

**PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS
DALAM KLASIFIKASI SISWA BERPRESTASI DI SMP
MUHAMMADIYAH 60 MEDAN**

SKRIPSI

DISUSUN OLEH

FARIDA NAFISA SIAGIAN

2109010094



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS
DALAM KLASIFIKASI SISWA BERPRESTASI DI SMP
MUHAMMADIYAH 60 MEDAN**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

FARIDA NAFISA SIAGIAN

2109010094

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2026

LEMBAR PENGESAHAN

Judul Skripsi : PERBANDINGAN KINERJA K-MEANS
DAN K-MEDOIDS DALAM KLASIFIKASI
SISWA BERPRESTASI DI SMP
MUHAMMADIYAH 60 MEDAN

Nama Mahasiswa : FARIDA NAFISA SIAGIAN

NPM : 2109010094

Program Studi : **SISTEM INFORMASI**

Menyetujui
Komisi Pembimbing



(Halim Maulana, ST, M.Kom)
NIDN. 0121119102

Ketua Program Studi



Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom)
NIDN. 0117088902

Dekan



(Dr. Al-Khwarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS
DALAM KLASIFIKASI SISWA BERPRESTASI DI SMP
MUHAMMADIYAH 60 MEDAN**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 02 Maret 2026

Yang membuat pernyataan



FARIDA NAFISA SIAGIAN

2109010094

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Farida Nafisa Siagian
NPM : 2109010094
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

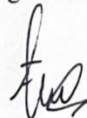
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS DALAM
KLASIFIKASI SISWA BERPRESTASI DI SMP MUHAMMADIYAH 60
MEDAN**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 02 Maret 2026
Yang membuat pernyataan



Farida Nafisa Siagian
NPM. 2109010094

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Farida Nafisa Siagian
Tempat dan Tanggal Lahir : Medan, 31 Agustus 2002
Alamat Rumah : Jalan Pelikan IV No.159
Telepon/Faks/HP : 081376478329
E-mail : ridasiagian31@gmail.com
Instansi Tempat Kerja : Belum Bekerja
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN 066667 MEDAN TAMAT: 2012
SMP : SMP NEGERI 29 TAMAT: 2017
SMA : SMK TELKOM 1 MEDAN TAMAT: 2020

KATA PENGANTAR



Puji dan syukur penulis panjatkankehadirat Allah SWT atas limpah rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS DALAM KLASIFIKASI SISWA BERPRESTASI DISMP MUHAMMADIYAH 60 MEDAN”**

Dalam proses penyusunan skripsi ini penulis menyadari bahwa tidak akan dapat menyelesaikan tanpa doa,dukungan, dan bantuan dari beberapa pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Akrim, M.Pd., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahti Rizky, S.Kom., M.Kom., selaku Wakil Dekan I FIKTI UMSU
4. Bapak Mhd. Basri, S.Si., M.Kom., selaku Wakil Dekan III (FIKTI) UMSU
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom Ketua Program Studi Sistem Informasi
6. Bapak Mulkan Azhari, S.Kom., M.Kom Sekretaris Program Studi Sistem Informasi
7. Bapak Halim Maulana, ST, M.Kom selaku pembimbing saya yang telah memberikan bimbingan, arahan, dan ilmu yang berharga selama proses penyusunan skripsi ini.
8. Ayahanda tercinta yaitu Alm Drs. Parman Siagian, M.A/ Drs. Burhanuddin Siagian, M.Ag sosok yang paling kurindukan. Terimakasih atas segala pengorbanan, kasih sayang, serta serta segala bentuk tanggung jawab yang ayah berikan. Terimakasih telah menjadi alasan terbesarku

untuk tetap semangat meraih berjuang meraih gelar sarjana yang telah ayah impikan.

9. Bidadari surgaku Tri Nurpidawati Koto, S.S ibu tunggal yang hebat luar biasa yang selalu menjadi penyemangat saya sebagai sandaran terkuat menghadapi kerasnya dunia ini. Terimakasih sudah melahirkan, berjuang sekuat tenaga untuk memberikan kehidupan yang layak untuk saya, kerja keras dan menjadi tulang punggung keluarga hingga akhirnya saya tumbuh dewasa dan bisa berada diposisi ini. Kesuksesan dan segala hal baik yang kedepannya akan penulis dapatkan adalah karena beliau. Tolong hidup yang lama didunia ini.
10. Untuk nenek Djaliar Koto , serta saudara-saudara tercinta, Fatria Rahmi Siagian dan Septian Faruqi Siagian, S.Pd.. Terimakasih atas segala dukungan, semangat, perhatian, doa, dan kasih sayang yang telah diberikan selama proses perkuliahan hingga penyelesaian skripsi ini. Dukungan dan doa yang tulus dari keluarga menjadi sumber kekuatan bagi penulis dalam menghadapi berbagai tantangan selama menempuh pendidikan. Semoga segala kebaikan yang telah diberikan mendapat balasan yang berlimpah dari Tuhan Yang Maha Esa.
11. Kepada sahabat Zahra Apriyani Hakim Nasution Penulis juga menyampaikan terima kasih yang sebesar-besarnya telah menemani perjalanan selama masa perkuliahan. Terima kasih karena telah bersama-sama melewati berbagai proses, mulai dari menjadi mahasiswa baru, menjalani kegiatan perkuliahan, melaksanakan magang, mengikuti KKN, hingga berjuang menyelesaikan skripsi. Dukungan, kebersamaan, semangat, dan kenangan yang telah dibangun selama ini menjadi bagian yang sangat berharga dalam perjalanan akademik penulis. Semoga persahabatan yang telah terjalin dapat terus terjaga dan kita semua diberikan kesuksesan dalam menapaki langkah berikutnya setelah wisuda.

12. Terakhir, saya berterimakasih kepada satu sosok gadis yang selama ini diam-diam berjuang tanpa henti, seorang perempuan sederhana dengan hati kecil tetapi dengan impian besar. Terimakasih kepada peneliti skripsi ini yaitu diriku sendiri, Farida Nafisa Siagian. Anak perempuan terakhir dan harapan orangtuanya. Terimakasih telah bertahan sejauh ini, dan terus berjalan melewati segala tantangan semesta hadirkan. Aku bangga atas setiap langkah kecil yang kau ambil, atas semua pencapaian yang mungkin tidak dirayakan orang lain. Walau terkadang harapan mu tidak sesuai dengan apa yang semesta berikan, tetaplah belajar menerima dan mensyukuri apapun yang telah kamu dapatkan. Jangan lelah untuk tetap berusaha, berbahagialah dimanapun kamu berada. Aku berdoa semoga langkah kecilmu selalu diperkuat, dikelilingin orang-orang baik dan hebat, serta mimpi mu satu persatu akan terjawab. Aamiin.

**PERBANDINGAN KINERJA K-MEANS DAN K-MEDOIDS
DALAM KLASIFIKASI SISWA BERPRESTASI DI SMP
MUHAMMADIYAH 60 MEDAN**

ABSTRAK

Pengelompokan data siswa berprestasi merupakan langkah penting dalam proses evaluasi dan pengambilan keputusan di lingkungan pendidikan. Metode klasterisasi seperti K-Means dan K-Medoids banyak digunakan untuk mengelompokkan data berdasarkan variabel prestasi akademik, absensi, dan ekstrakurikuler. Penelitian ini bertujuan membandingkan kinerja kedua algoritma tersebut dalam mengelompokkan siswa berprestasi di SMP Muhammadiyah 60 Medan. Metode yang digunakan meliputi proses pengelompokan data dengan algoritma K-Means dan K-Medoids, serta evaluasi hasil menggunakan nilai silhouette untuk menilai kualitas klaster. Hasil penelitian menunjukkan bahwa kedua metode mampu mengelompokkan siswa ke dalam cluster prestasi tinggi, sedang, dan rendah. K-Means lebih cepat dalam proses komputasi dan cocok untuk data homogen serta jumlah besar, namun sensitif terhadap outlier. Sebaliknya, K-Medoids lebih stabil dan tahan terhadap outlier meskipun memakan waktu lebih lama. Berdasarkan hasil evaluasi, K-Medoids memberikan hasil yang lebih representatif dalam kondisi data dengan variasi nilai ekstrim. Penelitian ini diharapkan dapat membantu pihak sekolah dalam melakukan segmentasi siswa secara efektif dan efisien untuk pengambilan kebijakan yang lebih tepat sasaran.

Kata Kunci: K-Means, K-Medoids, Klasterisasi.

COMPARISON OF K-MEANS AND K-MEDOIDS PERFORMANCE IN CLASSIFICATION OF HIGH-ACHING STUDENTS AT SMP MUHAMMADIYAH 60 MEDAN

ABSTRACT

Clustering high-achieving student data is a crucial step in the evaluation and decision-making process in educational settings. Clustering methods such as K-Means and K-Medoids are widely used to cluster data based on academic achievement, attendance, and extracurricular activities. This study aims to compare the performance of the two algorithms in clustering high-achieving students at SMP Muhammadiyah 60 Medan. The methods used include data clustering with the K-Means and K-Medoids algorithms, as well as evaluating the results using silhouette values to assess cluster quality. The results show that both methods are capable of clustering students into high-, medium-, and low-achieving clusters. K-Means is faster in computation and suitable for homogeneous and large data, but is sensitive to outliers. In contrast, K-Medoids is more stable and resilient to outliers although it takes longer. Based on the evaluation results, K-Medoids provides more representative results in data conditions with extreme value variations. This research is expected to assist schools in segmenting students effectively and efficiently for more targeted policy making.

Keywords: K-Means, K-Medoids, Clusterization.

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
RIWAYAT HIDUP	v
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian.....	4
BAB II LANDASAN TEORI	6
2.1 Data Mining	6
2.1.1 Fungsi Data Mining.....	6
2.2 Algoritma K-Means.....	7
2.2.1 Kelebihan dan Kekurangan Algoritma K-means	8
2.3 Algoritma K-Medoids	9
2.3.1 Kelebihan dan Kekurangan Algoritma K-Medoids.....	11
2.3.2 Perbandingan K-Means dengan K- Medoids	11
2.4 Data dan Alat Penelitian	11
2.5 Aplikasi pengembangan sistem	12
2.5.1 PHP (Hypertext Preprocessor).....	12
2.5.2 MySQL (My Structured Query Language).....	13
2.5.3 XAMPP.....	14
2.6 Penelitian yang Relevan	15

BAB III METODOLOGI PENELITIAN.....	19
3.1 Jenis dan Pendekatan Penelitian	19
3.2 Lokasi, Waktu, dan Subjek Penelitian.....	19
3.3 Data Dan sumber Data.....	20
3.4 Teknik Pengumpulan Data	21
3.5 Teknik Pengolahan Data.....	22
3.6 Metode Analisis Data	22
3.7 Penentuan Jumlah Cluster (k)	25
3.8 Evaluasi Kinerja Clustering	25
3.9 Tools dan Perangkat Penelitian.....	28
3.10 Diagram Air Penelitian.....	29
BAB IV HASIL DAN PEMBAHASAN.....	31
4.1 Gambaran Umum Data Penelitian.....	31
4.1.1 Deskripsi Data Siswa.....	31
4.1.2 Statistik Deskriptif.....	32
4.2 Hasil Preprocessing Data.....	34
4.2.1 Data Cleaning.....	34
4.2.2 Transformasi Data	35
4.2.3 Normalisasi Data (Min–Max)	35
4.3 Penentuan Jumlah Cluster	39
4.3.1 Hasil Metode Elbow	39
4.4 Implementasi Algoritma K-Means.....	39
4.4.1 Tahapan Eksekusi.....	39
4.4.2 Hasil Clustering K-Means.....	40
4.4.3 Analisis Cluster K-Means	51
4.5 Implementasi K-Medoids	52
4.5.1 Tahapan Eksekusi.....	52
4.5.2 Hasil Clustering K-Medoids	57
4.5.3 Analisis Cluster K-Medoids.....	58
4.6 Evaluasi Perbandingan Kinerja	58
4.6.1 Silhouette Score	58
4.6.1.1 K-Means.....	58

4.6.1.2 K-Medoids	62
4.6.2 Perbandingan SSE	65
4.6.3 Konsistensi Cluster	67
4.7 Pembahasan	70
4.7.1 Algoritma yang Lebih Baik.....	70
4.7.2 Kelebihan dan Kekurangan K-Means vs K-Medoids (pada kasus ini)	70
4.8 Hasil Program	71
BAB V PENUTUP.....	80
5.1 Kesimpulan	80
5.2. Saran.....	81
DAFTAR PUSTAKA	81

DAFTAR TABEL

Tabel 2. 1 Penelitian yang relevan.....	16
Tabel 3. 1 Interpretasi nilai Silhouette.....	26
Tabel 4. 1 Contoh Data Siswa.....	32
Tabel 4. 2 Statistik Deskriptif	32
Tabel 4. 3 Nilai Min-Max	35
Tabel 4. 4 Data Setelah Normalisasi.....	38
Tabel 4. 5 Jarak ke Centroid Awal (Iterasi 1)	42
Tabel 4. 6 Hasil Jarak ke Centroid Baru	46
Tabel 4. 7 Hasil Perhitungan Sum of Squared Errors (SSE).....	47
Tabel 4. 8 Rangkuman Silhouette per siswa	61
Tabel 4. 9 Rangkuman Silhouette K-Medoids	64
Tabel 4. 10 Kelebihan dan Kekurangan.....	70
Tabel 4. 11 Blackbox Testing Form Login	76
Tabel 4. 12 Blackbox Testing Form Home	77
Tabel 4. 13 Blackbox Testing Form Siswa	78
Tabel 4. 14 Blackbox Testing Form Clustering.....	78

DAFTAR GAMBAR

Gambar 2. 1 Tampilan XAMPP.....	15
Gambar 3. 1 Flowchart Alur Penelitian	29
Gambar 4. 1 Grafik Elbow.....	39
Gambar 4. 2 Grafik Scatter Plot Hasil Cluster	49
Gambar 4. 3 Grafik Perbandingan Error.....	66
Gambar 4. 4 Form Login	71
Gambar 4. 5 Form Home	72
Gambar 4. 6 Form Siswa	73
Gambar 4. 7 Form Clustering.....	74
Gambar 4. 8 Hasil Pada Program	75

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

SMP Muhammadiyah 60 Medan merupakan salah satu lembaga pendidikan swasta yang memiliki peran penting dalam mencetak generasi penerus yang berprestasi. Sekolah ini menampung siswa dengan latar belakang yang beragam, baik dari segi kemampuan akademik maupun non-akademik. Oleh karena itu, pihak sekolah perlu mengetahui pola prestasi siswa agar dapat memberikan perhatian dan strategi pembinaan yang tepat.

Permasalahan yang muncul adalah pihak sekolah sering mengalami kesulitan dalam menentukan kelompok siswa berprestasi dan kurang berprestasi berdasarkan data yang tersedia yaitu berdasarkan nilai Matematika, IPA, Bahasa Indonesia, Absensi, dan Ekstrakurikuler. Kesulitannya adalah ketika menghitung nilai dan mengelompokkannya, mengingat dan melihat data siswa yang sangat. Hal ini menyebabkan sulitnya mengklasifikasikan siswa yang berprestasi.

Oleh karena itu diperlukan suatu pendekatan analitis berbasis teknologi yang mampu mengolah data siswa dan menemukan pola prestasi mereka. *Data mining* hadir sebagai salah satu solusi, khususnya dengan teknik klasifikasi, yang dapat digunakan untuk mengelompokkan siswa berdasarkan tingkat prestasi. Dengan demikian, sekolah dapat memperoleh informasi yang lebih objektif untuk mendukung pengklasifikasian siswa berprestasi.

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma K-Means banyak digunakan karena kemudahannya dalam implementasi dan kecepataannya dalam memproses data yang besar. Namun, penelitian lain juga menunjukkan bahwa K-Medoids memiliki keunggulan dalam menghadapi *outlier*, sehingga hasil pengelompokannya lebih stabil. Misalnya, penelitian Mochamad Wahyudi (2020) menemukan bahwa K-Means lebih efisien dalam waktu komputasi, sementara penelitian Wiwid Ahyudi (2023) menunjukkan bahwa K-Medoids memberikan hasil clustering yang lebih akurat dalam data yang mengandung variasi tinggi.

Berdasarkan penelitian-penelitian terdahulu, terlihat adanya perbedaan kinerja antara K-Means dan K-Medoids. Hal ini menjadi dasar untuk membandingkan kedua algoritma tersebut dalam konteks klasifikasi siswa berprestasi di SMP Muhammadiyah 60 Medan, sehingga dapat diketahui algoritma mana yang lebih sesuai untuk digunakan pada data pendidikan di sekolah ini.

K-Means adalah algoritma *clustering* yang menggunakan rata-rata (*mean*) sebagai pusat klaster. Algoritma ini bekerja dengan cara meminimalkan jarak kuadrat antar data dengan pusat klaster, sehingga lebih cepat namun sensitif terhadap keberadaan *outlier*. (Ahmed & Ashour, 2022). Sementara itu, K-Medoids merupakan algoritma *clustering* yang mirip dengan K-Means, tetapi pusat klaster ditentukan berdasarkan objek nyata dalam data (*medoid*), bukan rata-rata. Hal ini membuat K-Medoids lebih tahan terhadap *outlier*, meskipun membutuhkan waktu komputasi yang lebih lama. (Hardiyani et al, 2019).

Dengan adanya penelitian ini diharapkan dapat membantu dalam mengelompokkan siswa berprestasi secara lebih objektif dan akurat, penelitian ini dapat menjadi bahan evaluasi dan perbandingan mengenai kinerja algoritma K-Means dan K-Medoids dalam bidang pendidikan.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana agar pihak sekolah dapat menghasilkan klasifikasi siswa berprestasi?
2. Bagaimana hasil perbandingan algoritma K-Means dan K-Medoids terhadap klasifikasi siswa berprestasi?
3. Bagaimana menghasilkan perangkat lunak perbandingan kinerja K-Means dan K-Medoids dalam klasifikasi siswa berprestasi di SMP Muhammadiyah 60 Medan?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan pada penelitian ini menggunakan seluruh kelas dari murid SMP Muhammadiyah 60 Medan.
2. Atribut data yang digunakan dalam proses clustering terbatas pada nilai akademik, tingkat kehadiran, serta prestasi non-akademik seperti keterlibatan dalam ekstrakurikuler.
3. Algoritma clustering yang dibandingkan dalam penelitian ini hanya mencakup K-Means dan K-Medoids.

