

Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means

Nia Dwi Rahayu¹, Abdul Halim Anshor², Irfan Afriantoro³

^{1,2,3} Universitas Pelita Bangsa, Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi, Indonesia
Email: rahayunia811@gmail.com

Abstrak. Pendidikan memiliki peranan penting sebagai modal utama membangun karakter bangsa. Keberhasilan murid dievaluasi berdasarkan pelajaran teoritis dan praktis, serta kehadiran murid selama di dalam kelas. Jumlah pemrosesan data yang terus meningkat mengharuskan penggunaan strategi dan metode sehingga dapat ditransformasikan menjadi informasi dan pengetahuan yang dapat dimanfaatkan oleh pendidik dalam proses pembuatan kebijakan. Tujuan penelitian ini yaitu mengelompokkan prestasi siswa berdasarkan nilai siswa menggunakan data mining dengan menerapkan algoritma K-Means. Dalam penelitian ini data yang digunakan adalah nilai raport PAS Ganjil SMK TON'S periode tahun 2023/2024. Perhitungan klasterisasi dengan algoritma K-Means berhasil mengelompokkan 24 siswa menjadi 3 cluster. Dimana siswa dengan kategori tinggi sebanyak 6 siswa, siswa dengan kategori cukup sebanyak 8 siswa, dan siswa dengan kategori rendah sebanyak 10 siswa. Berdasarkan hasil evaluasi, didapat nilai Davies Bouldin Index (DBI) 0.144 atau mendekati dengan 0 sehingga dapat disimpulkan bahwa hasil cluster yang dihasilkan cukup baik. Dengan hasil clustering yang telah diterapkan diharapkan bisa membantu dalam mengambil keputusan dengan tepat untuk menentukan siswa berprestasi.

Kata Kunci : Data Mining, Prestasi, Nilai Siswa, Clustering, K-Means

Abstract. Education plays an important role as the main capital in building the nation's character. Student success is evaluated based on theoretical and practical lessons, as well as student attendance during class. The increasing volume of data processing necessitates the use of strategies and methods to transform it into information and knowledge that educators can utilize in the policy-making process. The purpose of this research is to group student achievements based on student grades using data mining by applying the K-Means algorithm. In this study, the data used are the odd semester report card of TON'S Vocational High School for the 2023/2024 period. The clustering calculation with the K-Means algorithm successfully grouped 24 students into 3 clusters, with 6 students in the high category, 8 students in the fair category, and 10 students in the low category. Based on the evaluation results, a Davies Bouldin Index (DBI) value of 0.144 was obtained, which is close to 0, so it can be concluded that the resulting cluster results are quite good. With the clustering results that have been applied, it is hoped that they can help in making precise decisions to determine high-achieving students.

Keyword : Data Mining, Achievement, Student Grades, Clustering, K-Means

PENDAHULUAN

Pendidikan memiliki peranan penting sebagai modal utama membangun karakter bangsa. Pendidikan merupakan suatu cara pembelajaran untuk meningkatkan pengetahuan, sikap dan keterampilan seseorang. Sekolah menjadi sarana berlangsungnya pendidikan secara langsung, dengan bertemunya guru dan siswa[1]. Tingginya prestasi murid dan minimnya angka murid yang tidak berprestasi mencerminkan keunggulan sektor pendidikan. Secara umum, keberhasilan murid dievaluasi berdasarkan pelajaran teoritis dan praktis, serta kehadiran murid selama di dalam kelas. Penilaian dibagi menjadi tiga kategori yaitu pengetahuan, bakat, dan sikap. Pengajar menilai semua murid yang mengikuti pelajaran yang diberikan guna mengevaluasi dan menganalisis prestasi belajar murid[2].

Proses penggalian informasi berdasarkan data siswa terus bertambah setiap tahunnya dan diiringi dengan bertambahnya jumlah siswa, sehingga terjadi penumpukan data yang sangat besar dari data yang belum diolah secara optimal oleh pihak sekolah[3]. Jumlah pemrosesan data yang terus meningkat mengharuskan penggunaan strategi dan metode sehingga dapat ditransformasikan menjadi informasi dan pengetahuan yang dapat dimanfaatkan oleh pendidik dalam proses pembuatan kebijakan. Hal ini menjadi permasalahan bagi guru maupun wali kelas yaitu bagaimana cara menentukan tingkat prestasi murid yang rendah, cukup, dan tinggi



serta menemukan *top rank* murid unggulan dalam kelas, agar dapat membentuk kelas yang ideal untuk meningkatkan prestasi maupun memotivasi murid.

Di bidang pendidikan, pemetaan siswa berprestasi adalah langkah penting dalam meningkatkan kualitas pendidikan dan memberikan dukungan yang tepat kepada siswa. Identifikasi siswa berprestasi dan pemahaman terhadap faktor-faktor yang memengaruhi keberhasilan mereka dapat membantu sekolah dan lembaga pendidikan dalam merancang program-program pendukung yang lebih efektif. *Data Mining* sebagai cabang dari ilmu komputer yang berkaitan dengan penemuan pola dalam *dataset* besar, telah menjadi alat yang penting dalam menganalisis data pendidikan. Metode *clustering*, khususnya *K-Means*, adalah salah satu teknik yang umum digunakan dalam *data mining* untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan fitur.

