

Implementasi Algoritma K-Medoids Dalam Mengelompokkan Siswa Berdasarkan Keaktifan Dalam Proses Pembelajaran

Noor Oktavia Ih'Diati¹, Rosmini², Anto³

^{1,2,3}Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
Email: ¹nooroktavia7@gmail.com, ²rosmini@ppkia.ac.id, ³anto@ppkia.ac.id

Abstrak

Pengelompokan siswa berdasarkan tingkat keaktifan mereka merupakan salah satu strategi efektif untuk meningkatkan kualitas pembelajaran. SMP 9 Tarakan saat ini belum memiliki sistem yang mampu mengelompokkan siswa berdasarkan keaktifan mereka dalam proses pembelajaran, yang dapat membantu dalam mengevaluasi hasil pembelajaran. Pada tahap awal penerapan metode ini, data yang dikumpulkan berasal dari nilai rapor siswa kelas VIII I Semester 2 (Genap) tahun ajaran 2022/2023. Karakter yang digunakan dalam analisis meliputi nilai Agama, PPKn, Matematika, IPA, IPS, Bahasa Indonesia, Bahasa Inggris, Penjaskes, dan Seni Budaya Keterampilan, dengan total 31 data yang dianalisis. Langkah kedua adalah menentukan jumlah cluster. Tahap ketiga melibatkan pemilihan cluster secara acak dengan medoid awal. Tahap keempat adalah menghitung jarak masing-masing siswa menggunakan metode Euclidean distance, lalu menandai jarak terdekat serta menghitung total jarak. Tahap kelima adalah menghitung total simpang (S) dan menggunakan Davies-Bouldin Index (DBI) untuk mencari nilai k yang optimal, dengan melakukan pengujian sebanyak lima kali dengan k=3. Berdasarkan hasil perhitungan, analisis pengelompokan data siswa menghasilkan tiga cluster menggunakan perhitungan Euclidean distance dan Davies-Bouldin Index. Hasilnya menunjukkan bahwa 3 siswa masuk ke dalam cluster Sangat Berminat, 4 siswa dalam cluster Berminat, dan 24 siswa dalam cluster Kurang Berminat.

Kata Kunci: algoritma k-medoids, clustering, keaktifan, pembelajaran, siswa.

Implementation of the K-Medoids Algorithm for Grouping Students Based on Their Engagement in the Learning Process

Abstract

Grouping students based on their level of engagement is an effective strategy to improve the quality of learning. SMP 9 Tarakan currently does not have a system that can group students based on their engagement in the learning process, which could assist in evaluating learning outcomes. In the initial stage of applying this method, the data collected came from the report card grades of 8th-grade students (Class VIII I) in the 2nd semester (Even Semester) of the 2022/2023 academic year. The characteristics used in the analysis include grades in Religion, Civic Education (PPKn), Mathematics, Science (IPA), Social Studies (IPS), Indonesian Language, English, Physical Education (Penjaskes), and Cultural Arts and Skills, with a total of 31 data points analyzed. The second step is to determine the number of clusters. The third step involves randomly selecting clusters with an initial medoid. The fourth step is to calculate the distance for each student using the Euclidean distance method, then mark the nearest distance and calculate the total distance. The fifth step is to calculate the total deviation (S) and use the Davies-Bouldin Index (DBI) to find the optimal value of k by conducting tests five times with k=3. Based on the calculation results, the analysis of student data grouping produced three clusters using Euclidean distance and Davies-Bouldin Index calculations. The results show that 3 students fall into the Highly Interested cluster, 4 students into the Interested cluster, and 24 students into the Less Interested cluster.

Keywords: activity, clustering, k-medoids algorithm, learning, students.

I. PENDAHULUAN

Kemajuan dalam teknologi informasi mendorong adanya inovasi yang terus-menerus, terutama dalam bidang pendidikan. Salah satu bentuk inovasi ini adalah

pengembangan dan pemanfaatan aplikasi, baik untuk proses pembelajaran maupun administratif di sekolah, guna meningkatkan kualitas pendidikan. Peningkatan mutu pendidikan adalah prioritas utama dalam kebijakan pembangunan di Indonesia yang harus selalu menjadi

perhatian. Guru, sebagai faktor utama, memiliki peran besar dalam memberikan pengajaran yang efektif dan bermutu. Perkembangan kualitas serta keaktifan siswa dalam proses pembelajaran menjadi elemen dasar yang mencerminkan keberhasilan pendidikan.

