
Penerapan Clustering pada Sistem Pembagian Kelas Siswa Menggunakan Algoritma K-Means

Tri Martuti¹, Eviana Tjatur Putri², Roman Gusmana³

^{1,2,3}Program Studi Sistem Informasi, STMIK PPKIA Tarakanita Rahmawati, Tarakan, Kalimantan Utara
Email: ¹trimartuti356@gmail.com, ²eviana@ppkia.ac.id, ³roman@ppkia.ac.id

Abstrak

SMP Negeri 2 Malinau Utara merupakan sekolah di wilayah Kalimantan Utara tepatnya di Desa Putat Kecamatan Malinau Utara Kabupaten Malinau dan memiliki 127 siswa. Saat ini proses pembagian kelas belum efisien karena masih dilakukan secara acak tanpa pertimbangan apapun. Pembagian kelas dari proses clustering harus dapat memberikan hasil jumlah siswa yang seimbang pada setiap kelas. Pada penelitian ini penulis mengusulkan nilai yang digunakan untuk proses pengelompokan clustering adalah nilai mata pelajaran yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, dan Ilmu Pengetahuan Alam. Agar pembentukan kelas sesuai yang diharapkan, maka digunakan K-means untuk pengelompokan secara merata berdasarkan kriteria nilai yang sudah ditentukan. Metode K-Means Clustering adalah sebuah algoritma dalam analisis data yang digunakan untuk mengelompokkan sekumpulan data ke dalam beberapa kelompok berdasarkan kesamaan karakteristiknya. Pada proses clustering penelitian ini untuk menghitung jarak antara data dengan Centroid menggunakan perhitungan jarak Euclidean Distance. Penentuan centroid awal dan perhitungan jarak data dengan centroid awal dilakukan sampai anggota centroid tidak berubah, dalam penelitian ini iterasi rata-rata dilakukan sampai 3 kali. Penentuan centroid awal dilakukan dengan kombinasi sebanyak 1.081 kali diperoleh dari kombinasi 47 data untuk 2 kluster. Penelitian ini telah berhasil diterapkan untuk mengelompokkan siswa menggunakan metode K-Means Clustering dan memilih jumlah siswa yang seimbang antara satu kelas dengan kelas lain. Langkah selanjutnya adalah menggabungkan sebagian siswa pada setiap kluster dengan kluster lain, sehingga hasil akhir diperoleh setiap kelas memiliki tingkat kemampuan belajar yang berbeda-beda. Dengan perpaduan dua kluster dalam satu kelas maka diharapkan saat proses belajar dapat saling membantu satu siswa dengan yang lainnya.

Kata Kunci: K-Means, Clustering, Pembagian, Kelas, Siswa.

Application of K-Means Clustering for Student Class Division System

Abstract

SMP Negeri 2 Malinau Utara is a junior high school in Desa Putat, Malinau Utara, Malinau, Kalimantan Utara and has 127 students. Currently, the class division process is inefficient and random. On the other hand, the clustering process' class division must be able to provide each class a balanced number of students. This study proposes the grades of Indonesian and English languages, Mathematics, and Natural Sciences for the clustering. K-means is applied to evenly group students based on predetermined value criteria to achieve the expected class formation. K-Means Clustering is an algorithm in data analysis to group a set of data into several groups based on their similar characteristics. In the clustering process, the distance between the data and the Centroid was calculated using the Euclidean Distance. Initial centroid determination and data distance calculation with the initial centroid were performed until the centroid member remains unchanged. The initial centroid was determined using a combination of 1,081 times obtained from 47 data combinations for two clusters. This research has been successfully applied to classify students using the K-Means Clustering method and select a balanced number of students between one class and another. Next, combine some students in each cluster with other clusters, so that each class has different levels of learning ability. With the combination of two clusters in one class, it is expected that students can help each other during the learning process.

Keywords: K-Means, Clustering, Class Division, Students.

I. PENDAHULUAN

Data mining adalah bidang ilmu yang berfokus pada pengumpulan dan pengolahan data untuk mengekstrak informasi penting dari data tersebut. Salah satu teknik yang digunakan dalam data mining adalah clustering, di mana data dikelompokkan berdasarkan kemiripan. K-Means adalah metode clustering yang populer, yang bekerja dengan mencari rata-rata (mean) untuk menemukan pusat kluster, atau centroid. Algoritma K-Means mengelompokkan data yang memiliki jarak terdekat dengan centroid yang telah ditentukan, sehingga data dalam satu kluster sangat mirip, sementara data antar kluster berbeda.

