

OPTIMALISASI PRESTASI AKADEMIK SISWA MELALUI PENGELOMPOKAN INDEKS PRESTASI DENGAN *K-MEANS CLUSTERING*

Nana Suarna^{*1}, Nining Rahaningsih², Annisa Annastia Suarna³

¹Teknik Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia

²Komputerisasi Akuntansi, STMIK IKMI Cirebon, Cirebon, Indonesia

³Ilmu Komunikasi, Fakultas Ilmu Sosial dan Ilmu Politik, Universitas Gunung Djati, Cirebon, Indonesia

Email: [1nana.ikmi@gmail.com](mailto:nana.ikmi@gmail.com), [2niningr157@yahoo.co.id](mailto:niningr157@yahoo.co.id), [3annisa.annastia03@gmail.com](mailto:annisa.annastia03@gmail.com)

(Diterima : 23 April 2025, Direvisi : 22 Mei 2025, Disetujui : 25 Mei 2025)

Abstrak

Prestasi akademik merupakan indikator penting dalam menilai kemampuan dan perkembangan siswa selama proses pembelajaran. Namun, banyak lembaga pendidikan yang masih menghadapi tantangan dalam mengidentifikasi kelompok siswa berdasarkan performa akademik secara efisien, sehingga intervensi pembelajaran yang tepat sulit dilakukan. Penelitian ini bertujuan untuk mengoptimalkan prestasi akademik siswa melalui pengelompokan indeks prestasi dengan metode *K-Means Clustering*. Metode *K-Means* digunakan karena kemampuannya dalam mengelompokkan data ke dalam beberapa kluster homogen berdasarkan kesamaan nilai. Data yang digunakan berupa indeks prestasi siswa dari beberapa semester. Proses pengelompokan dilakukan dengan menetapkan sejumlah kluster awal, kemudian dilakukan iterasi untuk mengoptimalkan posisi centroid hingga mencapai konvergensi. Hasil penelitian menunjukkan bahwa siswa dapat dikelompokkan ke dalam tiga kategori utama, yaitu siswa dengan prestasi tinggi, sedang, dan rendah. Pengelompokan ini memberikan gambaran yang lebih jelas kepada pihak sekolah dalam menyusun strategi pembinaan dan pemberian program khusus yang sesuai dengan kebutuhan masing-masing kelompok siswa. Dengan penerapan *K-Means Clustering*, institusi pendidikan dapat melakukan pemantauan dan evaluasi akademik secara lebih terstruktur serta memberikan intervensi yang lebih tepat sasaran. Penelitian ini menyimpulkan bahwa teknik pengelompokan data dapat menjadi solusi inovatif dalam mendukung peningkatan mutu pendidikan melalui pendekatan berbasis data.

Keywords: evaluasi pendidikan, indeks prestasi, *k-means clustering*, pengelompokan data, prestasi akademik.

OPTIMIZATION OF STUDENTS' ACADEMIC PERFORMANCE THROUGH GRADE POINT INDEX CLUSTERING USING *K-MEANS CLUSTERING*

Abstract

*Academic performance is a key indicator in assessing students' abilities and progress throughout the learning process. However, many educational institutions still face challenges in efficiently identifying groups of students based on academic performance, making it difficult to implement targeted learning interventions. This study aims to optimize students' academic achievement through grade point index clustering using the *K-Means Clustering* method. *K-Means* is employed due to its ability to group data into several homogeneous clusters based on value similarities. The data used in this study consists of students' grade point indices from several semesters. The clustering process involves setting an initial number of clusters, followed by iterations to optimize the position of centroids until convergence is achieved. The results show that students can be categorized into three main groups: high achievers, average performers, and low achievers. This clustering provides schools with clearer insights for developing guidance strategies and implementing tailored programs that suit the needs of each student group. By applying *K-Means Clustering*, educational institutions can monitor and evaluate academic performance in a more structured manner and deliver more targeted interventions. This study concludes that data clustering techniques can serve as an innovative solution to support the improvement of education quality through a data-driven approach.*

Keywords: academic performance, data clustering, *k-means clustering*, grade point index, educational evaluation.

1. PENDAHULUAN

Prestasi akademik merupakan salah satu indikator utama dalam menilai keberhasilan proses pembelajaran di lingkungan pendidikan. Dalam praktiknya, banyak sekolah menghadapi tantangan dalam memetakan capaian siswa secara objektif dan efisien, yang berakibat pada kurang optimalnya strategi pembinaan individual [1]. Sementara itu, data indeks prestasi yang dimiliki seringkali belum dimanfaatkan secara maksimal untuk mendukung pengambilan keputusan berbasis data (*data-driven decision making*) [2].

