

BAB IV PEMBAHASAN DAN HASIL

4.1 Preprocessing Data

Dalam penelitian ini dataset yang digunakan adalah data yang digunakan berasal dari dataset public Kaggle yaitu data siswa <https://www.kaggle.com/spscientist/student-performance-in-exams/data>. Jumlah record data sebanyak 1.000 yang merupakan jumlah siswa dari berbagai latar belakang yang beraneka ragam, jumlah atribut dataset sebanyak 8 atribut yang merupakan faktor – faktor yang dimiliki oleh para siswa.

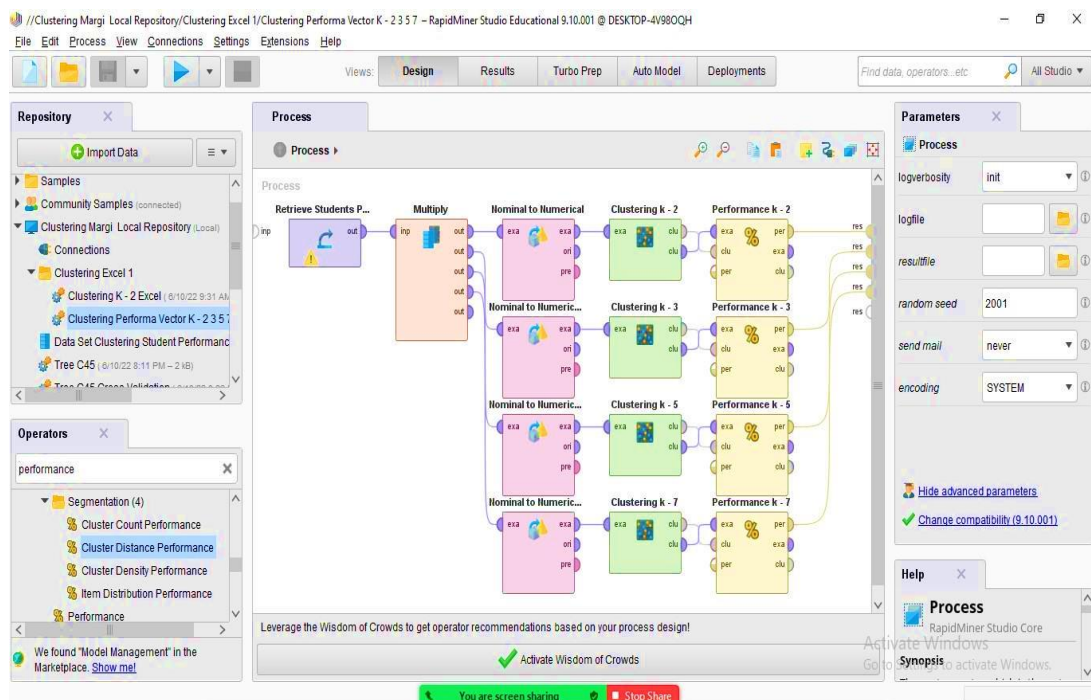
Tabel 4.1 Dataset Student Performance in Exams

1	Gender	Race/Ethnicity	Parental Level of Education	Lunch	Test Preparation Course	Math Score	Reading Score	Writing Score
2	female	group B	bachelor's degree	standard	none	72	72	74
3	female	group C	some college	standard	completed	69	90	88
4	female	group B	master's degree	standard	none	90	95	93
5	male	group A	associate's degree	free/reduced	none	47	57	44
6	male	group C	some college	standard	none	76	78	75
7	female	group B	associate's degree	standard	none	71	83	78
8	female	group B	some college	standard	completed	88	95	92
9	male	group B	some college	free/reduced	none	40	43	39
10	male	group D	high school	free/reduced	completed	64	64	67
11	female	group B	high school	free/reduced	none	38	60	50
12	male	group C	associate's degree	standard	none	58	54	52
13	male	group D	associate's degree	standard	none	40	52	43
14	female	group B	high school	standard	none	65	81	73
15	male	group A	some college	standard	completed	78	72	70
16	female	group A	master's degree	standard	none	50	53	58
17	female	group C	some high school	standard	none	69	75	78
18	male	group C	high school	standard	none	88	89	86
19	female	group B	some high school	free/reduced	none	18	32	28
20	male	group C	master's degree	free/reduced	completed	46	42	46
21	female	group C	associate's degree	free/reduced	none	54	58	61
22	male	group D	high school	standard	none	66	69	63
23	female	group B	some college	free/reduced	completed	65	75	70

Tabel diatas menunjukkan karakter data pada penelitian ini, data tersebut belum memiliki class label yang berarti data ini merupakan data Unsupervised. Alat yang digunakan pada penelitian ini adalah aplikasi Rapidminer.

4.2 Clustering Data

Clustering data diawali dengan menyiapkan dataset “student-performance-in-exams” selanjutnya mengimpor data tersebut kedalam tool Rapid Minner, menentukan berapa jenis kluster yang akan kita buat, dalam hal ini kita menentukan sebanyak 4 kluster dengan masing-masing nilai k adalah : $k = 2$, $k = 3$, $k = 5$, $k = 7$. Dari ke – 4 kluster tersebut akan kita ambil kluster dengan performance terbaik dimana hal tersebut dapat kita lihat dengan index *Davies Bouldin*. Disini kita memasukkan beberapa operator diantaranya : Multiply operator, Nominal to Numeric operator, Clustering Operator (Alogaritma K-Means), dan Performance operator dalam menjalankan pengujian.



Gambar 4.1. Dataset Clustering K-Means

Selain itu perlu beberapa pengaturan pada parameter masing – masing operator. Untuk Multiply dan Nominal to Numeric tidak ada pengaturan khusus pada parameter operator masing – masing. Sementara untuk operator Clustering k – 2 (K-Means), Clustering k – 3 (K-Means), Clustering k – 5 (K-Means), Clustering k – 7 (K-Means) masing –masing kluster disesuaikan dengan nilai k – nya.