Namun, tidak semua siswa memiliki kemampuan yang sama dalam memahami materi. Hal ini menjadi tantangan bagi sekolah dalam mencapai keberhasilan pendidikan secara keseluruhan. SMP Negeri 9 Tarakan, yang terletak di Jl. P Aji Iskandar (Perum Kopri), Juata Krilil, saat ini belum menerapkan sistem pengelompokan siswa berdasarkan keaktifan dalam pembelajaran. Sistem seperti ini sangat diperlukan untuk mengukur sejauh mana kemampuan siswa dalam mengikuti proses belajar serta menjadi bahan evaluasi bagi sekolah dalam menilai keberhasilan para guru. Proses pengelompokan ini didasarkan pada nilai akademik dan keaktifan siswa. *Data mining* adalah teknik untuk mengelompokkan data dengan karakteristik serupa menggunakan metode tertentu. Metode ini digunakan untuk mengekstraksi informasi prediktif tersembunyi dari basis data, dan dianggap sebagai teknologi yang sangat potensial bagi perusahaan dalam memanfaatkan data warehouse [1].

Clustering adalah proses pengelompokan di mana data dalam satu kelompok memiliki tingkat kemiripan yang tinggi. Algoritma K-Medoids, misalnya, mencari nilai tengah (medoid) yang mewakili data dalam kelompok tersebut dan menghitung jarak antar data dari medoid tersebut. Algoritma ini berbeda dengan metode lain yang menggunakan rata-rata sebagai nilai tengah, karena K-Medoids menggunakan data aktual sebagai pusat kelompok [2] [3].

II. METODOLOGI PENELITIAN

A. Data Mining

Data mining adalah proses analisis data yang bertujuan untuk menemukan pola atau keterkaitan yang belum terungkap sebelumnya melalui metode terkini, yang berguna bagi pemilik data [1]. Secara umum, Data Mining merupakan proses ekstraksi informasi dari kumpulan data yang besar dengan tujuan mengungkap pengetahuan tersembunyi dan memudahkan penggunaannya dalam aplikasi real-time. DM menggunakan berbagai algoritma untuk analisis data, dengan teknik utama yang mencakup Clustering, Association, dan Klasifikasi.

Clustering, atau pengelompokan, adalah teknik efektif untuk analisis data eksploratif dan telah diterapkan di berbagai bidang. Metode clustering umumnya dikategorikan ke dalam tiga jenis utama: partisi, hierarki, berbasis grid, dan metode berbasis model. Pengelompokan berbasis partisi menghasilkan pembagian data di mana objek dalam satu kelompok lebih mirip satu sama lain dibandingkan dengan objek di kelompok lain. Berikut gambar Metode penelitian yang digunakan.



Gambar 1. Metode Penelitian

B. Clustering

Clustering adalah teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan karakteristik atau kesamaan tertentu. Objek-objek dalam satu cluster memiliki kemiripan yang lebih tinggi satu sama lain dibandingkan dengan objek-objek di cluster lain. Metode ini sering digunakan untuk analisis eksploratif, membantu dalam mengidentifikasi pola atau struktur tersembunyi dalam data yang besar dan kompleks tanpa memerlukan label atau kategori yang sudah ditentukan sebelumnya[4].

C. K-Medoids

K-Medoids, atau yang dikenal juga sebagai PAM (Partitioning Around Medoid), adalah metode yang digunakan untuk menentukan pusat dari sebuah cluster[5]. Algoritma k-medoids menggunakan nilai rata-rata dari setiap kelompok sebagai titik fokus cluster. Selain itu, algoritma ini juga memanfaatkan informasi objek tertentu sebagai titik pusat dari cluster tersebut. K-medoids dapat dijadikan strategi pembagian, karena ia menggunakan objek yang dikumpulkan dalam cluster untuk membentuk titik pusat cluster berdasarkan nilai rata-rata objek dalam kumpulan cluster dalam penelitian ini menggunakan 3 cluster yaitu cluster 1 (C1), cluster 2 (C2) dan cluster 3 (C3)[6],[7]. Adapun tahapan perhitungan algoritma k-medoids sebagai berikut:

1. Inisialisasi pusat cluster sebanyak k
2. Alokasikan setiap data atau objek ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Disrace dengan rumus sebagai berikut.