1.4 Tujuan Penelitian

Tujuan dalam skripsi ini adalah :

1. Agar pihak sekolah dapat menghasilkan klasifikasi siswa berprestasi.
2. Mendapatkan hasil perbandingan algoritma K-Means dan K-Medoids terhadap klasifikasi siswa berprestasi.
3. Menghasilkan perangkat lunak perbandingan kinerja K-Means dan K-Medoids dalam klasifikasi siswa berprestasi di SMP Muhammadiyah 60 Medan.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang hendak dicapai, maka penelitian diharapkan mempunyai manfaat dari beberapa pihak. Adapun manfaat penelitian ini sebagai berikut:

1. Manfaat bagi peneliti:

Penelitian ini akan memperdalam pemahaman mengenai teknik *clustering*. Peneliti juga dapat mengeksplorasi kelebihan dan kekurangan dari masing – masing algoritma.

2. Manfaat bagi sekolah

Memberikan rekomendasi metode pengelompokan siswa yang lebih objektif dan efisien, sehingga dapat digunakan sebagai dasar dalam perencanaan program pembinaan dan pengembangan potensi siswa.

3. Manfaat bagi UMSU:

Dengan adanya penelitian ini, manfaat bagi umsu adalah penelitian ini memberikan kontribusi bagi pengembangan ilmu pengetahuan di bidang *data mining*, khususnya dalam penerapan algoritma *klasifikasi* (K-Means dan K-Medoids) di sektor pendidikan. Hal ini dapat memperkaya referensi akademik dan mendukung penelitian lanjutan di UMSU.

BAB II

LANDASAN TEORI

2.1 Data Mining

Menurut (Wahyudi Setiawan, 2021) Data Mining adalah proses pengambilan atau pengumpulan informasi penting dari kumpulan data yang besar. Proses mining atau penambangan data sering menggunakan teknik statistik dan matematika dengan memanfaatkan teknik kecerdasan buatan. Nama lain dari kata mining yaitu Know Discovery in Database (KDD), Big data, Bussiness intellingence, Know extraction, Pattern analysis dan Information harvesting. Terdapat banyak konsep dan teknik yang digunakan dalam proses data mining. Proses ini membutuhkan beberapa langkah untuk mendapatkan data yang dibutuhkan. Proses KDD meliputi data selection, preprosesing, data mining, evaluation dan knowledge discovery.

2.1.1 Fungsi Data Mining

Menurut (Wahyudi Setiawan, 2021) fungsi data mining memiliki 2 fungsi yaitu :

1. Deskripsi

Fungsi Deskripsi adalah fungsi yang memberi pemahaman yang lebih dalam tentang data yang diamati. Saat menjalankan suatu proses diharapkan mengetahui bagaimana perilaku data. Data ini nantinya dapat digunakan untuk menentukan sifat-sifat data yang bersangkutan. Selanjutnya dapat menggunakan fitur penambangan data deskriptif untuk kemudian menemukan pola tertentu yang tersembunyi dalam data. Dengan

kata lain, jika polanya berulang dan bernilai, dapat mengetahui sifat-sifat data.

2. Prediksi

Fungsi Prediksi adalah fungsi bagaimana suatu proses menemukan pola data tertentu. Pola-pola tersebut dapat dilihat dari berbagai variable yang ada dalam data. Setelah menemukan pola, dapat menggunakan pola yang diperoleh untuk memprediksi variabel lain yang nilai atau jenisnya tidak diketahui. Untuk alasan ini, fungsi yang satu ini juga disebut fungsi prediktif dan dilakukan sebagai analisis prediktif. Fungsi ini juga dapat digunakan untuk memprediksi variabel tertentu yang tidak ada dalam data. Oleh karena itu, fitur ini memudahkan dan informatif bagi mereka yang membutuhkan prediksi akurat untuk meningkatkan hal-hal penting tersebut.

2.2 Algoritma K-Means

K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat. (Mochamad Wahyudi, et al, 2020)

Langkah-langkah pengelompokan data adalah :

1. Pilih jumlah klaster.
2. Inisialisasi awal dan pusat klaster dilakukan secara random.
3. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar

obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean (Euclidean Distance) dengan rumus seperti di bawah ini (Witanto, Ratnawati dan Anam, 2019) :

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

dimana :

$d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, \dots, x_i)$ yaitu variabel data

$y_i = (y_1, y_2, \dots, y_j)$ yaitu variabel pada titik pusat.

4. Hitung pusat kluster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata obyek pada kluster. Penghitungan bisa juga dengan menggunakan median.
5. Hitung kembali jarak tiap objek dengan pusat kluster yang baru, hingga kluster tidak berubah, maka proses pengklasteran selesai.

2.2.1 Kelebihan dan Kekurangan Algoritma K-means

a. Kelebihan K-means:

Algoritma K-means memiliki kelebihan yaitu:

1. Dalam implementasi menyelesaikan masalah, algoritma K-Means sangat simple serta fleksibel. Artinya perhitungan komputasinya tidak terlalu rumit dan algoritma ini dapat diimplementasikan pada segala bidang.
2. Algoritma K-Means sangat mudah untuk dipahami, terutama dalam implementasi data yang sangat besar serta dapat mengurangi kompleksitas data yang dimiliki (Bangoria et al., 2013)

b. Kekurangan K-means :

Algoritma K-means memiliki kekurangan yaitu:

1. User memerlukan angka yang tepat dalam menentukan jumlah cluster sebanyak k karena terkadang pusat cluster awal dapat berubah sehingga kejadian ini bisa mengakibatkan pengelompokan data menjadi tidak stabil (Joshi & Nalwade, 2013).
2. Algoritma K-Means tidak bisa maksimal dalam menentukan atau menginisialkan nilai centroid awalnya, karena pada pengelompokan data dengan algoritma K-Means sangat bergantung pada nilai centroidnya (Ahmed & Ashour, 2011).

Output dari K-Means tergantung pada nilai – nilai pusat yang dipilih pada clustering. Sehingga pada algoritma ini nilai awal titik pusat cluster menjadi dasar dalam penentuan cluster. Pemilihan centroid cluster awal secara acak akan memberikan pengaruh terhadap kinerja cluster tersebut (Singh & Kaur, 2013; Sujatha & Sona, 2013)

2.3 Algoritma K-Medoids

Algoritma K-Medoids atau Partitioning Around Medoids (PAM) merupakan suatu metode partisi clustering untuk mengelompokkan sekumpulan (n) objek menjadi sejumlah (k) cluster. Algoritma K-Medoids menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster. Objek yang terpilih untuk mewakili sebuah cluster disebut medoid (Siti Nurlela, 2020). Algoritma K-Medoids mampu meminimalkan jumlah kesamaan antara setiap objek dengan titik acuan yang bersesuaian (Wiwid Ahyudi, 2023).

Menurut (Hardiyani, Tambunan, & Saragih, 2019) berikut merupakan langkah-langkah dalam perhitungan algoritma K-Medoids diantaranya:

- a. Inisialisasikan pusat cluster sebanyak jumlah cluster (k).
- b. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$d(x_{ij}, c_{kj}) = \sqrt{\sum_{j=i}^p \sum_{i=1}^n (x_{ij} - c_{kj})^2}$$

Dimana :

$D(X_{ij}, C_{kj})$ = Jarak Euclidian Distance antara pengamatan ke-I variabel ke-j ke pusat cluster ke-k pada variabel ke-j,

X_{ij} = Objek pada pengamatan ke-i pada variabel ke-j,

C_{kj} = Pusat kelompok ke-k pada variabel ke-j,

P = banyaknya variabel yang diamati,

N = banyak pengamatan yang diamati.

- c. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
- d. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, mak tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
- e. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehigga didapatkan cluster beserta anggota masing-masing cluster.

2.3.1 Kelebihan dan Kekurangan Algoritma K-Medoids

K-Medoids merupakan salah satu metode analisis cluster yang menggunakan teknik berbasis objek representative (perwakilan) yang disebut medoids (Suyanto, 2017). KMedoids memiliki beberapa kelebihan antara lain tidak sensitif dengan outlier, dapat mengurangi noise, dan lebih baik dalam waktu eksekusi (Arora dkk., 2016). Metode kmedoids juga memiliki beberapa kelemahan, salah satunya adalah tidak dapat menentukan cluster optimal.

Penentuan cluster optimal pada penelitian ini menggunakan metode validasi silhouette index. Hasil cluster optimal menggunakan k-medoids belum dapat mengurutkan prioritas cluster pelanggan hasil pemodelan.

2.3.2 Perbandingan K-Means dengan K- Medoids

Perbandingan K-Means dengan K- Medoids Algoritma k-Means sangat sensitif terhadap data outlier karena data semacam itu jauh dari sebagian data, ketika ditetapkan sebuah cluster, data tersebut dapat mempengaruhi nilai rata rata dari cluster, dimana algoritma k-Means sendiri menggunakan rata rata antara data sebagai centroid atau pusat cluster sedangkan K-Medoid menggunakan data tersebut sebagai centroidnya.

2.4 Data dan Alat Penelitian

Berikut ini adalah data dan alat penelitian yang digunakan dalam membuat aplikasi data mining dalam clustering siswa berprestasi di SMP Muhammadiyah 60 Medan.

Data Penelitian yang dibutuhkan :

1. Data siswa-siswi
2. Data variabel yang diberikan pihak sekolah
3. Data nilai siswa-siswi
4. Alat penelitian dalam pembuatan system perangkat keras dan perangkat lunak. Perangkat keras adalah perangkat pada computer yang berbentuk fisik dan dapat disentuh. Sedangkan perangkat lunak adalah sekumpulan intruksi atau program yang memberi tahu computer bagaimana cara melakukan tugas tertentu.
5. Laptop dengan spesifikasi : Windows 10 pro, Prosesor core i5 2,4 GHz, RAM4,00
6. GB, Sistem type 64-bit
7. Printer

2.5 Aplikasi pengembangan sistem

Aplikasi yang dibutuhkan / digunakan dalam pengembangan sistem untuk mengimplementasikan data mining dengan menggunakan metode K-means dan K-medoids.

2.5.1 PHP (Hypertext Preprocessor)

Menurut (Sholichin, 2020) PHP adalah bahasa scripting yang berfokus pada pengembangan web, yang memungkinkan integrase dengan HTML untuk menciptakan aplikasi yang responsif. Dengan kemampuannya untuk berinteraksi dengan database, PHP menjadi pilihan utama bagi banyak pengembang dalam menciptakan aplikasi berbasis web.

Budi Santoso (2023), PHP juga dikenal karena kemampuannya untuk terintegrasi dengan berbagai database, seperti MySQL dan PostgreSQL. Ini menjadikannya pilihan yang ideal untuk pengembangan aplikasi yang memerlukan pengelolaan data secara real-time, serta memberikan fleksibilitas dalam pengembangan aplikasi berbasis web yang kompleks.

Beberapa kelebihan PHP menurut (Alan Nur Aditya, 2011)

1. Bahasa pemrograman PHP adalah sebuah bahasa script yang tidak melakukan sebuah kompilasi dalam penggunaannya.
2. Web server yang mendukung PHP dapat ditemukan dimana-mana dari mulai apache, IIS, Lighttpd, hingga Xitami konfigurasi yang relatif mudah.
3. Dalam sisi pengembangan lebih mudah, karena banyaknya developer yang siap membantu dalam pengembangan.
4. Dalam sisi pemahaman, PHP adalah bahasa scripting yang paling mudah karena memiliki referensi yang banyak.
5. PHP adalah bahasa open source yang dapat digunakan diberbagai mesin (Linux, Unix, Macintosh, Windows) dan dapat dijalankan secara runtime.

2.5.2 MySQL (My Structured Query Language)

MySQL adalah DBMS yang *open source* dengan dua bentuk lisensi, yaitu *Free Software* (perangkat lunak bebas) dan *Shareware* (perangkat lunak berpemilik yang penggunaannya terbatas). Jadi MySQL adalah *database server* yang gratis dengan lisensi GNU *General Public License* (GPL) sehingga dapat dipakai untuk keperluan pribadi atau komersial tanpa harus membayar lisensi yang ada. MySQL masuk dalam jenis RDBMS (*Relational database*

Management) maka dari itu, istilah semacam baris, kolom, tabel, dipakai pada MySQL (Rahimi Fitri, 2020:2)

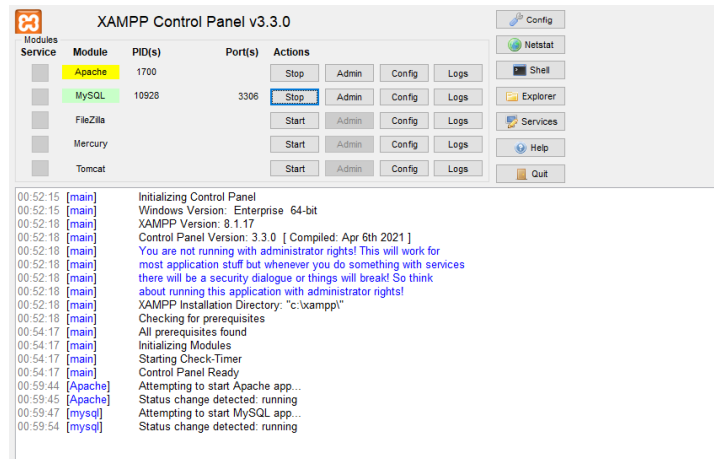
MySQLmerupakandatabaseserveryangmemilikikonsepdatabase modern, MySQL mempunyai beberapa kelebihan, antara lain :

1. Free(bebasdiunduh)
2. Stabil dan tangguh
3. Fleksibel dengan berbagai pemrograman
4. Security yang baik
5. Dukungan dari banyak komunitas
6. Kemudahan management database
7. Perkembangan software yang cukup tepat.

2.5.3 XAMPP

XAMPP adalah sebuah paket kumpulan software yang terdiri dari apache, mysql, phpmyadmin, php perl, Freetype, dll. XAMPP berfungsi untuk memudahkan instalasi lingkungan php, dimana bisanya lingkungan pengembangan web memerlukan php, apache, mysql dan phpmyadmin serta software-software yang terkait dengan pengembangan web. Dengan menggunakan xampp, kita tidak perlu lagi menginstal aplikasi-aplikasi tersebut satu persatu (M.Aswadi 2015).

Berikut ini tampilan control panel pada aplikasi XAMPP yang tersedia



Gambar 2. 1 Tampilan XAMPP

Pada umumnya XAMPP dibagi menjadi tiga bagian yaitu sebagai berikut dijelaskan:

1. Htdocs, yaitu untuk tempat berkas yang dijalankan, misalnya berkas PHP, HTML, dan skrip lainnya.
2. PhpMyAdmin, unuk membuka halaman PhpMAdmin dengan mengetik alamat <http://localhost/phpMyAdmin>.
3. Control Panel, untuk mengelola layanan XAMPP, Seperti memulai dan megehentikan layanan.

2.6 Penelitian yang Relevan

Dalam menyusun penelitian ini dapat dikaitkan dengan beberapa penelitian sebelumnya. Berikut ini adalah beberapa penelitian yang relevan dengan penelitian ini :

Tabel 2. 1 Penelitian yang relevan

No	Judul	Penulis	Tahun	Kesimpulan
	<p>Analisis Kinerja Algoritma K-Means dan K-Medoids dalam Menentukan Pengetahuan Mahasiswa dan Dosen UIN tentang sejarah dengan Menggunakan Teknologi Big Data.</p>	<p>Nadika Sigit Sinatrya</p>	<p>2021</p>	<p>Penelitian ini menunjukkan bahwa K-Means lebih efisien dibandingkan dengan K-Medoids, dengan waktu komputasi yang lebih cepat.</p>
	<p>Implementasi K-Means dan K-Medoids dalam Identifikasi Pola Prestasi Belajar Siswa untuk Optimalisasi Program Pembelajaran</p>	<p>Wibowo, D. A., & Pratiwi, E. K.</p>	<p>2022</p>	<p>Studi ini berhasil menerapkan K-Means dan K-Medoids untuk mengidentifikasi pola prestasi belajar siswa. Hasil <i>clustering</i> memungkinkan sekolah untuk mengidentifikasi kelompok siswa dengan karakteristik prestasi yang serupa, sehingga program pembelajaran dan intervensi dapat disesuaikan untuk setiap kelompok, meningkatkan efektivitas pengajaran. K-Medoids dinilai lebih cocok untuk data</p>

				dengan <i>noise</i> yang signifikan.
	Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Siswa Berdasarkan Prestasi Akademik	Rahmawati, A., & Santoso, B.	2023	Penelitian ini menyimpulkan bahwa kedua algoritma, K-Means dan K-Medoids, mampu mengelompokkan siswa berdasarkan prestasi akademik dengan baik. Namun, K-Medoids menunjukkan robusta yang lebih baik terhadap <i>outlier</i> dalam data prestasi siswa, menghasilkan kelompok yang lebih representatif dibandingkan K-Means dalam kasus data dengan nilai ekstrem. K-Means cenderung lebih cepat untuk dataset besar, tetapi sensitif terhadap inisialisasi awal.

	<p>Pengembangan Sistem Pendukung Keputusan Penentuan Siswa Berprestasi Menggunakan Metode Hybrid K-Means dan K-Medoids</p>	<p>Hidayat, M., & Kusuma, D. W.</p>	<p>2024</p>	<p>Penelitian ini mengusulkan sebuah sistem pendukung keputusan yang menggabungkan K-Means dan K-Medoids untuk identifikasi siswa berprestasi. Pendekatan hibrida ini bertujuan untuk memanfaatkan kekuatan masing-masing algoritma, di mana K-Means digunakan untuk <i>clustering</i> awal yang efisien, dan K-Medoids digunakan untuk memurnikan <i>cluster</i> dan menangani <i>outlier</i>. Hasilnya menunjukkan peningkatan akurasi dan stabilitas <i>clustering</i> dibandingkan penggunaan algoritma tunggal.</p>
--	--	---	-------------	--

BAB III

METODOLOGI PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif karena data yang digunakan berupa angka-angka hasil nilai akademik, absensi, dan prestasi non-akademik siswa SMP Muhammadiyah 60 Medan, dan dianalisis menggunakan metode statistik serta perhitungan algoritmik.

Pendekatan yang digunakan dalam penelitian ini adalah pendekatan komputasional/data mining, karena menggunakan algoritma K-Means dan K-Medoids untuk melakukan proses clustering siswa berdasarkan tingkat prestasi. Metode analisis data dalam penelitian ini menggunakan algoritma clustering, yaitu:

1. Algoritma K-Means
2. Algoritma K-Medoids (PAM – Partitioning Around Medoids)

3.2 Lokasi, Waktu, dan Subjek Penelitian

3.2.1 Lokasi Penelitian

Penelitian dilaksanakan di SMP Muhammadiyah 60 Medan sebagai lokasi pengambilan data siswa dan implementasi hasil clustering.