SMP Negeri 2 Malinau Utara, yang berlokasi di Desa Putat, Kalimantan Utara, saat ini membagi siswa ke dalam kelas secara acak tanpa mempertimbangkan tingkat kemampuan mereka. Untuk meningkatkan efektivitas pembelajaran, sekolah ingin membagi siswa ke dalam kelas-kelas yang memiliki variasi tingkat kemampuan. Dengan demikian, diharapkan siswa dapat saling membantu dan mendukung dalam proses belajar mengajar.

Penelitian sebelumnya dilakukan oleh Ari Sulistiyawati dan Eko Supriyanto pada tahun 2021 menunjukkan bahwa Algoritma K-Means dapat digunakan dalam penentuan kelas unggulan dengan mengolah atribut berupa nilai mata pelajaran[1].

Penelitian lainnya dilakukan Nurahman, Agung Purwanto, dan Sigit Mulyanto pada tahun 2022 menunjukkan bahwa Algoritma K-Means juga dapat digunakan dalam klusterisasi sekolah berdasarkan fasilitas, pendidik, dan tenaga pendidik[2]. Hasil pengolahan dan analisis data tersebut berupa rekomendasi apakah perlu dilakukan pendampingan dan pengadaan kebutuhan sekolah oleh dinas pendidikan setempat.

Dalam upaya ini, metode K-Means dapat digunakan untuk mengelompokkan siswa berdasarkan nilai akademik mereka di mata pelajaran utama seperti Bahasa Indonesia, Bahasa Inggris, Matematika, dan Ilmu Pengetahuan Alam. Nilai-nilai ini diambil dari hasil seleksi ujian tertulis, dan pengelompokan dilakukan saat kenaikan kelas, menggunakan nilai akhir semester genap pada raport.

Dengan menerapkan metode K-Means, sekolah diharapkan dapat membagi siswa ke dalam kelas-kelas yang lebih seimbang dan sesuai dengan kebutuhan pendidikan. Ini tidak hanya mempermudah pekerjaan administrasi sekolah, tetapi juga diharapkan meningkatkan kualitas pembelajaran dengan menciptakan lingkungan kelas yang lebih dinamis dan mendukung perkembangan akademik siswa.

II. METODOLOGI PENELITIAN

A. Clustering

Clustering atau klusterisasi merupakan salah satu metode utama dalam data mining yang berfungsi untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kemiripan karakteristik antar objek data[3]. Dalam konteks data mining, clustering mempartisi satu set objek data ke dalam himpunan-himpunan bagian yang disebut dengan kluster, di mana objek-objek dalam satu kluster memiliki kesamaan yang tinggi, sementara berbeda signifikan

dengan objek di kluster lainnya. Proses ini tidak dilakukan secara manual, melainkan menggunakan algoritma clustering, yang memungkinkan pengelompokan data secara otomatis.

Clustering dalam data mining sangat penting karena mampu menemukan pola distribusi tersembunyi dalam dataset yang berguna untuk analisis data lebih lanjut. Teknik ini memungkinkan deteksi group atau kelompok yang tidak dikenal sebelumnya dalam data, sehingga sering kali disebut sebagai data segmentasi. Selain itu, clustering juga dapat digunakan untuk outlier detection, yaitu mendeteksi data yang menyimpang dari pola umum.

Kualitas hasil clustering sangat bergantung pada metode yang digunakan dan pengukuran kesamaan antar objek data. Kesamaan ini biasanya dihitung secara numerik, misalnya menggunakan rumus jarak Euclidean (Euclidean Distance), yang mengukur jarak antara titik-titik data dalam ruang multi-dimensi[4]. Dengan demikian, clustering tidak hanya membantu dalam pengelompokan data, tetapi juga berperan penting dalam mengidentifikasi dan memahami struktur yang mendasari data dalam proses data mining.

Secara keseluruhan, clustering adalah alat yang sangat berguna dalam data mining untuk mengungkap pola tersembunyi, mempartisi dataset ke dalam kelompok yang bermakna, dan mendeteksi anomali, yang semuanya penting dalam pengambilan keputusan berbasis data.

B. Algoritma K-Means

Algoritma K-Means merupakan salah satu algoritma clustering yang sangat umum dalam mengelompokkan data sesuai dengan kesamaan karakteristik. Kelompok data yang dihasilkan disebut sebagai kluster[5]. Pada Algoritma K-Means setiap objek pengamatan dimiliki oleh sebuah kelompok yang memiliki mean terdekat, dimana mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh algoritma. Pada penelitian ini process mining dilakukan dengan menggunakan Algoritma K-Means yaitu dengan penerapan persamaan 1.

