

Implementasi Algoritma *K-Means Clustering* Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran

Falih Pramataning Dewi ^{(1)*}, Priskila Siwi Aryni ⁽²⁾, Yuyun Umaidah ⁽³⁾

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, Karawang
e-mail : {falih.pramataning18217,priskila.siwi18220}@student.unsika.ac.id,
yuyun.umaidah@staff.unsika.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 3 Januari 2022, direvisi 25 April 2022, diterima 25 April 2022, dan dipublikasikan 25 Mei 2022.

Abstract

The learning process through various interactions and learning experiences has a considerable influence on developing student activity to improve the quality of education. The teacher is the most important factor in determining the success of students in the implementation of the process. The development of the quality and activeness of students in learning is a basic element as a form of success in the learning process which of course not all students have a level of speed in understanding material. This is a concern for schools in improving the quality of education. The purpose of this study was to classify the level of activity of students at SMP ABC using the correlation between grades and the level of student activity who would be recommended to take part in competitions or prospective scholarship recipients. The data source that we used in this study came from the State Junior High School ABC which consists of several variables, including student attendance data, academic scores, psychomotor scores, and affective values. The method used in this research is the Clustering method with the K-Means Algorithm. The results of this study can be grouped into 3 clusters including cluster 0 indicating active students as many as 30 students, cluster 1 showing inactive students as many as 8 students, and cluster 2 indicating less active students as many as 21 students.

Keywords: *Data Mining, Student Activity, Clustering Predictions, K-Means, Davies Bouldin*

Abstrak

Proses pembelajaran melalui berbagai interaksi dan pengalaman belajar memiliki pengaruh yang cukup besar untuk mengembangkan keaktifan siswa guna meningkatkan mutu pendidikan. Pengajar merupakan faktor terpenting dalam penentu keberhasilan siswa pada implementasi proses tersebut. Perkembangan kualitas dan keaktifan siswa dalam pembelajaran merupakan unsur dasar sebagai bentuk keberhasilan dalam proses pembelajaran yang tentu saja tidak semua siswa memiliki tingkat kecepatan dalam memahami suatu materi. Hal ini menjadi perhatian bagi sekolah dalam meningkatkan mutu pendidikan. Adapun tujuan pada penelitian ini dilakukan untuk pengelompokan terhadap tingkat keaktifan siswa SMP ABC menggunakan korelasi antara nilai dengan tingkat keaktifan siswa yang akan direkomendasikan untuk mengikuti perlombaan atau calon penerima beasiswa. Sumber data yang kami gunakan pada penelitian ini berasal dari Sekolah Menengah Pertama ABC yang terdiri dari beberapa variabel, meliputi data kehadiran siswa, nilai akademik, nilai psikomotor, nilai afektif. Metode yang digunakan pada penelitian ini yaitu metode *clustering* dengan algoritma *K-Means*. Hasil penelitian ini dapat dikelompokkan menjadi 3 *cluster* meliputi *cluster 0* menunjukkan siswa aktif sebanyak 30 siswa, *cluster 1* menunjukkan siswa tidak aktif sebanyak 8 siswa, dan *cluster 2* menunjukkan siswa kurang aktif sebanyak 21 siswa.

Kata Kunci: *Data Mining, Keaktifan Siswa, Prediksi Clustering, K-Means, Davies Bouldin*

1. PENDAHULUAN

Peningkatan mutu pendidikan merupakan prioritas kebijakan umum pembangunan di Indonesia yang harus terus diperhatikan. Faktor yang memiliki pengaruh besar dalam pencapaian tersebut adalah guru yang memberikan pengajaran efektif dan berkualitas. Perkembangan kualitas dan keaktifan siswa dalam pembelajaran merupakan unsur dasar sebagai bentuk keberhasilan dalam



proses pembelajaran yang tentu saja tidak semua siswa memiliki tingkat kecepatan dalam memahami suatu materi (Triandini, Defit, and Nurcahyo, 2021). Tentu saja, hal ini menjadi perhatian khusus bagi pihak sekolah dalam mencapai keberhasilan.

Perkembangan dunia digital memiliki berbagai manfaat dalam pemenuhan kebutuhan manusia. Salah satunya adalah perkembangan *data science* (Primanda, Alwi, and Mustikasari, 2021). *Data science* memiliki fungsi signifikan dalam pengolahan data. Dalam penentuan keaktifan siswa dalam proses pembelajaran pada SMP ABC adalah dengan penerapan *data mining*. Korelasi antara pengolahan data dengan peningkatan mutu pendidikan dapat digunakan untuk mengklasterisasikan siswa berprestasi untuk menentukan keaktifan siswa dalam proses pembelajaran.

