

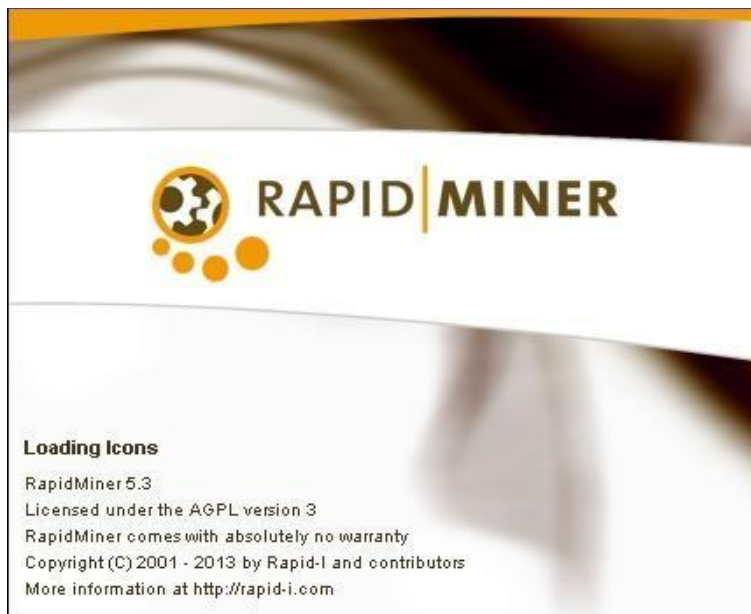
Nama : Shadam Alfito K P
NIM : 2106011
Kelas : Teknik Informatika - A

Tugas Praktikum Kecerdasan Buatan

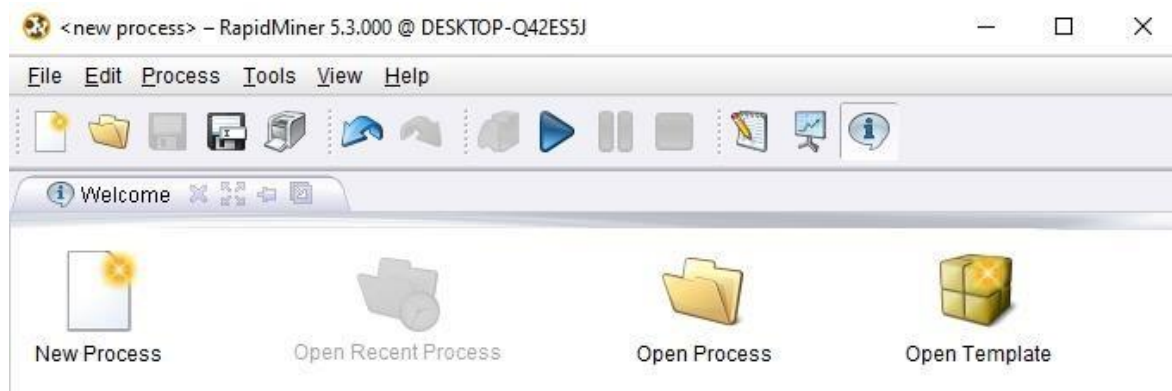
Klasifikasi dan Clustering Menggunakan Rapid Miner

Import Dataset

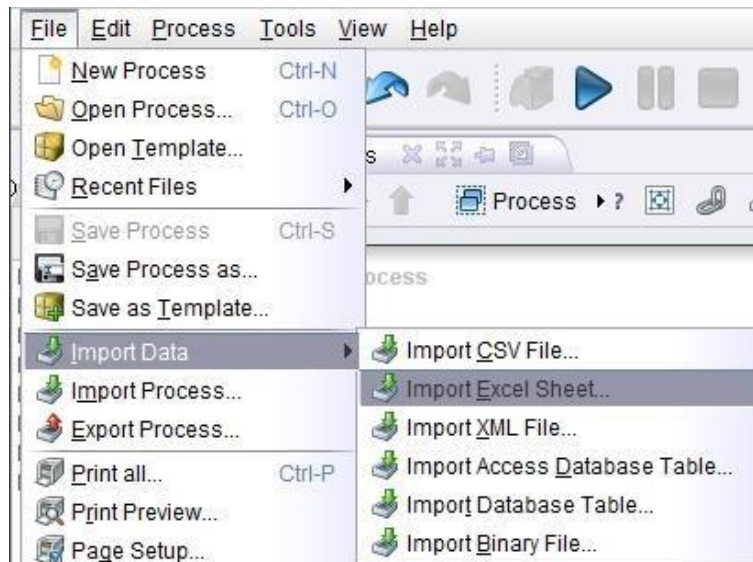
1. Buka RapidMiner 5.



