

BAB IV

HASIL DAN PEMBAHASAN

4.1. Analisis Data

Analisis data dalam penelitian ini dilakukan untuk mengetahui faktor-faktor yang memengaruhi prestasi akademik siswa dengan menggunakan pendekatan data mining melalui algoritma K-Means Clustering dan Naive Bayes. Data yang digunakan merupakan data siswa kelas X SMA Swasta Bhayangkari 2 Rantau Prapat dengan variabel nilai bahasa inggris (X1), kehadiran (X2), kursus (X3), minat (X4) dan prestasi (Y).

Sebelum dilakukan proses pengolahan, data terlebih dahulu melalui tahap preprocessing. Tahapan ini meliputi pemeriksaan kelengkapan data (data cleaning), penghapusan data yang tidak konsisten atau duplikat, serta transformasi data ke dalam bentuk yang sesuai agar dapat diproses oleh algoritma data mining. Selain itu, data juga disiapkan untuk proses analisis lanjutan.

Selanjutnya, dilakukan proses clustering menggunakan algoritma K-Means untuk mengelompokkan data siswa ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristik yang dimiliki. Hasil dari proses clustering ini akan menunjukkan pengelompokan siswa berdasarkan kategori tertentu, seperti tingkat prestasi rendah, sedang, dan tinggi.

Setelah proses clustering selesai, tahap berikutnya adalah klasifikasi menggunakan algoritma Naive Bayes. Pada tahap ini, model dibangun dengan menghitung probabilitas masing-masing variabel terhadap kelas prestasi siswa.

Model tersebut kemudian digunakan untuk mengidentifikasi faktor yang paling berpengaruh terhadap prestasi akademik serta melakukan prediksi terhadap data siswa.

4.2. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan mengambil data akademik siswa kelas X di SMA Swasta Bhayangkari 2 Rantau Prapat yang berkaitan dengan pembelajaran Bahasa Inggris. Data yang dikumpulkan meliputi nilai Bahasa Inggris, tingkat kehadiran siswa, kursus Bahasa Inggris, serta minat belajar siswa sebagai variabel yang memengaruhi prestasi akademik.

Data diperoleh dari dua sumber, yaitu data sekunder berupa dokumentasi sekolah seperti nilai rapor dan absensi siswa, serta data primer yang diperoleh melalui penyebaran kuesioner untuk mengukur minat belajar dan keikutsertaan kursus. Seluruh data yang digunakan merupakan data yang lengkap dan relevan dengan variabel penelitian, sehingga dapat mendukung proses analisis menggunakan algoritma K-Means Clustering dan Naive Bayes.

Tabel 4.1 Data siswa Semester Ganjil Siswa Kelas X

No	Nama Siswa	Nilai B. Inggris	Kehadiran%	Kursus	Minat
1	AD* AM**** PU***	78	75%	Ya	Sedang
2	AH*** FA***	82	83%	Ya	Sedang
3	AI*** SA****	77	75%	Ya	Sedang
4	AM**** AP*****	90	95%	Ya	Tinggi
5	AN**** DW*	92	95%	Ya	Tinggi
6	AN**** AL*****	88	95%	Ya	Tinggi
7	AL** SE*****	91	95%	Ya	Tinggi
8	AN*** FE*****	89	95%	Ya	Tinggi
9	AR*** DI***	78	80%	Tidak	Rendah
10	AU*** KA***	66	60%	Tidak	Sedang
11	AZ*** ER*****	79	75%	Ya	Sedang
12	AS**** NA***	93	95%	Ya	Tinggi

13	AU** SY*** SA*****	90	95%	Ya	Tinggi
14	AD* AY* LE*****	91	95%	Ya	Tinggi
15	AN***** AT****	87	91%	Ya	Tinggi
16	AN*** SI*****	80	83%	Ya	Sedang
17	AZ*** SY*****	88	91%	Ya	Tinggi
18	BA*** MO****	92	95%	Ya	Tinggi
19	B**** MO****	90	95%	Ya	Tinggi
20	BU*** WU*****	89	95%	Ya	Tinggi
21	BE*** AU***	91	95%	Ya	Tinggi
22	CI*** HI*****	93	95%	Ya	Tinggi
23	CL*** OK*****	88	95%	Ya	Tinggi
24	DI*** ZA*	90	95%	Ya	Tinggi
25	DE*** CI****	87	91%	Ya	Tinggi
26	DZ**** AR**	89	95%	Ya	Tinggi
27	ER** AU*****	91	95%	Ya	Tinggi
28	FA**** HA****	92	95%	Ya	Tinggi
29	FA*** GI*****	90	95%	Ya	Tinggi
30	FI*****	57	58%	Tidak	Rendah
31	FA*** AR*****	45	33%	Tidak	Rendah
32	GL** GA*****	76	75%	Ya	Sedang
33	GI**** PR*****	92	95%	Ya	Tinggi
34	HA*** FA****	88	95%	Ya	Tinggi
35	IR** SU*****	90	95%	Ya	Tinggi
36	JI*** FA****	86	91%	Ya	Tinggi
37	KE*** RI***	91	95%	Ya	Tinggi
38	LA*** DI**	89	95%	Ya	Tinggi
39	LI*** JA*****	90	95%	Ya	Tinggi
40	LU*** AD****	92	95%	Ya	Tinggi
41	MA**** PR*****	93	95%	Ya	Tinggi
42	MU***** HA*****	78	75%	Ya	Sedang
43	MU* DI**	91	95%	Ya	Tinggi
44	MU***** RE****	81	83%	Ya	Sedang
45	MEI*****	80	83%	Ya	Sedang
46	ME*****	70	66%	Tidak	Rendah
47	MU***** AZ****	50	41%	Tidak	Rendah
48	MU***** AL**	69	66%	Tidak	Rendah
49	MU**** SI*****	87	91%	Ya	Tinggi
50	NA***	92	95%	Ya	Tinggi
51	NA**** NA*****	91	95%	Ya	Tinggi
52	NA***** RA**	77	75%	Ya	Sedang
53	NA*** AU***	90	95%	Ya	Tinggi
54	NO*** ZA***	89	95%	Ya	Tinggi
55	NE*** AU*****	60	50%	Tidak	Rendah
56	NU**** SH****	91	95%	Ya	Tinggi
57	NU* ER*****	78	75%	Ya	Sedang
58	NA*** PU***	92	95%	Ya	Tinggi
59	NE*** YU***	81	83%	Ya	Sedang

60	NA*** SI*****	80	83%	Ya	Sedang
61	PU*** AM****	93	95%	Ya	Tinggi
62	RA*** NA*****	87	91%	Ya	Tinggi
63	RI* AN****	82	83%	Ya	Sedang
64	RI** RI*****	91	95%	Ya	Tinggi
65	RI*** AY****	80	83%	Ya	Sedang
66	RA*****	42	25%	Tidak	Rendah
67	RE*** WI****	90	95%	Ya	Tinggi
68	RI*** RA*****	91	95%	Ya	Tinggi
69	RA*** RA***	82	83%	Ya	Sedang
70	RE** KI****	56	58%	Tidak	Rendah
71	SU** RA*****	92	95%	Ya	Tinggi
72	SU***** NO****	90	95%	Ya	Tinggi
73	SA***** AN****	77	75%	Ya	Sedang
74	TI** RA*****	91	95%	Ya	Tinggi
75	TI*** AF****	92	95%	Ya	Tinggi
76	TO** RA*****	90	95%	Ya	Tinggi
77	TE*** AU***	89	95%	Ya	Tinggi
78	TR*** AY*	57	58%	Tidak	Rendah
79	UM** HA*****	81	83%	Ya	Sedang
80	WI*** AN****	70	66%	Tidak	Sedang
81	WI** RE*****	92	95%	Ya	Tinggi
82	YU**** AN****	91	95%	Ya	Tinggi
83	DA**** AD****	56	58%	Tidak	Rendah
84	IN*** RA*****	78	75%	Ya	Sedang
85	PA***	44	25%	Tidak	Rendah
86	TH***** KA*****	68	66%	Tidak	Sedang

