

## Peningkatan Efisiensi Pemantauan Kehadiran Siswa Melalui Analisis K-Means Clustering di Sekolah Menengah Pertama Negeri 3 Rancaekek, Kabupaten Bandung

Fitriani Agustina<sup>1\*</sup>, Rudi Kurniawan<sup>2</sup>, Tati Suprapti<sup>3</sup>

Program Studi Teknik Informatika, STMIK IKMI Cirebon

\*Email: agustinafitriani19@gmail.com

### Abstrak

Kehadiran siswa di sekolah merupakan faktor kunci dalam menentukan kualitas pendidikan dan efektivitas proses pembelajaran. Oleh karena itu, data kehadiran siswa dapat menjadi salah satu indikator untuk pihak sekolah dalam mengelola dan meningkatkan kualitas pendidikan. Yang menjadi permasalahan yaitu belum dilakukan proses analisis untuk mengelompokkan keaktifan siswa yang potensial berdasarkan kesamaan karakteristik dan pihak sekolah masih mengalami kesulitan dalam pengolahan data yang besar sehingga kualitas pendidikan belum maksimal. Penelitian ini melibatkan data kehadiran siswa dari SMP Negeri 3 Rancaekek selama satu tahun akademik sebagai dataset utama. Metode penelitian mencakup tahapan pengumpulan, pre-processing, dan analisis data. Data kehadiran siswa yang telah terkumpul diproses untuk menghilangkan outliers dan menciptakan dataset yang sesuai untuk analisis Clustering. Metode K-Means Clustering digunakan untuk mengelompokkan siswa menjadi beberapa kelompok berdasarkan pola kehadiran siswa tersebut. K-Means artinya prosedur pemecahan pengelompokan secara iteratif yang melakukan partisi untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek. K-Means sebagai metode data mining yang populer, merupakan prosedur pemecahan yang sering digunakan untuk mengidentifikasi kelompok alami dalam suatu kasus. Metode ini berfokus pada pengelompokan data yang memiliki kesamaan, sehingga hasilnya dapat dianalisis secara lebih mendalam. Hasil penelitian menunjukkan adanya beberapa kelompok Siswa dengan tingkat kehadiran yang berbeda-beda.

**Kata kunci:** Data Mining, Pola Kehadiran Siswa, K-Means Clustering, SMPN 3 Rancaekek

### Abstract

Student attendance at school is a key factor in determining the quality of education and the effectiveness of the learning process. Therefore, student attendance data can be one of the indicators for schools in managing and improving the quality of education. The problem is that the analysis process has not been carried out to group potential student activeness based on similar characteristics and the school still has difficulty in processing large data so that the quality of education is not optimal. This research involves student attendance data from SMP Negeri 3 Rancaekek for one academic year as the main dataset. The research method includes the stages of data collection, pre-processing, and analysis. The collected student attendance data was processed to remove outliers and create a dataset suitable for Clustering analysis. The K-Means Clustering method is used to group students into groups based on their attendance patterns. K-Means means an iterative clustering solution procedure that performs partitioning to classify or group a large number of objects. K-Means as a popular data mining method, is a solution procedure that is often used to identify natural groups in a case. This method focuses on grouping data that has similarities, so that the results can be analysed in more depth. The research results show that.

**Keywords:** Data Mining, Student Attendance Pattern, K-Means Clustering, SMPN 3 Rancaekek

### PENDAHULUAN

Sekolah merupakan tempat pendidikan setelah lingkungan keluarga. Di lingkungan sekolah siswa dapat terlibat dalam perilaku negatif yang berpotensi mempengaruhi perkembangan mental mereka. Oleh karena itu, menjaga disiplin siswa menjadi hal yang sangat krusial, karena disiplin merupakan salah satu faktor penentu keberhasilan mereka di masa

depan. Siswa yang dianggap disiplin di lingkungan sekolah adalah mereka yang patuh terhadap tata tertib, ditandai dengan kepatuhan seperti kehadiran di setiap kegiatan belajar mengajar (Mawarni and Budi, 2022).