Tahapan *data mining* yang digunakan pada penelitian ini meliputi himpunan data (*data set*), metode *data mining*, pengetahuan (*knowledge base*), dan evaluasi pola (*pattern evaluation*) (Vhallah, Sumijan, and Santony, 2018). KDD atau biasanya disebut *knowledge discovery database* merupakan metode *data mining* yang dapat digunakan untuk memperoleh pengetahuan dari *database* (Asroni, Fitri, and Prasetyo, 2018). Hasil pengetahuan (*knowledge base*) tersebut yang kemudian digunakan untuk keperluan penentuan keputusan (Gustientiedina & Desnelita, 2019). Adapun metode yang digunakan pada penelitian ini yaitu *clustering* dengan algoritma *K-Means* (Nainggolan and Purba, 2020).

Proses yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok agar setiap kelompok memuat data yang memiliki kemiripan yang maksimum atau biasa disebut *clustering*. Sedangkan *K-Means* merupakan metode yang banyak digunakan untuk mengidentifikasi suatu kelompok dari kasus. *K-Means* artinya prosedur pemecahan pengelompokan secara iteratif yang melakukan partisi untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek. Prosedur pemecahan *K-Means* adalah metode *data mining* yang banyak digunakan untuk mengidentifikasi suatu kelompok yang alami dari sebuah kasus yang berdasarkan pada pengelompokkan data yang memiliki kemiripan sehingga hasil dari pengelompokkan dapat dianalisis (Arofah and Marisa, 2018).

Pengelompokkan sekolah berdasarkan minat siswa guna meningkatkan kualitas dan kuantitas sekolah menggunakan sumber data Dapodikdasmen (Data Pokok Pendidikan Dasar dan Menengah) tahun 2019 dengan metode *K-Means clustering* dan menghasilkan 2 *cluster* (Oktarian, Defit, and Sumijan, 2020). Meningkatkan hasil belajar siswa serta mutu sekolah dengan pola pengelompokan siswa menggunakan nilai rata-rata kelas XI dengan algoritma *K-Means clustering* hingga menghasilkan *cluster* (Elda et al., 2021). Pengelompokan dalam menentukan kelayakan mengikuti *assessment center clustering* program SDP menggunakan metode *K-Means* sehingga menghasilkan 3 *cluster* dalam penelitian ini (Parlina et al., 2018).

Kelebihan dari penerapan *K-Means* yaitu mampu mengelompokkan objek besar serta dapat meningkatkan kecepatan proses pengelompokan (Parlambang and Fauziah, 2020). Tujuan algoritma ini merupakan untuk membagi data menjadi beberapa *cluster*. Dengan melakukan *clustering* prediksi terhadap tingkat keaktifan siswa serta korelasi antara nilai dengan tingkat keaktifan pada proses pembelajaran, akan membantu pengajar dalam mengelompokkan siswa yang akan direkomendasikan untuk dapat mengikuti perlombaan atau calon penerimaan beasiswa. Dalam penelitian ini dapat dilakukan prediksi terhadap tingkat keaktifan serta korelasi antara nilai dengan taraf keaktifan.

Pada penelitian (Kusuma & Aryati, 2019) dengan judul “Sistem Informasi Akademik Serta Penentuan Kelas Unggulan Menggunakan Metode Clustering dengan Algoritma *K-Means* Di SMP Negeri 3 Ubud” membahas mengenai efektifitas yang dihasilkan dari penerapan algoritma *K-Means clustering* dalam mengolah dan mengelompokkan nilai siswa berdasarkan nilai terdekat dengan titik pusat *cluster*. Berdasarkan hasil pengujian, penelitian ini menghasilkan nilai 83% menggunakan metode *precision* dan tingkat *accuracy* metode *K-Means* sebesar 92% dalam mengelompokkan siswa ke dalam kelas unggulan.



Pada penelitian (Sirait et al., 2019) dengan judul “Implementasi Algoritma *K-Means* untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA” membahas mengenai implementasi klasterisasi dalam pengolahan data dalam pemilihan calon peserta Olimpiade Sains Nasional (OSN). Hasil penelitian menunjukkan bahwa algoritma *K-Means clustering* cukup efektif membantu sekolah dalam mengolah data dalam jumlah besar untuk memilih siswa yang tepat dan dengan cara yang efektif.

