

PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN KELAS BERDASARKAN NILAI AKADEMIK

Eko Harli¹, Ahmad Fauzi², Tria Hadi Kusmanto³

*Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI
Jalan Raya Tengan No 80, Kelurahan Gedong, Pasar Rebo, Jakarta Timur*
ekoharli@gmail.com, ahmadfauzi.udzi@gmail.com, tria.hd.kusmanto@gmail.com

ABSTRAK

Mengorganisir data ke dalam satu atau lebih *cluster*, algoritma K-Means dapat digunakan dalam proses data *clustering*. Data dengan representasi yang serupa ditempatkan dalam satu kelompok, sedangkan data dengan representasi yang berbeda ditempatkan di kelompok lain. Pengelompokan data siswa, yang terdiri dari kelompok atau kelas siswa unggul, sedang, dan rendah dapat digunakan untuk membantu sekolah mendukung anak-anak berdasarkan kapasitas belajar dan partisipasi mereka yang beragam dalam pembelajaran. Perhitungan klaster K-Means digunakan untuk menyajikan laporan hasil belajar siswa berdasarkan data siswa dari penilaian terpusat. Karakteristik yang digunakan dalam proses perhitungan untuk mengidentifikasi siswa dalam kelas unggul, sedang, dan rendah diperoleh dari penilaian ini. Tujuan proyek ini adalah menggunakan bahasa pemrograman Python untuk pengujian algoritma K-Means dalam menangani data penilaian terpusat guna menyajikan laporan hasil belajar siswa dan mengklasifikasikan siswa ke dalam kelompok kelas. Hal ini memungkinkan penanganan dan pengelompokan data siswa secara efisien dan efektif.

Kata Kunci: K-Means, Siswa, Kelas

ABSTRACT

The K-Means algorithm can be used to group data into one or more clusters. Data instances with similar representations are grouped into the same cluster, while those with dissimilar representations are assigned to different clusters. The clustering of student data, which consists of superior, average, and low-performing student groups or classes, can be utilized to assist schools in supporting students based on their diverse learning capacities and levels of participation in the learning process. K-means cluster computation is employed to present reports on student learning outcomes based on centralized assessment data. This assessment determines the characteristics that the computational process uses to identify students within superior, average, and low-performing classes. The objective of this project is to implement the K-Means algorithm using the Python programming language to evaluate its effectiveness in processing centralized assessment data for generating student performance reports and classifying students into appropriate class groups. This approach enables efficient and effective management and clustering of student data.

Key Word: K-Means, Student, Classes

PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah berdampak pada pertumbuhan dan kemajuan banyak bidang korporasi, termasuk pendidikan. Kemajuan ini mendorong inovasi berkelanjutan. Penciptaan dan penggunaan aplikasi baik dalam proses pendidikan maupun manajemen sekolah merupakan salah satu jenis inovasi dalam pendidikan (Hasugian & Sagala, 2022). Kelompok belajar siswa dibentuk di dalam kelas sesuai dengan kriteria siswa untuk meningkatkan kualitas pendidikan melalui layanan pembelajaran. Di kelas unggulan, data siswa diorganisasikan ke dalam kelompok siswa. Penentuan penerimaan ke kelas unggulan sangat penting karena data yang dihasilkan dapat digunakan

sebagai panduan oleh sekolah untuk menerapkan langkah-langkah terorganisir guna meningkatkan kualitas dan standar pengajaran di SMK Wira Buana 2. Skor penilaian digunakan untuk menentukan penerimaan siswa ke kelas unggulan. Berbagai atribut penilaian ditampilkan dalam hasil identifikasi data rapor. Secara umum, terdapat perbedaan kriteria untuk jumlah siswa yang cenderung sama di antara siswa individu yang memiliki skor serupa. Setiap tahun ajaran, penempatan siswa kelas unggulan dapat bervariasi berdasarkan jumlah siswa yang terdaftar dan pertumbuhan sekolah. Memecahkan masalah pengelompokan siswa dengan volume data yang terus meningkat menjadi kurang efektif,

dan agar dapat diterima di kelas unggulan yang ditentukan, siswa juga harus dikelompokkan berdasarkan prestasi tinggi. Siswa akan belajar lebih efektif jika mereka dibagi menjadi kelompok-kelompok yang sesuai.