Mengawasi data kehadiran siswa bisa dijadikan sebagai salah satu tindakan preventif terjadinya pembolosan pada siswa dan sebagai tindakan peningkatan tingkat kedisiplinan.

Sekolah Menengah Pertama Negeri (SMPN) 3 Rancaekek berlokasi di Jl. Bumi Jl. Teratai Raya, Rancaekek Wetan, Kec. Rancaekek, Kabupaten Bandung, Jawa Barat 40394. Pada sekolah ini belum terdapat proses pemanfaatan basis data dalam mengukur tingkat kehadiran dan kedisiplinan siswa dalam berpendidikan, sehingga menimbulkan permasalahan terhadap proses pengajaran dan nilai akademik. Menurut (Mawarni and Budi, 2022) mendeskripsikan bahwa tantangan yang sulit diatasi adalah kekurangan kedisiplinan dalam hal kehadiran siswa saat mengikuti pelajaran tatap muka. Jumlah data siswa terus meningkat setiap tahunnya, menyebabkan akumulasi data yang belum diproses secara efektif. Setiap objek dalam suatu cluster memiliki kemiripan karakteristik, namun karakteristik tersebut berbeda dengan objek di cluster lainnya (Agusto, 2023).

Analisis cluster merupakan teknik data mining yang bertujuan untuk mengidentifikasi suatu kelompok dari objek yang memiliki karakteristik yang sama (Hartati, Nurdyawan and Wiyandi, 2021). Data tersebut dikelompokkan ke dalam suatu Cluster yang memiliki kesamaan, sementara setiap cluster dibedakan dari objek-objek yang termasuk dalam kelompok yang berbeda. Potensi Clustering memiliki kemampuan untuk mengidentifikasi struktur dalam data. Cluster dapat diterapkan dalam berbagai aplikasi, seperti klasifikasi, pengolahan gambar, dan pengenalan pola. Metode untuk melakukan Cluster dalam penelitian ini adalah K-Means. K-Means merupakan suatu Algoritma yang digunakan dalam pengelompokan secara partisi. Setiap kelompok dipisahkan datanya ke dalam kelompok yang berbeda-beda (Virgo, Defit and Yuhandri, 2020).

Perhitungan dengan K-Means clustering adalah proses yang dilakukan secara berulang-ulang (Elda *et al.*, 2021). K-Means Clustering adalah metode yang memiliki potensi untuk mengelompokkan data menjadi kelompok-kelompok yang memiliki pola kehadiran yang serupa. Pendekatan eksperimental ini memberikan landasan kuat untuk mengungkap wawasan penting dari data kehadiran siswa. Hasil penelitian ini dapat memberikan panduan bagi pihak sekolah dan pembuat kebijakan dalam meningkatkan manajemen kehadiran siswa. Selain itu, metode analisis data yang

dikembangkan dapat digunakan dalam berbagai konteks lain di bidang.

## TINJAUAN PUSTAKA

### 2.1 Data Mining

Data mining adalah proses eksplorasi untuk menemukan pola dan relasi melalui pengumpulan dan pemanfaatan data dengan tujuan untuk menentukan nilai gizi pada anak balita. Tahap ini merupakan tahap yang akan diterapkan menggunakan *Algoritma K-Means Clustering*.

Data mining, juga dikenal sebagai penemuan pengetahuan dalam basis data (KDD), adalah suatu proses yang melibatkan pengumpulan dan pemanfaatan data historis guna mengidentifikasi pola atau relasi yang terdapat dalam kumpulan data yang besar (Nafsiah Muthmainnah *et al.*, 2023).

Ditinjau dari segi objeknya, penelitian ini meneliti dan menganalisis sebuah data dari salah satu Sekolah Menengah Pertama Negeri yang ada di Kabupaten Bandung yaitu SMPN3 Rancaekek.

