

Implementation Of K-Means With Evaluation Of The Davies-Bouldin Index For Grouping The Leading Class Of Wijaya Sukodono SMP

Penerapan K-Means Dengan Evaluasi Davies Bouldin Index Untuk Pengelompokan Kelas Unggulan SMP Wijaya Sukodono

Feny Anggraeny¹⁾, Ade Eviyanti²⁾, Sumarno³⁾, Rohman Dijaya⁴⁾

^{1,2,3,4)} Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi : adeeviyanti@umsida.ac.id

Abstract. *This research was conducted at SMP Wijaya Sukodono, one of the largest educational institutions in Sukodono District, which aims to improve the quality of learning through the utilization of student academic data. The main focus of this study is to group students based on academic performance using the K-Means Clustering algorithm. This approach divides students into two main categories: the Superior Class, which includes students with high academic achievement, and the Regular Class, which consists of students with lower academic performance. The stages of the study include collecting report card data, processing and transforming data, and implementing the K-Means algorithm. Evaluation of the results was carried out using the Davies-Bouldin Index (DBI) to measure the effectiveness of the clustering. The analysis showed that out of 576 students, 488 students were included in the Superior Class, while 88 students were included in the Regular Class. Grouping with two clusters gave the best results with a DBI value of 0.337, indicating optimal separation between clusters. This study concludes that K-Means Clustering is an effective method in grouping students based on academic achievement. These findings provide valuable insights for schools in designing more focused learning programs to improve the quality of education. Future in-depth studies may consider non-academic variables or other clustering methods for more comprehensive results.*

Keywords - K-Means Clustering, Student Grouping, Academic Performance, Davies-Bouldin Index, Outstanding Class

Abstrak. *Penelitian ini dilakukan di SMP Wijaya Sukodono, salah satu institusi pendidikan terbesar di Kecamatan Sukodono, yang bertujuan meningkatkan mutu pembelajaran melalui pemanfaatan data akademik siswa. Fokus utama dari penelitian ini adalah mengelompokkan siswa berdasarkan performa akademik menggunakan algoritma K-Means Clustering. Pendekatan ini membagi siswa ke dalam dua kategori utama: Kelas Unggulan, yang mencakup siswa dengan prestasi akademik tinggi, dan Kelas Reguler, yang terdiri atas siswa dengan performa akademik lebih rendah. Tahapan penelitian meliputi pengumpulan data nilai raport, pengolahan dan transformasi data, serta Pengimplementasian K-Means. Analisis hasil menggunakan Davies-Bouldin Index (DBI) sebagai alat ukur efektivitas pengelompokan. Analisis menunjukkan bahwa dari 576 siswa, sebanyak 488 siswa masuk dalam Kelas Unggulan, sementara 88 siswa tergolong dalam Kelas Reguler. Pengelompokan dengan dua klaster memberikan hasil terbaik dengan nilai DBI sebesar 0,337, yang menandakan pemisahan antar-klaster yang optimal. Analisis ini menyimpulkan bahwasanya K-Means Clustering adalah teknik yang efektif untuk mengelompokkan siswa berdasarkan capaian akademik. Temuan ini memberikan wawasan berharga bagi sekolah dalam merancang program pembelajaran yang lebih terfokus untuk meningkatkan kualitas pendidikan. Kajian mendalam di masa mendatang dapat mempertimbangkan variabel non-akademik atau metode clustering lainnya untuk hasil yang lebih komprehensif.*

Kata Kunci - K-Means Clustering, Pengelompokan Siswa, Nilai Akademik, Davies-Bouldin Index, Kelas Unggulan

1. Pendahuluan

Pendidikan memiliki peran krusial dalam membangun generasi muda yang bermutu dan kompetitif di era globalisasi. Salah satu langkah yang dapat ditempuh untuk meningkatkan mutu pendidikan adalah dengan memanfaatkan teknologi data mining untuk mengidentifikasi dan mengelompokkan siswa berdasarkan prestasi akademik. Dalam konteks ini, SMP Wijaya Sukodono sebagai salah satu sekolah terbesar di Kecamatan Sukodono terus berupaya meningkatkan mutu pendidikan dengan memanfaatkan data akademik siswa sebagai dasar pengambilan keputusan strategis[1].

Namun, pengelompokan siswa berdasarkan nilai raport secara manual seringkali tidak efisien dan rentan terhadap bias subjektif. Dengan jumlah siswa yang besar, diperlukan metode objektif dan efisien untuk menganalisis data ini. Pendekatan data mining menawarkan solusi dalam mengubah data untuk menghasilkan informasi strategis yang dapat mendukung keputusan berbasis bukti[2].