Adapun beberapa referensi penelitian yang berkaitan dengan pengelompokan menggunakan algoritma *K-Means* sebagai berikut:

Penelitian yang dilakukan Sri Dewi, Sarjon Defit, Yuhandri Yunus (2021) dengan judul Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode *K-Means*[4]. Kesimpulan dari penelitian yang telah dilakukan adalah sampel data yang diolah dalam penelitian ini berjumlah 25 data siswa dan 10 mata pelajaran yang disajikan. Setelah mendapatkan hasil pengelompokan pada iterasi keenam dan dilanjutkan dengan iterasi ke tujuh tidak ada perubahan, maka proses pencarian dihentikan. Hasil *cluster* yang didapatkan pada *Cluster1* (K1) terdapat 3 siswa yang Sangat Berprestasi, pada *Cluster2* (K2) terdapat 10 siswa yang Berprestasi dan *Cluster3* (K3) terdapat 12 siswa yang Kurang Berprestasi.

Penelitian yang dilakukan Salim Kurniawan, Amril Mutoi Siregar, Hilda Yulia Novita (2023) dengan judul Penerapan Algoritma *K-Means* dan *Fuzzy C-Means* Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik[5]. Berdasarkan hasil perhitungan secara manual menggunakan *Microsoft Excel* 2019 dan Program *Python*. *Cluster* yang dihasilkan sama yaitu baik menggunakan algoritma *K-Means* maupun algoritma *Fuzzy C-Means*. Akan tetapi pada perhitungan menggunakan *tools Rapidminer* menghasilkan *cluster* yang berbeda. Dan kesimpulan hasil penelitian dari pengelompokan prestasi siswa berdasarkan nilai akademik menggunakan algoritma *K-Means* dan *Fuzzy C-Means* adalah perhitungan manual dan program *python* pada algoritma *K-Means* menghasilkan 52 siswa berprestasi, 25 siswa dengan prestasi sedang serta 28 siswa tidak berprestasi, adapun pada algoritma *Fuzzy C-Means* menunjukan 48 siswa berprestasi, 29 siswa dengan prestasi sedang dan 28 siswa tidak berprestasi. Perhitungan menggunakan *rapidminer studio* dengan algoritma *K-Means* menghasilkan 49 siswa berprestasi, 27 siswa dengan prestasi sedang serta 29 siswa tidak berprestasi, adapun algoritma *Fuzzy C-Means* tidak dapat diproses pada *rapidminer studio* karena tidak adanya algoritma tersebut dalam *rapidminer studio*. Hasil evaluasi *cluster* dalam mengelompokan prestasi siswa berdasarkan nilai akademik menggunakan algoritma *K-Means* dan *Fuzzy C-Means* dengan metode *Davies Bouldien Index* (DBI) dari setiap jenis perhitungan menunjukan algoritma *K-Means* merupakan algoritma terbaik karena memiliki nilai DBI yang mendekati 0 dan tidak negatif.

Penelitian yang dilakukan Diwa Oktario Dacwanda, Yessica Natalian (2021) dengan judul Implementasi *K-Means Clustering* untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan[6]. Kesimpulan dari penelitian ini penerapan algoritma *k-means* membagi *dataset* penilaian siswa yang meliputi nilai pengetahuan dan keterampilan menjadi tiga kelompok berdasarkan indeks DB, yaitu *cluster* pintar, sedang, dan cukup. Berdasarkan perhitungan *clustering* dengan nilai pengetahuan dan keterampilan didapatkan kelompok pintar dengan jumlah 72 siswa, sedang dengan jumlah 92 siswa, dan cukup dengan jumlah 103 siswa. Sementara itu, perhitungan *clustering* dengan nilai pengetahuan saja didapatkan kelompok pintar dengan jumlah 71 siswa, sedang dengan jumlah 87 siswa, dan cukup dengan jumlah 107 siswa. Dari perbandingan pengelompokan dengan nilai pengetahuan dan keterampilan dan pengelompokan dengan nilai pengetahuan saja didapatkan hasil bahwa terjadi perpindahan *cluster* siswa yang tidak terlalu signifikan, yaitu sebesar 10.15% dari keseluruhan siswa, dimana pengelompokan dengan nilai pengetahuan dan keterampilan lebih baik dibandingkan pengelompokan dengan nilai pengetahuan saja.

Penelitian yang dilakukan Yusma Elda, Sarjon Defit, Yuhandri Yunus, Raemon Syaljumairi (2021) dengan judul Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan *K-Means*[7]. Hasil perhitungan dengan menggunakan fungsi *SQRT* pada *Ms. Excel*, pada iterasi ke empat didapatkan nilai yang sama dengan perhitungan pada iterasi ketiga. Berdasarkan nilai yang sama tersebut, maka proses perhitungan dihentikan cukup sampai dengan iterasi ketiga. Data hasil pengelompokan pada iterasi ketiga menghasilkan 3 *cluster* yang dimana *Cluster1* (C1) berjumlah 47 siswa, *Cluster2* (C2) berjumlah 10 siswa, dan *Cluster3* (C3) berjumlah 33 siswa. Pengujian yang dilakukan dengan menggunakan *Software RapidMiner Studio* Versi 9.2 mendapatkan hasil klasterisasi yang sama dengan pengujian secara