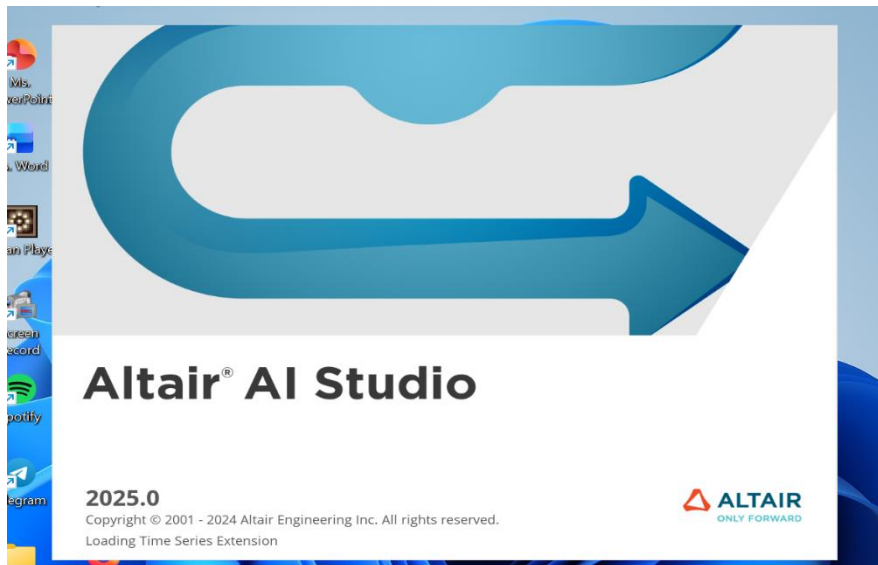
Tabel ini berisikan 86 data siswa yang terdiri dari variabel Nilai Bahasa Inggris, Kehadiran, Kursus Bahasa Inggris, Minat Belajar, dan Prestasi. Variabel kehadiran yang semula berbentuk persentase telah ditransformasikan ke dalam tiga kategori, yaitu 0-70% rendah (1), 71-85% sedang (2), dan 85-100% tinggi (3). Variabel kursus bahasa inggris yang semula berbentuk ya dan tidak telah ditransformasikan kedalam 2 kategori yaitu ya (1) dan tidak (2). Kemudian variabel minat yang semula berbentuk tinggi, sedang, rendah telah ditransformasikan ke dalam tiga kategori, yaitu tinggi (1), sedang (2), rendah (3).

Pada variabel kursus Bahasa Inggris, terdapat siswa yang mengikuti dan tidak mengikuti kursus, dengan jumlah yang relatif seimbang, meskipun sebagian siswa lebih banyak mengikuti kursus. Sementara itu, variabel minat belajar menunjukkan variasi yang cukup beragam, dimana siswa terbagi ke dalam kategori tinggi, sedang, dan rendah.

4.3. Penerapan Algoritma K-Means Dengan RapidMiner

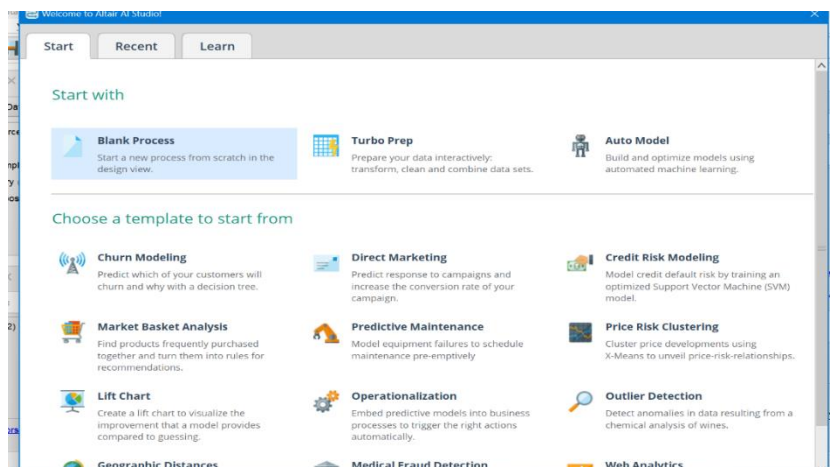
Pada penelitian ini, algoritma K-Means digunakan dengan bantuan aplikasi RapidMiner untuk mengidentifikasi dan mengelompokkan prestasi akademik siswa di SMA Swasta Bhayangkari 2 Rantau Prapat. Pemanfaatan RapidMiner dilakukan untuk mempermudah pengolahan data serta mempercepat proses clustering secara terstruktur. Dengan demikian, hasil pengelompokan prestasi akademik siswa dapat diperoleh secara lebih tepat, efektif, dan efisien, sehingga mampu memberikan gambaran pola serta faktor-faktor yang memengaruhi prestasi siswa.

Selain itu, penerapan algoritma K-Means dalam penelitian ini memungkinkan dilakukannya pengelompokan siswa ke dalam beberapa klaster berdasarkan tingkat kemiripan karakteristik data yang dimiliki, seperti nilai akademik, tingkat kehadiran, Kursus bahasa Inggris dan minat belajar. Hasil pengelompokan tersebut selanjutnya dapat dianalisis untuk mengidentifikasi karakteristik masing-masing klaster secara lebih mendalam.

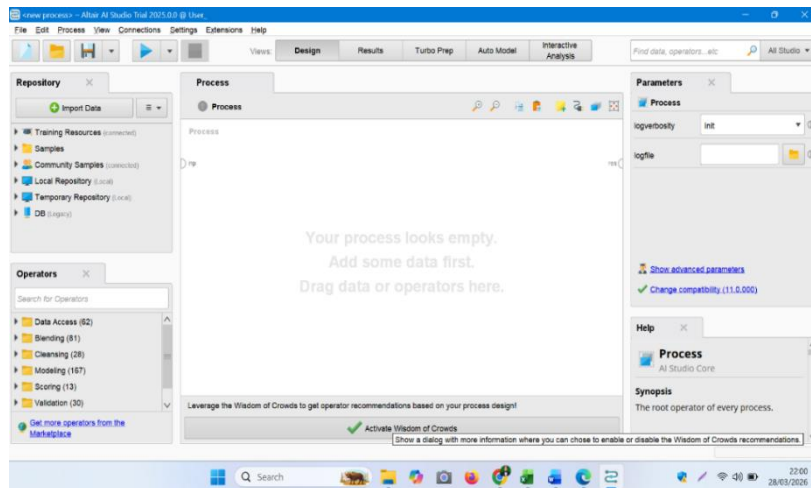


Gambar 4.1. RapidMiner

Langkah pertama dilakukan dalam proses menerapkan algoritma K-Means dengan RapidMiner untuk mengetahui prestasi akademik siswa di SMA Swasta Bhayangkari 2 Rantau Prapat yaitu membuka RapidMiner. Setelah itu muncul tampilan awal dari RapidMiner yaitu tampilan untuk memulai lembar kerja baru. Setelah itu klik “*Blank Process*” untuk memulai lembar kerja baru.