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

Keterangan:

- d = Distance
- x = Data testing
- y = Data training
- xi = Data testing ke-i
- yi = Data training ke-i

Proses pada klusterisasi dilakukan dengan cara mengidentifikasi data yang akan dikluster terlebih dahulu. Setelah melakukan indentifikasi data selanjutnya dapat menerapkan persamaan 1 untuk menghitung nilai jarak antara data x ke data y. Dimana diketahui bahwa xi merupakan data testing ke-i sedangkan yi adalah data training ke-i. Selanjutnya mencari nilai titik centroid dari kluster ke K dapat dilakukan dengan persamaan 2.

$$\mu_k = \frac{1}{N_k} \sum_{i=R_1}^{N_k} x_i \quad \dots\dots\dots (2)$$

Keterangan:

- μ_k = Nilai titik centroid
- N_k = Banyak data klaster
- x_i = Nilai data ke-i

Berdasarkan persamaan 2 diatas dapat dijelaskan bahwa μ_k merupakan nilai titik centroid dari klaster ke K. Kemudian simbol N_k digunakan untuk menunjukkan nilai banyaknya data pada klaster ke K. Selanjutnya simbol x_i digunakan untuk menunjukkan nilai data ke i pada klaster ke K. Dataset yang telah final akan dilakukan pengolahan data dengan menggunakan Algoritma K- Means. Jumlah nilai K akan ditentukan untuk membentuk jumlah klaster, kemudian pencarian nilai titik centroid, pengelompokan data berdasarkan jumlah klaster yang telah ditentukan. Jika titik centroid berubah maka akan dilakukan lagi pemilihan titik centroid hingga diketahui tidak ada perubahan pada titik centroid. Pada proses Clustering menggunakan K-Means jumlah penentuan centroid awal dilakukan sebanyak hasil perhitungan kombinasi dengan persamaan 3.

$$C(n, r) = \frac{n!}{r!(n-r)!} \quad \dots\dots\dots (3)$$

Keterangan:

- C = Kombinasi
- n = jumlah total objek (data)
- r = jumlah objek yang dipilih

Hasil klaster dalam algoritma K-Means berhenti beriterasi ketika salah satu dari kondisi berikut tercapai:

1. Centroid Tidak Berubah.
Iterasi berhenti jika posisi centroid tidak berubah antara dua iterasi berturut-turut, menandakan stabilitas klaster.
2. Jumlah Iterasi Maksimum
Iterasi berhenti jika jumlah iterasi maksimum tercapai, meskipun centroid mungkin masih berubah.
3. Konvergensi Fungsi Objektif
Iterasi berhenti jika perubahan dalam fungsi objektif sangat kecil, menunjukkan algoritma telah mencapai konvergensi.

Iterasi berhenti ketika algoritma tidak lagi menghasilkan perbaikan yang signifikan dalam pengelompokan data atau setelah mencapai batasan iterasi yang telah ditetapkan.

C. SMP Negeri 2 Malinau Utara

SMP Negeri 2 Malinau Utara, yang didirikan pada 2 April 2005, merupakan sekolah yang terletak di wilayah Kalimantan Utara, tepatnya di Desa Putat, Kecamatan Malinau Utara, Kabupaten Malinau. Hingga saat ini, sekolah tersebut masih beroperasi dan memiliki 127 siswa yang terbagi dalam 6 rombongan belajar. Setiap tingkat kelas memiliki dua kelas, yaitu Kelas VII A dan VII B, Kelas VIII A dan VIII B, serta Kelas IX A dan IX B.

III. HASIL DAN PEMBAHASAN

A. Studi Kasus

Hasil pengumpulan data di SMP Negeri 2 Malinau Utara menghasilkan informasi data nilai siswa kelas VII. Data nilai mata pelajaran yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Tabel Nilai

ID	Nama	Bhs. Idn.	Bhs. Ing.	Mtk.	IPA
1	Agrea Priskila	80	77	73	75
2	Alpius	80	77	72	74
3	Amelia Letesia	75	77	74	74
4	Arel Oktavianus	80	79	74	73
5	Arga	74	77	70	74
6	Ariska	75	70	75	77
7	Charly Okta Fiano	75	77	71	72
8	Crisfin Efendi	70	77	74	72
9	Cristiano Renaldo	75	78	72	73
10	Delia gonda Marliana Padu	87	79	80	85
11	Diki Saputra	75	76	71	70
12	Diori Faldo	73	77	70	74
13	Dominkus Sapio	73	77	70	73
14	Ferri	75	77	71	74
15	Fessy Alini	85	78	74	82
16	Fran Kifli Seven	80	76	70	73
17	Gali Emanuel	74	76	71	74
18	Gita Reston Teguh	78	75	72	75
19	Gressia	79	78	75	75
20	Jeshinta Laura	60	69	50	65
21	Jihan	80	78	72	76
22	Jose Sebastian	70	75	70	74
23	Julcristiano	80	78	72	74
24	Julita	79	78	71	74
25	Junikafreya Charlina	79	77	76	77
26	Juven Valentino	70	75	70	73
27	Kornalius	80	77	70	74
28	Lili Suriani	77	77	70	75
29	Majalia Rosa	79	78	71	74
30	Marsel Derossiwoda	77	76	71	73
31	Masrel	75	76	71	72
32	Melda	75	76	73	75
33	Muh. Fahrie Sam Arafah	70	78	78	75
34	Noviani	79	77	72	75

