JDSP
Jurnal Data Science Penusa
E-ISSN: 1234-5678

Analisis Pengelompokan Data Siswa Untuk Menentukan Siswa Berprestasi Menggunakan Algoritma *K-Means Clustering*

Desmy Maelani[1)*], Megaria Purba[2]
[1,2,] Sekolah Tinggi Manajemen Informatika dan Komputer Pelita Nusantara, Medan
[1,2,] Email : desmymaelani@gmail.com[1], megariapurba18@gmail.com[2]

Received: 30 October 2025 Accepted: 25 Juni 2026 Published: 10 July 2026
Pp. 243-254

***Penulis Koresponden**

Kata Kunci: Data Mining, K-Means Clustering, Kehadiran, Ekstrakurikuler, dan Prestasi Siswa

JDSP: Jurnal Data Science Penusa is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Penelitian ini bertujuan untuk mengidentifikasi siswa berprestasi di UPT SPF SD Negeri 106843 Jaharun B Dengan memanfaatkan metode K-Means Clustering dalam proses analisis. Variabel yang digunakan dalam penelitian meliputi kehadiran, nilai matematika, bahasa indonesia, agama dan ekstrakurikuler, karena kelima variabel tersebut dianggap sebagai indikator penting dalam menilai prestasi siswa secara menyeluruh. Melalui penerapan metode K-Means Clustering, proses analisis data dapat dilakukan secara lebih sistematis dan terstruktur. Proses analisis dimulai dari pengumpulan data siswa, pemberian bobot pada setiap variabel, hingga pengelompokan data ke dalam tiga kelompok utama, yaitu: Berpretasi sangat baik, Berprestasi baik dan Kurang berprestasi. Penelitian ini juga menghasilkan sistem berbasis web dengan memanfaatkan bahasa pemrograman PHP dan manajement basis data MySQL, yang memudahkan admin sekolah dalam menginput, mengelola, dan menganalisis data siswa secara otomatis. Berdasarkan hasil pengujian, sistem terbukti mampu berfungsi dengan baik sesuai dengan tujuan proses klasterisasi dengan baik pada iterasi ketiga dan mampu mengelompokkan siswa berdasarkan prestasi mereka. Keberhasilan metode ini terlihat dari kemampuannya menyajikan hasil clustering yang dapat menjadi dasar pengambilan keputusan strategis dalam pembinaan siswa. Keunggulan dari sistem ini adalah proses pengelompokan yang cepat, objektif, dan berpotensi dikembangkan lebih lanjut dalam konteks pendidikan berbasis data.

PENDAHULUAN

Teknologi informasi terus maju melalui kecepatan yang luar biasa dan memainkan peran penting dalam berbagai bidang, termasuk pendidikan. Penerapan teknologi seperti data mining memungkinkan setiap institusi pendidikan untuk menggali informasi penting dari data sebelumnya hanya tersimpan tanpa dimanfaatkan secara maksimal. Algoritma K-Means adalah salah satu teknik data mining yang paling umum digunakan untuk mengelompokkan data. Algoritma ini memungkinkan analisis berbasis klaster yang membantu dalam mengidentifikasi kelompok pola prestasi siswa berdasarkan kehadiran, nilai matematika, bahasa indonesia, agama dan ekstrakurikuler mereka, seperti siswa yang berprestasi sangat baik, berpretasi baik dan kurang berprestasi. Melalui pendekatan ini, pihak sekolah dapat mengambil Keputusan berbasis data guna meningkatkan kualitas proses belajar mengajar.

UPT SPF SD Negeri 106843 Jaharun B adalah salah satu sekolah dasar negeri yang terletak di Kecamatan Galang, Kabupaten Deli Serdang, Sumatera Utara. Sekolah ini didirikan pada 14 Februari 2018 dengan Nomor SK Pendirian 006 di bawah naungan Kementerian Pendidikan dan Kebudayaan. Data mengenai kehadiran, nilai Matematika, Bahasa Indonesia, Agama, serta kegiatan ekstrakurikuler memang sudah dicatat, namun pengelolaannya masih sebatas administrasi tanpa analisis lebih lanjut untuk mengidentifikasi tingkat prestasi siswa. Hal ini menyebabkan pihak sekolah kesulitan dalam mengidentifikasi siswa yang membutuhkan perhatian khusus secara lebih cepat dan objektif. Oleh karena itu, perlu dilakukan penelitian untuk menerapkan algoritma K-Means guna mengelompokkan siswa berdasarkan kombinasi data tersebut, sehingga dapat diketahui pola prestasi yang terbentuk dan membantu sekolah dalam proses pengambilan keputusan yang lebih tepat sasaran.

Dalam beberapa tahun terakhir, telah dilakukan berbagai penelitian yang berfokus pada analisis terhadap siswa berprestasi. Penelitian yang dilakukan oleh (Penda Sudarto Hasugian, 2022) bertujuan untuk mengelompokkan data siswa berdasarkan capaian akademik dengan menggunakan algoritma K-Means. Studi ini dilaksanakan di SMP Swasta Agia Sophia dan memanfaatkan nilai raport serta nilai ujian siswa sebagai atribut utama. Pengelompokan dilakukan ke dalam tiga kategori, yakni siswa unggul, sedang, dan rendah. Proses analisis dibantu dengan perangkat lunak RapidMiner, yang menunjukkan bahwa penerapan algoritma K-Means mampu

menghasilkan pengelompokan yang optimal, sehingga mempermudah pihak sekolah dalam menyusun strategi pembelajaran yang tepat sasaran. Di sisi lain, penelitian oleh (Syukron Ramadani & Fatah, 2024) dijelaskan bahwa metode *Clustering K-Means* efektif digunakan untuk menganalisis dan mengelompokkan data nilai siswa guna menentukan siswa berprestasi. Penelitian tersebut menggunakan data akademik dari berbagai mata pelajaran, dan hasilnya menunjukkan bahwa algoritma K-Means mampu membagi siswa ke dalam beberapa kelompok, seperti siswa berprestasi sangat baik, berprestasi baik dan kurang berprestasi. Hasil pengelompokan ini mempermudah pihak sekolah dalam mengambil keputusan berbasis data serta meningkatkan kualitas evaluasi pembelajaran siswa secara objektif dan efisien.