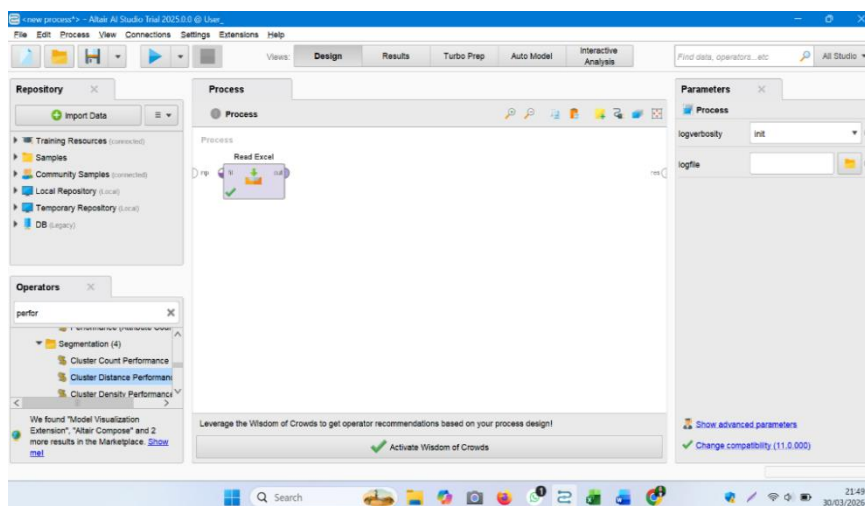


Gambar 4.2. Tampilan Untuk Memulai Lembar Kerja Baru Pada RapidMiner



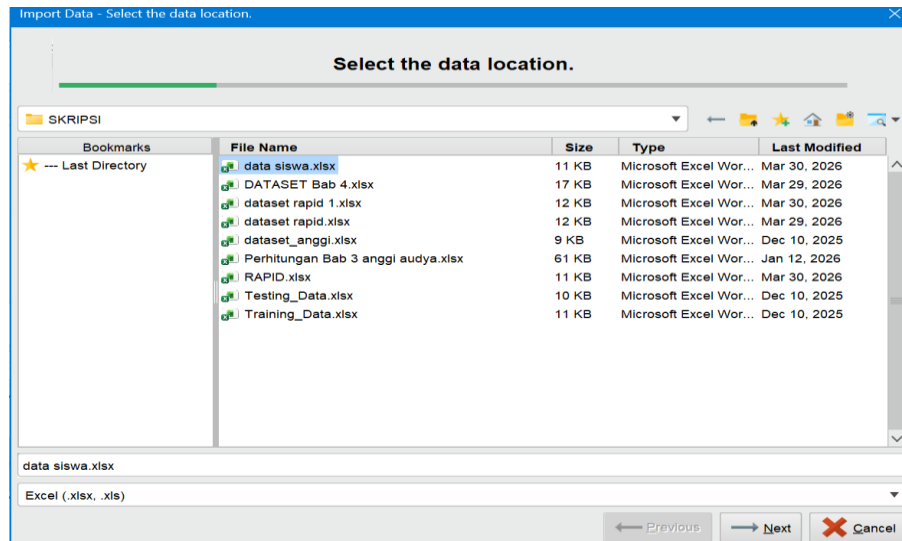
Gambar 4.3. Tampilan Kosong Halaman Lembar Kerja Baru Pada RapidMiner

Proses yang dilakukan selanjutnya adalah menginput data siswa kedalam lembar kerja pada RapidMiner. Maka dibutuhkan operator “*Read Excel*” yang didapatkan dari pencarian di bagian “Operators”. Kemudian drag and drop ke dalam proses.



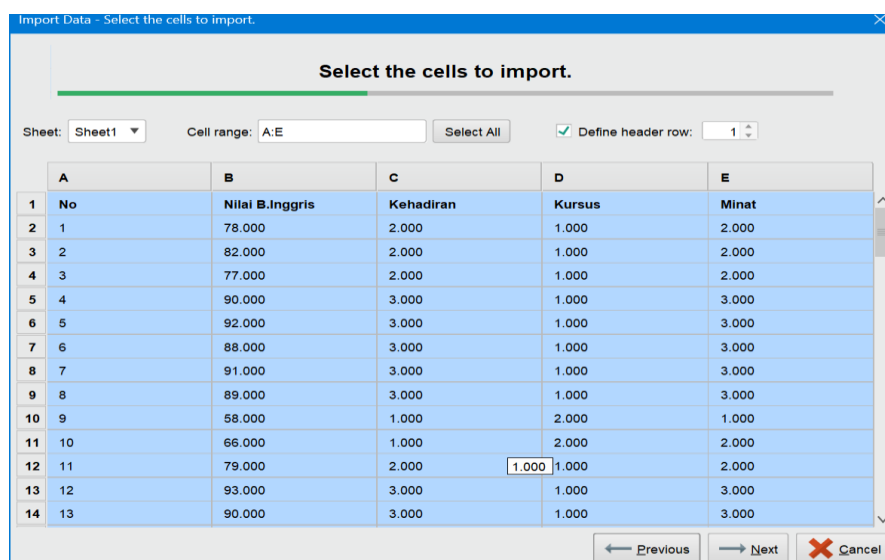
Gambar 4.4. Proses Menginput Operator K – Means Ke Dalam Lembar

Kerja Selanjutnya klik “*Import Configuration Wizard*” pada “*Parameters Read Excel*”. Setelah itu akan muncul tampilan “*Select the data location*” lalu pilih lokasi penyimpanan data nilai akademik siswa. Kemudian klik “Next”.



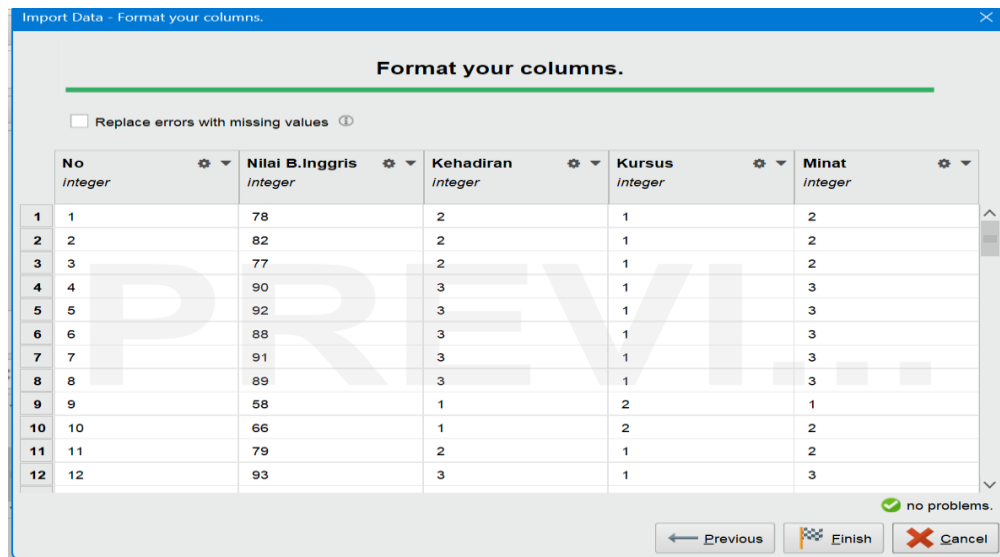
Gambar 4.5. Proses Memilih Lokasi Penyimpanan Data

Setelah mengklik “Next” maka muncul tampilan “*Select the cells to import*” lalu blok semua data nilai siswa. Lalu klik “Next”.