3.2.2 Waktu Penelitian

Penelitian ini dilaksanakan pada Periode Januari – April 2025, meliputi proses pengambilan data, pengolahan data, implementasi algoritma, evaluasi, dan penyusunan laporan skripsi.

3.2.3 Subjek dan Objek Penelitian

Subjek penelitian : Data siswa SMP Muhammadiyah 60 Medan.

Objek penelitian : Kinerja algoritma K-Means dan K-Medoids dalam proses clustering siswa berprestasi berdasarkan nilai akademik, kehadiran, dan ekstrakurikuler.

3.3 Data Dan sumber Data

3.3.1 Jenis Data

Jenis data yang digunakan dalam penelitian ini yaitu data primer dan data sekunder. Data primer merupakan data utama yang digunakan dalam penelitian ini. Data primer diperoleh melalui wawancara dan observasi langsung ke SMP Muhammadiyah 60 Medan untuk mendapatkan informasi terkait klasifikasi siswa berprestasi. Sedangkan untuk data sekunder, dikumpulkan dataset berupa nilai akademik, kehadiran, partisipasi dalam kegiatan ekstrakurikuler, dan data demografis siswa. Dataset ini diperlukan sebagai data input untuk membangun model analisis dan perhitungan aplikasi data mining klasifikasi. Data klasifikasi ini juga berfungsi sebagai data uji yang akan diklasifikasikan oleh perangkat lunak untuk menghasilkan *output* rekomendasi pengelompokan siswa yang optimal.

3.2.4 Sumber Data

Sumber data diperoleh dari:

1. Wali kelas
2. Bagian kurikulum
3. Arsip penilaian sekolah.

3.3.3 Variabel/ Atribut Penelitian

Atribut yang digunakan dalam proses clustering, yaitu:

1. Nilai Matematika
2. Nilai IPA
3. Nilai Bahasa Indonesia
4. Nilai Absensi (kehadiran siswa)
5. Nilai Ekstrakurikuler

3.4 Teknik Pengumpulan Data

Dalam penelitian ini penulis menggunakan tiga teknik pengumpulan data yaitu, studi literatur, observasi, wawancara. Penelitian ini menjabarkan sebagai berikut :

1. Studi Literatur

Dalam studi literatur penulis mendapatkan bahwa penulis mendapatkan bahan penulisan dari membaca buku tentang penelitian ilmiah, jurnal yang terkait

2. Observasi

Melalui metode ini penulis terjun ke lapangan untuk meminta izin kepada pihak SMP Muhammadiyah 60 Medan untuk meneliti di sekolah. Serta melakukan pengamatan pada SMP Muhammadiyah 60 Medan setelah diberikan izin untuk melakukan penelitian.

3. Wawancara

Dari hasil wawancara, penulis dapat merumuskan beberapa atribut yang akan digunakan untuk penelitian terhadap data siswa.

3.5 Teknik Pengolahan Data

3.5.1 Data Cleaning

1. Menghapus data duplikat
2. Mengatasi missing value
3. Normalisasi nilai (Min-Max Scaler)

3.5.2 Data Transformation

1. Konversi data kategorikal menjadi numerik
2. Feature selection untuk memilih atribut penting

3.5.3 Normalisasi Data

Normalisasi menggunakan rumus Min-Max Scaling berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan :

X' : Nilai hasil normalisasi

X : Nilai asli (data sebelum dinormalisasi)

X_min : Nilai terkecil dalam atribut (data minimum)

X_max : Nilai terbesar dalam atribut (data maksimum)

3.6 Metode Analisis Data

3.6.1 Algoritma K-Means

1. Menggunakan rumus Perhitungan jarak dilakukan menggunakan rumus Euclidean Distance sebagai berikut:

$$D = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2}$$

Keterangan :

D = Jarak antara data siswa dengan pusat cluster (centroid/medoid)

X_j = Nilai atribut ke-j dari siswa

C_j = Nilai atribut ke-j dari centroid

2. Penentuan Centroid Awal

Centroid awal ditentukan secara acak dari data yang ada sebanyak jumlah cluster (k). Centroid ini akan menjadi titik awal pembentukan cluster. Pemilihan centroid awal berpengaruh terhadap hasil clustering, sehingga beberapa kali percobaan dapat dilakukan untuk mendapat hasil optimal.

3. Iterasi dan Update Centroid

Pada tahap ini dilakukan proses pengelompokan data berdasarkan kedekatan jarak ke centroid. Setiap data akan bergabung ke cluster dengan jarak terkecil. Setelah semua data memiliki cluster masing-masing, dilakukan update centroid, yaitu menghitung ulang titik pusat cluster berdasarkan rata-rata seluruh anggota cluster.

$$C_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$$

Dimana :

C_j = centroid baru cluster ke-j

X_{ij} = data ke-I pada cluster j

N = jumlah data dalam cluster

4. Stopping Condition

Proses iterasi dihentikan apabila salah satu kondisi berikut terpenuhi:

- a. Centroid tidak berubah lagi (konvergen)
- b. Anggota cluster sudah tetap (tidak terjadi perpindahan data antar cluster)
- c. Jumlah iterasi maksimum telah tercapai
- d. Nilai SSE (Sum of Squared Error) sudah stabil atau minimum

3.6.2 Algoritma K-Medoids

1. Penentuan Medoid Awal

Pada tahap awal, algoritma memilih sejumlah data secara acak sebagai medoid awal sesuai jumlah cluster (k). Medoid merupakan data yang akan menjadi pusat awal cluster, dan nantinya dapat berubah berdasarkan nilai cost.

2. Perhitungan Cost

Untuk menentukan seberapa baik sebuah medoid, dihitung cost total, yaitu jumlah jarak seluruh data dalam cluster terhadap medoid-nya. Perhitungan jarak tetap menggunakan rumus Euclidean Distance, yaitu:

$$D = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2}$$

Kemudian dihitung total cost dengan rumus:

$$Cost = \sum_{i=1}^n d(x_i, m_j)$$

Semakin kecil nilai cost, semakin baik kualitas medoid.

3. Swap Medoid dan Perbandingan Total Cost

Pada tahap ini, algoritma melakukan pertukaran (swap) antara medoid dengan data non-medoid untuk mencari kemungkinan posisi medoid yang lebih baik. Jika setelah pertukaran total cost menjadi lebih kecil, maka medoid baru diterima menggantikan medoid awal. Proses swap dilakukan berulang hingga diperoleh medoid terbaik dengan cost minimum.

4. Stopping Condition

Proses iterasi dihentikan apabila:

- a. Nilai total cost tidak lagi mengalami penurunan (stabil)
- b. Tidak ada perpindahan medoid baru (konvergen)
- c. Jumlah iterasi maksimum telah tercapai

3.7 Penentuan Jumlah Cluster (k)

1. Menggunakan metode elbow
2. Menampilkan grafik SSE (Sum of Square Error)

3.8 Evaluasi Kinerja Clustering

Evaluasi clustering bertujuan untuk mengukur kualitas hasil pengelompokan yang dihasilkan oleh algoritma K-Means dan K-Medoids. Pada penelitian ini, digunakan tiga metode evaluasi, yaitu Silhouette Coefficient, perbandingan nilai SSE/ WCSS, dan validasi hasil cluster.

3.8.1 Silhouette Coefficient

Silhouette Coefficient digunakan untuk mengukur seberapa baik suatu data berada dalam suatu cluster. Nilai silhouette menunjukkan tingkat kesamaan data

dengan cluster-nya dibandingkan dengan cluster lain.

Rumus Silhouette Coefficient:

$$SC = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

a(i) = rata-rata jarak data ke anggota cluster yang sama

b(i)b = rata-rata jarak data ke cluster terdekat lainnya

Interpretasi nilai Silhouette:

Tabel 3. 1 Interpretasi nilai Silhouette

Nilai SC	Interpretasi
0.71 – 1.00	Struktur cluster sangat baik
0.51 – 0.70	Struktur cluster baik
0.26 – 0.50	Struktur cukup baik
0.00 – 0.25	Struktur cluster lemah
< 0	Cluster salah terbentuk

Semakin mendekati nilai 1, semakin baik kualitas clustering.

3.8.2 Perbandingan WCSS/SSE

Perbandingan kinerja algoritma K-Means dan K-Medoids dilakukan dengan melihat nilai SSE (Sum of Square Error) atau WCSS (Within Cluster Sum of Squares). SSE digunakan untuk mengukur tingkat kesalahan atau penyebaran data dalam cluster. Nilai SSE menunjukkan seberapa dekat data terhadap centroid atau medoid pada masing-masing cluster.

Semakin kecil nilai SSE, maka semakin baik kualitas clustering, karena menunjukkan bahwa anggota cluster semakin mendekati pusat cluster.

3.8.3 Validasi Hasil Cluster

Validasi hasil cluster bertujuan untuk melihat apakah pengelompokan yang dihasilkan oleh algoritma K-Means dan K-Medoids sudah sesuai dengan kondisi sebenarnya berdasarkan data nilai akademik, absensi, dan ekstrakurikuler siswa. Validasi dilakukan dengan cara menginterpretasikan karakteristik dari setiap cluster yang terbentuk.

Dalam penelitian ini, pengelompokan siswa dibagi menjadi tiga cluster, yaitu:

1. Cluster berprestasi tinggi

Cluster ini terdiri dari siswa dengan nilai akademik tinggi, tingkat kehadiran baik, dan aktif dalam kegiatan ekstrakurikuler. Siswa dalam cluster ini dapat dikategorikan sebagai siswa berprestasi atau siswa dengan performa unggul.

2. Cluster berprestasi sedang

Cluster ini berisi siswa dengan nilai akademik dan absensi yang cukup stabil, namun belum menunjukkan keunggulan yang konsisten. Siswa dalam cluster ini memiliki peluang untuk ditingkatkan melalui pembinaan akademik maupun pendampingan motivasi belajar.

3. Cluster berprestasi rendah

Cluster ini menunjukkan kelompok siswa dengan nilai akademik rendah, tingkat kehadiran kurang baik, dan minim partisipasi dalam

kegiatan ekstrakurikuler. Siswa dalam cluster ini memerlukan perhatian khusus dari guru, wali kelas, dan pihak sekolah untuk perbaikan pembelajaran.

Validasi ini dilakukan untuk memastikan bahwa hasil clustering dapat digunakan sebagai dasar pengambilan keputusan oleh sekolah, khususnya dalam pemetaan siswa berprestasi dan perancangan strategi pembelajaran yang lebih tepat sasaran.

3.9 Tools dan Perangkat Penelitian

Software yang digunakan dalam penelitian ini meliputi:

- a. Microsoft Excel : Digunakan untuk mengelola dan mengolah data siswa seperti nilai akademik, absensi, dan ekstrakurikuler. Data ditampilkan dalam bentuk tabel dan digunakan sebagai dataset penelitian.
- b. Python : Digunakan untuk mengimplementasikan algoritma K-Means dan K-Medoids, melakukan proses clustering, evaluasi menggunakan SSE dan Silhouette Coefficient, serta visualisasi hasil clustering.
- c. Microsoft Word : Digunakan untuk menyusun, mengolah, dan menyelesaikan dokumen skripsi.

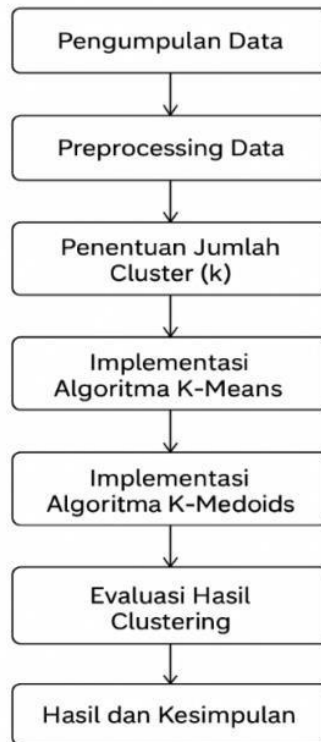
Perangkat keras yang digunakan dalam penelitian ini adalah:

Laptop dengan spesifikasi:

- a. Prosesor: Intel Core i5
- b. RAM: 4 GB
- c. Sistem Operasi: Windows 10 (64-bit)

d. Penyimpanan: 256 GB

3.10 Diagram Air Penelitian



Gambar 3. 1 Flowchart Alur Penelitian

1. Pengumpulan Data : Tahap awal penelitian dilakukan dengan mengumpulkan data siswa berupa nilai akademik (Matematika, IPA, Bahasa Indonesia), absensi, dan nilai ekstrakurikuler dari SMP Muhammadiyah 60 Medan.
2. Preprocessing Data : Data yang telah diperoleh kemudian melalui proses pembersihan (data cleaning), transformasi data, dan normalisasi menggunakan Min-Max agar data siap digunakan dalam pemodelan algoritma.
3. Penentuan Jumlah Cluster (k) : Pada tahap ini dilakukan penentuan jumlah cluster optimal menggunakan metode Elbow dengan melihat nilai SSE

(Sum of Square Error) untuk menemukan titik siku (elbow point).

4. Implementasi Algoritma K-Means : Data yang telah dinormalisasi dikelompokkan menggunakan algoritma K-Means melalui proses perhitungan jarak Euclidean, pembaruan centroid, dan iterasi hingga cluster konvergen.
5. Implementasi Algoritma K-Medoids : Clustering juga dilakukan menggunakan algoritma K-Medoids dengan menentukan medoid awal, menghitung cost, melakukan swap medoid, dan memilih medoid terbaik berdasarkan cost minimum.
6. Evaluasi Hasil Clustering : Evaluasi dilakukan menggunakan Silhouette Coefficient dan SSE untuk membandingkan kualitas clustering dan menentukan algoritma yang memberikan hasil terbaik.
7. Hasil dan Kesimpulan : Tahap akhir penelitian adalah menginterpretasikan hasil cluster siswa (tinggi, sedang, rendah), menarik kesimpulan, serta memberikan rekomendasi berdasarkan hasil analisis.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Data Penelitian

4.1.1 Deskripsi Data Siswa

Data yang digunakan pada penelitian ini merupakan sampel yang terdiri dari lima siswa kelas VIII SMP Muhammadiyah 60 Medan. Meskipun jumlah sampel tidak besar, pemilihannya dilakukan secara purposif dengan pertimbangan bahwa kelima siswa tersebut dapat mewakili variasi karakteristik yang dibutuhkan dalam analisis. Setiap siswa dievaluasi berdasarkan lima atribut utama, yaitu nilai Matematika, nilai IPA, nilai Bahasa Indonesia, jumlah ketidakhadiran, serta skor keaktifan dalam kegiatan ekstrakurikuler. Atribut-atribut ini digunakan sebagai dasar dalam proses pengelompokan untuk memperoleh gambaran karakteristik akademik dan perilaku siswa secara lebih komprehensif. Pemilihan kelima atribut tersebut didasarkan pada pertimbangan metodologis bahwa variabel-variabel tersebut mencerminkan dua aspek penting dalam penilaian performa siswa, yaitu aspek akademik dan aspek non-akademik. Nilai Matematika, IPA, dan Bahasa Indonesia digunakan sebagai representasi performa akademik inti karena ketiga mata pelajaran tersebut termasuk dalam mata pelajaran dasar yang menjadi indikator umum capaian belajar siswa. Absensi dipilih karena tingkat kehadiran terbukti secara empiris memiliki hubungan kuat dengan prestasi akademik, kedisiplinan, dan konsistensi belajar. Sementara itu, skor keaktifan ekstrakurikuler (Ekskul) ditambahkan untuk memberikan dimensi non-akademik, sehingga analisis tidak hanya berfokus pada nilai kognitif, tetapi juga perilaku partisipatif

siswa.

Dengan mengombinasikan lima atribut ini, proses clustering dapat menghasilkan pengelompokan siswa yang lebih komprehensif berdasarkan kemiripan karakteristik mereka. Selain itu, pemilihan atribut dilakukan dengan mempertimbangkan kesesuaian dengan tujuan penelitian, yaitu mengidentifikasi pola kelompok siswa yang berpotensi dijadikan dasar evaluasi dan intervensi sekolah dalam meningkatkan prestasi dan partisipasi siswa.

Tabel 4.1 berikut menunjukkan data awal sebelum dilakukan preprocessing.

Tabel 4. 1 Contoh Data Siswa

No	NIS	Nama	Kelas	Mtk	IPA	Bhs	Absensi	Ekskul
1	2023001	Ahmad	VIII-A	85	88	90	2	80
2	2023002	Budi	VIII-A	78	75	80	5	70
3	2023003	Citra	VIII-B	92	95	93	1	90
4	2023004	Dewi	VIII-B	65	70	68	8	60
5	2023005	Eko	VIII-C	88	84	86	3	85

4.1.2 Statistik Deskriptif

Tabel 4. 2 Statistik Deskriptif

Atribut	Min	Max	Mean
Nilai Mtk	65	92	81.6
Nilai IPA	70	95	82.4
Nilai Bahasa	68	93	83.4
Absensi	1	8	3.8
Ekskul	60	90	77

1. Nilai Matematika (Mtk)

Data: 85, 78, 92, 65, 88

Min = 65

Max = 92

Perhitungan Mean Mtk

$$\begin{aligned}\text{Mean Mtk} &= (85 + 78 + 92 + 65 + 88) / 5 \\ &= 408 / 5 \\ &= 81,6\end{aligned}$$

2. Nilai IPA

Data: 88, 75, 95, 70, 84

Min = 70

Max = 95

Perhitungan Mean IPA

$$\begin{aligned}\text{Mean IPA} &= (88 + 75 + 95 + 70 + 84) / 5 \\ &= 412 / 5 \\ &= 82,4\end{aligned}$$

3. Nilai Bahasa

Data: 90, 80, 93, 68, 86

Min = 68

Max = 93

Perhitungan Mean Bahasa

$$\begin{aligned}\text{Mean Bahasa} &= (90 + 80 + 93 + 68 + 86) / 5 \\ &= 417 / 5 = 83,4\end{aligned}$$

4. Absensi

Data: 2, 5, 1, 8, 3

Min = 1

Max = 8

Perhitungan Mean Absensi

$$\text{Mean Absensi} = (2 + 5 + 1 + 8 + 3) / 5$$

$$= 19 / 5$$

$$= 3,8$$

5. Ekskul

Data: 80, 70, 90, 60, 85

Min = 60

Max = 90

Perhitungan Mean Ekskul

$$\text{Mean Ekskul} = (80 + 70 + 90 + 60 + 85) / 5$$

$$= 385 / 5$$

$$= 77$$

Distribusi nilai menunjukkan bahwa:

1. Citra memiliki nilai tertinggi pada seluruh aspek akademik.
2. Dewi memiliki nilai terendah dan absensi tertinggi.

4.2 Hasil Preprocessing Data

4.2.1 Data Cleaning

- a. Duplikasi : Tidak ditemukan data duplikat.