Secara ringkas dapat dilihat pada tabel 4.2 :

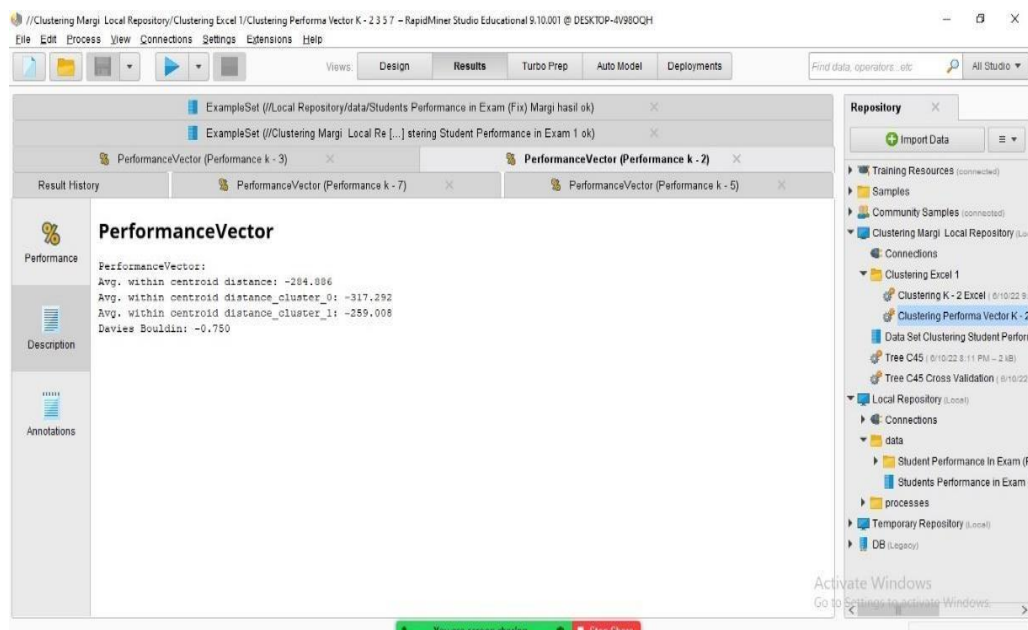
Tabel 4.2. Dataset Clustering K-Means

No	Segmentasi	Nilai k	Measure Types	Mixed Measure
1	Clustering $k - 2$	$k = 2$	<i>Mixed Measure</i>	<i>Mixed Euclidean Distance</i>
2	Clustering $k - 3$	$k = 3$	<i>Mixed Measure</i>	<i>Mixed Euclidean Distance</i>
3	Clustering $k - 5$	$k = 5$	<i>Mixed Measure</i>	<i>Mixed Euclidean Distance</i>
4	Clustering $k - 7$	$k = 7$	<i>Mixed Measure</i>	<i>Mixed Euclidean Distance</i>

Sementara untuk performance operator dari $k - 2$ hingga $k - 7$ diberi perlakuan sama untuk pengaturan parameternya yaitu *main criterion* diatur pada *Avg.within centroid distance*. Saat semuanya tersetting sesuai, maka hasil eksekusi tool RapidMiner dengan Algoritma K-Means yang didapat adalah :

1. Performance Vektor k-2

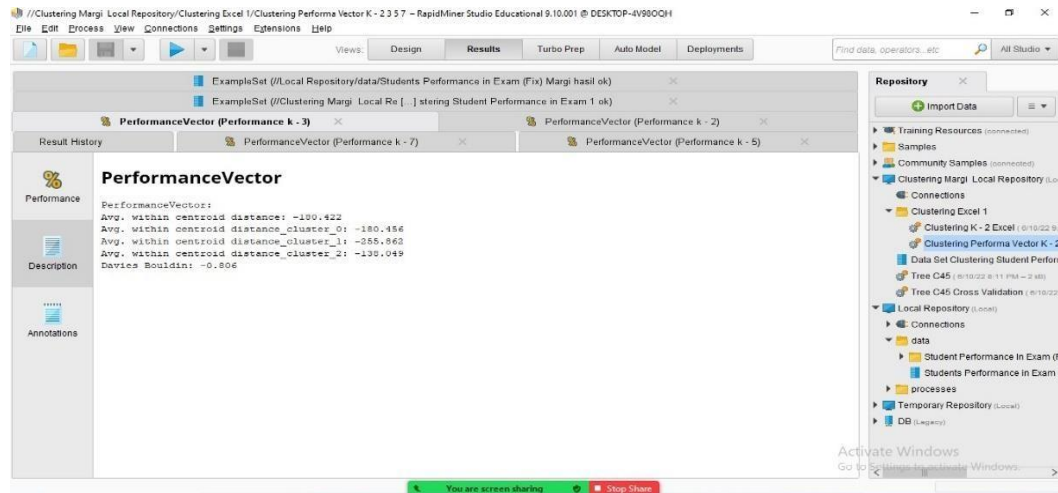
Dengan rata-rata jarak centroid sebesar : -284,886 ; rata-rata jarak centroid dalam kluster_0 sebesar : -317,292 ; rata-rata jarak centroid dalam kluster_1 sebesar : -259,006 dan Davies Bouldin index sebesar : -0,750.



Gambar 4.2. Performance Vektor k-2

2. Performance Vektor k-3

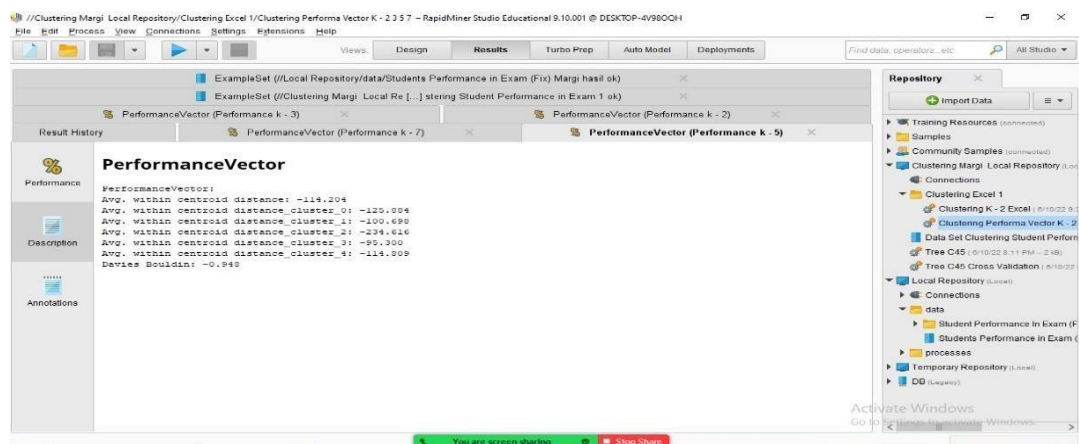
Dengan rata-rata jarak centroid sebesar : -100,422 ; rata-rata jarak centroid dalam kluster_0 sebesar : -180,456 ; rata-rata jarak centroid dalam kluster_1 sebesar : -255,862; rata-rata jarak centroid dalam kluster_2 sebesar : -138,049 dan Davies Bouldin index sebesar : -0,806.