Gambar 4.6. Proses Memilih Data Yang Akan Di Input

Selanjutnya menentukan label atau class dari dataset yang dipilih dengan memilih “*Change role*” dengan label. Lalu klik “*Finish*”.



Gambar 4.7. Proses Memilih Label Dari Dataset

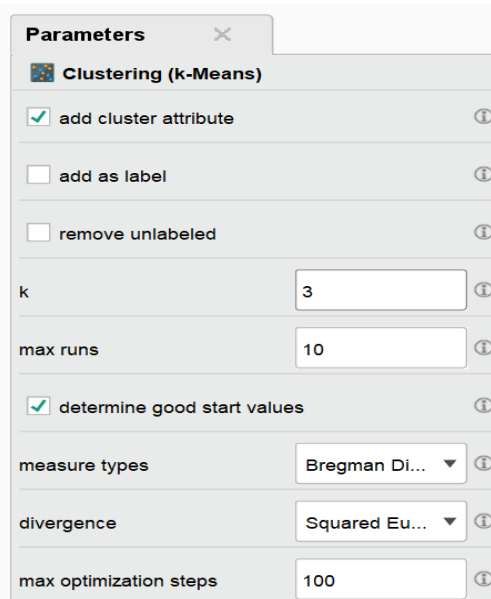
Setelah tahapan import data siswa selesai, maka selanjutnya yang dilakukan adalah menambahkan operator “*K-Means*” yang didapatkan dari pencarian di bagian “*Operators*”. Kemudian drag and drop ke dalam proses.



Gambar 4.8. Proses Menginput Operator K – Means Ke Dalam Lembar

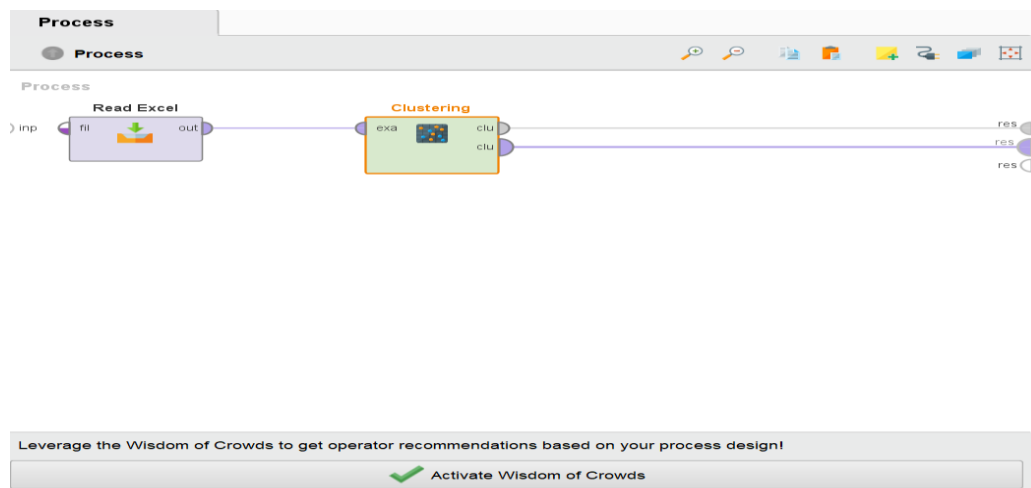
Kerja

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



Gambar 4.9. Proses Menentukan Jumlah Cluster

Selanjutnya sambungkan operator satu sama lain dengan menarik garis dari *output Read Excel* ke input Clustering (*K-Means*), lalu tarik garis dari Clustering (*K-Means*) ke res. Setelah semuanya selesai, klik tombol “Run” untuk menjalankan proses.



Gambar 4.10. 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.	id	cluster	Nilai B.Inggr...	Kehadiran%	Kursus	Minat
1	1	cluster_2	78	2	1	2
2	2	cluster_2	82	2	1	2
3	3	cluster_2	77	2	1	2
4	4	cluster_0	90	3	1	3
5	5	cluster_0	92	3	1	3
6	6	cluster_0	88	3	1	3
7	7	cluster_0	91	3	1	3
8	8	cluster_0	89	3	1	3
9	9	cluster_1	58	1	2	1
10	10	cluster_2	66	1	2	2
11	11	cluster_2	79	2	1	2
12	12	cluster_0	93	3	1	3
13	13	cluster_0	90	3	1	3

ExampleSet (86 examples, 2 special attributes, 4 regular attributes)

Gambar 4.11. 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.

Cluster Model

Description

Cluster 0: 52 items
 Cluster 1: 10 items
 Cluster 2: 24 items
 Total number of items: 86

