

BAB III

ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Sistem

Analisa mengenai sistem pengelompokan siswa SD Al Islam untuk mengikuti bimbingan belajar menggunakan metode K-Means adalah sistem yang digunakan untuk mengelompokkan siswa mana yang harus diberikan bimbingan khusus untuk persiapan menghadapi ujian sekolah. Data siswa tersebut didapatkan dari nilai mata pelajaran yang diujikan untuk ujian sekolah tahun ajaran 2022-2023 di SD Al Islam.

Sistem pembagian kelompok bimbingan belajar yang ada saat ini sebenarnya hanya berdasarkan daftar absen kelas yang ada. Kenyataannya tingkat belajar dan kemampuan siswa itu berbeda-beda, sehingga menyebabkan siswa yang kesulitan dalam memahami ilmu mempunyai potensi kegagalan yang tinggi. Dengan adanya sistem ini diharapkan pembagian kelompok belajar yang dihasilkan dapat lebih optimal dan siswa mendapat pengajaran materi yang lebih sesuai dengan tingkat pemahaman mereka.

Pengelompokan siswa dilakukan dengan menggunakan algoritma K-Means untuk menghasilkan cluster-cluster. Cluster-cluster tersebut ialah pengelompokan siswa berdasarkan nilai nilai mata pelajaran.

3.1.1 Representasi Data

Dalam melakukan pengujian sistem pengelompokan menggunakan data dari sekolah tahun ajaran 2022-2023. Penelitian ini menggunakan data nilai tugas, nilai sikap, nilai UTS dan nilai UAS dari siswa.

Tabel 3.1 Tabel Atribut Data.

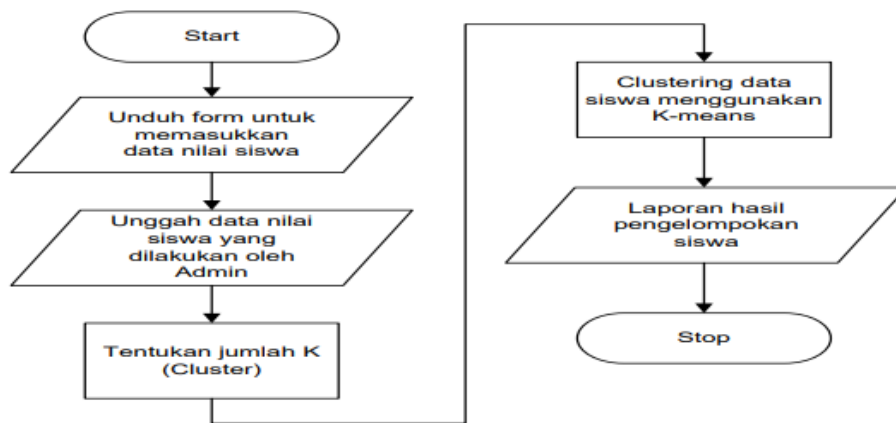
No	Atribut
1	Nilai Tugas
2	Nilai Sikap
3	Nilai UTS
4	Nilai UAS

3.2 Perancangan Sistem

Setelah analisis sistem selesai, dilanjutkan dengan perancangan sistem. Perancangan ini tersusun atas bentuk flowchart, diagram konteks, diagram berjenjang, data flow diagram (DFD).

3.2.1 Flowchart

Berikut adalah *flowchart* sistem pengelompokan kelas bimbingan belajar :



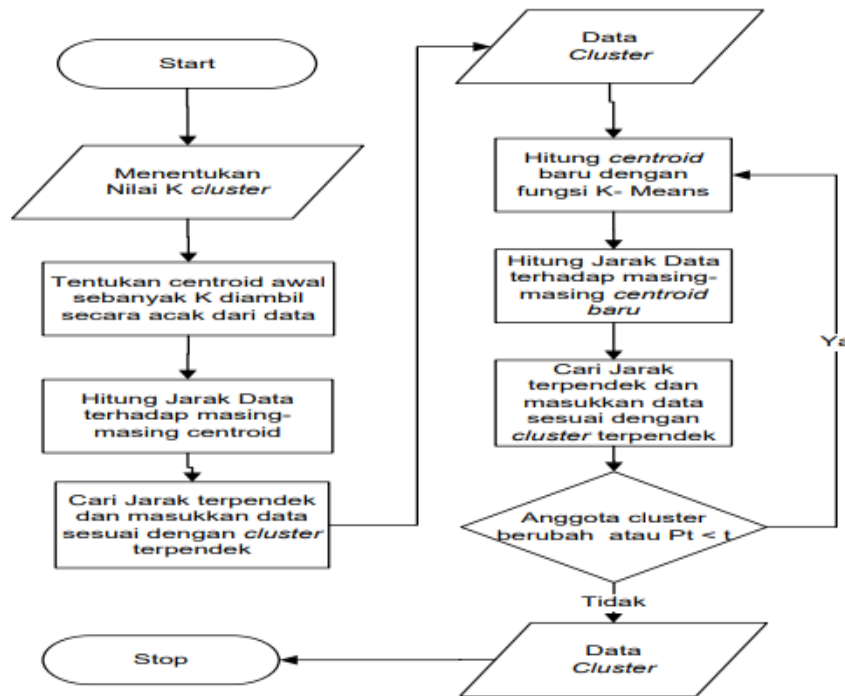
Gambar 3.1. Flowchart Sistem

Berikut adalah penjelasan dari gambar 3.1 yang merupakan flowchart dari sistem :

1. User mengunduh form isian nilai siswa dari halaman download form dan memasukkan data nilai siswa.
2. User mengunggah form data nilai siswa kedalam sistem.
3. Menentukan jumlah cluster (K).
4. Sistem ini akan memulai pengelompokan menggunakan metode K-Means. Sistem ini menampilkan hasil akhir dari pengelompokan, berupa laporan nama-nama siswa yang masuk kedalam kelompok.

3.2.1.1 Flowchart Sistem Menggunakan Metode K-Means

Berikut ini adalah *flowchart* dari sistem pengelompokan kelas bimbingan belajar yang menggunakan metode k-means



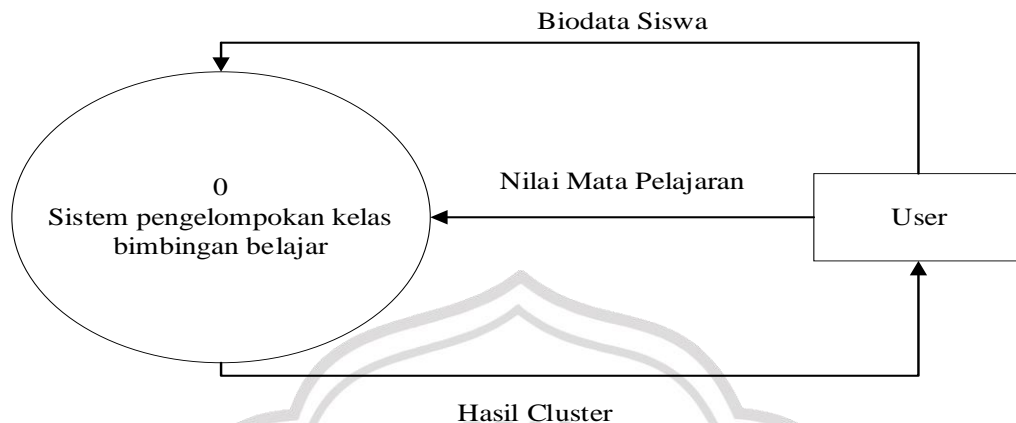
Gambar 3.2. Flowchart Sistem Menggunakan Metode K-Means

