, dan S. Mulyanto, "Klasterisasi sekolah menggunakan algoritma K-Means berdasarkan fasilitas, pendidik, dan tenaga pendidik," *matrik*, vol. 21, pp. 337–350, 2022. [Online]. Tersedia: <https://doi.org/10.30812/matrik.v21i2.1411>
- [10] N. Nurul Rhamadani, A. Fauzi, dan E. Nurlaelasari, "Penerapan algoritma K-Means dan K-Medoids dalam pengelompokan sekolah menengah kejuruan (SMK) di Kabupaten Karawang berdasarkan nilai ujian nasional," 2020.
- [11] S. Sindi, W. R. O. Ningse, dan I. A. Sihombing, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," 2020.
- [12] M. Tamba, "Analisa clustering laporan prestasi belajar siswa/siswi menggunakan K-Means (studi kasus: sekolah SMA N 21 Medan)," vol. 7, 2018.
- [13] L. Utari dan R. Gerald, "Pemetaan kompetensi siswa untuk mengikuti seleksi lomba kejuaraan multimedia menggunakan metode K-Means," *teknos. jurnal ilmiah teknologi informasi dan sains*, vol. 12, 2022.
- [14] A. Yani, Z. Azmi, dan D. Suherdi, "Implementasi data mining menganalisa data penjualan menggunakan algoritma K-Means clustering," *j. sist. inf. trig. dhar. JURSI TGD*, vol. 2, p. 315, 2023. [Online]. Tersedia: <https://doi.org/10.53513/jursi.v2i2.6357>
- [15] D. M. Yusup, A. M. Siregar, dan S. A. P. Lestari, "Penerapan data mining clustering dalam menentukan tingkat pembelian kredit tertinggi algoritma K-Means dan K-Medoids," 2023.