

PENERAPAN ALGORITME *K-MEANS CLUSTERING* UNTUK MENGELOMPOKKAN SISWA BERDASARKAN NILAI AKADEMIK DI SMP NEGERI 207 SSN

Reza Pahlevi Kurniawan^{1*}, Ferdiansyah²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}rezapahlv31@gmail.com, ²ferdiansyah@budiluhur.ac.id

(* : corresponding author)

Abstrak-Pendidikan merupakan salah satu bagian penting untuk memajukan suatu negara. Kualitas pendidikan dapat terlihat melalui tingkat keberhasilan siswa yang tinggi dan tingkat kegagalan yang rendah. Untuk memastikan keberhasilan dalam pendidikan, penting bagi lembaga pendidikan untuk memonitor dan menganalisis data siswa secara teratur. SMP Negeri 207 SSN merupakan salah satu lembaga pendidikan dengan jumlah siswa sebanyak 780 siswa pada tahun ajaran 2022/2023. Banyaknya jumlah siswa menyebabkan kurangnya pemahaman tentang potensi siswa dan karakteristiknya yang menjadi kendala dalam memberikan pelayanan pendidikan yang sesuai. Sehingga perlu dilakukan pengolahan data guna menggali informasi dan mengidentifikasi kebutuhan siswa untuk memberikan pelayanan pendidikan yang lebih efektif. Dengan adanya teknologi *data mining*, dapat dilakukan pengelompokan siswa berdasarkan nilai akademiknya. Pengelompokan ini dilakukan menggunakan metode *data mining* yaitu *K-Means Clustering* dan hasil *Clustering* dievaluasi menggunakan metode *Davies Bouldin Index*. Data selection yang digunakan berjumlah 15 atribut. Hasil dari pengujian yang dilakukan dengan memproses data menggunakan jumlah Cluster 2 sampai 7 didapatkan bahwa jumlah *Cluster* yang optimal untuk mengelompokkan data nilai siswa kelas 7 angkatan 2022 SMPN 207 SSN yaitu 3 *Cluster* yang menghasilkan 16 siswa dengan nilai terendah, 92 siswa dengan nilai menengah, dan 141 siswa dengan nilai tertinggi. Dari hasil *Clustering* tersebut dievaluasi menggunakan metode *Davies Bouldin Index* yang menghasilkan nilai 1,12388.

Kata Kunci: *Data Mining, Clustering, K-Means Clustering, Siswa*

IMPLEMENTATION OF *K-MEANS CLUSTERING* ALGORITHM TO GROUP STUDENTS BASED ON ACADEMIC VALUE IN SMP NEGERI 207 SSN

Abstract-Education is an important part of advancing a country. The quality of education can be seen through high student success rates and low failure rates. To ensure success in education, it is important for educational institutions to regularly monitor and analyze student data. SMP Negeri 207 SSN is an educational institution with a total of 780 students in the 2022/2023 academic year. The large number of students causes a lack of understanding of the potential of students and their characteristics which is an obstacle in providing appropriate educational services. So it is necessary to do data processing in order to gather information and identify student needs to provide more effective educational services. With data mining technology, it is possible to group students based on their academic grades. This grouping is done using the data mining method, namely *K-Means Clustering*. Data selection using 15 attributes. The results of the tests carried out by processing the data using the number of clusters 2 to 7 found that the optimal number of clusters for classifying grade 7 student grade data of 2022 SMPN 207 SSN is 3 clusters which produce 16 students with the lowest grades, 92 students with intermediate grades and 141 student with the highest score. From the results of the clustering it was evaluated using the *Davies Bouldin Index* method which produced a value of 1.12388.

Keywords: *Data Mining, Clustering, K-Means Clustering, Student*

1. PENDAHULUAN

Pendidikan merupakan salah satu bagian penting untuk memajukan suatu negara. Pendidikan memiliki peran penting dalam membentuk karakter dan kemampuan individu, serta membantu meningkatkan kualitas sumber daya manusia suatu negara. Melalui pendidikan, individu dapat memperoleh pengetahuan, keterampilan, dan nilai-nilai yang diperlukan untuk menjadi anggota masyarakat yang produktif dan berkontribusi positif terhadap pembangunan negara. Kualitas pendidikan dapat terlihat melalui tingkat keberhasilan siswa yang tinggi dan tingkat

kegagalan yang rendah. Untuk memastikan keberhasilan dalam pendidikan, penting bagi lembaga pendidikan untuk memonitor dan menganalisis data siswa secara teratur.

