



## **Implementasi RapidMiner dalam Optimasi Pembentukan Kelas Unggulan Menggunakan K-Means Clustering**

**Yurida Islahatul Muasaroh<sup>1\*</sup>, Zaehol Fatah<sup>2</sup>**

<sup>1,2</sup> Sistem Informasi, Universitas Ibrahimy, Indonesia

EMAIL: ([yuridaislahatul21@gmail.com](mailto:yuridaislahatul21@gmail.com), [zaeholfatah@gmail.com](mailto:zaeholfatah@gmail.com))

Diterima : 24 November 2024. Disetujui : 02 Desember 2024. Dipublikasikan : 05 Desember 2024.

**ABSTRACT** - Grouping students into superior classes is one of the important strategies for optimizing the learning process in schools. This study aims to implement the K-Means Clustering method in determining superior classes by considering various variables such as academic achievement, exam scores, parental income, KIP ownership, and home ownership status. The analysis was done using the RapidMiner Studio tool with 23 student data records. The implementation process includes the data preprocessing stage, the application of the K-Means algorithm with parameters  $k = 5$  and  $\text{max runs} = 10$ , and the evaluation of clustering performance using Cluster Distance Performance. The results showed the formation of 5 clusters with a Davies Bouldin Index value of  $-1.199$ , indicating good clustering quality. Clustering performance analysis based on the average value within centroid distance showed variations in the level of density in each cluster, where cluster\_2 had the highest density level with a value of  $-2.054$ . In contrast, cluster\_0 had the lowest density level with a value of  $-1.028$ . Cluster\_1, cluster\_3, and cluster\_4 have values of  $-1.824$ ,  $-1.203$ , and  $-1.772$ , respectively. The results of this clustering help identify groups of students with similar characteristics that can be used as a

basis for forming superior classes. The implementation of K-Means Clustering has proven effective in optimizing the process of grouping students objectively and measurably, providing a strong foundation for schools in making decisions related to the formation of superior classes.

**Keywords** : K-Means Clustering, First Class, Data Mining

**ABSTRAK** - Pengelompokan siswa ke dalam kelas unggulan merupakan salah satu strategi penting dalam mengoptimalkan proses pembelajaran di sekolah. Penelitian ini bertujuan untuk mengimplementasikan metode K-Means Clustering dalam penetapan kelas unggulan dengan mempertimbangkan berbagai variabel seperti prestasi akademik, nilai ujian, penghasilan orang tua, kepemilikan KIP, dan status kepemilikan rumah. Analisis dilakukan menggunakan tools RapidMiner Studio dengan dataset yang terdiri dari 23 record data siswa. Proses implementasi meliputi tahap preprocessing data, penerapan algoritma K-Means dengan parameter  $k=5$  dan  $\text{max runs}=10$ , serta evaluasi performa clustering menggunakan Cluster Distance Performance. Hasil penelitian menunjukkan pembentukan 5 cluster dengan nilai

Davies Bouldin Index sebesar -1.199, mengindikasikan kualitas clustering yang baik. Analisis performa clustering berdasarkan nilai rata-rata within centroid distance menunjukkan variasi tingkat kepadatan pada setiap cluster, dimana cluster\_2 memiliki tingkat kepadatan tertinggi dengan nilai -2.054, sementara cluster\_0 memiliki tingkat kepadatan terendah dengan nilai -1.028. Cluster\_1, cluster\_3, dan cluster\_4 masing-masing memiliki nilai -1.824, -1.203, dan -1.772. Hasil clustering ini membantu mengidentifikasi kelompok siswa dengan karakteristik serupa yang dapat dijadikan dasar dalam pembentukan kelas unggulan. Implementasi K-Means Clustering terbukti efektif dalam mengoptimalkan proses pengelompokan siswa secara objektif dan terukur, memberikan landasan yang kuat bagi pihak sekolah dalam pengambilan keputusan terkait pembentukan kelas unggulan

