

BAB IV
HASIL DAN PEMBAHASAN

4.1. Data Penelitian

Sumber Data Data yang digunakan adalah nilai akademik siswa kelas VI SD Negeri 05 Bilah Barat pada beberapa mata pelajaran utama. Jumlah Siswa Penelitian ini melibatkan seluruh siswa kelas VI SD Negeri 05 Bilah Barat yang memiliki data nilai lengkap. Total siswa yang dianalisis adalah 24 siswa. Tahun Ajaran Tahun ajaran yang digunakan dalam penelitian ini adalah 2025. Variabel yang Digunakan:

1. Bahasa Indonesia
2. Matematika
3. IPA (Ilmu Pengetahuan Alam)
4. IPS (Ilmu Pengetahuan Sosial)
5. PPKn (Pendidikan Pancasila dan Kewarganegaraan)

Tabel 4. 1 Data Nilai Rapor Semester Ganjil Siswa Kelas VI

No	Nama	PKN	B.INDO	MTK	IPA	IPS
1	A***** R*****	80	75	74	75	75
2	B***** H*****	70	70	70	70	70
3	A***** R*****	76	70	70	70	70
4	F***** Y*****	73	71	70	70	70
5	R***** M*****	82	80	75	75	76
6	N***** R*****	85	85	75	72	73
7	Y*** S*****	70	70	70	70	70
8	M***** A*****	80	70	73	75	75
9	A***** S***	76	72	73	75	74
10	A*****	70	70	70	70	70

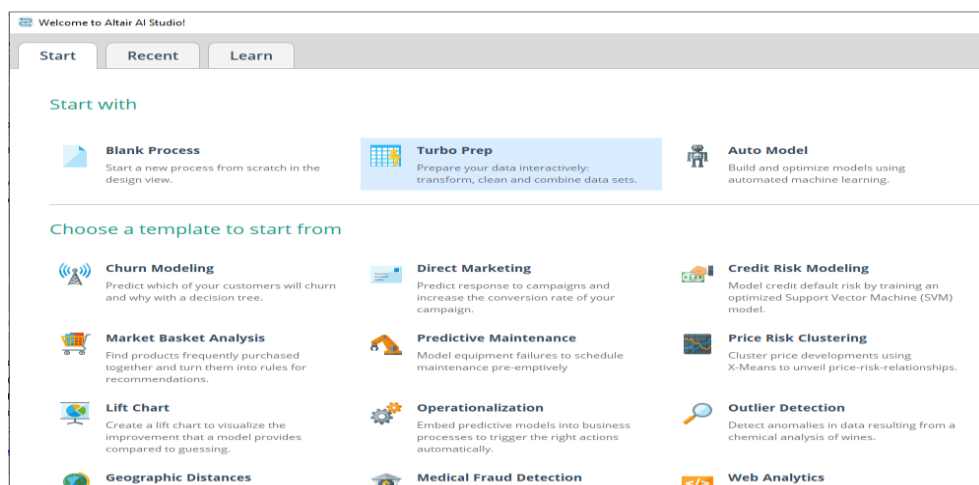
11	S***** R****	80	78	72	74	73
12	D*** M*****	75	72	70	70	70
13	H*** S*****	86	86	75	78	78
14	T**** R*****	72	72	70	72	72
15	W**** F*****	77	70	71	73	72
16	M***** R***	75	70	70	70	70
17	A***** H*****	74	70	70	72	70
18	S*** F*****	78	85	74	73	74
19	D*** P*****	75	72	71	73	73
20	F** A*****	75	70	70	70	70
21	M***** A***	70	70	70	72	70
22	N***** E***	72	73	72	73	72
23	R**** S*****	71	70	71	72	70
24	D**** P****	72	72	70	70	70
25	A**** T*****	73	70	70	70	70
26	R*****	70	68	65	60	60
27	L***** H***	70	65	66	60	60
28	S*** F*****	80	70	70	60	66
29	A**** H*****	80	75	70	70	60
30	F*** L***	88	78	70	72	69
31	A*** R*****	80	70	70	68	60
32	K**** S*****	80	70	70	68	60
33	F*** D*****	80	75	75	70	70
34	H***** C*****	82	70	70	65	68
35	R*****	70	65	68	65	65
36	W***** H*****	70	70	70	70	66
37	N*** N*****	78	70	70	70	67
38	Y*** B***	80	75	72	70	69
39	D**** M*****	82	72	71	70	60
40	D***** J*****	80	75	71	70	60
41	K**** H*****	85	80	78	78	75
42	M**	80	70	70	68	73
43	K*****	80	78	74	75	75
44	N***** K*****	78	70	70	70	70
45	R**** D*****	75	70	70	70	68

46	N*****	80	75	74	71	70
47	C*** I*****	82	80	75	75	76
48	O*** V*****	85	85	75	72	73
49	F*** B*****	70	70	70	70	70
50	M***** H*****	80	70	73	75	75

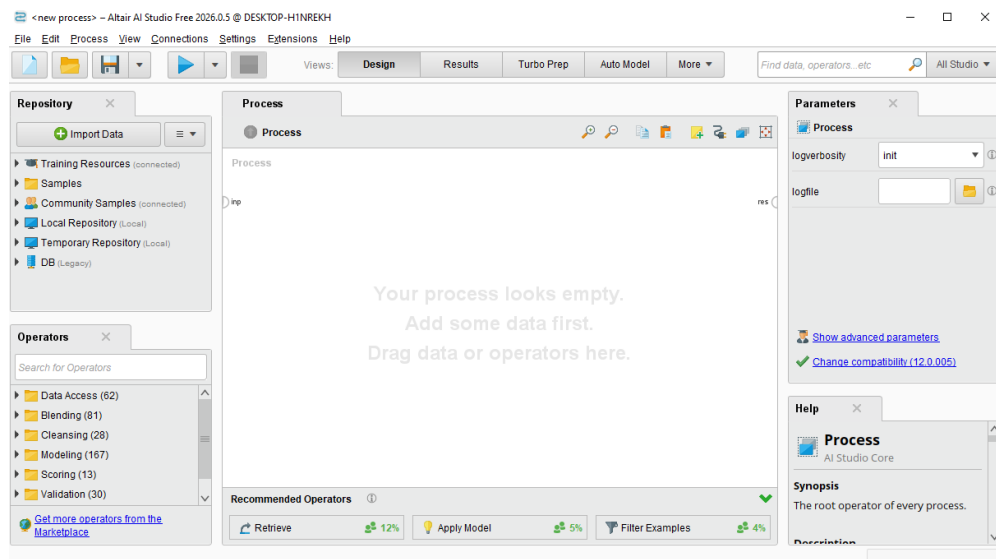
Dalam penelitian ini, algoritma K-Means diterapkan menggunakan perangkat lunak RapidMiner untuk menganalisis tingkat prestasi akademik siswa di SDN 05 Bilah Barat. Pemanfaatan RapidMiner memungkinkan proses clustering dilakukan secara lebih cepat dan akurat, sehingga hasil pengelompokan data prestasi akademik menjadi lebih efektif dan efisien.