Beberapa penelitian sebelumnya telah menerapkan metode *clustering* dalam pengelompokan siswa berdasarkan prestasi akademik mereka, seperti pada penelitian [3] berpendapat Proses pembelajaran melalui berbagai interaksi dan pengalaman belajar memiliki pengaruh yang cukup besar untuk mengembangkan keaktifan siswa guna meningkatkan mutu pendidikan. Penelitian sebelumnya [4] menyimpulkan bahwa penerapan metode *data mining* dalam bidang pendidikan, termasuk teknik klasifikasi dan *clustering* untuk analisis prestasi siswa. Namun, sebagian besar fokus pada aspek prediktif tanpa mengarahkan hasilnya pada strategi pembinaan konkret [5]. Penelitian [1] menunjukkan bahwa faktor jarak rumah ke sekolah memiliki pengaruh yang sangat terbatas terhadap prestasi akademik siswa, sementara penghasilan orang tua tidak berpengaruh signifikan. Melalui penerapan algoritma *K-Means*, data prestasi siswa dikelompokkan menjadi dua kategori, yaitu "Tinggi" dan "Rata-rata". Selanjutnya, model *Decision Tree* digunakan untuk mengklasifikasikan faktor-faktor yang memengaruhi prestasi tersebut. Namun, dari hasil evaluasi diketahui bahwa algoritma *Decision Tree* tidak cocok digunakan untuk memprediksi prestasi siswa dalam penelitian ini, karena tingkat akurasi hanya 68% dan nilai *AUC* sebesar 0.561, yang berada dalam kategori *failure* menurut skala performa *AUC*. Penelitian ini menyimpulkan bahwa jenis kelamin menjadi faktor utama dalam perbedaan prestasi, di mana siswa perempuan cenderung memiliki prestasi lebih tinggi dibandingkan siswa laki-laki. Sedangkan [6] Menyajikan hasil pengelompokan data nilai siswa SMAN 1 Tenganan berdasarkan rata-rata nilai mata pelajaran dan absensi selama semester satu hingga lima. Penelitian ini menggunakan algoritma *K-Means* dengan bantuan software RStudio dan menentukan jumlah *cluster* optimal melalui metode *Elbow*, *Silhouette*, dan *Gap*. Hasil menunjukkan bahwa siswa dikelompokkan ke dalam tiga kategori, yaitu rendah, cukup, dan tinggi. Dari pengelompokan tersebut, ditemukan 8 siswa pada kelompok tinggi, yang kemudian dianalisis lebih lanjut untuk menentukan *top rank* menggunakan metode *Simple Additive Weighting (SAW)*. Adapun [7] menyajikan hasil pengelompokan siswa berdasarkan nilai akademik yang terdiri dari dua kategori, yaitu *knowledge* (pengetahuan) dan *skill* (keterampilan), sesuai kurikulum 2013. Dengan menggunakan algoritma *K-Means*, siswa dikelompokkan menjadi tiga kelompok, yaitu pintar, sedang, dan cukup. Hasil menunjukkan bahwa dari total 266 siswa, terdapat 71 siswa dalam kelompok pintar, 92 siswa dalam kelompok sedang, dan 103 siswa dalam kelompok cukup. Selain itu, analisis menunjukkan bahwa nilai keterampilan memiliki pengaruh signifikan terhadap pengelompokan akademik siswa. Jurnal ini merekomendasikan agar sekolah memanfaatkan hasil pengelompokan untuk menyesuaikan strategi pengajaran, sehingga dapat meningkatkan keseimbangan antara nilai pengetahuan dan keterampilan.

Penerapan *K-Means Clustering* sendiri telah digunakan dalam berbagai sektor untuk segmentasi data, namun implementasinya secara spesifik dalam konteks pengelompokan prestasi akademik siswa masih terbatas dan belum banyak diadopsi oleh institusi pendidikan tingkat menengah [5], [8]. Penelitian ini dilakukan sebagai respons terhadap kebutuhan akan pendekatan analisis data yang lebih adaptif, guna mengoptimalkan pemetaan dan pembinaan prestasi siswa. Dengan menggunakan algoritma *K-Means*, diharapkan proses pengelompokan siswa berdasarkan indeks prestasi dapat memberikan dasar yang kuat bagi pengambilan keputusan yang lebih tepat sasaran, seperti pemberian bimbingan akademik, program remedial, atau percepatan belajar [9]. Tujuan penelitian ini adalah untuk menerapkan metode *K-Means Clustering* dalam mengelompokkan siswa berdasarkan indeks prestasi mereka, serta mengevaluasi efektivitas pendekatan ini dalam mendukung peningkatan kualitas pendidikan. Kebaruan dari penelitian ini terletak pada integrasi analisis kluster sebagai alat pendukung strategi akademik yang lebih terarah dan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan metode analisis data *K-Means Clustering* untuk mengelompokkan siswa berdasarkan Indeks Prestasi Akademik (IPA) yang dimiliki masing-masing. Proses dimulai dengan pengumpulan data IPA siswa sebagai basis data awal. Data yang diperoleh kemudian melalui tahap pra-proses, termasuk normalisasi, guna meminimalisir kesalahan dan memastikan keakuratan hasil analisis [10].

Setelah data dinormalisasi, algoritma *K-Means* diterapkan untuk melakukan pengelompokan (*clustering*) siswa ke dalam beberapa kategori berdasarkan kemiripan nilai akademik mereka. Penentuan jumlah kluster (*cluster*) yang optimal dilakukan menggunakan pendekatan *Knowledge Discovery in Database (KDD)* [9][10], dan analisis ini didukung oleh aplikasi RapidMiner sebagai alat bantu pemrosesan data [6]. Hasil dari pengelompokan ini diharapkan dapat memberikan informasi yang bermanfaat bagi pihak sekolah dalam merancang strategi pembelajaran yang lebih terarah dan efektif, sesuai dengan karakteristik dan kemampuan akademik masing-masing kelompok siswa.