35	Okta Fanli	75	77	70	72
36	Olivia	80	78	72	74
37	Ongki	77	78	70	74
38	Rahmat Rifai Sangadji	81	77	72	73
39	Ramdani	77	75	72	74
40	Renaldo	70	75	71	73
41	Reno Sabado	74	76	73	74
42	Syaparudin	77	75	72	73
43	Tiara Lusya Tulung	79	75	72	75
44	Varel	70	76	70	74
45	Veronika Adelia	79	77	75	75
46	Yuwidia Agustin	80	78	73	75
47	Zuliansyah	77	75	71	74

Selanjutnya adalah membagi siswa tersebut kedalam kluster berdasarkan nilai mata pelajaran dengan menggunakan Algoritma K-Means dengan langkah-langkah sebagai berikut:

1. Tentukan jumlah kluster yang ingin dibentuk. Pada penelitian ini, jumlah kluster yang akan dibentuk adalah sebanyak 2 (dua) kluster.
2. Menentukan centroid (titik pusat kluster) awal dengan mengambil data secara acak/random, diasumsikan:
 - a. Diambil data ke-1 sebagai pusat kluster ke-1 (C1) : {80, 77, 73, 75}
 - b. Diambil data ke-3 sebagai pusat kluster ke-2 (C2) : {75, 77, 74, 74}
3. Menghitung jarak antara data dengan Centroid awal menggunakan Euclidean Distance, dan mendapat jarak matriks sebagai berikut :

Iterasi 1 :

- a. Perhitungan data C1 terhadap data kluster

$$D1 = \sqrt{(80 - 80)^2 + (77 - 77)^2 + (73 - 73)^2 + (75 - 75)^2} = 0,00$$

$$D2 = \sqrt{(80 - 80)^2 + (77 - 77)^2 + (72 - 73)^2 + (74 - 75)^2} = 1,41$$

$$D3 = \sqrt{(75 - 80)^2 + (77 - 77)^2 + (74 - 73)^2 + (74 - 75)^2} = 5,20$$

$$D4 = \sqrt{(80 - 80)^2 + (79 - 77)^2 + (74 - 73)^2 + (73 - 75)^2} = 3,00$$

$$D5 = \sqrt{(74 - 80)^2 + (77 - 77)^2 + (70 - 73)^2 + (74 - 75)^2} = 6,78$$

Dilanjutkan perhitungan jarak data ke-6 sampai dengan data ke-47 dilakukan dengan cara yang sama.

- b. Perhitungan data C2 terhadap data kluster

$$D1 = \sqrt{(80 - 75)^2 + (77 - 77)^2 + (73 - 74)^2 + (75 - 74)^2} = 5,20$$

$$D2 = \sqrt{(80 - 75)^2 + (77 - 77)^2 + (72 - 74)^2 + (74 - 74)^2} = 5,39$$

$$D3 = \sqrt{(75 - 75)^2 + (77 - 77)^2 + (74 - 74)^2 + (74 - 74)^2} = 0,00$$

$$D4 = \sqrt{(80 - 75)^2 + (79 - 77)^2 + (74 - 74)^2 + (73 - 74)^2} = 5,48$$

$$D5 = \sqrt{(74 - 75)^2 + (77 - 77)^2 + (70 - 74)^2 + (74 - 74)^2} = 4,12$$

Dilanjutkan perhitungan jarak data ke-6 sampai dengan data ke-47 dilakukan dengan cara yang sama.

Perhitungan data C1 dan C2 terhadap pusat centroid, menghasilkan perhitungan pada Tabel 2.