4.1.1 Implementasi Rapid Miner

Langkah awal dalam penerapan algoritma K-Means menggunakan RapidMiner untuk menganalisis prestasi akademik siswa di SDN 05 Bilah Barat adalah membuka aplikasi RapidMiner. Setelah aplikasi dijalankan, akan muncul tampilan awal yang menyediakan opsi untuk memulai lembar kerja baru. Pada tahap ini, pengguna memilih menu “Blank Process” untuk memulai proses analisis.

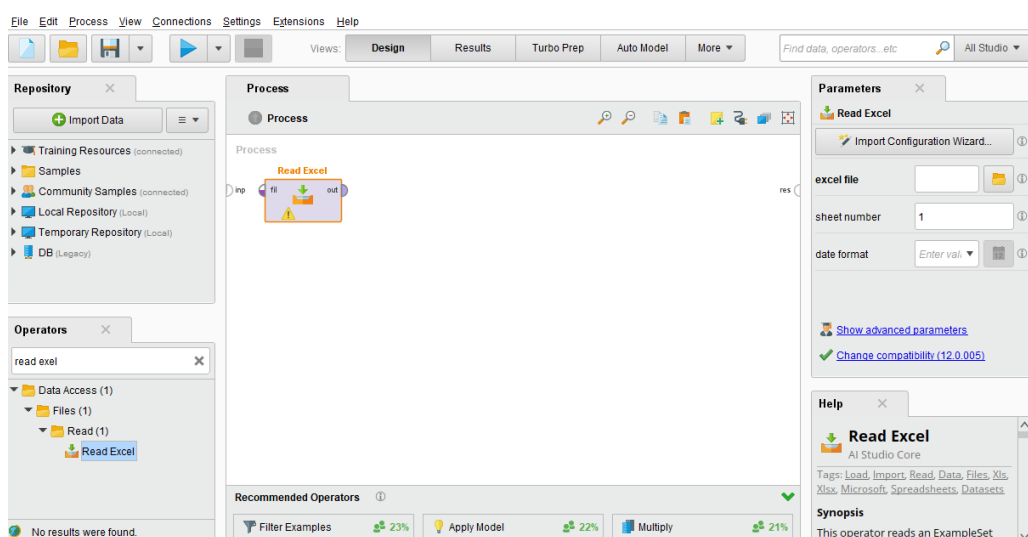


Gambar 4. 1 Tampilan Untuk Memulai Lembar Kerja Baru Pada RapidMiner



Gambar 4. 2 Tampilan Kosong Halaman Lembar Kerja Baru Pada RapidMiner

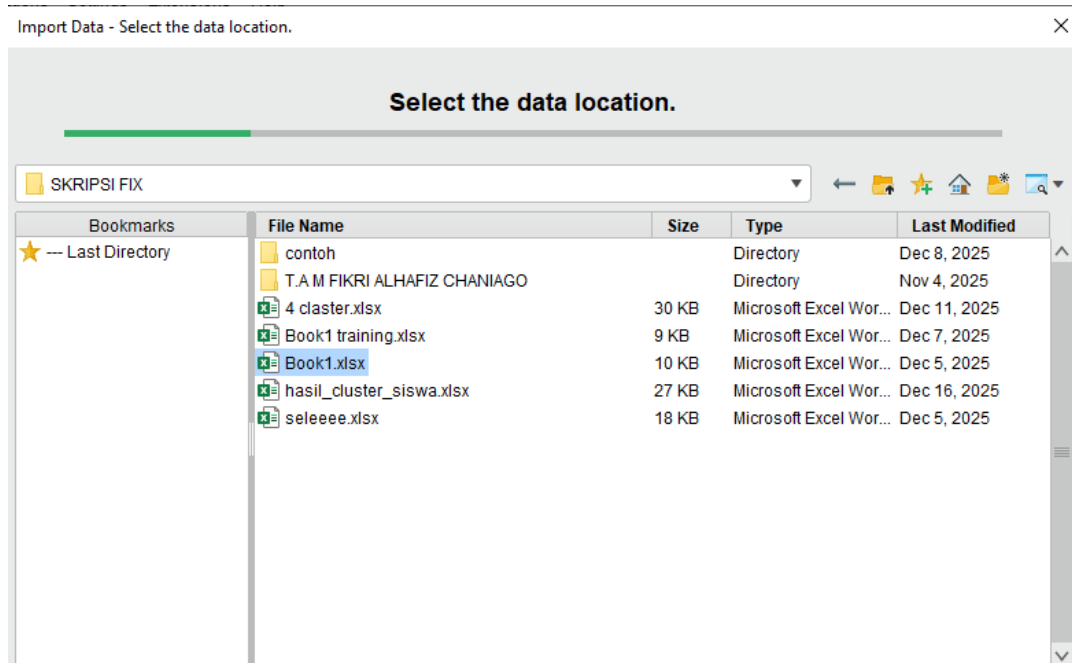
Tahap selanjutnya adalah mengimpor data nilai akademik siswa ke dalam lembar kerja RapidMiner. Pada tahap ini digunakan operator “*Read Excel*” yang dapat ditemukan melalui fitur pencarian pada bagian “*Operators*”. Setelah itu, operator tersebut dimasukkan ke dalam area proses dengan metode *drag and drop*.



Gambar 4. 3 Proses Menginput Operator *K – Means* Ke Dalam Lembar Kerja

Langkah berikutnya adalah memilih menu “*Import Configuration Wizard*” pada bagian “*Parameters*” dari operator “*Read Excel*”. Setelah itu, akan muncul tampilan