SMP Negeri 207 SSN merupakan salah satu lembaga pendidikan dengan jumlah siswa sebanyak 780 siswa pada tahun ajaran 2022/2023. Banyaknya jumlah siswa menyebabkan kurangnya pemahaman tentang potensi siswa dan kemampuan belajarnya yang menjadi kendala dalam memberikan pelayanan pendidikan yang sesuai. Sehingga perlu dilakukan pengolahan data guna menggali informasi dan mengidentifikasi kebutuhan siswa untuk memberikan pelayanan pendidikan yang lebih efektif. Teknologi *data mining* dapat digunakan dalam mengolah dan menganalisis data siswa SMP Negeri 207 SSN.

Menurut [1] *data mining* adalah proses menemukan dan menganalisis data dalam jumlah besar dengan tujuan menemukan pola atau informasi yang menarik dari sejumlah besar data yang tersimpan dengan menggunakan teknik atau metode tertentu. Menurut [2] *data mining* yang disebut juga dengan *Knowledge Discovery in Database (KDD)* adalah sebuah proses secara otomatis atas pencarian data didalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi hubungan (*Association*) atau pengelompokan (*Clustering*). Dengan adanya teknologi *data mining*, dapat dilakukan pengelompokan siswa berdasarkan nilai akademiknya. Pengelompokan siswa sangat penting dalam pendidikan, karena dapat membantu lembaga pendidikan dalam mengidentifikasi siswa-siswa yang berprestasi dan memberikan penghargaan kepada mereka, serta memberikan perhatian bagi siswa yang membutuhkan bantuan tambahan. Pengelompokan ini dilakukan dengan mengelompokkan siswa dari kemampuannya dalam memahami materi pelajaran. Pengelompokan siswa dilakukan menggunakan proses *Clustering* dengan membagi sekelompok siswa menjadi bagian yang disebut *Cluster*. Menurut [3] *Clustering* atau pengklasteran adalah suatu teknik *data mining* yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokan data atau lebih tepatnya mempartisi dari *dataset* kedalam *subset*. Menurut [4] *Clustering* atau klasterisasi adalah salah satu alat bantu atau salah satu metode dalam penerapan *data mining* yang bertujuan mengelompokkan objek-objek kedalam *Cluster-cluster*. Metode *Clustering* yang digunakan untuk mengelompokkan nilai siswa ini yaitu metode *K-Means Clustering*.

K-Means merupakan salah satu metode *data mining Clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *Cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *Cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya [5]. Pada penelitian sebelumnya [6] sudah ada permasalahan dalam mengelompokkan siswa berdasarkan nilainya menggunakan algoritme *K-Means*, penelitian tersebut menyatakan bahwa algoritme *K-Means* memang cukup efektif dalam mengelompokkan data nilai siswa.

2. METODE PENELITIAN

2.1 *K-Means Clustering*

Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain [7].

Langkah-langkah melakukan *Clustering* dengan metode *K-Means* adalah sebagai berikut [7]:

- Pilih jumlah *Cluster* k
- Inisialisasi k pusat *Cluster* ini bisa dilakukan dengan berbagai cara, namun yang paling sering dilakukan adalah dengan cara *random*. Pusat-pusat *Cluster* diberi nilai awal dengan angka-angka *random*.
- Tempatkan semua data atau objek ke *Cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *Cluster* tertentu ditentukan jarak antara data dengan pusat *Cluster*. Dalam tahap ini perlu dihitung jarak tiap data dengan satu *Cluster* tertentu akan menentukan suatu data masuk dalam *Cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *Cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai persamaan (1) berikut:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2} \quad (1)$$

Keterangan:

$D(i, j)$ = Jarak data ke i ke pusat *Cluster* j

X_{ki} = Data ke i pada atribut ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- Hitung kembali pusat *Cluster* dengan keanggotaan *Cluster* yang sekarang. Pusat *Cluster* adalah rata-rata dari semua data atau objek dalam *Cluster* tertentu yang dirumuskan sebagai persamaan (1) berikut:

$$C = \frac{\sum m}{n} \quad (2)$$

Keterangan:

C = Merupakan *Centroid* data

m = Anggota data yang termasuk ke dalam *Centroid* tertentu

n = Jumlah data yang menjadi anggota *Centroid* tertentu

e. Melakukan perulangan dari langkah ketiga dan keempat, sampai anggota tiap *Cluster* tidak ada yang berubah.