Penelitian ini memanfaatkan teknik K-Means Clustering dalam mengelompokkan siswa menjadi dua kategori utama, yaitu Kelas Unggulan dan Kelas Reguler. Kelas Unggulan terdiri dari siswa dengan performa akademik tinggi, sedangkan Kelas Reguler mencakup siswa dengan performa akademik yang lebih rendah. Algoritma K-Means didasarkan pada sifatnya yang sederhana dan efisien dalam mengolah data berukuran besar dengan atribut multidimensi[3], [4].

Sebagai tambahan, Python digunakan dalam implementasi metode ini karena keunggulannya dibandingkan platform lain. Python mendukung banyak pustaka (seperti Scikit-learn) yang memudahkan analisis data dan evaluasi model. Selain itu, Python memiliki efisiensi waktu pelatihan yang lebih baik dibandingkan beberapa platform serupa, sehingga mendukung proses analisis yang lebih cepat dan akurat[5].

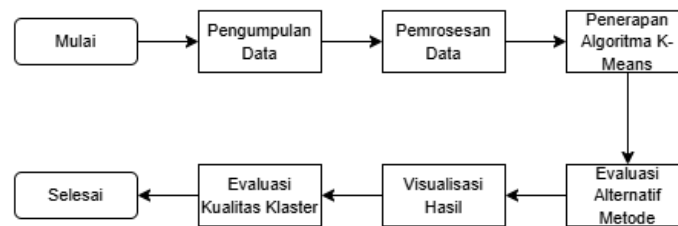
Beberapa penelitian sebelumnya telah memanfaatkan teknik clustering untuk keperluan pendidikan. Misalnya, penelitian [2] menunjukkan bahwa K-Means efektif untuk pengelompokan siswa berdasarkan nilai raport di SMK Muhammadiyah 3 Weleri. Namun, penelitian tersebut belum membahas evaluasi kualitas clustering menggunakan metrik seperti Davies-Bouldin Index (DBI). Di sisi lain, penelitian [6] menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi siswa berprestasi, tetapi tidak mengatasi heterogenitas data atau ketidakseimbangan antar kelompok.

Berdasarkan analisis ini, penelitian ini bertujuan untuk mengisi celah dalam literatur dengan menerapkan K-Means Clustering yang dievaluasi menggunakan DBI. Dengan demikian, penelitian ini tidak hanya memberikan gambaran distribusi akademik siswa di SMP Wijaya Sukodono tetapi juga menawarkan solusi praktis untuk meningkatkan efektivitas pembelajaran melalui pemanfaatan teknologi data mining.

2. Metode Penelitian

Diagram Alur Penelitian

Penelitian ini mengikuti serangkaian langkah yang dirancang untuk memastikan implementasi yang sistematis dan akurat. Gambar 1 menunjukkan diagram alur penelitian[7].



Gambar 1. Diagram Alur Penelitian

Pengumpulan Data

Data dikumpulkan melalui observasi langsung, yaitu dengan mengambil data secara langsung dari nilai rapor siswa di SMP Wijaya Sukodono, yang mencakup total 576 siswa. Selain itu, studi literatur juga dilakukan untuk mendukung penelitian ini, dengan memanfaatkan teori-teori terkait pengelompokan data akademik menggunakan algoritma K-Means. Studi literatur tersebut juga mencakup analisis kualitas hasil clustering, salah satunya melalui metode Davies-Bouldin Index (DBI), yang berfungsi untuk mengevaluasi kualitas kelompok yang dihasilkan [8].

Pemrosesan Data

Seleksi data dilakukan dengan memilih atribut-atribut penting, seperti nilai rapor yang relevan untuk proses clustering. Nilai-nilai ini kemudian dihitung rata-ratanya untuk setiap siswa sebagai dasar analisis[9]. Selanjutnya, tahap preprocessing data dilakukan untuk memproses dan membersihkan data dari elemen yang tidak relevan, seperti data kosong atau duplikat, guna meningkatkan kualitas analisis[10]. Selanjutnya, data ditransformasikan dengan menggunakan metode normalisasi MinMaxScaler untuk menyamakan skala antar atribut, sehingga proses clustering dapat berjalan dengan lebih akurat dan konsisten[11].

Penerapan Algoritma K-Means

Penentuan jumlah kluster dalam penelitian ini dilakukan dengan menetapkan dua kluster, yaitu Kelas Unggulan dan Kelas Reguler. Pemilihan jumlah kluster ini didasarkan pada kebutuhan sekolah serta hasil evaluasi menggunakan Davies-Bouldin Index (DBI). Proses clustering dilakukan dengan menerapkan algoritma K-Means menggunakan library Scikit-learn di Python. Parameter *n_clusters* diatur menjadi 2, sementara *random_state* diset ke 42 untuk menjaga konsistensi hasil[12]. Evaluasi model clustering dilakukan dengan menggunakan DBI, di mana nilai DBI dibandingkan dengan beberapa konfigurasi jumlah kluster lainnya untuk memastikan kualitas hasil yang optimal. Selain itu, evaluasi tambahan menggunakan *silhouette score* juga dilakukan untuk memperkuat keandalan hasil clustering yang diperoleh[13].