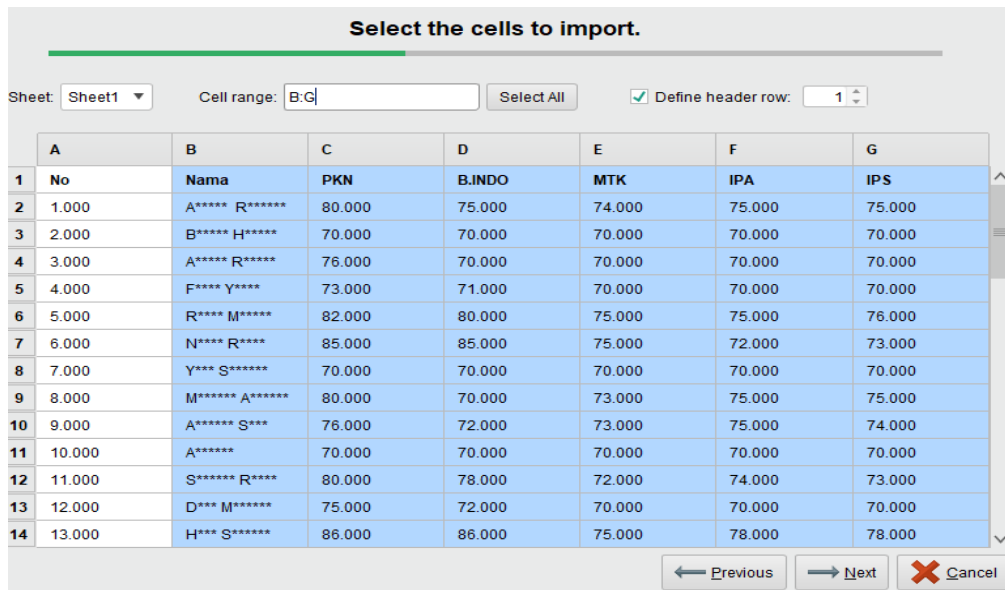
“Select the Data Location”, kemudian pengguna memilih lokasi penyimpanan data nilai akademik siswa. Selanjutnya, klik tombol “Next” untuk melanjutkan proses.



Gambar 4. 4 Proses Memilih Lokasi Penyimpanan Data

Setelah menekan tombol “Next”, sistem akan menampilkan *jendela “Select the Cells to Import”*. Selanjutnya, pengguna menyeleksi seluruh data nilai siswa yang akan digunakan, kemudian melanjutkan ke tahap berikutnya dengan mengklik

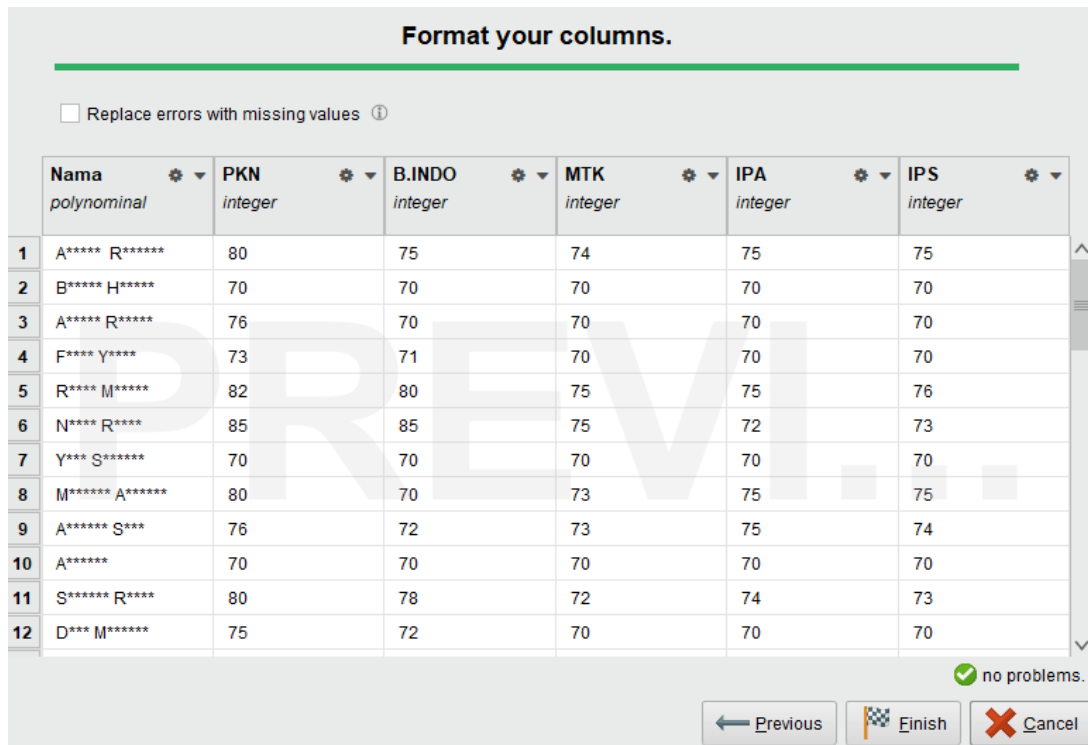
“Next”.



Gambar 4. 5 Proses Memilih Data Yang Akan Di Input

Tahap selanjutnya adalah menentukan label atau kelas dari dataset yang digunakan dengan memilih opsi “Change Role” dan menetapkannya sebagai label. Setelah itu,

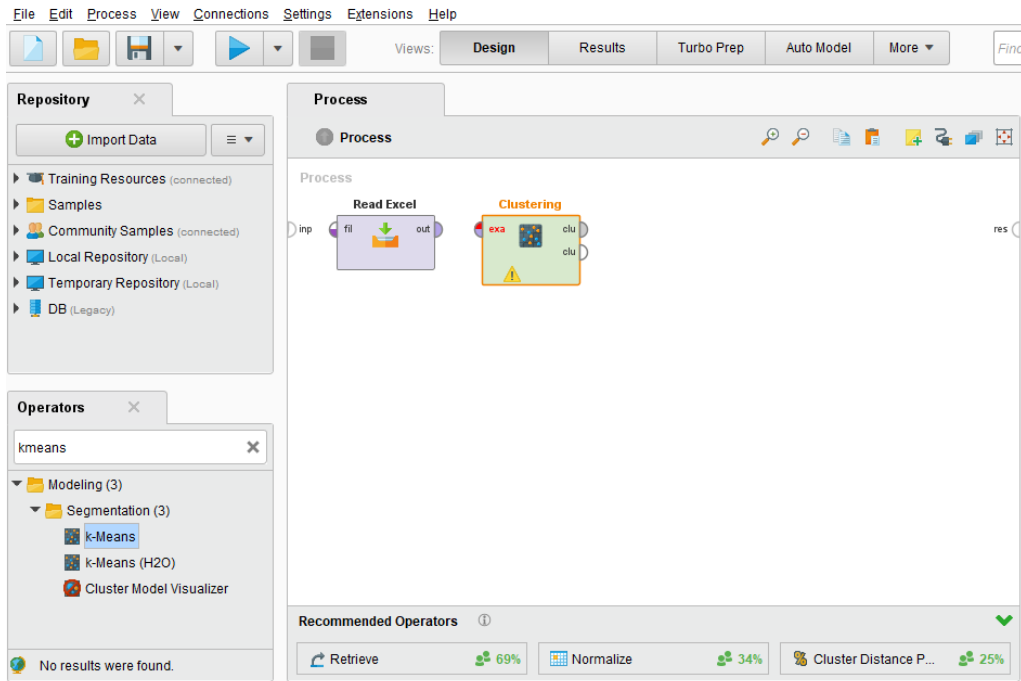
proses diakhiri dengan menekan tombol “Finish”