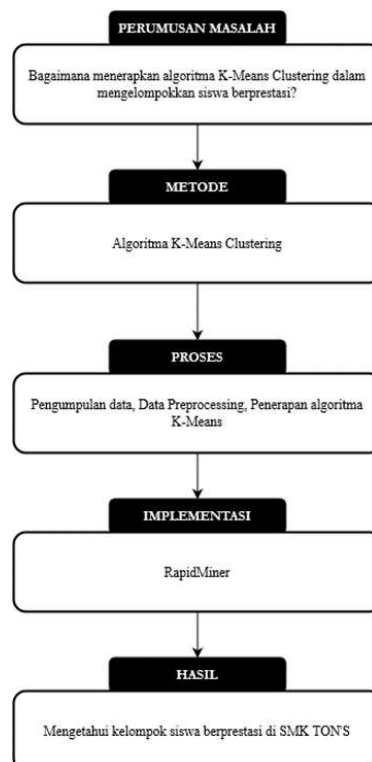
manual dengan *Ms. Excel* yang dimana *Cluster0* (C0) berjumlah 33 siswa, *Cluster1* (C1) berjumlah 47 siswa, dan *Cluster2* (C2) berjumlah 10 siswa. Berdasarkan hasil pengujian dengan menggunakan *Software RapidMiner* memiliki perbedaan penamaan dengan perhitungan menggunakan *Ms. Excel*. Didapatkan hasil berupa *cluster0*, *cluster1*, dan *cluster2*. Sedangkan anggota masing-masing *cluster* sama dengan hasil pengujian dengan menggunakan *Ms. Excel*. Data *cluster0* pada *RapidMiner* sama dengan data pada *cluster3* *Ms. Excel*. Sedangkan untuk *cluster1* dan *cluster2* tidak terdapat perbedaan. Hasil klasterisasi data 90 siswa dikelompokkan menjadi 3 *cluster* dengan jumlah data siswa untuk *cluster 1* dengan bobot sikap dan hasil belajar tinggi berjumlah 47 siswa, *cluster 2* dengan bobot sikap dan hasil belajar sedang berjumlah 10 siswa dan *cluster 3* dengan bobot sikap dan hasil belajar rendah berjumlah 33 siswa.

Penelitian yang dilakukan Melissa Triandini, Sarjon Defit, Gunadi Widi Nurcahyo (2021) dengan judul *Data Mining* dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia dengan Menggunakan Metode *K-Means Clustering*[8]. Dari hasil perhitungan manual dan pengujian dengan menggunakan *software rapidminer* maka diperoleh hasil kelompok yang terdiri 3 *Cluster* yang jumlah anggotanya sama dengan hasil perhitungan manual. Berdasarkan hasil *performance vector* diatas untuk mengukur tingkat keaktifan *cluster* menggunakan *Operator Cluster Distance Performance* mengambil model *cluster centroid* dan mengatur *cluster* sebagai *input* dan mengukur tingkat keaktifan berdasarkan *centroid cluster*. Adapun jarak yang terdekat dengan Avg. *within centroid distance* merupakan nilai yang sangat aktif dan itu terdapat pada *cluster2*, serta yang jarak yang terjauh dengan Avg. *within centroid distance* bernilai kurang aktif dan itu terdapat pada *cluster1*. Dan untuk *cluster0* bernilai aktif. Semakin kecil nilai *davies bouldin index* yang diperoleh (*non-negatif* ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan metode *clustering*. Hasil perhitungan menggunakan metode *k-means* menunjukan nilai -0.842. Angka tersebut memiliki arti masing-masing objek dalam *cluster* tersebut memiliki kesamaan yang cukup baik karena mendekati angka 0. Dari hasil dan pembahasan dapat disimpulkan bahwa data dapat dikelompokkan menjadi 3 *cluster* yaitu, *cluster 1* yang bernilai kurang aktif sebanyak 12 siswa, *cluster 0* yang bernilai Aktif sebanyak 10 siswa, dan *cluster 2* yang bernilai sangat aktif sebanyak 17 siswa.

METODOLOGI PENELITIAN

2.1. Kerangka Berfikir

Berikut kerangka berfikir dibuat untuk mempermudah dalam memahami arah penelitian dari proses hingga hasil yang akan didapatkan.



Gambar 1. Kerangka Berfikir



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 Berprestasi pada SMK TON'S.

2.2. Objek Penelitian

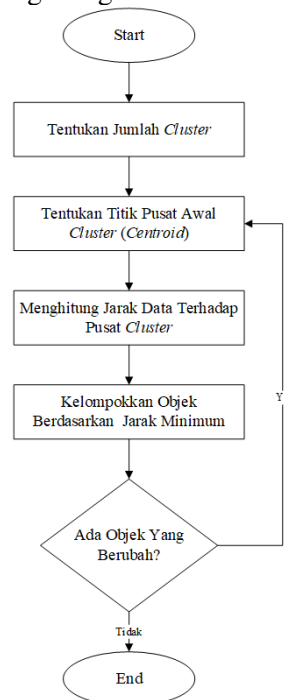
Objek penelitian ini yaitu berupa data nilai siswa tahun ajaran 2023/2024 yang didapatkan dari SMK TON'S. Dimana peneliti bertujuan untuk mengetahui hasil dari pengelompokkan nilai siswa berdasarkan cluster yang sudah ditentukan.

2.3. Data Yang Digunakan

Dalam penelitian ini, jenis data yang digunakan adalah data *primer*, karena data tersebut diambil dari SMK TON'S. Data *primer* berasal dari data yang didapat sendiri oleh perorangan atau suatu organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan dapat berupa wawancara langsung dengan pimpinan serta pada bagian yang menangani langsung permasalahan di lapangan serta observasi kegiatan sehari-hari suatu objek yang diteliti[9]. Data yang digunakan dalam penelitian ini merupakan data nilai siswa kelas XII – Multimedia semester ganjil tahun ajaran 2023/2024 yang terdiri dari 24 *record* data. Variabel data yang digunakan dalam penelitian ini adalah nama siswa dan nilai dari tiap pelajaran yang diikuti oleh siswa.