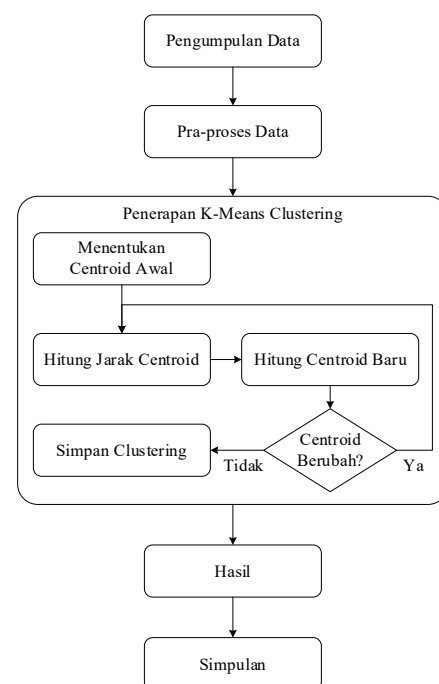
Sejak tahun 1990-an, penambangan data semakin populer karena meningkatnya penggunaan data di berbagai industri, termasuk bisnis dan pemasaran, penelitian dan teknik, serta seni dan hiburan. Menurut beberapa ahli, penambangan data, juga dikenal sebagai *Knowledge Discovery in Database* (KDD), adalah langkah dalam analisis pengetahuan dalam basis data (Mardi, 2016). Proses pengumpulan dan pemanfaatan data untuk mengidentifikasi pola atau hubungan dari kumpulan data yang besar dikenal sebagai penambangan data atau penemuan pengetahuan dalam basis data (KDD) (Tarigan et al., 2022). Pengambilan keputusan di masa depan dapat dievaluasi menggunakan hasil dari proses penambangan data (Harahap & Sulindawaty, 2019). Tujuan penambangan data adalah untuk mengekstrak, memeriksa, dan mengungkap pengetahuan tersembunyi dari suatu kumpulan data atau kumpulan data yang sangat besar. Pendekatan pengelompokan menggunakan algoritma K-Means adalah salah satu teknik yang dapat diterapkan pada penambangan data (Asmana et al., 2022). Teknik penambangan data yang disebut algoritma K-Means Clustering membagi data yang ada menjadi beberapa kluster, mengelompokkan data dengan kualitas serupa ke dalam satu kluster dan data dengan fitur yang berbeda ke dalam kluster lain. Menemukan pusat kluster (*centroid*) dan mencari anggota dari setiap kluster adalah proses yang dilakukan oleh algoritma K-Means.

Teknik penambangan data yang disebut K-Means dapat mengelompokkan atau mengelompokkan data ke dalam satu atau lebih kelompok sehingga data dengan kualitas serupa dikelompokkan bersama dan data dengan karakteristik berbeda dikelompokkan ke dalam berbagai kelompok (Dinata et al., 2020).

METODE PENELITIAN

Studi ini akan menjelaskan sejumlah teknik pengumpulan data, pengembangan perangkat

lunak dan pengujian sistem. Model dalam pengembangan sistem menggunakan waterfall model. Dalam penelitian yang dilakukan, *User Acceptance Testing* (UAT) digunakan sebagai uji coba sisi pengguna dan pengujian black-box sebagai uji fungsional untuk mengevaluasi sistem. Sementara itu, tinjauan pustaka, wawancara, dan observasi digunakan sebagai teknik pengumpulan data (Anggraini et al., 2025). Kerangka penelitian adalah tahapan yang diikuti seorang peneliti untuk menyelesaikan proyek penelitian. Langkah-langkah yang diambil dalam penelitian ini dapat dilihat pada Gambar 1 di bawah ini:.



Gambar 1. Kerangka Penelitian

1. Pengumpulan Data
Penelitian dimulai dari pengumpulan data siswa yaitu terhadap data diri siswa dan nilai akademik siswa yang diperoleh dari data nilai rapor siswa dan seleksi tes kemampuan siswa. Data akan digunakan dalam proses pengelompokan kelas siswa yang melalui tahap pra-proses data kemudian pengolahan menggunakan algoritma k-means clustering.
2. Pra-proses Data
Pra-proses Data merupakan langkah pertama dalam proses data *mining* untuk mengubah data mentah yang dikumpulkan dari berbagai sumber menjadi data yang lebih baik dan terhindar dari data *missing value*

- kemudian dilanjutkan ke tahap transformasi data (Awalia, 2024).
3. Penerapan K-Means Clustering
 Setelah data siap, selanjutnya pada tahap algoritma *KMeans Clustering* diterapkan menggunakan bahasa pemrograman Python. *K-Means Clustering* adalah algoritma yang digunakan untuk mengelompokkan data ke dalam *K cluster* berdasarkan karakteristik yang serupa. Dalam penelitian ini, nilai *K* ditentukan sebanyak 3 *cluster*, yaitu *cluster* unggul, sedang, dan rendah (Awalia, 2024).
 4. Hasil
 Dilakukan evaluasi pada proses clustering ditahap terakhir serta dilakukan analisa untuk memastikan bahwa pengelompokan yang dilakukan sesuai dengan tujuan penelitian yaitu pengelompokan kelas siswa menjadi 3 kelompok siswa.
 5. Simpulan
 Membentuk simpulan pada penelitian yang dilakukan yaitu penerapan algoritma k-means *clustering* dalam pengelompokan kelas siswa berdasarkan nilai akademik yang diharapkan lebih efektif dan menghemat waktu.