Gambar 4. 6 Proses Memilih Label Dari Dataset

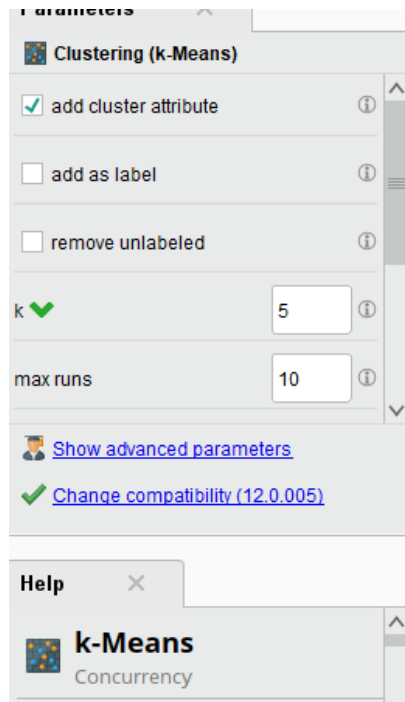
Gambar 4.6 menunjukkan proses pemilihan label dari dataset. Setelah tahap impor data nilai siswa selesai, langkah selanjutnya adalah menambahkan operator “K-Means” yang diperoleh melalui fitur pencarian pada bagian “Operators”. Operator

tersebut kemudian dimasukkan ke dalam area proses dengan *metode drag and drop*.



Gambar 4. 7 Proses Menginput Operator *K – Means* Ke Dalam Lembar Kerja

Selanjutnya pada *Parameters Clustering (K – Means)*, tentukan jumlah *cluster*. Pada penelitian ini, data dibagi dalam 3 *cluster*.

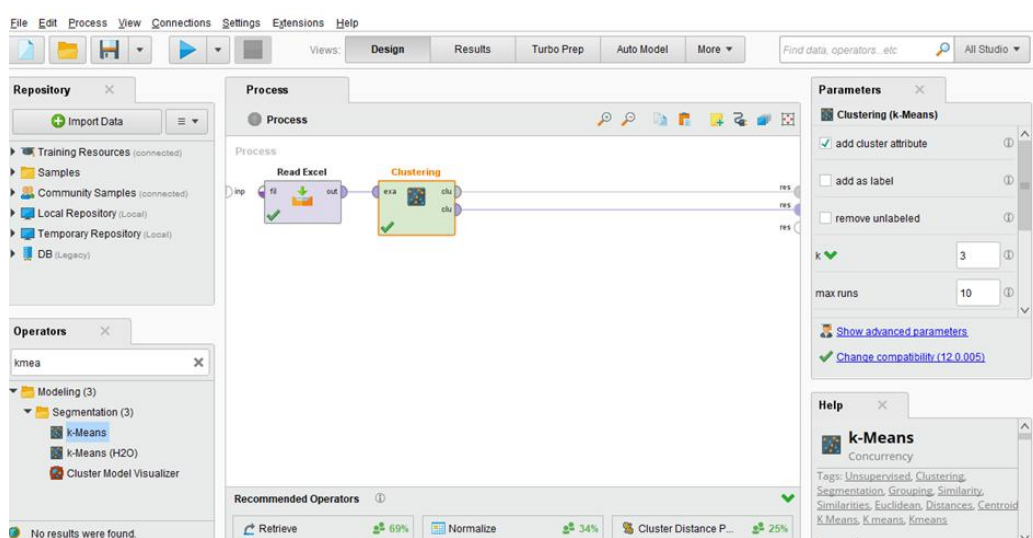


Gambar 4. 8 Proses Menentukan Jumlah Cluster

Tahap selanjutnya adalah menghubungkan antar operator dengan menarik garis dari output “*Read Excel*” ke input pada operator “*Clustering (K-Means)*”, kemudian dilanjutkan dengan menghubungkan output dari “*Clustering (K-Means)*” ke bagian “*res*”. Setelah seluruh rangkaian proses selesai, klik tombol “*Run*” untuk

menjalankan

analisis.



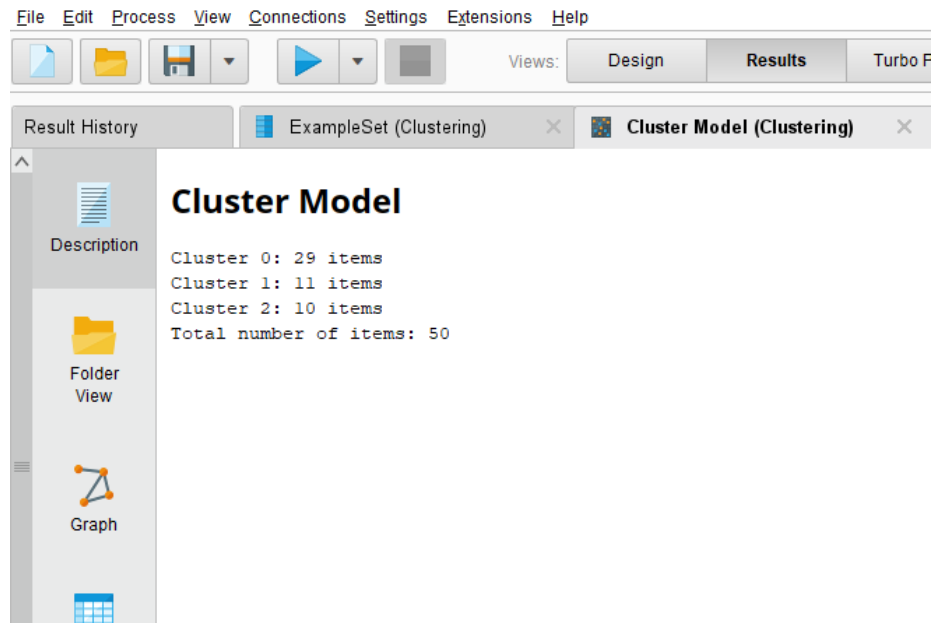
Gambar 4. 9 Proses Menyambungkan Operator

Hasil dari proses dapat dilihat di “ *Result* ” yang menampilkan data yang sudah diolah secara lengkap oleh *cluster* yang sudah ditentukan.