Berikut adalah penjelasan dari gambar 3.2 yang merupakan flowchart dari sistem menggunakan metode k-means :

1. Menentukan nilai K cluster, lalu menentukan nilai centroid awal secara random, kemudian menentukan jarak terdekat terhadap centroid pada setiap data dalam situasi ini menggunakan rumus Euclidian Distance.
2. Mencari jarak terdekat data dan memasukkan data kedalam cluster sesuai dengan centroid terdekat.
3. Melihat anggota cluster apakah anggota cluster berubah atau jika ada perubahan nilai fungsi objektif diatas nilai treshold. Jika terjadi perubahan mengulangi langkah menghitung centroid baru dengan fungsi K-Means sehingga anggota suatu cluster tidak berubah.

3.2.2 Diagram Konteks

Berikut ini adalah diagram konteks sistem pengelompokan kelas bimbingan belajar :



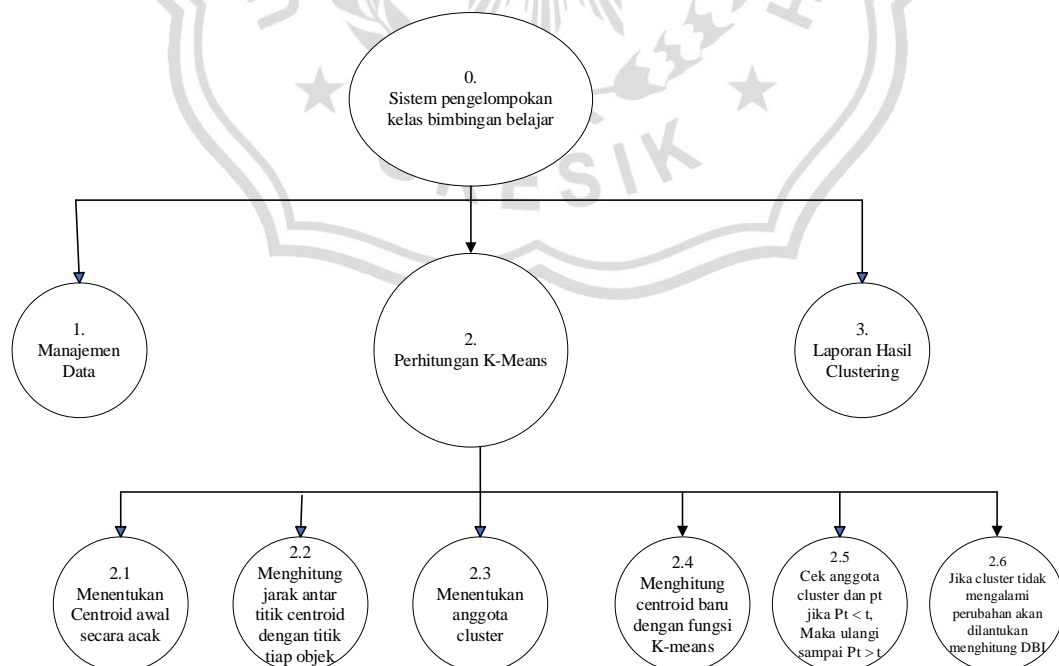
Gambar 3.3. Diagram Konteks Sistem

Keterangan Gambar 3.3

User menginputkan nilai mata pelajaran dan dapat melihat biodata siswa yang akan diclusterisasi.

3.2.2.1 Diagram Berjenjang

Berikut ini adalah diagram berjenjang sistem pengelompokan kelas bimbingan belajar :



Gambar 3.4. Diagram Berjenjang Sistem

Keterangan :

1. Top level : Sistem Pengelompokan Kelas Bimbingan Belajar Menggunakan Metode K-Means
2. Level 0: Merupakan sub proses dari Sistem Pengelompokan Kelas Bimbingan Belajar Menggunakan Metode K-Means yang sudah dibagi menjadi beberapa sub proses antara lain :
 1. Manajemen Data.
 2. Perhitungan K-Means.
 3. Laporan Hasil Clustering.
3. Level 1 : Merupakan sub proses dari Level 0 Sistem Pengelompokan Kelas Bimbingan Belajar Menggunakan metode K-Means yang sudah dibagi menjadi beberapa sub proses antara lain :

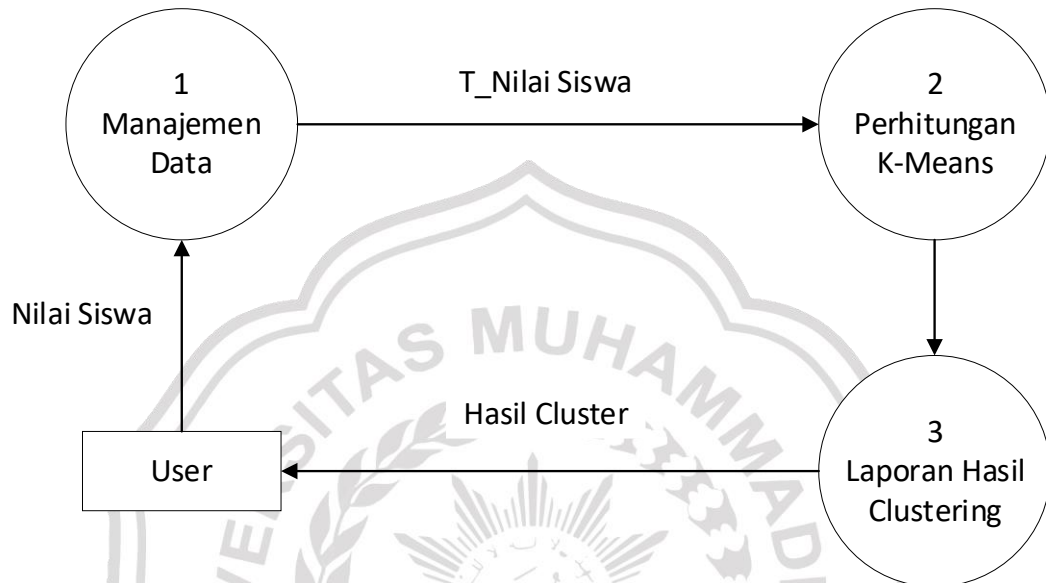
Pengelompokan Kelas Bimbingan Belajar Menggunakan metode K-Means meliputi :