Mengacu pada penjelasan latar belakang di atas, maka pentingnya dilakukan prediksi pengelompokan keaktifan siswa dengan menggunakan korelasi antara nilai dengan tingkat keaktifan siswa, yaitu untuk membantu sekolah dalam peningkatan mutu pendidikan dan dapat digunakan untuk mengklasterisasikan siswa berprestasi untuk menentukan keaktifan siswa dalam proses pembelajaran yang kemudian dapat direkomendasikan mengikuti perlombaan atau calon penerima beasiswa.

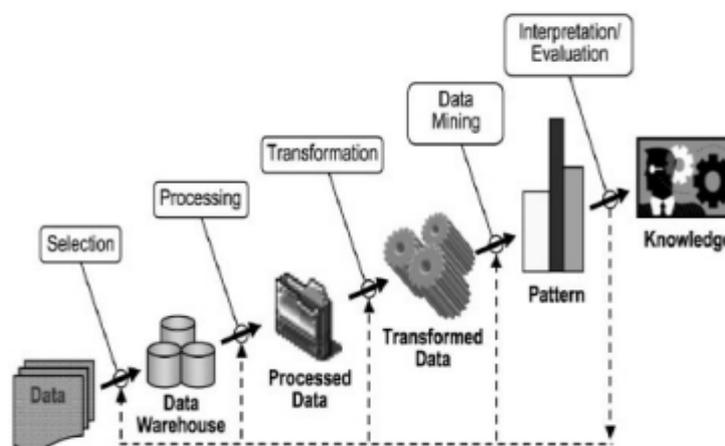
2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Maulida, 2018). *Data mining* memiliki banyak metode, metode *data mining* yang digunakan adalah *Knowledge Discovery in Database* (KDD) ataupun dapat disebut *pattern recognition*.

2.2 Knowledge Discovery in Database (KDD)

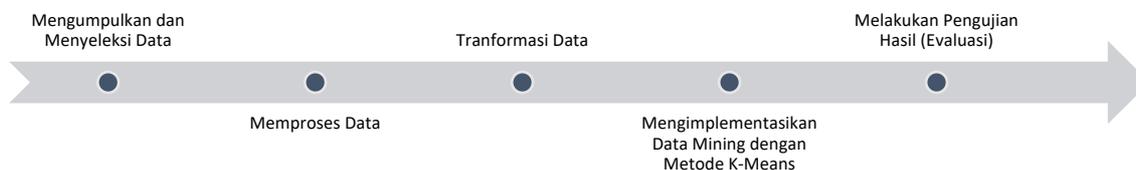
Knowledge discovery in databases (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti. Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap *pre-processing*, transformasi, data mining dan tahap evaluasi (Fatmawati and Windarto, 2018). Tahap seleksi dilakukan untuk menargetkan data yang digunakan untuk penelitian, tahap *pre-processing* dapat dilakukan integrasi data atau penggabungan data serta dilakukan *cleaning data*, yaitu dengan menghilangkan *noise*, data redundan, inkonsistensi data, serta data yang tidak relevan, tahap transformasi adalah penggabungan data dan penyesuaian format data agar dapat diproses pada tahap *data mining*, tahap *data mining* dilakukan menggunakan algoritma yang cocok untuk permasalahan dalam data, serta tahap evaluasi yang digunakan untuk pengujian dalam data (Nur Khormarudin, 2016). Gambar 1 merupakan tahapan alur metode KDD.



Gambar 1 Tahapan KDD



Pada Gambar 1 digambarkan alur mengenai tahapan KDD, setelah disesuaikan dengan tahapan KDD dihasilkan kerangka kerja penelitian yang dijabarkan pada Gambar 2.



Gambar 2 Kerangka Kerja

Sesuai dengan Gambar 2 mengenai langkah kerja, penelitian ini diawali dengan mengumpulkan data dan menyeleksi atribut data yang dijadikan data target, selanjutnya dilakukan pemrosesan data yang di dalamnya dilakukan pembersihan data, selanjutnya tranformasi data dilakukan untuk penggabungan data serta diubah format yang sesuai untuk dilakukan tahap selanjutnya yaitu *data mining*, dimana peneliti mengimplementasikan *data mining* dengan menggunakan *K-Means clustering*, serta tahap akhir dilakukan pengujian hasil menggunakan perhitungan manual dan dengan aplikasi RapidMiner.

2.3 Clustering

Analisis pengelompokan atau *clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar dari pada kesamaan data tersebut dengan data dalam kelompok lain (Pham et al., 2005). Potensi klasterisasi adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola (Ong, 2013).