2.4. Metode Yang Digunakan

Metode yang digunakan pada penelitian ini adalah menggunakan metode *K- Means Clustering*. *K-Means* adalah metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster*. Algoritma *K-Means* termasuk *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah[10]. Langkah-langkah melakukan *clustering* dengan algoritma *K-Means* sebagai berikut:



Gambar 2. Metode Yang Digunakan

Berikut uraian diagram diatas:

1. Menentukan jumlah *cluster*.
2. Menentukan *centroid* awal secara acak.
3. Menghitung jarak setiap data yang ada terhadap setiap pusat *cluster*.
4. Mengelompokkan data hasil perhitungan jarak data pada *centroid*.
5. Menentukan *centroid* baru dari hasil pengklasteran.
6. Menghitung kembali jarak setiap data menggunakan pusat *cluster* yang baru.

HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian

3.1.1. Persiapan *Dataset*

Penelitian ini menggunakan data nilai siswa SMK TON'S kelas XII – Multimedia semester ganjil tahun ajaran 2023/2024. *Dataset* terdiri dari atribut yang telah disederhanakan berupa nama siswa dan nilai siswa untuk 8 mata pelajaran yang telah ditempuh selama satu semester. *Dataset* sampel terdiri dari 24 *record* data. Berikut adalah data siswa yang akan digunakan untuk proses perhitungan dengan algoritma *K-Means*:

Tabel 1. *Dataset*

NAMA	PAI	PPKN	IND	MTK	ENG	TAV	DMI	PKK
Agnes Oktariani	82	81	84	88	80	82	83	84
Aji Sunarya	79	81	84	84	80	82	82	82
Alia Alkhansa	78	80	83	79	80	80	80	80
Arphan Chaerul	79	80	83	79	80	82	82	81
Arya Maulana Syarif	78	80	83	82	79	82	81	82
Farell Joan Susanto	78	80	84	82	79	81	81	80
Fauzan Zahy Zulfahmi	80	81	83	82	79	82	82	84
Hapsah	79	81	83	83	80	83	82	83
Iin Inayah Lutfia	78	81	83	81	80	80	80	80
Kerin Aprillia	81	81	84	81	80	81	83	83
Maulana Yusuf Ibrahim	80	80	83	79	79	82	80	83
Maydera Putri Amelia	82	81	84	89	80	83	83	84
Melany Putri	78	81	83	81	82	82	81	81
Muhamad Dafa Alfaritzi	78	80	83	78	79	82	80	82
Naisa Tiarani	78	80	83	78	79	79	80	80
Niken Aptasari	80	80	84	85	80	82	81	82
Rafli Alif Fikri	84	81	85	89	81	83	83	83
Ratna Damayanti	79	81	84	86	80	82	81	82
Reiza Firjatullah	79	81	83	87	80	83	82	82
Ridwan Nugraha	78	80	83	79	79	82	82	81
Sela Mita Ramjani	78	81	83	88	80	82	81	81
Sri Wahyuni	78	81	83	82	80	81	80	80
Suci	81	81	83	81	80	81	81	82
Tamimah Pajriyah	78	80	83	84	79	81	80	83

3.1.2. 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.

3.1.3. Menentukan *Centroid*

Menentukan *centroid* awal secara acak yang diperoleh dari data sampel. Pada penelitian ini, penulis memilih secara acak dari data nilai siswa untuk dijadikan titik pusat *cluster* pertama.

Tabel 2. Titik Pusat Awal *Cluster*

Centroid	PAI	PPKN	IND	MTK	ENG	TAV	DMI	PKK
C1	82	81	84	88	80	82	83	84
C2	80	81	83	82	79	82	82	84
C3	78	81	83	81	82	82	81	81



3.1.4. Menghitung Jarak Data ke Pusat Cluster

Langkah selanjutnya adalah menghitung jarak terdekat masing-masing *record data* dengan nilai awal *cluster* dengan menggunakan rumus *Euclidean Distance* sebagai berikut:

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

1. Data Pertama

Jarak dengan pusat *cluster* 1:

$$d_1 = \sqrt{(82 - 82)^2 + (81 - 81)^2 + (84 - 84)^2 + (88 - 88)^2 + (80 - 80)^2 + (82 - 82)^2 + (83 - 83)^2 + (84 - 84)^2}$$

= 0

Jarak dengan pusat *cluster* 2:

$$d_1 = \sqrt{(82 - 80)^2 + (81 - 81)^2 + (84 - 83)^2 + (88 - 82)^2 + (80 - 79)^2 + (82 - 82)^2 + (83 - 82)^2 + (84 - 84)^2}$$

= 6,557438524

Jarak dengan pusat *cluster* 3:

$$d_1 = \sqrt{(82 - 78)^2 + (81 - 81)^2 + (84 - 83)^2 + (88 - 81)^2 + (80 - 82)^2 + (82 - 82)^2 + (83 - 81)^2 + (84 - 81)^2}$$

= 9,110433579

Perhitungan dilakukan hingga data terakhir. Hasil perhitungan jarak data terhadap masing-masing *cluster* pada iterasi ke-1 dapat di lihat sebagai berikut:

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

No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
1	Agnes Oktariani	0	6,557438524	9,110433579	0	1
2	Aji Sunarya	5,477225575	3,31662479	4,123105626	3,31662479	2
3	Alia Alkhansa	11,3137085	6,244997998	3,872983346	3,872983346	3
4	Arphan Chaerul	10,09950494	4,582575695	3,31662479	3,31662479	3
5	Arya Maulana Syarif	7,937253933	3,16227766	3,464101615	3,16227766	2
6	Farell Joan Susanto	8,660254038	4,898979486	3,741657387	3,741657387	3
7	Fauzan Zahy Zulfahmi	6,557438524	0	4,898979486	0	2
8	Hapsah	6,164414003	2,236067977	3,872983346	2,236067977	2
9	Iin Inayah Lutfia	9,746794345	5,477225575	3,16227766	3,16227766	3
10	Kerin Aprillia	7,211102551	2,645751311	4,795831523	2,645751311	2

No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
11	Maulana Yusuf Ibrahim	9,8994949 37	3,8729833 46	4,7958315 23	3,8729833 46	2
12	Maydera Putri Amelia	1,4142135 62	7,5498344 35	9,9498743 71	1,4142135 62	1
13	Melany Putri	9,1104335 79	4,8989794 86	0	0	3
14	Muhamad Dafa Alfaritzi	11,489125 29	5,3851648 07	4,5825756 95	4,5825756 95	3
15	Naisa Tiarani	12,369316 88	7,0710678 12	5,4772255 75	5,4772255 75	3
16	Niken Aptasari	4,6904157 6	4,1231056 26	5,1961524 23	4,1231056 26	2
17	Rafli Alif Fikri	3	8,7177978 87	10,677078 25	3	1
18	Ratna Damayanti	4,5825756 95	4,8989794 86	5,6568542 49	4,5825756 95	1
19	Reiza Firjatullah	4,1231056 26	5,6568542 49	6,6332495 81	4,1231056 26	1
20	Ridwan Nugraha	10,488088 48	4,7958315 23	3,8729833 46	3,8729833 46	3
21	Sela Mita Ramjani	5,4772255 75	7,1414284 29	7,2801098 89	5,4772255 75	1
22	Sri Wahyuni	8,8881944 17	5,0990195 14	2,8284271 25	2,8284271 25	3
23	Suci	7,7459666 92	3	3,8729833 46	3	2
24	Tamimah Pajriyah	6,7823299 83	3,8729833 46	5	3,8729833 46	2

Jarak hasil perhitungan akan dilakukan perbandingan diantara 3 *cluster* data terdekat dengan pusat *cluster* dan mengambil nilai terkecil. Jarak ini menunjukkan bahwa data tersebut berada dalam 1 kelompok dengan pusat *cluster* terdekat. Berikut adalah anggota *cluster* yang terjadi pada iterasi ke-1 dilambangkan dengan simbol “T”.

Tabel 4. Anggota *Cluster* Iterasi ke-1

No	Nama	C1	C2	C3
1	Agnes Oktariani	T		
2	Aji Sunarya		T	
3	Alia Alkhansa			T
4	Arphan Chaerul			T
5	Arya Maulana Syarif		T	
6	Farell Joan Susanto			T
7	Fauzan Zahy Zulfahmi		T	
8	Hapsah		T	
9	Iin Inayah Lutfia			T
10	Kerin Aprillia		T	
11	Maulana Yusuf Ibrahim		T	
12	Maydera Putri Amelia	T		
13	Melany Putri			T



No	Nama	C1	C2	C3
14	Muhamad Dafa Alfaritzi			T
15	Naisa Tiarani			T
16	Niken Aptasari		T	
17	Rafli Alif Fikri	T		
18	Ratna Damayanti	T		
19	Reiza Firjatullah	T		
20	Ridwan Nugraha			T
21	Sela Mita Ramjani	T		
22	Sri Wahyuni			T
23	Suci		T	
24	Tamimah Pajriyah		T	

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 6 data, *cluster* 2 sebanyak 9 data, dan *cluster* 3 sebanyak 9 data.

Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

$$C_1 = \frac{(82 + 82 + 84 + 79 + 79 + 78)}{6} = 80,66666667$$

$$C_2 = \frac{(79 + 78 + 80 + 79 + 81 + 80 + 80 + 81 + 78)}{9} = 79,55555556$$

$$C_3 = \frac{(78 + 79 + 78 + 78 + 78 + 78 + 78 + 78 + 78)}{9} = 78,11111111$$

Berikut adalah nilai untuk *centroid* baru:

Tabel 5. Nilai *Centroid* Baru Iterasi ke-2

Centroid	PAI	PPKN	IND	MTK	ENG	TAV	DMI	PKK
C1	80,666 66667	81	83,833 33333	87,833 33333	80,166 66667	82,5	82,166 66667	82,666 66667
C2	79,555 55556	80,555 55556	83,333 33333	82,333 33333	79,555 55556	81,777 77778	81,333 33333	82,666 66667
C3	78,111 11111	80,333 33333	83,111 11111	79,888 88889	79,777 77778	81	80,666 66667	80,555 55556

Selanjutnya menghitung kembali jarak *euclidean distance* seperti pada langkah sebelumnya dan membandingkan hasil *cluster* dan diambil jarak minimum seperti langkah sebelumnya. Langkah tersebut dilakukan hingga tidak ada *cluster* yang berpindah. Berikut hasil perhitungan untuk iterasi ke-2 dapat dilihat pada tabel dibawah:

Tabel 6. Hasil Perhitungan Terhadap *Cluster* Pada Iterasi ke-2

No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
1	Agnes Oktariani	2,1408720 96	6,5977923 73	10,025276 7	2,1408720 96	1
2	Aji Sunarya	4,2720018 73	2,2054925 82	4,8825716 76	2,2054925 82	2
3	Alia Alkhansa	10,242883 71	5,1183523 58	1,6517854 16	1,6517854 16	3
4	Arphan Chaerul	9,2511260 58	3,9126259 69	2,1744873 1	2,1744873 1	3



No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
5	Arya Maulana Syarif	6,8007352 54	1,9657562 24	2,8974232 91	1,9657562 24	2
6	Farell Joan Susanto	7,3654599 31	3,3793125 17	2,5288459 28	2,5288459 28	3
7	Fauzan Zahy Zulfahmi	6,2115483 85	1,7881641 04	4,8711800 48	1,7881641 04	2
8	Hapsah	5,2201532 54	1,8189672 83	4,7674306 56	1,8189672 83	2
9	Iin Inayah Lutfia	8,5195852 79	4,0930534 35	1,8724777 27	1,8724777 27	3
10	Kerin Aprillia	7,0651727 98	2,8631330 5	4,7205879 52	2,8631330 5	2
11	Maulana Yusuf Ibrahim	9,3050165 68	3,7383564 09	3,5364067 7	3,5364067 7	3
12	Maydera Putri Amelia	2,4324199 2	7,5702471 54	10,987647 39	2,4324199 2	1
13	Melany Putri	7,8898669 19	3,6632981 16	2,8196838 98	2,8196838 98	3
14	Muhamad Dafa Alfaritzi	10,594810 05	4,9190985 82	2,7999118 15	2,7999118 15	3
15	Naisa Tiarani	11,441882 13	6,2073234 33	3,0102704 85	3,0102704 85	3
16	Niken Aptasari	3,4034296 43	2,9772802 24	5,8171733 83	2,9772802 24	2
17	Rafli Alif Fikri	3,9475730 94	8,5814902 73	11,773584 16	3,9475730 94	1
18	Ratna Damayanti	2,8722813 23	3,8984010 65	6,5281914 76	2,8722813 23	1
19	Reiza Firjatullah	2,2173557 83	4,9975302 54	7,7284147 83	2,2173557 83	1
20	Ridwan Nugraha	9,5524865 87	4,1869874 85	2,1227747 97	2,1227747 97	3
21	Sela Mita Ramjani	3,5	6,1624109 44	8,2229729 39	3,5	1
22	Sri Wahyuni	7,4777447 58	3,5398960 72	2,3934065 81	2,3934065 81	3
23	Suci	7,1821538 09	2,3517789 43	3,5048467 32	2,3517789 43	2
24	Tamimah Pajriyah	5,6494837 52	2,9016810 79	4,9052755 68	2,9016810 79	2

Diketahui bahwa ada perubahan pada data yang menghasilkan kelompok *cluster* yang baru dimana *cluster* 1 berjumlah 6 siswa, *cluster* 2 berjumlah 8 siswa, dan *cluster* 3 berjumlah 10 siswa. Berikut anggota kelompok *cluster* yang baru:

Tabel 7. Anggota *Cluster* Iterasi ke-2

No	Nama	C1	C2	C3
1	Agnes Oktariani	T		



No	Nama	C1	C2	C3
2	Aji Sunarya		T	
3	Alia Alkhansa			T
4	Arphan Chaerul			T
5	Arya Maulana Syarif		T	
6	Farell Joan Susanto			T
7	Fauzan Zahy Zulfahmi		T	
8	Hapsah		T	
9	Iin Inayah Lutfia			T
10	Kerin Aprillia		T	
11	Maulana Yusuf Ibrahim			T
12	Maydera Putri Amelia	T		
13	Melany Putri			T
14	Muhamad Dafa Alfartizi			T
15	Naisa Tiarani			T
16	Niken Aptasari		T	
17	Rafli Alif Fikri	T		
18	Ratna Damayanti	T		
19	Reiza Firjatullah	T		
20	Ridwan Nugraha			T
21	Sela Mita Ramjani	T		
22	Sri Wahyuni			T
23	Suci		T	
24	Tamimah Pajriyah		T	

Karena masih terdapat hasil *clustering* yang bergerak pada iterasi ke-2, maka perhitungan dilanjutkan pada iterasi ke-3, dan titik pusat *cluster* ditentukan kembali sesuai dengan rata-rata anggota pada *cluster* untuk mendapatkan hasil perhitungan *centroid* baru. Berikut merupakan nilai untuk *centroid* baru:

Tabel 8. Nilai *Centroid* Baru Iterasi ke-3

Centroid	PAI	PPKN	IND	MTK	ENG	TAV	DMI	PKK
C1	80,666 66667	81	83,833 33333	87,833 33333	80,166 66667	82,5	82,166 66667	82,666 66667
C2	79,5	80,625	83,375	82,75	79,625	81,75	81,5	82,625
C3	78,3	80,3	83,1	79,8	79,7	81,1	80,6	80,8