 - 2.1. Menentukan Centroid awal secara acak.
 - 2.2. Menghitung Jarak antar titik centroid dengan titik tiap cluster.
 - 2.3. Menentukan anggota cluster.
 - 2.4. Menghitung centroid baru dengan fungsi K-Means.
 - 2.5. Cek anggota cluster dan p_t jika $p_t < t$, Maka ulangi sampai $p_t > t$.
 - 2.6. Jika cluster tidak mengalami perubahan akan dihitung DBI

3.2.3 Data Flow Diagram (DFD)

3.2.3.1 DFD level 0

Berikut ini adalah DFD level 0 sistem pengelompokan kelas bimbingan belajar :

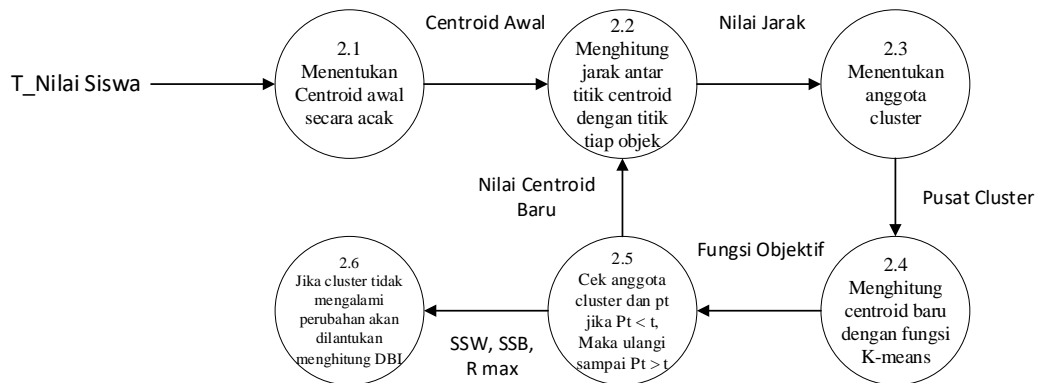


Gambar 3.5. DFD Level 0

Keterangan Gambar 3.5 :

1. Proses 1 manajemen data adalah nilai siswa dari user diproses dan menghasilkan hasil nilai siswa yang disimpan ditabel Nilai siswa.
2. Proses 2 perhitungan pengelompokan kelas bimbingan belajar dengan K-Means yaitu melakukan perhitungan pada tabel nilai.
3. Proses 3 pembuatan hasil laporan pengelompokan bimbingan belajar dengan menentukan cluster terbaik dan terendah kemudian menghasilkan hasil clustering siswa.

3.2.3.2 DFD level 1



Gambar 3.6 DFD Level 1

Keterangan Gambar 3.6 :

- 2.1 User memasukkan nilai k yaitu jumlah cluster yang diinginkan, lalu sistem akan memilih centroid secara acak dari data nilai pada data siswa sebanyak k dan akan disimpan di tabel centroid awal.
- 2.2 Selanjutnya dilakukan proses menghitung jarak tiap titik centroid dengan titik tiap objek yang didapat dari data nilai pada data siswa.
- 2.3 Setelah menghitung jarak, nilai jarak akan dikelompokkan sesuai jarak terpendek untuk menentukan anggota cluster.
- 2.4 Pada proses berikutnya menghitung centroid baru dengan fungsi K-means dan menghasilkan centroid baru. Kemudian kembali perhitungan jarak dengan menggunakan centroid baru, nilai jaraknya dikelompokkan sesuai jarak terpendek dan disimpan di tabel data laporan cluster
- 2.5 Proses selanjutnya adalah Memeriksa anggota cluster dan perubahan fungsi objektif, jika posisi cluster berubah atau perubahan fungsi objektif $> t$ maka proses diulangi hingga posisi cluster tidak berubah atau perubahan fungsi objektif $< t$.
- 2.6 Proses selanjutnya adalah menghitung nilai DBI dari setiap cluster jika cluster tidak mengalami perubahan

3.3 Perancangan Basis Data

Dalam proses ini dijelaskan tentang kebutuhan database atau tabel-tabel yang digunakan untuk keperluan sistem yang akan dibuat.

3.3.1 Tabel User

Tabel klien ini digunakan untuk menyimpan user_id, nama, username, password, dan status dengan user_id sebagai primary key yang menunjukkan urutan dari semua user user yang ada di database.

Tabel 3.2 Tabel user

No	Nama	Tipe	Panjang	Keterangan
1	User_id	Int	5	
2	Nama	Varchar	25	
3	Username	Varchar	25	
4	Password	Varchar	25	
5	Status	Int	25	

Variabel pada tabel user meliputi :

- a. Field User_id merupakan nama user_id.
- b. Field nama merupakan nama user pada saat login.
- c. Field username merupakan nama login.
- d. Field password merupakan password login.
- e. Field status merupakan kode user.

3.3.2 Tabel Siswa

Tabel siswa ini digunakan untuk menyimpan No Induk, Nama, Nilai Tugas, Nilai Sikap, UTS dan UAS, dengan No_induk sebagai primary key.

Tabel 3.3 Tabel siswa

No	Nama	Tipe	Panjang	Keterangan
1	No_Induk	Int	5	Primary key
2	Nama	Varchar	25	
3	Nilai Tugas	Double	10	
4	Nilai Sikap	Double	10	
5	UTS	Double	10	

No	Nama	Tipe	Panjang	Keterangan
6	UAS	Double	10	

Variabel pada tabel user meliputi :

- Field No induk merupakan nama No induk tiap siswa.
- Field nama merupakan nama siswa.
- Field Nilai Tugas, Nilai Sikap, UTS, dan UAS merupakan nilai dari tiap-tiap mata pelajaran.

3.3.3 Tabel Hasil Pengelompokan

Tabel hasil pengelompokan digunakan untuk menyimpan data data hasil pengelompokan

Tabel 3.4 Tabel Hasil Pengelompokan

No	Nama	Tipe	Panjang	Keterangan
1	Id_Hasil	Int	5	Primary Key
2	Jarak_Terdekat	Double	10	
3	Cluster	Int	5	
4	Iterasi	Int	5	

Variabel pada tabel laporan cluster meliputi :

- Field id_ hasil merupakan proses pengelompokan siswa.
- Field Jarak_Terdekat merupakan jarak terdekat pada cluster tersebut.
- Field Cluster merupakan cluster yang diikuti dari setiap siswa.
- Field Iterasi merupakan nilai iterasi yang sedang terjadi pada cluster tersebut.

3.4 Perancangan Pengujian

Data yang akan diproses untuk pengelompokan kelas bimbingan belajar adalah nilai mata pelajaran Matematika. Untuk siswa yang dikelompokkan berjumlah 33 siswa yang berasal dari siswa kelas 6. Berikut ini adalah contoh perhitungan pada penelitian ini.

Tabel 3.5 Tabel Data Nilai Mata Pelajaran Matematika.