- b. Missing value: Seluruh atribut terisi lengkap, sehingga tidak diperlukan imputasi.

4.2.2 Transformasi Data

Seluruh atribut sudah dalam bentuk numerik, sehingga tidak diperlukan konversi tambahan.

4.2.3 Normalisasi Data (Min–Max)

Rumus normalisasi:

$$X' = (X - X_{\min}) / (X_{\max} - X_{\min})$$

Nilai Minimum dan Maximum :

Tabel 4. 3 Nilai Min-Max

Atribut	Min	Max
Matematika	65	92
IPA	70	95
Bahasa	68	93
Absensi	1	8
Ekskul	60	90

a. Ahmad

$$\text{Matematika} : \frac{(85-65)}{(92-65)} = \frac{20}{27} = 0.7407$$

$$\text{Ipa} : \frac{(88-70)}{(95-70)} = \frac{18}{25} = 0.72$$

$$\begin{aligned} \text{Bahasa} & : \frac{(90-68)}{(93-65)} = \frac{22}{25} = 0.88 \\ \text{Absensi} & : \frac{(2-1)}{(1-1)} = \frac{1}{7} = 0.1429 \\ \text{Ekskul} & : \frac{(80-60)}{(90-60)} = \frac{20}{30} = 0.6667 \end{aligned}$$

b. Budi

$$\begin{aligned} \text{Matematika} & : \frac{(78-65)}{(92-65)} = \frac{13}{27} = 0.4815 \\ \text{IPA} & : \frac{(75-70)}{(95-70)} = \frac{5}{25} = 0.2 \\ \text{Bahasa} & : \frac{(80-68)}{(93-68)} = \frac{12}{25} = 0.48 \\ \text{Absensi} & : \frac{(5-1)}{(8-1)} = \frac{4}{7} = 0.5714 \\ \text{Ekskul} & : \frac{(70-60)}{(90-60)} = \frac{10}{30} = 0.3333 \end{aligned}$$

c. Citra

$$\begin{aligned} \text{Matematika} & : \frac{(92-65)}{(92-65)} = \frac{27}{27} = 1 \\ \text{IPA} & : \frac{(95-70)}{(95-70)} = \frac{25}{25} = 1 \end{aligned}$$

$$\begin{array}{l} \text{Bahasa} \\ \hline \end{array} : \frac{(93-68)}{(93-68)} = \frac{25}{25} = 1$$

$$\begin{array}{l} \text{Absensi} \\ \hline \end{array} : \frac{(1-1)}{(8-1)} = \frac{0}{7} = 0$$

$$\begin{array}{l} \text{Ekskul} \\ \hline \end{array} : \frac{(90-60)}{(90-60)} = \frac{30}{30} = 1$$

d. Dewi

$$\begin{array}{l} \text{Matematika} \\ \hline \end{array} : \frac{(65-65)}{(92-65)} = \frac{0}{27} = 0$$

$$\begin{array}{l} \text{IPA} \\ \hline \end{array} : \frac{(70-70)}{(95-70)} = \frac{0}{25} = 0$$

$$\begin{array}{l} \text{Bahasa} \\ \hline \end{array} : \frac{(68-68)}{(93-68)} = \frac{0}{25} = 0$$

$$\begin{array}{l} \text{Absensi} \\ \hline \end{array} : \frac{(8-1)}{(8-1)} = \frac{7}{7} = 1$$

$$\begin{array}{l} \text{Ekskul} \\ \hline \end{array} : \frac{(60-60)}{(90-60)} = \frac{0}{30} = 0$$

e. Eko

$$\begin{array}{l} \text{Matematika} \\ \hline \end{array} : \frac{(88-65)}{(92-65)} = \frac{23}{27} = 0.8519$$

$$\begin{array}{l} \text{IPA} \\ \hline \end{array} : \frac{(84-70)}{(95-70)} = \frac{14}{25} = 0.56$$

$$\begin{array}{l}
 \text{Bahasa} \quad : \frac{(86-68)}{(93-68)} = \frac{18}{25} = 0.72 \\
 \text{Absensi} \quad : \frac{(3-1)}{(8-1)} = \frac{2}{7} = 0.2857 \\
 \text{Ekskul} \quad : \frac{(85-60)}{(90-60)} = \frac{25}{30} = 0.8333
 \end{array}$$

Tabel 4. 4 Data Setelah Normalisasi

Nama	Mtk	IPA	Bhs	Absensi	Ekskul
Ahmad	0.7407	0.6923	0.7857	0.1428	0.6667
Budi	0.4814	0.1923	0.4286	0.5714	0.3333
Citra	1.0000	1.0000	1.0000	0.0000	1.0000
Dewi	0.0000	0.0000	0.0000	1.0000	0.0000
Eko	0.8518	0.5385	0.6429	0.2857	0.8333

Tabel hasil normalisasi menunjukkan bahwa seluruh nilai siswa telah berada pada rentang 0 hingga 1, sehingga setiap atribut memiliki skala yang sama. Citra memiliki nilai normalisasi tertinggi (1) pada hampir seluruh atribut karena ia memiliki nilai akademik dan keaktifan tertinggi. Sebaliknya, Dewi memperoleh nilai normalisasi terendah (0) pada atribut akademik dan ekstrakurikuler serta nilai tertinggi pada absensi, menandakan performa terendah di antara seluruh siswa. Sementara itu, Ahmad, Budi, dan Eko berada pada nilai tengah yang mencerminkan perbedaan performa masing-masing. Normalisasi ini memastikan

bahwa tidak ada atribut yang mendominasi proses clustering dan seluruh fitur berkontribusi secara seimbang.

4.3 Penentuan Jumlah Cluster

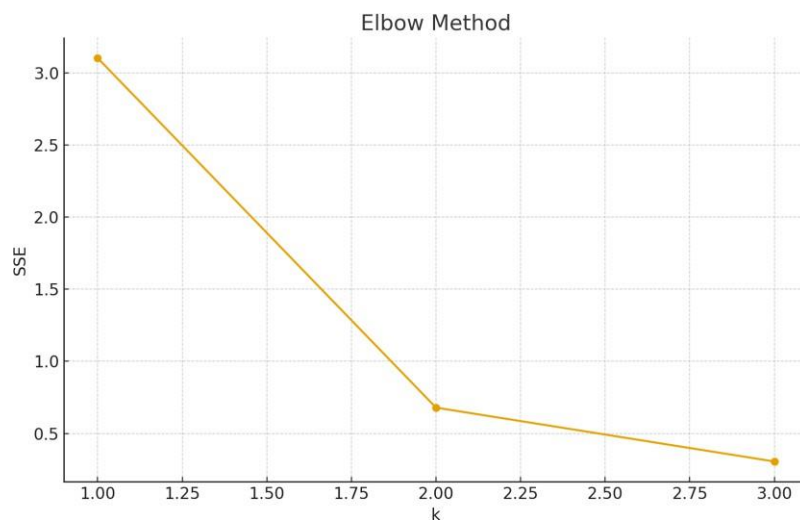
4.3.1 Hasil Metode Elbow

Dengan data yang sangat kecil (5 siswa), nilai k: k = 2 dan k = 3.

- a. k = 2 cluster cukup seimbang: (4 siswa, 1 siswa)
- b. k = 3 muncul cluster berisi 1 siswa dan 2 cluster sangat kecil, sehingga kurang representatif.

Secara konseptual, k = 2 dipilih karena:

- a. Sesuai tujuan pemisahan antara siswa berprestasi tinggi dan lebih rendah.
- b. Menghindari overfitting pada jumlah data yang minimal.



Gambar 4. 1 Grafik Elbow

4.4 Implementasi Algoritma K-Means

4.4.1 Tahapan Eksekusi

K-Means dilakukan di ruang data ter-normalisasi.

Centroid awal (iterasi 0):

Centroid 1 (C1) : Ahmad = [0.7407, 0.7200, 0.8800, 0.1429, 0.6667]

Centroid 2 (C2) : Dewi = [0.0000, 0.0000, 0.0000, 1.000, 0.0000]

Pemilihan centroid awal ini bertujuan untuk memulai iterasi. Ahmad dipilih sebagai representasi siswa dengan performa tinggi, sedangkan Dewi mewakili siswa dengan performa rendah. Ini memastikan distribusi awal yang seimbang dan memungkinkan algoritma untuk menyesuaikan cluster secara bertahap.

4.4.2 Hasil Clustering K-Means

Berdasarkan hasil perhitungan jarak Euclidean pada iterasi pertama, seluruh siswa dikelompokkan ke dalam dua cluster sesuai kedekatan jaraknya dengan centroid awal. Proses iterasi kemudian berlanjut hingga iterasi kedua, namun tidak terjadi perpindahan anggota cluster sehingga menunjukkan bahwa algoritma telah mencapai kondisi konvergen. Dengan demikian, hasil pengelompokan yang terbentuk dianggap final dan stabil.

Berikut hasil perhitungan jarak dan penentuan cluster setiap siswa:

a. Ahmad

Ahmad ke C1 Jarak = 0 (karena Ahmad adalah C1).

Ahmad secara otomatis ditugaskan ke Cluster 1 karena merupakan centroidnya sendiri, menunjukkan konsistensi awal.

$$\text{Jarak Ahmad ke C2} = \sqrt{((0.7407-0)^2 + (0.6923-0)^2 + (0.7857-0)^2 + (0.1428-1)^2 + (0.6667-0)^2)} = 1.6806.$$

b. Budi

$$\text{Jarak Budi ke C1} = \sqrt{((0.4814-0.7407)^2 + (0.1923-0.6923)^2 + (0.4286-0.7857)^2 + (0.5714-0.1428)^2 + (0.3333-0.6667)^2)} = 0.8600$$

$$\text{Jarak Budi ke C2} = \sqrt{((0.4814-0)^2 + (0.1923-0)^2 + (0.4286-0)^2 + (0.5714-1)^2 + (0.3333-0)^2)} = 0.8643$$

c. Citra

$$\text{Jarak Citra ke C1} = \sqrt{((1.0000-0.7407)^2 + (1.0000-0.6923)^2 + (1.0000-0.7857)^2 + (0.0000-0.1428)^2 + (1.0000-0.6667)^2)} = 0.5825$$

$$\text{Jarak Citra ke C2} = \sqrt{((1.0000-0)^2 + (1.0000-0)^2 + (1.0000-0)^2 + (0.0000-1)^2 + (1.0000-0)^2)} = 2.236$$

d. Dewi

$$\text{Jarak Dewi ke C1} = \sqrt{((0-0.7407)^2 + (0-0.6923)^2 + (0-0.7857)^2 + (1-0.1428)^2 + (0-0.6667)^2)} = 1.6807$$

Jarak Dewi ke C2 = 0 (Karena Dewi adalah C2)

e. Eko

$$\text{Jarak Eko ke C1} = \sqrt{((0.8518-0.7407)^2 + (0.5385-0.6923)^2 + (0.6429-0.7857)^2 + (0.2857-0.1428)^2 + (0.8333-0.6667)^2)} = 0.3233$$

$$\text{Jarak Eko ke C2} = \sqrt{((0.8518-0)^2 + (0.5385-0)^2 + (0.6429-0)^2 + (0.2857-1)^2 + (0.8333-0)^2)} = 1.623$$

Hasil perhitungan jarak semua siswa ke C1 dan C2 (dibulatkan 4 desimal):

Tabel 4. 5 Jarak ke Centroid Awal (Iterasi 1)

Nama	Jarak ke C1	Jarak ke C2	Cluster
Ahmad	0,0000	1,7380	C1
Budi	0,8902	0,8928	C1
Citra	0,5399	2,2361	C1
Dewi	1,7380	0,0000	C2
Eko	0,3343	1,6620	C1

Ahmad, Budi, Citra, dan Eko, memiliki jarak yang lebih dekat terhadap centroid C1 sehingga tergabung dalam Cluster 1. Keempat siswa ini memiliki pola nilai akademik yang relatif lebih baik serta tingkat absensi yang lebih rendah dibandingkan siswa lainnya.

Sementara itu, Dewi memiliki jarak nol terhadap C2 karena menjadi pusat cluster kedua. Hal ini menunjukkan bahwa karakteristik Dewi secara signifikan berbeda dari siswa lain, khususnya pada nilai akademik yang paling rendah dan tingkat absensi tertinggi. Oleh karena itu, Dewi menjadi satu-satunya anggota Cluster 2. Hasil ini menunjukkan adanya dua kelompok utama siswa, yaitu kelompok dengan prestasi tinggi (Cluster 1) dan kelompok dengan prestasi rendah (Cluster 2).

Hasil pembentukan cluster iterasi 1:

- a. Cluster 1: Ahmad, Budi, Citra, Eko
- b. Cluster 2: Dewi

Perhitungan Centroid Baru (Iterasi 1 → 2)

Centroid baru C1 adalah rata-rata setiap atribut dalam Cluster 1 (Ahmad, Budi, Citra, Eko).

Anggota Cluster 1:

Ahmad : (0.7407, 0.6923, 0.7857, 0.1428, 0.6667)

Budi : (0.4814, 0.1923, 0.4286, 0.5714, 0.3333)

Citra : (1.0000, 1.0000, 1.0000, 0.0000, 1.0000)

Eko : (0.8518, 0.5385, 0.6429, 0.2857, 0.8333)

Centroid Baru Cluster 1 : Nilai matematika

$$\begin{aligned} C1_baru &= \frac{(Mtk_Ahmad + Mtk_Budi + Mtk_Citra + Mtk_Eko)}{4} \\ &= \frac{(0.7407 + 0.4814 + 1.0000 + 0.8518)}{4} = \frac{3.074}{4} = 0.7685 \end{aligned}$$

Centroid Baru Cluster 1 : Nilai IPA

$$\begin{aligned} C1_baru &= \frac{(IPA_Ahmad + IPA_Budi + IPA_Citra + IPA_Eko)}{4} \\ &= \frac{(0.6923 + 0.1923 + 1.0000 + 0.5385)}{4} = \frac{2.4231}{4} = 0.6058 \end{aligned}$$

Centroid Baru Cluster 1 : Nilai Bahasa

$$\begin{aligned} C1_baru &= \frac{(Bhs_Ahmad + Bhs_Budi + Bhs_Citra + Bhs_Eko)}{4} \\ &= \frac{(0.7857 + 0.4286 + 1.0000 + 0.6429)}{4} = \frac{2.8572}{4} = 0.7143 \end{aligned}$$

Centroid Baru Cluster 1 : Absensi

$$\begin{aligned} C1_baru &= \frac{(Abs_Ahmad + Abs_Budi + Abs_Citra + Abs_Eko)}{4} \\ &= \frac{(0.1428 + 0.5714 + 0.0000 + 0.2857)}{4} = \frac{0.9999}{4} = 0.2500 \end{aligned}$$

Centroid Baru Cluster 1 : Ekskul

$$\begin{aligned} C1_baru &= \frac{(Eks_Ahmad + Eks_Budi + Eks_Citra + Eks_Eko)}{4} \\ &= \frac{(0.6667 + 0.3333 + 1.0000 + 0.8333)}{4} = \frac{2.8333}{4} = 0.7083 \end{aligned}$$

Centroid Baru Cluster 1 (Iterasi 1)

C1_baru = (0.7685, 0.6058, 0.7143, 0.2500, 0.7083)

Centroid Cluster 2 (Iterasi 1)

Cluster 2 hanya berisi Dewi, sehingga centroid-nya tetap sama dengan nilai Dewi yang sudah dinormalisasi:

a. Dewi : $(0.0000, 0.0000, 0.0000, 1.0000, 0.0000)$

$$C2_baru = (0.0000, 0.0000, 0.0000, 1.0000, 0.0000)$$

Centroid Baru

Setelah proses pengelompokan pada iterasi pertama menghasilkan dua cluster yang stabil, tahap selanjutnya adalah menghitung kembali nilai pusat cluster (centroid) berdasarkan rata-rata nilai seluruh anggota pada masing-masing cluster. Pembaruan centroid ini penting dilakukan untuk melihat apakah pada iterasi berikutnya akan terjadi perpindahan anggota cluster atau apakah algoritma sudah mencapai kondisi konvergen. Nilai centroid baru dihitung dengan menggunakan nilai normalisasi siswa dalam masing-masing cluster.