Algoritma K-Means digunakan untuk menganalisis dan menemukan pola pretasi siswa di kelas. Pembelajaran tanpa pengawasan *(unsupervised learning)* menggunakan algoritma K-Means untuk mengelompokkan data ke dalam beberapa klaster berdasarkan fitur tertentu. Algoritma K-Means akan digunakan untuk analisis data siswa yang terdiri dari berbagai kombinasi kehadiran, nilai matematika, bahasa indonesia, agama, dan ekstrakurikuler. Kelebihan dari algoritma K-Means adalah kesederhanaannya, kecepatan dalam memproses data dalam jumlah besar, serta kemampuannya menemukan pola tersembunyi secara otomatis dari data yang kompleks. Dengan pendekatan ini, data siswa di UPT SPF SDN 106843 Jaharun B dapat dikelompokkan menjadi beberapa kategori, seperti siswa berprestasi sangat baik,berprestasi baik dan kurang berprestasi, berdasarkan parameter yang dianalisis. Hasil pengelompokan ini diharapkan dapat membantu pihak sekolah dalam melakukan evaluasi dan merancang strategi pembinaan yang lebih tepat sasaran untuk meningkatkan kualitas pendidikan secara menyeluruh. Dari pembahasan diatas maka penulis menyimpulkan judul dari penelitian ini adalah: **"Analisis Pengelompokan Data Siswa untuk menentukan siswa berprestasi Menggunakan Algoritma K-Means Clustering "**

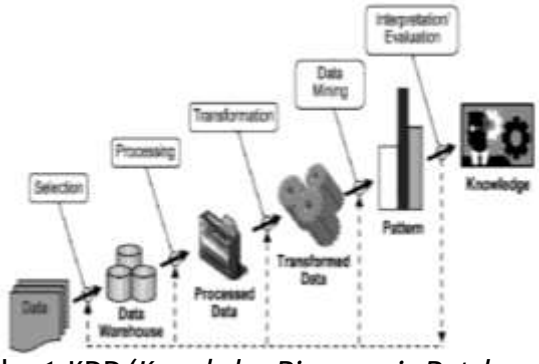
TINJAUAN LITERATUR

a) Prestasi Siswa

Prestasi merupakan hasil dari suatu usaha kegiatan yang telah dilakukan oleh seseorang dalam mencapai tujuan tertentu. Dalam konteks pendidikan, prestasi sering dikaitkan dengan pencapaian akademik siswa. Prestasi menunjukkan sejauh mana siswa menguasai materi pembelajaran yang diajarkan dalam kurikulum. Kehadiran siswa merupakan faktor penting dalam dunia pendidikan dan secara langsung berdampak pada kinerja akademis dan perkembangan sosial. Ada banyak faktor yang memengaruhi partisipasi, baik dalam diri siswa itu sendiri maupun lingkungannya. Menurut (Afniola et al., 2020), prestasi belajar atau prestasi akademik merupakan ungkapan keberhasilan seseorang dalam hal belajar, yang dapat diketahui melalui berbagai bentuk evaluasi seperti tes. Hasil belajar ini menjadi indikator penting dalam menilai efektivitas proses pembelajaran di sekolah.

b) Data Mining

Data mining merupakan sebuah pendekatan dalam pengolahan data yang bertujuan untuk menemukan pola-pola tersembunyi dalam suatu kumpulan data. Informasi yang diperoleh dari hasil proses ini dapat dijadikan sebagai landasan dalam pengambilan keputusan pada waktu yang akan datang. Secara konseptual, data mining merupakan salah satu tahap dalam proses *Knowledge Discovery in Databases* (KDD). Proses KDD sendiri merupakan rangkaian tahapan sistematis yang mencakup kegiatan *data cleaning*, *data integration*, *data selection*, *data transformation*, *pattern evaluation*, dan *knowledge presentation*.



Gambar 1. KDD (*Knowledge Discovery in Database)*

c) Algoritma K-Means Clustering

K-Means Clustering merupakan suatu metode analisis data yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan karakteristik atau properti yang dimiliki oleh masing-masing data. Metode ini termasuk ke dalam kategori pembelajaran tanpa pengawasan, di mana data yang digunakan tidak memiliki label atau kategori sebelumnya. K-Means Clustering secara luas diterapkan dalam berbagai bidang, antara lain segmentasi pasar, pengelompokan dokumen, pengenalan pola, serta analisis citra. Metode

pengelompokan k-means melibatkan dua fase iteratif: Fase penugasan atau inisialisasi melibatkan proses berulang untuk menetapkan setiap titik data ke titik pusat terdekatnya menggunakan metrik Euklides. Berikutnya adalah fase pembaruan pusat gravitasi. Pada fase ini, sentroid klaster diperbarui dengan mempertimbangkan partisi yang diperoleh pada fase sebelumnya. Proses iterasi berakhir ketika titik data tidak lagi mengalami perubahan klaster atau jumlah iterasi maksimum tertentu telah tercapai. (Oti et al., 2021)

d) UML *(Unified Modeling Languange)*

UML adalah bahasa pemodelan standar yang digunakan untuk menggambarkan, merancang dan mendokumentasikan sistem perangkat lunak. UML memungkinkan pengembang untuk menciptakan sistem yang dipahami oleh berbagai pemangku kepentingan yang terlibat dalam proyek pengembangan perangkat lunak. *Unified Modeling Language* (UML) adalah standar bahasa pemodelan visual yang digunakan dalam analisis, desain, dan sistem perangkat lunak berbasis objek. UML dikembangkan oleh Grady Booch, Ivar Jacobson dan James Rumbaust di tahun 90-an dan sekarang menjadi standar yang dikelola oleh *Object Management Group* (OMG)