No	Nama	Tugas	Sikap	UTS	UAS
1	AHMAD DINAN HAWARI	80	76	71	68
2	AIRA ZASKIA TANIA DEWI	80	83	78	75

No	Nama	Tugas	Sikap	UTS	UAS
3	ALIESHA NAMEERA AZKA	81	84	79	76
4	ASSIFA AULIYA SASONGKO	76	79	66	63
5	FAWWAZ ABI EL YOSI ARYASATYA	76	79	74	71
6	GHALLIH RANGGA ADILAH F	83	85	81	78
7	JIHAN NAWAL RAMADHANI	76	79	74	71
8	MIFTAH KASYAFANI FAEYZA	85	88	83	82
9	MOCHAMAD ARVANO IBNI SACHURI	72	75	70	67
10	MOCHAMMAD HAFIDZ RAIHAN	76	79	74	71
11	M. WAHYU VIRGIAWAN BAIHAQI	78	81	80	73
12	NUR AULIA AZIFAH ISLAMIYAH	90	93	88	85
13	RAUDHATUL KISWAH AL FAUZA	88	91	86	83
14	WAHYU PERDANA PUTRA KASTARI	80	83	78	75
15	MARIYATUL QIBTIYAH	82	85	84	87
16	ABID MUHAMMAD HABIBURROFI'EL	81	84	74	77
17	AGNETA KALYANI	84	87	77	80
18	ALFINO FABIANO VALENSIA	78	81	71	74
19	ATHAYA LIANDRA PUTRI ARDIYATI	76	79	69	72
20	AZZAM ABDILAH IRSYAD	84	87	77	80
21	DEA KAMILAH ZENITA RAHARDJANTO	82	85	75	78
22	MUHAMMAD ARIFIN ILHAM	83	84	76	77
23	MUHAMMAD FAIZ	81	84	74	77
24	MOHAMMAD MIRZA IBRAHIM	76	79	69	72
25	MUHAMMAD NABHANUL RIDHO	73	76	66	69
26	MUHAMMAD REZA	79	82	72	75
27	MUHAMAD Yafa ANUGRAH PRATAMA	78	81	71	74
28	MUHAMAD YOFI ANUGRAH PRATAMA	85	88	78	81
29	ZAHROTUS SYITA	84	86	80	82
30	PUTRA BAYU ZAIDAN	76	78	49	52
31	FILIA FATIN QAIS ELWafa	72	75	65	68
32	AKHMAL ARSYA FADILAH CAHYONO	84	87	77	80
33	MUHAMMAD FERRO AL AFGHANI	90	91	83	86

Tahap ini akan dilakukan proses utama, yaitu segmentasi data nilai yang menggunakan algoritma K-means untuk mengakses data dari database. Berikut ini merupakan asumsi bahwa inputan adalah jumlah data set sebanyak $N = 33$ data yaitu jumlah banyaknya siswa dan jumlah inisialisasi centroid $K = 3$.

Selanjutnya akan diterapkan algoritma K-Means untuk mengelompokkan data yang ada. Data yang ada akan di kelompokkan menjadi beberapa kelompok. Berikut pengelompokan datanya:

Penentuan cluster awal, untuk penentuan cluster awalnya dipilih secara acak. Disini kita mengambil centroid dari hasil nilai yang diberikan yaitu :

Diambil dari data ke 19 sebagai pusat cluster ke 1 : (76 , 79 , 69 , 72)

Diambil dari data ke 11 sebagai pusat cluster ke 2 : (78 , 81 , 80, 73)

Diambil dari data ke 8 sebagai pusat cluster ke 3 : (85 , 88 , 83 , 82)

Seperti terlihat pada tabel 3.6 berikut ini :

Tabel 3.6 Cluster Awal

C1	76	79	69	72
C2	78	81	80	73
C3	85	88	83	82

- a. Menghitung jarak setiap data yang ada terhadap setiap pusat cluster menggunakan rumus Euclidian Distance.

Jarak nilai terhadap C1

$$D1c1 = \sqrt{(80 - 76)^2 + (76 - 79)^2 + (71 - 69)^2 + (68 - 72)^2} = 6.70826$$

Jarak nilai terhadap C2

$$D1c2 = \sqrt{(80 - 78)^2 + (76 - 81)^2 + (71 - 80)^2 + (68 - 73)^2} = 11.619$$

Jarak nilai terhadap C3

$$D1c3 = \sqrt{(80 - 85)^2 + (76 - 88)^2 + (71 - 83)^2 + (68 - 82)^2} = 22.561$$

Perhitungan jarak data terhadap cluster selanjutnya sama dengan perhitungan jarak data terhadap cluster sebelumnya. Hasil yang diperoleh dapat dilihat pada tabel 3.10 sebagai berikut ini :

Tabel 3.7 Tabel jarak dan Cluster Iterasi 1 menggunakan Euclidean Distance

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
1	6.7082	11.619	22.561	1
2	11.0454	4	11.1355	2
3	12.8841	5.2915	9.1652	2
4	9.4868	17.4356	28.4956	1
5	5.099	6.9282	19.0788	1
6	16.2788	8.1854	5.7446	3
7	5.099	6.9282	19.0788	1
8	21.4009	13.7113	0	3
9	7.6158	14.4222	27.0555	1
10	5.099	6.9282	19.0788	1
11	11.4018	0	13.7113	2
12	30.3645	22.2711	9.1652	3
13	26.4197	18.3303	5.2915	3
14	11.0454	4	11.1355	2
15	22.8473	15.6205	6.6332	3
16	10	8.3666	11.7473	2
17	16	11.4018	6.4807	3
18	4	9.0554	17.4929	1
19	0	11.4018	21.4009	1
20	16	11.4018	6.4807	3
21	12	9.0554	9.8995	2
22	12.1655	8.124	9.6954	2
23	10	8.3666	11.7473	2
24	0	11.4018	21.4009	1
25	6	16.1864	27.313	1
26	6	8.3666	15.5563	1
27	4	9.0554	17.4929	1
28	18	12.8841	5.099	3
29	18.2757	11.9164	3.7417	3

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
30	28.3019	37.6165	47.2969	1
31	8	17.9444	29.2916	1
32	16	11.4018	6.4807	3
33	27.0555	20.5426	7.0711	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 14 siswa, cluster kedua ada 8 siswa dan cluster ketiga ada 11 siswa. Jika hasil cluster masih mengalami perubahan maka kita mengulangi langkah tersebut sampai posisi cluster tidak mengalami perubahan.

Dalam contoh perhitungan metode K-Means menggunakan Euclidean Distance ini berhenti pada iterasi 3 karena tidak ada data yang berpindah cluster. Berikut tabel hasil perhitungan jarak pada iterasi 3.