Evaluasi Alternatif Metode

Sebagai bagian dari pengembangan penelitian, metode lain seperti hierarchical clustering dan DBSCAN juga dipertimbangkan untuk perbandingan. Hasil clustering dari algoritma K-Means dibandingkan dengan algoritma-algoritma ini untuk menilai keunggulan relatifnya[14].

Visualisasi Hasil

Hasil clustering ditampilkan dalam bentuk visual menggunakan diagram scatter plot untuk mempermudah interpretasi dan analisis lebih lanjut. Tiap kluster ditampilkan dengan warna berbeda untuk menggambarkan distribusi siswa[15].

Evaluasi Kualitas Kluster

Penggunaan Davies-Bouldin Index (DBI) menjadi metode utama untuk menilai kualitas clustering dengan mengukur tingkat pemisahan dan kekompakan antar-kluster. Nilai DBI yang lebih rendah menunjukkan bahwa hasil clustering memiliki kualitas yang lebih baik, karena kluster-kluster lebih terpisah secara optimal dan data dalam kluster lebih homogen[16]. Selain itu, evaluasi tambahan dilakukan menggunakan silhouette score untuk memastikan bahwa hasil clustering tidak hanya menghasilkan kluster yang terpisah dengan baik, tetapi juga memiliki tingkat pemisahan yang optimal di antara data. Silhouette score memberikan perspektif tambahan dalam mengevaluasi kualitas struktur kluster yang dihasilkan[17].

Dengan alur metodologi ini, penelitian ini diharapkan menghasilkan pengelompokan siswa yang relevan, akurat, dan bermanfaat bagi pengambilan keputusan di SMP Wijaya Sukodono.

3. Hasil dan Pembahasan

Data yang diperoleh dari SMP Wijaya Sukodono merupakan nilai rapor siswa yang berjumlah 576 siswa yang dapat dilihat pada tabel dibawah ini.

Tabel 1. Sampel Data set

NO	NAMA	JK	AGAMA_P	AGAMA_K	PPKN_P	...	BTQ_K	TIK_P	TIK_K
1	Adam Surya Adlythina	L	83	83	90	...	85	86	86
2	Ahmad Raka Andreas Setyaria	L	78	79	82	...	78	80	80
3	Andi Apriliani Putri	P	90	92	90	...	0	90	92
4	Chelsea Safina Sabilal Ul'ya	P	81	83	90	...	85	80	80
...
573	Vernando	L	80	80	88	...	80	83	83
574	Mustafidah Dwi Agustina	P	77	79	77	...	75	84	87
575	Devina Putri Cahyani	P	80	80	85	...	85	80	80
576	Fera Yolanda	P	80	80	88	...	84	83	83

Seleksi Data

Setelah seluruh data berhasil dikumpulkan, langkah berikutnya adalah melakukan seleksi atribut yang akan digunakan dalam penerapan algoritma K-Means. Fitur ini berfungsi untuk menyaring atribut data yang relevan dan dibutuhkan untuk proses perhitungan lebih lanjut[18]. Berikut ini adalah tahapan seleksi atribut dari datasheet nilai raport SMP Wijaya Sukodono.

```
df = pd.read_csv('/content/hasil_rata_rata.csv')
```

```
df
```