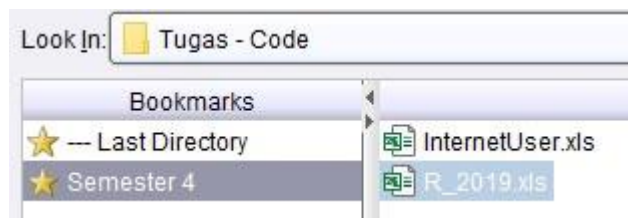
2. Buat proses baru. Pilih New Process.



- Klik Menu File, pilih Import Data lalu pilih Import Excel Sheet.



- Cari dataset yang akan diimpor. Lalu klik Next.



- Setelah terimpor, tampilan akan seperti di bawah ini. Klik Next dan Next lagi.

Report 2019							
A	B	C	D	E	F	G	H
GDP per cap	Social supp	Healthy life e	Freedom to	Generosity	Perceptions	Overall rank	Country or re
1.34	1.587	0.986	0.596	0.153	0.393	1	Finland
1.383	1.573	0.996	0.592	0.252	0.41	2	Denmark
1.488	1.582	1.028	0.603	0.271	0.341	3	Norway
1.38	1.624	1.026	0.591	0.354	0.118	4	Iceland
1.396	1.522	0.999	0.557	0.322	0.298	5	Netherlands
1.452	1.526	1.052	0.572	0.263	0.343	6	Switzerland
1.387	1.487	1.009	0.574	0.267	0.373	7	Sweden
1.303	1.557	1.026	0.585	0.33	0.38	8	New Zealand
1.365	1.505	1.039	0.584	0.285	0.308	9	Canada
1.376	1.475	1.016	0.532	0.244	0.226	10	Austria
1.372	1.548	1.036	0.557	0.332	0.29	11	Australia
1.034	1.441	0.963	0.558	0.144	0.093	12	Costa Rica
1.276	1.455	1.029	0.371	0.261	0.082	13	Israel
1.609	1.479	1.012	0.526	0.194	0.316	14	Luxembourg

Annotation	A	B	C	D	E	F	G	
Name	GDP per cap	Social supp	Healthy life e	Freedom to	Generosity	Perceptions	Overall rank	
-	1.34	1.587	0.986	0.596	0.153	0.393	1	F
-	1.383	1.573	0.996	0.592	0.252	0.41	2	D
-	1.488	1.582	1.028	0.603	0.271	0.341	3	N
-	1.38	1.624	1.026	0.591	0.354	0.118	4	I
-	1.396	1.522	0.999	0.557	0.322	0.298	5	N
-	1.452	1.526	1.052	0.572	0.263	0.343	6	S
-	1.387	1.487	1.009	0.574	0.267	0.373	7	S
-	1.303	1.557	1.026	0.585	0.33	0.38	8	N
-	1.365	1.505	1.039	0.584	0.285	0.308	9	C
-	1.376	1.475	1.016	0.532	0.244	0.226	10	A
-	1.372	1.548	1.036	0.557	0.332	0.29	11	A
-	1.034	1.441	0.963	0.558	0.144	0.093	12	C
-	1.276	1.455	1.029	0.371	0.261	0.082	13	I
-	1.609	1.479	1.012	0.526	0.194	0.316	14	L

6. Selanjutnya sesuaikan type data dari atribut dataset, apakah polynomial, binominal, integer dan sebagainya. Klik Next.