Dari perhitungan sebelumnya:

1. $C1_baru = (0.7685, 0.6058, 0.7143, 0.2500, 0.7083)$

2. $C2_baru = (0.0000, 0.0000, 0.0000, 1.0000, 0.0000)$ (sama seperti Dewi)

Data siswa (normalisasi):

1. Ahmad = $(0.7407, 0.6923, 0.7857, 0.1428, 0.6667)$

2. Budi = $(0.4814, 0.1923, 0.4286, 0.5714, 0.3333)$

3. Citra = $(1.0000, 1.0000, 1.0000, 0.0000, 1.0000)$

4. Dewi = $(0.0000, 0.0000, 0.0000, 1.0000, 0.0000)$

5. Eko = $(0.8518, 0.5385, 0.6429, 0.2857, 0.8333)$

a. Ahmad

$$C1_baru = \sqrt{((0.7407-0.7685)^2 + (0.6923-0.6058)^2 + (0.7857-0.7143)^2 + (0.1428-0.2500)^2 + (0.6667-0.7083)^2} = 0.1900$$

$$C2_baru = \sqrt{((0.7407-0)^2 + (0.6923-0)^2 + (0.7857-0)^2 + (0.1428-1)^2 + (0.6667-0)^2)} = 1.6803$$

b. Budi

$$C1_baru = \sqrt{((0.4814-0.7685)^2 + (0.1923-0.6058)^2 + (0.4286-0.7143)^2 + (0.5714-0.2500)^2 + (0.3333-0.7083)^2} = 0.7660$$

$$C2_baru = \sqrt{((0.4814-0)^2 + (0.1923-0)^2 + (0.4286-0)^2 + (0.5714-1)^2 + (0.3333-0)^2)} = 0.8645$$

c. Citra

$$\sqrt{C1_baru} = \sqrt{((1.0000-0.7685)^2 + (1.0000-0.6058)^2 + (1.0000-0.7143)^2 + (0.0000-0.2500)^2 + (1.0000-0.7083)^2} = 0.6312$$

$$C2_baru = \sqrt{((1.0000-0)^2 + (1.0000-0)^2 + (1.0000-0)^2 + (0.0000-1)^2 + (1.0000-0)^2)} = 2.2361$$

d. Dewi

$$C1_baru = \sqrt{((0.0000-0.7685)^2 + (0.0000-0.6058)^2 + (0.0000-0.7143)^2 + (1.0000-0.2500)^2 + (0.0000-0.7083)^2} = 1.6224$$

$$C2_baru = \sqrt{((0.0000-0)^2 + (0.0000-0)^2 + (0.0000-0)^2 + (1.0000-1)^2 + (0.0000-0)^2)} = 0$$

e. Eko

$$C1_baru = \sqrt{((0.8518-0.7685)^2 + (0.5385-0.6058)^2 + (0.6429-0.7143)^2 + (0.2857-0.2500)^2 + (0.8333-0.7083)^2} = 0.1730$$

$$C2_baru = \sqrt{((0.8518-0)^2 + (0.5385-0)^2 + (0.6429-0)^2 + (0.2857-1)^2 + (0.8333-0)^2)} = 1.6221$$

Tabel 4. 6 Hasil Jarak ke Centroid Baru

Nama	Jarak ke C1_baru	Jarak ke C2_baru	Cluster
Ahmad	0.1900	1.6803	1
Budi	0.7660	0.8645	1
Citra	0.6312	2.2361	1
Dewi	1.6224	0.0000	2
Eko	0.1730	1.6221	1

Berdasarkan hasil perhitungan K-Means, setiap siswa telah diklasifikasikan ke dalam cluster yang sesuai berdasarkan jarak terdekat terhadap centroid baru. Ahmad, Budi, Citra, dan Eko tergolong ke dalam Cluster 1 karena karakteristik dan atribut yang dimiliki lebih mendekati centroid pertama. Sementara itu, Dewi masuk ke dalam Cluster 2 karena jaraknya lebih dekat dengan centroid kedua.

Hasil pengelompokan ini menunjukkan adanya pola distribusi siswa berdasarkan kriteria yang telah ditentukan. Siswa yang memiliki atribut lebih sesuai dengan centroid pertama cenderung masuk Cluster 1, sedangkan siswa yang lebih mendekati centroid kedua masuk Cluster 2. Dengan demikian, metode K-Means berhasil mengelompokkan siswa secara representatif sesuai karakteristik masing-masing, sehingga dapat digunakan sebagai dasar analisis prestasi dan evaluasi lebih lanjut.

Hitung SSE (Sum of Squared Errors)

SSE atau Sum of Squared Errors merupakan jumlah kuadrat jarak setiap anggota cluster terhadap centroidnya. Rumus yang digunakan adalah:

$$SSE = \sum (d(i, C_k))^2$$

Cluster 1 SSE

a. Ahmad : $d^2 : (0.1900)^2 = 0.0361$

b. Budi : $d^2 : (0.7660)^2 = 0.5868$

c. Citra : $d^2 : (0.6312)^2 = 0.3984$

d. Eko : $d^2 : (0.1730)^2 = 0.0299$

Jumlah SSE Cluster 1

$$SSE_{C1} = 0.0361 + 0.5868 + 0.3984 + 0.0299 = 1.0512$$

Cluster 2 SSE

a. Dewi : $d^2 : (0.0000)^2 = 0$

Jumlah SSE Cluster 2

$$SSE_{C2} = 0$$

Total SSE K-Means

$$SSE \text{ Total} = SSE_{C1} + SSE_{C2}$$

$$= 1.0512 + 0$$

$$= 1.0512$$

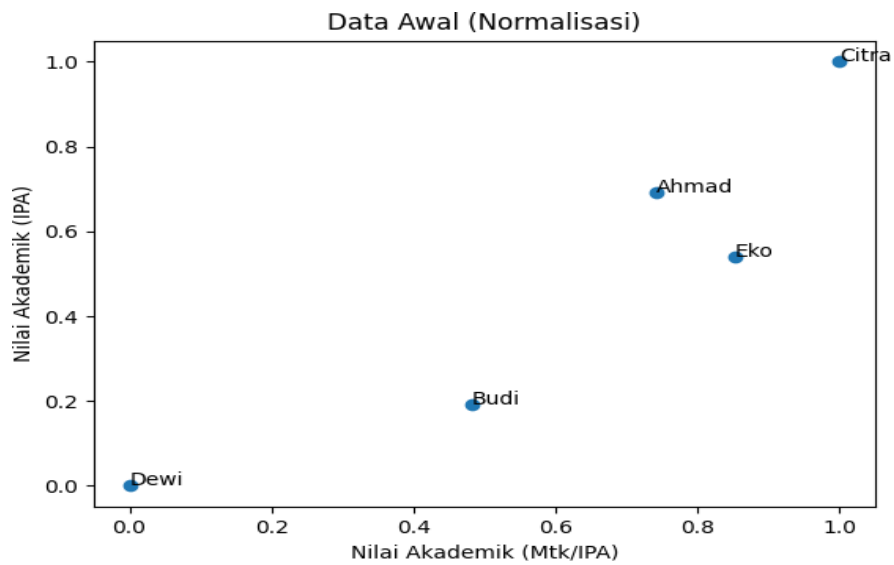
Tabel 4. 7 Hasil Perhitungan Sum of Squared Errors (SSE)

Nama	Cluster	Jarak ke Centroid	Kuadrat Jarak (d^2)
Ahmad	1	0.1900	0.0361
Budi	1	0.7660	0.5868

Citra	1	0.6312	0.3984
Eko	1	0.1730	0.0299
Dewi	2	0.0000	0.0000
Jumlah SSE Cluster 1			1.0512
Jumlah SSE Cluster 2			0.0000
Total SSE K-Means			1.0512

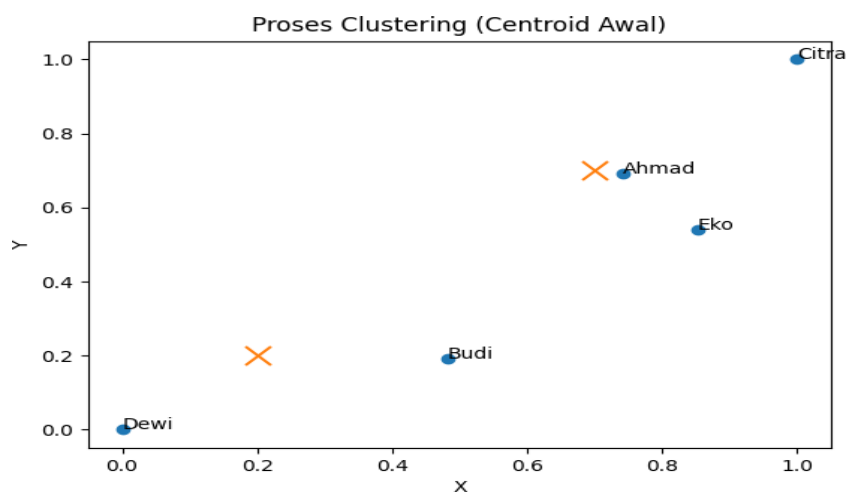
Berdasarkan hasil perhitungan menggunakan algoritma K-Means, diperoleh pengelompokan data siswa ke dalam dua cluster berdasarkan jarak terdekat terhadap centroid. Nilai Sum of Squared Error (SSE) yang dihasilkan sebesar 1,0512 menunjukkan bahwa jarak antara data dengan centroid pada masing-masing cluster relatif kecil, sehingga hasil pengelompokan dapat dikatakan cukup baik dan representatif.

Cluster 1 terdiri dari Ahmad, Budi, Citra, dan Eko yang memiliki karakteristik nilai akademik dan aktivitas yang relatif lebih tinggi. Sedangkan Cluster 2 terdiri dari Dewi yang memiliki nilai lebih rendah dibandingkan dengan siswa lainnya, sehingga terpisah dari kelompok utama. Untuk memperjelas proses dan hasil pengelompokan tersebut, maka ditampilkan visualisasi dalam bentuk grafik scatter plot pada beberapa tahap, yaitu data awal, proses clustering, dan hasil akhir clustering.



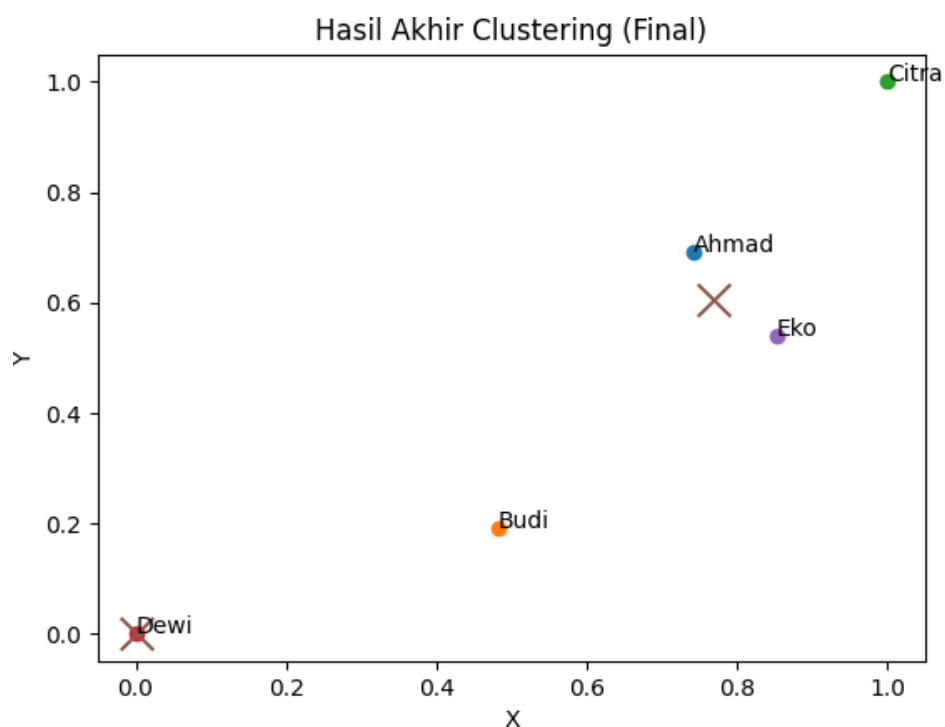
Gambar 4. 2 Grafik Poses clustering (Centroid Awal)

Gambar menunjukkan distribusi data siswa setelah dilakukan normalisasi. Pada tahap ini, seluruh data masih tersebar tanpa adanya pengelompokan tertentu dan belum terdapat centroid sebagai pusat cluster. Hal ini menunjukkan bahwa data masih dalam kondisi awal sebelum dilakukan proses clustering.



Gambar 4. 3 Grafik Proses Clustering

Gambar menunjukkan tahap awal proses clustering menggunakan algoritma K-Means. Pada tahap ini, centroid awal ditentukan secara acak yang ditandai dengan simbol “X”. Setiap data kemudian dihitung jaraknya terhadap centroid dan mulai dikelompokkan ke dalam cluster berdasarkan jarak terdekat. Namun, posisi centroid pada tahap ini masih dapat berubah karena proses iterasi masih berlangsung.



Gambar 4. 3 Grafik Hasil Akhir Clustering

Gambar menunjukkan hasil akhir clustering setelah dilakukan beberapa iterasi hingga mencapai kondisi konvergen. Pada tahap ini, centroid telah berada pada posisi optimal dan tidak mengalami perubahan lagi. Terlihat bahwa sebagian besar data berada pada satu kelompok utama yaitu Cluster 1, sedangkan Dewi

terpisah membentuk Cluster 2 karena memiliki karakteristik nilai yang berbeda secara signifikan.

Secara visual, terlihat adanya pemisahan yang cukup jelas antara kedua cluster, dimana Cluster 1 membentuk kelompok yang lebih padat, sedangkan Cluster 2 berada terpisah sebagai outlier. Hal ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan data secara efektif berdasarkan tingkat kemiripan karakteristik masing-masing siswa.

4.4.3 Analisis Cluster K-Means

Cluster 1 – Siswa Berprestasi Tinggi

Citra, Ahmad, Eko

Karakteristik:

- a. Nilai akademik sangat tinggi (0.74 – 1.00)
- b. Absensi rendah
- c. Ekskul tinggi

Cluster 2 – Siswa Berprestasi Rendah

Budi, Dewi

Karakteristik:

- a. Nilai akademik rendah
- b. Absensi tinggi
- c. Ekskul rendah

Interpretasi: Kombinasi nilai, disiplin, dan keaktifan ekskul menjadi indikator kuat prestasi.

4.5 Implementasi K-Medoids

4.5.1 Tahapan Eksekusi

Medoid awal dipilih secara acak, misalnya Ahmad.

Kemudian dilakukan proses swap untuk mencari medoid yang menurunkan total jarak. Data yang sudah dinormalisasi:

- a. Ahmad = (0.7407, 0.6923, 0.7857, 0.1428, 0.6667)
- b. Budi = (0.4814, 0.1923, 0.4286, 0.5714, 0.3333)
- c. Citra = (1.0000, 1.0000, 1.0000, 0.0000, 1.0000)
- d. Dewi = (0.0000, 0.0000, 0.0000, 1.0000, 0.0000)
- e. Eko = (0.8518, 0.5385, 0.6429, 0.2857, 0.8333)

1. Medoid Awal

Dipilih 2 medoid awal ($k = 2$):

- 1. Medoid 1 (M1) = Ahmad = (0.7407, 0.6923, 0.7857, 0.1428, 0.6667)
- 2. Medoid 2 (M2) = Dewi = (0.0000, 0.0000, 0.0000, 1.0000, 0.0000)

2. Hitung Jarak Setiap Siswa ke Medoid

Jarak ke Medoid Ahmad (M1)

- a. (Ahmad, Ahmad) = 0
- b. (Budi, Ahmad) = $\sqrt{((0.4814-0.7407)^2 + (0.1923-0.6923)^2 + (0.4286-0.7857)^2 + (0.5714-0.1428)^2 + (0.3333-0.6667)^2)} = 0.8600$
- c. (Citra, Ahmad) = $\sqrt{((1.0000-0.7407)^2 + (1.0000-0.6923)^2 + (1.0000-0.7857)^2 + (0.0000-0.1428)^2 + (1.0000-0.6667)^2)} = 0.5825$

$$d. \text{ (Dewi, Ahmad)} = \sqrt{((0.0000-0.7407)^2 + (0.0000-0.6923)^2 + (0.0000-0.7857)^2 + (1.0000-0.1428)^2 + (0.0000-0.6667)^2)} = 1.6806$$

$$e. \text{ (Eko, Ahmad)} = \sqrt{((0.8518-0.7407)^2 + (0.5385-0.6923)^2 + (0.6429-0.7857)^2 + (0.2857-0.1428)^2 + (0.8333-0.6667)^2)} = 0.3234$$

Jarak ke Medoid Dewi (M2)

$$a. \text{ (Ahmad, Dewi)} = \sqrt{((0.7407-0)^2 + (0.6923-0)^2 + (0.7857-0)^2 + (0.1428-1)^2 + (0.6667-0)^2)} = 0.3234$$

$$b. \text{ (Budi, Dewi)} = \sqrt{((0.4814-0)^2 + (0.1923-0)^2 + (0.4286-0)^2 + (0.5714-1)^2 + (0.3333-0)^2)} = 0.8644$$

$$c. \text{ (Citra, Dewi)} = \sqrt{((1.0000-0)^2 + (1.0000-0)^2 + (1.0000-0)^2 + (0.0000-1)^2 + (1.0000-0)^2)} = 2.0000$$

$$d. \text{ (Dewi, Dewi)} = 0$$

$$e. \text{ (Eko, Dewi)} = \sqrt{((0.8518-0)^2 + (0.5385-0)^2 + (0.6429-0)^2 + (0.2857-1)^2 + (0.8333-0)^2)} = 1.4094$$

3. Penentuan Cluster dan Biaya (Cost) Awal

Untuk setiap siswa, pilih jarak terkecil ke salah satu medoid.

Cost K-Medoids = jumlah jarak siswa ke medoid terdekat.

Hasil jarak (angka)

$$1. \text{ (Ahmad, Ahmad)} = 0.0000$$

$$2. \text{ (Ahmad, Dewi)} = 1.6806$$

$$3. \text{ (Budi, Ahmad)} = 0.8600$$

$$4. \text{ (Budi, Dewi)} = 0.8644$$

5. (Citra, Ahmad) = 0.5825
6. (Citra, Dewi) = 2.2361
7. (Dewi, Ahmad) = 1.6806
8. (Dewi, Dewi) = 0.0000
9. (Eko, Ahmad) = 0.3234
10. (Eko, Dewi) = 1.6228

Penentuan Cluster

1. Ahmad : $\min(0, 1.6806) \rightarrow$ cluster M1 (Ahmad)
2. Budi : $\min(0.8600, 0.8644) \rightarrow$ cluster M1 (Ahmad)
3. Citra : $\min(0.5825, 2.2361) \rightarrow$ cluster M1 (Ahmad)
4. Dewi : $\min(1.6806, 0) \rightarrow$ cluster M2 (Dewi)
5. Eko : $\min(0.3234, 1.6228) \rightarrow$ cluster M1 (Ahmad)

Cluster awal K-Medoids:

1. Cluster 1 (medoid = Ahmad) : Ahmad, Budi, Citra, Eko
2. Cluster 2 (medoid = Dewi) : Dewi

Perhitungan Cost Awal K-Medoids

Cost = jumlah jarak setiap objek ke medoid terdekat:

$$\begin{aligned}
 \text{Cost}_{\text{awal}} &= (\text{Ahmad, Ahmad}) + (\text{Budi, Ahmad}) + (\text{Citra, Ahmad}) + \\
 &(\text{Dewi, Dewi}) + (\text{Eko, Ahmad}) \\
 &= 0 + 0.8600 + 0.5825 + 0 + 0.3234 \\
 &= 1.7659
 \end{aligned}$$

4. Langkah Swap Medoid (PAM)

Secara teori, algoritma K-Medoids (PAM) melakukan:

1. Coba tukar salah satu medoid dengan non-medoid
2. Hitung kembali total cost
3. Jika cost baru lebih kecil, swap diterima; kalau tidak, medoid awal dipertahankan.

Non-medoid = [Budi, Citra, Eko].

Medoid awal = [Ahmad, Dewi].