Tabel 2. Hasil Perhitungan Iterasi ke-1

ID	C1	C2	ID	C1	C2
1	0,00	5,20	25	3,74	5,39
2	1,41	5,39	26	10,82	6,78
3	5,20	0,00	27	3,16	6,40
4	3,00	5,48	28	4,24	4,58
5	6,78	4,12	29	2,65	5,10
6	9,06	7,68	30	4,24	3,87
7	6,16	3,61	31	6,24	3,74
8	10,49	5,39	32	5,10	1,73
9	5,57	2,45	33	11,22	6,56
10	14,21	17,46	34	1,41	4,58
11	7,42	5,10	35	6,56	4,47
12	7,68	4,47	36	1,73	5,48
13	7,87	5,58	37	4,47	4,58
14	5,48	3,00	38	2,45	6,40
15	8,72	12,85	39	3,87	3,46
16	3,74	6,56	40	10,58	6,24
17	6,48	3,32	41	6,16	1,73
18	3,00	4,24	42	4,24	3,61
19	2,45	4,36	43	2,45	5,00
20	33,06	30,76	44	10,54	6,48
21	1,73	5,83	45	2,24	4,24
22	10,68	6,71	46	1,00	5,29
23	1,73	5,48	47	4,24	4,12
24	2,65	5,10			

4. Pengelompokan Data

Dilakukan pengelompokan data untuk mendapatkan hasil matriks, hasil pengelompokan terlihat pada Tabel 3.

Tabel 3. Pengelompokan

ID	C1	C2	ID	C1	C2
1	1		25	1	
2	1		26		2
3		2	27	1	
4	1		28	1	
5		2	29	1	
6		2	30		2
7		2	31		2

8	2	32	2
9	2	33	2
10	1	34	1
11	2	35	2
12	2	36	1
13	2	37	1
14	2	38	1
15	1	39	2
16	1	40	2
17	2	41	2
18	1	42	2
19	1	43	1
20	2	44	2
21	1	45	1
22	2	46	1
23	1	47	2
24	1		

Angka 1 merupakan anggota C1 dan angka 2 merupakan anggota C2.

5. Penentuan Pusat Centroid Baru

Proses ini bertujuan untuk memperbaiki akurasi pengelompokan, dengan membuat centroid lebih dekat ke pusat sebenarnya dari data dalam kluster, sehingga iterasi selanjutnya bisa menghasilkan pengelompokan yang lebih optimal.

a. Perhitungan centroid baru C1 :

Bahasa Indonesia

$$\frac{80 + 80 + 80 + 87 + 85 + 80 + 78 + 79 + 80 + 80 + 79 + 79 + 80 + 77 + 79 + 79 + 80 + 77 + 81 + 79 + 79 + 80}{22}$$

$$= 79.91$$

Bahasa Inggris

$$\frac{77 + 77 + 79 + 79 + 78 + 76 + 75 + 78 + 78 + 78 + 78 + 77 + 77 + 77 + 78 + 77 + 78 + 78 + 77 + 75 + 77 + 78}{22}$$

$$= 77.36$$

Matematika

$$\frac{73 + 72 + 74 + 80 + 74 + 70 + 72 + 75 + 72 + 72 + 71 + 76 + 70 + 70 + 71 + 72 + 72 + 70 + 72 + 72 + 75 + 73}{22}$$

$$= 72.64$$

IPA

$$\frac{75 + 74 + 73 + 85 + 82 + 73 + 75 + 75 + 76 + 74 + 74 + 77 + 74 + 75 + 74 + 75 + 74 + 74 + 73 + 75 + 75 + 75}{22}$$

$$= 75.32$$

Sehingga diperoleh centroid baru C1 = {79.91, 77.36, 72.64, 75.32}

b. Perhitungan centroid baru C2 :

Bahasa Indonesia

$$\frac{75 + 74 + 75 + 75 + 70 + 75 + 75 + 73 + 73 + 75 + 74 + 60 + 70 + 70 + 77 + 75 + 75 + 70 + 75 + 77 + 70 + 74 + 77 + 70 + 7}{25}$$

$$= 73.24$$

Bahasa Inggris

$$\frac{77 + 77 + 70 + 77 + 77 + 78 + 76 + 77 + 77 + 77 + 76 + 69 + 75 + 75 + 76 + 76 + 76 + 78 + 77 + 75 + 75 + 76 + 75 + 76 + 75}{25}$$

$$= 75.72$$

Matematika

$$\frac{74 + 70 + 75 + 71 + 74 + 72 + 71 + 70 + 70 + 71 + 71 + 50 + 70 + 70 + 71 + 71 + 73 + 78 + 70 + 72 + 71 + 73 + 72 + 70 + 71}{25}$$