HASIL DAN PEMBAHASAN

Situasi seorang siswa dapat dipetakan untuk menilai apakah mereka termasuk dalam kelompok tertentu dengan mengelompokkan mereka berdasarkan prestasi akademik dan hasil tes kecerdasan. Selain itu, algoritma K-Means menggunakan jenis kelamin sebagai variabel tambahan dalam pengelompokan (Hasugian & Sagala, 2022).

Analisa

Informasi yang akan digunakan dalam prosedur perhitungan algoritma k-means berasal dari SMK Wira Buana 2 dan ditunjukkan pada tabel 1 di bawah ini.

Tabel 1. Data Awal

No	Nama	Kelas	Rapor	Tes	Sikap	JK
1.	Abdulah	1A	90	88	5	1
2.	Biah	1A	85	82	4	0
3.	Ciko	1A	78	75	4	1
4.	Dita	1B	70	72	3	0
5.	Eko	1B	68	65	3	1
6.	Fika	1B	60	62	2	0
7.	Gema	1A	58	60	2	1
8.	Hikmah	1B	50	55	2	0
9.	Ivan	1A	45	50	1	1
10.	Jelita	1B	40	45	1	0

Penerapan Algoritma K-Means Clustering

Identifikasi siswa dilakukan melalui segmentasi dengan menggunakan analisis kluster untuk mengelompokkan siswa. Analisis kluster dapat dilakukan dengan berbagai cara, mulai dari teknik sederhana hingga teknik kompleks yang menggunakan bantuan kecerdasan tiruan seperti jaringan saraf tiruan (Harli et al., 2016).

Teknik K-Means digunakan untuk mengelompokkan data sesuai dengan kelompok yang akan dibuat. Langkah-langkah dalam menggunakan metode K-Means adalah sebagai berikut:

1. Menentukan Jumlah *Cluster*
 Ketentuan untuk jumlah *cluster* dalam penelitian ini adalah 3 *cluster*, yaitu C1 = Unggul, C2 = Sedang, dan C3 = Rendah.
2. Menentukan Centroid Awal
 Centroid awal diperoleh dari data awal yang dipilih secara acak. Menyesuaikan dengan jumlah *cluster* sebanyak 3. Tabel 2 menampilkan centroid awal.

Tabel 2. Centroid Awal

Cluster	Nama	Rapor	Tes	Sikap	JK
C1	Abdulah	90	88	5	1
C2	Dita	70	72	3	0
C3	Hikmah	50	55	2	0

3. Menghitung Jarak
 Jarak dihitung menggunakan rumus Euclidean sebagai berikut:

$$d = \sqrt{(R_i - R_c)^2 + (T_i - T_c)^2 + (S_i - S_c)^2 + (JK_i - JK_c)^2} \quad (1)$$

4. Hitung Centroid Baru
 Jarak data awal 1 terhadap C1:
 $d = \sqrt{(90 - 90)^2 + (88 - 88)^2 + (5 - 5)^2 + (1 - 1)^2} = 0$

Jarak data awal 1 terhadap C2:

$$d = \sqrt{(20)^2 + (16)^2 + (2)^2 + (1)^2} = 25,71$$

Jarak data awal 1 terhadap C3:

$$d = \sqrt{(40)^2 + (33)^2 + (3)^2 + (1)^2} = 51,95$$

Hasil perhitungan untuk setiap data disajikan dalam tabel 3.

Tabel 3. Centroid Awal

Nama	d(C1)	d(C2)	d(C3)	Cluster
Abdulah	0.00	25.71	51.95	Unggul
Biah	7.55	17.20	44.17	Unggul
Ciko	17.69	7.48	34.37	Sedang
Dita	25.71	0.00	26.23	Sedang
Eko	31.80	7.94	20.43	Sedang
Fika	41.77	14.87	12.10	Rendah
Gema	44.50	17.65	9.31	Rendah
Hikmah	51.95	26.23	0.00	Rendah

Nama	d(C1)	d(C2)	d(C3)	Cluster
Ivan	58.88	33.54	7.17	Rendah
Jelita	65.92	40.85	14.19	Rendah

Perhitungan jarak yang dilakukan, diperoleh hasil pengelompokan data berdasarkan nilai terkecil dari masing-masing *cluster*.

