ALGORITMA K-MEANS UNTUK MENINGKATKAN MODEL KLASTERISASI DATA SISWA SMK SAMUDRA NUSANTARA KABUPATEN CIREBON BERDASARKAN NILAI AKADEMIK

Bimo Alif Prayudha, Rudi Kurniawan, Yudhistira Wijaya, Umi Hayati
Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat, Indonesia
kiki009123@gmail.com

ABSTRAK

Pengelolaan data nilai akademik siswa yang besar dan kompleks menjadi tantangan signifikan dalam dunia pendidikan, khususnya di SMK yang fokus mempersiapkan siswa untuk dunia kerja. Data yang tidak terorganisir dengan baik sering kali menghambat proses pengambilan keputusan berbasis data. Algoritma K-Means dipilih dalam penelitian ini karena kemampuannya yang efektif dalam menganalisis data dan mengelompokkan pola tersembunyi. Penelitian ini menerapkan metodologi Knowledge Discovery in Databases (KDD), meliputi seleksi data, praproses, transformasi, klasterisasi, evaluasi menggunakan Davies-Bouldin Index (DBI), dan interpretasi hasil. Dataset terdiri dari nilai akademik siswa pada mata pelajaran utama, seperti Matematika, Bahasa Inggris, dan Kimia. Hasil penelitian menunjukkan klaster ideal terdiri dari dua kelompok dengan nilai DBI 0.519, di mana atribut Kimia memiliki pengaruh paling signifikan. Klasterisasi ini memberikan wawasan yang mendalam tentang pola akademik siswa, mendukung strategi pembelajaran berbasis data, serta membantu sekolah menyusun kebijakan pendidikan yang lebih efektif.

Kata kunci : *Algoritma K-Means, klasterisasi, Davies-Bouldin Index, nilai akademik siswa, Knowledge Discovery in Databases (KDD).*

1. PENDAHULUAN

Teknologi informasi dan komunikasi (TIK) telah membawa perubahan signifikan dalam cara pengelolaan data di berbagai sektor, termasuk pendidikan. Dalam konteks pendidikan di SMK, pengelolaan data nilai akademik siswa menjadi tantangan utama. Dengan tingginya jumlah siswa dan data akademik yang terus bertambah, analisis manual menjadi kurang efektif dan cenderung menghasilkan keputusan yang subyektif. Oleh karena itu, diperlukan pendekatan berbasis teknologi, seperti data mining, untuk membantu mengelompokkan data akademik berdasarkan pola tertentu.

Salah satu metode paling umum untuk data mining adalah clustering, yaitu teknik untuk mengelompokkan data berdasarkan hal-hal yang sebanding. Algoritma K-Means adalah salah satu algoritma clustering yang paling umum dan banyak digunakan dalam industri seperti analisis data, transaksi penjualan [4], pengelompokan pengiriman paket [2], serta analisis data medis [3].Metode ini dapat digunakan untuk meningkatkan pemahaman tentang pola-pola dalam data siswa, khususnya dalam pengelompokkan berdasarkan nilai akademik mereka.

Diharapkan dapat menghasilkan model klasterisasi yang lebih akurat dengan menggunakan algoritma K-Means, yang pada gilirannya dapat membantu pihak sekolah dalam merencanakan kebijakan pendidikan yang lebih baik. Sebagai contoh, penerapan clustering pada data nilai akademik siswa SMK Samudra Nusantara di Kabupaten Cirebon dapat mengidentifikasi kelompok siswa dengan performa yang serupa, memberikan wawasan tentang efektivitas pengajaran, serta merancang intervensi pendidikan yang lebih tepat sasaran.

Pengelolaan dan analisis data akademik siswa merupakan salah satu tantangan utama dalam pendidikan, terutama di SMK, yang memiliki tujuan untuk menghasilkan profesional yang kompeten dan bersiap menghadapi dunia kerja. Data akademik siswa yang terkumpul selama proses pembelajaran menyimpan informasi yang sangat berharga, namun pengolahan data dalam jumlah besar dan kompleks sering kali menjadi kendala.

Dalam konteks ini, algoritma K-Means yang telah terbukti efektif di berbagai bidang dapat menjadi solusi yang relevan untuk mengelompokkan siswa berdasarkan pola nilai akademik mereka. Meskipun algoritma ini sudah banyak digunakan dalam berbagai aplikasi seperti pengelompokan barang berdasarkan penjualan [4], penentuan eligibilitas bantuan sosial, serta pengelompokan menu makanan dan minuman [5], penerapannya dalam konteks pendidikan, khususnya pengelompokan data akademik siswa, masih minim.