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

Keterangan :

$d(x,y)$ = jarak
 x_i = Data training
 y_i = Data testing
 I = variabel data
 N = dimensi data

3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoids baru.
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoids baru.
5. Hitung total simpangan (S) dengan rumus

$$S = (B - A) \quad (2)$$

Keterangan :

S = Simpangan Baku
 B = Total Distance Baru
 A = Total Distance Lama

Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoids.

6. Lakukan perulangan langkah 3 sampai dengan 5 hingga tidak terjadi perubahan medoids sehingga didapatkan cluster beserta anggota cluster masing-masing.

D. Davies-Bouldin Index (DBI)

Davies-Bouldin Index adalah metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin, yang dikenal dengan nama Davies-Bouldin Index (DBI), digunakan untuk mengevaluasi kualitas cluster. Evaluasi dengan Davies-Bouldin Index mengikuti skema evaluasi internal cluster, di mana kualitas cluster dinilai berdasarkan kuantitas dan kedekatan antar cluster yang dihasilkan. Davies-Bouldin Index merupakan salah satu metode yang digunakan untuk mengukur validitas hasil pengelompokan suatu cluster, dengan mengukur "cohesion" sebagai penjumlahan hubungan data dengan titik pusat cluster yang diikuti. Pemisahan antar cluster diukur berdasarkan jarak antara titik pusat cluster satu dengan yang lainnya. Pengukuran menggunakan Davies-Bouldin Index berusaha memaksimalkan jarak antar cluster (misalnya, antara cluster C_i dan C_j) sambil meminimalkan jarak antara titik-titik dalam satu cluster. Jika jarak antar cluster maksimal, ini menunjukkan bahwa kesamaan karakteristik antar cluster rendah, sehingga perbedaan antar cluster menjadi lebih jelas. Sebaliknya, jika jarak intra-cluster minimum, berarti setiap objek dalam cluster memiliki tingkat kesamaan karakteristik yang tinggi. Berikut langkah-langkah perhitungan DBI[2], [5].

1. Sum Of Square Within- Cluster (SSW)

Untuk mengetahui kohesi dalam sebuah cluster ke- i salah satunya adalah dengan menghitung nilai dari Sum Of Square Within-Cluster (SSW). Sebelum menghitung nilai SSW terlebih menghitung jarak dengan menggunakan rumus sebagai berikut :

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3)$$

Keterangan:

P, q = Dua titik di ruang n Euclidean

Q_i, P_i = Vector Euclidean, dimulai dari asal ruang (titik awal)

n = ruang- n

Rumus perhitungan SSW persamaan berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (4)$$

Keterangan:

M_i = jumlah data dalam cluster ke- i ,

C_i = centroid cluster ke- i

D = jarak setiap data kecentroid menggunakan jarak euclidean

2. Sum of square between cluster (SSB)

Perhitungan Sum Of Square Between-Cluster (SSB) bertujuan untuk mengetahui separasi atau jarak antar cluster. dengan rumus perhitungan sebagai berikut

$$SSB_{ij} = d(X_i, X_j) \quad (5)$$

Keterangan:

$d(X_i, X_j)$ = Jarak antara data ke I dengan data ke j di cluster lain.

3. Ratio (Rasio)

Perhitungan rasio ($R_{i,j}$) ini bertujuan untuk mengetahui nilai perbandingan antara cluster ke- i dan cluster ke- j untuk menghitung nilai rasio yang dimiliki oleh masing-masing cluster. indeks I dan j merupakan merepresentasikan jumlah cluster, dimana jika terdapat 4 cluster maka terdapat indeks sebanyak 4 yaitu i, j, k dan l . untuk menentukan nilai rasio dengan rumus sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (6)$$

Keterangan:

SSW_i = Sum Of Square Within- Cluster pada centroid i

$SSB_{i,j}$ = Sum of Square Between Cluster data ke i dengan j pada cluster yang berbeda

4. Davies Bouldin Index (DBI)

Nilai rasio yang diperoleh dari hasil perhitngan menggunakan rumus diatas berikutnya mencari nilai DBI dengan menggunakan perhitungan sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (7)$$

Keterangan:

K = jumlah cluster

Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh.