☒ Reload data
 ☒ Guess value types
 ☒ Preview uses only first 100 rows.
 Date format

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
GDP per cap	Social supp	Healthy life e	Freedom to	Generosity	of corruption	Overall rank	ity or region
real	real	real	real	real	real	integer	polyno...
attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute

1.340	1.587	0.986	0.596	0.153	0.393	1	Finland
1.383	1.573	0.996	0.592	0.252	0.410	2	Denmark
1.488	1.582	1.028	0.603	0.271	0.341	3	Norway
1.380	1.624	1.026	0.591	0.354	0.118	4	Iceland
1.396	1.522	0.999	0.557	0.322	0.298	5	Netherlands
1.452	1.526	1.052	0.572	0.263	0.343	6	Switzerland
1.387	1.487	1.009	0.574	0.267	0.373	7	Sweden
1.303	1.557	1.026	0.585	0.330	0.380	8	New Zealand
1.365	1.505	1.039	0.584	0.285	0.308	9	Canada
1.376	1.475	1.016	0.532	0.244	0.226	10	Austria
1.372	1.548	1.036	0.557	0.332	0.290	11	Australia
1.034	1.441	0.963	0.558	0.144	0.093	12	Costa Rica
1.276	1.455	1.029	0.371	0.261	0.082	13	Israel
1.609	1.479	1.012	0.526	0.194	0.316	14	Luxembourg