Gambar 4.3. Performance Vektor k-3

3. Performance Vektor k-5

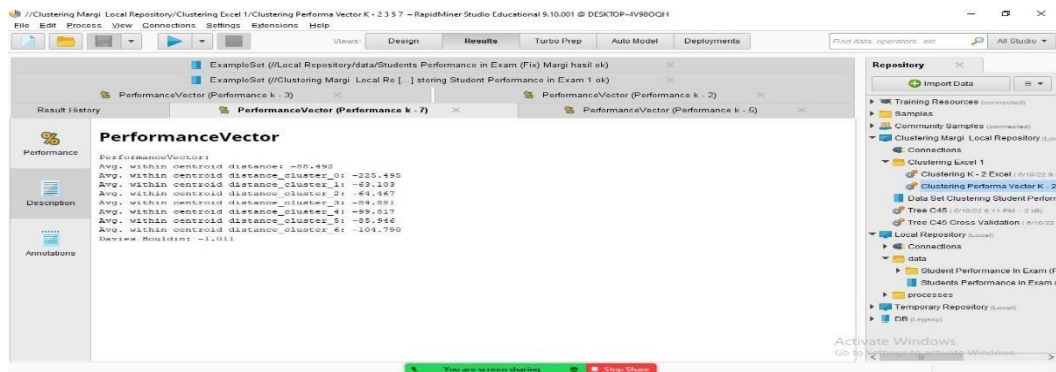
Dengan rata-rata jarak centroid sebesar : -114,204 ; rata-rata jarak centroid dalam kluster_0 sebesar : -125,004; rata-rata jarak centroid dalam kluster_1 sebesar : -100,690; rata-rata jarak centroid dalam kluster_2 sebesar : -234,616; rata-rata jarak centroid dalam kluster_3 sebesar : -95,300; rata-rata jarak centroid dalam kluster_4 sebesar : -114,809 dan Davies Bouldin index sebesar : -0,948.



Gambar 4.4. Performance Vektor k-5

4. Performance Vektor k-7

Dengan rata-rata jarak centroid sebesar $-55,492$; rata-rata jarak centroid dalam kluster₀ sebesar $-255,495$; rata-rata jarak centroid dalam kluster₁ sebesar $-63,101$; rata-rata jarak centroid dalam kluster₂ sebesar $-64,467$ - $234,616$; rata-rata jarak centroid dalam kluster₃ sebesar $-84,581$; rata-rata jarak centroid dalam kluster₄ sebesar $-99,017$; rata-rata jarak centroid dalam kluster₅ sebesar $-85,496$; rata-rata jarak centroid dalam kluster₆ sebesar $-104,790$ dan Davies Bouldin index sebesar $-1,011$



Gambar 4. 5. Performance Vektor k-7

secara ringkas hasil performance Vektor adalah sebagai berikut :

Tabel 4.3. Dataset Performance Vector K-Means

No	Performance Vector	Nilai k	Davies Bouldin Index
1	Clustering k – 2 (K-Means)	k = 2	-0,705
2	Clustering k – 3 (K-Means)	k = 3	-0,806
3	Clustering k – 5 (K-Means)	k = 5	-0,948
4	Clustering k – 7 (K-Means)	k = 7	-1,011

Dari tabel didapat Index Davies Bouldin terkecil (nilai mutlak) dengan nilai sebesar 0.705 pada Clustering dengan Nilai k = 2 dan nilai Index Davies Bouldin terbesar yaitu 1,011. Sehingga clustering /segmentasi dataseta student performance in exams yang dianggap paling ideal adalah clustering/segmentasi dataset menjadi 2 cluster.

Sementara itu auto model cluster data k – 2 didapat hasil berupa tabel clustering :

1. K-Means –Clustered Data Table

Dengan jumlah attribut dataset sebanyak 8 terdiri dari atribut math score, reading score, writing score, gender, lunch, parental education, race/ethnic, dan test preparation yang merupakan faktor – faktor yang dimiliki oleh para siswa dan terbagi menjadi 2 kluster ; kluster_0 dan kluster_1.