III. HASIL DAN PEMBAHASAN

A. Observasi dan Pengumpulan Data

Pada penelitian ini data yang digunakan adalah nilai raport siswa kelas VIII (delapan) semester 2 (genap) tahun ajaran 2022/2023 dengan total data sebanyak 31 Siswa. Data tersebut akan diproses untuk mengetahui hasil pengelompokan siswa berdasarkan berbagai kriteria di sekolah. Pengelompokan data ini dapat menjadi informasi baru bagi pihak sekolah dan berguna untuk membuat keputusan terkait kebijakan pendidikan ke depannya. Dalam melakukan clustering. Sebanyak 9 atribut yang digunakan dalam perhitungan yaitu Agama, PPKN, Matematika, Ilmu Pengetahuan Alam, Ilmu Pengetahuan Sosial, Bahasa Indonesia, Bahasa Inggris, Penjaskes, Seni budaya Keterampilan

B. Implementasi Algoritma K-Medoids

Penerapan Metode K-Medoids pada proses pengelompokkan siswa Bimbingan Belajar penelitian ini menggunakan nilai mata pelajaran Agama, Pendidikan Pancasila Kewarganegaraan, Matematika, Ilmu Pengetahuan Alam, Ilmu Pengetahuan Sosial, Bahasa Indonesia, Bahasa Inggris, penjaskes, Seni Budaya Keterampilan. Berikut data nilai siswa.

Tabel 1. Data Siswa

Kode	1	2	3	4	5	6	7	8	9
A001	82	92	84	87	83	84	90	88	93
A002	65	76	68	60	65	73	66	75	70
A003	78	78	74	71	73	78	73	88	81
A004	78	78	69	70	70	71	70	75	73
A005	79	82	74	78	71	83	83	88	82
A006	85	81	74	71	83	83	78	88	89
A007	88	88	97	96	94	88	89	88	96
A008	78	79	70	70	70	81	70	88	76
A009	80	82	80	75	75	79	83	88	86
A010	78	79	80	82	78	77	78	92	88
A011	76	83	73	70	70	75	70	88	86
A012	75	87	72	73	70	83	70	88	85
A013	82	84	84	84	73	81	73	88	85
A014	88	82	85	91	83	91	88	88	93
A015	78	76	68	70	69	76	71	90	81
A016	78	84	77	74	73	82	76	88	83
A017	80	82	89	81	76	79	83	88	86
A018	80	82	72	71	75	79	71	88	83
A019	78	79	70	70	68	74	70	88	78
A020	75	76	69	66	70	75	70	88	73
A021	80	79	72	70	68	74	70	88	78
A022	78	77	74	70	69	77	73	88	76
A023	80	79	76	74	73	78	74	88	80
A024	78	81	73	71	73	77	74	90	79
A025	82	85	75	74	75	79	79	88	83
A026	78	78	72	72	73	80	71	88	83
A027	75	82	73	73	70	79	81	88	85
A028	78	76	73	72	74	74	70	88	79
A029	78	86	71	71	73	78	70	88	78
A030	82	80	76	71	72	75	71	92	77
A031	80	82	74	79	72	85	71	88	81

Dari data siswa akan dilakukan proses pengclusteringan penilaian keaktifan siswa seperti langkah-langkah berikut ini

1. Inisiasi pusat cluster sebanyak k (jumlah cluster). Menentukan nilai centroid awal yang ditentukan secara random diasumsikan A007, A002, A018
2. Perhitungan jarak Euclian Distance medoids awal uji coba 1 dengan menggunakan rumus [1] sebagai berikut
A007 =

$$\sqrt{((82 - 88)^2 + (92 - 88)^2 + (84 - 97)^2 + (87 - 96)^2 + (83 - 94)^2 + (84 - 88)^2 + (90 - 89)^2 + (88 - 88)^2 + (93 - 96)^2)}$$