Algoritma *K-Means* bekerja dengan cara membagi data ke dalam k kelompok (*cluster*), di mana setiap data akan dimasukkan ke dalam kluster berdasarkan kedekatan nilai terhadap pusat kluster (*centroid*)

Rumus utama yang digunakan untuk menghitung jarak (biasanya jarak Euclidean) antara data x dan pusat kluster μ adalah [13].

$$d(x, \mu) = \sqrt{\sum_{i=1}^n (x_i - \mu_i)^2} \quad (1)$$

Di mana:

- x_i adalah nilai fitur ke-iii dari data,
- μ_i adalah nilai fitur ke-iii dari pusat kluster,
- $d(x, \mu)$ adalah jarak *Euclidean* antara titik data dan pusat kluster.

Tujuan dari *K-Means* adalah meminimalkan total jarak kuadrat antara setiap titik data dan pusat klasternya [14].

$$J = \sum_{j=1}^k \sum_{x_i \in C_j} ||x_i - \mu_j||^2 \quad (2)$$

Dengan:

- J adalah fungsi objektif yang ingin diminimalkan,
- C_j adalah kluster ke- j
- μ_j adalah pusat dari kluster C_j .
- x_i adalah data dalam kluster tersebut [15].

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, digunakan data siswa sebanyak 279 yang belum diproses, setelah melalui proses transformasi data jumlahnya berkurang menjadi 168 data siswa dengan 9 atributnya. Pembahasan dalam penelitian akan difokuskan pada tiga aspek yang melibatkan menerapkan algoritma *K-Means* untuk mengelompokkan siswa berdasarkan nilai indeks prestasi akademik mereka, mengidentifikasi kelompok-kelompok prestasi akademik siswa berdasarkan hasil pengelompokan menggunakan algoritma *K-Means* dan menyediakan informasi yang dapat digunakan oleh pihak sekolah untuk membuat keputusan terkait strategi peningkatan kualitas pendidikan berdasarkan hasil pengelompokan indeks prestasi akademik siswa.

3.1. Penerapan algoritma *K-Means* pada pengelompokan siswa berdasarkan hasil indeks prestasi akademik siswa.

Penerapan algoritma *K-Means* dalam penelitian ini bertujuan untuk mengelompokkan siswa berdasarkan pola nilai akademik masing-masing. Pengelompokan dilakukan menggunakan data nilai rapor guna memperoleh gambaran yang lebih jelas mengenai performa akademik siswa di SMK Veteran, serta untuk memahami perbedaan tingkat kinerja siswa secara lebih terstruktur. Proses pengelompokan diawali dengan melakukan serangkaian uji coba terhadap jumlah kluster, dimulai dari 2 hingga 10 *cluster*, untuk menentukan jumlah kluster yang optimal. Evaluasi hasil pengelompokan dilakukan menggunakan *Davies Bouldin Index* (DBI), yang merupakan salah satu metrik validasi internal dalam pengukuran kualitas kluster. Berdasarkan hasil evaluasi tersebut, diperoleh bahwa konfigurasi dua kluster menghasilkan nilai DBI terendah, yaitu sebesar 0,738, yang menunjukkan bahwa pengelompokan pada jumlah dua kluster memiliki kualitas terbaik dibandingkan jumlah kluster lainnya. Dua kluster tersebut memiliki karakteristik akademik yang berbeda secara signifikan, sehingga dapat dikategorikan sebagai *Cluster* "Baik" dan *Cluster* "Sangat Baik". Pengelompokan ini diharapkan dapat membantu pihak sekolah dalam menyusun strategi pembelajaran yang lebih efektif dan sesuai dengan kebutuhan masing-masing kelompok siswa.

3.2. Hasil identifikasi kelompok-kelompok prestasi akademik siswa berdasarkan hasil pengelompokan menggunakan algoritma *K-Means*.

Berdasarkan hasil yang diperoleh dari penerapan model pengelompokan tingkat prestasi akademik siswa menggunakan algoritma *K-Means*, diperoleh dua kluster siswa dengan karakteristik nilai rata-rata yang berbeda. *Cluster_0*, yang dikategorikan sebagai *kelompok baik*, terdiri dari 141 siswa dengan nilai rata-rata berada dalam rentang 70 hingga 80. Nilai tertinggi dalam kelompok ini mencapai 80,8, sedangkan nilai terendah berada pada angka 70,8. Sebagian besar siswa dalam kluster ini memiliki nilai rata-rata antara 73 hingga 78, yang menunjukkan bahwa mereka memiliki capaian akademik yang cukup baik. Sementara itu, *Cluster_1*, yang dikategorikan sebagai *kelompok sangat baik*, terdiri dari 27 siswa dengan nilai rata-rata lebih tinggi, yakni dalam rentang 80 hingga di atas 90. Nilai tertinggi dalam kluster ini mencapai 90,7, dan sebagian besar siswa memiliki nilai rata-rata antara 82 hingga 87. Hal ini menunjukkan bahwa siswa dalam kelompok ini memiliki prestasi akademik yang lebih unggul dibandingkan dengan siswa di *Cluster_0*. Melalui penerapan model ini, siswa dapat dikelompokkan ke dalam dua kategori utama berdasarkan prestasi akademik mereka, yaitu kelompok baik (*Cluster_0*) dan kelompok sangat baik (*Cluster_1*). Hasil pengelompokan ini diharapkan dapat menjadi acuan dalam perancangan strategi pembelajaran yang lebih tepat sasaran sesuai dengan karakteristik masing-masing kelompok siswa.