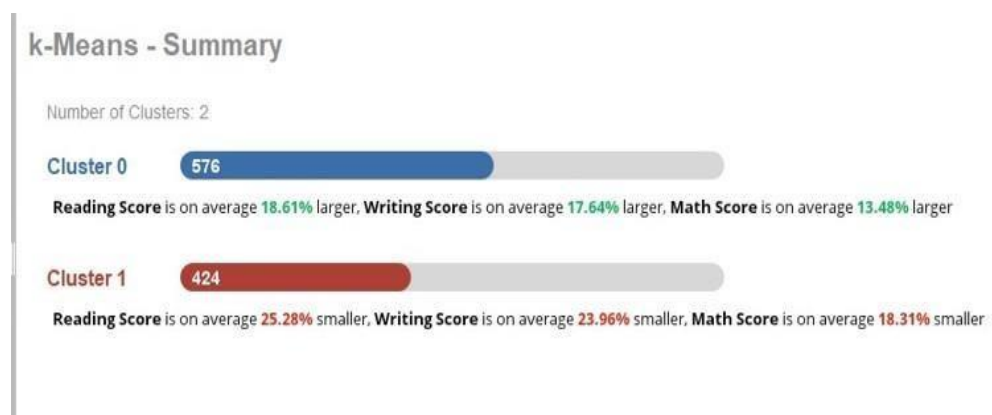
Tabel 4.4. K-Means Clustered Data

k-Means - Clustered Data

Row No.	id	cluster	Math Score	Reading Sc...	Writing Sco...	Gender	Lunch	Parental Le...	Race/Ethni...	Test Prepa...
1	1	cluster_0	72.000	72.000	74.000	female	standard	bachelor's d...	group B	none
2	2	cluster_0	69.000	90.000	88.000	female	standard	some college	group C	completed
3	3	cluster_0	90.000	95.000	93.000	female	standard	master's de...	group B	none
4	4	cluster_1	47.000	57.000	44.000	male	free/reduced	associate's ...	group A	none
5	5	cluster_0	76.000	78.000	75.000	male	standard	some college	group C	none
6	6	cluster_0	71.000	83.000	78.000	female	standard	associate's ...	group B	none
7	7	cluster_0	88.000	95.000	92.000	female	standard	some college	group B	completed
8	8	cluster_1	40.000	43.000	39.000	male	free/reduced	some college	group B	none

2. K-Means – Summary Data

K-Means Summary Data (catatan data) menampilkan jumlah kluster yaitu kluster_0 terdiri dari 576 record data siswa dengan rata-rata reading score 18,61% lebih besar, rata-rata writing score 17,64% lebih besar, dan math score 13,48% lebih besar, sementara kluster_1 yang terdiri dari 424 record data siswa rata-rata reading score 25,28% lebih kecil, rata-rata writing score 23,96% lebih kecil, dan math score 18,31% lebih kecil.



Gambar 4.6. K-Means Summary Data

3. K-Means – Centroid Table Data

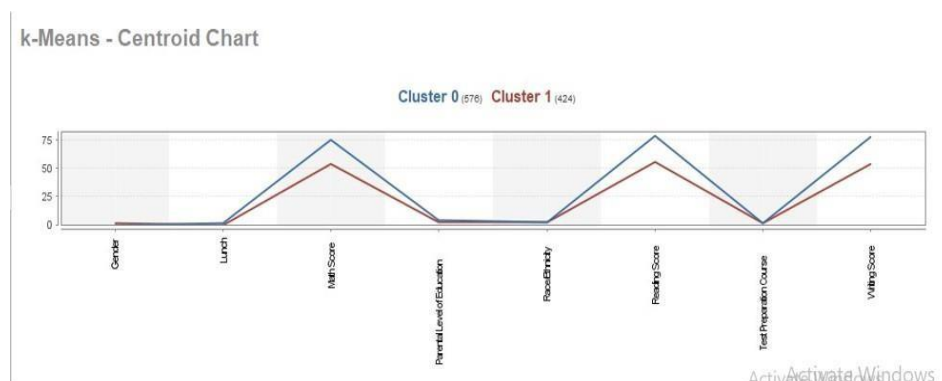
K-Means Centroid Data Table (tabel 4.5) menampilkan titik pusat kluster_0 terdiri dari attribute gender dengan nilai 0 , lunch dengan nilai 1, math score sebesar 74,995, Parental Level Education dengan nilai 4, race dengan nilai 2, reading score 78,878, test preparation 1, dan writing score 78,295 dan titik pusat kluster_0 terdiri dari attribute gender dengan nilai 1, lunch dengan nilai 0, math score sebesar 55,991, Parental Level Education dengan nilai 2, race dengan nilai 2, reading score 55,979, test preparation 1, dan writing score 54,142.

Tabel 4.5. K-Means Centroid Tabel

Cluster	Gender	Lunch	Math Score	Parental Level ...	Race/Ethnicity	Reading Score	Test Preparatio...	Writing Score
Cluster 0	0	1	74.995	4	2	78.878	1	78.295
Cluster 1	1	0	55.991	2	2	55.979	1	54.142

4. K-Means – Centroid Chart Data

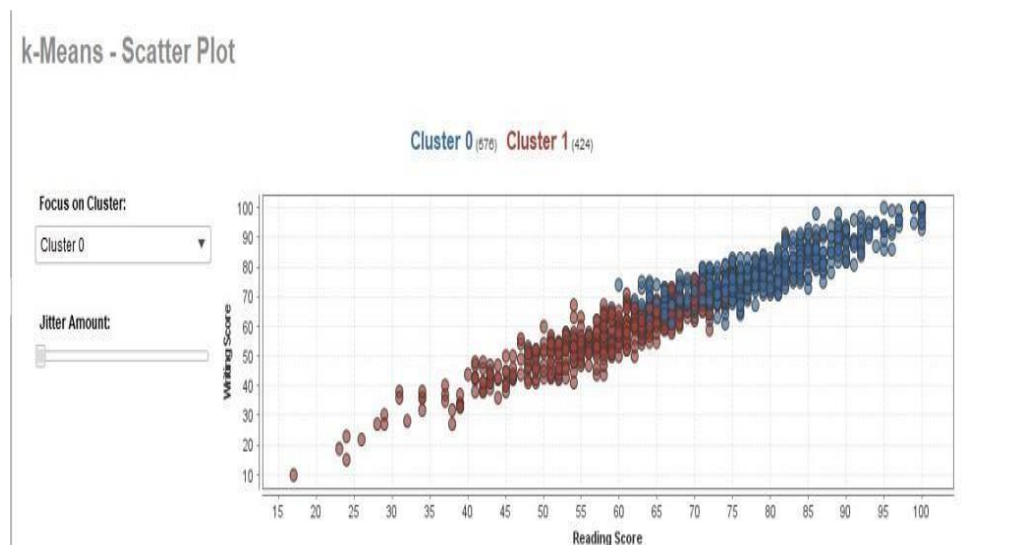
K-Means Centroid Data chart menampilkan chart atau bagan nilai dari atribut baik kluster_0 dan kluster _1 dimana dalam bagan ini menunjukkan bahwa nilai Lunch, Math Score, Parental Level Education, Race, Reading score, Test preparation score, dan writing score yang terlihat sangat berpengaruh terhadap student performance.