Tabel 3.8 Tabel jarak dan Cluster iterasi 3 menggunakan Euclidean Distance

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
1	5.455	12.4495	22.4031	1
2	13.2138	1.9468	11.3181	2
3	15.1248	2.3643	9.4074	2
4	6.1635	18.1491	28.2294	1
5	6.1635	8.7516	19.1442	1
6	18.6111	5.4028	6.237	2
7	6.1635	8.7516	19.1442	1
8	23.9773	10.6014	1.9235	3
9	5.4196	16.655	27.0721	1
10	6.1635	8.7516	19.1442	1
11	12.9313	5.8813	14.3143	2
12	32.8046	19.5394	9.5969	3
13	28.8416	15.5625	5.9414	3
14	13.2138	1.9468	11.3181	2
15	25.8366	13.2586	7.1063	3
16	12.7818	2.8965	11.149	2
17	18.7039	6.1474	5.4314	3

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
18	6.9937	7.1407	17.0617	1
19	3.5501	10.9266	21.0309	1
20	18.7039	6.1474	5.4314	3
21	14.7488	3.1922	9.2033	2
22	14.6546	2.5278	9.2033	2
23	12.7818	2.8965	11.149	2
24	3.5501	10.9266	21.0309	1
25	4.2593	16.7926	27.0019	1
26	8.8919	5.3656	15.0831	2
27	6.9937	7.1407	17.0617	1
28	20.6879	7.9743	3.7283	3
29	21.0564	7.8861	3.1464	3
30	25.5326	37.4351	46.668	1
31	5.999	18.7667	28.9948	1
32	18.7039	6.1474	5.4314	3
33	29.7216	16.691	6.4109	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 13 siswa, cluster kedua ada 9 siswa dan cluster ketiga ada 11 siswa. Karena posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster yaitu :

- a. Cluster pertama ada 13 anggota siswa, yaitu data ke : 1,4,5,7,9,10,18,19,24,25,27,30,31
- b. Cluster kedua ada 9 anggota siswa, yaitu data ke : 2,3,11,14,16,21,22,23,26
- c. Cluster ketiga ada 11 anggota siswa, yaitu data ke : 6,8,12,13,15,17,20,28,29,32,33

Dengan merata-ratakan seluruh fitur setiap nilai centroid dari iterasi sebelumnya, maka dapat diketahui nilai cluster terkecil, nilai cluster sedang, dan nilai cluster terbesar pada ketiga cluster tersebut. Tabel di bawah menampilkan nilai rata-rata untuk cluster terakhir:

Tabel 3.9 Hasil rata-rata centroid terakhir

C1	72,73
C2	79,28
C3	84,23

Pada tabel 3.9 menunjukkan bahwa hasil rata-rata setiap centroid yang diperoleh dari data akhir centroid adalah sebagai berikut, yang mendapatkan nilai rata-rata terbesar adalah Centroid 3 yaitu 84,23. Kemudian yang mendapatkan nilai rata-rata sedang adalah Centroid 2 yaitu 79,28. Dan yang mendapatkan nilai rata-rata terkecil adalah Centroid 1 yaitu 72,73. Maka dapat disimpulkan bahwa siswa yang mengikuti Cluster 3 merupakan siswa dengan kelompok nilai siswa yang tinggi, dan Cluster 2 merupakan siswa dengan kelompok nilai siswa yang sedang, dan Cluster 1 merupakan siswa dengan kelompok nilai siswa yang rendah.

b. Menghitung jarak setiap data yang ada terhadap setiap pusat cluster menggunakan rumus Manhattan Distance.

Jarak nilai terhadap C1

$$D1c1 = |80 - 76| + |76 - 79| + |71 - 69| + |68 - 72| = 13$$

Jarak nilai terhadap C2

$$D1c2 = |80 - 78| + |76 - 81| + |71 - 80| + |68 - 73| = 21$$

Jarak nilai terhadap C3

$$D1c3 = |80 - 85| + |76 - 88| + |71 - 83| + |68 - 82| = 43$$

Perhitungan jarak data terhadap cluster selanjutnya sama dengan perhitungan jarak data terhadap cluster sebelumnya. Hasil yang diperoleh dapat dilihat pada tabel 3.10 sebagai berikut ini :

Tabel 3.10 Tabel jarak dan Cluster Iterasi 1 menggunakan Manhattan Distance

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
1	13	21	43	1
2	20	8	22	2
3	24	10	18	2
4	12	28	54	1
5	6	12	38	1
6	31	15	11	3

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
7	6	12	38	1
8	42	26	0	3
9	14	28	54	1
10	6	12	38	1
11	16	0	26	2
12	60	44	18	3
13	52	36	10	3
14	20	8	22	2
15	42	26	12	3
16	20	16	22	2
17	32	22	10	3
18	8	10	34	1
19	0	16	42	1
20	32	22	10	3
21	24	18	18	2
21	24	18	18	3
22	24	16	18	2
23	20	16	22	2
24	0	16	42	1
25	12	28	54	1
26	12	12	30	2
26	12	12	30	1
27	8	10	34	1
28	36	24	6	3
29	36	20	6	3
30	41	57	83	1
31	16	32	58	1
32	32	22	10	3
33	54	38	12	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 14 siswa, cluster kedua ada 7 siswa dan cluster ketiga ada 12 siswa. Jika hasil cluster masih mengalami perubahan maka kita mengulangi langkah tersebut sampai posisi cluster tidak mengalami perubahan. Dalam contoh perhitungan metode K-Means menggunakan Manhattan Distance ini berhenti pada iterasi 4 karena tidak ada data yang berpindah cluster. Berikut tabel hasil perhitungan jarak pada iterasi 4.

Tabel 3.11 Tabel jarak dan Cluster iterasi 4 menggunakan Manhattan Distance

No	d_nC1	d_nC2	d_nC3	Cluster yang diikuti
1	9.7273	18.8182	41.8181	1
2	27.4545	3.4546	20.8181	2
3	31.4545	6.1818	16.8181	2
4	8.5455	29.8182	52.8181	1
5	11.4545	13.8182	36.8181	1
6	38.4545	13.1818	9.8181	3
7	11.4545	13.8182	36.8181	1
8	49.4545	24.1818	2.2727	3
9	8.7273	29.8182	52.8181	1
10	11.4545	13.8182	36.8181	1
11	23.4545	11.2728	24.8181	2
12	67.4545	42.1818	19.1819	3
13	59.4545	34.1818	11.1819	3
14	27.4545	3.4546	20.8181	2
15	49.4545	24.1818	13.9091	3
16	27.4545	4.7272	20.8181	2
17	39.4545	14.1818	8.8181	3
18	15.4545	9.8182	32.8181	2
19	7.4545	17.8182	40.8181	1
20	39.4545	14.1818	8.8181	3
21	31.4545	6.7272	16.8181	2
22	31.4545	6.1818	16.8181	2
23	27.4545	4.7272	20.8181	2