**Kata kunci :** K-Means Clustering, Kelas Unggulan, Data Mining

## I. PENDAHULUAN

Pengelolaan kelas unggulan dalam sistem pendidikan modern memerlukan pendekatan yang lebih sistematis dan berbasis data. Saat ini, banyak sekolah masih mengandalkan metode manual dalam penentuan kelas unggulan dengan hanya mempertimbangkan nilai akademis sebagai faktor utama. Padahal, terdapat berbagai faktor penting lainnya yang dapat mempengaruhi potensi dan prestasi siswa, seperti latar belakang ekonomi keluarga, kepemilikan bantuan pendidikan (KIP), dan status kepemilikan rumah yang perlu dipertimbangkan dalam proses pengelompokan siswa[1].

Berdasarkan data yang diperoleh dari penelitian ini, terdapat beberapa variabel kunci yang dianalisis, meliputi prestasi akademik, nilai ujian, penghasilan orang tua, status kepemilikan KIP, dan kepemilikan rumah. Kompleksitas dari berbagai variabel ini menyebabkan pihak sekolah mengalami kesulitan dalam melakukan pengelompokan siswa secara objektif dan efisien. Terlebih lagi, pengelompokan manual seringkali menghasilkan distribusi yang kurang optimal dan memakan waktu yang cukup lama dalam prosesnya[2].

Untuk mengatasi permasalahan tersebut, penerapan metode K-Means Clustering dengan bantuan tools

RapidMiner menawarkan solusi yang lebih terstruktur dan berbasis data. K-Means Clustering merupakan algoritma pengelompokan yang dapat menganalisis multiple variabel secara simultan, sehingga mampu menghasilkan klasifikasi yang lebih akurat dan komprehensif. Metode ini tidak hanya mempertimbangkan aspek akademik, tetapi juga mengintegrasikan faktor sosial-ekonomi dalam prosesnya, yang tercermin dari variabel-variabel seperti penghasilan orang tua dan kepemilikan KIP[3].

Penelitian ini bertujuan untuk mengoptimalkan pembentukan kelas unggulan dengan mengimplementasikan algoritma K-Means Clustering pada dataset yang mencakup 1000 sampel data siswa dengan berbagai karakteristik. Melalui pendekatan ini, diharapkan dapat menghasilkan model pengelompokan yang lebih objektif dan dapat dipertanggungjawabkan secara ilmiah. Hasil clustering yang diperoleh nantinya dapat membantu pihak sekolah dalam mengambil keputusan yang lebih tepat terkait pembentukan kelas unggulan, sekaligus memberikan pemahaman yang lebih mendalam tentang karakteristik dan kebutuhan siswa berdasarkan berbagai aspek yang dianalisis[4].

Inti dari penelitian ini terletak pada upaya mengintegrasikan teknologi data mining dalam proses pendidikan, khususnya dalam optimalisasi pembentukan kelas unggulan. Dengan mempertimbangkan berbagai variabel yang ada dalam dataset, penelitian ini diharapkan dapat memberikan kontribusi praktis bagi sekolah dalam mengembangkan sistem pengelompokan siswa yang lebih efektif dan berkeadilan[5].

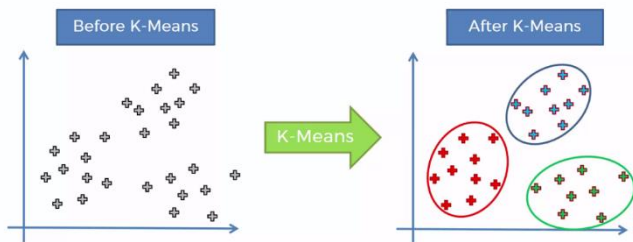
## II. TEORI DASAR

Dalam penelitian informatika, terdapat beberapa teori dasar yang menjadi landasan untuk memahami dan mengembangkan bidang ini. Berikut adalah beberapa teori dasar yang relevan dalam penelitian informatika:

1. Data Student: Data siswa merupakan kumpulan informasi yang berkaitan dengan karakteristik dan atribut dari setiap siswa di sekolah. Dalam konteks penelitian ini, data siswa mencakup beberapa variabel penting seperti prestasi akademik,

nilai ujian, penghasilan orang tua, status kepemilikan KIP (Kartu Indonesia Pintar), dan status kepemilikan rumah. Data-data ini menjadi dasar penting dalam proses analisis untuk pembentukan kelas unggulan karena mencerminkan tidak hanya kemampuan akademik siswa tetapi juga kondisi sosial-ekonomi yang dapat mempengaruhi proses pembelajaran mereka[6].

2. K-Means Clustering: K-Means Clustering adalah salah satu algoritma unsupervised learning yang paling populer dalam analisis pengelompokan data. Algoritma ini bekerja dengan membagi sejumlah data ke dalam k kelompok (cluster) berdasarkan kemiripan karakteristik data. Proses clustering dilakukan dengan menentukan titik pusat (centroid) untuk setiap cluster, kemudian menghitung jarak setiap data ke masing-masing centroid. Data akan dikelompokkan ke dalam cluster yang memiliki jarak terdekat dengan centroid-nya. Proses ini diulang secara iteratif hingga tidak ada perubahan signifikan pada posisi centroid atau telah mencapai jumlah iterasi maksimum yang ditentukan[7].



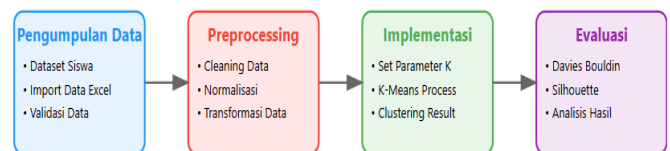
Gambar 1. Hasil K-Means Clustering

3. RapidMiner: RapidMiner merupakan platform perangkat lunak data science yang menyediakan lingkungan terintegrasi untuk persiapan data, machine learning, deep learning, text mining, dan predictive analytics. Perangkat lunak ini menawarkan antarmuka visual yang intuitif dan memudahkan pengguna dalam melakukan analisis data tanpa perlu menulis kode program yang kompleks. RapidMiner mendukung berbagai operator untuk pengolahan data, termasuk algoritma clustering seperti K-Means, yang membuatnya menjadi alat yang efektif untuk implementasi teknik data mining dalam berbagai konteks, termasuk dalam bidang pendidikan[8].

4. Data Mining: Data Mining adalah proses ekstraksi dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Dalam konteks pendidikan, data mining berperan penting dalam menganalisis pola-pola tersembunyi dari data siswa untuk menghasilkan wawasan yang berguna bagi pengambilan keputusan. Teknik data mining meliputi berbagai metode seperti klasifikasi, clustering, asosiasi, dan prediksi[9]. Dalam penelitian ini, data mining digunakan untuk menganalisis data siswa menggunakan metode clustering untuk mengoptimalkan pembentukan kelas unggulan berdasarkan berbagai kriteria yang telah ditentukan.

### III. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode clustering dengan pendekatan kuantitatif untuk menganalisis pengelompokan siswa dalam pembentukan kelas unggulan. Data yang digunakan dalam penelitian ini terdiri dari 1000 record data siswa dengan variabel yang meliputi Prestasi, Nilai Ujian, Penghasilan Ortu, Punya Sejenis KIP, dan Kepemilikan Rumah. Pengolahan data dilakukan menggunakan software RapidMiner Studio dengan menerapkan algoritma K-Means Clustering sebagai metode analisis utama. Tahapan metodologi penelitian dimulai dengan proses pengumpulan data yang mencakup pengambilan dataset dari sumber data sekolah dalam format Excel. Selanjutnya dilakukan preprocessing data menggunakan operator Set Role untuk menentukan atribut target dan Remove Duplicates untuk memastikan tidak ada data ganda. Tahap preprocessing ini penting untuk memastikan kualitas data yang akan dianalisis. Setelah data siap, dilanjutkan dengan proses normalisasi untuk menyeragamkan skala data menggunakan operator



Normalize[10].