Folder View

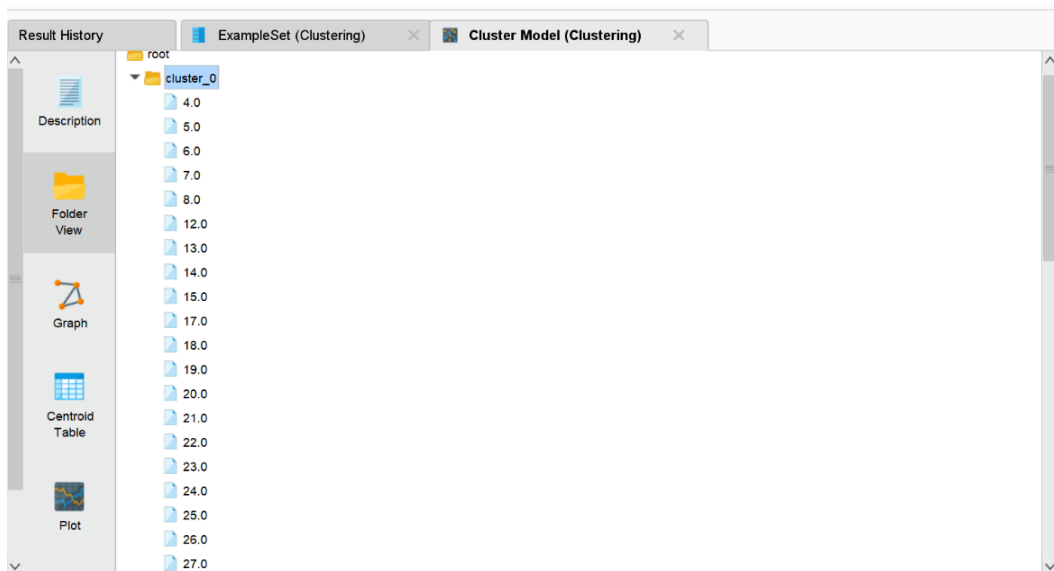
Graph

Centroid Table

Plot

Gambar 4.12. 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*

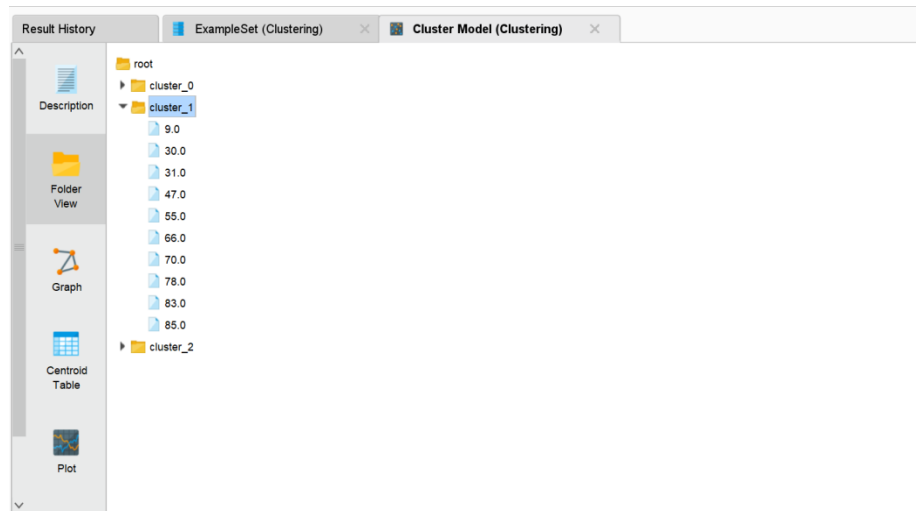


Gambar 4.13. Hasil Pengelompokan Cluster_0

Berdasarkan hasil clustering pada tampilan *Cluster Model*, terlihat bahwa seluruh data berasal dari satu root sebagai induk utama, yang kemudian terbagi ke dalam beberapa kelompok, salah satunya adalah *cluster_0*. Cluster ini berisi sejumlah data yang memiliki kemiripan karakteristik sehingga dikelompokkan bersama oleh algoritma clustering. Adapun data yang termasuk ke dalam *cluster_0* adalah data dengan nomor 4, 5, 6, 7, 8, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 49, 50, 51, 53, 54, 56, 58, 61, 62, 64, 67, 68, 71, 72, 74, 75, 76, 77, 81, dan 82. Hal ini menunjukkan bahwa data-data tersebut memiliki pola atau atribut yang serupa dibandingkan dengan data lain yang kemungkinan masuk ke cluster berbeda. Dengan demikian,

cluster_0 merepresentasikan satu kelompok data hasil segmentasi dari keseluruhan data pada root berdasarkan tingkat kemiripannya.

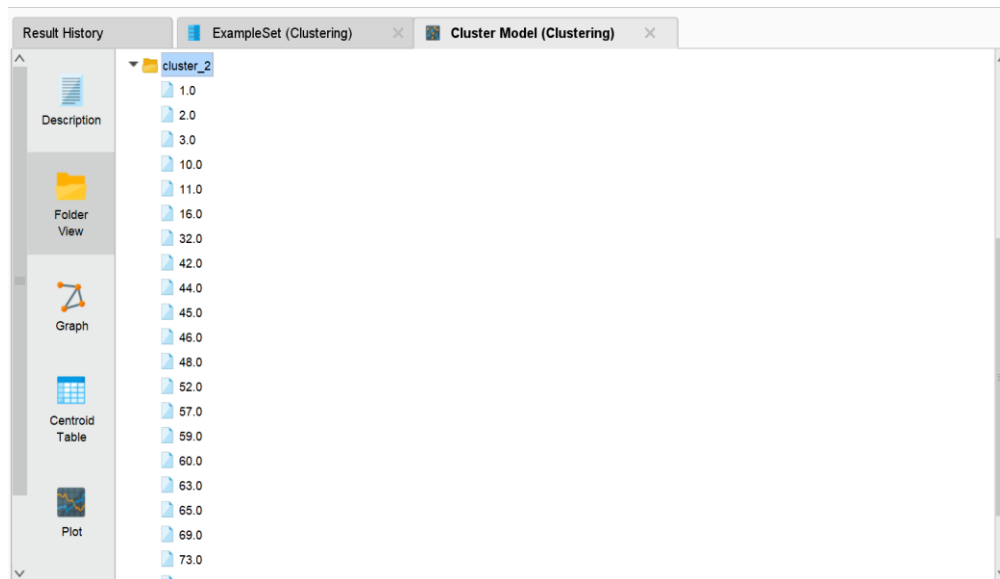
Berikut ini gambaran dari *Folder View* untuk *Cluster 1*.



Gambar 4.14. Hasil Pengelompokan Cluster_1

Berdasarkan hasil clustering pada tampilan *Cluster Model*, seluruh data awal berada pada root sebagai induk utama, kemudian terbagi ke dalam beberapa kelompok, salah satunya adalah cluster_1. Cluster ini berisi sekumpulan data yang memiliki kemiripan karakteristik sehingga dikelompokkan bersama oleh algoritma clustering. Adapun data yang termasuk ke dalam cluster_1 adalah data dengan nomor 9, 30, 31, 47, 55, 66, 70, 78, 83, dan 85. Hal ini menunjukkan bahwa data-data tersebut memiliki pola atau atribut yang serupa satu sama lain, namun berbeda dengan data pada cluster lainnya seperti cluster_0 atau cluster_2. Dengan demikian, cluster_1 merepresentasikan salah satu segmen data hasil pengelompokan dari keseluruhan data pada root berdasarkan tingkat kemiripan karakteristik yang dimiliki.

Berikut ini gambaran dari Folder View untuk Cluster 2.



Gambar 4.15 Hasil Pengelompokan Cluster_2

Berdasarkan hasil clustering pada tampilan *Cluster Model*, seluruh data awal berada pada root sebagai induk utama yang kemudian terbagi menjadi beberapa kelompok, salah satunya adalah cluster_2. Cluster ini berisi sekumpulan data yang memiliki kemiripan karakteristik sehingga dikelompokkan bersama oleh algoritma clustering. Adapun data yang termasuk ke dalam cluster_2 adalah data dengan nomor 1, 2, 3, 10, 11, 16, 32, 42, 44, 45, 46, 48, 52, 57, 59, 60, 63, 65, 69, 73, 79, 80, 86, dan 84. Hal ini menunjukkan bahwa data-data tersebut memiliki pola atau atribut yang serupa satu sama lain, namun berbeda dengan data pada cluster lainnya seperti cluster_0 dan cluster_1. Dengan demikian, cluster_2 merepresentasikan salah satu segmen data hasil pengelompokan dari keseluruhan data pada root berdasarkan tingkat kemiripan karakteristik yang dimiliki.

4.4. Hasil Penerapan Algoritma K - Means Dengan RapidMiner

Hasil penerapan algoritma K-Means menggunakan RapidMiner menunjukkan bahwa data berhasil dikelompokkan ke dalam beberapa cluster sesuai dengan jumlah yang telah ditentukan sebelumnya. Berdasarkan hasil proses clustering, diperoleh bahwa pada Cluster_0 terdapat 52 siswa dengan kategori nilai prestasi akademik bahasa inggris tinggi, pada Cluster_1 terdapat 10 siswa dengan kategori nilai prestasi akademik bahasa inggris sedang, dan pada Cluster_2 terdapat 24 siswa dengan kategori nilai prestasi akademik bahasa inggris rendah. Hasil pengelompokan tersebut menunjukkan adanya perbedaan tingkat prestasi akademik siswa. Hasil dari pengelompokan atau hasil dari clustering dapat dilihat pada tabel – tabel dibawah ini:

Tabel 4.2 Hasil Cluster_0 (Tinggi)

No	Nama Siswa
1	AM**** AP*****
2	AN**** DW*
3	AN**** AL*****
4	AL** SE*****
5	AN*** FE*****
6	AS**** NA***
7	AU** SY*** SA*****
8	AD* AY* LE*****
9	AN***** AT****
10	AN*** SI*****
11	AZ*** SY*****
12	BA*** MO****
13	B**** MO****
14	BU*** WU*****
15	BE*** AU***
16	CI*** HI*****
17	CL*** OK*****
18	DI*** ZA*
19	DE*** CI*****
20	DZ**** AR**
21	ER** AU*****
22	FA**** HA****

