



ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN SISWA UNGGULAN BERDASARKAN HASIL UJIAN DI SEKOLAH

Ainul Fadil^{a*}, Zaehol Fatah^b

^a Sains Dan Teknologi / Teknologi Informasi, ainulfadil765@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

^b Sains Dan Teknologi / Teknologi Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur

* korespondensi

ABSTRACT

Determining classes for outstanding students based on exam results is a crucial step in promoting the improvement of learning quality. This study applied a data mining method using the K-Means Clustering algorithm to group students based on their exam results. The process includes collecting exam score data, preprocessing the data, and applying the K-Means algorithm to form several student groups based on their achievement levels. Through this algorithm, students are clustered into groups with similar characteristics, such as excellent, average, and those requiring more attention. The study's results indicate that the K-Means Clustering approach can provide an accurate representation of the distribution of student abilities, serving as a basis for designing more effective and equitable learning strategies. This implementation is expected to help schools identify students' potential more objectively and enhance overall educational quality.

Keywords: *Data Mining, K-Means Clustering, Outstanding Students, Exam Results, Education.*

Abstrak

Penentuan kelas siswa unggulan berdasarkan hasil ujian merupakan langkah penting untuk mendorong peningkatan kualitas pembelajaran. Dalam penelitian ini, diterapkan metode data mining menggunakan algoritma K-Means Clustering untuk mengelompokkan siswa berdasarkan hasil ujian mereka. Proses ini mencakup pengumpulan data nilai ujian, pra-pemrosesan data, dan penerapan algoritma K-Means untuk membentuk beberapa kelompok siswa berdasarkan tingkat pencapaian mereka. Dengan algoritma ini, siswa dikelompokkan ke dalam kluster dengan karakteristik serupa, seperti kelompok unggul, sedang, dan membutuhkan perhatian lebih. Hasil penelitian menunjukkan bahwa pendekatan K-Means Clustering dapat memberikan gambaran yang akurat mengenai distribusi kemampuan siswa, yang berguna sebagai dasar dalam menyusun strategi pembelajaran yang lebih efektif dan berkeadilan. Implementasi ini diharapkan dapat mendukung sekolah dalam mengidentifikasi potensi siswa secara lebih objektif dan meningkatkan mutu pendidikan secara keseluruhan.

Kata Kunci: *Data Mining, K-Means Clustering, Siswa Unggulan, Hasil Ujian, Pendidikan.*

1. PENDAHULUAN

Dalam era digital, teknologi informasi telah menjadi bagian tak terpisahkan dari berbagai bidang, termasuk dunia pendidikan. Sekolah sebagai institusi pendidikan menghasilkan banyak data, mulai dari nilai ujian, tingkat kehadiran, hingga aktivitas belajar siswa. Namun, sering kali data tersebut tidak dimanfaatkan secara optimal untuk mendukung pengambilan keputusan, khususnya dalam menentukan siswa unggulan. Pendekatan konvensional, seperti evaluasi manual oleh guru, memiliki kelemahan, termasuk memakan waktu, kurang efisien, dan rentan terhadap subjektivitas. Oleh karena itu, diperlukan solusi berbasis teknologi yang mampu menganalisis data secara objektif dan akurat.

Salah satu pendekatan yang relevan adalah data mining, yaitu serangkaian proses untuk menemukan informasi atau pengetahuan baru dari kumpulan data besar yang sebelumnya tidak diketahui. Proses ini melibatkan ekstraksi data dan pemilihan atribut penting yang dapat memengaruhi prediksi. Dengan menggunakan teknik statistik, kecerdasan buatan, dan pembelajaran mesin, data mining mampu mengolah data menjadi informasi yang berguna (Hendrastuty Nirwana et al., 2024; Yogiarto et al., 2024). Teknologi ini banyak digunakan untuk menganalisis data dalam jumlah besar dan mendukung pengambilan keputusan yang strategis (Fatah et al., 2024). Data mining juga membantu mengidentifikasi pola atau hubungan tersembunyi dalam data (Praja et al., 2023).