2.2 Davies Bouldin Index

Menurut [8] *Davies Bouldin Index* (DBI) merupakan salah satu metode evaluasi *Cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *Centroid* dari *Cluster* yang diikuti, sedangkan separasi didasarkan pada jarak antar *Centroid* dari *Cluster*-nya. Perhitungan evaluasi dimulai dari menentukan nilai SSW, SSB dari titik *Centroid* awal lalu akan menentukan nilai DBI nya [9]. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster* (SSW) yaitu sebagai persamaan (3) berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan rumus:

m = jumlah data dalam *Cluster* ke i

x = data pada *Cluster*

c = *Centroid*

d(x,c) = jarak data dengan *Centroid*

Sum of Square Between-cluster (SSB) bertujuan untuk mengetahui separasi antar cluster dengan persamaan sebagai persamaan (4) berikut:

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Keterangan rumus:

c_i = *Cluster* satu

c_j = *Cluster* lain

d(c_i,c_j) = jarak *Centroid* dengan *Centroid* lain

Selanjutnya untuk menghitung nilai rasio yang dimiliki oleh masing-masing *Cluster*, digunakan persamaan (5) berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

R_{ij} adalah ukuran rasio seberapa baik nilai perbandingan antara *Cluster* ke I dengan *Cluster* ke j. *Cluster* yang baik adalah *Cluster* dengan kohesi sekecil mungkin dan separasi sebesar mungkin. Persamaan DBI yaitu sebagai persamaan (6) berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Keterangan rumus:

R_{ij} = rasio SSW dan SSB dari dua *Cluster*

K = jumlah *Cluster* terbentuk

2.3 Pengumpulan Data

Dalam proses penelitian, penulis memerlukan data yang relevan dengan topik yang sedang dibahas. Untuk mencapai tujuan ini, dilakukan metode pengumpulan data guna mendapatkan informasi yang diperlukan:

2.3.1 Wawancara

Wawancara dilakukan dengan Wakil Kepala Sekolah SMPN 207 untuk memperoleh informasi mengenai kriteria penilaian, sistem perhitungan nilai rapor, dan pengetahuan mengenai prestasi serta karakteristik siswa. Wawancara ini dilakukan untuk memperoleh wawasan yang lebih mendalam tentang sistem penilaian yang digunakan dan pengetahuan guru mengenai prestasi serta karakteristik siswa untuk digunakan sebagai acuan dalam pembuatan sistem aplikasi.

2.3.2 Dokumentasi

Dalam proses dokumentasi, penulis meminta kepada Wakil Kepala Sekolah terkait data yang diperlukan sebagai masukan dalam proses pembuatan sistem aplikasi.

2.3.3 Tinjauan Pustaka

Tinjauan Pustaka dilakukan dengan tujuan menghimpun informasi dan pengetahuan mengenai *data mining* dan *K-Means Clustering* dari jurnal-jurnal terpercaya. Nantinya, informasi ini akan dijadikan referensi penting dalam pelaksanaan penelitian.

2.4 Data Penelitian

Data penelitian adalah informasi atau data yang bersifat asli, tidak mengalami manipulasi, dan bebas dari campur tangan pihak lain. Dengan demikian, data tersebut dapat digunakan untuk analisis dalam suatu penelitian. Penelitian dilakukan di SMP Negeri 207 SSN yang merupakan sekolah menengah pertama standar nasional yang terletak di DKI Jakarta. SMPN 207 memiliki siswa sebanyak 780 siswa pada tahun ajaran 2022/2023 yang terdiri dari 249 siswa kelas 7, 252 siswa kelas 8, dan 279 siswa kelas 9. Data yang penulis dapat yaitu data nilai rapor semester ganjil siswa kelas 7 dalam bentuk tujuh file microsoft excel kelas 7A sampai 7G yang jika digabungkan berisi 249 baris data.