Tantangan utama untuk penggunaan algoritma K-Means adalah menghitung jumlah cluster yang terbaik dan identifikasi atribut yang paling berpengaruh dalam membentuk klaster. Hal ini menjadi semakin penting karena pemilihan jumlah cluster yang tidak tepat atau atribut yang kurang relevan dapat menghasilkan model yang kurang akurat dan tidak dapat diandalkan.

2. TINJAUAN PUSTAKA

Penelitian ini fokus pada penerapan algoritma K-Means untuk meningkatkan model klasterisasi data siswa SMK Samudra Nusantara Kabupaten Cirebon berdasarkan nilai akademik. K-Means adalah metode data mining untuk mengelompokkan data berdasarkan kesamaan karakteristik atau nilai tertentu, yang telah banyak diterapkan di berbagai bidang. Berikut adalah beberapa penelitian terkait.

2.1. Algoritma K-Means dalam Data Mining

Davies-Bouldin Index (DBI) digunakan untuk mengukur kualitas klaster hasil algoritma K-Means. DBI yang rendah menunjukkan klaster yang lebih baik, dengan klaster yang terpisah jelas dan anggota klaster yang lebih homogen. Dalam penelitian ini, DBI digunakan untuk menilai keefektifan pengelompokan siswa berdasarkan nilai akademik, di mana nilai DBI yang rendah menunjukkan bahwa klaster siswa memiliki karakteristik yang serupa dan terpisah dengan baik. Algoritma K-Means adalah metode clustering yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. K-Means sering digunakan dalam berbagai bidang, termasuk analisis data pendidikan, karena kemampuannya dalam mengelompokkan data yang besar dan kompleks [6].

2.2. Penerapan K-Means untuk Klasterisasi Siswa

Penelitian menunjukkan penerapan algoritma K-Means dalam menentukan jurusan siswa SMA, dengan tujuan untuk mengelompokkan siswa berdasarkan minat dan kemampuan mereka [7]. Pendekatan ini relevan dalam mengelompokkan siswa berdasarkan nilai akademik mereka.

2.3. Penerapan K-Means dalam Bidang Pendidikan

Algoritma K-Means digunakan untuk mengelompokkan data rekam medis pasien, namun metode ini juga banyak digunakan dalam pendidikan untuk mengelompokkan siswa berdasarkan nilai, kemampuan, dan kinerja akademik [3].

2.4. Penggunaan K-Means untuk Penerima Beasiswa

K-Means diimplementasikan untuk mengelompokkan data penerima beasiswa Bank Indonesia (BI). Hal ini relevan dengan penelitian ini karena mengelompokkan siswa berdasarkan nilai akademik juga dapat digunakan untuk mendeteksi pola penerimaan beasiswa atau bantuan pendidikan lainnya [6].

2.5. Normalisasi Data dalam K-Means

Normalisasi data disarankan untuk meningkatkan efisiensi algoritma K-Means. Dalam penelitian ini, normalisasi nilai akademik siswa penting untuk memastikan hasil klasterisasi yang akurat dan adil [8].

2.6. Pengelompokan Berdasarkan Kinerja Akademik

K-Means diterapkan untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiskinan. Meskipun berbicara tentang variabel sosial-ekonomi, penerapan K-Means pada nilai akademik siswa dapat mengungkap pola-pola yang sebelumnya tidak terlihat [9].

2.7. Penerapan K-Means pada Pengelompokan Kasus Kesehatan

K-Means digunakan untuk mengelompokkan tingkat risiko penyakit jantung. Penggunaan K-Means untuk pengelompokan tingkat risiko, baik dalam konteks kesehatan maupun pendidikan, menunjukkan fleksibilitas metode ini dalam berbagai disiplin ilmu [10].

2.8. Implementasi K-Means dalam Pengelompokan Siswa Berdasarkan Prestasi

K-Means diterapkan dalam menentukan karyawan tetap berdasarkan kinerja. Metode ini dapat diterapkan pada siswa dengan cara mengelompokkan mereka berdasarkan prestasi akademik untuk tujuan evaluasi dan pengembangan [11].

2.9. Pengelompokan Berdasarkan Tingkat Kemiskinan dan Pendidikan

K-Means digunakan untuk mengelompokkan provinsi berdasarkan indikator pendidikan. Penerapan metode serupa pada data siswa akan membantu dalam memetakan prestasi akademik dan membantu dalam pemberian beasiswa atau program pengembangan lainnya [12].