Row No.	Nama	cluster	PKN	B.INDO	MTK	IPA	IPS
7	Y*** S*****	cluster_0	70	70	70	70	70
8	M***** A*****	cluster_0	80	70	73	75	75
9	A***** S***	cluster_0	76	72	73	75	74
10	A*****	cluster_0	70	70	70	70	70
11	S***** R****	cluster_1	80	78	72	74	73
12	D*** M*****	cluster_0	75	72	70	70	70
13	H*** S*****	cluster_1	86	86	75	78	78
14	T**** R*****	cluster_0	72	72	70	72	72
15	W**** F*****	cluster_0	77	70	71	73	72
16	M***** R***	cluster_0	75	70	70	70	70
17	A***** H*****	cluster_0	74	70	70	72	70
18	S*** F*****	cluster_1	78	85	74	73	74
19	D*** P*****	cluster_0	75	72	71	73	73

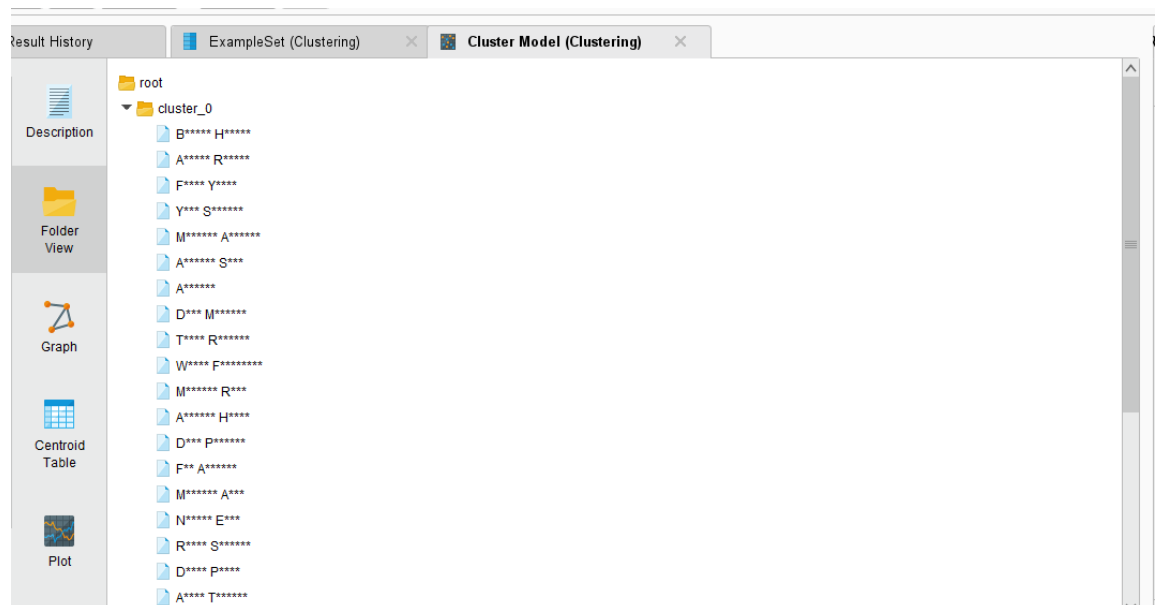
Gambar 4. 10 Hasil Clustering Data

Cluster Model merupakan data yang menampilkan jumlah total data yang telah diolah secara lengkap. Berikut ini gambaran hasil dari *Cluster Model*. Berikut ini gambaran dari hasil *Cluster Model*.



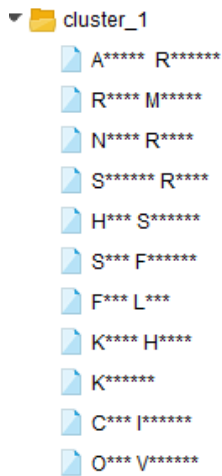
Gambar 4. 11 Hasil dari Cluster Model

Folder View merupakan tampilan halaman isi dari data yang diolah secara keseluruhan lengkap dengan pengelompokan *clusternya*. Berikut ini gambaran dari *Folder View* untuk *Cluster 0*.



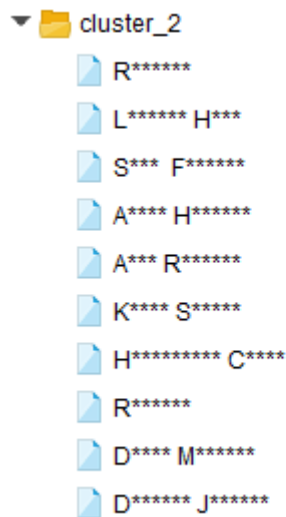
Gambar 4. 12 Hasil Pengelompokan Cluster_0

Berikut ini gambaran dari *Folder View* untuk *Cluster 1*.



Gambar 4. 13 Hasil Pengelompokan *Cluster_1*

Berikut ini gambaran dari *Folder View* untuk *Cluster 2*.



Gambar 4. 14 Hasil Pengelompokan *Cluster_2*

Hasil dari penerapan algoritma *K – Means* dengan RapidMiner menunjukkan bahwa data berhasil dikelompokkan ke dalam jumlah *cluster* yang telah ditentukan. Hasil yang didapatkan dari penerapan algoritma *K – Means* dengan RapidMiner menunjukkan bahwa ada 29.siswa dengan kategori nilai kurang berprestasi yang berada di *Cluster_0* (C0), 11 siswa dengan kategori nilai rata – rata

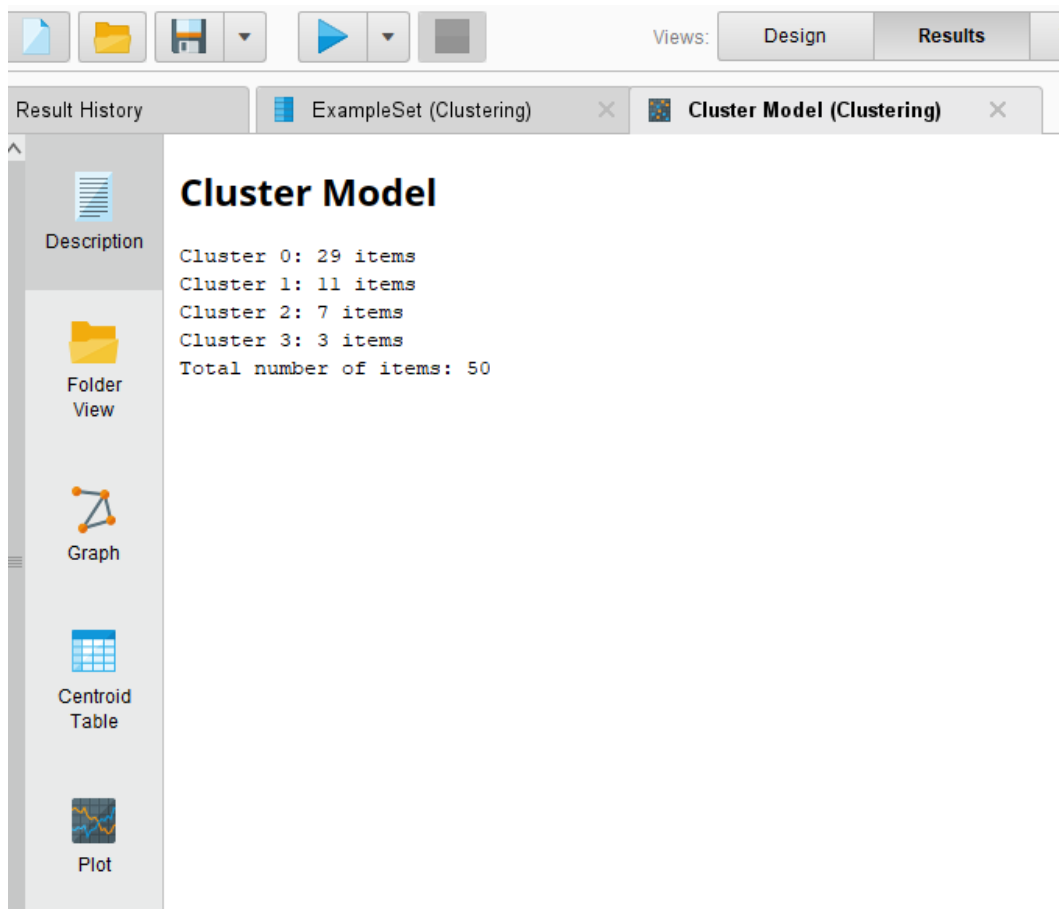
yang berada di *Cluster_1* (C2), dan 10 siswa dengan kategori nilai sangat berprestasi yang berada di *Cluster_2* (C2).