2.5 Tahapan *Knowledge Discovery in Database (KDD)*

2.5.1 *Data Selection*

Data selection merupakan suatu langkah yang bertujuan untuk memilih atribut-atribut yang relevan dari kumpulan data, sehingga analisis data dapat dilakukan dengan lebih efektif. Pada tahap ini digabungkan file excel dari setiap kelas dan di tambahkan atribut “kelas” sebagai informasi kelas pada setiap siswa. Dari beberapa atribut yang telah disebutkan sebelumnya, dipilih 15 atribut seperti berikut:

- a. NIS
- b. Nama
- c. Kelas
- d. SBD (Seni Budaya)
- e. IPS (Ilmu Pengetahuan Sosial)
- f. IPA (Ilmu Pengetahuan Alam)
- g. BING (Bahasa Inggris)
- h. MAT (Matematika)
- i. PAI (Pendidikan Agama Islam)
- j. PAK (Pendidikan Agama Kristen)
- k. PAKH (Pendidikan Agama Katolik)
- l. PJOK (Pendidikan Jasmani, Olahraga, dan Kesehatan)
- m. PRA (Prakarya)
- n. BIND (Bahasa Indonesia)
- o. PKN (Pendidikan Kewarganegaraan)

2.5.2 *Pre-Processing / Cleaning*

Pada tahap ini Teknik yang diterapkan adalah data cleaning. Data cleaning merupakan suatu langkah yang harus dilakukan untuk menghilangkan data noise ataupun missing values [10]. Pada tahap pre-processing ini,

penulis menggabungkan mata pelajaran PAI (Pendidikan Agama Islam), PAK (Pendidikan Agama Kristen), dan PAKH (Pendidikan Agama Katolik) menjadi satu atribut yaitu PA (Pendidikan Agama) jadi total atribut yang digunakan yaitu 13 atribut. Penulis juga menghapus kolom kompetensi pada setiap atribut mata pelajaran, karena yang akan digunakan hanya nilai dari setiap mata pelajaran. Penulis juga mengubah data nilai siswa yang berisi (-) menjadi (0).

2.5.3 Transformation

Transformasi data merupakan tahapan yang dilakukan untuk mengubah data yang telah dipilih sehingga sesuai dengan kebutuhan dalam proses data mining. Proses transformasi dalam KDD (Knowledge Discovery in Databases) bersifat kreatif dan sangat tergantung pada jenis atau pola informasi yang ingin diambil dari database. Meskipun demikian, pada tahap ini penulis memutuskan untuk tidak melakukan transformasi data karena data yang diperoleh dari SMPN 207 sudah sesuai dan siap untuk dilakukan proses data mining.

2.5.4 Data Mining

Setelah proses pembersihan dan modifikasi data, langkah selanjutnya adalah mengolah data menggunakan algoritme *K-Means Clustering* untuk mendapatkan informasi yang terkandung dalam data tersebut. Data hasil olahan akan disimpan dalam *database* dan diproses secara otomatis melalui aplikasi yang telah dikembangkan dengan menerapkan teknik dan perhitungan algoritme *K-Means Clustering*. Data yang di proses hanya atribut nilai mata pelajarannya saja.

2.5.5 Interpretation / Evaluation

Setelah data diolah dengan proses *data mining*, kemudian data dievaluasi untuk mengetahui keakuratan informasi yang dihasilkan. Pada tahap ini digunakan metode *Davies Bouldin Index (DBI)* yang bertujuan untuk menguji kualitas dari *Cluster* yang dihasilkan.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Algoritme *K-Means Clustering*

Pada sub bab ini akan dijelaskan proses *Clustering* menggunakan algoritme *K-Means Clustering*. Data Tabel 1 yang digunakan untuk pemodelan *Clustering* adalah data nilai rapor semester ganjil kelas 7 angkatan 2022 yang berjumlah 249 baris data, berikut cuplikan datanya:

Tabel 1. Data *Clustering*

NO	NIS	NAMA	KELAS	SBD	IPS	IPA	BING	MAT	PA	...	PKN
1	11414	ABDURRAHMAN HUSIN THOYYIB	7A	85	76	75	76	78	76	...	78
2	11434	ALBERTUS AGUNG NOVALDO	7A	88	82	87	76	83	91	...	85
3	11436	ALFIANO BAYU FEBRIAN	7A	75	75	75	75	76	75	...	75
...
247	11641	SOFI DELIA ARIFAH	7G	83	75	77	89	78	78	...	76
248	11658	YULIANITA ISTIQOMAH	7G	84	75	77	89	78	75	...	76
249	11660	ZAKY KAMALUDDIN	7G	85	81	86	93	82	80	...	84