Gambar 2. Metodologi Penelitian

Implementasi K-Means Clustering dilakukan dengan mengatur parameter  $k=5$  untuk membentuk 5 cluster dan  $\text{max runs}=10$  sebagai batas maksimal iterasi. Operator K-Means dikonfigurasi dengan mengaktifkan "add cluster attribute" untuk menambahkan informasi keanggotaan cluster pada setiap record data. Hasil clustering kemudian dievaluasi menggunakan Cluster Distance Performance untuk mengukur kualitas cluster yang terbentuk, dengan memperhatikan nilai Davies Bouldin Index dan rata-rata within centroid distance sebagai indikator utama performa clustering.[11]

#### IV. IMPLEMENTASI

##### 1. Preprocessing

Preprocessing data merupakan tahapan krusial dalam analisis data mining untuk memastikan kualitas data yang akan diolah. Dalam penelitian ini, preprocessing data dilakukan pada dataset siswa yang terdiri dari variabel Alias, Prestasi, Nilai Ujian, Penghasilan Ortu, Punya Sejenis KIP, Kepemilikan Rumah, dan Target. Tahap preprocessing dimulai dengan pemeriksaan kelengkapan data untuk memastikan tidak ada missing value atau data yang kosong pada setiap record. Dataset yang digunakan memiliki 1000 record data siswa yang sudah terisi lengkap sehingga tidak memerlukan proses penanganan missing value.

A	B	C	D	E	F	G
Alias	Prestasi	NilaiUjian	PenghasilanOrtu	PunyaSejenisKIP	KepemilikanRumah	Target
1	0	65	2,1	0	2	0
2	0	70	2,7	1	2	0
3	1	59	2,6	1	2	0
4	1	80	2,7	0	2	1
5	0	60	2,5	1	2	0
6	0	67	2,9	0	2	0
7	0	78	2,1	1	2	1
8	0	80	2,9	0	2	0
9	0	80	2,6	1	2	0
10	0	80	2,9	0	2	0
11	0	78	2,2	1	2	1
12	1	59	2,2	1	2	1
13	0	80	2,9	0	2	0
14	1	80	2,9	0	2	0
15	0	78	2,2	0	2	0
16	1	77	2,2	1	2	1
17	1	80	2,6	0	2	0
18	0	78	2,8	0	2	0
19	0	60	2,8	0	2	0

Gambar 3.

Data Setelah *convert* Excel

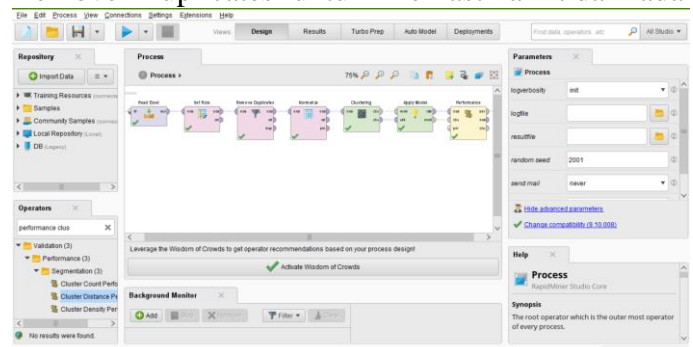
Selanjutnya, dilakukan pengecekan konsistensi tipe data untuk setiap variabel. Variabel Alias berfungsi sebagai identitas unik siswa dengan tipe data numerik, Prestasi menggunakan tipe data biner (0 dan 1), Nilai Ujian berupa data numerik, Penghasilan Ortu dalam format numerik (dalam jutaan rupiah), Punya Sejenis KIP dan Kepemilikan

Rumah menggunakan tipe data biner, serta Target sebagai label kelas juga menggunakan tipe data biner. Pada tahap ini tidak ditemukan inkonsistensi tipe data sehingga tidak diperlukan transformasi khusus.