4.1.2 Perbandingan Hasil Clustering pada K=4 dan K=5

Selain pengujian pada jumlah cluster sebanyak 3, penelitian ini juga melakukan percobaan clustering dengan jumlah cluster sebanyak 4 dan 5 sebagai bahan perbandingan. Pengujian ini dilakukan untuk melihat pengaruh penambahan jumlah cluster terhadap hasil pengelompokan data prestasi akademik siswa. Dengan adanya perbandingan ini, maka keputusan bahwa jumlah cluster optimal adalah 3 tidak hanya didasarkan pada hasil Elbow Method dan Silhouette Score, tetapi juga didukung oleh hasil pengelompokan pada jumlah cluster lain

4.1.3 Hasil Clustering Menggunakan 4 Cluster

Pada percobaan clustering dengan jumlah cluster sebanyak 4, data siswa berhasil dikelompokkan ke dalam empat kelompok sesuai karakteristik nilai akademiknya. Hasil ini menunjukkan bahwa semakin banyak jumlah cluster yang digunakan, maka pembagian data menjadi semakin rinci. Namun, pembagian yang terlalu rinci belum tentu memberikan hasil pengelompokan yang lebih baik.



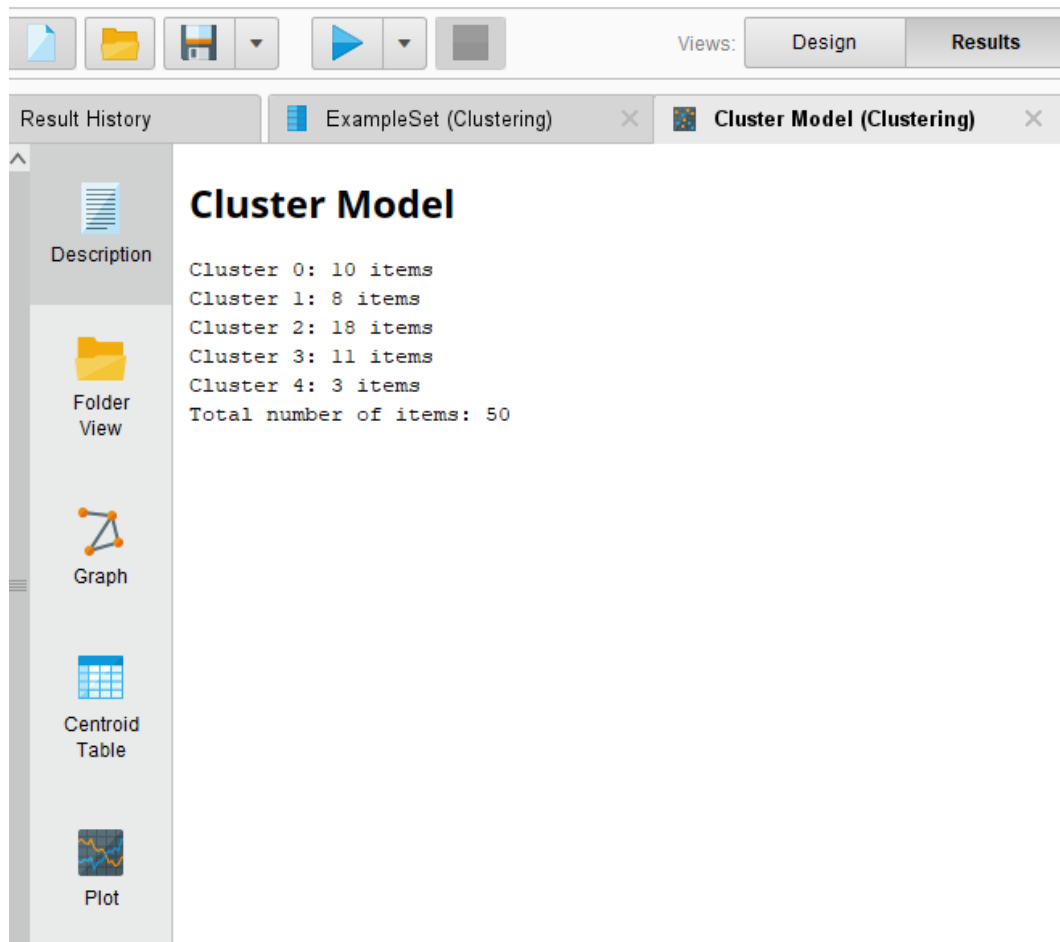
Gambar 4. 15 hasil pengolahan data jika cluster 4 menggunakan *RapidMiner*

Pada pengujian dengan jumlah cluster sebanyak 4 ($K=4$), data siswa terbagi ke dalam empat cluster, yaitu Cluster 0 sebanyak 29 data, Cluster 1 sebanyak 11 data, Cluster 2 sebanyak 7 data, dan Cluster 3 sebanyak 3 data. Hasil ini menunjukkan bahwa data dapat dipisahkan menjadi kelompok yang lebih rinci. Namun, jika dibandingkan dengan $K=3$, penggunaan $K=4$ menghasilkan pembagian cluster yang kurang seimbang karena terdapat cluster dengan jumlah anggota yang sangat sedikit, yaitu hanya 3 data pada Cluster 3. Kondisi ini menunjukkan bahwa penambahan jumlah cluster menjadi 4 cenderung membuat data terpecah terlalu detail. Oleh karena itu, $K=3$ dinilai lebih baik karena mampu

menghasilkan pengelompokan yang lebih optimal, lebih seimbang, dan lebih mudah diinterpretasikan.

4.1.4 Hasil Clustering Menggunakan 5 Cluster

Pada percobaan clustering dengan jumlah cluster sebanyak 5, data siswa terbagi menjadi lima kelompok. Hasil ini memperlihatkan bahwa data menjadi semakin terpecah ke dalam kelompok-kelompok yang lebih kecil. Secara teknis, pengelompokan ini mampu membagi data secara lebih rinci, tetapi dari sisi interpretasi hasil, pembagian tersebut berpotensi menimbulkan over-clustering.