Langkah pertama yang dilakukan yaitu menentukan jumlah *Cluster*. Pada proses perhitungan ini ditentukan 3 *Cluster*. Setelah menentukan jumlah *Cluster*, kemudian menentukan nilai *Centroid* awal secara acak. Pada proses perhitungan ini, nilai *Centroid* awal yang dipilih dapat dilihat pada Tabel 2 di bawah ini:

Tabel 2. *Centroid* Awal

<i>Centroid</i>	NIS	NAMA	SBD	IPS	IPA	BING	MAT	PA	...	PKN
-----------------	-----	------	-----	-----	-----	------	-----	----	-----	-----

C1	11436	ALFIANO BAYU FEBRIAN	75	75	75	75	76	75	...	75
C2	11439	ALIFAH NURUL FADHA	88	85	84	77	81	88	...	83
C3	11472	CHELSEA AZZAHRA PUTRI SUHARTO	92	95	94	90	86	93	...	85

Setelah menentukan nilai *Centroid*, kemudian hitung jarak tiap data dengan tiap *Centroid* menggunakan rumus *Euclidean Distance* yang akan menentukan suatu data masuk dalam *Cluster* mana. Proses perhitungannya sebagai berikut dan hasilnya dijabarkan pada Tabel 3:

ABDURRAHMAN HUSIN THOYYIB

$$\begin{aligned} \text{Jarak ke C1} &= \sqrt{\frac{(85 - 75)^2 + (76 - 75)^2 + (75 - 75)^2 + (76 - 75)^2 + (78 - 76)^2 + (76 - 75)^2 + \dots + (78 - 75)^2}{(76 - 75)^2 + \dots + (78 - 75)^2}} = 17,34935 \\ \text{Jarak ke C2} &= \sqrt{\frac{(85 - 88)^2 + (76 - 85)^2 + (75 - 84)^2 + (76 - 77)^2 + (78 - 81)^2 + (76 - 88)^2 + \dots + (78 - 83)^2}{(76 - 88)^2 + \dots + (78 - 83)^2}} = 21,79449 \\ \text{Jarak ke C3} &= \sqrt{\frac{(85 - 92)^2 + (76 - 95)^2 + (75 - 94)^2 + (76 - 90)^2 + (78 - 86)^2 + (76 - 93)^2 + \dots + (78 - 85)^2}{(76 - 93)^2 + \dots + (78 - 85)^2}} = 40,9634 \end{aligned}$$

Tabel 3. Hasil Perhitungan Iterasi Ke-1

NO	NIS	NAMA	KELAS	C1	C2	C3	Hasil
1	11414	ABDURRAHMAN HUSIN THOYYIB	7A	17,34935	21,79449	44,23456	C1
2	11434	ALBERTUS AGUNG NOVALDO	7A	32,68027	6,16441	23,76973	C2
3	11436	ALFIANO BAYU FEBRIAN	7A	0	30,88689	50,80354	C1
...
247	11641	SOFI DELIA ARIFAH	7G	17,4356	25,88436	41,02438	C1
248	11658	YULIANITA ISTIQQOMAH	7G	17,86057	26,85144	41,68933	C1
249	11660	ZAKY KAMALUDDIN	7G	31,63858	20,12461	26,45751	C2

Kemudian untuk menentukan nilai *Centroid* pada iterasi selanjutnya, dilakukan perhitungan dengan persamaan (2), hasilnya adalah Tabel 4 sebagai berikut:

Tabel 4. Centroid Iterasi Ke-2

<i>Centroid</i>	SBD	IPS	IPA	BING	MAT	PA	...	PKN
C1	80,11765	72,23529	71,52941	76	76,88235	75,88235	...	74,70588
C2	83,14286	82,69481	84,87013	84,62987	82,13636	85,93506	...	84,28571
C3	86,79487	87,03846	89,0641	94,23077	84,07692	89,84615	...	87,05128

Hasil dari iterasi ke-2 dapat dilihat pada Tabel 5 di bawah ini:

Tabel 5. Hasil Perhitungan Iterasi Ke-2

NO	NIS	NAMA	KELAS	C1	C2	C3	Hasil
1	11414	ABDURRAHMAN HUSIN THOYYIB	7A	13,08186	21,8817	33,66427	C1

2	11434	ALBERTUS AGUNG NOVALDO	7A	30,91192	11,95243	20,80313	C2
3	11436	ALFIANO BAYU FEBRIAN	7A	7,89748	30,50298	43,10127	C1
...
247	11641	SOFI DELIA ARIFAH	7G	15,06905	22,36801	31,77124	C1
248	11658	YULIANITA ISTIQOMAH	7G	15,21859	23,35421	32,37443	C1
249	11660	ZAKY KAMALUDDIN	7G	30,11087	12,56426	17,03394	C2