METODE PENELITIAN

Metodologi penelitian ini diawali dengan identifikasi masalah, yaitu penentuan siswa berprestasi di SDN 106843 Jaharun B yang masih bersifat subjektif dan hanya berfokus pada nilai akademik. Selanjutnya dilakukan analisis masalah untuk mengkaji kelemahan sistem manual yang belum memanfaatkan teknologi dan indikator penilaian yang terbatas, sehingga dipilih solusi menggunakan algoritma K-Means Clustering. Data penelitian dikumpulkan melalui studi pustaka untuk memperkuat landasan teori, observasi langsung di sekolah, serta wawancara dengan guru wali kelas dan tata usaha guna menentukan variabel penilaian siswa, meliputi kehadiran, nilai matematika, bahasa Indonesia, agama, dan ekstrakurikuler. Data tersebut kemudian diproses melalui tahap preprocessing dan dianalisis dengan algoritma K-Means untuk membentuk klaster prestasi siswa menjadi kategori rendah, menengah, dan tinggi. Hasil analisis digunakan dalam perancangan sistem yang dibuat dengan pendekatan UML (use case, class, sequence, dan activity diagram), serta desain database dan antarmuka. Sistem kemudian dibangun menggunakan bahasa pemrograman PHP dan basis data MySQL, diuji menggunakan metode black box testing untuk memastikan fungsionalitas berjalan sesuai kebutuhan, dan diimplementasikan di sekolah dengan melibatkan guru dan staf melalui pelatihan serta pemantauan agar sistem dapat berjalan optimal.

HASIL PENELITIAN

A Analisis Algoritma *K-Means*

Salah satu metode yang diterapkan dalam proses pengelompokan data pada data mining adalah algoritma K-Means. Metode ini digunakan untuk membentuk pohon keputusan dengan menganalisa data yang sudah masuk sehingga memudahkan untuk melakukan pengelompokan untuk data berikutnya.

Tabel 1. Dataset Siswa

No	Nis	Nama	Jenis kelamin	Kelas	Alamat	Kehadiran	Matematika	B_Indonesia	Agama	Ekstrakurikuler
1	2182	Citra ayu kencana	Perempuan	5	Jl. Pahlawan No. 17	98	95	90	98	95
2	2196	Ananda oktavianta	Laki-laki	5	Jl. Perjuangan No. 22	85	75	98	85	75
3	2200	Claudya nathaline zebua	Perempuan	5	Jl. Teratai No. 7	98	80	80	98	80
4	2205	Aldiansyah	Laki-laki	5	Jl. Kembang Raya No. 3	98	90	85	98	90
5	2207	Fadlan alathif	Laki-laki	5	Jl. Asia No. 2	98	100	88	98	100
6	2211	Fardhan twieky swatman	Laki-laki	5	Jl. Bunga Raya No. 5	95	75	90	95	75
7	2220	Ashera queenara eklesia m.	Perempuan	5	Jl. Pancing no. 23	75	80	80	75	80
8	2223	Avina syakilah	Perempuan	5	Jl. Pemuda No.56	85	85	75	85	85
9	2224	Cinta nur ramadhani	Perempuan	5	Jl. Pancasila No. 31	94	90	90	94	90
10	2232	Aylin raisa a. Ginting	Perempuan	4	Jl. Perjuangan No. 11	90	90	90	90	90
11	2233	Abdullah	Laki-laki	4	Jl. Mekar No. 1 Pakam	95	95	90	95	95
12	2234	Andika pratama	Laki-laki	4	Jl. Pertahanan NegaraNo. 2	75	80	80	75	80

13	2235	Diego samuel malau	Laki-laki	4	Jl. Percobaan no. 1	75	85	75	75	85
14	2246	Alfah syahri r d f	Laki-laki	4	Jl. Pahlawan No. 322	95	100	90	95	100
15	2251	David sebastian	Laki-laki	4	Jl.Sisingamangaraja No. 2	90	85	75	90	85
16	2254	Abel jocelin br. Tobing	Perempuan	4	Jl. Pintu Air No. 45	99	90	90	99	90
17	2255	Bastian marcello	Laki-laki	4	Jl. Terang Sari No. 2	85	95	85	85	95
18	2264	Delima mikhaila putri	Perempuan	4	Jl. Asia Barat No.2	95	100	90	95	100
19	2270	Adli fhairus sagala	Laki -laki	4	Jl. Gerbong Kereta No.1	85	80	75	85	80
20	2271	Ariel purba	Laki-laki	5	Jl. Lintas Sumatera No.8	95	80	85	95	80

Tabel diatas menyatakan data siswa yang akan digunakan penelitian ini, yaitu calon siswa yang akan dinyatakan memiliki prestesi sangat baik, baik dan berpretasi kurang baik. Pada tahap selanjutnya dilakukan pembersihan data (data cleaning) yaitu dengan seleksi atribut-atribut yang relavan dan sesuai kebutuhan analisis

Tabel 2. Dataset yang telah di cleaning

No	Nis	Nama	Kehadiran	Matematika	B_Indonesia	Agama	Ekstrakurikuler
1	2182	Citra ayu kencana	98	95	90	98	95
2	2196	Ananda oktavianta	85	75	98	85	75
3	2200	Claudya nathaline zebua	98	80	80	98	80
4	2205	Aldiansyah	98	90	85	98	90
5	2207	Fadlan alathif	98	100	88	98	100
6	2211	Fardhan twieky swatman	95	75	90	95	75
7	2220	Ashera queenara eklesia m.	75	80	80	75	80
8	2223	Avina syakilah	85	85	75	85	85
9	2224	Cinta nur ramadhani	94	90	90	94	90
10	2232	Aylin raisa a. Ginting	90	90	90	90	90
11	2233	Abdullah	95	95	90	95	95
12	2234	Andika pratama	75	80	80	75	80
13	2235	Diego samuel malau	75	85	75	75	85
14	2246	Alfah syahri r d f	95	100	90	95	100
15	2251	David sebastian	90	85	75	90	85
16	2254	Abel jocelin br. Tobing	99	90	90	99	90
17	2255	Bastian marcello	85	95	85	85	95
18	2264	Delima mikhaila putri	95	100	90	95	100
19	2270	Adli fhairus sagala	85	80	75	85	80
20	2271	Ariel purba	95	80	85	95	80

Tabel di atas ini merupakan hasil dari proses data yang telah dibersihkan, yang hanya terdiri dari atribut NIS, Nama, Nilai Matematika, Bahasa Indonesia dan Ekstrakurikuler sebagai variable utama yang akan digunakan dalam proses analisis selanjutnya.