3.3. Hasil model pengelompokan tingkat prestasi akademik siswa.

Berdasarkan hasil pengelompokan indeks prestasi akademik siswa di SDN Lebakwangi menggunakan algoritma *K-Means*, diperoleh dua kategori utama yang mencerminkan pola capaian akademik siswa, yaitu *Cluster_0* (Baik) dan *Cluster_1* (Sangat Baik). *Cluster_0* mencakup 141 siswa dengan nilai rata-rata 76.242, yang sebagian besar berada dalam rentang nilai 71–81. Siswa dalam kelompok ini menunjukkan hasil yang cukup baik, terutama pada mata pelajaran PJOK (80.284) dan SBDP (76.766), namun masih terdapat mata pelajaran inti seperti Matematika yang nilainya relatif lebih rendah (75), sehingga memerlukan perhatian dan penguatan pembelajaran. Sementara itu, *Cluster_1* terdiri dari 27 siswa dengan rata-rata nilai 85.181, menunjukkan performa akademik yang sangat baik secara menyeluruh, terutama pada mata pelajaran PKN (89.074), Bahasa Indonesia (88.963), dan Matematika (83.889). Kategori ini terdiri dari siswa dengan nilai 80–90, yang menggambarkan pemahaman akademik yang mendalam di berbagai mata pelajaran. Pengelompokan ini mengacu pada ketentuan nilai Kriteria Ketuntasan Minimal (KKM), dan dapat menjadi dasar bagi sekolah untuk memberikan intervensi yang lebih tepat, seperti dukungan belajar tambahan bagi siswa di *Cluster_0*, serta program pengayaan atau pembelajaran lanjutan yang sesuai potensi bagi siswa di *Cluster_1*. Seperti pada tabel 5 berikut.

Tabel 1. Pengkategorian Nilai KKM

No	Rentang Nilai	Kategori
1.	87 - 100	Sangat Baik
2.	76 - 87	Baik
3.	65-76	Cukup
4.	64	Kurang

3.4. Analisis dan Penerapan Metode KDD

Berdasarkan hasil penelitian yang dilakukan, dengan mengumpulkan data nilai rapor siswa dalam bentuk file excel. Langkah selanjutnya menganalisis dan menerapkan metode *Knowledge Discovery in Database (KDD)* yang mencakup lima tahapan, yaitu *Data Selection, Preprocessing Data, transformation Data, Data Mining* dan *Evaluation*[16].

a. Data

Informasi yang diperoleh berupa hasil nilai rapor siswa semester genap tahun 2024 di dengan total 279 data siswa. Seperti pada tabel 1 berikut :

Tabel 1. Data siswa

No	Nama	Kelas	Pkn	Pai	B. Indo	Mt k	Sbk p	Pjo k	B. Sunda	B. Ing	Ip s	Ip a	Rata-rata
1	RAFI HARDIANTSYAH	8B	78	86	89	85	88	85	78	84	77	79	80,25
2	AISAH PUTRI RAMADHANI	8B	84	85	89	87	79	85	76	85	78	79	83,25
3	ANDINI TASYA RAHMA	8B	78	84	87	85	78	83	75	81	77	75	80,67
4	ANTON FEBRIANTO	8B	75	83	84	80	79	82	74	79	75	76	78,92
5	MEITRA RAMADHANI	8B	79	79	84	81	80	78	72	80	75	75	77,75
6	AJENG KARTIKA	8B	77	79	83	78	75	79	71	80	75	74	76,75
7	DWI PUTRA	8B	77	78	80	77	74	78	70	79	73	74	75,83
8	ALFIA KHOIRUNNISA	8B	76	78	79	76	73	77	69	77	73	72	74,83
9	ILHAM RIZKY SAPUTRA	8B	76	77	79	75	72	76	68	75	71	70	73,58
10	OCTAVIA RIANA	8B	75	77	77	73	72	75	67	75	70	70	72,5
11	CARYTA VIRGINIA	8B	75	76	76	72	71	75	66	73	69	69	71,67
12	PUTRI APRILIANA	8B	74	75	75	71	70	74	65	72	68	68	70,67
13	SAYYIDA AMELIA YASMIN	8B	74	74	74	70	69	73	64	71	67	67	69,75
14	ZAKKY AHMAD ZIKRY	8B	73	74	74	70	69	72	63	70	66	67	69,25
15	MUHAMAD MARUF	8B	73	73	73	69	68	71	62	69	65	66	68,5