2.4 Algoritma K-Means

K-Means merupakan metode yang banyak digunakan untuk mengidentifikasi suatu kelompok dari kasus. *K-Means* artinya prosedur pemecahan pengelompokan secara iteratif yang melakukan partisi buat mengklasifikasikan atau mengelompokkan sejumlah besar objek (Triandini et al., 2021). Prosedur pemecahan *K-Means* adalah metode *data mining* yang banyak digunakan buat mengidentifikasi suatu kelompok yang alami berasal sebuah kasus yang berdasarkan pada pengelompokan data yang memiliki kemiripan sehingga hasilnya dari pengelompokan dapat dianalisis (Arofah and Marisa, 2018). Kelebihan dari penerapan *K-Means* merupakan mampu mengelompokkan objek besar serta meningkatkan kecepatan proses pengelompokan.

Adapun di bawah ini merupakan uraian dari algoritma menggunakan metode *K-Means clustering*.

- 1) Tentukan sebanyak k kelompok yang kita inginkan
- 2) Inisiasi *centroid* untuk setiap kelompok. *Centroid* ibarat seperti "*leader group*", yang mempresentasikan kelompok.
- 3) Hitung kedekatan antara suatu data dengan *centroid*, selanjutnya masukkan data tersebut pada kelompok *centroid* yang memiliki sifat terdekat dengan dirinya
- 4) Pilih *centroid* baru untuk masing-masing kelompok dari anggota kelompok itu sendiri dan lakukan perhitungan kedekatan antara suatu data dengan *centroid*.
- 5) Ulangi langkah-langkah sebelumnya hingga tidak ada perubahan anggota untuk semua kelompok.

2.5 Davies Bouldin

Davies Bouldin merupakan metode yang biasa digunakan dalam proses *clustering*, berdasarkan kegunaannya metode ini dipakai untuk menentukan seberapa efektif *clustering* yang dilakukan serta untuk mengevaluasi *cluster* secara umum (Adhitama et al., 2020). Dengan ketentuan, jika



semakin kecil nilai Davies Bouldin maka akan semakin baik *cluster* yang dihasilkan, kohesi didefinisikan sebagai jumlah kedekatan data dengan *centroid* pada *cluster* (Rahma, 2020).

3. HASIL DAN PEMBAHASAN

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 nilai rapor siswa SMP ABC kelas 8 semester dua. Dilakukan menggunakan cara tanya jawab secara langsung dengan pihak sekolah atau pengajar.
- 2) Studi pustaka dilakukan dengan cara mempelajari serta mencari referensi pada jurnal maupun literatur lain yang berafiliasi dengan penelitian yang dilakukan yaitu pengelompokan 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 Seleksi Data

Berdasarkan kerangka pemikiran pada Gambar 2. Berikut adalah uraian mengenai penjelasan dari masing-masing langkah. Mengumpulkan data merupakan tahapan pertama yang digunakan pada penelitian ini. Dalam tahap pengumpulan ini, peneliti juga menyeleksi data yang perlu dan sesuai untuk digunakan dalam penelitian ini. Tahap ini berisikan pengumpulan nilai rapor dengan rincian nilai keaktifan, nilai pengetahuan, nilai psikomotor, nilai afektif, dan nilai kehadiran data siswa SMP ABC kelas 8 semester dua. Selain itu, juga dilakukan observasi lapangan melalui tanya jawab secara langsung dengan pihak sekolah atau pengajar.

3.3 Pemrosesan Data dan Tranformasi Data

Setelah dilakukannya seleksi data yang dibutuhkan dalam penelitian, dilanjutkan dengan pemrosesan data dan tranformasi data. Pemrosesan pada data dapat dilakukan dengan *cleaning* (pembersihan) dari data yang *missing value* atau redundan serta dilakukan juga intergrasi data. Namun, dikarenakan data yang sudah diberikan sudah lolos tahap *cleaning* dan integrasi, maka dilanjut dengan tahap tranformasi data, yaitu penggabungan data yang sudah sesuai dengan data yang digunakan dalam proses *data mining* nantinya. Dilihat pada Tabel 1 merupakan data yang sudah selesai dalam tahap transformasi dan sudah sesuai untuk dilakukannya proses *data mining*.