$$= 70.84$$

IPA

$$\frac{74 + 74 + 77 + 72 + 72 + 73 + 70 + 74 + 73 + 74 + 74 + 65 + 74 + 73 + 73 + 72 + 75 + 75 + 72 + 74 + 73 + 74 + 73 + 74 + 74}{25}$$

$$= 73.12$$

Sehingga diperoleh centroid baru C2 = {73.24, 75.72, 70.84, 73.12}

Iterasi 2 :

a. Telah ditentukan titik pusat centroid baru, yaitu :

$$C1 = \{79.91, 77.36, 72.64, 75.32\}$$

$$C2 = \{73.24, 75.72, 70.84, 73.12\}$$

b. Perhitungan jarak pusat kluster

1. Perhitungan data C1 terhadap data kluster

$$D1 = \sqrt{(80 - 79.91)^2 + (77 - 77.36)^2 + (73 - 72.64)^2 + (75 - 75.32)^2} = 0,61$$

$$D2 = \sqrt{(80 - 79.91)^2 + (77 - 77.36)^2 + (72 - 72.64)^2 + (74 - 75.32)^2} = 1,51$$

$$D3 = \sqrt{(75 - 79.91)^2 + (77 - 77.36)^2 + (74 - 72.64)^2 + (74 - 75.32)^2} = 5,28$$

$$D4 = \sqrt{(80 - 79.91)^2 + (79 - 77.36)^2 + (74 - 72.64)^2 + (73 - 75.32)^2} = 3,15$$

$$D5 = \sqrt{(74 - 79.91)^2 + (77 - 77.36)^2 + (70 - 72.64)^2 + (74 - 75.32)^2} = 6,61$$

Dilanjutkan perhitungan jarak data ke-6 sampai dengan data ke-47 dilakukan dengan cara yang sama.

2. Perhitungan data C2 terhadap data kluster

$$D1 = \sqrt{(80 - 73,24)^2 + (77 - 75,72)^2 + (73 - 70,84)^2 + (75 - 73,12)^2} = 7,45$$

$$D2 = \sqrt{(80 - 73,24)^2 + (77 - 75,72)^2 + (72 - 70,84)^2 + (74 - 73,12)^2} = 7,03$$

$$D3 = \sqrt{(75 - 73,24)^2 + (77 - 75,72)^2 + (74 - 70,84)^2 + (74 - 73,12)^2} = 3,94$$

$$D4 = \sqrt{(80 - 73,24)^2 + (79 - 75,72)^2 + (74 - 70,84)^2 + (73 - 73,12)^2} = 8,15$$

$$D5 = \sqrt{(74 - 73,24)^2 + (77 - 75,72)^2 + (70 - 70,84)^2 + (74 - 73,12)^2} = 1,92$$

Dilanjutkan perhitungan jarak data ke-6 sampai dengan data ke-47 dilakukan dengan cara yang sama.

Perhitungan data C1 dan C2 terhadap pusat centroid, menghasilkan perhitungan pada Tabel 4.

Tabel 4. Hasil Perhitungan Iterasi ke-2

ID	C1	C2	ID	C1	C2
1	0,61	7,45	25	3,89	8,75
2	1,51	7,03	26	10,78	3,43
3	5,28	3,94	27	2,97	6,99
4	3,15	8,15	28	3,96	4,47
5	6,61	1,92	29	2,38	6,26
6	9,31	8,26	30	4,29	3,78
7	6,16	2,45	31	6,30	2,11
8	10,54	4,83	32	5,12	3,37
9	5,50	3,11	33	11,29	8,40
10	14,17	20,62	34	1,21	6,30
11	7,54	3,60	35	6,50	2,59
12	7,52	1,78	36	1,60	7,28
13	7,76	1,55	37	4,19	4,56
14	5,35	2,35	38	2,66	7,95
15	8,53	15,24	39	4,02	4,10
16	3,77	6,82	40	10,75	3,33
17	6,42	1,21	41	6,22	2,47
18	3,12	5,30	42	4,45	4,00
19	2,63	7,70	43	2,63	6,21
20	32,94	26,85	44	10,43	3,47
21	1,13	7,78	45	2,58	7,46
22	10,60	3,53	46	0,80	7,69
23	1,60	7,28	47	4,30	3,93
24	2,38	6,26			

c. Pengelompokan Data

Dilakukan pengelompokan data untuk mendapatkan hasil matriks, hasil pengelompokan terlihat pada Tabel 5.