### 2.2 Algoritma K-Means Clustering

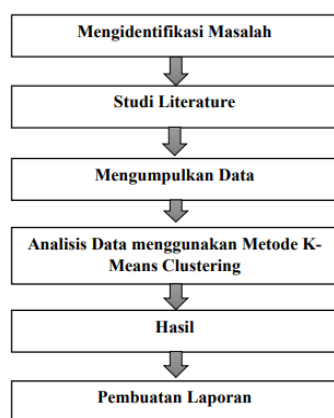
*Clustering* merupakan metode dalam data mining yang bertujuan untuk menganalisis data dengan lebih akurat, memecahkan masalah pengelompokan data, atau membagi kumpulan data menjadi beberapa bagian. Tujuan utama dari clustering adalah mengelompokkan data sehingga hubungan antar anggota dalam satu kelompok menjadi kuat, sementara hubungan antar kelompok menjadi lemah (Agusto, 2023). K-Means adalah suatu metode pemecahan masalah pengelompokan yang dilakukan secara iteratif.

Penelitian ini memutuskan untuk menerapkan Algoritma K-Means Clustering karena algoritma tersebut sering digunakan dalam pengelompokan data mining.

## METODE PENELITIAN

Metode penelitian adalah suatu metode atau langkah-langkah yang diterapkan untuk menghimpun data dengan menggunakan teknik tertentu. Dalam penelitian ini akan menggunakan metode penelitian yaitu proses *Knowledge Discovery in Databases* (KDD)

dengan pendekatan kuantitatif. Pada tahap ini akan diarahkan pada langkah-langkah teknik *Clustering* terdiri dari beberapa tahap, antara lain:



Gambar 1. Metode Penelitian

### 3.1 Sumber Data

Berikut merupakan sumber data yang diperoleh serta metode pengumpulan data yang digunakan dalam penelitian ini:

1. Sumber Data Primer berasal dari kehadiran dan nilai siswa SMPN 3 Rancaekek Kabupaten Bandung kelas 8 semester dua. Didapatkan secara langsung dari sumber data yang berhubungan dengan penelitian yang dilakukan pihak sekolah atau pengajar.
2. Data Sekunder dan studi pustaka didapatkan dengan cara mempelajari serta mencari referensi pada jurnal maupun literatur lain yang bersumber dari instansi atau internet sebagai penunjang landasan penelitian yang dilakukan yaitu pengelompokkan siswa dengan menggunakan Algoritma K-Means.
3. Dokumentasi serta pengumpulan data yang berafiliasi dengan penelitian ini, misalnya foto maupun data pendukung lainnya yang mendukung untuk melakukan analisis dalam menentukan keaktifan siswa dalam proses pembelajaran.

### 3.2 Populasi dan Tempat

Populasi dalam penelitian ini berjumlah 310 Siswa yang ada di kelas delapan pada Sekolah Menengah Pertama Negeri (SMPN) 3 Rancaekek. Berdasarkan data yang diterima pada tahun pelajaran 2022/2023. Tempat yang

dijadikan sebagai objek penelitian ini berada di Jl. Bumi Teratai Raya, Rancaekek Wetan, Kec. Rancaekek, Kabupaten Bandung, Jawa Barat 40394.

### 3.3 Data Selection

Data seringkali tidak dimanfaatkan sepenuhnya, sehingga hanya data yang relevan untuk analisis yang akan diekstrak dari dataset. Data dalam bentuk file excel yang diperoleh dari SMPN 3 Rancaekek dengan jumlah record sebanyak 310 record.

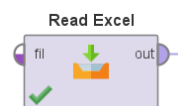
Tabel 1. Atribut Dataset

No.	Atribut	Type Data
1.	Nama Siswa	<b>Id</b>
2.	Alpa	<b>Integer</b>
3.	Izin	<b>Integer</b>
4.	Sakit	<b>Integer</b>

Tabel 2. Sample Data

NAMA SISWA	SAKIT	IZIN	ALPA
ABUZAR A.	19	19	<b>17</b>
ALMAIRA KH.	10	13	<b>6</b>
ANGGA RIPANA	10	13	<b>7</b>
ANISA YASMA	9	2	<b>6</b>
ARIA RASPANJI	11	23	<b>10</b>

Pilih operator read excel dan drag pada bagan process.