2.10. K-Means untuk Pengelompokan Berdasarkan Data Pendidikan dan Ekonomi

K-Means digunakan untuk mengelompokkan data kepadatan penduduk. Penggunaan data yang mencakup variabel pendidikan dan ekonomi dapat memberikan wawasan lebih luas tentang pengelompokan siswa berdasarkan kinerja akademik dan potensi masa depan mereka [13].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk menguji efektivitas algoritma K-Means dalam mengklasifikasikan nilai akademik siswa SMK Samudra Nusantara, Kabupaten Cirebon. Data yang digunakan mencakup nilai mata pelajaran utama. Metodologi yang diterapkan mengikuti tahapan KDD, yaitu pemilihan data, prapemrosesan, transformasi, penambangan data dengan K-Means, evaluasi menggunakan indeks Davies-Bouldin (DBI), dan interpretasi hasil. Penelitian ini bertujuan menentukan jumlah klaster optimal dan mengidentifikasi atribut nilai yang berpengaruh. Hasilnya diharapkan membantu sekolah merancang strategi pembelajaran yang lebih efektif.

Alat bantu seperti Google Colab dan RapidMiner digunakan untuk mengolah data.

3.1. Sumber Data

Data yang digunakan dalam penelitian ini dikumpulkan langsung dari SMK Samudra Nusantara yang terletak di Kabupaten Cirebon. Dataset yang diperoleh mencakup nilai akademik siswa dari mata pelajaran utama, seperti Matematika, Bahasa Indonesia, dan Kimia, serta informasi lainnya yang relevan, termasuk nomor urut (No), NIS, dan Nama siswa. Proses pengambilan data dilakukan dengan mengajukan permohonan resmi kepada bagian Ke-Siswaan di SMK Samudra Nusantara, dengan prosedur yang telah ditetapkan untuk memastikan bahwa data yang digunakan dalam penelitian ini memenuhi standar kualitas dan kelengkapan yang diperlukan. Dataset ini akan digunakan untuk melakukan analisis menggunakan algoritma K-Means, yang bertujuan untuk mengelompokkan data berdasarkan pola yang ada pada nilai akademik siswa

3.2. Teknik Pengumpulan Data

Dalam penelitian ini, data dikumpulkan melalui metode observasi dokumentasi dengan mengajukan permohonan langsung kepada pihak Ke-Siswaan SMK Samudra Nusantara Kabupaten Cirebon. Permohonan tersebut meminta akses data nilai akademik siswa dari mata pelajaran utama, seperti No, Nis, Nama, Matematika, Bahasa Indonesia, dan Kimia. Pihak Ke-Siswaan kemudian menyediakan data dalam format yang sesuai. Teknik observasi dokumentasi ini memastikan pengumpulan informasi yang akurat dan relevan, yang selanjutnya dianalisis menggunakan algoritma K-Means untuk klasterisasi.

3.3. Teknik Analisis Data

Penelitian ini menggunakan teknik KDD (Knowledge Discovery in Databases) untuk menganalisis data dengan tahapan berikut: 1) Pemilihan Data: Data nilai akademik siswa SMK Samudra Nusantara Kabupaten Cirebon diidentifikasi dan dipilih dari dataset yang relevan. Data mencakup nilai dari beberapa mata pelajaran utama. 2) Praproses Data melibatkan Pembersihan data dan normalisasi data: yaitu dengan menghapus data duplikasi dan mengisi nilai yang kosong. Serta mengonversi atribut ke dalam skala yang seragam untuk memastikan keakuratan analisis. 3) Transformasi Data: Atribut-atribut dataset diubah ke dalam format yang sesuai untuk algoritma K-Means, termasuk konversi data kategorikal menjadi numerik menggunakan operator Nominal to Numeric. 4) Proses Clustering: Data diolah menggunakan algoritma K-Means untuk membagi siswa ke dalam klaster berdasarkan pola nilai akademik. Penentuan jumlah klaster optimal dilakukan menggunakan Davies-Bouldin Index (DBI). 5) Evaluasi Model: Kualitas klaster dievaluasi dengan indeks DBI, yang menilai jarak antar klaster dan konsistensi dalam satu klaster. Atribut yang paling berpengaruh diidentifikasi dengan analisis sensitivitas melalui penghapusan atribut satu per satu. 6) Interpretasi Hasil: Hasil klasterisasi diinterpretasikan untuk memahami karakteristik setiap klaster, mendukung pengambilan keputusan berbasis data, dan membantu strategi pembelajaran.