Dalam implementasinya menggunakan RapidMiner, proses preprocessing melibatkan beberapa operator penting. Pertama, menggunakan operator 'Set Role' untuk menentukan atribut Target sebagai label yang akan menjadi acuan dalam evaluasi hasil clustering. Kemudian dilakukan pengecekan duplikasi data menggunakan operator 'Remove Duplicates' untuk memastikan keunikan setiap record, meskipun dalam dataset ini tidak ditemukan data yang terduplikasi. Dengan demikian, dataset telah siap untuk diproses ke tahap selanjutnya yaitu implementasi algoritma

##### 2. K-Means Clustering

Implementasi K-Means Clustering dalam penelitian ini dilakukan menggunakan RapidMiner Studio dengan menerapkan beberapa tahapan proses. Tahap pertama dimulai dengan mempersiapkan dataset yang terdiri dari 23 record data siswa dengan variabel Prestasi, NilaiUjian, PenghasilanOrtu, PunyaSejenisKIP, dan KepemilikanRumah. Setelah data dimuat menggunakan operator Read Excel, dilanjutkan dengan tahap preprocessing menggunakan operator Set Role untuk menentukan atribut target dan Remove Duplicates untuk memastikan tidak ada



Gambar 4.

Tahapan Proses RapidMiner

Proses clustering menggunakan operator K-Means dengan konfigurasi parameter  $k=5$  yang menentukan jumlah cluster yang akan dibentuk. Parameter max runs diatur pada nilai 10 yang menentukan jumlah maksimal iterasi dalam proses

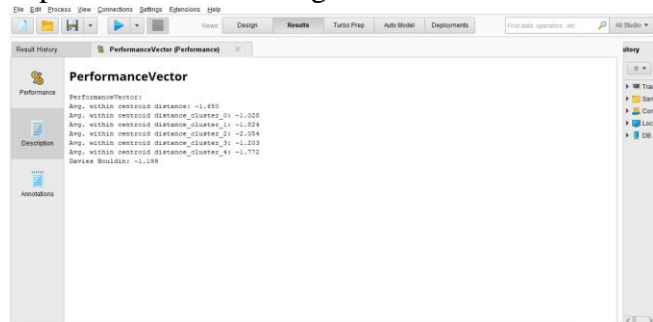
pembentukan cluster. Operator K-Means juga dikonfigurasi dengan mengaktifkan "add cluster attribute" untuk menambahkan informasi keanggotaan cluster pada setiap record data. Penggunaan numerical measure type dalam proses clustering memungkinkan penanganan data numerik dengan lebih baik, mengingat sebagian besar variabel dalam dataset bersifat numerik.

Setelah proses clustering selesai, operator Apply Model digunakan untuk menerapkan model clustering yang telah terbentuk pada dataset. Hasil implementasi menunjukkan pembagian data siswa ke dalam 5 cluster berbeda berdasarkan kemiripan karakteristik dari variabel-variabel yang dianalisis. Setiap cluster merepresentasikan kelompok siswa dengan karakteristik yang serupa, yang dapat digunakan sebagai dasar dalam pembentukan kelas unggulan.

### 3. Evaluasi Model

Evaluasi model clustering dalam penelitian ini dilakukan dengan menggunakan operator Cluster Distance Performance pada RapidMiner untuk mengukur kualitas cluster yang terbentuk. Hasil evaluasi menunjukkan nilai rata-rata within centroid distance sebesar -1.650, yang mengindikasikan tingkat kepadatan (cohesion) data dalam setiap cluster. Nilai negatif yang lebih kecil menunjukkan bahwa anggota dalam setiap cluster memiliki karakteristik yang relatif homogen dan tergabung dengan baik.