Tabel 1 *Dataset Keaktifan Siswa*

No.	Nama	nAktif	nPeng	nPsi	nAfek	nAbsen
1	SISWA01	48	80,8	82	80	100
2	SISWA02	83	81,9	83	80	100
3	SISWA03	82	81,2	82,1	80	100
4	SISWA04	78	81,9	82,5	80	100
5	SISWA05	61	81,5	82,3	80	100
6	SISWA06	61	79,9	80,7	80	100
7	SISWA07	88	80,3	81,4	80	100
8	SISWA08	65	79,7	80,4	80	100
9	SISWA09	57	79,8	80,7	80	100
10	SISWA10	57	83,1	84,1	80	100
...
56	SISWA56	90	83,4	83,1	80	100
57	SISWA57	62	84	84,3	80	100
58	SISWA58	95	85,2	85,6	80	100
59	SISWA59	85	82,1	82	80	100



Berdasarkan pada Tabel 1 data di atas terdapat 59 data yang akan diolah menggunakan metode *data mining* dengan algoritma *K-Means clustering* dengan cara mengelompokkan nilai siswa dari beberapa kelompok ke dalam bentuk *cluster* sehingga dapat dilakukan analisis pola pembentukan *cluster*.

3.4 Metode K-Means Clustering

Setelah data bersih dan dapat diproses, dilanjutkan dengan tahap *data mining* menggunakan algoritma *clustering*, yaitu dengan cara mengelompokkan nilai siswa dari beberapa kelompok ke dalam bentuk *cluster* dapat dilakukan analisa terhadap pola pembentukan *cluster*. Kemudian dapat menetapkan nilai *centroid* secara acak, selanjutnya mengestimasi jarak pada data ke setiap *centroid* dengan menggunakan metode korelasi antara dua buah objek yang terdapat pada rumus *Euclidean Distance* pada Pers. (1).

$$D_{(i,j)} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Di mana $D_{(i,j)}$ merupakan jarak data ke i ke pusat *cluster* j , dengan X_{ki} adalah data ke i pada atribut data ke k , dan X_{kj} adalah titik pusat ke j pada atribut ke k .

Berdasarkan penjabaran pada metode penelitian, langkah-langkah dalam *K-Means clustering* yaitu,

a) Menentukan k kelompok.

Jumlah *cluster* dalam penelitian ini adalah $k = 3$.

b) Inisiasi *centroid*

Adapun *centroid* dipilih secara acak seperti pada contoh Tabel 2.

Tabel 2 Inisiasi Centroid

K	nAktif	nPeng	nPsi	nAfek	nAbsen
C0	48	80.8	82	80	100
C1	100	80.9	81.6	80	100
C2	83	79.5	80.2	80	100

c) Hitung jarak pusat *cluster* dengan *Euclidean Distance* menggunakan rumus yang tertera di Tabel 3 merupakan contoh hasil perhitungan sampai data ke-3

Tabel 3 Jarak Pusat Cluster dengan Euclidean Distance

D(1,1)	D(1,2)	D(1,3)
0.00	52.00	35.07
35.03	17.09	3.69
34.00	18.01	2.74

Perhitungan dilakukan dari data ke-4 sampai data ke-59 dengan menggunakan perhitungan di Microsoft Excel sehingga diperoleh hasil seperti pada Tabel 4.

Tabel 4 Hasil 3 Data C0 Iterasi Ke-1

nAktif	nPeng	nPsi	nAfek	nAbsen	Kelompok
48	80.8	82	80	100	C0
61	81.5	82.3	80	100	C0
61	79.9	80.7	80	100	C0

Berdasarkan Tabel 4 diperoleh contoh 3 data dari hasil kelompok *cluster* 0 pada iterasi ke-1 dan terdapat 19 data untuk jumlah keseluruhan *cluster* 0.



Tabel 5 Hasil 3 Data C1 Iterasi Ke-1

nAktif	nPeng	nPsi	nAfek	nAbsen	Kelompok
100	80.9	81.6	80	100	C1
100	82.3	83.5	80	100	C1
100	82.1	83.2	80	100	C1

Berdasarkan Tabel 5 diperoleh contoh 3 data dari hasil kelompok *cluster 1* pada iterasi ke-1 dan terdapat 5 data untuk jumlah keseluruhan *cluster 1*.

Tabel 6 Hasil 3 Data C2 Iterasi Ke-1

nAktif	nPeng	nPsi	nAfek	nAbsen	Kelompok
83	81.9	83	80	100	C2
82	81.2	82.1	80	100	C2
78	81.9	82.5	80	100	C2

Berdasarkan Tabel 6 diperoleh contoh 3 data dari hasil kelompok *cluster 2* pada iterasi ke-1 dan terdapat 35 data untuk jumlah keseluruhan *cluster 2*.

d) Menentukan *cluster* baru dari hasil iterasi ke-1

Cara menentukan *cluster* baru adalah menghitung rata-rata yang di dapat dari setiap *cluster* yaitu, *cluster 0*, *cluster 1*, dan *cluster 2* dengan hasil seperti pada Tabel 7.