Gambar 4.6. K-Means Centroid Chart

5. K-Means – Scatter Plot Reading Score Data

K-Means Scatter Plot atau Plot sebar (alias bagan sebar, grafik sebar) menggunakan titik untuk mewakili nilai untuk dua variabel numerik yang berbeda. Posisi setiap titik pada sumbu horizontal dan vertikal menunjukkan nilai untuk titik data individual. Plot pencar kita digunakan untuk mengamati hubungan antar `kluster_0` dan `kluster_1` untuk Reading Score.



Gambar 4.7. K-Means Scatter Plot

4.2.1 Hasil Clustering

Setelah melakukan pengolahan data set dengan algoritma clustering dengan hingga memperoleh hasil berupa data set jumlah atribut dataset sebanyak 8 terdiri dari atribut math score, reading score, writing score, gender, lunch, parental education, race/ethnic, dan test preparation dengan penambahan satu class label yang terbagi menjadi 2 kluster ; `kluster_0` (Low Performance Student Cluster) dan `kluster_1` (High Performance Student Cluster). Selanjutnya secara dapat dilihat pada tabel 4.6. Hasil processing data set dengan label `cluster_0` (tabel 4.7) dan `cluster_1` (tabel 4.8), maka dataset tersebut kita export dan disimpan dalam bentuk file Excel untuk selanjutnya dapat kita gunakan sebagai dataset prediksi.

Tabel 4.6. K-Means Clustered Data

k-Means - Clustered Data

Row No.	id	cluster	Math Score	Reading Sc...	Writing Sco...	Gender	Lunch	Parental Le...	Race/Ethni...	Test Prepa...
1	1	cluster_0	72.000	72.000	74.000	female	standard	bachelor's d...	group B	none
2	2	cluster_0	69.000	90.000	88.000	female	standard	some college	group C	completed
3	3	cluster_0	90.000	95.000	93.000	female	standard	master's de...	group B	none
4	4	cluster_1	47.000	57.000	44.000	male	free/reduced	associate's ...	group A	none
5	5	cluster_0	76.000	78.000	75.000	male	standard	some college	group C	none
6	6	cluster_0	71.000	83.000	78.000	female	standard	associate's ...	group B	none
7	7	cluster_0	88.000	95.000	92.000	female	standard	some college	group B	completed
8	8	cluster_1	40.000	43.000	39.000	male	free/reduced	some college	group B	none

Tabel 4.7. Low Performance Student Cluster

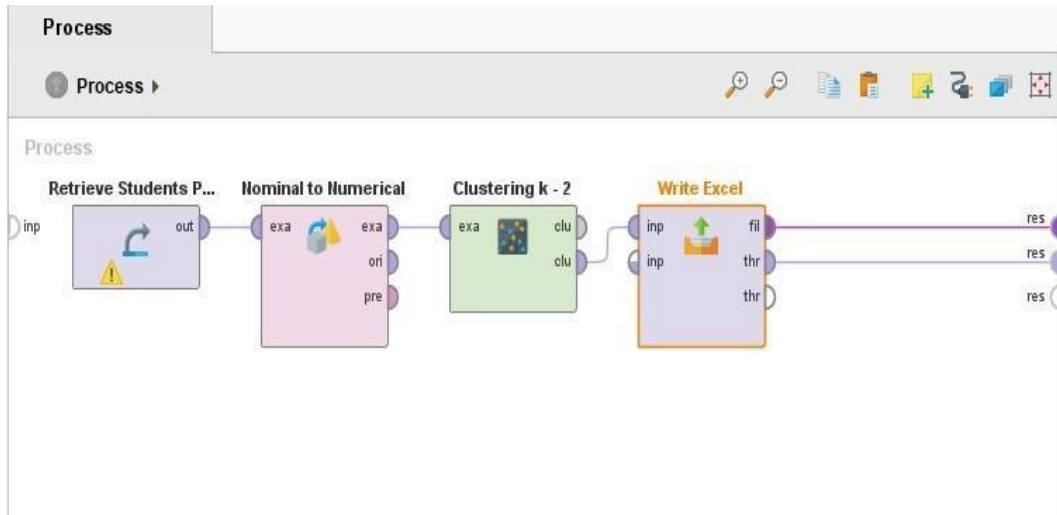
Math Score	Reading Score	Writing Score	Gender	Lunch	Parental Level of Education	Race/Ethnicity	Test Preparation Course	cluster	id
72,0	72,0	74,0	female	standard	bachelor's degree	group B	none	cluster_0	1,0
69,0	90,0	88,0	female	standard	some college	group C	completed	cluster_0	2,0
90,0	95,0	93,0	female	standard	master's degree	group B	none	cluster_0	3,0
76,0	78,0	75,0	male	standard	some college	group C	none	cluster_0	5,0
71,0	83,0	78,0	female	standard	associate's degree	group B	none	cluster_0	6,0
88,0	95,0	92,0	female	standard	some college	group B	completed	cluster_0	7,0
65,0	81,0	73,0	female	standard	high school	group B	none	cluster_0	13,0
78,0	72,0	70,0	male	standard	some college	group A	completed	cluster_0	14,0
69,0	75,0	78,0	female	standard	some high school	group C	none	cluster_0	16,0
88,0	89,0	86,0	male	standard	high school	group C	none	cluster_0	17,0
65,0	75,0	70,0	female	free/reduced	some college	group B	completed	cluster_0	22,0
69,0	73,0	73,0	female	standard	some high school	group C	none	cluster_0	24,0
74,0	71,0	80,0	male	free/reduced	bachelor's degree	group D	completed	cluster_0	25,0
73,0	74,0	72,0	male	free/reduced	master's degree	group A	none	cluster_0	26,0
67,0	69,0	75,0	female	standard	bachelor's degree	group C	none	cluster_0	28,0
70,0	70,0	65,0	male	standard	high school	group C	none	cluster_0	29,0
62,0	70,0	75,0	female	standard	master's degree	group D	none	cluster_0	30,0
69,0	74,0	74,0	female	standard	some college	group D	none	cluster_0	31,0
63,0	65,0	61,0	female	standard	some college	group B	none	cluster_0	32,0