Salah satu algoritma yang sering digunakan dalam data mining adalah Algoritma K-Means. Algoritma ini termasuk metode klusterisasi yang bekerja dengan cara mengelompokkan data berdasarkan kedekatannya dengan titik pusat kluster (centroid). Tujuan dari algoritma ini adalah mengelompokkan data dengan tingkat kemiripan maksimum dalam kluster yang sama, sambil meminimalkan kemiripan antara kluster yang berbeda. Prosesnya menggunakan fungsi jarak sebagai ukuran kemiripan, di mana jarak terpendek antara data dan centroid menentukan pengelompokan (Jeanne Clarisa Wetik, 2017).

Sebelum proses data mining dilakukan, tahap data cleaning menjadi langkah penting dalam Knowledge Discovery in Databases (KDD). Data cleaning adalah proses untuk mendeteksi, memperbaiki, atau menghapus data yang tidak akurat, tidak lengkap, tidak konsisten, atau salah dalam kumpulan data, tabel, maupun database. Data yang bermasalah, sering disebut dirty data, dapat diganti, dimodifikasi, atau dihapus sesuai kebutuhan. Tahap ini sangat penting untuk mencegah duplikasi data, menghindari ambiguitas, serta memperbaiki kesalahan data sehingga kualitas data yang diolah menjadi lebih baik (Novi & Mubarak, 2021).

Algoritma K-Means pertama kali diperkenalkan oleh J.B. MacQueen pada tahun 1976 dan telah menjadi salah satu metode klusterisasi paling populer. Teknik ini digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Data dengan kesamaan karakteristik ditempatkan dalam satu kelompok (cluster), sementara data dengan perbedaan karakteristik berada di cluster lain. K-Means Clustering adalah teknik klusterisasi non-hierarki yang bertujuan membagi data ke dalam sejumlah kelompok (K) yang telah ditentukan sebelumnya. Proses algoritma ini dimulai dengan pemilihan acak K titik pusat cluster, kemudian setiap data dimasukkan ke dalam cluster dengan pusat terdekat (Butsianto Sufajar et al., 2021; Novi & Mubarak, 2021).

Dengan memanfaatkan algoritma K-Means Clustering, data hasil ujian siswa dapat dikelompokkan menjadi kategori seperti unggul, sedang, dan rendah. Pengelompokan ini tidak hanya membantu sekolah dalam mengidentifikasi siswa unggulan, tetapi juga memberikan wawasan untuk merancang strategi pembelajaran yang lebih tepat sasaran bagi semua kategori siswa.

2. TINJAUAN PUSTAKA

Sebelum proses data mining dilakukan, tahap data cleaning menjadi langkah penting dalam Knowledge Discovery in Databases (KDD). Data cleaning adalah proses untuk mendeteksi, memperbaiki, atau menghapus data yang tidak akurat, tidak lengkap, tidak konsisten, atau salah dalam kumpulan data, tabel, maupun database. Data yang bermasalah, sering disebut dirty data, dapat diganti, dimodifikasi, atau dihapus sesuai kebutuhan. Tahap ini sangat penting untuk mencegah duplikasi data, menghindari ambiguitas, serta memperbaiki kesalahan data sehingga kualitas data yang diolah menjadi lebih baik [1].

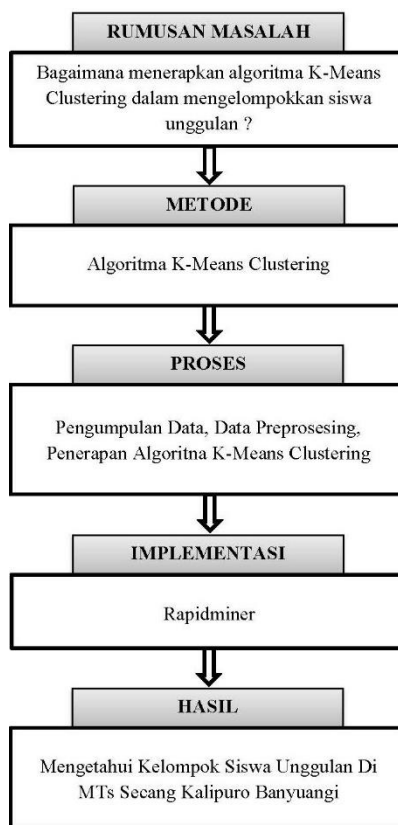
Algoritma K-Means pertama kali diperkenalkan oleh J.B. MacQueen pada tahun 1976 dan telah menjadi salah satu metode klusterisasi paling populer. Teknik ini digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Data dengan kesamaan karakteristik ditempatkan dalam satu kelompok (cluster), sementara data dengan perbedaan karakteristik berada di cluster lain. K-Means Clustering adalah teknik klusterisasi non-hierarki yang bertujuan membagi data ke dalam sejumlah kelompok (K) yang telah ditentukan sebelumnya. Proses algoritma ini dimulai dengan pemilihan acak K titik pusat cluster, kemudian setiap data dimasukkan ke dalam cluster dengan pusat terdekat [2][1].