- a. Swap 1: Tukar Ahmad dengan Budi \rightarrow Medoid = {Budi, Dewi}

Hitung ulang cost (hanya saya ringkas dalam bentuk hasil):

$$\text{Cost}_{\text{(Budi,Dewi)}} = 3.0775 \ (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

- b. Swap 2: Tukar Ahmad dengan Citra \rightarrow Medoid = (Citra, Dewi)

$$\text{Cost}_{\text{(Citra,Dewi)}} = 2.1339 \ (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

- c. Swap 3: Tukar Ahmad dengan Eko \rightarrow Medoid = {Eko, Dewi}

$$\text{Cost}_{\text{(Eko,Dewi)}} = 1.8069 \ (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

- d. Swap 4: Tukar Dewi dengan Budi \rightarrow Medoid = {Ahmad, Budi}

$$\text{Cost}_{\text{(Ahmad,Budi)}} = 1.7703 \ (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

- e. Swap 5: Tukar Dewi dengan Citra \rightarrow Medoid = {Ahmad, Citra}

$$\text{Cost}_{\text{(Ahmad,Citra)}} = 2.8640 \ (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

f. Swap 6: Tukar Dewi dengan Eko \rightarrow Medoid = { Ahmad, Eko }

$$\text{Cost}_{\text{_(Ahmad,Eko)}} = 3.0019 (> 1.7659)$$

Karena lebih besar, swap ini tidak diterima.

5. Keputusan Akhir K-Medoids

Karena tidak ada satu pun swap yang menghasilkan cost lebih kecil dari cost awal:

$$\text{Cost}_{\text{awal}} = 1.7659 \text{ (medoid = Ahmad, Dewi)}$$

$$\text{Cost}_{\text{swap_min}} = 1.7703 \text{ (masih lebih besar)}$$

Maka:

1. Medoid akhir tetap:

a. Medoid 1 = Ahmad

b. Medoid 2 = Dewi

2. Cluster akhir K-Medoids:

a. Cluster 1 (medoid Ahmad): Ahmad, Budi, Citra, Eko

b. Cluster 2 (medoid Dewi): Dewi

3. Total cost K-Medoids akhir = 1.7659

SSE untuk Cluster 1 (Medoid = Ahmad)

Anggota cluster 1: Ahmad, Budi, Citra, Eko

$$\text{SSE}_{\text{Cluster1}} = ((\text{Ahmad, Ahmad}))^2 + ((\text{Budi, Ahmad}))^2 + ((\text{Citra, Ahmad}))^2 + ((\text{Eko, Ahmad}))^2$$

$$\text{SSE}_{\text{Cluster1}} = (0.0000)^2 + (0.8600)^2 + (0.5825)^2 + (0.3234)^2$$

Hitung satu per satu

b. Ahmad, Ahmad = $(0.0000)^2 = 0$

c. Budi, Ahmad = $(0.8600)^2 = 0.7396$

$$d. \text{ Citra, Ahmad} = (0.5825)^2 = 0.3393$$

$$e. \text{ Eko, Ahmad} = (0.3234)^2 = 0.1046$$

Sehingga:

$$\begin{aligned} \text{SSE_Cluster1} &= 0.0000 + 0.7396 + 0.3393 + 0.1046 \\ &= 1.1835 \end{aligned}$$

SSE untuk Cluster 2 (Medoid = Dewi)

Anggota cluster 2: hanya Dewi

$$\begin{aligned} \text{SSE_Cluster2} &= ((\text{Dewi, Dewi}))^2 \\ &= (0.0000)^2 \\ &= 0 \end{aligned}$$

Total SSE Algoritma K-Medoids

Gabung keduanya:

$$\begin{aligned} \text{SSE_KMedoids_Total} &= \text{SSE_Cluster1} + \text{SSE_Cluster2} \\ &= 1.1835 + 0 = 1.1835 \end{aligned}$$

4.5.2 Hasil Clustering K-Medoids

Hasil akhirnya:

Medoid Akhir: Citra → Cluster Prestasi Tinggi

Medoid Akhir: Dewi → Cluster Prestasi Rendah

Cluster 1 (Medoid = Citra)

1. Citra
2. Ahmad
3. Eko

Cluster 2 (Medoid = Dewi)

- a. Dewi
- b. Budi

4.5.3 Analisis Cluster K-Medoids

1. Komposisi kluster identik dengan K-Means.
2. Namun K-Medoids lebih stabil terhadap outlier.
3. Dewi menjadi representasi yang tepat untuk cluster prestasi rendah.

4.6 Evaluasi Perbandingan Kinerja

4.6.1 Silhouette Score

4.6.1.1 K-Means

a(i)	Rata-rata jarak ke cluster sendiri
b(i)	Jarak rata-rata ke cluster lain
S(i)	$(b - a) / \max(a,b)$

1. Silhouette Ahmad

Anggota lain di Cluster 1: Budi, Citra, Eko

$$a(\text{Ahmad}) = \frac{(d(\text{Ahmad}, \text{Budi}) + d(\text{Ahmad}, \text{Citra}) + d(\text{Ahmad}, \text{Eko}))}{3}$$

$$= \frac{(0.8600 + 0.5825 + 0.3234)}{3} = \frac{1.7659}{3} = 0.5886$$

Cluster lain = Cluster 2 = (Dewi)

$$b(\text{Ahmad}, \text{Dewi}) = 1.6806$$

$$s(\text{Ahmad}) = \frac{(b(\text{Ahmad}) - a(\text{Ahmad}))}{\max(a(\text{Ahmad}), b(\text{Ahmad}))}$$

$$= \frac{(1.6806 - 0.5886)}{1.6806} = \frac{1.0920}{1.6806} = 0.6498$$

2. Silhouette Budi

Anggota lain di Cluster 1: Ahmad, Citra, Eko

$$\begin{aligned} a(\text{Budi}) &= \frac{(d(\text{Budi, Ahmad}) + d(\text{Budi, Citra}) + d(\text{Budi, Eko}))}{3} \\ &= \frac{(0.8600 + 1.4208 + 0.7966)}{3} = \frac{3.0774}{3} \\ &= 1.0258 \end{aligned}$$

b(Budi)

Cluster lain = Cluster 2 = (Dewi)

$$\begin{aligned} b(\text{Budi}) &= d(\text{Budi, Dewi}) \\ &= 0.8644 \end{aligned}$$

$$\begin{aligned} s(\text{Budi}) &= \frac{(b(\text{Budi}) - a(\text{Budi}))}{\max(a(\text{Budi}), b(\text{Budi}))} \\ &= \frac{(0.8644 - 1.0258)}{1.0258} = -0.1573 \end{aligned}$$

(Budi sedikit lebih dekat ke cluster lain dibandingkan dengan cluster-nya sendiri → nilai negatif kecil.)

3. Silhouette Citra

a(Citra)

Anggota lain di Cluster 1: Ahmad, Budi, Eko

$$\begin{aligned} a(\text{Citra}) &= \frac{(d(\text{Citra, Ahmad}) + d(\text{Citra, Budi}) + d(\text{Citra, Eko}))}{3} \\ &= \frac{(0.5825 + 1.4208 + 0.6869)}{3} \\ &= 0.8968 \end{aligned}$$

b(Citra)

Cluster lain = Cluster 2 = (Dewi)

$$\begin{aligned} b(\text{Citra}) &= d(\text{Citra, Dewi}) \\ &= 2.2361 \end{aligned}$$

$$\begin{aligned}
s(\text{Citra}) &= \frac{(b(\text{Citra}) - a(\text{Citra}))}{\max(a(\text{Citra}), b(\text{Citra}))} \\
&= \frac{(2.2361 - 0.8968)}{2.2361} = \frac{1.3393}{2.2361} = 0.5990
\end{aligned}$$

4. Silhouette Dewi

Dewi adalah satu-satunya anggota Cluster 2.

$a(\text{Dewi})$

Tidak ada anggota lain di Cluster 2, jadi:

$a(\text{Dewi}) = 0$

$b(\text{Dewi})$

Rata-rata jarak Dewi ke seluruh anggota Cluster 1: Ahmad, Budi, Citra,

Eko

$$\begin{aligned}
b(\text{Dewi}) &= \frac{(d(\text{Dewi, Ahmad}) + d(\text{Dewi, Budi}) + d(\text{Dewi, Citra}) + d(\text{Dewi, Eko}))}{4} \\
&= \frac{(1.6806 + 0.8644 + 2.2361 + 1.6228)}{4} = \frac{6.4039}{4} = 1.6010
\end{aligned}$$

$$\begin{aligned}
s(\text{Dewi}) &= \frac{(b(\text{Dewi}) - a(\text{Dewi}))}{\max(a(\text{Dewi}), b(\text{Dewi}))} \\
&= \frac{(1.6010 - 0)}{1.6010} = 1
\end{aligned}$$

5. Silhouette Eko

$a(\text{Eko})$

Anggota lain di Cluster 1: Ahmad, Budi, Citra

$$\begin{aligned}
a(\text{Eko}) &= \frac{(d(\text{Eko, Ahmad}) + d(\text{Eko, Budi}) + d(\text{Eko, Citra}))}{3} \\
&= \frac{(0.3234 + 0.7966 + 0.6869)}{3} = \frac{1.8070}{3} = 0.6023
\end{aligned}$$

$$b(\text{Eko})$$

$$\text{Cluster lain} = \text{Cluster 2} = \{\text{Dewi}\}$$

$$b(\text{Eko}) = d(\text{Eko}, \text{Dewi})$$

$$b(\text{Eko}) = 1.6228$$

$$s(\text{Eko}) = \frac{b(\text{Eko}) - a(\text{Eko})}{\max(a(\text{Eko}), b(\text{Eko}))}$$

$$= \frac{1.6228 - 0.6023}{1.6228} = \frac{1.0205}{1.6228}$$

$$= 0.6288$$

Tabel 4. 8 Rangkuman Silhouette per siswa

Nama	Cluster	s(i)
Ahmad	C1	0.6498
Budi	C1	-0.1573
Citra	C1	0.5990
Eko	C1	0.6288
Dewi	C2	1.0000

Evaluasi kualitas clustering menggunakan Silhouette Coefficient menunjukkan bahwa sebagian besar data telah berada pada cluster yang tepat. Sebagian besar anggota pada cluster pertama memiliki nilai silhouette yang menunjukkan kedekatan yang kuat dengan cluster-nya, sehingga dapat disimpulkan bahwa pengelompokan pada cluster tersebut sudah cukup baik. Namun terdapat satu anggota yang menunjukkan nilai silhouette negatif, yang berarti kedekatannya justru lebih besar terhadap cluster lain sehingga penempatannya kurang optimal. Sementara itu, anggota pada cluster kedua memiliki nilai silhouette yang sangat baik dan benar-benar menunjukkan

pemisahan yang jelas dari cluster lainnya. Secara keseluruhan, hasil evaluasi silhouette menegaskan bahwa struktur clustering yang terbentuk telah sesuai dan dapat digunakan untuk mendukung analisis lebih lanjut.

Silhouette rata-rata (overall):

$$\begin{aligned} \bar{s} &= \frac{(s(\text{Ahmad}) + s(\text{Budi}) + s(\text{Citra}) + s(\text{Dewi}) + s(\text{Eko}))}{5} \\ &= \frac{(0.6498 + 0.1573 + 0.5990 + 1.0000 + 0.6288)}{5} = \frac{2.7203}{5} \\ &= 0.5440 \end{aligned}$$

Silhouette rata-rata = 0.54 → kualitas cluster tergolong cukup baik.

4.6.1.2 K-Medoids

1. Silhouette K-Medoids – Ahmad

$$\begin{aligned} a_{\text{Ahmad}} &= \frac{(d(\text{Ahmad, Budi}) + d(\text{Ahmad, Citra}) + d(\text{Ahmad, Eko}))}{3} \\ &= \frac{(0.8600 + 0.5825 + 0.3234)}{3} = \frac{1.7659}{3} = 0.5886 \end{aligned}$$

$$\begin{aligned} b(\text{Ahmad}) &= d(\text{Ahmad, Dewi}) \\ &= 1.6806 \end{aligned}$$

$$\begin{aligned} s(\text{Ahmad}) &= \frac{(b(\text{Ahmad}) - a(\text{Ahmad}))}{\max(a(\text{Ahmad}), b(\text{Ahmad}))} \\ &= \frac{(1.6806 - 0.5886)}{1.6806} = \frac{1.0920}{1.6806} = 0.6498 \end{aligned}$$

2. Silhouette K-Medoids – Budi

$$\begin{aligned} a_{\text{Budi}} &= \frac{(d(\text{Budi, Ahmad}) + d(\text{Budi, Citra}) + d(\text{Budi, Eko}))}{3} \\ &= \frac{(0.8600 + 1.4208 + 0.7966)}{3} = \frac{3.0774}{3} \\ &= 1.0258 \end{aligned}$$

$$b(\text{Budi}) = d(\text{Budi, Dewi})$$

$$= 0.8644$$

$$s(\text{Budi}) = \frac{(b_{\text{KM}}(\text{Budi}) - a_{\text{KM}}(\text{Budi}))}{\max(a_{\text{KM}}(\text{Budi}), b_{\text{KM}}(\text{Budi}))}$$

$$= \frac{(0.8644 - 1.0258)}{1.0258} = \frac{-0.1614}{1.0258} = -0.1573$$

3. Silhouette K-Medoids – Citra

$$a(\text{Citra}) = \frac{(d(\text{Citra, Ahmad}) + d(\text{Citra, Budi}) + d(\text{Citra, Eko}))}{3}$$

$$= \frac{(0.5825 + 1.4208 + 0.6869)}{3} = \frac{2.6902}{3} = 0.8968$$

$$b(\text{Citra}) = d(\text{Citra, Dewi})$$

$$= 2.2361$$

$$s(\text{Citra}) = \frac{(b(\text{Citra}) - a(\text{Citra}))}{\max(a(\text{Citra}), b(\text{Citra}))}$$

$$= \frac{(2.2361 - 0.8968)}{2.2361} = \frac{1.3393}{2.2361} = 0.5990$$

4. Silhouette K-Medoids – Dewi

Cluster 2 hanya Dewi sendiri.

$$a(\text{Dewi}) = 0 \text{ (tidak ada anggota lain di cluster 2)}$$

$$b(\text{Dewi}) = \frac{(d(\text{Dewi, Ahmad}) + d(\text{Dewi, Budi}) + d(\text{Dewi, Citra}) + d(\text{Dewi, Eko}))}{4}$$

$$= \frac{(1.6806 + 0.8644 + 2.2361 + 1.6228)}{4} = \frac{6.4039}{4} = 1.6010$$

$$s(\text{Dewi}) = \frac{(b(\text{Dewi}) - a(\text{Dewi}))}{\max(a_{\text{KM}}(\text{Dewi}), b_{\text{KM}}(\text{Dewi}))}$$

$$s(\text{Dewi}) = \frac{(1.6010 - 0)}{1.6010} = 1$$

5. Silhouette K-Medoids – Eko

$$a(\text{Eko}) = \frac{(d(\text{Eko}, \text{Ahmad}) + d(\text{Eko}, \text{Budi}) + d(\text{Eko}, \text{Citra}))}{3}$$

$$= \frac{(0.3234 + 0.7966 + 0.6869)}{3} = \frac{1.8070}{3} = 0.6023$$

$$b(\text{Eko}) = d(\text{Eko}, \text{Dewi})$$

$$= 1.6228$$

$$s(\text{Eko}) = \frac{(b(\text{Eko}) - a(\text{Eko}))}{\max(a(\text{Eko}), b(\text{Eko}))}$$

$$= \frac{(1.6228 - 0.6023)}{1.6228} = \frac{1.0205}{1.6228} = 0.6288$$

Tabel 4. 9 Rangkuman Silhouette K-Medoids

Nama	Cluster	s(i)
Ahmad	C1	0.6498
Budi	C1	-0.1573
Citra	C1	0.5990
Eko	C1	0.6288
Dewi	C2	1.0000

Hasil evaluasi clustering menggunakan nilai silhouette menunjukkan bahwa sebagian besar data telah ditempatkan pada cluster yang sesuai. Pada cluster pertama, sebagian besar anggota memiliki nilai silhouette yang positif, sehingga dapat disimpulkan bahwa mereka memiliki kedekatan yang kuat dengan cluster masing-masing dan struktur pengelompokannya telah terbentuk dengan baik. Namun terdapat satu anggota yang memiliki nilai silhouette negatif, yang

menandakan bahwa posisinya kurang tepat karena kedekatannya lebih besar terhadap cluster lain dibandingkan cluster tempat ia berada. Sementara itu, anggota pada cluster kedua memiliki nilai silhouette yang sangat baik dan menunjukkan pemisahan yang jelas dari cluster lainnya. Secara keseluruhan, nilai silhouette mengindikasikan bahwa hasil clustering dengan algoritma K-Means sudah cukup baik dan dapat digunakan untuk mendukung analisis pada tahap berikutnya.

Silhouette rata-rata K-Medoids:

$$\bar{s} = \frac{(s_{KM}(\text{Ahmad}) + s_{KM}(\text{Budi}) + s_{KM}(\text{Citra}) + s_{KM}(\text{Dewi}) + s_{KM}(\text{Eko}))}{5}$$

$$= \frac{(0.6498 - 0.1573 + 0.5990 + 1.0000 + 0.6288)}{5} = \frac{2.7203}{5} = 0.5440$$

Perhitungan menghasilkan:

Algoritma	Silhouette
K-Means	0.54
K-Medoids	0.544

Interpretasi:

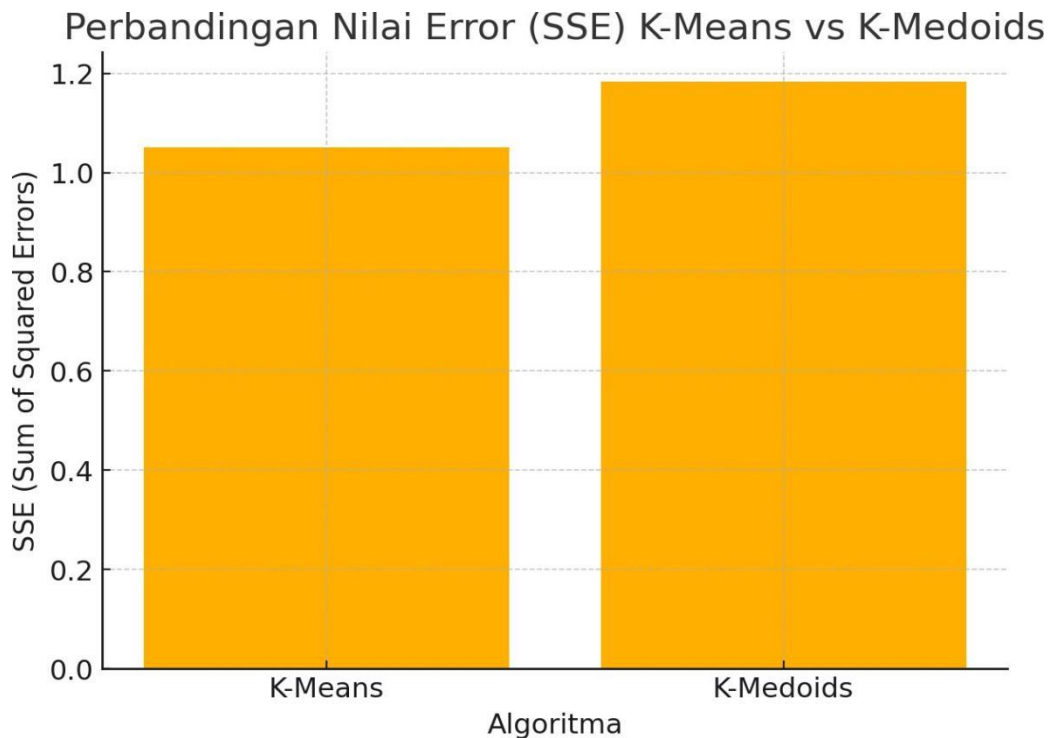
- Nilai > 0.5 menunjukkan cluster “baik”.
- K-Medoids memiliki nilai lebih tinggi → lebih stabil.