23	FA*** GI*****
24	GI***** PR*****
25	HA*** FA****
26	IR** SU*****
27	JI*** FA****
28	KE*** RI***
29	LA*** DI**
30	LI*** JA*****
31	LU*** AD****
32	MA**** PR*****
33	MU* DI**
34	MU**** SI*****
35	NA***
36	NA**** NA*****
37	NU**** SH***
38	NA*** PU***
39	PU*** AM****
40	RA*** NA*****
41	RI** RI*****
42	RE*** WI****
43	RI*** RA*****
44	SU** RA*****
45	SU***** NO****
46	TI** RA*****
47	TI*** AF****
48	TO** RA*****
49	TE*** AU***
50	WI** RE*****
51	YU**** AN****
52	NO*** ZA***

Tabel 4.3 Hasil Cluster_1 (Sedang)

No	Nama Siswa
1	AR*** DI***
2	FI*****
3	FA*** AR*****
4	MU***** AZ****
5	NE*** AU*****
6	RA*****
7	RE** KI*****
8	TR*** AY*
9	DA**** AD****
10	PA***

Tabel 4.4 Hasil Cluster_2 (Rendah)

No	Nama Siswa
1	AD* AM**** PU***
2	AH*** FA***
3	AI*** SA****
4	AU*** KA***
5	AZ*** ER*****
6	AN*** SI*****
7	GL** GA*****
8	MU***** HA*****
9	MU***** RE*****
10	MEI*****
11	ME*****
12	MU***** AL**
13	NA***** RA**
14	NU* ER*****
15	NE*** YU***
16	NA*** SI*****
17	RI* AN*****
18	RI*** AY*****
19	RA*** RA***
20	SA***** AN*****
21	UM** HA*****
22	WI*** AN*****
23	IN*** RA*****
24	TH***** KA*****

Hasil dari pengelompokan ini Bagi siswa dengan kategori prestasi rendah, disarankan untuk diberikan bimbingan tambahan seperti pengulangan materi maupun program remedial. Upaya ini bertujuan agar siswa dapat memperdalam pemahaman terhadap materi pembelajaran, mengejar ketertinggalan, serta meningkatkan hasil belajar secara lebih optimal.

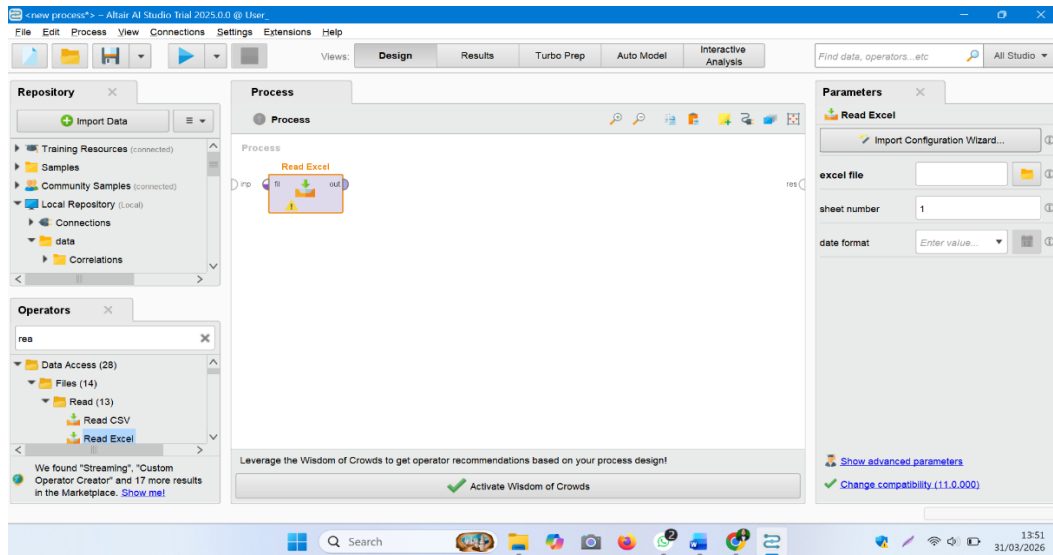
Selanjutnya, bagi siswa dengan kategori prestasi sedang, perlu adanya peningkatan motivasi belajar. Hal ini penting karena motivasi memiliki peran dalam mendorong intensitas dan kualitas usaha belajar siswa, sehingga dapat berdampak pada peningkatan prestasi akademik.

Sementara itu, bagi siswa dengan kategori prestasi tinggi, perlu diberikan tantangan pembelajaran yang lebih lanjut. Hal ini bertujuan untuk mendorong kemampuan berpikir kritis, meningkatkan kreativitas, serta menjaga konsistensi motivasi belajar sehingga potensi siswa dapat berkembang secara maksimal.

4.5. Penerapan Algoritma *Naïve Bayes* Dengan RapidMiner

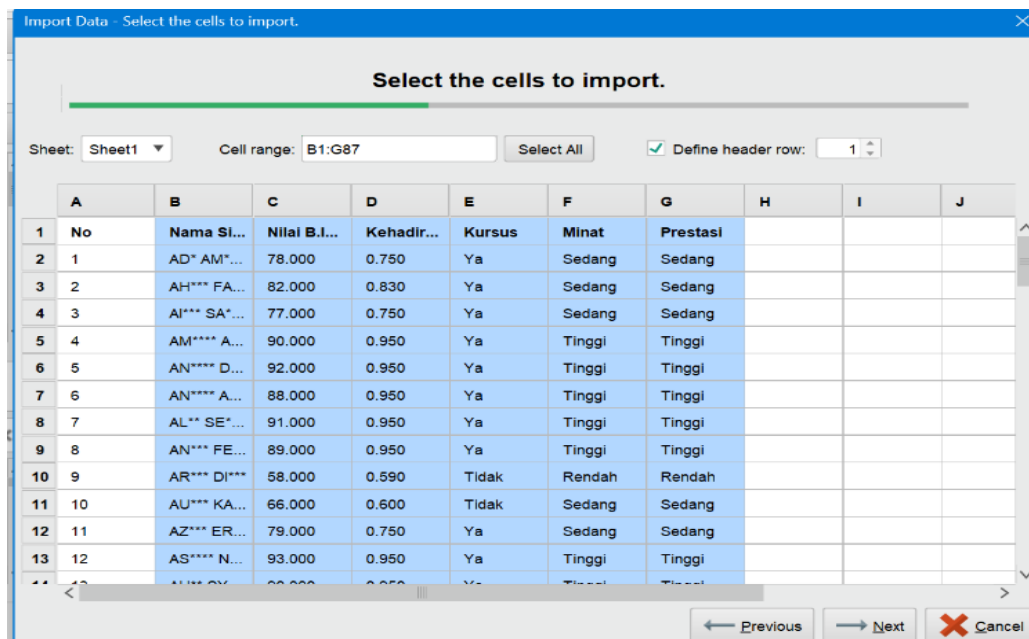
Metode *Naïve Bayes* merupakan salah satu teknik klasifikasi yang digunakan untuk menganalisis serta memprediksi faktor-faktor yang memengaruhi prestasi akademik siswa. Metode ini didasarkan pada konsep probabilitas yang merujuk pada Teorema Bayes, dengan asumsi bahwa setiap atribut bersifat independen satu sama lain. Dalam penelitian ini, atribut yang digunakan meliputi kehadiran, keikutsertaan dalam kursus, dan minat belajar, yang kemudian dianalisis untuk menentukan kemungkinan siswa berada pada kategori prestasi tertentu. Hasil klasifikasi ditentukan berdasarkan nilai probabilitas tertinggi, sehingga mampu menghasilkan prediksi yang objektif, sistematis, dan terukur terhadap prestasi akademik siswa.