Tabel 5. Tabel Pengelompokan

ID	C1	C2	ID	C1	C2
1	1		25	1	
2	1		26		2
3		2	27	1	
4	1		28	1	
5		2	29	1	
6		2	30		2
7		2	31		2
8		2	32		2
9		2	33		2
10	1		34	1	
11		2	35		2
12		2	36	1	
13		2	37	1	
14		2	38	1	
15	1		39	1	
16	1		40		2
17		2	41		2
18	1		42		2
19	1		43	1	
20		2	44		2
21	1		45	1	
22		2	46	1	
23	1		47		2
24	1				

Angka 1 merupakan anggota C1 dan angka 2 merupakan anggota C2.

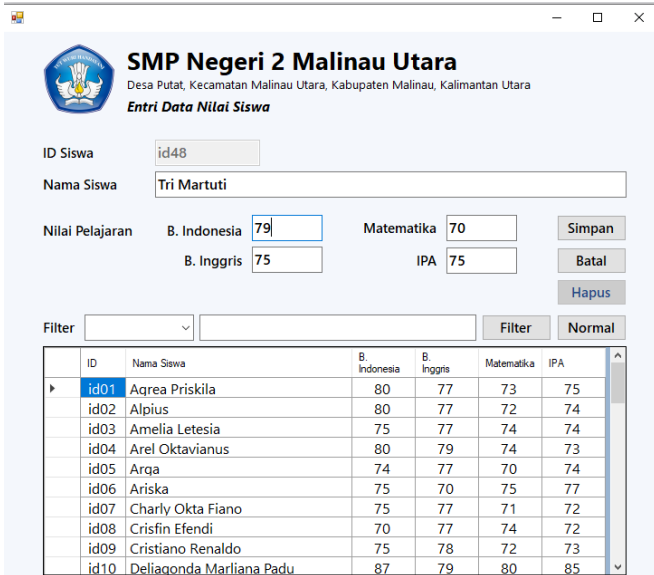
Lakukan pengulangan hingga kondisi perhentian terpenuhi. Diketahui iterasi berhenti pada iterasi-3. Proses perhitungan berhenti, maka data akan dikelompokkan berdasarkan masing-masing kluster, seperti pada tabel 3.6 sebagai berikut :

Klaster 1 anggota : Data { 1, 2, 4, 10, 15, 16, 18, 19, 21, 23, 24, 25, 27, 28, 29,34, 36, 37, 38, 39, 43, 45,46}

Klaster 2 anggota : Data { 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 17, 20, 22, 26, 30, 31, 32, 33, 35, 40, 41, 42, 44, 47}

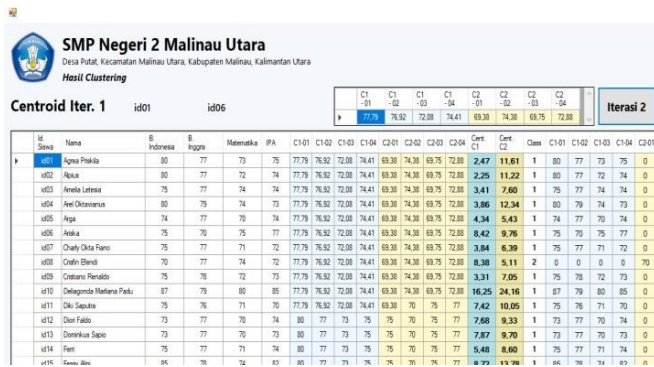
B. Uji Coba

Untuk menunjukkan bagaimana Algoritma K-Means dapat bekerja, maka penulis membangun sebuah program berbasis desktop.

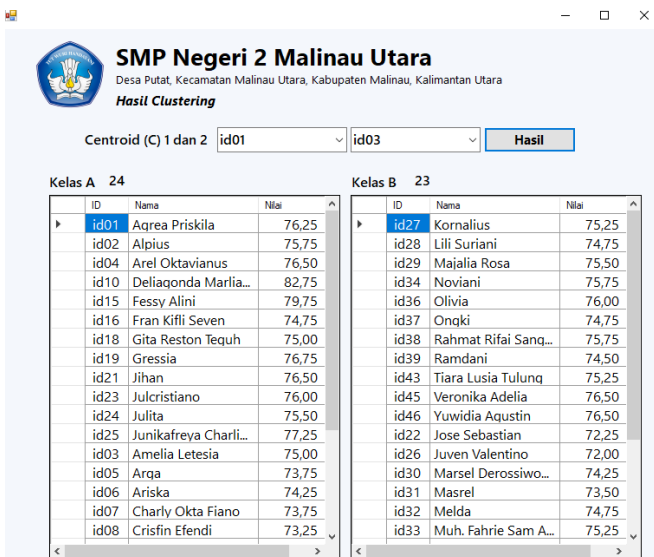


Gambar 1. Halaman Entri Data Nilai

Halaman entri data nilai digunakan oleh pengguna untuk mengolah data nilai siswa. Selanjutnya, data tersebut diolah untuk diketahui hasil klasternya.