Selanjutnya menghitung kembali jarak *eucclidean distance* seperti pada langkah sebelumnya dan membandingkan hasil *cluster* dan diambil jarak minimum seperti langkah sebelumnya. Berikut hasil perhitungan untuk iterasi ke-3 dapat dilihat pada tabel dibawah:

Tabel 9. Hasil Perhitungan Terhadap *Cluster* Pada Iterasi ke-3

No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
1	Agnes Oktariani	2,140872096	6,219927652	9,956404974	2,140872096	1
2	Aji Sunarya	4,272001873	1,785357107	4,871344784	1,785357107	2
3	Alia Alkhansa	10,24288371	5,402545696	1,769180601	1,769180601	3
4	Arphan Chaerul	9,251126058	4,235268587	2,032240143	2,032240143	3
5	Arya Maulana Syarif	6,800735254	2,106537443	2,816025568	2,106537443	2
6	Farell Joan Susanto	7,365459931	3,418698583	2,670205985	2,670205985	3
7	Fauzan Zahy Zulfahmi	6,211548385	1,920286437	4,661544808	1,920286437	2
8	Hapsah	5,220153254	1,639359631	4,661544808	1,639359631	2



No	Nama	C1	C2	C3	Jarak Terdekat	Cluster
9	Iin Inayah Lutfia	8,519585279	4,235268587	2,080865205	2,080865205	3
10	Kerin Aprillia	7,065172798	2,989565186	4,553020975	2,989565186	2
11	Maulana Yusuf Ibrahim	9,305016568	4,20565096	3,182766093	3,182766093	3
12	Maydera Putri Amelia	2,43241992	7,189401922	10,92382717	2,43241992	1
13	Melany Putri	7,889866919	3,766629793	2,886173938	2,886173938	3
14	Muhamad Dafa Alfaritzi	10,59481005	5,332682252	2,555386468	2,555386468	3
15	Naisa Tiarani	11,44188213	6,51440711	3,05450487	3,05450487	3
16	Niken Aptasari	3,403429643	2,633913438	5,77321401	2,633913438	2
17	Rafli Alif Fikri	3,947573094	8,242420761	11,71025192	3,947573094	1
18	Ratna Damayanti	2,872281323	3,491060011	6,536818798	2,872281323	1
19	Reiza Firjatullah	2,217355783	4,575751304	7,741446893	2,217355783	1
20	Ridwan Nugraha	9,552486587	4,493050189	2,032240143	2,032240143	3
21	Sela Mita Ramjani	3,5	5,760859311	8,302409289	3,5	1
22	Sri Wahyuni	7,477744758	3,596873642	2,555386468	2,555386468	3
23	Suci	7,182153809	2,633913438	3,306055051	2,633913438	2
24	Tamimah Pajriyah	5,649483752	2,772634127	4,850773134	2,772634127	2

Diketahui bahwa tidak ada perubahan pada data yang menghasilkan kelompok *cluster* yang dimana *cluster* 1 berjumlah 6 siswa, *cluster* 2 berjumlah 8 siswa, dan *cluster* 3 berjumlah 10 siswa. Berikut anggota kelompok *cluster* yang baru:

Tabel 10. Anggota *Cluster* Iterasi ke-3

No	Nama	C1	C2	C3
1	Agnes Oktariani	T		
2	Aji Sunarya		T	
3	Alia Alkhansa			T
4	Arphan Chaerul			T
5	Arya Maulana Syarif		T	
6	Farell Joan Susanto			T
7	Fauzan Zahy Zulfahmi		T	
8	Hapsah		T	
9	Iin Inayah Lutfia			T
10	Kerin Aprillia		T	
11	Maulana Yusuf Ibrahim			T
12	Maydera Putri Amelia	T		
13	Melany Putri			T
14	Muhamad Dafa Alfaritzi			T
15	Naisa Tiarani			T
16	Niken Aptasari		T	
17	Rafli Alif Fikri	T		
18	Ratna Damayanti	T		
19	Reiza Firjatullah	T		
20	Ridwan Nugraha			T
21	Sela Mita Ramjani	T		
22	Sri Wahyuni			T
23	Suci		T	
24	Tamimah Pajriyah		T	

Pada perhitungan ini iterasi berhenti di iterasi ke-3 karena tidak ada *cluster* yang berpindah lagi. Hal ini menunjukkan bahwa hasil *cluster* telah mencapai stabil dan *konvergen*.

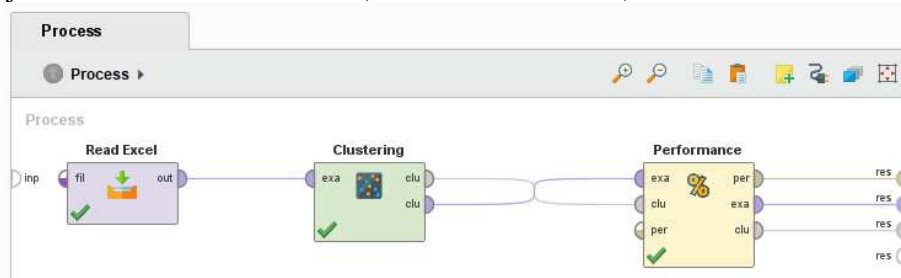


3.2. Pembahasan

3.2.1. Proses Pengujian Menggunakan *RapidMiner*

Pada proses ini, metode klusterisasi dengan algoritma *K-Means* diterapkan untuk pembentukan kelompok *cluster* dengan akurat. Dalam penelitian ini menggunakan pengujian dengan *tools RapidMiner Studio* versi 10.3.1. Adapun langkah-langkahnya sebagai berikut:

1. Sebelum melakukan pengolahan data menggunakan *RapidMiner*, perlu mengunggah *dataset* yang akan digunakan. Operator yang digunakan adalah *Read Excel* karena format data yang digunakan adalah **xlsx (Microsoft Excel Worksheet)*.
2. Kemudian pada kolom “Nama” ubah *type role* datanya menjadi *id* karena kolom ini bersifat *non-numeric* sehingga algoritma *K-Means* tidak akan bekerja jika *type role* kolom “Nama” masih *attribute*.
3. Lalu pada Operator *Modeling – Segmentation – drag Operator k-Means*. Ubah jumlah *k* sesuai dengan jumlah *cluster* yang sudah ditentukan adalah sebanyak 3 *cluster*. Atur *Measure Types* menjadi *BregmanDivergences* karena data yang akan diuji hanya atribut nilai siswa saja yang memiliki *type data integer*. Untuk *Divergence* pilih *SquaredEuclideanDistance* karena perhitungan *clustering* menggunakan rumus *EuclideanDistance* untuk menghitung jarak objek ke pusat *cluster*.
4. Terakhir, *drag Operator Performance (Cluster Distance Performance)*. Tahapan *performance* ini bertujuan untuk mencari nilai DBI (*Davies Bouldin Index*).



Gambar 3. Proses Pengujian *RapidMiner*

3.2.2. Hasil Pengujian Algoritma *K-Means*

Dari hasil pengujian dengan 3 *cluster* maka selanjutnya akan dicari nilai DBI (*Davies Bouldin Index*). Nilai DBI digunakan untuk mengukur kedekatan antar data dalam satu kelompok data. Berikut nilai *Davies Bouldin Index* yang dihasilkan adalah:

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.832
Avg. within centroid distance_cluster_0: 0.796
Avg. within centroid distance_cluster_1: 1.073
Avg. within centroid distance_cluster_2: 0.695
Davies Bouldin: 0.144
```

Gambar 4. Hasil Nilai *Davies Bouldin Index*

Dari gambar diatas dapat dilihat hasil yang diperoleh mendekati 0 yang berarti semakin kecil nilai *Davies Bouldin Index* yang diperoleh maka semakin baik *cluster* yang dihasilkan.

KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan, dapat diambil kesimpulan bahwa penerapan algoritma *K-Means* dalam mengelompokkan siswa berprestasi berdasarkan nilai siswa dapat diterapkan dengan baik. Dari proses pengujian yang dilakukan mendapatkan hasil *cluster_1* (C1) kategori tinggi sebanyak 25% dengan jumlah 6 siswa, hasil *cluster_2* (C2) kategori cukup sebanyak 35% dengan jumlah 8 siswa, dan hasil *cluster_0* (C3) kategori rendah sebanyak 40% dengan jumlah 10 siswa. Berdasarkan hasil pencarian nilai *Davies Bouldin Index* (DBI), hasil dari perhitungan yang telah dilakukan menghasilkan nilai DBI 0.144 atau

mendekati dengan 0 sehingga dapat disimpulkan bahwa hasil *cluster* yang dihasilkan cukup baik. Dengan hasil *clustering* yang telah diterapkan diharapkan bisa membantu dalam mengambil keputusan dengan tepat untuk menentukan siswa berprestasi.

DAFTAR PUSTAKA

- [1] I. Amirulloh, "PEMETAAN KELOMPOK KERJA SISWA DENGAN METODE CLUSTERING K-MEANS DAN ALGORITMA GREEDY," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, Sep. 2019, doi: 10.36499/jinrpl.v1i2.2953.
- [2] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, pp. 424–439, Oct. 2021, doi: 10.51519/journalisi.v3i3.164.
- [3] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [4] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *Jurnal Sistim Informasi dan Teknologi*, pp. 28–33, Mar. 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [5] S. Kurniawan, A. M. Siregar, and H. Y. Novita, "Penerapan Algoritma K-Means dan Fuzzy C-Means Dalam Mengelompokan Prestasi Siswa Berdasarkan Nilai Akademik," vol. IV, no. 1, 2023.
- [6] D. O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *AITI*, vol. 18, no. 2, pp. 125–138, Nov. 2021, doi: 10.24246/aiti.v18i2.125-138.
- [7] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *Jurnal Informasi dan Teknologi*, pp. 103–108, Sep. 2021, doi: 10.37034/jidt.v3i3.130.
- [8] M. Triandini, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia," *Jurnal Informasi dan Teknologi*, pp. 167–173, Sep. 2021, doi: 10.37034/jidt.v3i3.120.
- [9] N. Barkah, E. Sutinah, and N. Agustina, "Metode Asosiasi Data Mining Untuk Analisa Persediaan Fiber Optik Menggunakan Algoritma Apriori," *Jurnal Kajian Ilmiah*, vol. 20, no. 3, pp. 237–248, Sep. 2020, doi: 10.31599/jki.v20i3.288.
- [10] K. I. Wardoyo, M. Maryaningsih, and J. Fredricka, "KLAUSTERISASI SISWA PENYANDANG DISABILITAS BERDASARKAN TINGKAT TUNAGRAHITA MENGGUNAKAN ALGORITMA K-MEANS," *JURNAL MEDIA INFOTAMA*, vol. 19, no. 1, pp. 1–10, Apr. 2023, doi: 10.37676/jmi.v19i1.3307.