Dengan memanfaatkan algoritma K-Means Clustering, data hasil ujian siswa dapat dikelompokkan menjadi kategori seperti unggul, sedang, dan rendah. Pengelompokan ini tidak hanya membantu sekolah dalam mengidentifikasi siswa unggulan, tetapi juga memberikan wawasan untuk merancang strategi pembelajaran yang lebih tepat sasaran bagi semua kategori siswa.

3. METODOLOGI PENELITIAN

3.1 Alur kerja penelitian

Alur kerja penelitian merupakan beberapa tahapan yang dilakukan oleh peneliti dalam menyelesaikan penelitian. Berikut adalah beberapa Langkah yang dilakukan di dalam penelitian ini dapat di lihat pada gambar 1 berikut ini:



Gambar 1. Alur Kerja Penelitian

Berdasarkan gambar diatas didapat permasalahan yang ada sebagai latar belakang penelitian dan tujuan yang akan dicapai. Algoritma yang akan digunakan adalah K-Means dengan bantuan tools RapidMiner Studio. Hasil yang didapatkan dari penelitian ini adalah mengetahui kelompok Siswa unggulan.

3.2 Objek Penelitian

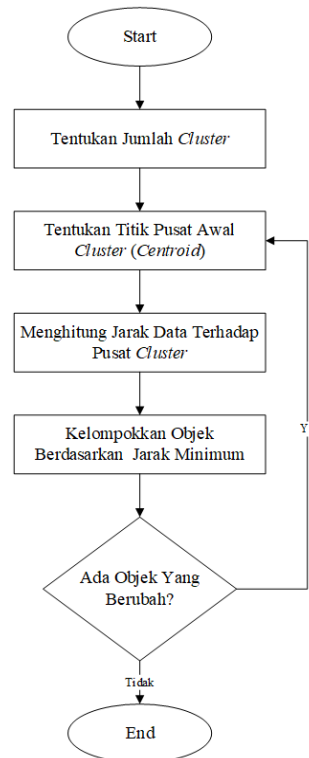
Objek penelitian ini yaitu berupa data nilai siswa tahun ajaran 2022/2023 yang didapatkan dari MTs Secang Kalipuro Banyuwangi. Dimana peneliti bertujuan untuk mengetahui hasil dari pengelompokkan nilai siswa berdasarkan cluster yang sudah ditentukan.

3.3 Objek Penelitian

Dalam penelitian ini, jenis data yang digunakan adalah data primer, karena data tersebut diambil dari MTs Secang Kalipuro Banyuwangi. Data primer adalah data yang diperoleh langsung dari sumber asli oleh individu atau organisasi untuk tujuan penelitian tertentu. Data ini dihasilkan melalui berbagai metode, seperti wawancara langsung dengan pihak-pihak yang relevan, seperti pimpinan atau staf yang terlibat langsung dalam masalah yang diteliti, serta melalui observasi kegiatan sehari-hari dari objek penelitian. Pendekatan ini memastikan data yang dikumpulkan sesuai dengan kebutuhan spesifik studi yang sedang dilakukan[3]. Data yang digunakan dalam penelitian ini merupakan data nilai siswa kelas IX MTs Secang Kalipuro Banyuwangi semester genap tahun ajaran 2022/2023 yang terdiri dari 74 record data. Variabel data yang digunakan dalam penelitian ini adalah nama siswa dan nilai dari tiap pelajaran yang diikuti oleh siswa.