7. Simpan dataset di local repository dan berikan nama. Lalu klik Finish



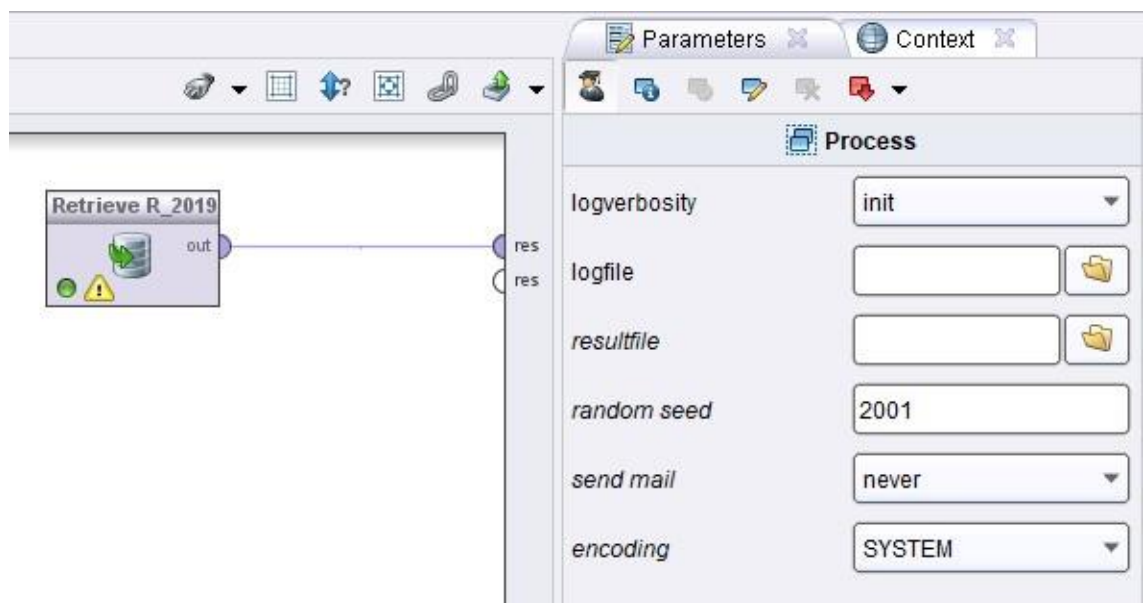
Preprocessing Data

a. Mengecek Missing Value dan Noise

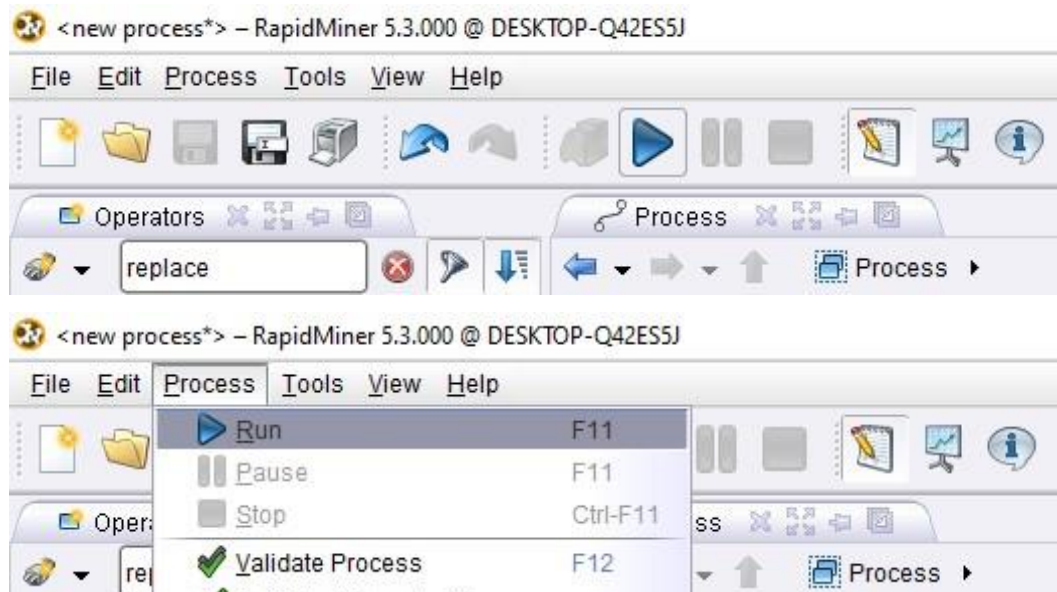
1. Drag and drop dataset yang sudah di simpan di local repository ke workspace.



2. Setelah dataset dipindahkan ke workspace, kita lihat apakah dataset ini memiliki missing value dan noise atau tidak dengan menarik garis dari out yang ada di operator dataset ke res yang ada di atas kanan seperti berikut.



3. Jika sudah, lalu klik tombol Run atau dengan menekan F11.



4. Setelah di Run, maka dataset akan langsung menampilkan seperti gambar di bawah ini.

Result Overview | ExampleSet (Retrieve R_2019) | ExampleSet (//Local Repository/data/R_2019)