	NO	URUT	NAMA	JK	AGAMA_P	AGAMA_K	PPKN_P	PPKN_K	SIN_P	SIN_K	NAT_P	...	PRAKARYA_P	PRAKARYA_K	JAWA_P	JAWA_K	BTQ_P	BTQ_K	TIK_P	TIK_K	Rata-rata Prestasi	Rata-rata Keterampilan
0	1	Adam Surya Adlythina	L	83	83	90	90	88	88.0	78	...	93.0	93.0	83.0	83.0	83.0	85.0	80.0	80.0		85.307692	86.923077
1	2	Ahmad Raka Andreas Setyaria	L	78	79	82.0	83	78	78.0	79	...	85.0	85.0	80.0	82.0	78.0	78.0	80.0	80.0		80.153846	81.538462
2	3	Andi Apriliani Putri	P	90	92	90.0	90	90	92.0	93	...	90.0	92.0	90.0	92.0	0.0	0.0	90.0	92.0		83.692308	84.615385
3	4	Chelsea Safina Sabilal Ul'ya	P	81	83	90.0	90	80	82.0	80	...	85.0	85.0	85.0	87.0	85.0	85.0	80.0	80.0		83.000000	84.789231
4	5	Chintya Putri Novitasari	P	80	83	90.0	85	82	82.0	80	...	90.0	90.0	82.0	83.0	80.0	80.0	80.0	79.0		81.923077	83.692308
...
571	572	David Peratama	L	90	92	90.0	90	93	93.0	92	...	87.0	90.0	88.0	89.0	90.0	89.0	90.0	90.0		90.153846	91.538462
572	573	Vernando	L	80	80	88.0	90	80	80.0	80	...	79.0	79.0	80.0	80.0	80.0	80.0	83.0	83.0		82.153846	82.538462
573	574	Mustafidah Dwi Agustina	P	77	79	77.0	78	75	75.0	75	...	75.0	75.0	79.0	79.0	75.0	75.0	84.0	87.0		77.076923	77.461538
574	575	Devina Putri Cahyani	P	80	80	85.0	85	80	82.0	80	...	85.0	85.0	85.0	87.0	85.0	85.0	80.0	80.0		82.538462	84.153846
575	576	Fera Yolanda	P	80	80	88.0	90	85	85.0	80	...	85.0	85.0	78.0	78.0	84.0	84.0	83.0	83.0		82.692308	84.153846

576 rows x 31 columns

Gambar 2. Hasil seleksi data

Setelah dilakukan proses seleksi, data tersebut ditambah dengan kolom baru yang menunjukkan rata-rata nilai setiap siswa. Hasil seleksi ini memberikan gambaran yang lebih ringkas dan memudahkan analisis performa siswa secara keseluruhan, karena hanya kolom-kolom penting yang disajikan. Kolom "Rata-rata" membantu untuk menilai tingkat pencapaian siswa secara menyeluruh.

Processing Data

Setelah tahap seleksi data nilai selesai dilakukan, langkah selanjutnya adalah proses processing data dari hasil seleksi sebelumnya. Proses ini bertujuan untuk membersihkan data dari komponen yang tidak diperlukan. Berikut ini adalah tahapan processing pada datasheet nilai raport SMP Wijaya Sukodono.



```
[8] df = df[['NO.URUT', 'NAMA', 'Rata-rata Prestasi', 'Rata-rata Keterampilan']]
df.head(4)
```

	NO.URUT	NAMA	Rata-rata Prestasi	Rata-rata Keterampilan
0	1	Adam Surya Adlythina	85.307692	86.923077
1	2	Ahmad Raka Andreas Setyaria	80.153846	81.538462
2	3	Andi Apriliani Putri	83.692308	84.615385
3	4	Chelsea Safina Sabila Ul'ya	83.000000	84.769231

Langkah berikutnya: [Buat kode dengan df](#) [Lihat plot yang direkomendasikan](#) [New interactive sheet](#)

Gambar 3. Processing Data

Gambar 3 menunjukkan proses processing data dari dataset nilai rapor siswa SMP Wijaya Sukodono. Pada tahap ini, data hasil seleksi sebelumnya dibersihkan dengan cara menghapus komponen atau kolom yang tidak relevan dengan tujuan analisis. Dataset yang ditampilkan memuat informasi no. urut, nama, rata-rata prestasi, rata-rata keterampilan. Proses ini bertujuan untuk menyederhanakan data sehingga hanya kolom-kolom penting yang digunakan dalam analisis lebih lanjut. Data yang telah diproses ini siap digunakan untuk tahap berikutnya, seperti clustering atau klasifikasi untuk menentukan kelas unggulan dan kelas reguler.

Transformasi Data

Setelah proses processing data selesai, langkah berikutnya adalah melakukan transformasi data pada hasil tahapan sebelumnya. Transformasi data adalah langkah penting dalam proses analisis, di mana data diubah menjadi format yang sesuai atau tepat untuk diolah. Berikut ini adalah tahapan transformasi data pada datasheet nilai raport SMP Wijaya Sukodono.

```

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
x_train = scaler.fit_transform(x_train)

x_train
array([[0.59784367, 0.57682491],
       [0.56172507, 0.54109239],
       [0.58652291, 0.56151097],
       ...,
       [0.54016173, 0.51403777],
       [0.57843666, 0.55844819],
       [0.57951482, 0.55844819]])

```

Gambar 4. Transformasi Data

Gambar 4 menunjukkan transformasi data pada dataset nilai rapor siswa SMP Wijaya Sukodono menggunakan MinMaxScaler. Transformasi ini mengubah data menjadi rentang 0 hingga 1 untuk menyamakan skala antar fitur, sehingga analisis lebih akurat. Prosesnya menggunakan `fit_transform`, yang menghitung nilai minimum dan maksimum data, lalu menerapkannya ke dataset. Hasil transformasi berupa array dengan nilai yang telah dinormalisasi, seperti 0.566237 dan 0.510499, memastikan data siap untuk analisis lebih lanjut.