3.4 Metode Yang Digunakan

Metode yang digunakan pada penelitian ini adalah menggunakan metode K- Means Clustering. Langkah-langkah melakukan clustering dengan algoritma K-Means sebagai berikut:



Gambar 2. Metode Yang Digunakan

Berikut uraian diagram diatas:

- Menentukan jumlah cluster.
- Menentukan centroid awal secara acak.
- Menghitung jarak setiap data yang ada terhadap setiap pusat cluster.
- Mengelompokkan data hasil perhitungan jarak data pada centroid.
- Menentukan centroid baru dari hasil pengklasteran.
- Menghitung kembali jarak setiap data menggunakan pusat cluster yang baru.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengujian

4.1.1 Persiapan Dataset

Penelitian ini menggunakan data nilai siswa MTs Secang Kalipuro Banyuangi kelas IX –semester Genap tahun ajaran 2022/2023. Dataset terdiri dari atribut yang telah disederhanakan berupa nama siswa dan nilai siswa untuk 17 mata pelajaran yang telah ditempuh selama satu semester. Dataset sampel terdiri dari 74 record data. Berikut adalah data siswa yang akan digunakan untuk proses perhitungan dengan algoritma K-Means:

Tabel 1. Dataset

Nama	QH	AA	FIK	SKI	PPKn	BIND	BAR	MTK	IPA	IPS	BING	SB	PJOK	PRKTI	Aswaj	BK	TIK
ADITIA PU	84	83	83	88	83	87	87	80	83	82	83	86	82	89	80	86	86
ALFIN DW	80	84	81	87	81	84	85	80	80	82	76	86	84	81	80	80	86
ANDIKAP	80	82	85	84	81	84	84	79	83	82	76	87	80	86	80	84	86
APRIYAN	80	80	80	82	80	84	84	78	85	79	76	85	81	81	80	82	86
ARILMULI	87	82	82	87	83	87	88	81	84	85	81	87	85	86	81	86	87
DIMASSRI	80	85	84	86	83	86	87	82	82	85	81	86	84	86	80	84	86
FAJRIBAI	83	81	87	84	82	85	88	80	80	82	81	86	81	86	80	84	86
FIKRI NAB	80	82	81	81	82	83	86	80	83	81	76	85	81	81	80	80	86
KUSVAMI	80	81	83	83	80	83	85	79	80	78	76	86	82	81	80	82	86
M. HARIYA	80	80	81	81	79	83	84	79	81	79	76	85	80	81	80	84	86
M. MURSI	91	83	90	85	81	85	87	81	80	81	83	86	83	89	83	85	87
MOH. ALD	80	81	83	83	80	83	84	79	81	83	81	84	81	81	80	80	86
MOH. RIZI	89	86	91	88	83	87	90	84	88	85	81	86	85	87	81	87	86
MOHAMAD	80	83	82	84	85	86	86	81	80	83	81	88	84	89	80	87	86
MOHAMAM	80	84	82	88	83	87	86	80	85	84	81	88	83	89	80	87	86
MOHAMAM	91	84	91	85	84	86	88	81	85	85	86	87	81	90	82	86	86
MUHAMAM	80	84	90	89	83	87	87	80	84	84	81	87	84	85	80	82	86
MUHAMAM	80	81	89	86	83	84	85	80	84	77	76	86	84	88	80	83	86
QUSDAFI	80	83	80	84	80	84	82	78	81	81	76	86	81	81	80	76	86
SAWISLILA	80	81	88	84	83	84	86	81	81	80	76	86	81	84	80	86	86
SLAMETDI	80	84	90	86	83	86	86	80	84	81	85	87	81	86	80	82	86
TRI SUMA	82	82	89	84	82	86	86	81	80	85	87	84	83	89	80	82	87