Tabel 4.8. High Performance Student Cluster

Math Score	Reading Score	Writing Score	Gender	Lunch	Parental Level of Education	Race/Ethnicity	Test Preparation Course	cluster	id
47,0	57,0	44,0	male	free/reduced	associate's degree	group A	none	cluster_1	4,0
40,0	43,0	39,0	male	free/reduced	some college	group B	none	cluster_1	8,0
64,0	64,0	67,0	male	free/reduced	high school	group D	completed	cluster_1	9,0
38,0	60,0	50,0	female	free/reduced	high school	group B	none	cluster_1	10,0
58,0	54,0	52,0	male	standard	associate's degree	group C	none	cluster_1	11,0
40,0	52,0	43,0	male	standard	associate's degree	group D	none	cluster_1	12,0
50,0	53,0	58,0	female	standard	master's degree	group A	none	cluster_1	15,0
18,0	32,0	28,0	female	free/reduced	some high school	group B	none	cluster_1	18,0
46,0	42,0	46,0	male	free/reduced	master's degree	group C	completed	cluster_1	19,0
54,0	58,0	61,0	female	free/reduced	associate's degree	group C	none	cluster_1	20,0
66,0	69,0	63,0	male	standard	high school	group D	none	cluster_1	21,0
44,0	54,0	53,0	male	standard	some college	group D	none	cluster_1	23,0
69,0	54,0	55,0	male	standard	some college	group B	none	cluster_1	27,0
56,0	72,0	65,0	female	free/reduced	master's degree	group E	none	cluster_1	33,0
40,0	42,0	38,0	male	standard	some college	group D	none	cluster_1	34,0
50,0	64,0	59,0	female	free/reduced	some high school	group D	none	cluster_1	38,0
57,0	56,0	57,0	male	free/reduced	associate's degree	group B	none	cluster_1	40,0
55,0	61,0	54,0	male	free/reduced	associate's degree	group C	none	cluster_1	41,0
53,0	58,0	65,0	female	standard	associate's degree	group B	none	cluster_1	43,0

4.2.2 Export Data Excel

Tahap selanjutnya didapatkan segmentasi dataset student performance in exams yang menjadi 2 cluster yang akan diexport dan disimpan dalam format file excel. Hasil eksekusi dataset sebagai berikut :



Gambar 4.8. Processing Cluster Data Set to Excel

The screenshot shows the 'Result History' window in Orange Data Mining software. The window title is 'ExampleSet (Clustering k - 2)'. The table displays the following data:

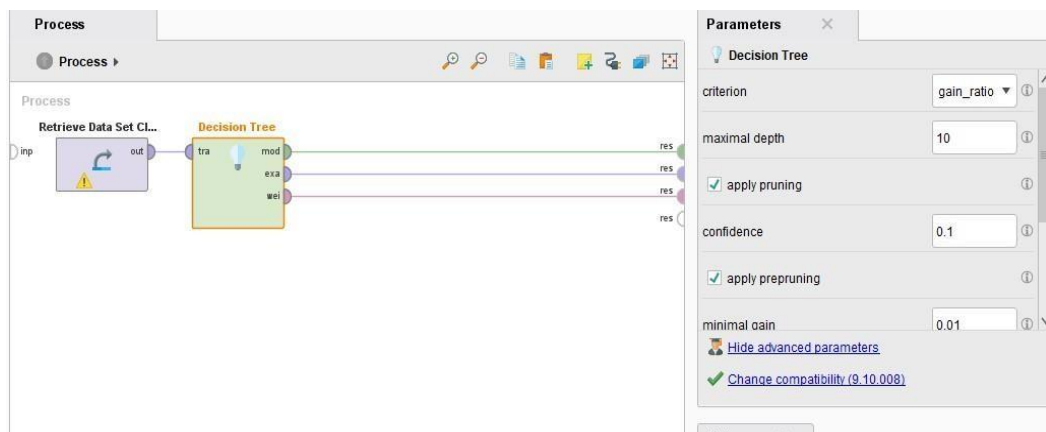
Row No.	id	label	Gender = fe...	Gender = m...	Race/Ethnici...	Race/Ethnici...	Race/Ethnici...	Race/Ethnici...	Race/Ethn...
1	1	cluster_0	1	0	1	0	0	0	0
2	2	cluster_0	1	0	0	1	0	0	0
3	3	cluster_0	1	0	1	0	0	0	0
4	4	cluster_1	0	1	0	0	1	0	0
5	5	cluster_0	0	1	0	1	0	0	0
6	6	cluster_0	1	0	1	0	0	0	0
7	7	cluster_0	1	0	1	0	0	0	0
8	8	cluster_1	0	1	1	0	0	0	0
9	9	cluster_1	0	1	0	0	0	1	0
10	10	cluster_1	1	0	1	0	0	0	0
11	11	cluster_1	0	1	0	1	0	0	0
12	12	cluster_1	0	1	0	0	0	1	0
13	13	cluster_0	1	0	1	0	0	0	0
14	14	cluster_0	0	1	0	0	1	0	0

Gambar 4.9. Clustered Data Set

Dari hasil processing diperoleh data set dengan label cluster_0 dan cluster_1, maka dataset tersebut kita export dan disimpan dalam bentuk file Excel. Untuk selanjutnya dapat kita gunakan sebagai dataset prediksi.

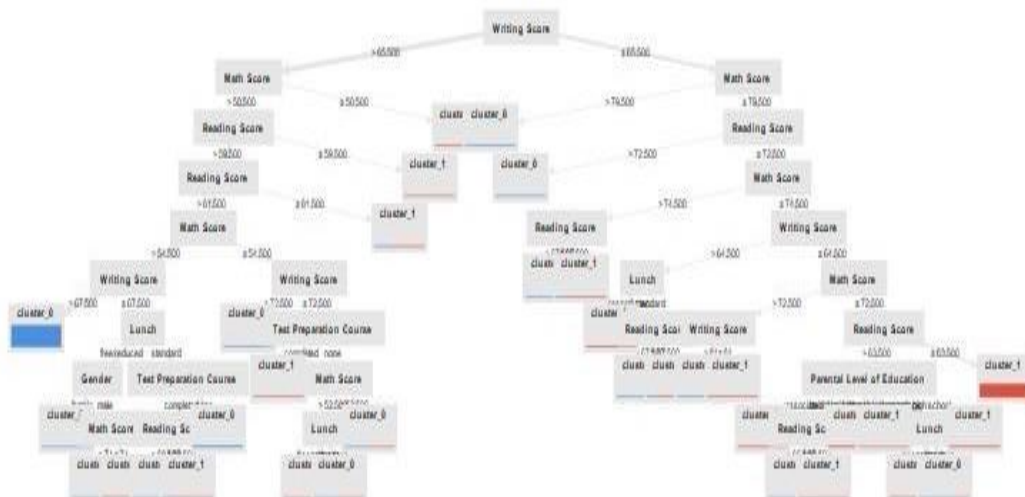
4.3 Prediction Data Algoritma Decision Tree C 4.5