4. HASIL DAN PEMBAHASAN

Penelitian ini menjelaskan penerapan algoritma K-Means untuk mengelompokkan nilai akademik siswa di SMK Samudra Nusantara Kabupaten Cirebon, dengan tujuan membantu merancang strategi pembelajaran yang lebih efektif.

4.1. Data Undestanding

Proses ini meliputi pemilihan data siswa, praproses data, dan analisis pengelompokan berdasarkan tingkat prestasi nilai mata pelajaran. Dataset yang digunakan mencakup data siswa tahun ajaran 2023-2024, termasuk atribut seperti Nomor Induk Siswa (NIS), Nama Siswa, dan nilai mata pelajaran. Hasil pengelompokan ini akan membantu pihak sekolah dalam merancang strategi yang lebih sesuai untuk setiap kelompok.

4.2. Data Selection

Selection adalah tahap awal untuk mengidentifikasi dan mengumpulkan data relevan guna mencapai tujuan penelitian. Data yang dipilih berasal dari "DatasetSMKBaru.xlsx", berisi 396 entri dengan 12 atribut terkait nilai akademik siswa SMK Samudra Nusantara Kabupaten Cirebon pada periode 2023-2024, mencakup No, Nis, Nama, Kelas, dan Mata Pelajaran. Gambar 1 menunjukkan operator Read Excel.

Gambar 1. Operator read excel

Gambar 1 menunjukkan tugas operator Read Excel untuk membaca dan mengimpor data nilai siswa dari file eksternal ke dalam lingkungan pemrosesan data, mempersiapkannya untuk tahap pra-proses. Berikut adalah Tabel 4.1 Parameter Operator Read Excel.

Tabel 1. Parameter operator read excel

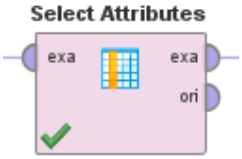
No	Parameter	Isi
1	Sheet Selection	Sheet Number
2	Sheet Number	1
3	Import Cell	A1
4	Encoding	SYSTEM
5	Header Row	1
6	Data Format	
7	Time Zone	SYSTEM

Berikut Tabel 1 Merupakan isi parameter operator read excel pengaturan untuk memprosesan import data. Berikut Tabel 2 hasil operator read excel.

Tabel 2. Hasil Read Excel (Sumber Data)

No	NIS	Nama	Kelas	P.Agama	…	Informatika
1	2122.1.001	Aan Suparman	Xii Titl 1	82	…	83
2	2122.1.002	Abil Putra Anggara	Xii Titl 1	8	…	83
3	2122.1.003	Adi Muhamad Fazri	Xii Titl 1	83	…	87
…	…	…	…	…	…	…
396	2122.1.446	Yoga Budi Utomo	Xii Tkr 2	77	…	60

Tahapan Selanjutnya Menggunakan Operator Select Atributtes Untuk Memilih Atribut Yang akan di gunakakan agar lebih relevan dan signifikan terhadap tujuan pengelompokan Nilai Akademik Siswa Berikut Gambar 2 Operator Select Attributes.



Gambar 2. Operator Select Attributes

Gambar 2 Menunjukan Operator memilih atribut, operator ini memilih beberapa atribut yang spesifik dari Kumpulan data yang di gunakan untuk pengelompokan. Berikut Tabel 3.

Tabel 3. Parameter Operator Select Attributes

No	Parameter	Isi
1	Type	Include Attributes
2	Attribute Filter Type	A Subset
…	…	…
	Attributes	Selected Attributes
	Kelas	Bahasa Indonesia
	Nis	Bahasa Inggris
		Informatika
		Kimia
		Matematika
		Nama
		No
		Pendidikan Agama Islam
		Pendidikan Jasmani Olahraga
		Pendidikan Pancasila

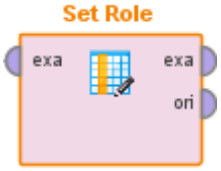
Tabel 3 menghasilkan Pemilihan atribut relevansi dan kontribusinya terhadap analisis klasterisasi Data Nilai Akademik Siswa SMK Samudra Nusantara Kabupaten Cirebon. Attribute “Nis” dan “Kelas” Tidak di gunakan karena menuju Nilai yang sama Dengan “Nama” menunjukan identitas siswa.

Tabel 4. Hasil Operator Select Attributes.