16	RAIHAN ARDI	8B	72	72	72	68	67	70	61	68	64	65	67,75
17	DESTIA AHMAD	8B	71	71	71	67	66	69	60	67	63	64	70
18
19

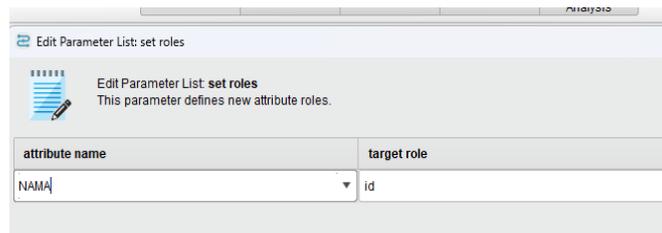
b. Data Selection

Dari data rapor nilai siswa yang terdiri dari 12 atribut seperti kelas, mata pelajaran Pendidikan agama Islam, Pendidikan Kewarganegaraan, Bahasa Indonesia, Matematika, Seni Budaya dan Prakarya, Pendidikan Jasmani olahraga dan kesehatan, Bahasa Sunda, Bahasa Inggris, IPA, IPS, dan Nilai rata-rata. Selanjutnya dilakukan penyeleksian data termasuk atribut yang akan digunakan dengan menghilangkan atribut IPA, IPS dan kelas. Di dalam data *selection* melibatkan penggunaan operator read excel berfungsi untuk mengakses dan membaca informasi data yang terdapat dalam file MS-Excel. Untuk import data dan memasukan operator read excel, Langkahnya sebagai berikut:



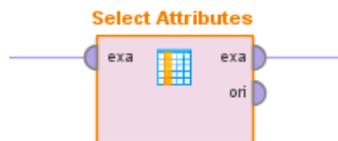
Gambar 1. Operator read excel

Setelah memasukan data langkah selanjutnya menambahkan operator *set role* yang bertujuan untuk menjadikan nama sebagai *id* pada atribut, dengan cara mengklik *edit list* pada parameter lalu pilih dan *apply*. Gambar *set role* seperti tampilan gambar 2.

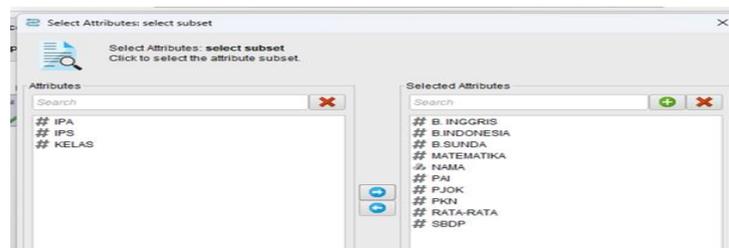


Gambar 2. Hasil Set Role

Langkah selanjutnya menambahkan operator *select attribute* untuk menentukan atribut yang digunakan, dengan mengklik di parameter *attribute filter type* kemudian pilih *a subset*. Pada tahap ini tidak menggunakan attribute IPA, IPS dan kelas karena kelas 1 dan kelas 2 tidak memiliki nilai IPA dan IPS sedangkan tidak menggunakan atribut kelas karena pengelompokan ini satu sekolah tidak pengelompokan perkelas. Terlihat pada gambar 3 berikut.

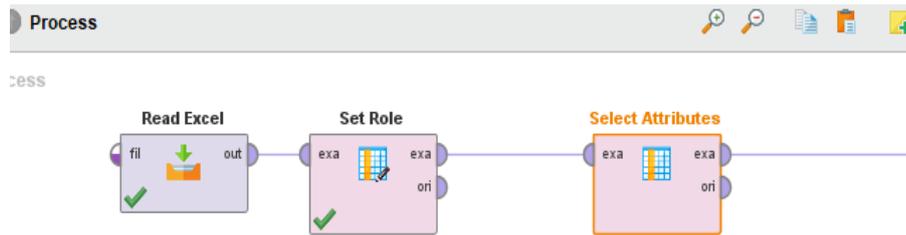


Gambar 3. Select Attribute



Gambar 4. Hasil Select Attribute

Model pemrosesan pada tahap Data Selection rapidminer terlihat pada gambar 5 berikut.



Gambar 5. Hasil *Set Role*

Gambar 5 menunjukkan alur proses dalam aplikasi RapidMiner, yang merupakan salah satu platform *data science* visual. Berikut penjelasan ringkas dari setiap komponen dalam alur tersebut:

1. *Read Excel*: Operator ini digunakan untuk membaca data dari file Excel. Output dari proses ini berupa dataset yang akan digunakan untuk analisis lebih lanjut.
2. *Set Role*: Operator ini digunakan untuk menentukan peran (role) dari atribut tertentu, seperti menetapkan atribut mana yang menjadi *label* (target variabel) atau *id*. Ini penting untuk membedakan fitur dengan target dalam proses pembelajaran mesin.
3. *Select Attributes*: Operator ini memungkinkan pengguna untuk memilih atribut mana saja yang akan digunakan dalam proses selanjutnya. Atribut yang tidak dipilih akan dikeluarkan dari dataset.

c. *Data Preprocessing*

Pada tahap *preprocessing*, menggunakan penggunaan operator *remove duplicates* dengan tujuan untuk menghapus data yang duplikat agar hasil *clustering* akurat. Berdasarkan hasil dari operator *remove duplicates* yang telah dilakukan, terdapat data yang duplikat sebanyak 3. Jadi, data yang awalnya 309 data menjadi 279 data. Terlihat pada gambar 6 berikut.