1) Penerapan Algoritma K-Means

Berikut tahapan dalam penerapan algoritma k-means clustering :

a. Menentukan Jumla cluster
   Jumlah cluster yang akan dibentuk sebanyak 3 cluster, yaitu : C1 = Berprestasi Sangat Baik, C2 = Berprestasi Baik dan C3= Kurang Berprestasi.

b. Menentukan titik pusat cluster
Titik pusat cluster diambil secara acak yaitu data ke 1 sebagai *centroid* 1, ke 3 sebagai *centroid* 2 dan data ke 13 sebagai *centroid* 3.

Tabel 3. Tabel centroid awal

Data ke	Centeroid	Kehadiran	Matematika	B_Indonesia	Agama	Estrakurikuler
1	1	98	95	90	98	95
3	2	98	80	80	98	80
13	3	75	85	75	75	85

c. Menghitung jarak dari setiap data terhadap pusat cluster
Rumus untuk menghitung jarak antara satu data (x) dengan pusat cluster (c) menggunakan metode Euclidean adalah sebagai berikut:

$$d(x,c) = \sqrt{\{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2\}}$$

Perhitungan *Euclidean Distance*:

Untuk data baris-1:

$d(1,1)=\sqrt{(98-98)^2+(95-95)^2+(90-90)^2+(98-98)^2+(95-95)^2} = 0$

$d(1,2)=\sqrt{(98-98)^2+(95-80)^2+(90-80)^2+(98-98)^2+(95-80)^2} = 23{,}4521$

$d(1,3)=\sqrt{(98-75)^2+(95-85)^2+(90-75)^2+(98-75)^2+(95-85)^2} = 38{,}5097$

Tabel 4 Hasil Iterasi 1

No	Nis	Nama	Kehadiran, Matematika, Bahasa Indonesia, Agama dan Ektrakurikuler	C1	C2	C3	Minimum	Cluster
1	2182	Citra ayu kencana	98, 95, 90, 98, 95	0.0000	23.4521	38.5097	0	1
2	2196	Ananda oktavianta	85, 75, 98, 85, 75	34.6699	26.6833	30.4795	26.68333	2
3	2200	Claudya nathaline zebua	98, 80, 80, 98, 80	23.4521	0.0000	33.6601	0	2
4	2205	Aldiansyah	98, 90, 85, 98, 90	8.6603	15.0000	34.7563	8.660254	1
5	2207	Fadlan alathif	98, 100, 88, 98, 100	7.3485	29.3939	40.9512	7.348469	1
6	2211	Fardhan twieky swatman	95, 75, 90, 95, 75	28.6007	12.9615	35.0000	12.96148	2
7	2220	Ashera queenara eklesia m.	75, 80, 80, 75, 80	40.0999	32.5269	8.6603	8.660254	3
8	2223	Avina syakilah	85, 85, 75, 85, 85	27.6225	20.3224	14.1421	14.14214	3
9	2224	Cinta nur ramadhani	94, 90, 90, 94, 90	9.0554	18.2209	31.5753	9.055385	1
10	2232	Aylin raisa a. Ginting	90, 90, 90, 90, 90	13.3417	20.6882	26.9258	13.34166	1
11	2233	Abdullah	95, 95, 90, 95, 95	4.2426	23.8328	35.0000	4.242641	1
12	2234	Andika pratama	75, 80, 80, 75, 80	40.0999	32.5269	8.6603	8.660254	3
13	2235	Diego samuel malau	75, 85, 75, 75, 85	38.5097	33.6601	0.0000	0	3
14	2246	Alfah syahri r d f	95, 100, 90, 95, 100	8.2462	30.2985	38.4057	8.246211	1
15	2251	David sebastian	90, 85, 75, 90, 85	23.5160	14.2478	21.2132	14.24781	2
16	2254	Abel jocelin br. Tobing	99, 90, 90, 99, 90	7.2111	17.3781	37.7757	7.211103	1
17	2255	Bastian marcello	85, 95, 85, 85, 95	19.0526	28.5132	22.3607	19.05256	1
18	2264	Delima mikhaila putri	95, 100, 90, 95, 100	8.2462	30.2985	38.4057	8.246211	1
19	2270	Adli fhairus sagala	85, 80, 75, 85, 80	31.8277	19.0526	15.8114	15.81139	3
20	2271	Ariel purba	95, 80, 85, 95, 80	22.2036	6.5574	30.8221	6.557439	2

Dari tabel di atas didapat pemetaan cluster sebagai berikut, data yang berada di cluster 1 maka diberi tanda angka 1 di kolom C1, data yang berada pada cluster 2 diberi tanda angka 1 di kolom C2, dan data yang berada pada cluster 3 diberi tanda angka 1 di kolom C3, maka dapat dilihat pemetaan sebagai berikut:

d. Hitung pusat cluster baru
Penentuan nilai *centroid* yang baru didapat dari rata-rata data berdasarkan kelompok data tersebut sesuai dengan pemetaan data diatas.

$$v = \sum_{i=1}^{n} xi \,/n$$

Keterangan
V : centroid pada cluster
Xi : objek ke-i
N : adalah banyaknya objek atau jumlah objek yang menjadi anggota cluster
Centroid 1
untuk variabel kehadiran=(98+98+98+94+90+95+95+99+85+95)/10=94.7
untuk variabel matematika=(95+90+100+90+90+95+100+90+95+100)/10=94.5
untuk variabel B.Indonesia=(90+85+88+90+90+90+90+90+85+100)/10=88.80
untuk variabel Agama=(98+98+98+94+90+95+95+99+85+95)/10=94.7
untuk variabel ekstrakurikuler =(95+90+100+90+90+95+100+90+95+100)/10=94.5
Centroid 2
untuk variabel kehadiran = (85+98+95+90+95)/5=92.6
untuk variabel matematika= (75+80+75+85+80)/5=79.00
untuk variabel B. Indonesia= (98+80+90+75+85)/5=85.60
untuk variabel Agama= (85+98+95+90+95)/5=92.6
untuk variabel ekstrakurikuler= (75+80+75+85+80)/5=79.00
Centroid 3
untuk variabel kehadiran= (75+85+75+75+85)/5=79
untuk variabel matematika= (80+85+80+85+80)/5=82
untuk variabel B. Indonesia = (85+75+80+75+75)/5=77
untuk variabel Agama= (75+85+75+75+85)/5=79