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
24	7.4545	17.8182	40.8181	1
25	7.2727	29.8182	52.8181	1
26	19.4545	5.8182	28.8181	2
27	15.4545	9.8182	32.8181	2
28	43.4545	18.1818	4.8181	3
29	43.4545	18.1818	4.8181	3
30	35.5455	58.8182	81.8181	1
31	9.2727	33.8182	56.8181	1
32	39.4545	14.1818	8.8181	3
33	61.4545	36.1818	13.1819	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 11 siswa, cluster kedua ada 11 siswa dan cluster ketiga ada 11 siswa. Karena posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster yaitu :

- a. Cluster pertama ada 11 anggota siswa, yaitu data ke :
1,4,5,7,9,10,19,24,25,30,31
- b. Cluster kedua ada 11 anggota siswa, yaitu data ke :
2,3,11,14,16,18,21,22,23,26,27
- c. Cluster ketiga ada 11 anggota siswa, yaitu data ke :
6,8,12,13,15,17,20,28,29,32,33

Dengan merata-ratakan seluruh fitur setiap nilai centroid dari iterasi sebelumnya, maka dapat diketahui nilai cluster terkecil, nilai cluster sedang, dan nilai cluster terbesar pada ketiga cluster tersebut. Tabel di bawah menampilkan nilai rata-rata untuk cluster terakhir:

Tabel 3.12 Hasil rata-rata centroid terakhir

C1	72,25
C2	78,50
C3	84,05

Pada tabel 3.12 menunjukkan bahwa hasil rata-rata setiap centroid yang diperoleh dari data akhir centroid adalah sebagai berikut, yang mendapatkan nilai rata-rata terbesar adalah Centroid 3 yaitu 84,05. Kemudian yang mendapatkan nilai rata-rata

sedang adalah Centroid 2 yaitu 78,50. Dan yang mendapatkan nilai rata-rata terkecil adalah Centroid 1 yaitu 72,25. Maka dapat disimpulkan bahwa siswa yang mengikuti Cluster 3 merupakan siswa dengan kelompok nilai siswa yang tinggi, dan Cluster 2 merupakan siswa dengan kelompok nilai siswa yang sedang, dan Cluster 1 merupakan siswa dengan kelompok nilai siswa yang rendah.

c. Menghitung jarak setiap data yang ada terhadap setiap pusat cluster menggunakan rumus Minkowski Distance dengan Nilai $P = 3$.

Jarak nilai terhadap C1

$$D1c1 = \sqrt[3]{(80 - 76)^3 + (76 - 79)^3 + (71 - 69)^3 + (68 - 72)^3} = 5.4626$$

Jarak nilai terhadap C2

$$D1c2 = \sqrt[3]{(80 - 78)^3 + (76 - 81)^3 + (71 - 80)^3 + (68 - 73)^3} = 9.9565$$

Jarak nilai terhadap C3

$$D1c3 = \sqrt[3]{(80 - 85)^3 + (76 - 88)^3 + (71 - 83)^3 + (68 - 82)^3} = 18.4935$$

Perhitungan jarak data terhadap cluster selanjutnya sama dengan perhitungan jarak data terhadap cluster sebelumnya. Hasil yang diperoleh dapat dilihat pada tabel 3.13 sebagai berikut ini :

Tabel 3.13 Tabel jarak dan Cluster Iterasi 1 menggunakan Minkowski Distance

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
1	5.4626	9.9565	18.4935	1
2	9.5973	3.1748	8.9545	2
3	10.953	4.3445	7.4169	2
4	9.1098	15.55	23.6512	1
5	5.0133	6.2145	15.2089	1
6	13.5775	6.8041	4.7475	3
7	5.0133	6.2145	15.2089	1
8	17.327	11.2977	0	3
9	6.333	11.8119	21.5199	1
10	5.0133	6.2145	15.2089	1
11	11.0466	0	11.2977	2
12	24.4096	17.859	7.3803	3
13	21.3267	14.7606	4.3445	3

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
14	9.5973	3.1748	8.9545	2
15	19.2937	14.3192	5.6462	3
16	7.937	6.9382	9.9396	2
17	12.6992	9.2909	6.0912	3
18	3.1748	9.0041	14.3029	1
19	0	11.0466	17.327	1
20	12.6992	9.2909	6.0912	3
21	9.5244	7.2304	8.5726	2
22	9.7819	6.5421	8.1433	2
23	7.937	6.9382	9.9396	2
24	0	11.0466	17.327	1
25	4.7622	14.5148	21.9434	1
26	4.7622	8.0517	12.818	1
27	3.1748	9.0041	14.3029	1
28	14.2866	10.6443	5.0133	3
29	14.7146	10.2281	3.3019	3
30	25.1989	33.9373	40.8232	1
31	6.3496	15.7835	23.4952	1
32	12.6992	9.2909	6.0912	3
33	21.5156	17.0449	6	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 14 siswa, cluster kedua ada 8 siswa dan cluster ketiga ada 11 siswa. Jika hasil cluster masih mengalami perubahan maka kita mengulangi langkah tersebut sampai posisi cluster tidak mengalami perubahan. Dalam contoh perhitungan metode K-Means menggunakan Minkowski Distance ini berhenti pada iterasi 3 karena tidak ada data yang berpindah cluster. Berikut tabel hasil perhitungan jarak pada iterasi 3.

Tabel 3.14 Tabel jarak dan Cluster iterasi 3 menggunakan Minkowski Distance

No	d_nC1	d_nC2	d_nC3	Cluster yang diikuti
1	4.7003	10.4434	18.4117	1
2	11.0215	1.6089	9.2809	2
3	12.4759	2.3084	7.8107	2
4	5.7614	15.4317	23.3923	1
5	5.7614	7.0744	15.3813	1
6	15.2308	4.6476	5.3236	2
7	5.7614	7.0744	15.3813	1
8	19.3641	8.5299	1.7514	3
9	4.5363	13.2949	21.6212	1
10	5.7614	7.0744	15.3813	1
11	11.901	4.6922	11.9697	2
12	26.2559	15.5829	7.956	3
13	23.141	12.4461	5.1601	3
14	11.0215	1.6089	9.2809	2
15	21.9561	11.9094	5.7232	3
16	10.3523	2.7389	9.0511	2
17	14.985	5.1298	4.6932	3
18	5.9266	6.1449	13.6741	1
19	3.4092	8.9925	16.7992	1
20	14.985	5.1298	4.6932	3
21	11.885	2.565	7.5477	2
22	11.7179	2.2795	7.4982	2
23	10.3523	2.7389	9.0511	2
24	3.4092	8.9925	16.7992	1
25	3.5528	13.5347	21.5143	1
26	7.3566	4.8539	12.1212	2
27	5.9266	6.1449	13.6741	1
28	16.5458	6.5813	3.4278	3
29	17.1061	6.6169	2.6531	3

No	$d_n C1$	$d_n C2$	$d_n C3$	Cluster yang diikuti
30	22.812	32.8745	40.0683	1
31	4.9856	15.0791	23.0904	1
32	14.985	5.1298	4.6932	3
33	23.739	13.4274	5.3025	3