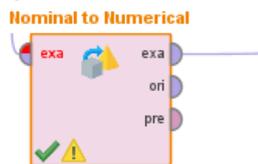
Gambar 6. Memasukan Operator *Remove Duplicates*

Row No.	NAMA	PAI	PKN	BLINDONESIA	MATEMATIKA	SDP	PJOK	BSUNDA	B.INGGRIS	RATA-RATA
1	ADIVA RAIN	80	85	85	80	80	80	75	70	79.375
2	ALISHA PUT...	80	80	80	75	75	85	75	70	77.500
3	ANCI TAHILA	80	80	80	80	75	85	75	70	78.125
4	ANDINI SRI F.	80	80	75	75	75	85	75	70	76.875
5	ARSYLA MEL	75	75	75	75	73	80	70	70	74.125
6	ARUMI RIZKI	85	90	95	90	85	90	80	70	85.625
7	ARYA KUSU...	75	75	75	75	74	80	70	80	75.500
8	ASYIFA RAYS	77	75	80	75	70	85	72	70	75.500
9	AZKA ADITYA	74	75	75	75	70	80	70	75	74.250
10	AZKA KHARL	80	75	80	80	75	90	72	70	77.750
11	CAHYA WIGU	75	75	75	75	70	80	70	70	73.750
12	EKA PUTRA	73	70	70	70	70	80	70	75	72.250
13	FARADIBA AS	85	85	90	80	85	90	85	70	85
14	KEYSHA APR	75	70	75	80	70	85	70	75	75
15	LUCKY MILH	80	80	85	85	80	85	75	70	80

Gambar 7. Hasil dari *Remove Duplicates*

d. *Data Transformation*

Pada tahap *data transformation* ini menggunakan operator *nominal to numerical* untuk merubah jika ada nilai yang nominal akan dirubah kedalam format numerik agar dapat digunakan dalam proses analisis data tampilan operator *nominal to numeric* seperti pada gambar 8.



Gambar 8. Operator *nominal to numerical*

Row No.	NAMA	PAI	PKN	B.INDONESIA	MATEMATIKA	SBEP	PUKJ	B.SUNDA	B.INGGRIS	RATA-RATA
1	ADHA RANI	80	85	85	80	80	80	75	70	79,375
2	ALISHA PUT.	80	80	80	75	75	85	75	70	77,500
3	ANGI TAPLA.	80	80	80	80	75	85	75	70	78,125
4	ANDRI SRI P.	80	80	75	75	75	85	75	70	76,875
5	ARSYA MEI	75	75	75	75	73	80	70	70	74,125
6	ARUM RIZKI	85	90	95	90	85	90	80	70	85,625
7	ARYA KUSU.	75	75	75	75	74	80	70	80	75,500
8	ASWIA RAYIS	77	75	80	75	70	85	72	70	75,500
9	AZKA ADITYA.	74	75	75	75	70	80	70	75	74,250
10	AZKA KHAR.	80	75	80	80	75	90	72	70	77,750
11	CAHYA WIGU.	75	75	75	75	70	80	70	70	73,750
12	EKA PUTRA.	73	70	70	70	70	80	70	75	72,250
13	FARINDA AS.	85	85	80	90	85	90	85	70	85
14	KEVISHA APRIL	75	70	75	80	70	85	70	75	75
15	LUCKY MUKH	80	80	85	85	80	85	75	70	80

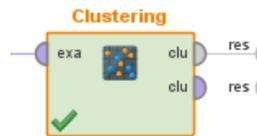
Gambar 9. Hasil dari operator *nominal to numerical*

Gambar 9 menampilkan tampilan *ExampleSet* dari RapidMiner setelah data berhasil dimuat, kemungkinan dari file Excel, seperti yang ditunjukkan pada gambar sebelumnya. Berikut penjelasan ringkas:

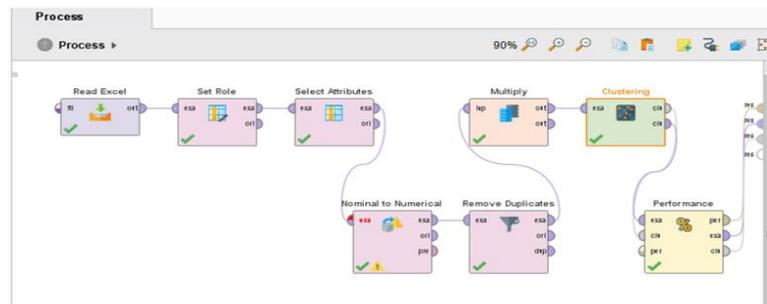
1. Data: Merupakan nilai-nilai akademik siswa dalam berbagai mata pelajaran, seperti PAI, PKN, B.INDONESIA, MATEMATIKA, IPA, IPS, B.SUNDA, B.INGGRIS, dan RATA-RATA.
2. Kolom pertama (Nama): Menunjukkan nama-nama siswa.
3. Kolom-kolom lainnya: Berisi nilai dalam bentuk numerik dari masing-masing mata pelajaran.
4. Kolom RATA-RATA: Merupakan nilai rata-rata dari semua mata pelajaran untuk setiap siswa.

e. *Data Mining*

Setelah data dinormalisasikan dengan parameter *multiply*, lalu proses data *mining* ini menggunakan operator *K-Means* untuk mengelompokkan siswa berdasarkan nilai indeks prestasi akademik mereka. Dengan cara menghitung *cluster* dari 2 sampai 10 dengan membandingkan nilai *DBI* yang paling kecil, lalu setelah mendapatkan nilai *DBI* yang paling kecil maka *cluster* tersebut merupakan *cluster* yang optimal. Dengan bantuan operator *performance* untuk mengevaluasi kinerja dari model *K-Means*, yang mengukur seberapa baik data tersebut dikelompokkan dalam *cluster*, dengan memilih *main criterion Davies Bouldin* dan *maximize* pada parameter.