Tabel 7 Nilai Centroid Baru

K	\bar{x} nAktif	\bar{x} nPeng	\bar{x} nPsi	\bar{x} nAfek	\bar{x} nAbsen
C0	56,63	82,18	82,64	80,00	100,00
C1	97,40	82,42	83,32	80,00	100,00
C2	78,43	82,60	83,22	80,00	100,00

Jika sudah dihasilkan nilai *centroid* baru, setelahnya akan dihitung kembali untuk iterasi ke-2. Jika iterasi ke-2 selesai, maka akan dihitung galat dengan cara kelompok *cluster* iterasi ke-2 dikurangi dengan kelompok *cluster* iterasi ke-1. Iterasi akan terus dilakukan hingga seluruh nilai galat sama dengan nol. Dari hasil penelitian di atas, dilakukan iterasi sebanyak 3 kali untuk mendapatkan hasil galat sama dengan nol. Hasil akhir perhitungan manual *K-Means clustering* ditunjukkan pada Tabel 8.

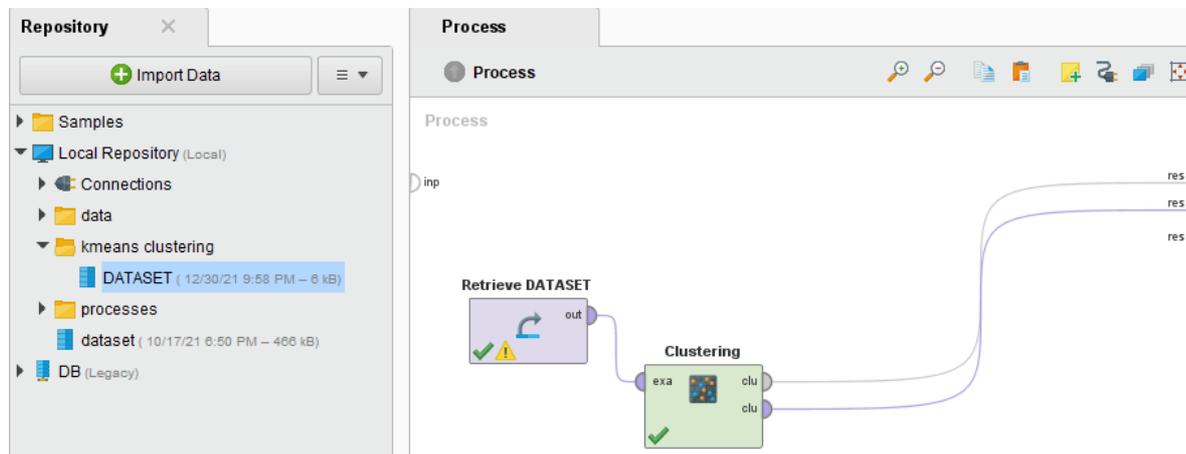
Tabel 8 Hasil Perhitungan Manual K-Means Clustering Iterasi Ke-3

No	D(3,1)	D(3,2)	D(3,3)	Cluster
1	9.72	46.28	30.23	0
2	25.38	11.26	4.95	2
3	24.40	12.33	4.36	2
4	20.38	16.26	1.23	2
5	3.45	33.27	17.21	0
6	4.46	33.41	17.57	0
7	30.46	6.74	10.36	1
8	8.07	29.47	13.81	0
9	3.05	37.40	21.50	0
10	1.90	37.28	21.15	0
...
57	5.07	32.32	16.21	0
58	37.63	4.02	17.19	1
59	27.39	9.30	7.03	2



3.5 Implementasi Algoritma *K-Means Clustering* Menggunakan RapidMiner

Data inputan yang digunakan berupa data excel dengan langkah-langkah pengerjaannya dapat dilihat pada Gambar 3.



Gambar 3 Tampilan Input *Dataset*

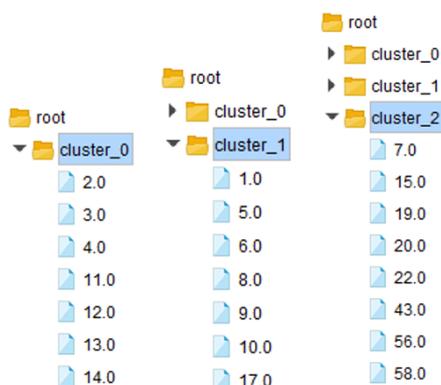
Setelah *input data* dan operator *K-Means* diletakkan pada kolom *design* dan diatur dengan jumlah *k* yang diinginkan, peneliti menghubungkan *connector* satu dengan yang lainnya agar tidak terjadi *error* dan menjalankan program dengan hasil dari data proses ditunjukkan pada Gambar 4.