Maka hasil dari tabel diatas, untuk cluster pertama ada 13 siswa, cluster kedua ada 10 siswa dan cluster ketiga ada 10 siswa. Karena posisi cluster tidak berubah maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster yaitu :

- a. Cluster pertama ada 13 anggota siswa, yaitu data ke : 1,4,5,7,9,10,18,19,24,25,27,30,31
- b. Cluster kedua ada 10 anggota siswa, yaitu data ke : 2,3,6,11,14,16,21,22,23,24,26
- c. Cluster ketiga ada 10 anggota siswa, yaitu data ke : 8,12,13,15,17,20,28,29,32,33

Dengan merata-ratakan seluruh fitur setiap nilai centroid dari iterasi sebelumnya, maka dapat diketahui nilai cluster terkecil, nilai cluster sedang, dan nilai cluster terbesar pada ketiga cluster tersebut. Tabel di bawah menampilkan nilai rata-rata untuk cluster terakhir:

Tabel 3.15 Hasil rata-rata centroid terakhir

C1	72,73
C2	79,28
C3	84,48

Pada tabel 3.15 menunjukkan bahwa hasil rata-rata setiap centroid yang diperoleh dari data akhir centroid adalah sebagai berikut, yang mendapatkan nilai rata-rata terbesar adalah Centroid 3 yaitu 84,48. Kemudian yang mendapatkan nilai rata-rata sedang adalah Centroid 2 yaitu 79,28. Dan yang mendapatkan nilai rata-rata terkecil adalah Centroid 1 yaitu 72,73. Maka dapat disimpulkan bahwa siswa yang mengikuti Cluster 3 merupakan siswa dengan kelompok nilai siswa yang tinggi, dan Cluster 2 merupakan siswa dengan kelompok nilai siswa yang sedang, dan Cluster 1 merupakan siswa dengan kelompok nilai siswa yang rendah.

3.5 Evaluasi Cluster Davies Bouldin Index

Skenario evaluasi cluster sistem ini menggunakan evaluasi validitas internal. Evaluasi validitas internal adalah proses menganalisis hasil cluster tanpa menggunakan data luar atau eksternal. Evaluasi validitas cluster akan membandingkan hasil cluster dengan nilai $K=3$ menggunakan metode validitas Davies Bouldin Index yang pertama kali diperkenalkan oleh Donald W. Bouldin dan David L. Davies pada tahun 1972. Dimana nilai DBI yang paling rendah adalah letak cluster yang terbaik/valid.

Langkah-langkah menghitung Evaluasi Cluster Davies Bouldin Index yaitu:

1. Menghitung SSW (Sum of Square Within cluster). Perhitungan Sum of Square Within cluster (SSW) dengan menghitung jarak setiap data ke centroid masing – masing, dari hasil cluster terakhir yang didapatkan dalam proses akhir clustering iterasi 1

Rumus SSW yaitu :

$$SSW = \frac{1}{m} \sum_{i=1}^m ||x_i - c_{pi}||^2$$

Rata – rata jarak data cluster 1 dengan centroid 1 menggunakan Euclidean Distance

$$SSW_1 = 6.815$$

Rata – rata jarak data cluster 2 dengan centroid 2 menggunakan Euclidean Distance

$$SSW_2 = 5.9005$$

Rata – rata jarak data cluster 3 dengan centroid 3 menggunakan Euclidean Distance

$$SSW_3 = 5.6535$$

Rata – rata jarak data cluster 1 dengan centroid 1 menggunakan Manhattan Distance

$$SSW_1 = 10.956$$

Rata – rata jarak data cluster 2 dengan centroid 2 menggunakan Manhattan Distance

$$SSW_2 = 5.6049$$

Rata – rata jarak data cluster 3 dengan centroid 3 menggunakan Manhattan Distance

$$SSW_3 = 9.4167$$

Rata – rata jarak data cluster 1 dengan centroid 1 menggunakan Minkowski Distance

$$SSW_1 = 5.9548$$

Rata – rata jarak data cluster 2 dengan centroid 2 menggunakan Minkowski Distance

$$SSW_2 = 4.7929$$

Rata – rata jarak data cluster 3 dengan centroid 3 menggunakan Minkowski Distance

$$SSW_3 = 4.9734$$

2. Langkah selanjutnya yaitu menghitung SSB (Sum of Square Between cluster).

Rumus SSB yaitu :

$$SSB = \frac{2}{m(m-1)} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \|c_i - c_j\|^2$$

Jarak antara centroid 1 dengan centroid 2

$$SSB_{1,2} = 11,08279759$$

Jarak antara centroid 1 dengan centroid 3

$$SSB_{1,3} = 20,94823149$$

Jarak antara centroid 2 dengan centroid 3

$$SSB_{2,3} = 13,7113092$$

3. Selanjutnya yaitu menghitung R dan DBI. Rumus dari R yaitu :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Sedangkan rumus dari DBI yaitu :

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

Tabel 3.16 Hasil perhitungan uji validitas DBI menggunakan Euclidean Distance Iterasi 1

	R	Cluster ke-i			R max	DBI
		1	2	3		
Cluster ke-i	1	0	1.1152	0.5826	1.1152	1.0244
	2	1.1152	0	0.8427	1.1152	
	3	0.5826	0.8427	0	0.8427	

Tabel 3.17 Hasil perhitungan uji validitas DBI menggunakan Manhattan Distance Iterasi 1

	R	Cluster ke-i			R max	DBI
		1	2	3		
Cluster ke-i	1	0	1.9782	0.9929	1.9782	1.8489
	2	1.9782	0	1.5903	1.9782	
	3	0.9929	1.5903	0	1.5903	

Tabel 3.18 Hasil perhitungan uji validitas DBI menggunakan Minkowski Distance Iterasi 1

	R	Cluster ke-i			R max	DBI
		1	2	3		
Cluster ke-i	1	0	0.9426	0.5106	0.9426	0.8658
	2	0.9426	0	0.7123	0.9426	
	3	0.5106	0.7123	0	0.7123	

Dari hasil perhitungan menggunakan 33 data dalam 3 cluster pengelompokan, hasil uji validitas pengelompokan dengan DBI pada iterasi 1 menggunakan Euclidean Distance adalah 1.0244, kemudian untuk hasil uji validitas pengelompokan dengan DBI pada iterasi 1 menggunakan Manhattan Distance adalah 1.8489, dan untuk hasil uji validitas pengelompokan dengan DBI pada iterasi 1 menggunakan Manhattan Distance adalah 0.8658. Dapat ditunjukkan bahwa ketika nilai SSW semakin kecil, maka hasil clustering yang didapat akan lebih baik. Secara esensial, DBI menginginkan nilai sekecil (non-negatif ≥ 0) mungkin untuk

menilai baiknya cluster yang didapat. Pada contoh perhitungan metode K-Means ini, nilai DBI juga dihitung tiap iterasi. Berikut nilai DBI untuk tiap iterasi disajikan pada tabel 3.19.