TRI SUMAI	82	82	89	84	82	86	86	81	80	85	87	84	83	89	80	82	87
ACH. FAH	83	82	86	85	82	85	85	81	83	79	76	86	82	86	80	80	87
AHMAD S	80	85	88	87	83	85	85	80	85	83	83	87	84	89	80	86	86
AHMAD S	81	84	92	85	83	86	87	82	85	76	87	87	82	89	80	86	86
AHMAD D	80	83	92	85	81	86	83	84	85	82	83	87	82	89	80	85	86
AHMAD H	80	84	91	86	82	87	87	81	85	86	83	87	83	89	80	86	86
AHMAD SI	80	82	81	84	81	84	84	79	81	77	76	84	80	81	80	80	86
ANDREAN	80	84	89	87	82	86	86	80	80	75	83	83	82	83	80	82	86
DAFFA EK	80	85	91	86	81	84	85	80	82	84	86	88	83	89	80	87	86
DIESTHA C	80	84	80	84	80	85	84	79	81	78	76	83	81	81	80	80	86
DIMAS RE	80	83	86	82	81	84	85	80	83	77	77	85	82	85	80	84	86
DIRGA	80	82	85	82	83	84	86	81	82	76	76	84	82	83	80	82	86
FERDIANS	80	82	81	80	82	83	84	79	85	81	76	85	80	81	80	78	86
M. KHAI RI	80	85	92	84	83	87	87	83	85	83	76	88	85	87	80	84	86
MUHAWIM	80	83	84	82	81	84	85	80	80	75	76	86	83	83	80	85	86
MUHAWIM	80	80	87	85	83	84	86	80	82	81	76	86	81	84	80	80	86
RADEN BA	80	82	84	82	82	83	85	80	83	77	76	86	82	83	80	80	86
RYANDIKI	82	84	90	85	83	85	86	81	85	76	76	86	82	86	80	87	87
SAIPULAN	80	82	82	84	81	83	85	80	83	77	81	85	81	83	80	80	86
SYAMSULLI	80	83	89	88	83	85	85	80	82	79	81	86	83	85	80	82	86
TALRIKUR	80	82	91	84	82	85	86	80	85	81	76	85	83	86	80	83	87
ZAKI KHOI	81	84	90	85	83	87	86	81	84	82	83	85	84	87	80	84	86
ALFI NURI	80	80	84	85	82	83	86	79	83	83	77	86	80	81	80	80	86
ALIKANUR	95	83	94	93	85	89	90	86	88	92	88	89	84	90	85	89	88
ANI LESTA	80	83	90	85	81	85	85	84	86	89	83	87	83	89	80	86	87
ANNISA PI	81	81	89	88	82	86	87	81	85	88	83	87	83	88	80	85	87
AULYANI	80	81	86	85	82	85	87	80	84	85	83	86	81	88	80	85	86
DEVI AMAN	80	80	87	87	81	85	86	80	84	83	76	86	81	85	80	78	86
DEWI NUR	80	82	85	85	82	86	86	80	86	85	81	86	80	88	80	86	86
DYAHANU	80	80	91	91	82	86	88	83	86	86	83	87	84	89	80	85	87
FITRI NUR	80	82	88	85	82	86	89	80	86	83	83	86	81	85	80	86	86
FITRIADA	80	81	86	86	82	85	85	80	88	84	83	86	81	84	80	82	86
FITRIYANI	80	83	90	88	81	86	88	81	88	87	83	86	83	88	80	85	86
GHONIYA	80	82	89	88	82	87	89	80	88	86	83	86	81	84	80	84	87
HINTAN NU	80	84	93	93	82	87	87	86	88	93	88	89	84	90	80	89	86
KHUSNUL	80	83	93	93	83	88	88	83	84	92	87	88	84	86	80	88	87
KUSNIAH	90	84	92	91	84	88	87	83	85	90	89	88	84	90	84	88	87
LAILATULI	80	82	89	87	82	86	85	80	82	87	77	86	81	86	80	86	86
MAULIDA	80	80	88	85	82	84	87	81	83	87	81	86	80	86	80	82	86
MAULIDAT	85	85	93	94	85	90	87	90	88	95	89	89	85	90	84	90	88
NAILA FITRI	86	82	88	88	83	85	87	80	85	83	83	86	82	84	80	85	86
PUTRI HAR	86	84	94	94	84	89	88	87	84	94	88	89	83	90	83	90	88
RIFKA ALF	80	82	88	86	82	86	87	81	84	87	77	87	81	84	80	84	86
RISALATU	80	82	92	85	81	87	86	81	88	90	81	86	82	87	80	86	86
ROHMAIN	80	83	91	84	81	88	89	80	83	87	81	86	82	86	80	85	86
ROSDATI	80	81	87	85	82	85	89	81	84	87	81	87	81	85	80	85	86
SALSABILA	82	80	91	88	82	86	88	81	86	90	81	87	84	85	80	87	86
SELIANA	82	82	92	87	83	88	88	82	86	92	88	87	83	90	80	86	87
SHINTANI	80	82	87	85	83	85	88	80	88	86	81	86	81	83	80	84	86
SUCI RAMI	82	82	87	84	82	86	88	80	83	86	77	87	81	88	80	85	86
SUCI WULU	80	83	88	85	81	86	88	81	83	89	81	85	81	89	80	85	86
TANTRI RI	80	83	89	92	83	88	86	85	84	86	85	84	82	87	80	85	86