Tabel 4. Centeroid kedua

Penentuan centeroid baru	Kehadiran	Matematika	B.Indonesia	Agama	Ekstrakurikuler
centeroid baru yang ke -1	94.70	94.50	88.80	94.70	94.50
centeroid baru yang ke -2	92.60	79.00	85.60	92.60	79.00
centeroid baru yang ke -3	79.00	82.00	77.00	79.00	82.00

e. Menghitung Kembali jarak terhadap masing -masing pusat cluster yang baru
Perhitungan *Euclidean Distance*:

Untuk data baris-1:

d(1,1)=$\sqrt{(98-94.70)^2+(95-94.50)^2+(90-88.80)^2+(98-94.70)^2+(95-94.50)^2}$= 4,8703

d(1,2)=$\sqrt{(98-92.6)^2+(95-79)^2+(90-85.60)^2+(98-92.6)^2+(95-79)^2}$17= 24,2833

d(1,3)=$\sqrt{(98-79)^2+(95-82)^2+(90-77)^2(98-79)^2+(95-82)^2}$= 35, 0571

Berikut hasil perhitungan jarak dari setiap data :

Tabel 5. Hasil iterasi 2

No	Nis	Nama	Kehadiran, Matematika, Bahasa Indonesia, Agama dan Ektrakurikuler	C1	C2	C3	Minimum	Cluster
1	2182	Citra ayu kencana	98, 95, 90, 98, 95	4.8703	24.2833	35.0571	4.8703	1
2	2196	Ananda oktavianta	85, 75, 98, 85, 75	32.1453	17.3574	24.7184	17.3574	2
3	2200	Claudya nathaline zebua	98, 80, 80, 98, 80	22.7974	9.5750	27.1846	9.5750	2
4	2205	Aldiansyah	98, 90, 85, 98, 90	8.7590	17.3401	30.2324	8.7590	1
5	2207	Fadlan alathif	98, 100, 88, 98, 100	9.1060	30.7584	38.6135	9.1060	1
6	2211	Fardhan twieky swatman	95, 75, 90, 95, 75	27.6065	7.9297	27.9106	7.9297	2
7	2220	Ashera queenara eklesia m.	75, 80, 80, 75, 80	35.6948	25.5515	7.0000	7.0000	3
8	2223	Avina syakilah	85, 85, 75, 85, 85	23.6457	17.3170	9.6954	9.6954	3
9	2224	Cinta nur ramadhani	94, 90, 90, 94, 90	6.5513	16.2874	27.3313	6.5513	1
10	2232	Aylin raisa a. Ginting	90, 90, 90, 90, 90	9.2801	16.5795	23.2164	9.2801	1
11	2233	Abdullah	95, 95, 90, 95, 95	1.4560	23.2998	31.9218	1.4560	1
12	2234	Andika pratama	75, 80, 80, 75, 80	35.6948	25.5515	7.0000	7.0000	3
13	2235	Diego samuel malau	75, 85, 75, 75, 85	33.8692	28.3528	7.3485	7.3485	3
14	2246	Alfah syahri r d f	95, 100, 90, 95, 100	7.8816	30.2139	36.4555	7.8816	1
15	2251	David sebastian	90, 85, 75, 90, 85	20.3745	14.0670	16.2481	14.0670	2
16	2254	Abel jocelin br. Tobing	99, 90, 90, 99, 90	8.8837	18.5278	33.1210	8.8837	1
17	2255	Bastian marcello	85, 95, 85, 85, 95	14.2520	25.0575	21.7715	14.2520	1
18	2264	Delima mikhaila putri	95, 100, 90, 95, 100	7.8816	30.2139	36.4555	7.8816	1
19	2270	Adli fhairus sagala	85, 80, 75, 85, 80	28.2687	15.1618	9.1652	9.1652	3
20	2271	Ariel purba	95, 80, 85, 95, 80	20.8595	3.7256	24.1661	3.7256	2

Dari tabel di atas didapat pemetaan cluster sebagai berikut sebagai berikut,data yang berada di cluster 1 maka diberi tanda angka 1 di kolom C1, data yang berada pada cluster 2 diberi tanda angka 1 di kolom C2, dan data yang berada pada cluster 3 diberi tanda angka 1 di kolom C3, maka dapat dilihat pemetaan sebagai berikut:

1. Hasil perbandingan pemetaan iterasi 1 dan iterasi 2

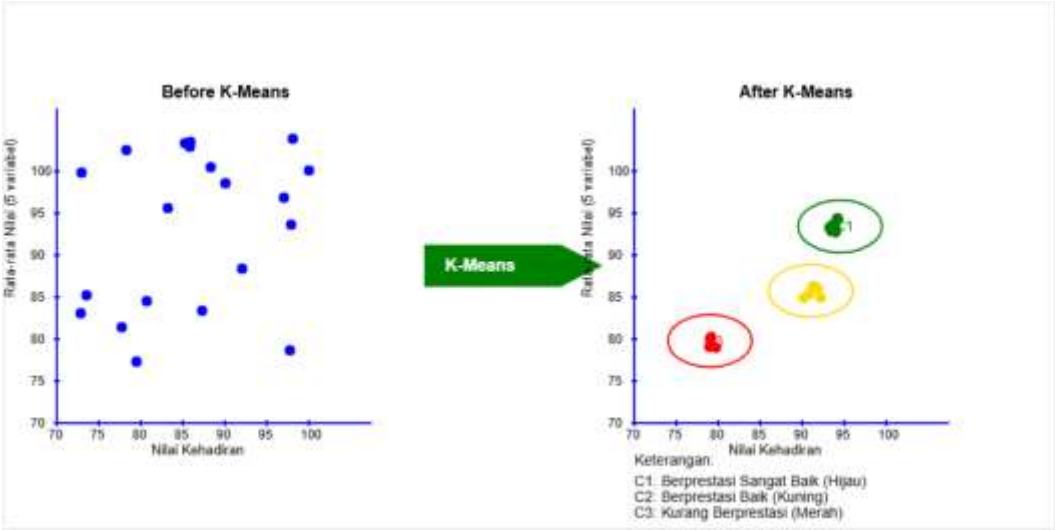
Tabel 6. Hasil K-Means

No	Nis	Nama	Kehadiran, Matematika, Bahasa Indonesia, Agama dan Ektrakurikuler	Cluster	Hasil
1	2182	Citra ayu kencana	98, 95, 90, 98, 95	1	Berprestasi Sangat Baik
4	2205	Aldiansyah	98, 90, 85, 98, 90	1	Berprestasi Sangat Baik
5	2207	Fadlan alathif	98, 100, 88, 98, 100	1	Berprestasi Sangat Baik
9	2224	Cinta nur ramadhani	94, 90, 90, 94, 90	1	Berprestasi Sangat Baik
10	2232	Aylin raisa a. Ginting	90, 90, 90, 90, 90	1	Berprestasi Sangat Baik
11	2233	Abdullah	95, 95, 90, 95, 95	1	Berprestasi Sangat Baik
14	2246	Alfah syahri r d f	95, 100, 90, 95, 100	1	Berprestasi Sangat Baik
16	2254	Abel jocelin br. Tobing	99, 90, 90, 99, 90	1	Berprestasi Sangat Baik
17	2255	Bastian marcello	85, 95, 85, 85, 95	1	Berprestasi Sangat Baik
18	2264	Delima mikhaila putri	95, 100, 90, 95, 100	1	Berprestasi Sangat Baik
20	2271	Ariel purba	95, 80, 85, 95, 80	2	Berprestasi Baik
2	2196	Ananda oktavianta	85, 75, 98, 85, 75	2	Berprestasi Baik

3	2200	Claudya nathaline zebua	98, 80, 80, 98, 80	2	Berprestasi Baik
15	2251	David sebastian	90, 85, 75, 90, 85	2	Berprestasi Baik
6	2211	Fardhan twieky swatman	95, 75, 90, 95, 75	2	Berprestasi Baik
7	2220	Ashera queenara eklesia m.	75, 80, 80, 75, 80	3	Kurang Berprestasi
8	2223	Avina syakilah	85, 85, 75, 85, 85	3	Kurang Berprestasi
12	2234	Andika pratama	75, 80, 80, 75, 80	3	Kurang Berprestasi
13	2235	Diego samuel malau	75, 85, 75, 75, 85	3	Kurang Berprestasi
19	2270	Adli fhairus sagala	85, 80, 75, 85, 80	3	Kurang Berprestasi

Berdasarkan hasil pemetaan Cluster 1 (C1) dengan pusat cluster (94.70, 94.50, 88.80, 94.70 dan 94.50) yang diartikan cluster Berpretasi Sangat Baik yaitu ada 10 siswa, siswa data ke 1,4,5,9,10,11,14,16,17,dan 18. Sedangkan Cluster 2 (C2) dengan pusat cluster (92.60, 79.00, 85.60, 92.60 dan 79.00) yang diartikan cluster Berpretasi Baik yaitu ada 5 siswa, siswa data ke 2,3,6, 15 dan 20. Sehingga Cluster 3 (C3) dengan pusat cluster (79.00, 82.00, 77.00, 79.00 dan 82.00) yang diartikan cluster Kurang Berpretasi yaitu ada 5 siswa, siswa data ke 7,8,12,13 dan 19. Dari tabel tersebut menunjukkan bahwa mayoritas siswa sudah berprestasi sangat baik, sedangkan sisanya terbagi rata antara kategori baik dan kurang berprestasi.

2. Grafik



Gambar 2. Grafik Hasil

Dari grafik tersebut menunjukkan bahwa mayoritas siswa sudah berprestasi sangat baik, sedangkan sisanya terbagi rata antara kategori baik dan kurang berprestasi.

PEMBAHASAN

Implementasi sistem dilakukan melalui serangkaian prosedur untuk merealisasikan rancangan yang telah disusun dan diuji, mulai dari instalasi hingga penerapan sistem guna memastikan aplikasi benar-benar dapat mencapai tujuan penelitian. Pada tahap ini, terdapat beberapa tampilan utama yang mendukung fungsionalitas sistem. Halaman login administrator menjadi gerbang awal bagi admin untuk masuk ke sistem dengan memasukkan username dan password yang valid. Setelah berhasil login, admin diarahkan ke halaman beranda yang berisi menu utama, seperti data siswa, data penilaian, proses K-Means, dan hasil pengelompokan. Halaman data siswa menampilkan seluruh informasi siswa yang tersimpan, dilengkapi fitur tambah, edit, dan hapus data, sedangkan formulir pengisian data siswa digunakan untuk memasukkan data baru yang akan diproses dalam pengelompokan K-Means. Selanjutnya, terdapat halaman nilai kriteria yang menampilkan data kriteria penilaian siswa dengan opsi tambah, ubah, dan hapus. Pada halaman proses K-Means, admin dapat menjalankan algoritma untuk melakukan pengelompokan berdasarkan data siswa dan kriteria yang ada, yang kemudian hasilnya ditampilkan dalam halaman hasil proses K-Means. Untuk memberikan visualisasi yang lebih jelas, sistem juga menyediakan halaman grafik yang menampilkan hasil pengelompokan dalam bentuk diagram sehingga memudahkan pihak sekolah dalam menganalisis prestasi siswa secara lebih objektif. Dengan implementasi ini, sistem siap digunakan sebagai alat bantu evaluasi akademik di sekolah.

1. Tampilan Halaman Login Administrator

   Halaman Login Administrator merupakan halaman login bagi administrator agar dapat memproses data. Administrator harus memasukkan username dan password yang benar agar dapat masuk dalam Beranda administrator.