Gambar 10. Memasukan operator *K-Means* pada *Rapidminer*



Gambar 11. Model pemrosesan pada tahap *Data Mining*

Gambar 11 menampilkan proses analisis data menggunakan RapidMiner dengan alur yang lebih kompleks. Berikut penjelasan ringkas tiap komponen:

1. Read Excel: Membaca data dari file Excel.
2. Set Role: Menentukan peran dari atribut (misalnya, mana yang menjadi ID, label, dll).
3. Select Attributes: Memilih atribut yang relevan untuk analisis.
4. Multiply: Menggandakan dataset agar bisa digunakan di dua jalur proses yang berbeda secara paralel.
5. Nominal to Numerical: Mengubah atribut bertipe nominal (kategori) menjadi numerik agar bisa digunakan dalam algoritma yang membutuhkan data numerik.
6. Remove Duplicates: Menghapus data duplikat untuk memastikan keunikan dan akurasi hasil.
7. Clustering: Melakukan pengelompokan data menggunakan algoritma *clustering* seperti *K-Means*.
8. Performance: Mengevaluasi performa dari hasil *clustering*.

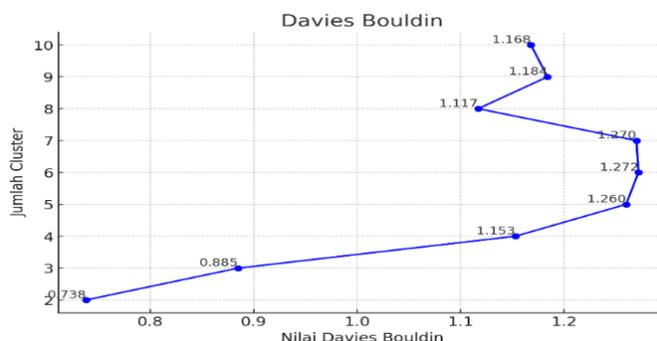
3.5. Evaluasi Hasil

Pada tahap ini telah dilakukan pengujian terhadap penentuan jumlah *cluster* menggunakan operator *K-Means*. Pengujian *cluster* dimulai dari *cluster* 2 sampai dengan *cluster* 10 dengan melihat nilai *DBI* yang paling kecil, dengan nilai *DBI* yang semakin kecil (*non negatif*) jumlah *cluster* yang optimal akan semakin baik. Hal ini dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Hasil Uji *Cluster*

No	Cluster	Anggota Cluster	Nilai DBI
1.	Cluster 2	Total number of items: 168	0.738
2.	Cluster 3	Total number of items: 168	0.885
3.	Cluster 4	Total number of items: 168	1.153
4.	Cluster 5	Total number of items: 168	1.260
5.	Cluster 6	Total number of items: 168	1.272
6.	Cluster 7	Total number of items: 168	1.270
7.	Cluster 8	Total number of items: 168	1.117
8.	Cluster 9	Total number of items: 168	1.184
9.	Cluster 10	Total number of items: 168	1.168

Supaya lebih jelas melihat nilai hasil *Clusternya*, tabel 2 yang menampilkan nilai *DBI* dari hasil *Cluster* 2 sampai 10. Dari gambar tersebut terlihat bahwa *Cluster* 2 memiliki nilai *DBI* yang paling terkecil diantara *Cluster* lainnya, yaitu sebesar 0.738. Seperti tampilan gambar 12.



Gambar 12. Grafik Hasil Evaluasi Nilai Indeks *DBI*

Pada *cluster* tabel 3 merupakan hasil *Cluster* yang menampilkan nilai titik pusat (*centroid*) setiap *Clusternya*, dimana titik pusat tersebut menggambarkan posisi rata-rata data dalam setiap kelompok.

Tabel 3. Anggota *Cluster 0*

Cluster 0			
NO	NAMA	NILAI RATA-RATA	KATEGORI
1	Adiva Rain Destria	79.375	Baik
2	Alisha Putri Ramadhani	77.5	Baik
3	Andi Tahlana Pratama	78.125	Baik
4	Andini Sri Febriani	76.875	Baik
5	Arsyla Meidisa Putri	74.125	Cukup
6	Arya Kusuma Jaya	75.5	Cukup
7	Asyifa Raysha Putri	75.5	Cukup
8	Azka Aditya Irawan	74.25	Cukup
9	Azka Khairunnisa	77.75	Baik
10	Cahaya Wiguna	73.75	Cukup
..	Muhamad Denies A	79.3	Baik
...	Bintang Cahya Putra	78.3	Baik
141	Rasya Purnama Saputra	78.1	Baik