Jika hasil *Clustering* masih mengalami perubahan, maka langkah selanjutnya yaitu menentukan nilai *Centroid* untuk iterasi berikutnya. Nilai *Centroid* untuk iterasi ke-3 dapat dilihat pada Tabel 6 di bawah ini:

Tabel 6. Centroid Iterasi Ke-3

<i>Centroid</i>	SBD	IPS	IPA	BING	MAT	PA	...	PKN
C1	80,64706	72,23529	71,52941	75,64706	76,94118	75,88235	...	74,47059
C2	82,80741	82,57778	84,66667	83,57778	82,02222	85,6	...	84,0963
C3	86,45361	86,35052	88,52577	93,87629	83,84536	89,54639	...	86,81443

Hasil dari iterasi ke-3 dapat dilihat pada Tabel 7 di bawah ini:

Tabel 7. Hasil Perhitungan Iterasi Ke-3

NO	NIS	NAMA	KELAS	C1	C2	C3	Hasil
1	11414	ABDURRAHMAN HUSIN THOYYIB	7A	13,25032	21,11926	32,59467	C1
2	11434	ALBERTUS AGUNG NOVALDO	7A	30,90755	11,54063	20,04917	C2
3	11436	ALFIANO BAYU FEBRIAN	7A	8,159101	29,68927	41,90453	C1
...
247	11641	SOFI DELIA ARIFAH	7G	15,29265	22,16606	30,55315	C1
248	11658	YULIANITA ISTIQOMAH	7G	15,48188	23,14545	31,21228	C1
249	11660	ZAKY KAMALUDDIN	7G	30,32055	13,12649	15,94966	C2

Pada proses perhitungan ini, hasil perhitungan yang sama berada di iterasi ke-8 dan iterasi ke-9, hal itu dapat dilihat dari hasil *Cluster* tiap data tidak berpindah lagi, Maka iterasi selanjutnya tidak perlu dilakukan. proses perhitungannya tidak dimasukkan semua karena terlalu banyak. Hasil dari perhitungan *Clustering* ini menunjukkan bahwa ada 16 siswa dengan nilai terendah yang berada di *Cluster* 1 (C1), 92 siswa dengan nilai menengah yang berada di *Cluster* 2 (C2), dan 141 siswa dengan nilai tertinggi yang berada di *Cluster* 3 (C3).

3.2 Evaluasi Hasil *Clustering* Menggunakan DBI

Untuk menghitung nilai DBI, tahap pertama yang harus dilakukan yaitu menghitung nilai SSW dengan persamaan (3) yaitu menghitung rata-rata dari nilai jarak data ke *Cluster* sehingga didapatkan hasil sebagai berikut:

$$SSW_1 = 17,30174$$

$$SSW_2 = 9,43105$$

$$SSW_3 = 8,83229$$

Tahap kedua yaitu menghitung nilai SSB dengan persamaan (4) sehingga didapatkan hasil Tabel 8 sebagai berikut:

$$SSB_{12} = 27,80447$$

$$SSB_{13} = 38,669$$

$$SSB_{23} = 15,1551$$

Tabel 8. Matriks Hasil Perhitungan SSB

SSB	1	2	3
1	0	27,80447	38,669
2	27,80447	0	15,1551
3	38,669	15,1551	0

Tahap ketiga yaitu menghitung nilai *ratio* dan menentukan *max ratio* dengan persamaan (5) sehingga didapatkan hasil Tabel 9 sebagai berikut:

$$R_{12} = \frac{17,30174 + 9,43105}{27,80447} = 0,96146$$

$$R_{13} = \frac{17,30174 + 8,83229}{38,669} = 0,67584$$

$$R_{23} = \frac{9,43105 + 8,83229}{15,1551} = 1,20509$$

Tabel 9. Matriks Hasil Perhitungan Max Ratio

R	1	2	3	Max Ratio
1	0	0,96146	0,67584	0,96146
2	0,96146	0	1,20509	1,20509
3	0,67584	1,20509	0	1,20509

Setelah mendapatkan nilai *max ratio*, Langkah selanjutnya yaitu menghitung nilai *Davies Bouldin Index* (DBI) menggunakan persamaan (6) sehingga didapatkan hasil sebagai berikut:

$$DBI = \frac{0,96146 + 1,20509 + 1,20509}{3} = 1,12388$$

Evaluasi hasil *Clustering* menggunakan perhitungan *Davies Bouldin Index* (DBI) menghasilkan nilai 1,12388. Semakin mendekati 0 nilai DBI maka akan semakin baik hasil *Clustering* yang diperoleh. Hal tersebut menunjukkan bahwa nilai DBI yang dihasilkan dari perhitungan ini relatif sangat baik.