Performa clustering juga dapat dilihat dari nilai rata-rata within centroid distance untuk masing-masing cluster, dimana cluster\_0 memiliki nilai -1.028, cluster\_1 sebesar -1.824, cluster\_2 sebesar -2.054, cluster\_3 sebesar -1.203, dan cluster\_4 sebesar -1.772. Variasi nilai ini menunjukkan tingkat kepadatan yang berbeda pada setiap cluster, dengan cluster\_2 memiliki tingkat kepadatan tertinggi yang ditunjukkan oleh nilai terkecil (-2.054), sedangkan cluster\_0 memiliki tingkat kepadatan terendah dengan nilai -1.028.



Gambar 5.

### Hasil Evaluasi Model K-Means

Evaluasi kualitas clustering secara keseluruhan ditunjukkan oleh nilai Davies Bouldin Index sebesar -1.199. Nilai Davies Bouldin yang relatif kecil ini mengindikasikan bahwa hasil clustering memiliki kualitas yang baik, dengan pemisahan antar cluster yang jelas (separation) dan kepadatan dalam cluster yang tinggi (cohesion). Hal ini menunjukkan bahwa algoritma K-Means telah berhasil mengelompokkan data siswa ke dalam cluster-cluster yang memiliki karakteristik yang mirip di dalam cluster dan berbeda antar cluster.

### V. KESIMPULAN

Berdasarkan penelitian ini bahwa implementasi K-Means Clustering untuk penetapan kelas unggulan menunjukkan hasil yang efektif dalam mengelompokkan siswa berdasarkan karakteristik yang dimiliki. Analisis terhadap 1000 record data siswa menghasilkan 5 cluster dengan nilai Davies Bouldin Index sebesar -1.199, menandakan kualitas clustering yang baik. Performa clustering dapat dilihat dari nilai rata-rata within centroid distance untuk masing-masing cluster, dimana cluster\_0 memiliki nilai -1.028, cluster\_1 sebesar -1.824, cluster\_2 sebesar -2.054, cluster\_3 sebesar -1.203, dan cluster\_4 sebesar -1.772. Variasi nilai ini menunjukkan tingkat kepadatan yang berbeda pada setiap cluster, dengan cluster\_2 memiliki tingkat kepadatan tertinggi yang ditunjukkan oleh nilai terkecil (-2.054), sedangkan cluster\_0 memiliki tingkat kepadatan terendah dengan nilai -1.028. Hasil ini menunjukkan bahwa metode K-Means berhasil memisahkan siswa ke dalam kelompok-kelompok yang memiliki kemiripan karakteristik berdasarkan variabel yang dianalisis.

### REFERENSI

- [1] I. N. Rosmita Sari Siregar, Iskandar Kato, *Dasar-Dasar Pendidikan*. 2021.
- [2] M. Z. I. M. Sumari Mawardi, *Manajemen Kesiswaan di Sekolah*, vol. 7146, no. September. 2020.
- [3] I. Yuni Franata Sinurat, Masrizal, *Data Mining Pengelompokan Siswa Berprestasi Menggunakan Metode Clustering*. 2024. [Online].

Available:

- [4] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.
- [5] Ismai, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. 2017.
- [6] E. Kristanto, *Sekolah Untuk Semua*. 2013.
- [7] Nurhayati, *Pemodelan K- Means Algoritma Dan Big Data Analysis (Pemetaan Data Mustahiq)*. 2022.
- [8] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [9] I. Safira, R. Salkiawati, and W. Priatna, "Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi," *J. Inform. Inf. Secur.*, vol. 3, no. 1, pp. 99–110, 2022, doi: 10.31599/jiforty.v3i1.1253.
- [10] H. Alrasyid, A. Homaidi, M. Kom, Z. Fatah, and M. Kom, "Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes," vol. 1, no. 1, pp. 447–453, 2024.
- [11] E. A. Novia, W. I. Rahayu, and C. Prianto, *Sistem Perbandingan Algoritma K-Means dan Naive Bayes Untuk Memprediksi Priotas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan*. 2020.

*Halaman ini sengaja dikosongkan*