☒ Data View ☐ Meta Data View ☐ Plot View ☐ Advanced Charts ☐ Annotations

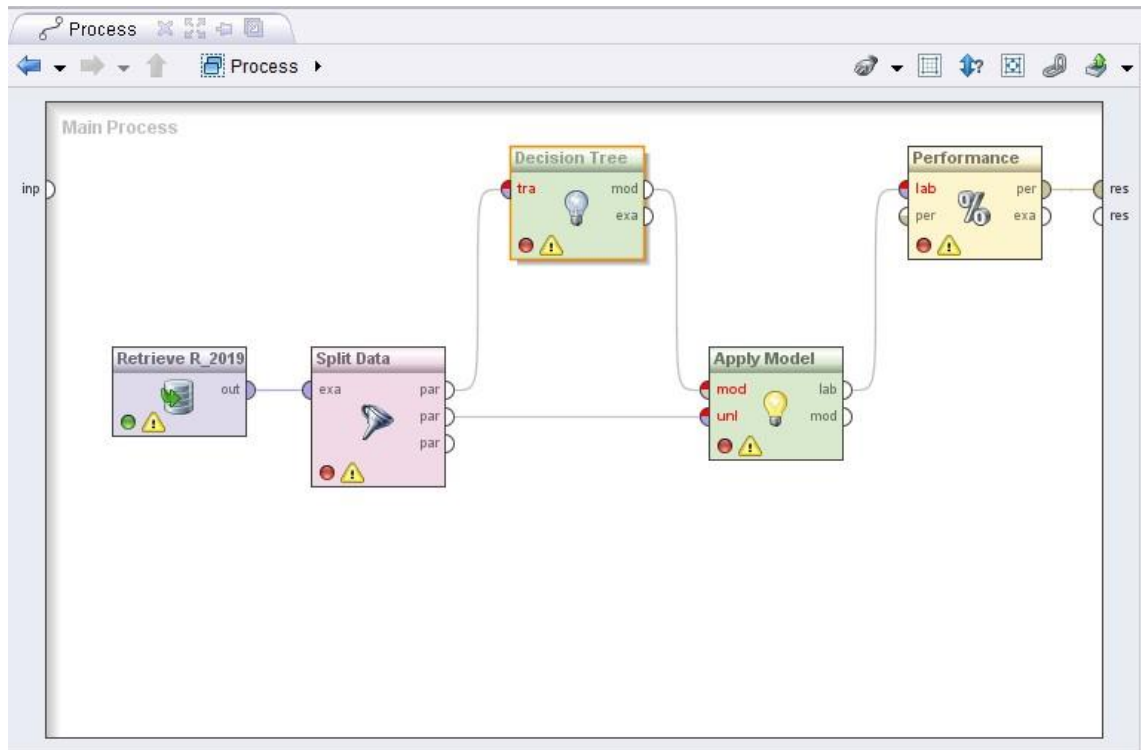
ExampleSet (156 examples, 0 special attributes, 8 regular attributes)

Row No.	GDP per ca...	Social supp...	Healthy life ...	Freedom to ...	Generosity	Perceptions...	Overall rank	Country or r...
1	1.340	1.587	0.986	0.596	0.153	0.393	1	Finland
2	1.383	1.573	0.996	0.592	0.252	0.410	2	Denmark
3	1.488	1.582	1.028	0.603	0.271	0.341	3	Norway
4	1.380	1.624	1.026	0.591	0.354	0.118	4	Iceland
5	1.396	1.522	0.999	0.557	0.322	0.298	5	Netherlands
6	1.452	1.526	1.052	0.572	0.263	0.343	6	Switzerland
7	1.387	1.487	1.009	0.574	0.267	0.373	7	Sweden
8	1.303	1.557	1.026	0.585	0.330	0.380	8	New Zealand
9	1.365	1.505	1.039	0.584	0.285	0.308	9	Canada
10	1.376	1.475	1.016	0.532	0.244	0.226	10	Austria
11	1.372	1.548	1.036	0.557	0.332	0.290	11	Australia
12	1.034	1.441	0.963	0.558	0.144	0.093	12	Costa Rica
13	1.276	1.455	1.029	0.371	0.261	0.082	13	Israel
14	1.609	1.479	1.012	0.526	0.194	0.316	14	Luxembourg
15	1.333	1.538	0.996	0.450	0.348	0.278	15	United Kingd
16	1.499	1.553	0.999	0.516	0.298	0.310	16	Ireland
17	1.373	1.454	0.987	0.495	0.261	0.265	17	Germany
18	1.356	1.504	0.986	0.473	0.180	0.310	18	Belgium

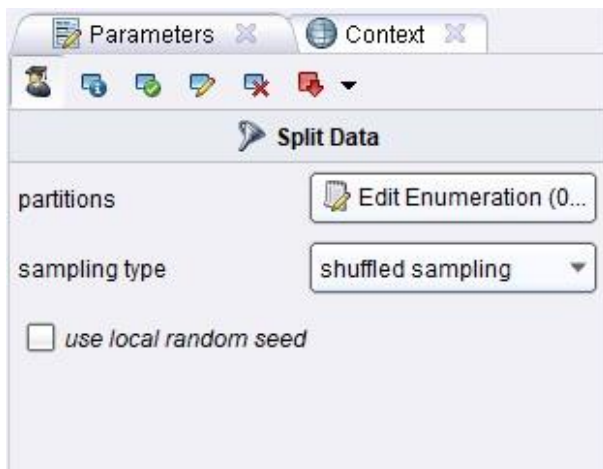
5. Dan setelah dilihat dataset-nya tidak memiliki missing value dan noise sehingga dapat dilanjutkan ke tahapan selanjutnya.