Gambar 4. 16 hasil pengolahan data jika cluster 5 menggunakan RapidMiner

Berdasarkan hasil percobaan clustering pada $K=3$, $K=4$, dan $K=5$, dapat diketahui bahwa semakin besar nilai K , maka data akan terbagi menjadi kelompok yang semakin rinci. Namun, hasil yang lebih rinci tidak selalu menunjukkan kualitas clustering yang lebih baik. Pada $K=4$, distribusi data terbagi menjadi 29, 11, 7, dan 3 data, sedangkan pada $K=5$ distribusi data terbagi menjadi 10, 8, 18, 11, dan 3 data. Pada kedua percobaan tersebut masih terdapat cluster dengan jumlah anggota yang sangat sedikit, yaitu 3 data. Hal ini menunjukkan bahwa penambahan jumlah cluster cenderung menyebabkan data terpecah terlalu detail dan menghasilkan cluster kecil yang kurang representatif. Oleh karena itu, $K=3$ tetap dinilai sebagai jumlah cluster yang paling baik karena mampu memberikan hasil pengelompokan yang lebih sederhana, lebih seimbang, dan lebih mudah diinterpretasikan.

4.1.5 Analisis Perbandingan $K=3$, $K=4$, dan $K=5$

Berdasarkan hasil percobaan clustering pada $K=3$, $K=4$, dan $K=5$, dapat diketahui bahwa setiap penambahan jumlah cluster akan membuat pembagian data menjadi lebih rinci. Namun, hasil yang lebih rinci tidak selalu menunjukkan kualitas clustering yang lebih baik. Pada penelitian ini, hasil evaluasi Elbow Method menunjukkan bahwa titik siku terbentuk pada $K=3$, yang menandakan bahwa penambahan cluster setelah $K=3$ tidak lagi memberikan peningkatan yang signifikan

Selain itu, hasil evaluasi Silhouette Score pada $K=3$ menunjukkan nilai rata-rata sebesar 0,415, yang menandakan bahwa kualitas pengelompokan sudah cukup baik dan mendukung pemilihan 3 cluster sebagai jumlah cluster optimal

Jika dibandingkan dengan $K=4$ dan $K=5$, maka $K=3$ memiliki kelebihan dalam hal:

1. mampu membentuk kelompok yang lebih sederhana dan mudah diinterpretasikan;
2. tetap dapat merepresentasikan variasi prestasi akademik siswa;
3. tidak menyebabkan data terpecah ke dalam terlalu banyak kelompok kecil.

Dengan demikian, meskipun pengujian pada $K=4$ dan $K=5$ menunjukkan bahwa data tetap dapat dikelompokkan, hasil tersebut belum mampu memberikan kualitas pengelompokan yang lebih baik dibandingkan $K=3$. Oleh karena itu, jumlah cluster yang paling optimal dan paling sesuai untuk penelitian ini tetap adalah 3 cluster.

Tabel 4. 2 hasil clustering pada $K=3$, $K=4$, dan $K=5$

Nilai K	Jumlah Cluster	Keterangan Hasil
3	3 cluster	Paling optimal, mudah diinterpretasikan
4	4 cluster	Lebih rinci, tetapi beberapa cluster masih berdekatan
5	5 cluster	Terlalu rinci, berpotensi over-clustering

Berdasarkan perbandingan hasil clustering pada $K=3$, $K=4$, dan $K=5$, dapat disimpulkan bahwa $K=3$ merupakan jumlah cluster yang paling optimal. Hal ini karena $K=3$ mampu memberikan keseimbangan antara kualitas pemisahan data, kemudahan interpretasi, dan representasi variasi prestasi akademik siswa.

4.1.6. Evaluasi Menggunakan Elbow Method

Grafik Elbow Method menunjukkan hubungan antara jumlah cluster (K) dengan nilai Within Cluster Sum of Squares (WCSS) atau inertia. Nilai WCSS menggambarkan total variasi jarak data terhadap centroid dalam masing-masing cluster. Semakin kecil nilai WCSS, maka semakin kompak dan homogen cluster yang terbentuk.

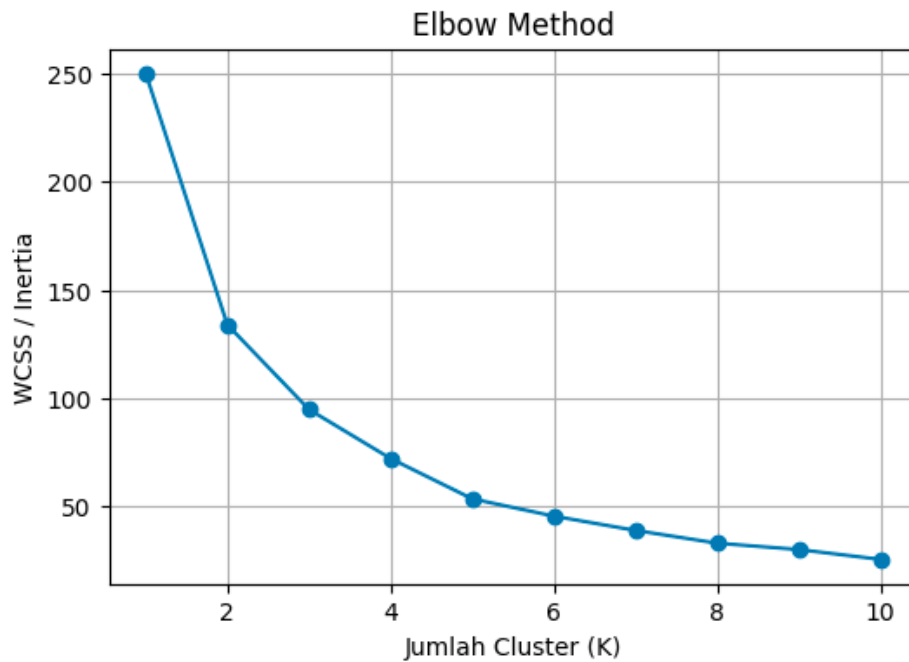
Berdasarkan grafik, terlihat bahwa:

1. Terjadi penurunan WCSS yang sangat signifikan dari $K=1$ ke $K=2$.
2. Penurunan masih cukup tajam hingga $K=3$.
3. Setelah $K=3$, penurunan nilai WCSS mulai melandai dan tidak terlalu signifikan.

Perubahan pola penurunan tersebut membentuk titik siku (elbow) pada $K=3$. Hal ini menunjukkan bahwa penambahan jumlah cluster setelah $K=3$ tidak lagi memberikan peningkatan kualitas pengelompokan yang signifikan. Dengan demikian, berdasarkan metode Elbow, jumlah cluster yang optimal dalam penelitian ini adalah 3 cluster.