Iterasi ke 2, nilai data cluster tidak berubah karena sama dengan iterasi 1 sehingga perhitungan dihentikan.

Pengujian K-Means

Penelitian yang dilakukan menggunakan bahasa pemrograman Python dalam pengelompokan siswa menggunakan algoritma K-Means Clustering. Sebanyak 10 data uji dalam pengelompokan, data tersebut dikomputasi untuk menentukan pengelompokan data berdasarkan 3 *cluster*.

Berikut tahapan yang dilakukan dalam perhitungan data menggunakan rumus k-means.

1. Pengumpulan Data

Gambar 2 menampilkan data awal siswa yakni nama, kelas, nilai rapor, nilai tes, nilai sikap dan jenis kelamin sebagai data input.

No	Nama	Kelas	Rapor	Tes	Sikap	JK
1	Abdulah	1A	90	88	5	Laki-laki
2	Biah	1A	85	82	4	Perempuan
3	Ciko	1A	78	75	4	Laki-laki
4	Dita	1B	70	72	3	Perempuan
5	Eko	1B	68	65	3	Laki-laki
6	Fika	1B	60	62	2	Perempuan
7	Gema	1A	58	60	2	Laki-laki
8	Hikmah	1B	50	56	2	Perempuan
9	Ivan	1A	45	50	1	Laki-laki
10	Jelita	1B	40	45	1	Perempuan

Gambar 2. Data Awal

2. Proses Data

Tahapan berikutnya yaitu perhitungan jarak centroid yang menggunakan rumus euclidean terhadap data awal yang telah dimasukan untuk menentukan cluster yang terdiri dari unggul, sedang dan rendah.

No	Nama	Kelas	Jarak Unggul	Jarak Sedang	Jarak Rendah	Cluster
1	Abdulah	1A	3.97	25.05	51.90	Unggul
2	Biah	1A	3.97	17.27	44.17	Unggul
3	Ciko	1A	13.81	7.44	34.37	Sedang
4	Dita	1B	21.86	2.52	26.23	Sedang
5	Eko	1B	27.98	6.95	20.43	Sedang
6	Fika	1B	35.94	14.88	12.10	Rendah
7	Gema	1A	38.75	17.65	9.51	Rendah
8	Hikmah	1B	48.09	27.05	1.62	Rendah
9	Ivan	1A	55.17	34.08	7.17	Rendah
10	Jelita	1B	62.20	41.09	14.19	Rendah

Gambar 3. Jarak ke Centroid

Pengujian dilakukan sampai iterasi ke 3 dan menampilkan proses perhitungan yang sama mulai dari iterasi ke 2 sampai iterasi ke 3. Gambar 4 menampilkan centroid per iterasi.

Iterasi	Centroid Unggul				Centroid Sedang				Centroid Rendah			
	Rapor	Tes	Sikap	JK	Rapor	Tes	Sikap	JK	Rapor	Tes	Sikap	JK
0	90.00	88.00	5.00	1.00	70.00	72.00	3.00	0.00	50.00	55.00	2.00	0.00
1	87.50	85.00	4.50	0.50	72.00	70.87	3.33	0.87	50.60	54.40	1.60	0.40
2	87.50	85.00	4.50	0.50	72.00	70.87	3.33	0.87	50.60	54.40	1.60	0.40

Gambar 4. Centroid per Iterasi

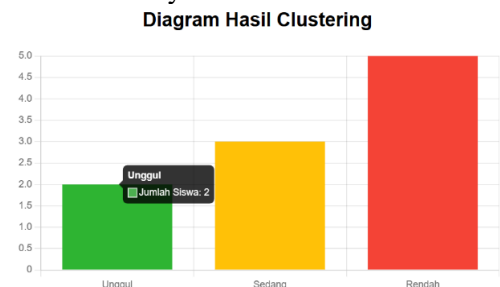
3. Hasil Pengujian

Clustering dari 10 data dengan 3 *cluster* yaitu unggul, sedang, dan rendah diperoleh 2 siswa termasuk dalam *cluster* unggul, 3 siswa termasuk dalam *cluster* sedang, dan 5 siswa termasuk dalam *cluster* rendah. Mayoritas siswa pada data diatas berada pada *cluster* rendah. Gambar 5 menampilkan data hasil clustering.