No	Nama	P.Agama	P.Pancasila	…	Informatika
1	Aan suparman	82	79	…	83
2	Abil putra anggara	83	87	…	83
3	Adi muhamad fazri	83	84	…	87
…	…	…	…	…	…
396	Yoga budi utomo	77	80	…	80

4.3. Preprocessing

Preprocessing data bertujuan menyiapkan dataset sebelum digunakan dalam model pengelompokan. Setelah penentuan peran (Set Role) untuk setiap atribut, data menjadi lebih terstruktur dan siap untuk proses clustering. Implementasi tahap ini dapat dilihat pada Gambar 3 Operator Set Role.



Gambar 3. Operator Set Role

Gambar 3. Menunjukakan Operator Set Role bertujuan menyederhanakan analisis dan interpretasi hasil klasterisasi. Berikut Tabel 5 Menunjukan Parameter Operator Set Role.

Tabel 5. Parameter Set Role

No	Uraian	Keterangan
1	Attributes Name	No.
2	Target Role	Id

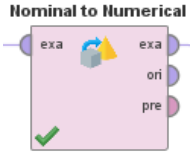
Tabel 5 menunjukkan parameter Operator Set Role, di mana "NO" ditetapkan sebagai ID. Proses ini menyederhanakan analisis dan interpretasi hasil klasterisasi, dengan setiap cluster dapat diidentifikasi berdasarkan karakteristik yang ada. Berikut adalah Tabel 6 yang menampilkan isi parameter Set Role.

Tabel 6. Hasil operator set role

No	Nama	P.Agama	…	Informatika
1	Aan suparman	82	…	83
2	Abil putra anggara	83	…	83
3	Adi muhamad fazri	83	…	87
…	…	…	…	…
396	Yoga budi utomo	77	…	80

4.4. Transformation

Transformasi data dilakukan untuk mengonversi atribut nominal menjadi numerik menggunakan operator Nominal To Numeric. Konversi ini bertujuan agar data kategori dapat di analisis secara optimal. Gambar 4. menunjukan proses transformasi data.



Gambar 4. Operator nominal to numerical

Gambar 4 menunjukan penggunaan operator Nominal To Numeric dengan pengaturan parameter tertentu untuk mengkonversi atribut nominal (NAMA) menjadi bentuk numerik. Tabel 7 ini menunjukan parameter dari operator Nominal To Numerical.

Tabel 7. Parameter operator nominal to numerical

No	Uraian	Keterangan
1	Attribute Filter Type	Single
2	Attribute	NAMA
3	Coding Type	Unique Integres

Tabel 7 menunjukkan konfigurasi Nominal To Numerical untuk atribut (NAMA), mengonversi nilai kategori menjadi angka integer berbeda. Transformasi ini memungkinkan data digunakan dalam pengelompokan, seperti yang ditunjukkan pada Tabel 8.

Tabel 8. Hasil Operator Nominal To Numerical

No	Nama	P. Agama	…	Informatika
1	0	82	…	83
2	1	83	…	83
3	2	83	…	87
…	…	…	...	…
396	395	77	…	80

Tabel 8 menunjukkan hasil transformasi kolom (NAMA) menjadi numerik setelah konversi dengan operator Nominal To Numerical. Nilai kategori diubah menjadi angka integer berbeda untuk setiap kategori unik, memungkinkan data teks digunakan dalam model pembelajaran mesin seperti clustering.

4.5. Data Mining

Proses data mining dengan K-Means mengelompokkan data berdasarkan kesamaan fitur untuk mengidentifikasi pola tersembunyi, membedakan karakteristik antar cluster.

Gambar 5. Operator Clustering K-Means

Gambar 5 menunjukkan Operator Clustering K-Means yang digunakan untuk memulai proses pengelompokan data. Parameter konfigurasi yang digunakan untuk operator ini ditampilkan pada Tabel 9.

Tabel 9. Isi Parameter Operator Clustering K-Means

No	Uraian	Keterangan
1	Nilai K	2-16
2	Max Runs	10
3	Measure Types	Bregman Divergences
4	Divergence	Squared Euclidian Distance
5	Max Optimize Steps	100

Operator K-Means mengelompokkan data berdasarkan kesamaan fitur, dengan parameter seperti jumlah cluster (K=2 hingga 16), maksimum pengulangan (10), jarak Bregman Divergences menggunakan Squared Euclidean Distance, dan langkah optimalisasi maksimum (100). Tabel 10 Hasil Cluster_0 dengan K=2 mengungkap pola data yang mendukung analisis lebih lanjut.