Tabel 3.19 Hasil Uji Validitas DBI Tiap Iterasi

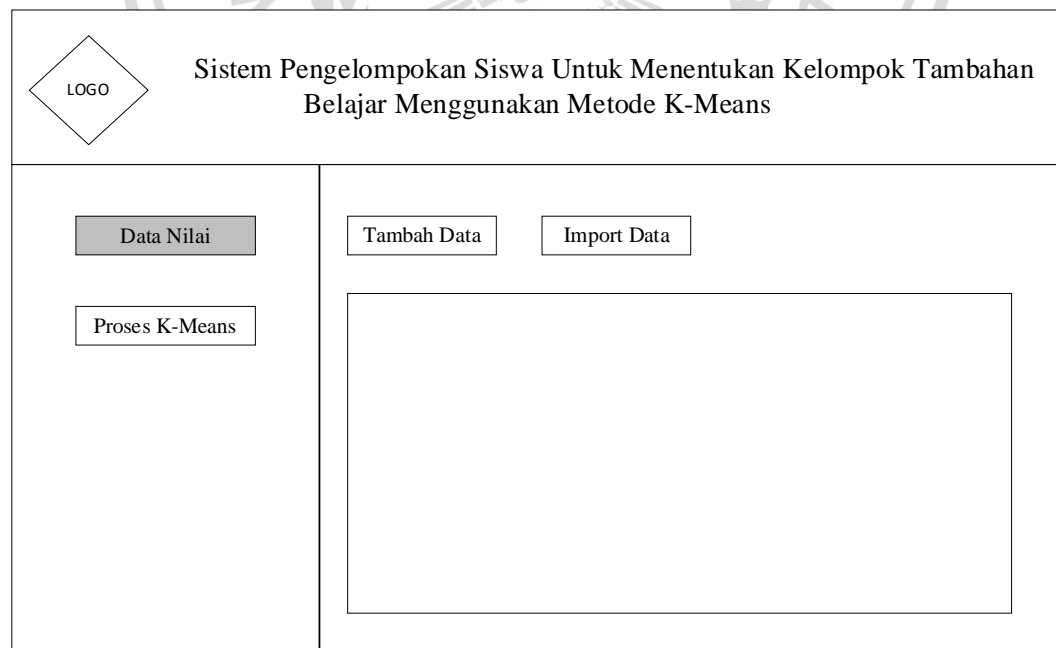
Manhattan Distance		Euclidean Distance		Minkowski Distance	
Iterasi ke-	Hasil DBI	Iterasi ke-	Hasil DBI	Iterasi ke-	Hasil DBI
Iterasi 1	1.8489	Iterasi 1	1.0244	Iterasi 1	0.8658
Iterasi 2	1.4756	Iterasi 2	0.8595	Iterasi 2	0.7484
Iterasi 3	1.50350	Iterasi 3	0.8271	Iterasi 3	0.7173
Iterasi 4	1.4066				

Dari pengujian hasil DBI untuk tiap iterasi didapatkan nilai terkecil menggunakan Minkowski Distance.

3.6 Perancangan Antarmuka Sistem

Perancangan antarmuka adalah bagian dari program yang menciptakan hubungan antara pengguna dan aplikasi. Interface sistem ditulis dalam bahasa PHP berbasis web. Sistem ini membagi datanya ke dalam beberapa menu.

3.6.1 Halaman Utama



Gambar 3.7 Interface Halaman Home

3.6.2 Halaman Proses K-Means

The interface for the K-Means process is titled "Sistem Pengelompokan Siswa Untuk Menentukan Kelompok Tambahan Belajar Menggunakan Metode K-Means". It includes a logo placeholder and several interactive elements: a "Data Nilai" button, a "Tambah Proses K-Means" button, a "Proses K-Means" button, and input fields for "Mata Pelajaran" and "Jumlah Cluster". A large empty rectangular area is provided for displaying the results of the clustering process.

Gambar 3.8 Interface Halaman Proses K-Means

3.7 Skenario Pengujian Sistem

Berikut ini adalah skenario pengujian sistem :

- Penginputan data kesistem berupa nilai Tugas, nilai Sikap, Nilai UTS dan nilai UAS para siswa kelas 6 SD.
- Dilakukan perhitungan clustering permata pelajaran yang diujikan. Ada 12 mata pelajaran yang akan diujiakan yaitu PAI, IPA, IPS, Matematika, Kemuhammadiyah, Bahasa Arab, Bahasa Jawa, PLH, B.Ingggris, B.Indonesia, PPKN, PJOK
- Perhitungan dilakukan menggunakan metode K-Means dengan rumus menghitung jarak berbeda-beda yaitu, Euclidean Distance, Manhattan Distance, dan Minkowski Distance.
- Kemudian hasil perhitungan cluster yang telah ditentukan melalui perhitungan, dilakukan proses voting menggunakan rata-rata terakhir atau centroid terakhir untuk menentukan cluster yang Tinggi, Sedang, dan Rendah.

- e. Setelah mendapatkan hasil cluster dari perhitungan, langkah selanjutnya adalah melakukan evaluasi menggunakan Davies Bouldin Index (DBI) untuk mengetahui cluster terbaik dan centroid yang lebih valid.
- f. Percobaan perhitungan dilakukan beberapa kali untuk mencari nilai DBI terkecil. Secara esensial, DBI menginginkan nilai sekecil (non-negatif ≥ 0) untuk menilai baiknya cluster yang didapat.
- g. Pada percobaan kali ini dilakukan perhitungan menggunakan $K = 2$ sampai 7 menggunakan jarak Euclidean Distance, Manhattan Distance, dan Minkowski Distance. Hasil yang diperoleh dapat dilihat pada tabel 3.20.

Tabel 3.20 Hasil Perhitungan

K	Nilai DBI Mata Pelajaran Matematika		
	Euclidean Distance	Manhattan Distance	Minkowski Distance
2	0.7172	1.2229	0.6225
3	0.8271	1.4066	0.7173
4	0.8067	1.4151	0.7202
5	0.8498	1.4473	0.6905
6	0.7685	1.3321	0.7039
7	0.8133	1.3723	0.6893

- h. Hasil perhitungan cluster tiap data akan dibandingkan, pada contoh kali ini menggunakan 2 Cluster. Dapat dilihat pada Tabel 3.21.

Tabel 3.21 Hasil perhitungan menggunakan 2 Cluster

Data	Cluster yang diikuti		
	Euclidean Distance	Manhattan Distance	Minkowski Distance
1	1	1	1
2	2	2	2
3	2	2	2
...
31	1	1	1
32	2	2	2
33	2	2	2

- i. Dibawah ini contoh menggunakan 7 cluster. Dapat dilihat pada Tabel 3.22.

Tabel 3.22 Hasil perhitungan menggunakan 7 Cluster

Data	Cluster yang diikuti		
	Euclidean Distance	Manhattan Distance	Minkowski Distance
1	4	2	4
2	5	5	5
3	5	5	5
4	1	1	1
5	4	2	4
...
31	1	1	1
32	6	6	6
33	7	7	7