Selesai proses clustering menggunakan algoritma K-Means , tahap berikutnya adalah menguji dataset yang telah didapat untuk mendapatkan suatu prediksi dengan algoritma Decision Tree C 4.5. Langkah uji diawali dengan import dataset clustering ke dalam Tool Rapid Minner kemudian dillakukan pengujian dengan algoritma Decision Tree C 4.5. Tidak lupa atur parameter untuk algoritma Decision Tree C 4.5. Perhatikan gambar di bawah :



Gambar 4.10. Predicting Processing

Didapatkan pohon keputusan sebagai berikut :



Gambar 4.11. Patern Prediction

Selanjutnya didapatkan deskripsi pohon keputusan sebagai berikut :

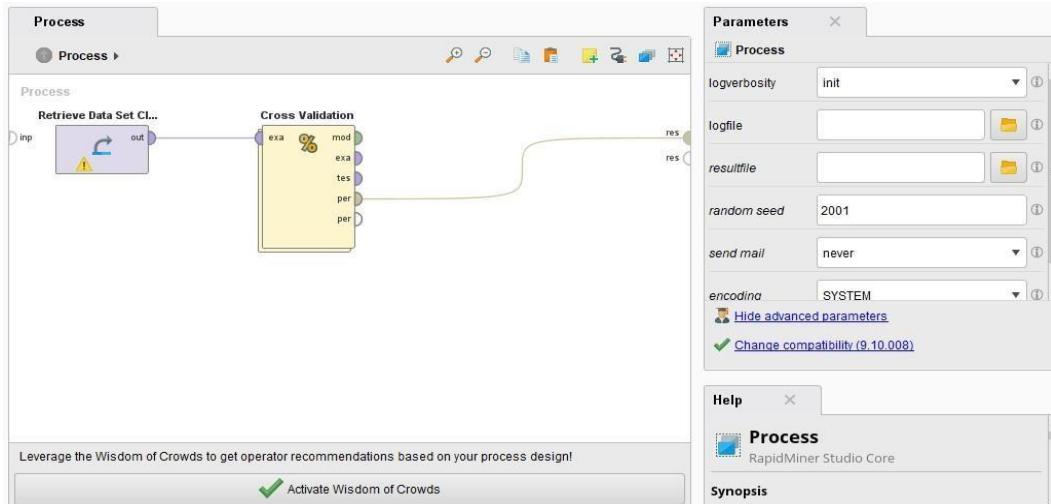


Gambar 4.12. Deskripsi Pohon Keputusan

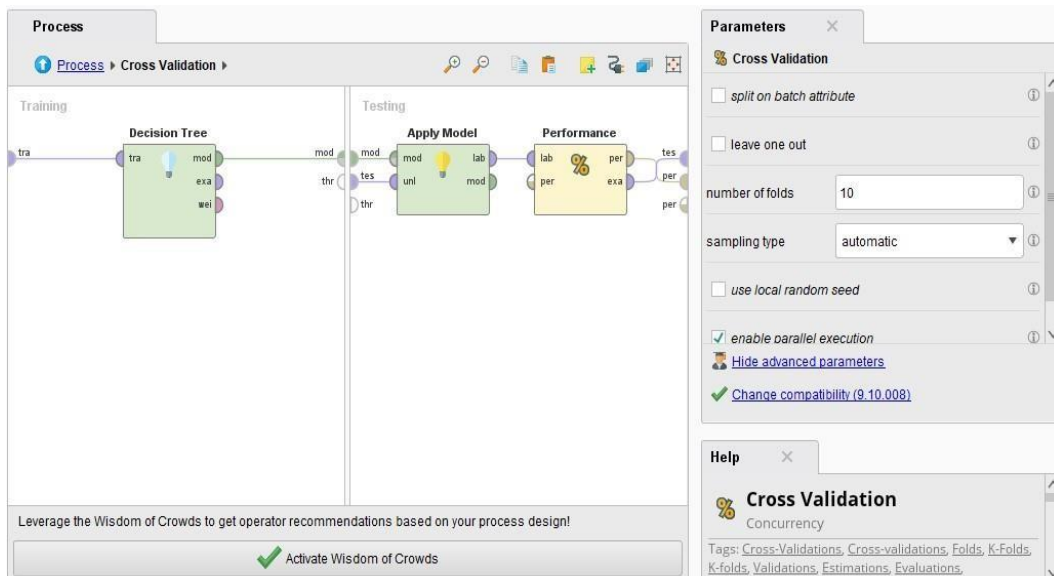
Setelah diketahui pola pohon keputusan berdasarkan dataset, kemudian akan dilakukan uji prediksi performance dan uji tingkat akurasi.

4.3.1 Prediction Performance

Tahap selanjutnya adalah pengujian dataset prediksi dengan menggunakan cross validasi untuk mengetahui seberapa besarnya nilai akurasi, presisi, confusion matrix atau nilai recall dan nilai AUC, disajikan melalui gambar berikut :



Gambar 4.13. Processing Algoritma Decision Tree C4.5 Cross Validation



Gambar 4.14. Processing Algoritma Decision Tree C4.5 Cross Validation

Pada gambar 4.13 dan gambar 4.14 data yang sudah disiapkan diproses dengan aplikasi rapidminer klasifikasi dengan melakukan pengujian menggunakan cross validation operator digunakan adalah decision tree untuk mendapatkan performance akurasi hingga AUC.