Cluster Model

```
Cluster 0: 30 items  
Cluster 1: 21 items  
Cluster 2: 8 items  
Total number of items: 59
```

Gambar 4 Hasil Data Proses

Dari hasil perhitungan menggunakan RapidMiner dan perhitungan manual ditemukan kesamaan dalam jumlah *cluster* yang dihasilkan. Walaupun *cluster* yang dihasilkan tidak sesuai, namun penyesuaian dilihat dari jumlah setiap *cluster* yang sama datanya. Hasil lengkap menggunakan aplikasi ditampilkan dalam Gambar 5 dan Gambar 6.

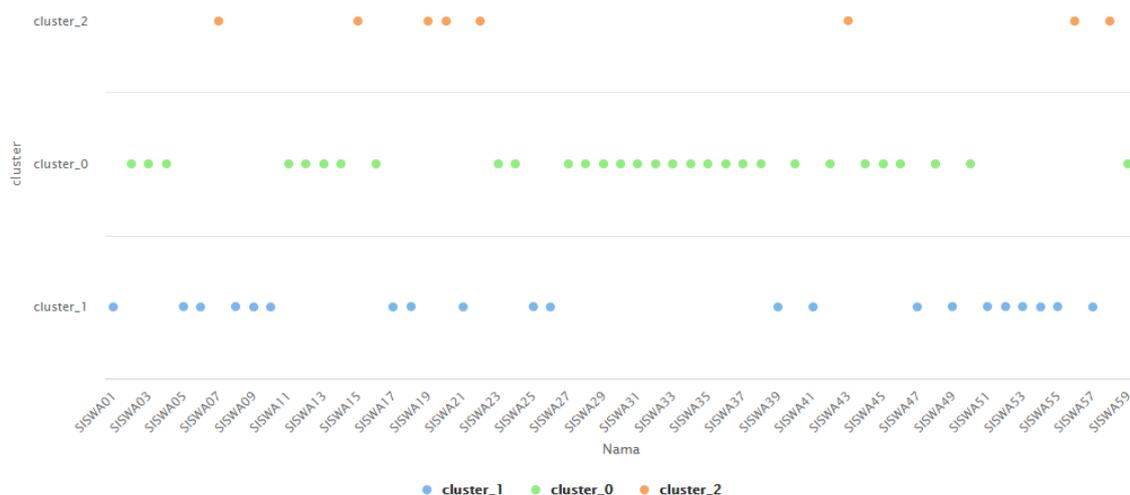


Gambar 5 Hasil per *Cluster*



Attribute	cluster_0	cluster_1	cluster_2
Nilai Keaktifan Pribadi	78.133	57.619	94.250
Nilai Akademik	82.743	82.090	82.312
Nilai Psikomotor	83.380	82.695	82.912
Nilai Afektif	80	80	80
Nilai Absensi	100	100	100

Gambar 6 Hasil Centroid Table



Gambar 7 Hasil Visualisasi Cluster

3.6 Pengujian Hasil (Evaluasi)

Setelah hasil perhitungan *K-Means clustering* melalui RapidMiner dihasilkan, maka akan dilakukan pengujian hasil dari perhitungan melalui aplikasi dengan perhitungan secara manual menunjukkan hasil yang sama.

Dengan *performance vector*-nya adalah sebagai berikut:

Avg. within centroid distance: -36.635
Avg. within centroid distance_cluster_0: -33.268
Avg. within centroid distance_cluster_1: -45.019
Avg. within centroid distance_cluster_2: -27.252
 Davies Bouldin: -0.609

Hasil menunjukkan jarak terdekat dengan *Avg. within centroid distance* adalah C0 yang menunjukkan kelompok siswa aktif. Sedangkan jarak terjauh dengan *Avg. within centroid distance* adalah C1 yang menunjukkan kelompok siswa tidak aktif. Semakin kecil nilai Davies Bouldin maka akan semakin baik nilai *cluster* yang diperoleh dari metode *K-Means*. Davies Bouldin: -0.609 menunjukkan bahwa masing-masing objek dalam *cluster* tersebut memiliki kesamaan yang cukup baik karena mendekati angka 0.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa data prediksi terhadap tingkat keaktifan siswa SMP ABC kelas 8 dapat dikelompokkan menjadi 3 cluster yaitu, *cluster 0* yang menunjukkan siswa aktif sebanyak 30 siswa, *cluster 1* yang menunjukkan siswa tidak aktif sebanyak 8 siswa, dan *cluster 2* yang menunjukkan siswa kurang aktif sebanyak 21 siswa.