K-Means Clustering

Setelah tahap transformasi data selesai, langkah selanjutnya adalah menerapkan proses K-Means Clustering pada hasil transformasi tersebut. Tujuan dari tahap ini adalah mengelompokkan data ke dalam cluster berdasarkan kesamaan nilai atribut yang telah dipilih, sesuai dengan tujuan analisis yang diinginkan. Berikut ini adalah tahapan K-Means Clustering pada datasheet nilai raport SMP Wijaya Sukodono.

```

from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(x_train)

KMeans
KMeans(n_clusters=2, random_state=42)

```

Gambar 5. Menentukan jumlah cluster

Gambar 5 menunjukkan proses menentukan jumlah cluster menggunakan metode K-Means Clustering. Pada langkah ini, library `KMeans` diimpor dari `sklearn.cluster`, dan model K-Means diinisialisasi dengan parameter `n_clusters=2`, yang berarti data akan dikelompokkan ke dalam 2 cluster. Parameter `random_state=42` digunakan untuk memastikan hasil pengelompokan konsisten setiap kali algoritma dijalankan. Setelah itu, model dilatih menggunakan data yang telah ditransformasikan dengan perintah `kmeans.fit(x_train)`, sehingga algoritma dapat mengidentifikasi pola dan menentukan cluster berdasarkan kemiripan data.

```

#menentukan titik centorid
kmeans.cluster_centers_

array([[0.59047159, 0.56259561],
       [0.41738177, 0.35853636]])

```

Gambar 6. Menentukan titik centroid

Gambar 6 menunjukkan proses menentukan titik centroid dari masing-masing kluster yang terbentuk dari K-Means Clustering. Centroid merupakan titik pusat setiap kluster, yang dihitung berdasarkan rata-rata posisi data dalam cluster tersebut. Dalam gambar, perintah `kmeans.cluster_centers_` digunakan untuk memperoleh koordinat centroid[19]. Hasilnya adalah dua titik centroid dengan koordinat $[0.59047159, 0.56259561]$ untuk cluster pertama dan $[0.41738177, 0.35853636]$ untuk cluster kedua. Titik-titik ini berfungsi sebagai representasi pusat dari masing-masing cluster dan membantu dalam memahami karakteristik utama dari data dalam tiap kelompok.

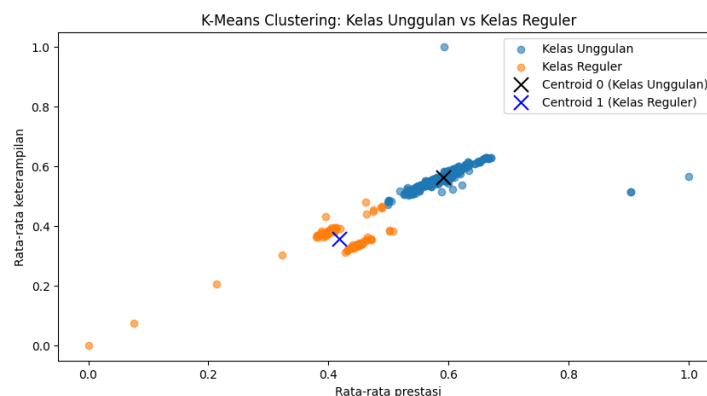
Hasil Clustering

Algoritma K-Means digunakan untuk melakukan clustering dengan konfigurasi dua kluster yaitu Kelas Unggulan dan Kelas Reguler. Data yang digunakan mencakup 576 siswa dari SMP Wijaya Sukodono. Untuk memberikan gambaran yang lebih jelas, tabel berikut menyajikan distribusi jumlah siswa berdasarkan hasil pengelompokan dalam setiap kluster:

Tabel 2. Hasil clustering

Cluster	Jumlah
0 (Unggulan)	488
1 (Reguler)	88

Visualisasi hasil clustering ditampilkan dalam grafik scatter plot, di mana setiap kluster diwakili oleh warna berbeda, dan centroid ditunjukkan dengan simbol khusus. Analisis ini menunjukkan bahwa mayoritas siswa tergolong dalam Kelas Unggulan.

**Gambar 7.** Visualisasi hasil clustering

Evaluasi Kualitas Clustering

Evaluasi DBI menunjukkan nilai 0,337 untuk konfigurasi dua kluster, yang merupakan nilai terbaik dibandingkan konfigurasi kluster lainnya. Nilai DBI yang rendah menunjukkan bahwa pemisahan antar-kluster sudah optimal dengan tingkat kompaksi yang tinggi.