PerformanceVector (Performance) X

Criterion

- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Table View Plot View

accuracy: 94.50% +/- 2.51% (micro average: 94.50%)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	554	33	94.38%
pred. cluster_1	22	391	94.67%
class recall	96.18%	92.22%	

Gambar 4.15. Nilai accuracy Algoritma Decision Tree C4.5 Cross Validation

PerformanceVector (Performance) X

Criterion

- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Table View Plot View

precision: 94.77% +/- 3.74% (micro average: 94.67%) (positive class: cluster_1)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	554	33	94.38%
pred. cluster_1	22	391	94.67%
class recall	96.18%	92.22%	

Gambar 4.16. Nilai precision Algoritma Decision Tree C4.5 Cross Validation

PerformanceVector (Performance) X

Criterion

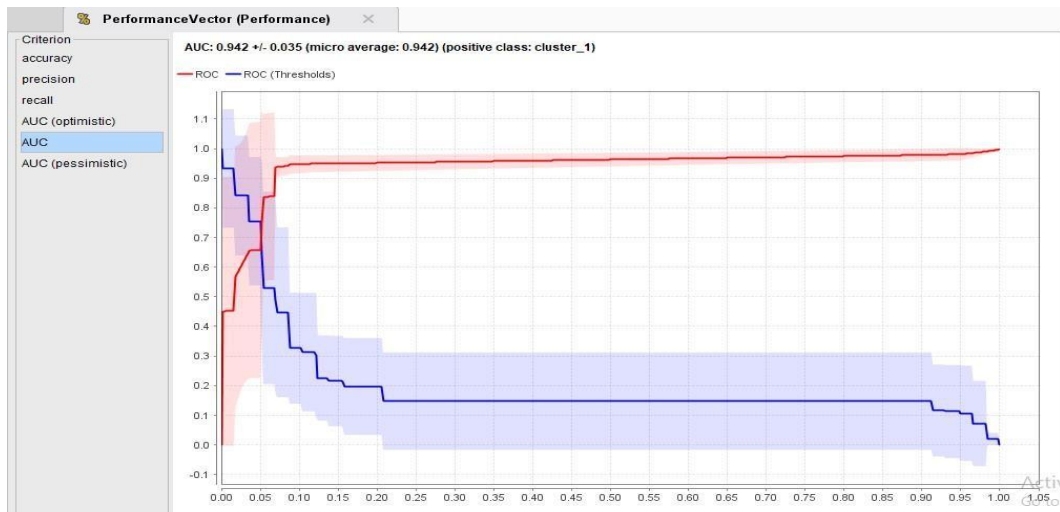
- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Table View Plot View

recall: 92.19% +/- 3.89% (micro average: 92.22%) (positive class: cluster_1)

	true cluster_0	true cluster_1	class precision
pred. cluster_0	554	33	94.38%
pred. cluster_1	22	391	94.67%
class recall	96.18%	92.22%	

Gambar 4.17. Nilai Recall Algoritma Decision Tree C4.5 Cross Validation



Gambar 4.18. Nilai dan Curva AUC Decision Tree C4.5 Cross Validation

Confusion matrix yang menunjukkan hasil eksperimen, didalam confusion matrix kita dapat melihat hasil akurasi, class presisi, dan class recall dan nilai AUC Akurasi yang dihasilkan adalah 94,50 % class presisi 94,77 % dan class recall 92,19 % Nilai AUC 0.942. Perbandingan hasil akurasi pada penelitian student performance beberapa penelitian sebelumnya yang menggunakan metode berbeda dengan penelitian student performance dengan menggunakan metode clustering dan decision tree C 4.5 untuk prediksi student performance ditunjukkan pada tabel 4.9 :

Tabel 4.9. Perbandingan akurasi Penelitian dulu dengan penelitian saat ini

No	Penelitian	Metode	Akurasi
1.	“Data Mining Prestasi Akademik Dengan Naïve Bayes Berdasarkan <i>Attribut Importance</i> (AI)/Ni Komang Sri Julyantari, I Ketut Dedy Suryawan(Komanget al., 2013)/2017	Naïve Bayes, AI	-
2.	“Educational Data Mining & Student Performance Prediction Using SVM Techniques/Mrs. Jamuna dan Mrs.S.A. Shoba(Jamuna & Shoba, 2017)/2017	MLP, NN, NB Tree, REPTree	74,80 %
3.	“Analysis of Student Data base Using Classification Techniques/K. Sumathi, Ph.D, S. Kannan, Ph.D, K.Nagarajan(Sumathi et al., 2016)/2017	DT, J48	91,90%
4.	“Educational Data Mining in Predicting	K-NN with	77,36%

	StudentFinal Grades/ William Willibrobus Damopoli, NathanPriyasadie, Amalia Zahra(The & Academy, 2021)/2021	rapid miner	
5.	“Student Performance Prediction Using Data Mining Techniques”/Durges Ugale, Jeet Pawar, Sacim Yadav, dan Dr.Chandra Sekar Raut/2018	DT, K-NN, Naive Baye	79,00%
6.	Prediksi Kinerja Siswa Menggunakan Algoritma K-Means Dan Alogaritma Decision Tree C4.5	K-Means, DT C 4.5	94,50%

4.4 Evaluasi

Dari hasil pengujian pada penelitian ini, dapat dilihat bahwa alogaritma K-Means membagi dataset menjadi dua kluster yaitu cluster_0 (Low Student Performance Cluster) sebagai cluster rendah dan cluster_1 (High Student Performance Cluster) sebagai cluster tinggi. Akan tetapi sesuai dengan tampilan diagram pencar atau scatter plot diagram (gambar 4.7) nampak beberapa titik data yang kedudukannya masih tersebar jauh dari centroid sehingga terlihat sebagai pencilan data dan juga beberapa titik data dari kedua cluster yang saling tumpang tindih membentuk irisan cluster.

Decision Tree C4.5 meraih performance akurasi yang tinggi dengan nilai accuracy 94,50 %, class presisi 94,77 % dan class recall 92,19 %, dan Nilai AUC 0.942, kinerja algoritma pada penelitian ini menunjukkan performa yang sangat baik sekali dan tingkat akurasi yang sangat tinggi.