Penggunaan RapidMiner bertujuan untuk memfasilitasi proses pengolahan data secara terstruktur, mulai dari tahap praproses data, pembangunan model, hingga evaluasi hasil klasifikasi



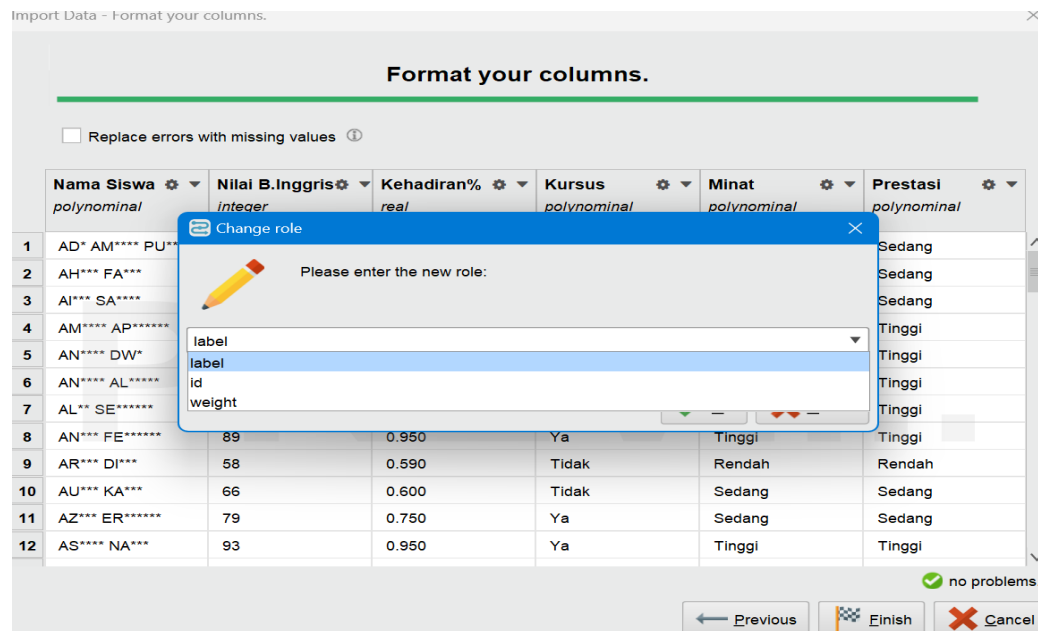
Gambar 4.16. Operator Excel

Pada tampilan utama, terlihat sebuah operator “*Read Excel*” yang telah ditambahkan ke dalam area *Process*. Operator ini berfungsi untuk mengimpor data dari file Excel ke dalam RapidMiner sebagai tahap awal pengolahan data.



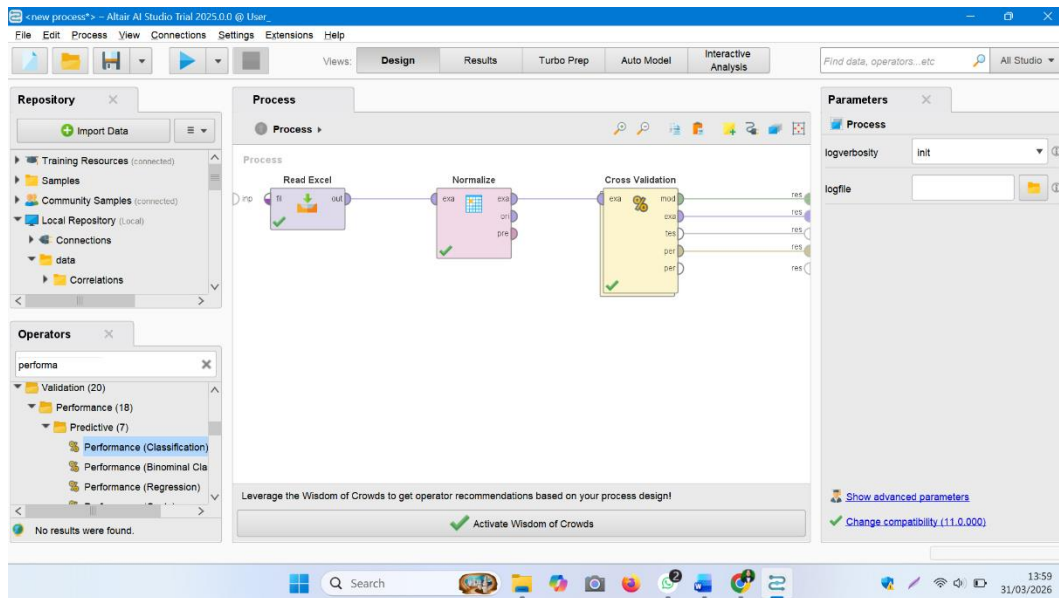
Gambar 4.17. Impor Data

Pada Gambar 4.17 menggambarkan proses seleksi data sebelum dilakukan pengolahan lebih lanjut, yang merupakan tahap penting dalam *data preprocessing* pada penelitian data mining.



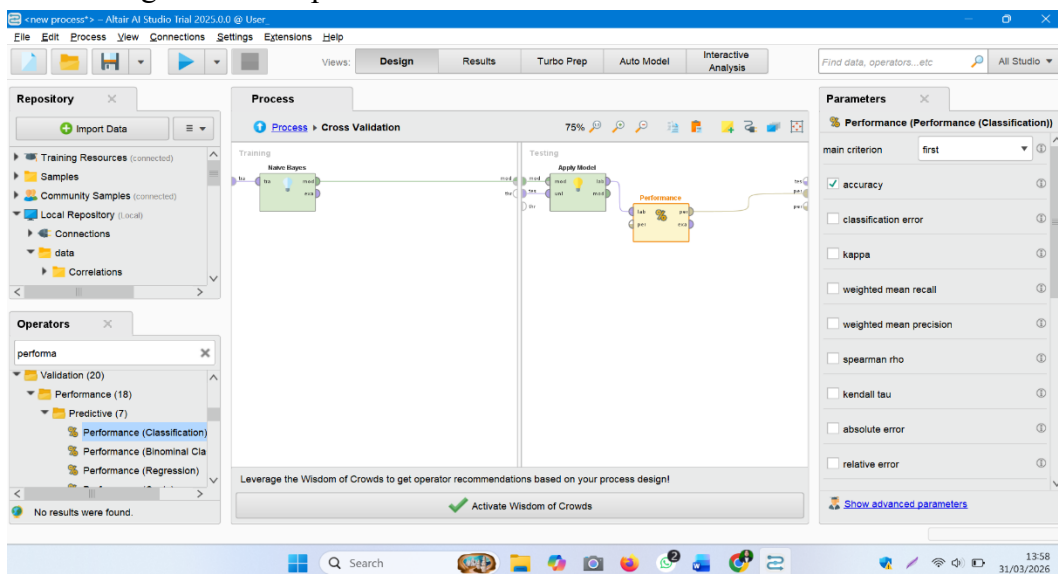
Gambar 4.18. Set Role Variabel Targe

Pada gambar 4.18 terdapat jendela pop-up “Change role” yang digunakan untuk menentukan peran dari suatu atribut. Dalam gambar tersebut, atribut sedang diatur menjadi “label”, yang berarti kolom tersebut dijadikan sebagai target atau variabel utama yang akan diprediksi dalam proses klasifikasi (misalnya pada metode Naive Bayes). Penentuan label ini sangat penting karena akan menjadi acuan dalam proses pembelajaran model. Dengan penetapan yang tepat, model dapat menghasilkan klasifikasi yang lebih akurat dan sesuai dengan tujuan penelitian.