Sumber : MTs secang kalipuro [1]

4.1.2 Menentukan Jumlah Cluster

Pada penelitian ini, jumlah cluster yang akan dibentuk adalah tiga cluster, sehingga nilai k = 3. Cluster yang akan dibentuk yaitu cluster_1 (C1) diartikan sebagai tinggi, cluster_2 (C2) diartikan sebagai cukup, dan cluster_0 (C3) diartikan sebagai rendah.

Tabel 2. Titik Pusat Awal Cluster

Centroid	QH	AA	FIK	SKI	PPKn	BIND	BAR	MTK	IPA
C1	81	83	90	86	82	86	87	81	84
C2	80	82	84	84	81	84	85	80	82
C3	85	84	93	92	84	88	88	85	86

4.1.3 Menghitung Jarak Data ke Pusat Cluster

Langkah selanjutnya adalah menghitung jarak setiap data ke pusat cluster dengan menggunakan rumus Euclidean Distance. Proses ini dilakukan untuk menentukan jarak terdekat antara data dan nilai awal cluster, yang dirumuskan sebagai berikut[4]:

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

a. Data pertama

Jarak dengan pusat cluster 1:

$$d1 = \sqrt{(81 - 81)^2 + (83 - 83)^2 + (90 - 90)^2 + (86 - 86)^2 + (82 - 82)^2 + (86 - 86)^2 + (87 - 87)^2 + (81 - 81)^2 + (84 - 84)^2}$$

= 0

b. Jarak dengan pusat *cluster* 2:

$$d1 = \sqrt{(81 - 80)^2 + (83 - 82)^2 + (90 - 84)^2 + (86 - 84)^2 + (82 - 81)^2 + (86 - 84)^2 + (87 - 85)^2 + (81 - 80)^2 + (84 - 82)^2}$$

$$= 7.48.$$

c. Jarak dengan pusat *cluster* 3:

$$d1 = \sqrt{(81 - 85)^2 + (83 - 84)^2 + (90 - 93)^2 + (86 - 92)^2 + (82 - 84)^2 + (86 - 88)^2 + (87 - 88)^2 + (81 - 85)^2 + (84 - 86)^2}$$

$$= 9.53.$$

Perhitungan dilakukan hingga semua data selesai diproses. Hasil penghitungan jarak setiap data terhadap masing-masing cluster pada iterasi pertama dapat dilihat berikut ini:

Tabel 3. Hasil Perhitungan Terhadap Cluster Pada Iterasi ke-1

Nama	d(C1)	d(C2)	d(C3)	Cluster	Jarak Terdekat
ADITIA PUTRA	8.366.666.666.666.660	8.214.285.714.285.710	88.0	C3	88.0
ALFIN DWI RADITYA	8.066.666.666.666.660	8.071.428.571.428.570	8.428.571.428.571.420	C1	8.066.666.666.666.660
ANDIKA PANDU MAULANA	83.0	7.971.428.571.428.570	8.514.285.714.285.710	C2	7.971.428.571.428.570
APPRIYANTO	82.0	7.928.571.428.571.420	8.314.285.714.285.710	C2	7.928.571.428.571.420
ARIL MAULANA PUTRA	85.0	8.285.714.285.714.280	8.771.428.571.428.570	C2	8.285.714.285.714.280
DIMAS SANTOSO	8.366.666.666.666.660	8.214.285.714.285.710	8.728.571.428.571.420	C2	8.214.285.714.285.710
FAJAR BAITUL MAKMUN	82.0	8.114.285.714.285.710	86.0	C2	8.114.285.714.285.710
FIKRI NABILUL MUSTHOFA	8.133.333.333.333.330	8.014.285.714.285.710	8.328.571.428.571.420	C2	8.014.285.714.285.710
KUSMAWIYANTO	80.0	7.971.428.571.428.570	8.385.714.285.714.280	C2	7.971.428.571.428.570
M. HARIYADI	8.133.333.333.333.330	7.914.285.714.285.710	83.0	C2	7.914.285.714.285.710

M. MURSID	82.0	8.357.142.857.1 42.850	87.0	C2	8.357.142.857.142.85 0
MOH. ALDI SETIAWAN	8.133.333.333. 333.330	8.028.571.428.5 71.420	8.342.857.142.857.140	C2	8.028.571.428.571.42 0
...