Gambar 2. Read Excel

### 3.4 Preprocessing

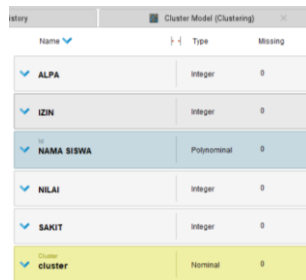
Proses pembersihan dan pengolahan data yang telah terkumpul melibatkan langkah-langkah seperti menghapus data yang tidak valid. Tindakan ini mencakup eliminasi duplikat data, pemeriksaan terhadap nilai yang hilang, dan penanganan data kosong untuk mencegah dampak negatif pada akurasi hasil yang akan diperoleh.



Gambar 3. Parameter Set Role

### 3.5 Transformasi (Transformation)

Langkah di mana data diubah dan digabungkan menjadi format yang sesuai untuk proses yang dapat diatasi dengan menggunakan operator split data pada RapidMiner. Operator ini memungkinkan pembagian dataset menjadi bagian data pelatihan dan data uji dengan proporsi yang telah ditentukan.



Name	Type	Missing
ALPA	Integer	0
IZIN	Integer	0
NAMA SISWA	Polynomial	0
NILAI	Integer	0
SAKIT	Integer	0
Cluster	Nominal	0

Gambar 4. Hasil Transformasi

### 3.6 Data Mining

Penentuan jumlah cluster merupakan langkah yang harus dilakukan. Nilai kluster atau k yang digunakan adalah (jumlah cluster optimal, jumlah anggota optimal dalam cluster). Hasil dari pengelompokan tersebut dimana akan dijadikan sebagai acuan untuk melakukan evaluasi peningkatan keaktifan dari data kehadiran Sekolah Menengah Pertama Negeri (SMPN) 3 Rancaekek. Berikut adalah desain proses untuk membuat cluster K-Means data siswa kelas delapan di SMPN 3 Rancaekek menggunakan tools Rapid Miner versi 10.2.



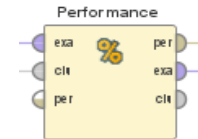
Gambar 5. Desain Proses Clustering



Gambar 6. Cluster K-Means

Clustering merupakan operator Cluster K-Means, operator ini adalah operator utama dari pemodelan proses ini agar dapat mengetahui hasil dari pengklasteran tersebut. Algoritma K-

Means menentukan satu set k cluster dan menetapkan setiap Contoh ke satu cluster yang tepat.



Gambar 7. Performance

Performance digunakan untuk mengukur Cluster Distance Performance dari proses K-Means tersebut, untuk mengukur peformaa dari operator K-Means. Operator Cluster Distance Performance mengambil model cluster centroid ini dan clustered set sebagai input dan mengevaluasi kinerja model berdasarkan centroid cluster. Dua ukuran kinerja didukung: Rata-rata dalam jarak cluster dan indeks Davies-Bouldin. Ukuran kinerja ini dijelaskan dalam parameter.

#### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 22.366
Avg. within centroid distance_cluster_0: 17.057
Avg. within centroid distance_cluster_1: 41.777
Avg. within centroid distance_cluster_2: 26.546
Avg. within centroid distance_cluster_3: 19.038
Davies Bouldin: 0.997
```

Gambar 8. Hasil Performance Vektor

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
SAKIT	8	11.419	9.270	7.983
IZIN	3.805	10.233	21.811	11.325
ALPA	8.947	15.505	9.514	7.137

Gambar 9. Hasil Cluster

Hasil dari pengelompokan menggunakan algoritma K-Means dengan menggunakan tool Rapid Miner versi 9.10 untuk mencari K2 - K20 memperoleh hasil K mana yang paling baik performanya yaitu yang mendekati angka 0. K yang mendekati angka 0 adalah K2. Dari K2=0,619 dan iterasi ke 1 dari jumlah iterasi 1 - 50, untuk mengetahui peforma dari algoritma K-Means dapat dilihat dari performance. Dimana davies bouldin apabila mendekati angka 0 nilai berarti algoritma semakin baik, maka jumlah cluster terbaik disini ada 2 yang terbaik atau yang mendekati 0.

## HASIL DAN PEMBAHASAN

Dalam tahap awal, peneliti melakukan persiapan dataset. Setelah itu, empat model dirancang dengan membagi data latih dan uji berdasarkan rasio tertentu, dengan tujuan untuk mengevaluasi kinerja setiap model. Peneliti telah secara rinci menjelaskan proses implementasinya dari awal hingga akhir, dan pembahasan mengenai tahap implementasi tersebut diuraikan. Dalam hasil penelitian, dilakukan tujuh percobaan untuk menentukan jumlah kluster, di mana fokus utama adalah pada identifikasi nilai terendah dari DBI, karena nilai DBI yang rendah dianggap dapat meningkatkan kinerja algoritma K-Means clustering.

Tabel 3. *Davies Bouldin Index (DBI)*

Jumlah kluster	Nilai DBI
2	1.076
3	1.197
4	0.997
5	0.997
6	0.997
7	1.008

Dari Tabel 3 menjelaskan nilai DBI terbaik terdapat Pada kluster 3 maka penulis menggunakan Kluster 3 dalam melakukan klusterisasi dalam penelitian ini.

## SIMPULAN

Berdasarkan hasil dari penelitian dan pengamatan yang telah dilakukan, peneliti mendapatkan kesimpulan yang dimana Metode Algoritma *K-Means Clustering* dapat diterapkan dalam melakukan penilaian terhadap Tingkat Kehadiran Siswa pada SMPN 3 Rancaekek Kabupaten Bandung. Penerapan Algoritma K-Means Clustering pada Rapidminer menampilkan hasil anggota *Cluster 0* sebanyak 113 anggota, *Cluster 1* sebanyak 43 anggota, *Cluster 2* sebanyak 37 anggota dan *Cluster 3* sebanyak 117 dengan total data 310 anggota. Hasil dari *Performance Evaluasi Davies Bouldin Index (DBI)* terbaik setelah dilakukan 7 percobaan clustering yang telah didapatkan pada penelitian ini 4 Clustering dengan Nilai DBI nya adalah 0.997.

Dari temuan yang diperoleh dalam studi ini, diharapkan dapat menjadi landasan bagi

penelitian berikutnya dalam mengeksplorasi tingkat kehadiran yang optimal di Sekolah Menengah Pertama (SMP) dan lembaga pendidikan lainnya. Pendekatan yang lebih maju, seperti penggabungan algoritma K-means dengan metode clustering lainnya, diharapkan dapat diterapkan untuk analisis yang lebih mendalam.

## DAFTAR PUSTAKA

- Agusto, S.D. (2023) '... Data Nilai Siswa Kelas 8 Berbasis Nilai Pengetahuan Untuk Menentukan Siswa Berprestasi Menggunakan Metode K-Means Clustering (Studi Kasus: SMP Negeri 4 ...', 7(September), pp. 630–638. Available at: [https://repository.uksw.edu/handle/123456789/30778%0Ahttps://repository.uksw.edu/bitstream/123456789/30778/7/T1\\_672018414\\_Isi.pdf](https://repository.uksw.edu/handle/123456789/30778%0Ahttps://repository.uksw.edu/bitstream/123456789/30778/7/T1_672018414_Isi.pdf).
- Elda, Y. et al. (2021) 'Klusterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means', *Jurnal Informasi dan Teknologi*, 3, pp. 103–108. Available at: <https://doi.org/10.37034/jidt.v3i3.130>.
- Hartati, T., Nurdiawan, O. and Wiyandi, E. (2021) 'Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari', *Jurnal Sains Teknologi Transportasi Maritim*, 3(1), pp. 1–7. Available at: <https://doi.org/10.51578/j.sitektransmar.v3i1.30>.
- Mawarni, Q.I. and Budi, E.S. (2022) 'Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa', *Jurnal Sistem Komputer dan Informatika (JSON)*, 3(4), p. 522. Available at: <https://doi.org/10.30865/json.v3i4.4242>.
- Nafsiah Muthmainnah, T. et al. (2023) 'Jurnal Informatika dan Rekayasa Perangkat Lunak Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat', 5(2), pp. 122–129.
- Virgo, I., Defit, S. and Yuhandri, Y. (2020) 'Klusterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering', *Jurnal Sistim Informasi dan Teknologi*, 2, pp. 23–28. Available at: <https://doi.org/10.37034/jsisfotek.v2i1.17>.