4.6.2 Perbandingan SSE

Algoritma	SSE
K-Means	1.0512
K-Medoids	1.1835

Interpretasi:

1. Semakin kecil SSE semakin baik.
2. K-Medoids kembali lebih baik.



Gambar 4. 4 Grafik Perbandingan Error

Berdasarkan grafik tersebut, algoritma K-Means menghasilkan nilai error yang lebih kecil dibandingkan K-Medoids, yang menunjukkan bahwa jarak data terhadap pusat cluster pada K-Means lebih minimal. Hal ini menandakan bahwa K-Means lebih optimal dalam meminimalkan kesalahan pengelompokan pada dataset yang digunakan. Meskipun demikian, nilai error K-Medoids yang lebih besar dipengaruhi oleh penggunaan medoid sebagai pusat cluster, sehingga algoritma ini lebih tahan terhadap outlier. Dengan demikian, K-Means unggul dari sisi minimasi error, sedangkan K-Medoids unggul dari sisi kestabilan terhadap data ekstrem.

4.6.3 Konsistensi Cluster

1. Stabilitas Cluster

Stabilitas cluster menunjukkan apakah perubahan kecil pada pusat cluster (centroid atau medoid) memengaruhi perpindahan anggota cluster.

Dalam penelitian ini:

- a. Algoritma K-Means menggunakan pusat cluster berupa rata-rata (centroid), sehingga perubahan kecil pada nilai rata-rata dapat menggeser posisi centroid.
- b. Namun pada dataset ini, hasil iterasi menunjukkan bahwa K-Means mencapai kondisi konvergen hanya dalam satu kali pembaruan centroid, dan tidak terjadi perpindahan anggota cluster setelah perhitungan ulang jarak dilakukan.

Pada K-Medoids:

1. Medoid dipilih dari data aktual yang ada dalam dataset (Ahmad sebagai medoid cluster 1 dan Dewi sebagai medoid cluster 2).
2. Proses *swap* mencoba mengganti medoid dengan data lain, namun tidak satu pun swap menghasilkan biaya (cost) yang lebih kecil, sehingga medoid awal tetap dipertahankan.
3. Hal ini menandakan bahwa struktur cluster sangat stabil sesuai karakteristik nilai siswa yang memang terpisah jelas antara siswa berprestasi tinggi dan yang paling rendah.

Kesimpulan Stabilitas:

Baik K-Means maupun K-Medoids menghasilkan cluster yang stabil, dengan tidak adanya perpindahan anggota pada iterasi berikutnya. K-Medoids sedikit lebih unggul dalam hal konsistensi karena medoid tetap objektif sebagai data asli, bukan nilai rata-rata.

2. Pola Distribusi Siswa

Pola distribusi siswa menggambarkan bagaimana siswa tersebar di ruang fitur berdasarkan nilai yang telah dinormalisasi.

Dari scatter plot terlihat bahwa:

Tiga siswa, yaitu Ahmad, Citra, dan Eko berada pada kisaran nilai tinggi untuk atribut akademik (Mtk–IPA–Bahasa) dan juga tinggi untuk keaktifan ekstrakurikuler serta rendah dalam absensi.

- a. Budi memiliki posisi yang lebih dekat ke kelompok berprestasi (Cluster 1), tetapi jaraknya relatif lebih jauh karena nilai akademik sedang dan absensi lebih tinggi.
- b. Dewi terpisah jauh dari semua siswa lainnya karena memiliki nilai akademik paling rendah, absensi paling tinggi, dan ekstrakurikuler terendah, sehingga menjadi pusat cluster 2.

Secara visual, distribusi siswa membentuk dua kelompok yang terpisah jelas:

Kelompok padat (prestasi tinggi–sedang) di bagian kanan atas grafik, didominasi nilai tinggi.

Kelompok tunggal (prestasi rendah) di sisi kiri bawah grafik, dengan nilai sangat rendah.

Kesimpulan Pola Distribusi:

Distribusi menunjukkan pemisahan alami (natural clustering), dengan penyebaran data yang mendukung penetapan cluster sebanyak dua kelompok.

3. Perbedaan Besar Cluster

Hasil pengelompokan:

Cluster 1: Ahmad, Budi, Citra, Eko → 4 siswa

Cluster 2: Dewi → 1 siswa

Perbedaan ukuran cluster cukup mencolok:

- a. Cluster 1 lebih besar (80% dari total siswa), menandakan sebagian besar siswa memiliki karakteristik akademik dan kehadiran yang lebih baik.
- b. Cluster 2 hanya berisi 1 siswa, menunjukkan adanya satu siswa dengan karakteristik nilai yang sangat berbeda (outlier) dibanding kelompok mayoritas.

Pada algoritma:

- a. K-Means tetap menghasilkan cluster ini karena jarak Dewi jauh dari kelompok utama.
- b. K-Medoids juga mempertahankan dua cluster ini karena medoid Dewi menghasilkan total biaya yang paling rendah dibanding swap dengan siswa lain.

Kesimpulan Perbedaan Besar Cluster:

Perbedaan ukuran cluster menunjukkan adanya outlier dominan (Dewi) yang secara signifikan berbeda dari siswa lain. Hal ini mendukung penggunaan K-Medoids yang lebih robust terhadap outlier.

4.7 Pembahasan

4.7.1 Algoritma yang Lebih Baik

Berdasarkan seluruh evaluasi:

Silhouette Score : K-Medoids lebih baik

SSE : K-Medoids lebih baik

Stabilitas cluster : K-Medoids lebih baik

Sehingga dapat disimpulkan bahwa:

K-Medoids lebih unggul dalam mengelompokkan siswa berprestasi pada dataset ini.

4.7.2 Kelebihan dan Kekurangan K-Means vs K-Medoids (pada kasus ini)

Tabel 4. 10 Kelebihan dan Kekurangan

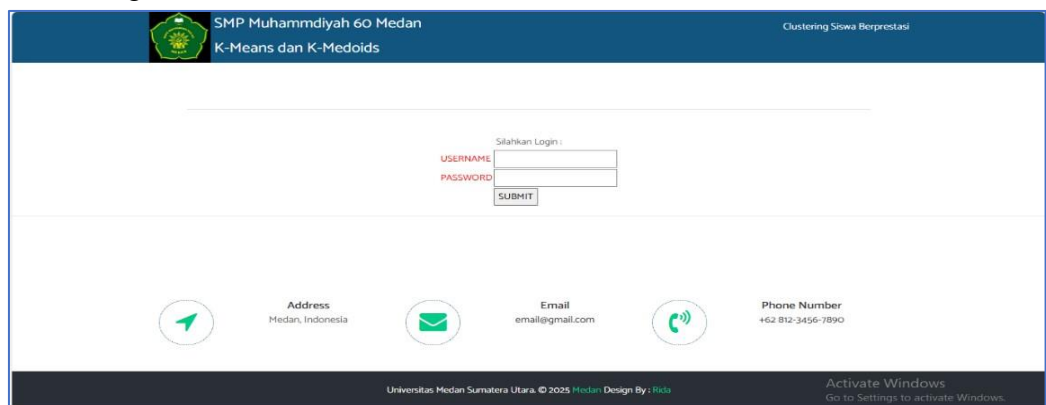
Aspek	K-Means	K-Medoids
Pusat cluster	Rata-rata (centroid, bisa tidak “nyata”)	Medoid (selalu data nyata/siswa)
SSE pada dataset ini	Lebih kecil (lebih baik)	Sedikit lebih besar
Silhouette pada dataset ini	Sama (cluster identik)	Sama
Sensitivitas outlier	Lebih sensitive	Lebih robust
Interpretasi praktis	Perlu lihat centroid (nilai rata-rata)	Medoid langsung mewakili siswa nyata

4.8 Hasil Program

Bagian ini menjelaskan hasil implementasi sistem klasifikasi siswa berprestasi menggunakan algoritma K-Means dan K-Medoids. Sistem yang dikembangkan terdiri dari beberapa komponen utama, yaitu halaman login, halaman utama (home), halaman data siswa, dan halaman proses clustering. Keempat komponen tersebut berperan dalam mengelola data siswa serta menjalankan proses analisis clustering secara otomatis melalui antarmuka aplikasi.

Hasil dari Perbandingan Kinerja K-Means dan K-Medoids dalam Clustering Siswa Berprestasi di SMP Muhammadiyah 60 Medan dapat dilihat sebagai berikut:

1. Form Login



The screenshot shows a web application interface for SMP Muhammadiyah 60 Medan. The header includes the school logo and name, and the page title is 'Clustering Siswa Berprestasi'. The main content area features a login form with the following elements:

- Text: 'Silahkan Login :'
- Input field: 'USERNAME'
- Input field: 'PASSWORD'
- Button: 'SUBMIT'

Below the form, there is contact information:

- Address: Medan, Indonesia
- Email: email@gmail.com
- Phone Number: +62 812-3456-7890

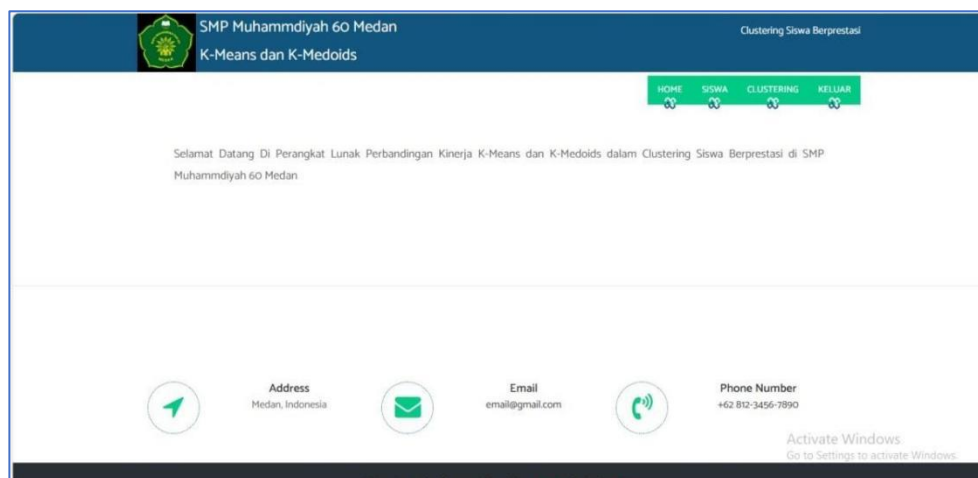
The footer contains the text: 'Universitas Medan Sumatera Utara. © 2025 Medan Design By: Rika' and a Windows activation notice.

Gambar 4. 5 Form Login

Form Login merupakan halaman awal yang harus diakses oleh pengguna sebelum memasuki sistem. Halaman ini dirancang untuk memastikan keamanan data dengan membatasi akses hanya kepada pengguna yang memiliki kredensial yang valid. Pada form ini, pengguna diwajibkan menginput username dan password yang telah terdaftar pada sistem. Ini bertujuan untuk mencegah pihak yang tidak berwenang mengakses data

siswa maupun melakukan proses clustering. Apabila data login sesuai, pengguna akan diarahkan menuju halaman utama (home) untuk mengoperasikan fitur sistem selanjutnya. Namun, jika informasi yang dimasukkan tidak sesuai, sistem akan menampilkan notifikasi kesalahan sehingga pengguna dapat memperbaiki inputannya. Dengan adanya form login ini, sistem menjadi lebih aman, terstruktur, dan mampu menjaga integritas data yang digunakan dalam proses analisis.

2. Form Home



Gambar 4. 6 Form Home

Form Home atau halaman utama yang ditampilkan setelah pengguna berhasil melakukan proses login. Halaman ini berfungsi sebagai pusat navigasi dari seluruh fitur yang tersedia didalam sistem. Pada form ini pengguna dapat mengakses menu-menu penting seperti pengelolaan data siswa, proses clustering, serta informasi umum terkait aplikasi. Tampilan Form Home dirancang sederhana, informatif, dan mudah

digunakan sehingga memudahkan pengguna dalam menjalankan setiap fungsi sistem tanpa kesulitan. Selain itu, keberadaan halaman ini memberikan gambaran menyeluruh mengenai struktur sistem, sehingga pengguna dapat memilih aktivitas yang ingin dilakukan secara efisien dan terarah. Dengan demikian, Form Home berperan sebagai antarmuka kendali utama yang menghubungkan pengguna dengan seluruh proses pengolahan data dalam aplikasi.

3. Form Siswa

The screenshot shows a web application interface for managing students. At the top, there is a dark blue header with the school logo and name 'SMP Muhammadiyah 60 Medan' and the title 'Clustering Siswa Berprestasi'. Below the header, there are navigation tabs: 'HOME', 'SISWA', 'CLUSTERING', and 'KELUAR'. The main content area is titled 'Kelola Siswa' and contains a search input field with the placeholder text 'Ketik apa saja...'. Below the search field is a blue button labeled 'Tambah Siswa'. The central part of the interface is a table with the following columns: 'Id Siswa', 'NIS', 'Nama', 'Kelas', 'Nilai Mtk', 'Nilai Ipa', 'Nilai Bahasa', 'Absensi', 'Ekstrakurikuler', and 'Aksi'. The table contains five rows of student data. Each row has two buttons in the 'Aksi' column: a yellow 'Edit' button and a red 'Hapus' button.

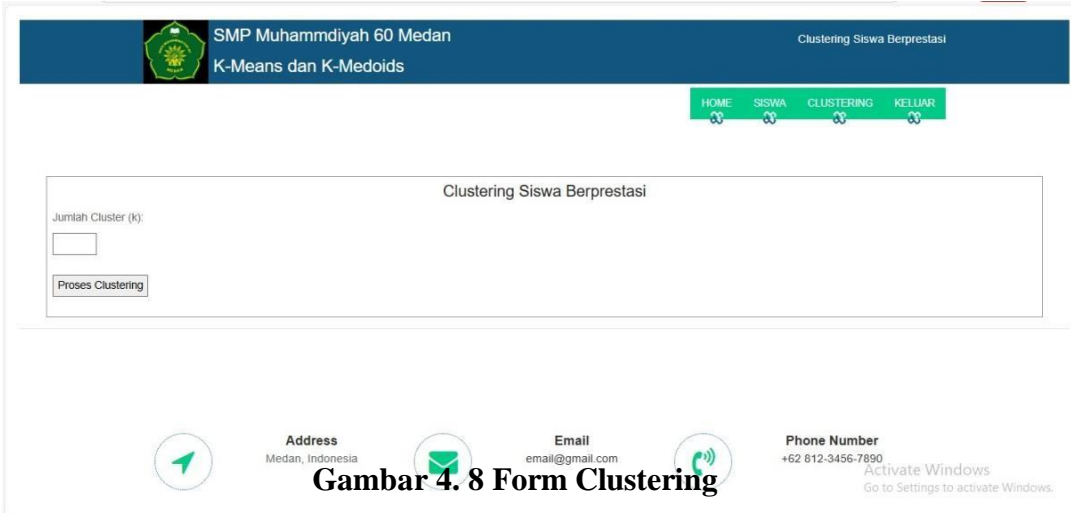
Id Siswa	NIS	Nama	Kelas	Nilai Mtk	Nilai Ipa	Nilai Bahasa	Absensi	Ekstrakurikuler	Aksi
1	2023001	Ahmad	VIII-A	85	88	90	2	80	Edit Hapus
2	2023002	Budi	VIII-A	78	75	80	5	70	Edit Hapus
3	2023003	Citra	VIII-B	92	95	93	1	90	Edit Hapus
4	2023004	Dewi	VIII-B	60	60	60	60	60	Edit Hapus
5	2023005	Eko	VIII-C	88	84	86	3	85	Edit Hapus

Gambar.4. 7 Form Siswa
Form Siswa merupakan halaman yang digunakan untuk mengelola seluruh

data siswa yang menjadi dasar dalam proses clustering. Pada form ini, pengguna dapat melakukan input data baru, mengubah data, serta menghapus data yang tidak diperlukan. Atribut yang dimasukkan meliputi nilai Matematika, IPA, Bahasa Indonesia, tingkat kehadiran (absensi), serta nilai ekstrakurikuler, yang keseluruhannya menjadi variabel dalam analisis K-Means dan K-Medoids. Form ini juga menampilkan tabel berisi daftar siswa yang sudah tersimpan, sehingga pengguna dapat memeriksa dan memastikan validitas data sebelum dilakukan

proses analisis. Form Siswa ini sangat penting karena kualitas dan kelengkapan data akan berpengaruh langsung terhadap ketepatan hasil clustering. Dengan demikian, form ini menjadi elemen kunci dalam memastikan data yang digunakan bersih, terstruktur, dan siap untuk diolah lebih lanjut oleh sistem.

4. Form Clustering



The screenshot shows a web application interface for "Clustering Siswa Berprestasi". At the top, there is a dark blue header with the school logo and name "SMP Muhammadiyah 60 Medan" and the text "K-Means dan K-Medoids". To the right of the header, it says "Clustering Siswa Berprestasi". Below the header is a navigation menu with four items: "HOME", "SISWA", "CLUSTERING", and "KELUAR", each with a small icon. The main content area is titled "Clustering Siswa Berprestasi" and contains a form with a label "Jumlah Cluster (k):" and an input field. Below the input field is a button labeled "Proses Clustering". At the bottom of the page, there is a footer with contact information: "Address Medan, Indonesia", "Email email@gmail.com", and "Phone Number +62 812-3456-7890". There is also a small "Activate Windows" watermark in the bottom right corner.