Jarak hasil perhitungan akan dibandingkan antara tiga cluster data terdekat dengan pusat cluster dan dipilih nilai terkecil. Jarak ini menunjukkan bahwa data tersebut tergolong dalam satu kelompok dengan pusat cluster yang paling dekat. Berikut adalah anggota cluster yang terbentuk pada iterasi pertama, yang ditandai dengan simbol "T".

Tabel 4. Anggota Cluster Iterasi ke-1

NO	Nama	C1	C2	C3
1	ADITIA PUTRA			T
2	ALFIN DWI RADITYA	T		
3	ANDIKA PANDU MAULANA		T	
4	APPRIYANTO		T	
5	ARIL MAULANA PUTRA		T	
6	DIMAS SANTOSO		T	
7	FAJAR BAITUL MAKMUN		T	
8	FIKRI NABILUL MUSTHOFA		T	
9	KUSMAWIYANTO		T	
10	M. HARIYADI		T	
11	M. MURSID		T	
12	MOH. ALDI SETIAWAN		T	
...

4.1.4 Proses Pengujian Menggunakan RapidMiner

Proses ini menggunakan metode klusterisasi dengan algoritma K-Means untuk membentuk kelompok cluster secara akurat. Penelitian ini dilakukan menggunakan RapidMiner Studio versi 10.3 dengan langkah-langkah sebagai berikut:

a. Unggah Dataset

Sebelum memulai pengolahan data di RapidMiner, unggah dataset yang akan digunakan. Gunakan operator Read Excel karena format data yang digunakan adalah .xlsx (Microsoft Excel Worksheet).

b. Atur Type Role Kolom "Nama"

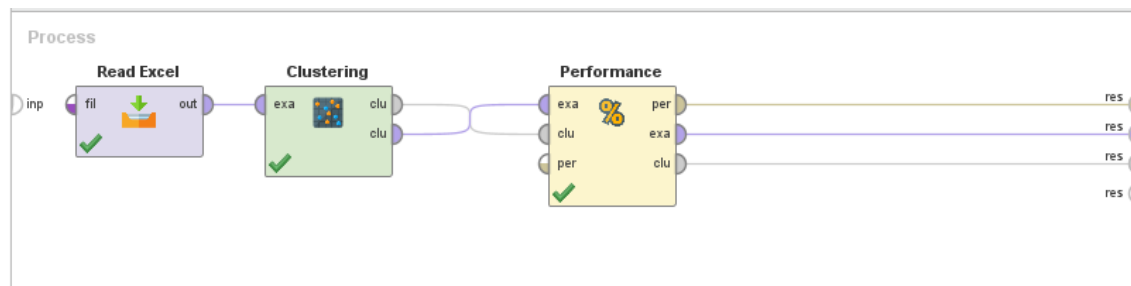
Ubah type role kolom "Nama" menjadi id, karena kolom ini berisi data non-numerik. Algoritma K-Means hanya dapat bekerja dengan data numerik, sehingga peran kolom "Nama" harus diubah dari attribute ke id.

c. Konfigurasi Operator K-Means

Pada bagian Modeling – Segmentation, seret (drag) operator k-Means ke area kerja. Atur jumlah k sesuai dengan jumlah cluster yang diinginkan, yaitu 3 cluster. Pilih Measure Types menjadi BregmanDivergences, karena data yang digunakan hanya berupa atribut nilai siswa dengan tipe data integer. Untuk Divergence, gunakan SquaredEuclideanDistance, karena metode ini menghitung jarak antar objek ke pusat cluster menggunakan rumus Euclidean Distance.

d. Evaluasi Kinerja Klasterisasi

Seret operator Performance (Cluster Distance Performance) untuk mengukur kinerja klasterisasi. Langkah ini bertujuan menghitung nilai DBI (Davies-Bouldin Index) sebagai indikator kualitas klasterisasi.