$$= 21.19$$

$$A018 =$$

$$\sqrt{((82 - 80)^2 + (92 - 82)^2 + (84 - 72)^2 + (87 - 71)^2 + (83 - 75)^2 + (84 - 79)^2 + (90 - 71)^2 + (88 - 88)^2 + (93 - 83)^2)}$$

$$= 57.00$$

$$A002 =$$

$$\sqrt{((82 - 65)^2 + (92 - 76)^2 + (84 - 68)^2 + (87 - 60)^2 + (83 - 65)^2 + (84 - 73)^2 + (90 - 66)^2 + (88 - 75)^2 + (93 - 70)^2)}$$

$$= 57.00$$

Perhitungan diatas merupakan sampel data A001 untuk mendapatkan nilai C1, C2 dan C3 menggunakan medoids awal perhitungan nilai cost seluruhnya, sehingga didapatkan total cost lama sebesar 341.92, seperti tabel berikut.

Tabel 2. Hasil Perhitunga Nilai Cost

Kode	C1	C2	C3	Kedekatan	Cluster
A001	21.19	57.00	57.00	21.19	C1
A002	72.46	29.95	0.00	0.00	C3
A003	48.74	6.08	27.53	6.08	C2
A004	59.87	19.72	18.11	18.11	C3
A005	41.90	15.20	36.47	15.20	C2
A006	39.09	13.96	40.99	13.96	C2
A007	0.00	47.80	72.46	0.00	C1
A008	54.52	9.85	24.33	9.85	C2
A009	37.52	15.26	38.47	15.26	C2
A010	35.27	17.35	41.81	17.35	C2
A011	51.25	8.37	27.66	8.37	C2
A012	49.06	10.00	30.68	10.00	C2
A013	35.10	18.36	41.90	18.36	C2
A014	18.57	35.07	60.37	18.57	C1
A015	55.52	10.49	25.79	10.49	C2
A016	42.60	8.94	33.06	8.94	C2
A017	30.50	23.30	44.88	23.30	C2
A018	47.80	0.00	29.95	0.00	C2
A019	56.06	10.86	23.26	10.86	C2
A020	60.27	15.39	18.97	15.39	C2
A021	54.80	10.49	24.68	10.49	C2
A022	53.18	11.27	24.33	11.27	C2
A023	45.74	7.55	30.27	7.55	C2
A024	49.30	6.56	28.11	6.56	C2
A025	41.52	9.75	35.47	9.75	C2
A026	48.98	5.10	28.44	5.10	C2
A027	46.39	12.61	31.78	12.61	C2
A028	51.40	9.22	26.19	9.22	C2
A029	51.29	7.21	27.15	7.21	C2
A030	50.20	10.05	30.10	10.05	C2
A031	44.27	10.82	34.15	10.82	C2

3. Setelah itu memilih objek medoid baru yang ditentukan secara random diasumsikan A014, A004, A016.
4. Perhitungan Euclian Distance medoids baru digunakan untuk mendapatkan hasil iterasi kedua dengan menggunakan rumus [1] :

$$\begin{aligned}
 A014 &= \sqrt{((82 - 88)^2 + (92 - 82)^2 + (84 - 85)^2 + (87 - 91)^2 + (83 - 83)^2 + (84 - 91)^2 + (90 - 88)^2 + (88 - 88)^2 + (93 - 93)^2)} \\
 &= 14.35
 \end{aligned}$$

$$\begin{aligned}
 A004 &= \sqrt{((82 - 78)^2 + (92 - 78)^2 + (84 - 70)^2 + (87 - 70)^2 + (83 - 70)^2 + (84 - 71)^2 + (90 - 70)^2 + (88 - 75)^2 + (93 - 73)^2)} \\
 &= 45.09
 \end{aligned}$$

$$\begin{aligned}
 A016 &= \sqrt{((82 - 78)^2 + (92 - 84)^2 + (84 - 77)^2 + (87 - 74)^2 + (83 - 73)^2 + (84 - 82)^2 + (90 - 76)^2 + (88 - 88)^2 + (93 - 83)^2)} \\
 &= 26.42
 \end{aligned}$$

Perhitungan diatas merupakan sampel data A001 untuk mendapatkan nilai C1, C2, dan C3 menggunakan medoids baru untuk iterasi kedua. Perhitungan nilai cost nilai kedekatan diambil dari satu nilai dan tiga cluster dengan nilai paling terkecil maka hasil membaca nilai paling terkecil juga. Dari perhitungan diatas mendapatkan total cost baru sebesar 343.50.