No	Nama	Kelas	Rapor	Tes	Sikap	Cluster
1	Abdulah	1A	90	88	5	Unggul
2	Biah	1A	85	82	4	Unggul
3	Ciko	1A	78	75	4	Sedang
4	Dita	1B	70	72	3	Sedang
5	Eko	1B	68	65	3	Sedang
6	Fika	1B	60	62	2	Rendah
7	Gema	1A	58	60	2	Rendah
8	Hikmah	1B	50	55	2	Rendah
9	Ivan	1A	45	50	1	Rendah
10	Jelita	1B	40	45	1	Rendah

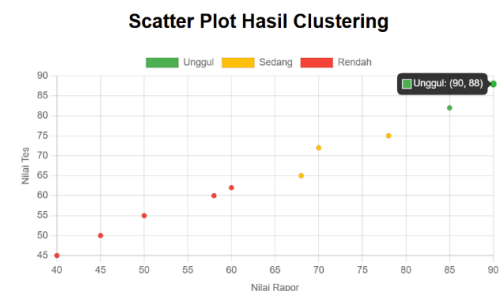
Gambar 5. Hasil Clustering

Gambar 6 menampilkan diagram batang hasil clustering dengan skala satuan diperoleh cluster unggul sebanyak 2, cluster sedang sebanyak 3, dan cluster rendah sebanyak 5.



Gambar 6. Diagram Hasil Clustering

Grafik hasil *clustering* ditampilkan pada gambar 7 disajikan dalam bentuk *scatter plot* berdasarkan nilai rapor dan nilai tes skala puluhan. Grafik menampilkan skala tinggi pada *cluster* unggul 90 dan 80 terhadap nilai rapor dan nilai tes.



Gambar 7. Scatter Plot Clustering

SIMPULAN DAN SARAN

Penerapan tahapan dengan algoritma k-means clustering untuk menggabungkan data atau mengelompokkan data siswa ke dalam tiga kelompok yakni kelompok siswa unggul, kelompok menengah, dan kelompok rendah berdasarkan nilai akademik diperoleh simpulan dimana hasil yang sama diperoleh ketika algoritma k-means dihitung secara manual dan diuji menggunakan aplikasi yang dibangun menggunakan bahasa pemrograman Python dari data nilai siswa, dimana *cluster* 1 (C1) merupakan kelompok unggul dengan dua data siswa. *Cluster* 2 (C2) adalah kelompok menengah dengan tiga data siswa. Kemudian terdapat lima data siswa di *cluster* 3 (C3) yang merupakan kelompok rendah. Hasil klasifikasi pengelompokan data pada proses *clustering* memberikan hasil yang efektif dan lebih hemat waktu dalam pengelompokan kelas siswa.

DAFTAR PUSTAKA

- Anggraini, F., Suprpta, I., Warno, W., & Fahrezi, H. (2025). Sistem Manajemen Kelas Menggunakan K-Means Clustering untuk Pengelompokan Kelas Unggulan pada Sekolah Dasar Negeri Neglasari. *Jurnal Kridatama Sains dan Teknologi*, 7(02), 720–738.
- Asmana, Wijaya, Y. A., & Martanto. (2022). K-Means di Sekolah Menengah Kejuruan Wahidin Kota Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 552–559.
- Awalia, R. (2024). Penerapan Metode K-Means Clustering untuk Pengelompokan Prestasi Siswa menggunakan Orange Data Mining: Studi Kasus di MTs Muhammadiyah Tallo Makassar. *MAPLE*, 6(2), 36–41.
- Dinata, R. K., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. *INFORMAL Informatics Journal*, 5(1), 10–17.
- Harahap, P. N., & Sulindawaty. (2019). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah). *MATICS*, 11(2), 46–50.
- Harli, E., Fauzi, A., & Kusmanto, T. H. (2016). *Pengelompokan Kelas Menggunakan Self Organizing Map*

Neural Network pada SMK N 1 Depok. 2(2), 90–95.

- Hasugian, P. S., & Sagala, J. R. (2022). *Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means*. 3(3), 262–268.
- Mardi, Y. (2016). Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Tarigan, P. M. S., Hardinata, J. T., Qurniawan, H., Safii, M., & Winanjaya, R. (2022). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap) Implementation of Data Mining Using Apriori Algorithm in Determining Inventory (Case Study: Sinar Harahap Store). *Jurnal Janitra Informatika Dan Sistem Informasi*, 2(1), 9–19. <https://doi.org/10.25008/janitra.v2i1.14>