Tabel 10 Hasil cluster_0 K=2

No	cluster	NAMA	P. Agama	…	Informatika
198	cluster_0	197	79	…	81
199	cluster_0	198	80	…	83
200	cluster_0	199	81	…	85
201	cluster_0	200	76	…	78
…	…	…	…	…	…
396	cluster_0	395	77	…	80

Hasil clustering menunjukkan bahwa siswa nomor 198 hingga 396 tergabung dalam Cluster 0 dengan nilai rata-rata antara 70 hingga 82. Pola ini dapat digunakan untuk menganalisis metode pembelajaran dan merancang strategi pendidikan yang lebih efektif.

Tabel 11. Hasil cluster_1 K=2

No	cluster	NAMA	P. Agama	…	Informatika
1	cluster_1	0	82	…	83
2	cluster_1	1	83	…	83
3	cluster_1	2	83	…	87
4	cluster_1	3	80	…	78
…	…	…	…	…	…
197	cluster_1	196	80	…	83

Operator K-Means menghasilkan Cluster 0 dan Cluster 1 dengan K=2. Cluster 1 mencakup siswa nomor 1 hingga 197 dengan rata-rata nilai antara 78 hingga 90, dikelompokkan berdasarkan pola nilai tertentu untuk mendukung strategi pembelajaran.

Tabel 12. Hasil Operator Clustering K=2

Attribute	Cluster_0	Cluster_1
NAMA	199	197
Pendidikan Agama Islam dan Budi Pekerti	80.191	82.360
Pendidikan Pancasila	82.844	84.457
Bahasa Indonesia	81.492	81.726
MATEMATIKA	79.824	87.533
Bahasa Inggris	79.171	80.985
Pendidikan Jasmani, Olahraga dan Kesehatan	82.618	85.914
Kimia	80.201	80.437
Informatika	84.528	85.574

Dalam Cluster_0, siswa memiliki minat tertinggi pada Informatika (84.528) dan terendah pada Bahasa Inggris (79.171). Sementara itu, di Cluster_1, minat tertinggi ada pada MATEMATIKA (87.533) dan terendah pada Kimia (80.437). Cluster_0 cenderung pada minat teknologi, sedangkan Cluster_1 pada minat numerik.

Gambar 6. Operator Performance

Operator ini dalam K-Means mengevaluasi kualitas cluster dengan mengukur jarak antar data dan centroid. Semakin kecil jarak, semakin baik kualitas cluster. Operator ini juga membantu menentukan jumlah cluster optimal. Parameter Operator Performance dapat dilihat pada Tabel 13.

Tabel 13. Parameter Operator Performance

No	Uraian	Keterangan
1	Main Criterion	Davies Bouldin
2		Maximize

Tabel 13 menjelaskan bahwa parameter Operator Performance menggunakan Davies Bouldin sebagai kriteria utama untuk mengevaluasi kualitas cluster. Parameter Maximize digunakan untuk menentukan apakah suatu nilai perlu dimaksimalkan. Pada Tabel 14 Menunjukan hasil dari K=2 pada operator performance.

Tabel 14. Hasil Operator Performance K=2

No	Uraian	Keterangan
1	Avg. Within Centroid Distance	3386.404
2	Avg. Within Centroid Distance Cluster_0	3409.810
3	Avg. Within Centroid Distance Cluster_1	3362.720
4	Davies Bouldin	0.519

Tabel 14 menunjukan hasil dari K=2 yang dimana mendapatkan nilai DBI 0.519. Pada Tabel 14 menunjukkan nilai DBI terbaik adalah 0.519 pada K=2 sampai K=16, seperti yang ditampilkan pada Tabel 15.

Tabel 15. Hasil Nilai DBI K 2 sampai K 16

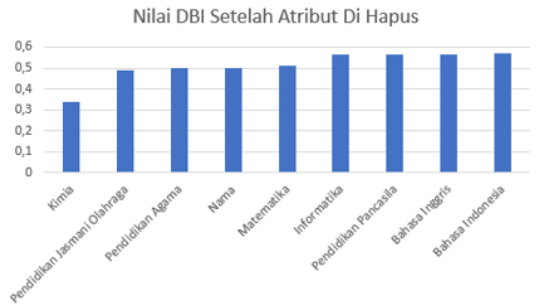
K	Davies Bouldin Index
2	0.519
3	0.539
4	0.558
5	0.580
6	0.609
7	0.638
8	0.672
9	0.715
10	0.750
11	0.725
12	0.771
13	0.778
14	0.831
15	0.868
16	0.906

Tabel 15 menunjukkan bahwa nilai DBI terbaik diperoleh saat K=2, dengan nilai sebesar 0.519. Hasil ini menunjukkan bahwa pemilihan K=2 memberikan klasterisasi yang paling efektif berdasarkan kriteria DBI, yang merupakan ukuran kualitas cluster.