5. Menghitung total Simpangan (S) menggunakan rumus [2] sebagai berikut :

$$\begin{aligned}
 S &= (Total\ Cost\ baru - Total\ Cost\ Lama) \\
 &= 343.50 - 341.92 \\
 &= 1.58
 \end{aligned}$$

Dari hasil perhitungan simpang diperoleh hasil simpangan 1.58 karena Simpangan > 0, maka proses pengelompokan berhenti pada iterasi kedua namun apabila Simpangan < 0, makan ulangi langkah 3 sampai 5 hingga nilai Simpangan > 0.

6. Hasil output dari penelitian yaitu berupa 3 cluster, yaitu Sangat Berminat (SB), Berminat (B), Kurang Berminat (KB)

Tabel 3. Hasil Akhir

Kode	C1	C2	C3	Kedekatan	Medoids	Hasil
A001	14.35	45.09	26.42	14.35	C1	SB
A002	60.37	18.11	33.06	18.11	C2	B
A003	35.71	18.06	9.11	9.11	C3	KB
A004	47.70	0.00	23.47	0.00	C2	B
A005	26.93	25.88	9.22	9.22	C3	KB
A006	26.66	29.77	14.73	14.73	C3	KB
A007	18.57	59.87	42.60	18.57	C1	SB
A008	40.71	16.73	13.60	13.60	C3	KB

A009	25.04	27.60	9.43	9.43	C3	KB
A010	24.02	30.72	13.89	13.89	C3	KB
A011	39.09	19.97	11.83	11.83	C3	KB
A012	36.17	23.77	9.70	9.70	C3	KB
A013	24.06	30.07	13.38	13.38	C3	KB
A014	0.00	47.70	29.70	0.00	C1	SB
A015	41.52	17.92	15.68	15.68	C3	KB
A016	29.70	23.47	0.00	0.00	C3	KB
A017	21.14	33.88	16.64	16.64	C3	KB
A018	35.07	19.72	8.94	8.94	C3	KB
A019	42.87	14.46	15.49	14.46	C2	B
A020	47.28	14.63	19.87	14.63	C2	B
A021	41.79	14.87	14.83	14.83	C3	KB
A022	39.91	15.81	13.15	13.15	C3	KB
A023	32.82	19.03	7.68	7.68	C3	KB
A024	36.57	18.65	9.11	9.11	C3	KB
A025	28.69	23.58	6.56	6.56	C3	KB
A026	35.44	19.31	9.70	9.70	C3	KB
A027	32.60	23.41	8.77	8.77	C3	KB
A028	39.13	15.94	14.18	14.18	C3	KB
A029	39.12	17.92	11.22	11.22	C3	KB
A030	38.20	19.90	13.00	13.00	C3	KB
A031	30.32	23.66	9.00	9.00	C3	KB

C. Evaluasi Davies Boudin Index (DBI)

DBI adalah Pengukuran evaluasi yang digunakan pada penelitian adalah Davies Bouldin Index, pengukuran dilakukan untuk mengetahui apakah jumlah cluster sudah optimal untuk digunakan pada penelitian menggunakan K=3 jadi titik pusat cluster diambil sebanyak 3. Data awal yang digunakan sebelum proses perhitungan DBI yaitu menggunakan data pengelompokan pada iterasi terakhir, untuk contoh peneliti menggunakan data pada bulan Januari hasil iterasi sebanyak 2 iterasi. Adapun langkah-langkah perhitungan DBI sebagai berikut.