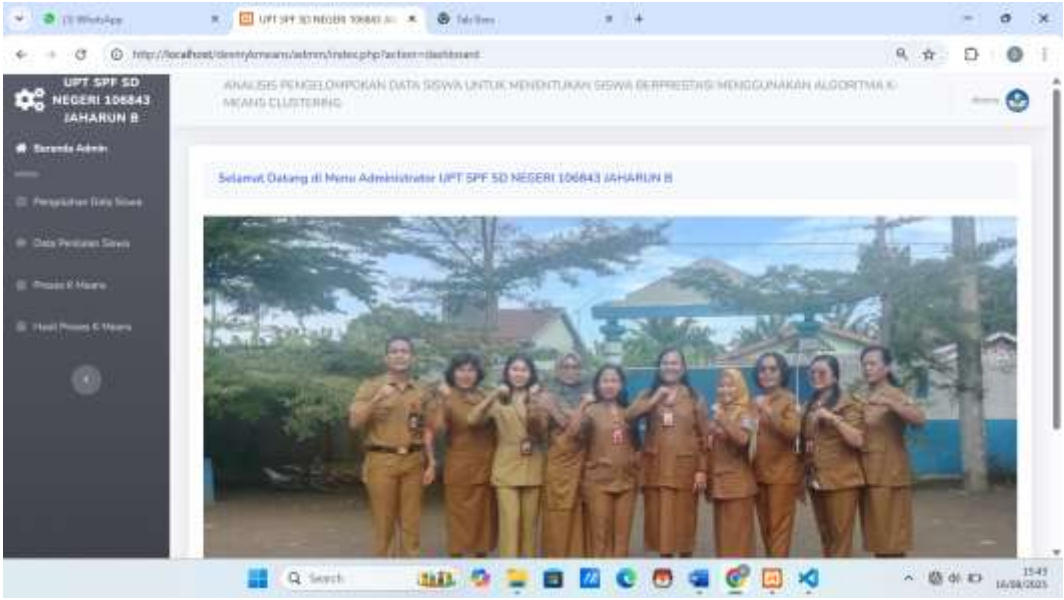


Gambar 5. 1 Halaman Login Administrator

Halaman ini digunakan sebagai formulir untuk melakukan login ke dalam sistem oleh administrator sistem dengan memasukkan username dan password yang benar oleh admin sistem.

2. Tampilan Halaman Beranda

   Halaman beranda merupakan halaman awal menu administrator.

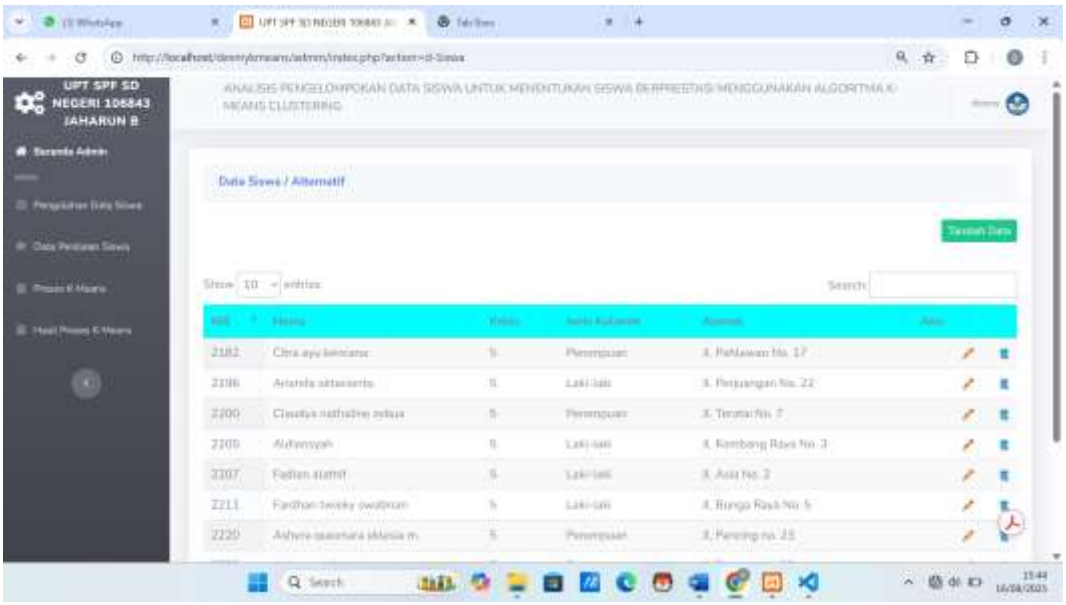


Gambar 5. 2 Halaman Beranda Administrator

Halaman ini berfungsi untuk menampilkan halaman aawal beranda administrator, halaman ini berisi menu beranda admin, data siswa, data penilaian siswa, proses K-Means dan hasil proses k-means.

3. Tampilan halaman data siswa

Halaman data siswa adalah halaman yang menampilkan data-data siswa serta tombol operasi seperti tambah data, edit dan hapus data siswa.



Gambar 5. 3 Tampilan data siswa

Halaman ini berfungsi untuk menampil data-data siswa yang sudah pernah dimasukkan ke dalam sistem, dan juga terpadat menu tambah data, edit data dan hapus sesuai dengan kebutuhan sistem.

4. Tampilan Hasil Proses K-Means

Setelah menekan menu hasil proses K_Means maka akan muncul hasil pengelompokan data seperti berikut ini dengan jumlah iterasi yang sesuai dengan data yang sudah dimasukkan ke dalam sistem baik itu jumlah siswa dan data nilai kriteria .