Tabel 4. Anggota Cluster 1

Cluster 1			
NO	NAMA	NILAI RATA-RATA	KATEGORI
1	Arumi Rizkiya Ramadhani	85.625	Baik
2	Faradiba Assyauqia Sya'bani	85	Baik
3	Nazwan Annafi	81.25	Baik
4	Alcina Kianinara Putri Gumi	85.375	Baik
5	Anandara Farrania H	84.375	Baik
6	Salsabila Kalifa Fauni	84.375	Baik
7	Syaqila Anggita	82	Baik
8	Agni Nurfitri Ramadhani	90.2	Sangat Baik
9	Alsahda Elviana Nur P	90.7	Sangat Baik
10	Anisa Dinda Septiani	87.8	Sangat Baik
...	Rafi Padilah	80.8	Baik
...	Sabila Indriani	83.8	Baik
27	Anindia Keysa Putri	81.6	Baik

4. KESIMPULAN

Setelah melakukan eksperimen dan menganalisis data dalam penelitian ini, penulis dapat menyimpulkan bahwa penerapan algoritma *K-Means* dalam pengelompokan data nilai rapor siswa berhasil menghasilkan dua *cluster* optimal berdasarkan nilai *Davies Bouldin Index* (DBI) terendah sebesar 0.738, yang menunjukkan perbedaan karakteristik akademik yang jelas antar kelompok. Hasil analisis menunjukkan terbentuknya dua kelompok, yaitu *Cluster_0* (Baik) dengan 141 siswa yang memiliki rata-rata nilai antara 74.922 hingga 80.284, dan *Cluster_1* (Sangat Baik) dengan 27 siswa yang rata-rata nilainya berkisar antara 80.111 hingga 89.407. *Cluster_0* mencerminkan siswa dengan capaian akademik cukup baik namun masih memerlukan peningkatan, khususnya dalam mata pelajaran inti seperti Matematika. Sementara itu, siswa dalam *Cluster_1* menunjukkan performa akademik yang lebih tinggi dan stabil di berbagai mata pelajaran. Model ini dapat dimanfaatkan oleh pihak sekolah untuk memberikan intervensi pendidikan yang lebih tepat sasaran, seperti bimbingan tambahan bagi siswa di *Cluster_0* serta program pembelajaran yang lebih menantang bagi siswa di *Cluster_1*.

DAFTAR PUSTAKA

- [1] A. Majid, M. Bari, Y. M. Cani, and U. Enri, "Penerapan Algoritma *K-Means* dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 355, 2022, doi: 10.30865/json.v4i2.5299.
- [2] B. Sembiring, S. Natalia, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma *K-Means*," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [3] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma *K-Means* Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022, doi: 10.14421/jiska.2022.7.2.111-121.
- [4] S. Haviyola, Susilawati, and M. Jajuli, "Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma *K-Means*," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2786–2791, 2024, doi: 10.36040/jati.v7i4.7200.
- [5] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma *K-Means* Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [6] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering *K-Means*," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [7] D. O. Dacwanda and Y. Nataliani, "Implementasi *K-Means* Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *Aiti (Jurnal Teknol. Informasi)*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [8] Q. I. Mawarni and E. S. Budi, "Implementasi Algoritma *K-Means* Clustering Dalam Penilaian Kedisiplinan Siswa," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 522, 2022, doi: 10.30865/json.v3i4.4242.
- [9] N. K. K. M. J. and F. A., "Penerapan Algoritma *K-Means* Untuk Pengelompokan Penilaian Akhir Semester

- Di Sdn Kadokan 01 Sukoharjo,” *INFOTECH J.*, vol. 9, no. 1, pp. 190–197, 2023, doi: 10.31949/infotech.v9i1.5343.
- [10] N. Nurahman, A. Purwanto, and S. Mulyanto, “Klasterisasi Sekolah Menggunakan Algoritma *K-Means* berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.
- [11] P. S. Hasugian and J. Raphita Sagala, “KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Pengelompokan Siswa Berdasarkan Nilai Akademik dengan Algoritma *K-Means*,” *Media Online*, vol. 3, no. 3, pp. 262–268, 2022, doi: <https://djournals.com/klik>.
- [12] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, “Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering *K-Means*,” *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 71–83, 2024, doi: 10.53842/juki.v6i1.474.
- [13] A. Sulistiyawati and E. Supriyanto, “Implementasi Algoritma *K-Means* Clustering dalam Penentuan Siswa Kelas Unggulan,” *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [14] I. Suputra, I. Candiasa, and I. Suryawan, “Klasterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma *K-Means*,” *Wahana Mat. dan Sains J. Mat. Sains, dan Pembelajarannya*, vol. 15, no. 1, pp. 22–30, 2021, doi: <https://ejournal.undiksha.ac.id/index.php/JPM/article/view/25380>.
- [15] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma *K-Means* Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *Skatika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skatika.v6i1.2982.
- [16] D. K. Yestiani and N. Zahwa, “Peran Guru dalam Pembelajaran pada Siswa Sekolah Dasar,” *Fondatia*, vol. 4, no. 1, pp. 41–47, 2020, doi: 10.36088/fondatia.v4i1.515.