1. Sum of Square Within – Cluster (SSW)

Menghitung nilai SSW berdasarkan data hasil iterasi terakhir, hasil iterasi terakhir terdapat pada tabel hasil akhir diatas, selanjutnya menentukan titik pusat cluster baru dengan menghitung SSW. Nilai rata-rata jarak tiap cluster

Tabel 4. Titik Pusat Cluster DBI

Centroid	1	2	3	4	5	6	7	8	9
C1	78	81	74	73	73	78	74	88	81
C2	88	88	97	96	94	88	89	88	96
C3	88	82	85	91	83	91	88	88	93

Selanjutnya titik pusat diatas digunakan pada perhitungan nilai SSW setiap cluster menggunakan menggunakan perhitungan DBI, perhitungan SSW 1 untuk cluster 1 (C1). Hasil perhitungan seperti berikut ini

Tabel 5. Hasil Perhitungan SSW C1

Kode	1	2	3	4	5	6	7	8	9	Hsl
A010	78	79	80	82	78	77	78	92	88	11.59
A012	75	87	72	73	70	83	70	88	85	10.24
A013	82	84	84	84	73	81	73	88	85	11.70
A016	78	84	77	74	73	82	76	88	83	3.50
A017	80	82	89	81	76	79	83	88	86	16.15
A006	80	82	72	71	75	79	71	88	83	8.53
A008	80	79	76	74	73	78	74	88	80	6.32
A009	75	82	73	73	70	79	81	88	85	9.42
A011	78	86	71	71	73	78	70	88	78	11.29
A012	80	82	74	79	72	85	71	88	81	8.09

Untuk memperoleh nilai distance pada data pertama pada tabel yaitu A010 dilakukan pengukuran euclidiance menggunakan rumus [3] sebagai berikut.

$$d = \sqrt{((78 - 78)^2 + (83 - 81)^2 + (77 + 74)^2 + (76 - 78)^2 + (73 - 73)^2 + (80 - 78)^2 + (75 - 74)^2 + (88 - 88)^2 + (83 - 81)^2)} = 11.59$$

Perhitungan jarak tersebut menghasilkan nilai jarak data 1 yaitu 11.59 untuk jarak data ke 2 sampai dengan data ke 10, cluster 1 dilakukan perhitungan dengan cara yang sama data 1 kemudian hasil keseluruhan perhitungan jarak pada cluster 1 dihitung nilai SSW menggunakan rumus [4] sebagai berikut :

$$SSW_1 = \frac{1}{3} * (11.29 + 10.24 + 11.70 + 3.50 + 16.15 + 8.53 + 6.32 + 9.42 + 11.29 + 8.09) = 9.68$$

Selanjutnya menghitung nilai SSW 2 untuk cluster 2 sebagai menggunakan titik pusat cluster 2 (C2). Dengan menggunakan rumus yang sama pada cluster 1 (C1) Hasil perhitungan sebagai berikut

Tabel 6. Hasil Perhitungan SSW C2

Kode	1	2	3	4	5	6	7	8	9	Hsl
A002	65	76	68	60	65	73	66	75	70	33.28
A003	78	78	74	71	73	78	73	88	81	8.38
A004	78	78	69	70	70	71	70	75	73	22.91
A008	78	79	70	70	70	81	70	88	76	13.68
A011	76	83	73	70	70	75	70	88	86	11.09

A015	78	76	68	70	69	76	71	90	81	14.77
A019	78	79	70	70	68	74	70	88	78	14.67
A020	75	76	69	66	70	75	70	88	73	19.65
A021	80	79	72	70	68	74	70	88	78	14.00
A022	78	77	74	70	69	77	73	88	76	12.83
A024	78	81	73	71	73	77	74	90	79	8.75
A026	78	78	72	72	73	80	71	88	83	8.77
A028	78	76	73	72	74	74	70	88	79	12.50
A030	82	80	76	71	72	75	71	92	77	12.09

Hasil perhitungan SSW 2 diperoleh 14.81 dihitung dengan cara yang sama pada perhitungan SSW 1. Selanjutnya menghitung SSW 3 sebagai berikut.

Tabel 7. Hasil Perhitungan SSW C3

Kode	1	2	3	4	5	6	7	8	9	Hsl
A001	82	92	84	87	83	84	90	88	93	26.58
A005	79	82	74	78	71	83	83	88	82	9.86
A006	85	81	74	71	83	83	78	88	89	15.22
A007	88	88	97	96	94	88	89	88	96	42.22
A009	80	82	80	75	75	79	83	88	86	9.80
A014	88	82	85	91	83	91	88	88	93	29.48
A025	82	85	75	74	75	79	79	88	83	7.23

Hasil perhitungan SSW 3 diperoleh 20.06 dihitung dengan cara yang sama pada perhitungan SSW 1.