Tabel 3. Evaluasi DBI

Jumlah Klaster	Davies-Bouldin Index
2	0.337075975
3	0.623500165
4	0.560577797
5	0.471238263

Sebagai evaluasi tambahan, silhouette score digunakan untuk memastikan tingkat pemisahan antar-klaster. Nilai silhouette score untuk konfigurasi dua klaster adalah 0,802, yang menunjukkan bahwa anggota klaster memiliki tingkat kedekatan tinggi dengan klasternya masing-masing dan cukup terpisah dari klaster lain.

Tabel 4. Evaluasi Silhouette Score

Jumlah Klaster	Nilai Silhouette Score
2	0,802
3	0,501
4	0,510
5	0,530

Analisis Karakteristik Klaster

Kelas Unggulan (Cluster 0): Siswa dalam klaster ini memiliki rata-rata nilai akademik yang lebih tinggi. Centroid klaster ini menunjukkan karakteristik umum siswa unggulan, seperti konsistensi pada nilai raport di semua mata pelajaran.

Kelas Reguler (Cluster 1): Siswa dalam klaster ini memiliki rata-rata nilai akademik yang lebih rendah. Analisis lebih lanjut menunjukkan beberapa faktor yang mungkin memengaruhi performa mereka, seperti variasi signifikan pada nilai beberapa mata pelajaran.

Perbandingan dengan Algoritma Lain

Sebagai langkah pengembangan, algoritma DBSCAN dan hierarchical clustering diuji pada dataset yang sama. Hasil menunjukkan bahwa K-Means menunjukkan performa terbaik dengan nilai Silhouette Score tertinggi (0,802) yang menjadikannya algoritma paling cocok untuk penelitian ini karena menghasilkan klasterisasi dengan pemisahan yang jelas dan homogenitas tinggi, sementara DBSCAN unggul dalam DBI terendah (0,162), dan Hierarchical Clustering memberikan hasil yang seimbang, namun tidak unggul di kedua metrik[20], seperti yang dirangkum di bawah ini:

Tabel 5. Perbandingan Algoritma

Algoritma	DBI	Silhouette Score
K-Means	0,337	0,802
DBSCAN	0,162	0,790
Hierarchical	0,321	0,801

Pembahasan

Evaluasi Alternatif dan Kualitas

Evaluasi kualitas clustering menggunakan Davies-Bouldin Index menunjukkan bahwa konfigurasi dua kluster memberikan hasil terbaik dengan nilai DBI terendah dibandingkan dengan konfigurasi kluster lainnya. Silhouette score memperkuat hasil ini dengan skor yang tinggi, menegaskan homogenitas dalam setiap kluster dan pemisahan yang jelas antar-kluster.

Sebagai pembanding, algoritma lain seperti DBSCAN dan hierarchical clustering diuji pada dataset yang sama. Hasil perbandingan menunjukkan bahwa K-Means unggul dalam menghasilkan struktur kluster yang jelas dengan nilai silhouette score tertinggi. Namun, DBSCAN menunjukkan performa lebih baik dalam mengelola data dengan distribusi tidak seragam, yang menjadi peluang untuk penelitian selanjutnya.

Refleksi Berdasarkan Penelitian Terdahulu

Penelitian ini memiliki beberapa perbedaan signifikan dengan penelitian terdahulu yang menjadi rujukan, baik dari segi pendekatan, metodologi, maupun hasil yang dihasilkan. Salah satu perbedaan utama adalah pada evaluasi kualitas clustering. Dalam penelitian ini, evaluasi dilakukan dengan menggunakan Davies-Bouldin Index (DBI) dan silhouette score, yang memberikan pengukuran kuantitatif terhadap pemisahan dan kekompakan antar kluster, memastikan hasil clustering yang optimal. Hal ini melangkah lebih jauh dibandingkan penelitian [2], yang menggunakan algoritma K-Means tanpa mengevaluasi kualitas klusterisasi dengan metrik tertentu. Selain itu, penelitian ini menitikberatkan pada pengelompokan siswa berdasarkan performa akademik untuk mendukung strategi pendidikan, berbeda dengan penelitian [6] yang menggunakan algoritma K-Nearest Neighbor (KNN) untuk memperkirakan siswa berprestasi. Penelitian ini tidak hanya mengelompokkan siswa tetapi juga memberikan wawasan praktis untuk perencanaan pendidikan yang lebih baik.