4. KESIMPULAN

Setelah melakukan serangkaian pengujian dan evaluasi pada aplikasi yang telah dibangun menggunakan *Dataset* dan Algoritme yang diusulkan, maka dapat disimpulkan bahwa jumlah *Cluster* yang optimal untuk mengelompokkan data nilai siswa kelas 7 angkatan 2022 SMPN 207 SSN yaitu 3 *Cluster* yang menghasilkan 16 siswa dengan nilai terendah, 92 siswa dengan nilai menengah dan 141 siswa dengan nilai tertinggi. Dari hasil *Clustering* tersebut dievaluasi menggunakan metode *Davies Bouldin Index* yang menghasilkan nilai 1,12388. Penelitian ini juga mendapatkan bahwa *K-Means* memiliki kelemahan yang disebabkan oleh penentuan pusat *Cluster* awal atau *Centroid* yang dilakukan secara acak. Hasil *Clustering* yang diperoleh dari algoritme *K-Means* sangat bergantung pada inisialisasi nilai *Centroid* awal yang ditentukan, hal ini menyebabkan sulitnya mendapatkan nilai *Centroid* awal yang unik, sehingga peneliti menggunakan metode DBI untuk mengevaluasi hasil *Clustering* agar dapat menentukan jumlah *Cluster* yang optimal dan mengetahui kualitas dari hasil *Clustering* yang diperoleh. Dari kelemahan algoritme *K-Means* yang telah diketahui, diharapkan pada penelitian selanjutnya untuk menggunakan algoritme lainnya dan bandingkan hasilnya.

DAFTAR PUSTAKA

- [1] Asmana, Y. A. Wijaya, and Martanto, “CLUSTERING DATA CALON SISWA BARU MENGGUNAKAN METODE K-MEANS DI SEKOLAH MENENGAH KEJURUAN WAHIDIN KOTA CIREBON,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 6, no. 2, pp. 552–559, 2022.
- [2] F. N. R. F. J. Aziz, B. D. Setiawan, and I. Arwani, “Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.,* vol. 2, no. 6, pp. 2243–2251, 2018.
- [3] A. Ali, “Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumpat Rejo Sukodono Sidoarjo,” *J. Tek. Inform. dan Sist. Inf.,* vol. 7, no. 3, pp. 395–407, 2020.
- [4] H. Yuwafi, F. Marisa, and I. D. Wijaya, “IMPLEMENTASI DATA MINING UNTUK MENENTUKAN SANTRI BERPRESTASI DI PP . MANAARULHUDA DENGAN METODE CLUSTERING ALGORITMA K-MEANS,” *J. SPIRIT,* vol. 11, no. 1, pp. 22–29, 2019.
- [5] S. A. Rahmah, “KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA),” *J. Inf. Technol. Res.,* vol. 1, no. 1, pp. 1–5, 2020.
- [6] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *J. Inf. Syst. Informatics,* vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [7] S. P. Tamba, F. T. Kesuma, and Feryanto, “PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING,” *J. Sist. Inf. Ilmu Komput. Prima (JUSIKOM PRIMA),* vol. 2, no. 2, pp. 67–72, 2019.
- [8] R. J. Hablum, A. Khairan, and Rosihan, “CLUSTERING HASIL TANGKAP IKAN DI PELABUHAN PERIKANAN NUSANTARA (PPN) TERNATE MENGGUNAKAN ALGORITMA K-,” *JIKO (Jurnal Inform. dan Komputer),* vol. 02, no. 1, pp. 26–33, 2019.
- [9] T. M. Dista and F. F. Abdulloh, “Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization,” *J. Media Inform. Budidarma,* vol. 6, no. 3, pp. 1339–1348, 2022, doi: 10.30865/mib.v6i3.4172.
- [10] T. Kurniawan and M. Jajuli, “Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means,” *Gener. J.,* vol. 6, no. 1, pp. 1–12, 2022.