2. Sum of Square Between – Cluster (SSB)

Menghitung nilai SSB untuk setiap cluster menggunakan nilai centroid setiap cluster dan nilai dari SSW 1, 2 dan 3. Hasil SSB berupa Matriks sebagai berikut.

Tabel 8. Hasil Matrik SSB

SSB	1	2	3
1	0	13.10	16.5
2	13.10	0	28.37
3	16.5	28.37	0

Pada tabel di atas, sebagai contoh data pada SSB (1,2) di peroleh dari perhitungan perhitungan jarak pada titik pusat C1 dan C2 dengan menggunakan rumus [5], seperti berikut :

$$\begin{aligned}SSB_{1,2} &= \sqrt{(78-77)^2 + (83-78)^2 + (77-72)^2 + (76-70)^2 + (73-80)^2 + (80-76)^2 + (75-71)^2 + (88-87)^2 + (83-78)^2} \\ &= 13.10\end{aligned}$$

Untuk (C1,C3) dan (C2,C3) dilakukan dengan cara yang sama pada perhitungan (C1,C2).

3. Ratio (Rasio)

Perhitungan ratio menggunakan rumus [6] sebagai berikut

$$R = \frac{(9.68 + 14.81)}{13.10} = 4$$

Nilai Rmax R (1,1), (1,2), (1,3) yaitu 4 diambil nilai tertinggi pada baris R1.

4. Davies Bouldin Index (DBI)

Perhitungan DBI menggunakan rumus [7] sebagai berikut

$$DBI = \frac{1}{3} * (10.81 + 10.81.1.48 + 14) = 7.70$$

Semakin kecil nilai DBI yang diperoleh (non-negatif >= 0), maka semakin baik cluster, berdasarkan perhitungan DBI maka jumlah cluster K=3 sudah optimal.

IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan di SMP Negeri 9 Tarakan, implementasi metode k-medoids diterapkan untuk mengelompokkan siswa berdasarkan tingkat keaktifan mereka. Dalam studi ini, 31 siswa kelas VIII-I Tahun Ajaran Genap 2022/2023 dianalisis menggunakan nilai akademik sebagai parameter utama. Hasil akhir pengelompokan menghasilkan tiga cluster: 3 siswa masuk ke dalam cluster "Sangat Berminat," 4 siswa tergabung dalam cluster "Berminat," dan 24 siswa masuk ke dalam cluster "Kurang Berminat.". Metode k-medoids dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kesamaan, di mana setiap cluster memiliki satu objek sebagai pusat (medoid). Namun, dalam proses pemilihan medoids yang dilakukan secara acak, hasil pengelompokan dapat bervariasi setiap kali algoritma. Dengan demikian, meskipun metode k-medoids efektif dalam mengelompokkan siswa, hasil akhir dari pengelompokan sangat dipengaruhi oleh pemilihan awal medoids. Oleh karena

itu, hasil ini memberikan gambaran umum, namun mungkin tidak konsisten pada setiap pengulangan analisis dan berdasarkan perhitungan DBI menyatakan bahwa K=3 merupakan cluster optimal.

REFERENSI

- [1] I. Fatma, H. S. Tambunan, and F. Rizki, "Analisis Metode K-Medoids Cluster Dalam Mengelompokkan Siswa Yang Berprestasi," *Bull. Informatics Data Sci.*, vol. 1, no. 1, p. 14, 2022, doi: 10.61944/bids.v1i1.4.
- [2] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1464-1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [3] B. Riyanto, "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 562-568, 2019, doi: 10.30865/komik.v3i1.1659.
- [4] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166-173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [5] G. B. Kaligis and S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai," *IT-Explore J. Penerapan Teknol. Inf. dan Komun.*, vol. 1, no. 3, pp. 179-193, 2022, doi: 10.24246/itexplore.v1i3.2022.pp179-193.
- [6] E. T. Ena Tasia and M. Afdal, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 65-73, 2023, doi: 10.57152/ijirse.v3i1.523.
- [7] H. Pohan, M. Zarlis, E. Irawan, H. Okprana, and Y. Pranayama, "Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia," *JUKI J. Komput. dan Inform.*, vol. 3, no. 2, pp. 97-104, 2021, doi: 10.53842/juki.v3i2.69.