Tabel 16. Nilai Davies Bouldin Index Setelah Menghapus Atribute

No	Atribut	DBI K=2
1	Kimia	0.337
2	Pendidikan Jasmani Olahraga	0.488
3	Pendidikan Agama	0.500
4	Nama	0.500
5	Matematika	0.511
6	Informatika	0.563
7	Pendidikan Pancasila	0.566
8	Bahasa Inggris	0.567
9	Bahasa Indonesia	0.571

Tabel 16 menunjukkan nilai DBI setelah penghapusan atribut. Setelah menentukan K optimal, analisis dilakukan untuk mengidentifikasi atribut paling berpengaruh dengan menghitung DBI setelah menghapus atribut satu per satu. Atribut yang menyebabkan penurunan DBI terbesar dianggap paling berpengaruh.



Gambar 7. Nilai DBI Setelah Di Hapus

Gambar 7 ini menunjukkan nilai DBI atribut Kimia jauh lebih rendah (0.337) dibandingkan atribut lainnya yang rata-rata sekitar 500. Hal ini menegaskan bahwa atribut Kimia memiliki peran signifikan dalam kualitas pengelompokan.

4.6. Evaluation & Interpretation

Karakteristik setiap cluster dianalisis melalui rata-rata dan standar deviasi atributnya, dirangkum dalam Tabel 17 untuk evaluasi pola.

Tabel 17. Hasil Data setiap cluster

Attribute	Cluster_0	Cluster_1
NAMA	199	197
Pendidikan Agama Islam dan Budi Pekerti	80.191	82.360
Pendidikan Pancasila	82.844	84.457
Bahasa Indonesia	81.492	81.726
MATEMATIKA	79.824	87.533
Bahasa Inggris	79.171	80.985
Pendidikan Jasmani, Olahraga dan Kesehatan	82.618	85.914
Kimia	80.201	80.437
Informatika	84.528	85.574

Tabel 17 menunjukkan nilai rata-rata mata pelajaran di setiap cluster. Cluster 0 memiliki nilai rata-rata lebih rendah, dengan Bahasa Inggris terendah (79.171) dan Informatika tertinggi (84.528). Cluster 1 memiliki Matematika tertinggi (87.533) dan Kimia terendah (80.437). Data ini mengungkapkan perbedaan kontribusi atribut terhadap karakteristik masing-masing cluster. Pada Setiap Cluster Memiliki Nilai Karakteristik tertentu Cluster_1 terdiri dari 197 data siswa, menunjukkan pola nilai yang lebih konsisten dan cenderung lebih tinggi, mengindikasikan performa akademik yang lebih baik dibandingkan Cluster_0.

4.7. Pembahasan

Hasil analisis menunjukkan nilai K optimal untuk data nilai akademik siswa adalah 2, dengan nilai DBI 0.519. DBI mendekati nol menunjukkan clustering yang baik, sedangkan mendekati 1 mengindikasikan clustering yang kurang tepat ([14]; [1]). Atribut Kimia paling berpengaruh dalam pembentukan cluster, dengan DBI 0.337 setelah dihapus. Atribut Pendidikan Jasmani Olahraga, Informatika, dan Matematika juga berperan penting dalam membedakan cluster ([15]). Karakteristik cluster menunjukkan bahwa Cluster 0 mencakup siswa dengan nilai rendah, terutama pada Bahasa Inggris (79.167) dan Matematika (79.823), sementara Cluster 1 terdiri dari siswa dengan nilai lebih tinggi, khususnya pada Matematika (87.495) dan Pendidikan Jasmani Olahraga (85.899). Pengelompokan ini memberikan wawasan untuk strategi pembelajaran yang lebih terarah.