Gambar 3. Tahapan Pengujian dengan RapidMiner

4.1.1 Hasil Pengujian Algoritma K-Means

Berdasarkan hasil pengujian dengan 3 cluster, langkah selanjutnya adalah menghitung nilai DBI (Davies Bouldin Index). DBI digunakan sebagai ukuran untuk menilai kedekatan data dalam satu cluster. Adapun nilai Davies Bouldin Index yang diperoleh adalah sebagai berikut:

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 50.253
Avg. within centroid distance_cluster_0: 52.482
Avg. within centroid distance_cluster_1: 44.499
Avg. within centroid distance_cluster_2: 57.429
Davies Bouldin: 1.060
```

Gambar 4. Hasil Nilai *Davies-Bouldin Index*

5. KESIMPULAN DAN SARAN

Dalam era digital, pemanfaatan teknologi informasi menjadi krusial untuk mengoptimalkan pengelolaan data pendidikan, khususnya dalam mendukung pengambilan keputusan yang lebih objektif dan efisien. Pendekatan data mining, seperti Algoritma K-Means, memungkinkan pengelompokan data siswa berdasarkan kesamaan karakteristik, sehingga dapat digunakan untuk mengidentifikasi siswa unggulan serta memberikan wawasan yang mendukung strategi pembelajaran yang lebih terarah. Dengan didukung proses data cleaning yang baik, analisis data menjadi lebih akurat dan relevan. Hasil penelitian ini menunjukkan bahwa implementasi teknologi seperti Algoritma K-Means dapat menjadi solusi efektif bagi sekolah dalam memanfaatkan data besar untuk keperluan pendidikan.

Ucapan Terima Kasih

Saya mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua saya, Buhari dan Fatimah, atas doa dan dukungannya, serta kepada dosen pembimbing saya, Bapak Zaehol Fatah, M.Kom, atas bimbingan dan arahnya dalam penyusunan jurnal ini. Semoga segala kebaikan yang diberikan mendapatkan balasan terbaik dari Allah SWT..

DAFTAR PUSTAKA

- [1] Fatah, Z., Informasi, T., Sukorejo, U. I., Timur, S. J., Informasi, P. S., Sukorejo, U. I., & Timur, S. J. (2024). *IMPLEMENTASI RAPIDMINER PADA KLASTERISASI GEMPA BUMI DI INDONESIA BERDASARKAN KEDALAMAN MENGGUNAKAN K-MEANS*. 1(6), 84–91.
- [2] Hasugian, P. S., & Raphita Sagala, J. (2022). KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma K-Means. *Media Online*, 3(3), 262–268. <https://djournals.com/klik>
- [3] Jeanne Clarisa Wetik, W. (2017). Kata kunci *g*. *Kinabalu*, 11(2), 50–57.
- [4] K-means, M. D. M., Butsianto, S., Kom, S., Kom, M., & Kom, N. S. S. (2021). *Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means*. July. <https://doi.org/10.32672/jnkti.v3i1.2008>
- [5] Novi, ., & Mubarak, A. (2021). Penerapan Algoritma K-Means untuk Menentukan Kelas Unggulan di SMP Pelita Bandung. *Infomatek*, 32(2), 97–106. <https://doi.org/10.23969/infomatek.v23i2.4351>
- [6] *Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa Nirwana Hendrastuty Sistem Informasi, Universitas Teknokrat Indonesia, Indonesianirwanahendrastuty@teknokrat.ac.id AbstrakKat.pdf*. (n.d.).
- [7] Praja, S., Maruli, R., Informatika, T., Cipta, S., & Informatika, K. (2023). *Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Jalur Prestasi Akademik Di SMA Negeri 13 Jakarta Dengan Menggunakan Algoritma Random Forest*. 3, 10065–10079.
- [8] Rahayu, N. D., Anshor, A. H., Afriantoro, I., Bangsa, U. P., Selatan, C., & Siswa, N. (2024). *Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means*. 6, 71–83.
- [9] Suputra, I., Candiasa, I., & Suryawan, I. (2021). Klasterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma K-Means. *Wahana Matematika Dan Sains: Jurnal Matematika, Sains, Dan Pembelajarannya*, 15(1), 22–30. <https://ejournal.undiksha.ac.id/index.php/JPM/article/view/25380>
- [10] Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>