Penelitian ini juga mengeksplorasi algoritma lain seperti DBSCAN dan hierarchical clustering sebagai pembanding. Hasil evaluasi menunjukkan bahwa K-Means menghasilkan performa terbaik dalam hal silhouette score, sementara DBSCAN memberikan nilai DBI terendah untuk distribusi data tertentu. Hal ini membuat penelitian ini lebih komprehensif dibandingkan penelitian terdahulu yang hanya menggunakan satu algoritma tanpa eksplorasi lebih lanjut. Implementasi penelitian ini di SMP Wijaya Sukodono dengan fokus spesifik pada pengelompokan siswa berdasarkan nilai rapor memberikan solusi yang langsung dapat diterapkan pada institusi pendidikan tersebut, berbeda dengan penelitian terdahulu yang cenderung bersifat lebih generik tanpa fokus khusus pada kebutuhan atau konteks institusi tertentu.

Selain itu, penelitian ini memberikan rekomendasi strategis untuk sekolah dalam menyusun program pembelajaran, seperti program percepatan untuk siswa di Kelas Unggulan dan program remedial untuk siswa di Kelas Reguler. Temuan ini berbeda dari penelitian sebelumnya yang umumnya hanya berfokus pada penerapan algoritma tanpa memberikan solusi strategis yang dapat diimplementasikan langsung oleh pihak sekolah. Dalam hal inovasi pada pemanfaatan data akademik, penelitian ini memproses data siswa secara menyeluruh, mulai dari seleksi atribut hingga normalisasi data menggunakan MinMaxScaler, untuk memastikan akurasi hasil clustering.

Proses ini menambah nilai dibandingkan penelitian sebelumnya yang kurang menekankan pada tahap preprocessing data.

Dengan pendekatan-pendekatan ini, penelitian ini tidak hanya melengkapi kekurangan dalam penelitian terdahulu tetapi juga memberikan kontribusi baru, baik dalam hal metodologi maupun penerapan praktis di bidang pendidikan. Penelitian ini menunjukkan bahwa algoritma K-Means, bila dilengkapi dengan evaluasi yang tepat, dapat menjadi alat yang efektif untuk meningkatkan kualitas pengambilan keputusan strategis di institusi pendidikan.

Menghubungkan Temuan dengan Literatur

Penemuan penelitian ini memberikan rekomendasi strategis bagi sekolah untuk merancang program pembelajaran yang disesuaikan dengan kebutuhan setiap klaster. Kelas Unggulan dapat difokuskan pada program percepatan, sementara Kelas Reguler membutuhkan program remedial yang lebih intensif. Pendekatan ini sejalan dengan literatur yang menekankan pentingnya segmentasi berbasis data dalam meningkatkan kualitas pendidikan.

Dengan hasil clustering yang akurat, sekolah dapat mengambil langkah proaktif dalam menyusun kebijakan pendidikan yang inklusif dan berbasis bukti. Selain itu, integrasi variabel non-akademik di masa depan, seperti tingkat kehadiran atau partisipasi ekstrakurikuler, dapat memberikan wawasan yang lebih komprehensif.

4. Penutup

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan siswa SMP Wijaya Sukodono berdasarkan nilai akademik mereka. Hasil clustering menunjukkan bahwa dua klaster, yaitu Kelas Unggulan dan Kelas Reguler, merupakan konfigurasi optimal dengan distribusi masing-masing 488 siswa dan 88 siswa. Evaluasi menggunakan Davies-Bouldin Index (DBI) menghasilkan nilai 0,337, yang menunjukkan kualitas clustering yang baik dengan pemisahan antar-klaster yang jelas. Evaluasi tambahan menggunakan silhouette score memberikan nilai 0,802, yang menunjukkan tingkat kompaksi dan pemisahan klaster yang optimal. Analisis karakteristik klaster menunjukkan bahwa siswa dalam Kelas Unggulan memiliki rata-rata nilai akademik yang lebih tinggi, sedangkan siswa dalam Kelas Reguler cenderung memiliki performa yang lebih rendah. Informasi ini memberikan wawasan strategis bagi sekolah untuk meningkatkan kualitas pendidikan. Melalui penelitian ini, algoritma K-Means terbukti efektif dalam mengelompokkan siswa berdasarkan performa akademik, yang dapat menjadi dasar untuk pengambilan keputusan strategis dalam menyusun program pembelajaran.

Penelitian selanjutnya disarankan untuk memasukkan variabel non-akademik, seperti tingkat kehadiran, partisipasi ekstrakurikuler, dan faktor sosial-ekonomi, untuk menghasilkan pengelompokan yang lebih komprehensif. Sebagai langkah pengembangan, algoritma clustering lain seperti DBSCAN atau hierarchical clustering dapat dicoba untuk membandingkan kualitas hasil pengelompokan, terutama pada dataset yang lebih kompleks. Metode ini dapat diterapkan di sekolah lain dengan dataset serupa untuk mengidentifikasi siswa berdasarkan performa akademik. Penyesuaian jumlah klaster dan parameter evaluasi perlu dilakukan sesuai kebutuhan masing-masing sekolah. Berdasarkan hasil clustering, sekolah dapat merancang program pembelajaran yang sesuai, seperti program percepatan untuk siswa dalam Kelas Unggulan dan program remedial serta mentoring untuk siswa dalam Kelas Reguler. Sekolah juga disarankan untuk melakukan evaluasi berkala terhadap keefektifan strategi yang diimplementasikan.

berdasarkan hasil clustering ini. Hal ini akan memastikan bahwa program yang dijalankan memberikan dampak positif terhadap seluruh siswa.