5. KESIMPULAN DAN SARAN

Penelitian K-Means pada data akademik siswa SMK Samudra Nusantara menghasilkan dua klaster optimal dengan nilai Davies-Bouldin Index (DBI) sebesar 0,519. Klaster pertama (Cluster_0) terdiri dari siswa dengan nilai rata-rata rendah, seperti Bahasa Inggris 79,167 dan Matematika 79,823, sementara klaster kedua (Cluster_1) mencakup siswa dengan nilai rata-rata tinggi, seperti Matematika 87,495 dan Pendidikan Jasmani 85,899. Atribut Kimia terbukti paling signifikan dalam memengaruhi kualitas klasterisasi, yang terlihat dari penurunan nilai DBI menjadi 0,337 ketika atribut tersebut dihapus. Hasil penelitian ini mendukung penerapan strategi pembelajaran yang disesuaikan, seperti program remedial untuk siswa di Cluster_0 dan tantangan tambahan bagi siswa di Cluster_1. Namun, karena analisis penelitian ini terbatas pada nilai akademik, penelitian lanjutan disarankan untuk menyertakan atribut tambahan, seperti kehadiran siswa, partisipasi dalam kegiatan ekstrakurikuler, atau latar belakang sosial, guna menghasilkan klasterisasi yang lebih komprehensif dan mendalam.

DAFTAR PUSTAKA

[1] A'yuni, Q., Nazir, A., Handayani, L., & Afrianty, I. (2023). Penerapan Algoritma K-Means Clustering untuk Mengetahui Pola Penerima Beasiswa Bank Indonesia (BI). Journal of Computer System and Informatics (JoSYC), 4(3), 530–539. https://doi.org/10.47065/josyc.v4i3.3343

[2] Abineno, R. T., & Pekuwali, A. A. (2024). Penerapan Algoritma K-Means Clustering Nilai. 720–734.

[3] Adiyanto, A., & Arie Wijaya, Y. (2023). Penerapan Algoritma K-Means Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023. JATI (Jurnal Mahasiswa Teknik Informatika), 7(2), 1344–1350. https://doi.org/10.36040/jati.v7i2.6849

[4] Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. Jurnal SAINTEKOM, 13(1), 100–110. https://doi.org/10.33020/saintekom.v13i1.402

[5] Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa

Tegalwangi. Hello World Jurnal Ilmu Komputer, 2(1), 20–33. https://doi.org/10.56211/helloworld.v2i1.230

[6] Fikri Sallaby, A., Tri Alinse, R., Novita Sari, V., & Ramadani, T. (2022). Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil Penjualan Di Toko Widya Bengkulu. Jurnal Media Infotama, 18(1), 2022.

[7] Febrian, A., Nana Suarna, & Gifthera Dwilestari. (2022). Penerapan Algoritma K-Means Untuk Mengelompokkan Data Pengiriman Paket Di Kantor Pos Cirebon. Jurnal Teknologi Technoscientia, 15(1), 23–27. https://doi.org/10.34151/technoscientia.v15i1.3858

[8] Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021). Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari. Jurnal Sains Teknologi Transportasi Maritim, 3(1), 1–7. https://doi.org/10.51578/j.sitektransmar.v3i1.30

[9] Haryadi, D., & Atmaja, D. M. U. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung. Journal of Informatics and Communication Technology (JICT), 3(2), 51–66. https://doi.org/10.52661/j_ict.v3i2.85

[10] Rady Putra, I., & Anggrawan, M. (2021). Pengelompokan Kelayakan Penerima Bantuan Sosial Menggunakan Algoritma K-Means Clustering. Jurnal Informatika Universitas Pembangunan Nasional, 22(3), 88-95. https://doi.org/10.33028/jinfotech.v22i3.453

[11] Suryadi, S. (2019). Penerapan Metode Clustering K-Means Untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi. Jurnal Informatika, 6(1), 52–72. https://doi.org/10.36987/informatika.v6i1.738

[12] Triyandana, D., Wibowo, A., & Syamsuddin, M. (2022). Pengelompokan Menu Makanan dan Minuman Menggunakan K-Means Clustering. Jurnal Teknologi dan Sistem Informasi, 17(2), 112–121. https://doi.org/10.33020/jtsinformasi.v17i2.721

[13] Yunitasari, M., Maharani, T., & Hikmahwan, B. (2022). Implementasi Metode K-Means Untuk Pengelompokan Data Jamaah Pada Biro Umroh Jabal Rahmah Pacitan. KLIK: Kumpulan Jurnal Ilmu Komputer, 09(1), 1–9. https://doi.org/10.20527/klik.v9i1

[14] Permadi, A., & Wijaya, Y. A. (2023). Pengelompokkan Dataset Bus Menggunakan Algoritma K-Means. INFORMATICS FOR EDUCATORS AND PROFESSIONAL: Journal of Informatics, 7(2), 138. https://doi.org/10.51211/itbi.v8i1.2259

[15] Kristanto, B., Turmudi Zy, A., & M. Fatchan. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. Bulletin of Information Technology (BIT), 4(1), 112–120. https://doi.org/10.47065/bit.v4i1.521