Gambar 2. Proses Perhitungan Algoritma K-Means



Gambar 3. Hasil Pembagian Klaster

Gambar 2 menunjukkan proses perhitungan jarak antara titik centroid dan data nilai siswa, di mana setiap data nilai diuji dan dibandingkan dengan jarak ke centroid lainnya. Proses ini akan selesai setelah semua klaster telah dihitung dan berhenti ketika iterasi mencapai batas yang telah ditentukan.

Proses iterasi untuk mendapatkan kombinasi 2 centroid dari 47 data siswa memerlukan 1.081 iterasi. Berdasarkan uji coba yang dilakukan, waktu yang dibutuhkan untuk menyelesaikan iterasi tersebut adalah sekitar 4 jam, menggunakan perangkat keras dengan spesifikasi Processor Intel(R) Core(TM) i5 dan Memori RAM 4 GB. Iterasi ini diperlukan untuk menemukan pembagian klaster yang optimal, di mana setiap iterasi berusaha meminimalkan jarak antara data siswa dan centroid yang dihasilkan.

Gambar 3 menampilkan hasil keseluruhan perhitungan pada form K-Means, yang menunjukkan jumlah siswa serta nama-nama siswa yang masuk ke dalam kelas A dan kelas B. Pengguna juga dapat mengatur titik centroid yang diambil, sehingga hasil pengelompokan dapat disesuaikan sesuai kebutuhan.

Berdasarkan uji coba yang telah dilakukan, metode K-Means Clustering terbukti dapat diterapkan dengan baik dalam kasus pembagian kelompok siswa berdasarkan nilai dari empat mata pelajaran, yaitu Bahasa Indonesia, Bahasa Inggris, Matematika, dan Ilmu Pengetahuan Alam.

Hasil clustering ini menunjukkan bahwa pengelompokan siswa tanpa menggunakan nilai rata-rata dapat memberikan alternatif pembagian siswa dengan pendekatan berbeda, yakni berdasarkan kemiripan kemampuan. Sesuai dengan tujuan Clustering, yang membagi data menjadi kelompok-kelompok (kluster) di mana data dalam satu kelompok memiliki karakteristik yang mirip dan berbeda dari kelompok lainnya, hasil pengelompokan ini menunjukkan bahwa siswa dengan tingkat pemahaman yang serupa (perbedaan nilai yang kecil) dikelompokkan bersama.

Namun, untuk memenuhi tujuan penelitian yang menginginkan adanya variasi karakteristik dalam setiap kelompok agar siswa dapat saling membantu, dilakukan pengelompokan ulang dengan mengambil sebagian siswa dari klaster A dan sebagian dari klaster B untuk membentuk kelompok atau kelas baru. Dalam kelas atau kelompok baru ini, terdapat tingkatan kemampuan yang berbeda, sehingga dalam proses pembelajaran, siswa dapat saling membantu.

IV. KESIMPULAN

Dari hasil penelitian dan pengujian program ini, dapat disimpulkan bahwa metode K-Means Clustering efektif dalam mengelompokkan siswa berdasarkan jarak dengan centroid, sehingga setiap kelas memiliki tingkat kemampuan belajar yang berbeda dan memungkinkan siswa saling membantu dalam proses belajar. Metode ini juga direkomendasikan untuk diterapkan di SMP Negeri 2 Malinau Utara, khususnya untuk digunakan dalam pembagian kelas setelah kenaikan kelas, guna mencapai pembagian yang lebih optimal dan efektif.

REFERENSI

- [1] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," vol. 15, no. 2.
- [2] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 337–350, Mar. 2022, doi: 10.30812/matrik.v21i2.1411.
- [3] Siti Hardianti, Sinawati, and Dikky Praseptian M., "Implementasi Clustering dengan Metode Minimum Spanning Tree untuk Pengelompokan Siswa berdasarkan Nilai Hasil Studi," *Journal of Big Data Analytic and Artificial Intelligence (JBIDAI)*, vol. 4, no. Vol. 4 No. 1 (2018): JBIDAI, pp. 23–28, 2018.
- [4] R. Nainggolan and G. Lumbantoruan, "METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi OPTIMASI PERFORMA CLUSTER K-MEANS MENGGUNAKAN SUM OF SQUARED ERROR (SSE) 1," vol. 2, no. 2, 2018, doi: 10.46880/jmika.Vol2No2.pp103-108.
- [5] C. Selvi, D. Sembiring, L. Hanum, and S. Parsaoran Tamba, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022.