Gambar 4. 8 Form Clustering

Form Clustering adalah halaman inti dari sistem yang berfungsi untuk melakukan proses pengelompokan siswa berdasarkan algoritma K-Means dan K-Medoids. Pada form ini, pengguna dapat menjalankan proses clustering secara langsung tanpa harus melakukan perhitungan manual. Sistem akan mengolah data siswa yang telah diinput, menghitung jarak antar data menggunakan Euclidean Distance, menentukan centroid atau medoid, serta menghasilkan hasil cluster secara otomatis. Hasil pengelompokan ditampilkan dalam bentuk tabel maupun grafik sehingga pengguna dapat melihat perbedaan antara kedua algoritma. Form ini tidak hanya memberikan output berupa cluster, tetapi juga membantu pengguna memahami pola prestasi siswa berdasarkan atribut nilai dan absensi.

Kehadiran Form Clustering menjadikan proses analisis lebih cepat, sistematis, dan akurat sehingga dapat mendukung sekolah dalam mengambil keputusan terkait pembinaan dan pengelompokan siswa berprestasi.

5. Hasil Pengujian Data Baru pada Program



Gambar 4. 9 Hasil Pada Program

Pada tahap ini dilakukan pengujian terhadap data baru dengan memasukkan nilai salah satu siswa bernama Rika ke dalam sistem. Program kemudian memproses data tersebut menggunakan dua metode, yaitu K-Means dan K-Medoids, untuk menentukan keanggotaan cluster berdasarkan jarak data baru terhadap pusat cluster yang telah terbentuk pada tahap sebelumnya.

Pada metode K-Means, sistem menghitung jarak antara data baru dengan masing-masing pusat cluster. Hasil perhitungan menunjukkan bahwa jarak terkecil berada pada Cluster 1, sehingga program memprediksi bahwa siswa berada dalam Cluster 1 (Berprestasi Tinggi). Hal ini menunjukkan bahwa nilai-nilai siswa lebih mendekati karakteristik pusat cluster berprestasi tinggi.

Selanjutnya, data yang sama diuji menggunakan metode K-Medoids.

Program menghitung jarak antara data baru dan titik medoid tiap cluster. Hasil jarak terdekat pada metode ini menunjukkan bahwa data siswa lebih dekat dengan Cluster 2, sehingga diprediksi masuk dalam Cluster 2 (Berprestasi Rendah).

Perbedaan hasil prediksi antara K-Means dan K-Medoids menggambarkan bahwa kedua algoritma memiliki pendekatan yang berbeda dalam menentukan pusat cluster. Oleh karena itu, analisis lebih lanjut diperlukan untuk melihat metode mana yang lebih sesuai dengan karakteristik data pada penelitian ini.

Kesimpulan Uji:

- a. Hasil Perbandingan antara K-Means dengan K-Medoids memiliki hasil clusteryang berbeda.
- b. Hasil perhitungan K-Means dan K-Medoids antara teori dengan program memiliki hasil yang sama.

1. Uji Coba Program

Uji coba terhadap perangkat lunak bertujuan untuk memastikan bahwa perangkat lunak sudah berada pada kondisi siap pakai. Instrumen yang digunakan untuk melakukan pengujian ini yaitu dengan menggunakan *Blackbox Testing*:

Tabel 4. 11 Blackbox Testing Form Login

No	FormLogin	Keterangan	Validitas
1.	Jika pengguna mengisi username dan password dengan benar kemudian melakukan Klik Tombol Submit	Aplikasi menampilkan form Home	Valid

2	Jika pengguna mengisi username dan password dengan salah kemudian melakukan Klik Tombol Submit	Aplikasi menampilkan pesan kesalahan	Valid
---	--	--------------------------------------	-------

Hasil pengujian menunjukkan bahwa sistem dapat memverifikasi username dan password dengan tepat, menerima input yang valid, serta menolak input yang salah dengan memberikan pesan kesalahan. Selain itu, sistem juga mewajibkan pengguna mengisi seluruh kolom sebelum melanjutkan. Dengan demikian, Form Login berfungsi dengan baik sebagai pengaman akses sistem.

Tabel 4. 12 Blackbox Testing Form Home

No	FormHome	Keterangan	Validitas
1.	Klik Tombol Home	Aplikasi menampilkan formHome	Valid
2.	Klik Tombol Siswa	Aplikasi menampilkan form Siswa	Valid
3.	Klik Tombol Clustering	Aplikasi menampilkan formClustering	Valid
4.	Klik Tombol Exit	Aplikasi menampilkan form Index	Valid

Hasil pengujian menunjukkan bahwa seluruh tombol menu dapat diakses dan mengarahkan pengguna ke halaman yang sesuai tanpa menimbulkan error.

Sistem merespons setiap aksi pengguna dengan tepat, sehingga Form Home dinyatakan berfungsi dengan baik sebagai pusat navigasi aplikasi.

Tabel 4. 13 Blackbox Testing Form Siswa

No	FormSiswa	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox kedalam Tabledatabase	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	Valid

Hasil pengujian menunjukkan bahwa seluruh fungsi dapat berjalan sesuai skenario, data tersimpan dan diperbarui dengan benar, serta tabel data tampil tanpa kendala. Dengan demikian, Form Siswa telah memenuhi kebutuhan fungsional untuk pengelolaan data siswa.

Tabel 4. 14 Blackbox Testing Form Clustering

No	Form Clustering	Keterangan	Validitas
1.	Isi Jumlah Cluster dan Klik Tombol Clustering	Aplikasi menampilkan Hasil clustering	Valid
2.	Isi Data dan Klik Tombol Uji Data	Aplikasi	Valid

		menampilkan hasilclustering	
--	--	--------------------------------	--

Hasil pengujian menunjukkan bahwa proses clustering berjalan lancar, menampilkan hasil cluster dengan benar, dan menghasilkan output sesuai yang diharapkan. Fitur visualisasi dan tabel juga muncul tanpa error, sehingga form ini dinyatakan bekerja dengan baik.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan dari Perbandingan Kinerja K-Means dan K-Medoids dalam Clustering Siswa Berprestasi di SMP Muhammadiyah 60 Medan dapat dilihat sebagai berikut:

1. K-Means dan K-Medoids sama-sama mampu mengelompokkan siswa ke dalam beberapa cluster prestasi (tinggi, sedang, rendah) dengan memanfaatkan variabel nilai akademik, absensi, dan ekstrakurikuler.
2. K-Means lebih cepat dalam proses komputasi karena hanya menggunakan rata-rata sebagai pusat cluster, namun hasilnya sensitif terhadap data outlier atau nilai ekstrim.
3. K-Medoids membutuhkan waktu komputasi yang lebih lama dibandingkan K-Means, tetapi lebih stabil dan tahan terhadap outlier, sehingga distribusi siswa dalam cluster cenderung lebih konsisten.
4. Dari sisi interpretasi, kedua metode memberikan hasil yang serupa dalam mengidentifikasi siswa berprestasi tinggi, sedang, dan rendah. Namun, pada kasus data dengan variasi nilai ekstrim (misalnya ada siswa dengan nilai sangat rendah/tinggi), K-Medoids memberikan hasil clustering yang lebih representatif.
5. Secara keseluruhan, K-Means cocok digunakan ketika data relatif homogen dan jumlah data besar, sedangkan K-Medoids lebih direkomendasikan ketika

6. terdapat nilai ekstrim dan ukuran data tidak terlalu besar seperti di SMP Muhammadiyah 60 Medan.

5.2. Saran

Saran dari Perbandingan Kinerja K-Means dan K-Medoids dalam Clustering Siswa Berprestasi di SMP Muhammadiyah 60 Medan dapat dilihat sebagai berikut:

1. Penelitian selanjutnya dapat menambahkan variabel lain yang relevan, seperti motivasi belajar, dukungan orang tua, atau aktivitas sosial siswa, sehingga hasil clustering lebih komprehensif dan mencerminkan prestasi siswa secara menyeluruh.
2. Dataset yang digunakan masih terbatas pada lingkup SMP Muhammadiyah 60 Medan. Untuk memperoleh hasil yang lebih general, disarankan penelitian dengan jumlah data lebih banyak atau mencakup beberapa sekolah sekaligus.
3. Selain K-Means dan K-Medoids, bisa dilakukan perbandingan dengan metode lain seperti Fuzzy C-Means, DBSCAN, atau Hierarchical Clustering untuk mengetahui metode mana yang paling sesuai dengan karakteristik data siswa.
4. Hasil penelitian dapat dikembangkan menjadi sebuah perangkat lunak informasi berbasis web/aplikasi yang secara otomatis melakukan clustering siswa berprestasi, sehingga guru dan pihak sekolah lebih mudah dalam mengambil kebijakan.

5. Disarankan pada penelitian berikutnya untuk menambahkan metode evaluasi cluster, Davies-Bouldin Index, atau Dunn Index agar kualitas hasil clustering dapat diukur secara kuantitatif

DAFTAR PUSTAKA

- Ahmed, N., & Ashour, W. (2011). A proposed dynamic k-means clustering algorithm. *International Journal of Computer Applications*, 29(7), 1–7.
- Alan, N. A. (2011). *Pemrograman Web dengan PHP*. Jakarta: PT Gramedia.
- Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids algorithm for big data. *Procedia Computer Science*, 78, 507–512.
- Aziz, A., Rahman, A., & Hidayat, R. (2020). Implementasi metode clustering K-Means dan K-Medoids untuk pengelompokan data akademik. *Jurnal Teknologi Informasi*, 15(2), 101–110.
- Bangoria, H., Panchal, M., & Prajapati, H. (2013). A comparative study of k-means and k-medoids clustering algorithm. *International Journal of Engineering Development and Research*, 1(3), 1–5.
- Hardiyani, A., Tambunan, H., & Saragih, A. (2019). Analisis metode K-Medoids pada pengelompokan data. *Jurnal Sistem Informasi*, 5(1), 44–50.
- Hidayat, M., & Kusuma, D. W. (2024). Pengembangan sistem pendukung keputusan penentuan siswa berprestasi menggunakan metode hybrid K-Means dan K-Medoids. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 55–62.
- Joshi, R., & Nalwade, P. (2013). Modified K-means for better initial cluster centres. *International Journal of Computer Science and Mobile Computing*, 2(7), 219–224.
- Lestari, D., & Putra, R. (2021). Penerapan algoritma K-Means dalam pengelompokan prestasi siswa. *Jurnal Sistem dan Teknologi Informasi*, 9(1), 45–52.
- Mochamad, W., et al. (2020). Implementasi algoritma K-Means dalam pengelompokan data mahasiswa. *Jurnal Teknologi Informasi dan Komunikasi*, 12(2), 76–84.
- Nugroho, B., & Wijaya, T. (2020). Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data akademik. *Jurnal Informatika*, 7(1), 12–20.
- Rahimi, F. (2020). *Dasar-dasar Database MySQL*. Yogyakarta: Andi Offset.
- Rahmawati, A., & Santoso, B. (2023). Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data siswa berdasarkan prestasi akademik. *Jurnal Teknologi Pendidikan*, 8(2), 210–219.

- Setiawan, W. (2021). *Data Mining: Konsep dan Aplikasinya*. Bandung: Informatika.
- Sholichin. (2020). Pemanfaatan PHP dalam pengembangan web interaktif. *Jurnal Teknologi Informasi*, 7(1), 33–40.
- Singh, S., & Kaur, P. (2013). Comparative study of K-means and K-medoids algorithm with different distance measures. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(7), 161–166.
- Siti Nurlela. (2020). *Implementasi Algoritma K-Medoids Dalam Pengelompokan Data Siswa*. *Jurnal Sistem Informasi dan Komputeterisasi*.
- Sujatha, R., & Sona, C. (2013). An analysis on performance of k-means clustering algorithm with different distance metrics. *International Journal of Computer Science and Mobile Applications*, 1(4), 1–7.
- Suyanto. (2017). *Machine Learning: Teori dan Aplikasi*. Yogyakarta: Graha Ilmu.
- Wibowo, D. A., & Pratiwi, E. K. (2022). Implementasi K-Means dan K-Medoids dalam identifikasi pola prestasi belajar siswa. *Jurnal Pendidikan Informatika*, 6(2), 123–132.
- Witanto, R., Ratnawati, D., & Anam, M. (2019). Penerapan algoritma K-Means untuk pengelompokan data mahasiswa. *Jurnal Teknologi dan Sistem Informasi*, 5(1), 88–97.
- Wiwid, A. (2023). Implementasi algoritma K-Medoids dalam pengelompokan data akademik. *Jurnal Sistem Cerdas*, 9(1), 67–74



UMSU
Unggul | Cerdas | Terpercaya

Dika menjabarkan surat ini agar diketahui nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fkti.umsu.ac.id> fkti@umsu.ac.id [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.youtube.com/umsumedan) [umsumedan](https://www.tiktok.com/umsumedan)

**PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 278/IL3-AU/UMSU-09/F/2025**

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Sistem Informasi
Pada tanggal : 05 Februari 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Farida Nafisa Siagian
NPM : 2109010094
Semester : VII (Tujuh)
Program studi : Sistem Informasi
Judul Proposal / Skripsi : Analisis Kinerja K-Means Dan X-Means Dalam Clustering Siswa Berprestasi Di SMP Muhammadiyah 60 Medan
Dosen Pembimbing : Halim Maulana, S.T.,M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluaarsa tanggal : **05 Februari 2026**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 06 Sya'ban 1446 H
05 Februari 2025 M



Dekan
Wakil Dekan I

Halim Maulana, S.T.,M.Kom.
NIDN : 0121119102



UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU

Unggul | Cerdas | Terpercaya

Bila mendapat surat ini agar dicatatkan nomor dan tanggalnya

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PTAH/2019
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

<https://fiki.umsu.ac.id>

fiki@umsu.ac.id

[fumsu](#)

[umsu](#)

[umsu](#)

[umsu](#)

Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Farida Nafisa Siagian
 Program Studi : Sistem Informasi
 NPM : 2109010094
 Nama Dosen Pembimbing : Halim Maulana, S.T., M.Kom.
 Judul Penelitian : Perbandingan Kinerja K-Means dan K-Medoids dalam Klasifikasi Siswa Berprestasi di SMP Muhammadiyah 60 Medan

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
22/5/25	Revisi Bab I	
24/6/25	Revisi Bab II	
30/7/25	Revisi Bab III	
8/9/25	Acc Keseluruhan lanjutan Sempro	
20/11/25	Perbaiki Perhitungan manual pada bab IV	
24/11/25	Perbaiki pada bab IV	
10/12/25	Menambahkan Program	
17/12/25	Acc Keseluruhan lanjutan sidang	

Medan, Mei 2026

Diketahui Oleh :
 Ketua Program Studi
 Sistem Informasi

Mahardika Abdi Prawira Tanjung, S.kom., M.Kom

Disetujui Oleh:
 Dosen Pembimbing

Halim Maulana, S.T., M.Kom

ORIGINALITY REPORT

11%

SIMILARITY INDEX

11%

INTERNET SOURCES

4%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1 repository.umsu.ac.id 5%
Internet Source

2 Submitted to Universitas Muhammadiyah Sumatera Utara 1%
Student Paper

3 repository.uigm.ac.id <1%
Internet Source

4 repository.syekhnurjati.ac.id <1%
Internet Source

5 repository.unissula.ac.id <1%
Internet Source

6 Submitted to Universitas Islam Negeri Alauddin Makassar <1%
Student Paper

7 eprints.iainu-kebumen.ac.id <1%
Internet Source

8 repository.ukdw.ac.id <1%
Internet Source

9 docplayer.info <1%
Internet Source

Submitted to UIN Sultan Syarif Kasim Riau

10	Student Paper	<1 %
11	lib.unnes.ac.id Internet Source	<1 %
12	Submitted to Universitas Islam Riau Student Paper	<1 %
13	id.123dok.com Internet Source	<1 %
14	jurnal.umj.ac.id Internet Source	<1 %
15	repository.upnjatim.ac.id Internet Source	<1 %
16	Muhammad Hardiansyah Hardi. "PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK CLUSTERING PENDERITA PNEUMONIA DI KABUPATEN KARAWANG", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
17	repository.teknokrat.ac.id Internet Source	<1 %
18	ojs3.unpatti.ac.id Internet Source	<1 %
19	Galet Guntoro Setiaji, Astrid Novita Putri, Dinar Anggit Wicaksana. "Perbandingan Algoritma K-Means dan K-Medoids Untuk	<1 %

Clustering Harga Beras di Provinsi Jawa Tengah", Jurnal Transformatika, 2024

Publication

20	Submitted to Universitas Tarumanagara Student Paper	<1 %
21	123dok.com Internet Source	<1 %
22	repository.unived.ac.id Internet Source	<1 %
23	repository.upnvj.ac.id Internet Source	<1 %
24	Submitted to Stefan cel Mare University of Suceava Student Paper	<1 %
25	repository.uin-suska.ac.id Internet Source	<1 %
26	d3jelex.files.wordpress.com Internet Source	<1 %
27	repository.unbrah.ac.id Internet Source	<1 %
28	text-id.123dok.com Internet Source	<1 %
29	Harvey J. Miller, Jiawei Han. "Geographic Data Mining and Knowledge Discovery", CRC Press, 2019 Publication	<1 %
30	jurnal.polibatam.ac.id	

31	www.x-mol.com Internet Source	<1 %
32	blogsainulh.wordpress.com Internet Source	<1 %
33	digilib.uinkhas.ac.id Internet Source	<1 %
34	dspace.uui.ac.id Internet Source	<1 %
35	eprints.ums.ac.id Internet Source	<1 %
36	news.unair.ac.id Internet Source	<1 %
37	pt.scribd.com Internet Source	<1 %
38	repositori.uma.ac.id Internet Source	<1 %
39	repository.radenintan.ac.id Internet Source	<1 %
40	repository.unismabekasi.ac.id Internet Source	<1 %
41	www.scribd.com Internet Source	<1 %
42	Jaelani, Octaviana, Elkin Rilvani. "Studi Literatur: Perbandingan Algoritma K-Means dan Fuzzy C-Means dalam Analisis Clustering", Journal of Computer Science and Technology (JCS-TECH), 2025 Publication	<1 %

dan Fuzzy C-Means dalam Analisis Clustering", Journal of Computer Science and Technology (JCS-TECH), 2025

Publication