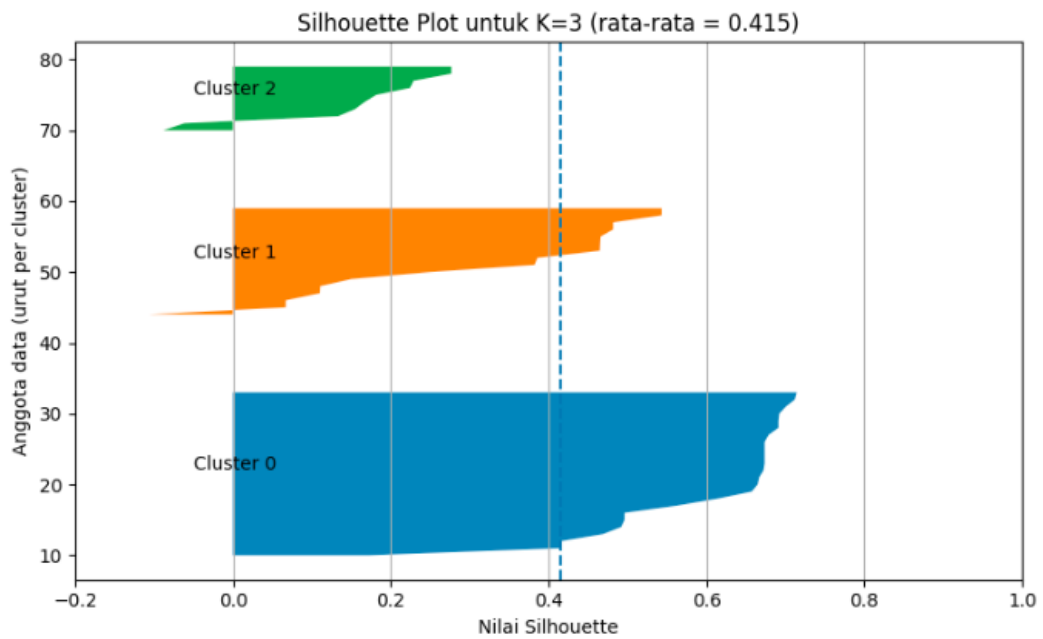
Pemilihan tiga cluster ini dinilai mampu merepresentasikan variasi prestasi akademik siswa secara efektif tanpa menyebabkan over-clustering (terlalu banyak

cluster) maupun under-clustering (terlalu sedikit cluster).



Gambar 4. 17 Gambar Hasil Evaluasi Elbow Method Menggunakan Python

4.1.7. Evaluasi Menggunakan Silhouette Score



Gambar 4. 18 Gambar hasil Evaluasi silhouette score menggunakan python

hasil evaluasi Silhouette untuk $K=3$ dengan nilai rata-rata sebesar 0,415. Garis putus-putus vertikal menunjukkan nilai Silhouette rata-rata keseluruhan. Sebagian besar anggota cluster memiliki nilai Silhouette positif, yang menunjukkan bahwa pengelompokan data sudah cukup baik dan tiap anggota cenderung lebih dekat dengan cluster-nya dibandingkan cluster lain.

Cluster 0 memiliki distribusi nilai Silhouette yang relatif tinggi dan stabil, menunjukkan kualitas pemisahan yang baik. Cluster 1 juga menunjukkan nilai positif dengan variasi yang moderat. Sementara itu, Cluster 2 memiliki beberapa nilai yang mendekati nol, menandakan adanya beberapa data yang berada di batas antar cluster. Secara keseluruhan, dengan nilai rata-rata Silhouette sebesar 0,415, kualitas pengelompokan termasuk dalam kategori cukup baik dan mendukung pemilihan jumlah cluster sebanyak 3.

4.2 Pembahasan

4.2.1 Analisis Hasil Clustering

Berdasarkan hasil pengelompokan menggunakan algoritma K-Means, siswa terbagi menjadi tiga kelompok berdasarkan tingkat prestasi belajar.

1. Cluster 3 (Sangat Berprestasi)

Siswa dalam cluster ini memiliki nilai tinggi pada hampir semua mata pelajaran. Hal ini menunjukkan bahwa siswa memiliki kemampuan akademik yang konsisten dan unggul.

2. Cluster 1 (Rata-rata)

Kelompok ini menunjukkan siswa dengan kemampuan sedang. Siswa

dalam cluster ini masih memiliki potensi untuk ditingkatkan melalui strategi pembelajaran yang tepat.

3. Cluster 2 (Kurang Berprestasi)

Siswa dalam cluster ini memiliki nilai yang relatif rendah. Kelompok ini memerlukan perhatian khusus seperti pembelajaran remedial.

Hasil ini menunjukkan bahwa kemampuan siswa tidak homogen dan perlu pendekatan pembelajaran yang berbeda.

4.2.2 Analisis Penentuan Jumlah Cluster

Berdasarkan hasil Elbow Method, titik optimal berada pada $K=3$ karena setelah titik tersebut penurunan WCSS tidak signifikan.

Hal ini menunjukkan bahwa:

1. Model dengan 3 cluster sudah cukup merepresentasikan data
2. Penambahan cluster hanya menambah kompleksitas tanpa meningkatkan kualitas

4.2.3 Analisis Kualitas Clustering

Nilai Silhouette Score menunjukkan bahwa:

1. Data dalam satu cluster cukup mirip (homogen)
2. Antar cluster memiliki perbedaan yang jelas

Hal ini menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data dengan baik.

4.2.4 Kaitan dengan Teori

Hasil penelitian ini sesuai dengan teori K-Means Clustering yang menyatakan bahwa data dengan karakteristik serupa akan dikelompokkan dalam satu cluster berdasarkan jarak terdekat.

Selain itu:

1. Elbow Method digunakan untuk menentukan jumlah cluster optimal
2. Silhouette Score digunakan untuk mengevaluasi kualitas cluster

Pendekatan ini sejalan dengan konsep *Educational Data Mining (EDM)* dalam analisis data pendidikan.

4.2.5 Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini sejalan dengan penelitian sebelumnya:

1. Hendrastuty (2024): K-Means efektif dalam clustering siswa
2. Flomina (2025): clustering membantu pemetaan prestasi

Perbedaan penelitian ini adalah:

1. Menggunakan kombinasi Elbow dan Silhouette
2. Fokus pada pendidikan dasar

4.2.6 Implikasi Penelitian

Hasil clustering memberikan manfaat bagi sekolah, yaitu:

1. **Cluster tinggi**
→ Diberikan pengayaan
2. **Cluster sedang**
→ Pembelajaran penguatan

3. Cluster rendah

→ Program remedial

Dengan demikian, pembelajaran dapat dilakukan secara lebih efektif dan berbasis data.