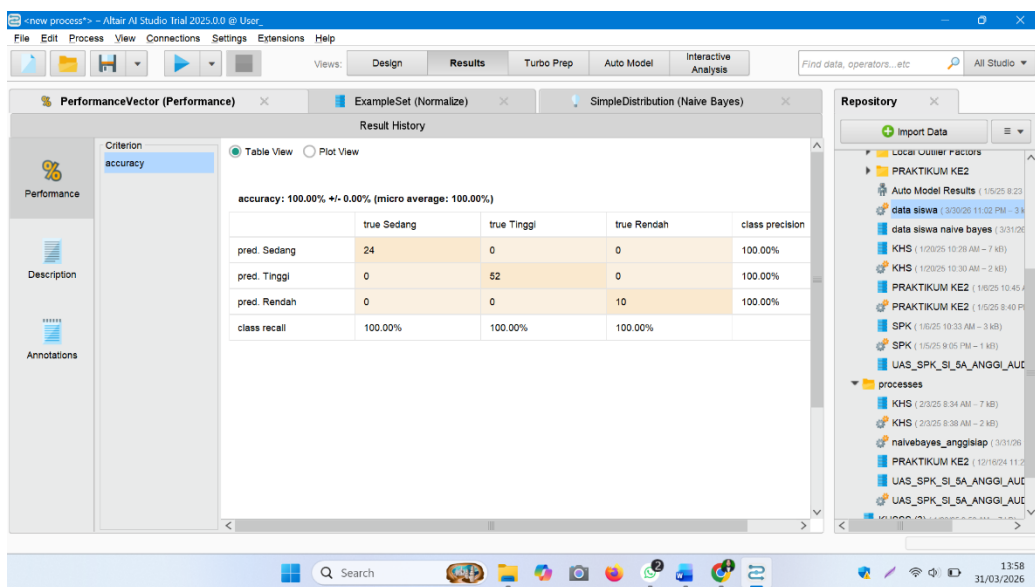
Gambar 4.19. Design Process

Gambar tersebut menampilkan tampilan *Design Process* pada aplikasi RapidMiner yang menunjukkan alur awal pengolahan data. Terlihat dua operator yang digunakan, yaitu Read Excel dan Normalize, cross validation yang saling terhubung dalam satu proses.



Gambar 4.20. Design Process

Tampilan *training* dan *testing* pada proses pengolahan data tersebut berada di dalam operator Cross Validation pada aplikasi RapidMiner. Dengan adanya pembagian ini, proses evaluasi model menjadi lebih objektif dan mampu mengukur tingkat akurasi serta performa metode yang digunakan secara lebih valid.



Gambar 4.21. Design Process

4.6. Hasil Penerapan Algoritma Naïve Bayes Dengan RapidMiner

Berdasarkan hasil pengujian menggunakan algoritma Naïve Bayes pada data yang ditampilkan, diperoleh nilai akurasi sebesar 100%, yang menunjukkan bahwa seluruh data berhasil diklasifikasikan dengan sangat baik tanpa kesalahan. Pada confusion matrix terlihat bahwa data dengan kelas Sedang sebanyak 24 siswa berhasil diprediksi dengan benar sebagai Sedang, data Tinggi sebanyak 52 siswa juga seluruhnya diprediksi dengan tepat sebagai Tinggi, serta data Rendah sebanyak 10 siswa seluruhnya sesuai dengan prediksi. Tidak terdapat kesalahan klasifikasi pada setiap kelas (nilai 0 pada semua kolom selain diagonal utama).

Selain itu, nilai precision dan recall untuk masing-masing kelas juga mencapai 100%, yang berarti model memiliki kemampuan sangat baik dalam mengidentifikasi setiap kelas.

4.7. Hasil Pengelompokan K-Means dan Klasifikasi Naïve Bayes

Berdasarkan hasil pengolahan data menggunakan algoritma K-Means, siswa dikelompokkan ke dalam tiga kategori, yaitu prestasi tinggi 52 siswa, sedang 10 siswa, dan rendah 24 siswa. Kelompok prestasi tinggi didominasi oleh siswa dengan nilai tinggi, kehadiran baik, mengikuti kursus, dan memiliki minat belajar tinggi. Sebaliknya, kelompok prestasi rendah umumnya memiliki kehadiran rendah, tidak mengikuti kursus, serta minat belajar yang kurang.

Selanjutnya, hasil pengelompokan dianalisis menggunakan algoritma Naïve Bayes untuk mengetahui faktor yang memengaruhi prestasi siswa. Hasil klasifikasi menunjukkan tingkat akurasi sebesar 100%, yang berarti model mampu mengklasifikasikan data dengan sangat baik.

Berdasarkan hasil tersebut, diketahui bahwa faktor utama yang memengaruhi prestasi siswa adalah kehadiran, keikutsertaan dalam kursus, dan minat belajar. Ketiga faktor ini saling mendukung dalam meningkatkan prestasi akademik siswa.

4.8. Hasil Analisis Faktor yang Mempengaruhi Prestasi Siswa

Berdasarkan hasil klasifikasi menggunakan algoritma Naïve Bayes yang didasarkan pada pengelompokan K-Means, diketahui bahwa faktor yang paling memengaruhi prestasi Bahasa Inggris siswa adalah minat belajar Bahasa Inggris

(lihat gambar 4.1). Hal ini ditunjukkan pada Gambar 4.1 yang menampilkan hasil pengolahan data menggunakan RapidMiner, di mana variabel minat belajar memiliki kontribusi yang paling dominan dalam proses klasifikasi. Dominasi tersebut terlihat dari peran minat belajar dalam menentukan pembentukan model klasifikasi, baik melalui nilai probabilitas yang lebih signifikan dibandingkan variabel lain maupun pengaruhnya terhadap hasil akhir prediksi prestasi siswa. Dengan kata lain, perubahan pada nilai minat belajar memberikan dampak yang lebih besar terhadap hasil klasifikasi dibandingkan variabel kehadiran maupun kursus, sehingga minat belajar dapat diidentifikasi sebagai faktor utama yang memengaruhi prestasi siswa.

Hasil analisis menunjukkan bahwa siswa dengan prestasi tinggi umumnya memiliki minat belajar yang tinggi, sehingga mereka lebih aktif dalam memahami materi dan berusaha meningkatkan kemampuan. Sementara itu, siswa dengan prestasi sedang memiliki tingkat minat belajar yang cukup, namun belum konsisten sehingga hasil yang diperoleh masih bervariasi. Adapun siswa dengan prestasi rendah cenderung memiliki minat belajar yang rendah, sehingga kurang termotivasi dalam mengikuti proses pembelajaran.

Dengan demikian, dapat disimpulkan bahwa minat belajar merupakan faktor utama dalam menentukan prestasi siswa. Minat belajar yang tinggi akan mendorong siswa untuk lebih fokus, aktif, dan memiliki keinginan kuat untuk memahami materi Bahasa Inggris. Sebaliknya, rendahnya minat belajar dapat menyebabkan kurangnya perhatian dan usaha dalam belajar. Oleh karena itu, upaya peningkatan prestasi siswa sebaiknya difokuskan pada peningkatan minat

belajar, misalnya melalui metode pembelajaran yang menarik, pemberian motivasi, serta lingkungan belajar yang mendukung, sehingga siswa lebih terdorong untuk belajar secara optimal.