Berdasarkan kesimpulan pengelompokan di atas dapat digunakan sebagai acuan bagi pihak sekolah dalam menentukan siswa yang akan direkomendasikan sekolah untuk mengikuti perlombaan atau calon penerimaan beasiswa. Selain itu, dapat dilakukan evaluasi proses pembelajaran pada hasil yang menunjukkan *cluster* siswa tidak aktif agar dapat dilakukan penguatan dalam hal belajar (penguatan afektif, kognitif maupun psikomotorik) di sekolah.

DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Ananda, R. (2020). Penentuan Jumlah Cluster Ideal SNK di Jawa Tengah dengan Metode X-Means Clustering dan K-Means Clustering. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 1–5. <https://doi.org/10.33387/jiko.v3i1.1635>
- Arofah, S. N., & Marisa, F. (2018). Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering. *JOINTECS (Journal of Information Technology and Computer Science)*, 3(2), 85–90. <https://doi.org/10.31328/jointecs.v3i2.787>
- Aroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1), 60–64. <https://doi.org/10.18196/st.211211>
- Elda, Y., Defit, S., Yunus, Y., & Syaljumairi, R. (2021). Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means. *Jurnal Informasi Dan Teknologi*, 3, 103–108. <https://doi.org/10.37034/jidt.v3i3.130>
- Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan RapidMiner dengan K-Means Cluster pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, 3(2), 173. <https://doi.org/10.24114/cess.v3i2.9661>
- Kusuma, A. S., & Aryati, K. S. (2019). Sistem Informasi Akademik Serta Penentuan Kelas Unggulan Dengan Metode Clustering Dengan Algoritma K-Means Di Smp Negeri 3 Ubud. *Jurnal Sistem Informasi Dan Komputer Terapan Indonesia (JSIKTI)*, 1(3), 143–152. <https://doi.org/10.33173/jsikti.29>
- Maulida, L. (2018). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167. <https://doi.org/10.14421/jiska.2018.23-06>
- Nainggolan, R., & Purba, E. (2020). Cluster Analysis of Online Shop Product Reviews Using K-Means Clustering. *IJEBD (International Journal Of Entrepreneurship And Business Development)*, 3(02), 142–151. <https://doi.org/10.29138/ijebd.v3i02.977>
- Nur Khormarudin, A. (2018). *Teknik Data Mining: Algoritma K-Means Clustering*. IlmuKomputer.Com. <https://ilmukomputer.org/2018/05/08/teknik-data-mining-algoritma-k-means-clustering/>
- Oktarian, S., Defit, S., & Sumijan. (2020). Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method. *Jurnal Informasi Dan Teknologi*, 2, 68–75. <https://doi.org/10.37034/jidt.v2i3.65>
- Ong, J. O. (2013). Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri*, 12(1), 10–20. <https://doi.org/10.23917/jiti.v12i1.651>
- Parlambang, B., & Fauziah. (2020). Implementasi Algoritma K-Means dalam Proses Penilaian Kuesioner kepada Dosen Guna Mendukung Kepuasan Mahasiswa Terhadap Dosen. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(2), 161–173. <https://doi.org/10.35760/tr.2020.v25i2.2719>
- Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Assessment Center untuk Clustering Program SDP. *Computer Engineering, Science and System Journal*, 3(1), 87. <https://doi.org/10.24114/cess.v3i1.8192>
- Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K -means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103–119. <https://doi.org/10.1243/095440605X8298>
- Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). Data Mining Seleksi Siswa Berprestasi Untuk



- Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering (Studi Kasus di MTS Darul Fikri). *KOMPUTEK*, 5(1), 88. <https://doi.org/10.24269/jkt.v5i1.686>
- Rahma, L. M. (2020). *Implementasi Algoritma K-Means Berbasis Android Untuk Clustering Tingkat Pemahaman Mahasiswa*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Sirait, W., Defit, S., & Nurcahyo, G. W. (2019). Algoritma K-Means Untuk Klasterisasi Tugas Akhir Mahasiswa Berdasarkan Keahlian. *Jurnal Sistim Informasi Dan Teknologi*, 1(3), 25–30. <https://doi.org/10.35134/jsisfotek.v1i3.6>
- Triandini, M., Defit, S., & Nurcahyo, G. W. (2021). Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia. *Jurnal Informasi Dan Teknologi*, 3, 167–173. <https://doi.org/10.37034/jidt.v3i3.120>
- Vhallah, I., Sumijan, S., & Santony, J. (2018). Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(2), 572–577. <https://doi.org/10.29207/resti.v2i2.308>