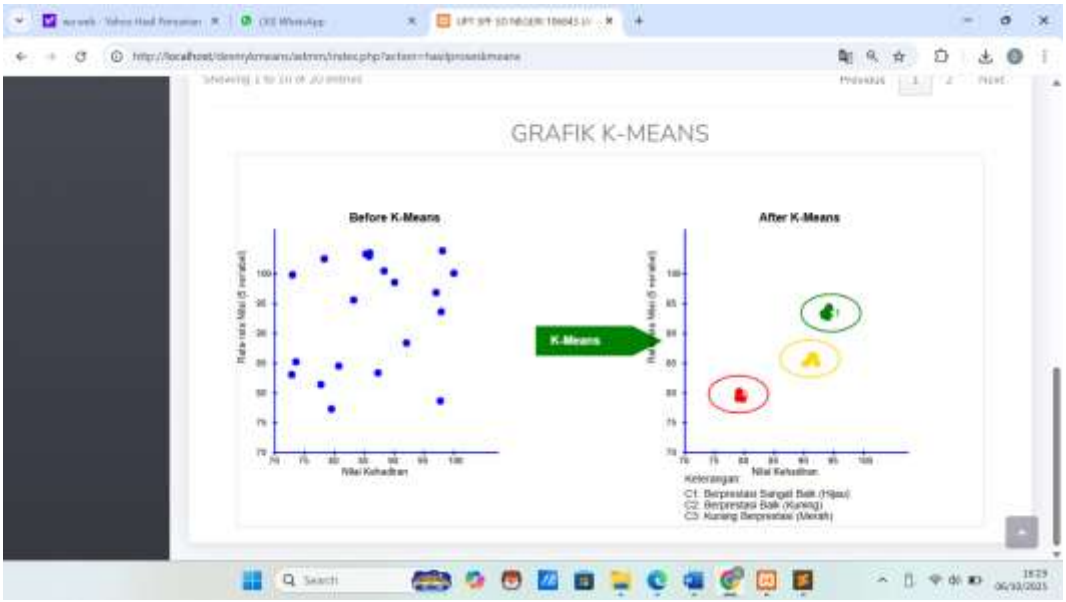


Gambar 5. 4 Hasil Proses K-Means

5. Tampilan Grafik

Setelah adanya hasil proses k-means maka muncul lah grafik dari hasil proses k-means dengan jumlah iterasi yang sesuai dengan data yang sudah dimasukkan ke dalam sistem baik itu jumlah siswa dan

data nilai kriteria .



Gambar 5. 5 Tampilan Grafik

KESIMPULAN

Dari hasil pembuatan tugas akhir *Analisis Pengelompokan Data Siswa untuk Menentukan Siswa Berprestasi Menggunakan Algoritma K-Means Clustering*, dapat disimpulkan bahwa algoritma K-Means berhasil diterapkan untuk mengelompokkan siswa berdasarkan variabel kehadiran, nilai matematika, bahasa Indonesia, agama, dan ekstrakurikuler. Proses pengelompokan berhenti pada iterasi kedua, yang menunjukkan bahwa data telah stabil dalam kelompoknya masing-masing, yaitu kategori berprestasi sangat baik, berprestasi baik, dan kurang berprestasi. Selain itu, sistem berbasis web yang dibangun dengan PHP dan MySQL telah mampu menjalankan fungsi pengelompokan secara otomatis, sehingga admin sekolah dapat dengan mudah memasukkan data serta memperoleh hasil pengelompokan secara cepat dan akurat.

Saran yang dapat diberikan untuk pengembangan lebih lanjut adalah agar pihak sekolah dapat memanfaatkan sistem pengelompokan siswa berbasis algoritma K-Means ini secara maksimal dalam proses evaluasi dan pembinaan, karena sistem ini mampu menjadi alat bantu yang objektif dalam menentukan siswa berprestasi. Bagi pengembang sistem berikutnya, disarankan agar aplikasi ini dapat dikembangkan ke platform mobile (Android) sehingga lebih fleksibel dan mudah diakses melalui berbagai perangkat. Selain itu, untuk penelitian lanjutan, pengembangan dapat dilakukan dengan membandingkan algoritma K-Means dengan metode lain seperti DBSCAN atau Hierarchical Clustering, guna mengetahui metode yang paling akurat dalam mengpengelompokankan prestasi siswa di lingkungan pendidikan dasar

REFERENCES

Abdul mukhsyi, S., Irma Purnamaari, A., Bahtiar, A., & Kaslani. (2025). Improving Student Achievement Clustering Model Using K-Means Algorithm in Pasundan Majalaya Vocational School. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, *4*(2), 977–985. https://doi.org/10.59934/jaiea.v4i2.793

Afniola, S., Ruslana, R., & Artika, W. (2020). Intelegensi Dan Bakat Pada Prestasi Siswa. *Al-Din: Jurnal Dakwah dan Sosial Keagamaan*, *6*(1), 1–10. https://doi.org/10.35673/ajdsk.v6i1.844

Andini, Y., Hardinata, J. T., & Purba, Y. P. (2022). Penerapan Data Mining Terhadap Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar Menggunakan Metode Apriori. *Jurnal TIMES*, *11*(1), 9–15. https://doi.org/10.51351/jtm.11.1.2022661

Apriandi, D., Sari, R. M., & Sarif, M. I. (2024). Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means. *Jurnal Minfo Polgan*, *13*(1), 1117–1129. https://doi.org/10.33395/jmp.v13i1.13959

Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, *07*(08), 64–69. https://doi.org/10.31695/ijasre.2021.34050

Penda Sudarto Hasugian, J. R. S. (2022). Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means. *KLIK: Kajian Ilmiah Informatika dan Komputer*, *3*(3), 262–268. https://djournals.com/klik

Srirahmawati, E. O., Purnamasari, A. I., Bahtiar, A., & Tohidi, E. (2025). Pengelompokan Prestasi Akademik Siswa Sd Menggunakan Algoritma K-Means. *Jurnal Informatika & Rekayasa Elektronika)*, *8*(1), 80–86. http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900

Sultan Alalawi, S. J., Mohd Shaharanee, I. N., & Mohd Jamil, J. (2023). CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS. *Journal of Computational Innovation and Analytics (JCIA)*, *2*(1), 41–55. https://doi.org/10.32890/jcia2023.2.1.3

Syukron Ramadani, M., & Fatah, Z. (2024). Riset Sistem Informasi Analisis Pengelompokan Data Nilai Siswa Untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Syukron Ramadani, Mochammad Fatah, Zaehol*, *1*(4), 103–110.

Widyanti, T., Hilabi, S. S., Hananto, A., Tukino, & Novalia, E. (2023). Implementasi K-Means dan K-Nearest Neighborspada Kategori Siswa Berprestasi. *Jurnal Informasi dan Teknologi*, *5*(1), 75–82. https://doi.org/10.37034/jidt.v5i1.255