5. Referensi

- [1] N. D. Rahayu, A. H. Anshor, I. Afriantoro, and A. Halim Anshor, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means Oleh : Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI : Jurnal Komputer dan Informatika*, vol. 6, 2024.
- [2] H. Afif and M. Arif, "Jurnal Teknik Informatika dan Desain Komunikasi Visual Penerapan Data Mining Untuk Pengelompokan Kelas Siswa Unggulan Berdasarkan Nilai Raport Menggunakan Metode K-Means Clustering (Studi Kasus: SMK Muhammadiyah 3 Weleri)," *Universitas Selamat Sri*, vol. 1, no. 2, 2022.
- [3] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [4] E. Desi, S. Aliyah, S. Lestari, and W. Dari, "Implementasi Algoritma K-Means Untuk Penerimaan Siswa Baru Di SMANPAS Berdasarkan Nilai Rapot dan Hasil Tes," *IT (INFORMATIC TECHNIQUE) JOURNAL*, vol. 10, no. 1, p. 01, Apr. 2022, doi: 10.22303/it.10.1.2022.01-10.
- [5] D. A. Manalu and G. Gunadi, "IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI," *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [6] S. Widaningsih and S. Yusuf, "Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 3, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [7] N. F. Hilmi and F. Irwiensyah, "Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naïve Bayes," *SMATIKA JURNAL*, vol. 14, no. 01, pp. 146–156, Jul. 2024, doi: 10.32664/smatika.v14i01.1210.
- [8] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa."
- [9] S. Haviyola and M. Jajuli, "PENGELOMPOKAN PRESTASI SISWA GUNA KUALIFIKASI BEASISWA BERDASARKAN DATA NILAI MENGGUNAKAN ALGORITMA K-MEANS."
- [10] M. Aqshal Al Fachrizy and Hendri, "Klasifikasi Pencari Kerja pada Disnaker Menggunakan Metode K-Means Clustering," *Bulletin of Computer Science Research*, vol. 4, no. 2, pp. 196–206, Feb. 2024, doi: 10.47065/bulletincsr.v4i2.334.
- [11] M. Alexander Justin Audison Sibarani, I. Gede Susrama Mas Diyasa, P. Studi Sains Data, F. Ilmu Komputer, and U. Pembangunan, "PENGUNAAN K-MEANS DAN HIERARCHICAL CLUSTERING SINGLE LINKAGE DALAM PENGELOMPOKAN STOK OBAT," vol. 5, no. 2, 2024, doi: 10.46306/lb.v5i2.

- [12] W. Kokoh Andriyan, D. Palembang, J. A. Jenderal Yani No, and P. Sumatera Selatan, "Penerapan Data Mining Dengan Menggunakan Metode K-Means Clustering Dalam Pengelompokan Data Nilai Pada SMA YKPP PENDOPO Untuk Menentukan Jurusan Ipa Dan Ips," 2023.
- [13] B. Kristanto, A. T. Zy, and M. Fatchan, "Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 112–120, 2023, doi: 10.47065/bit.v3i1.
- [14] A. Kristianto, "Implementasi DBSCAN dalam Clustering Data Minat Mahasiswa Setelah Pandemi Covid19," 2022.
- [15] G. Triyandana, L. A. Putri, and Y. Umaidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [16] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [17] F. Yunita, "PENERAPAN DATA MINING MENGGUNKAN ALGORITMA K-MEANS CLUSTERING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS: UNIVERSITAS ISLAM INDRAGIRI)," 2018.
- [18] R. Ishak, "Volume 5 Nomor 2 Juli 2023 Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah," *Jambura Journal of Electrical and Electronics Engineering*, vol. 192.
- [19] Y. Hasan, "PENGUKURAN SILHOUETTE SCORE DAN DAVIES-BOULDIN INDEX PADA HASIL CLUSTER K-MEANS DAN DBSCAN," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5001.
- [20] S. Mutiah, Y. Hasnataeni, A. Fitrianto, E. Erfiani, and L. M. R. D. Jumansyah, "Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat," *Teorema: Teori dan Riset Matematika*, vol. 9, no. 2, p. 247, Sep. 2024, doi: 10.25157/teorema.v9i2.16290.