b. Decision Tree (Classifiaction)

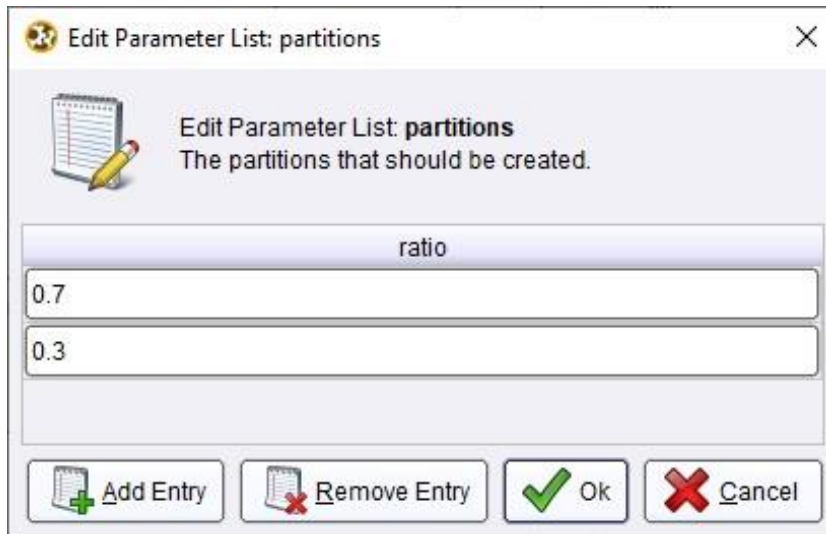
1. Drag and drop operator Decision Tree, Apply Model, Split Data dan Performance dari panel operator di sebelah kiri ke proses di workspace , lalu hubungkan antar operatornya sehingga terlihat seperti berikut.



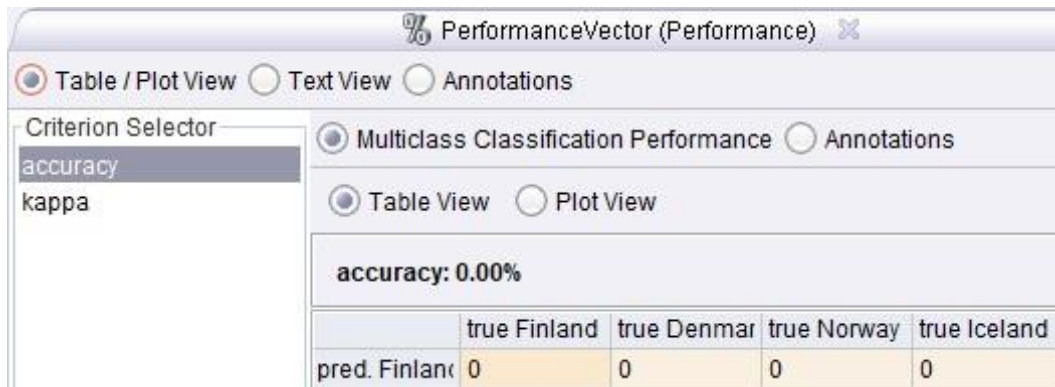
2. Lalu, klik operator Split Data sehingga di menu parameters terlihat seperti berikut



3. Setelah itu klik Edit Enumeration untuk menentukan dataset ini akan dibagi menjadi berapa persen untuk data training dan data testingnya. Untuk kali ini, dataset akan dibagi 70% untuk data training dan 30% untuk data testing, maka buat entry seperti berikut.

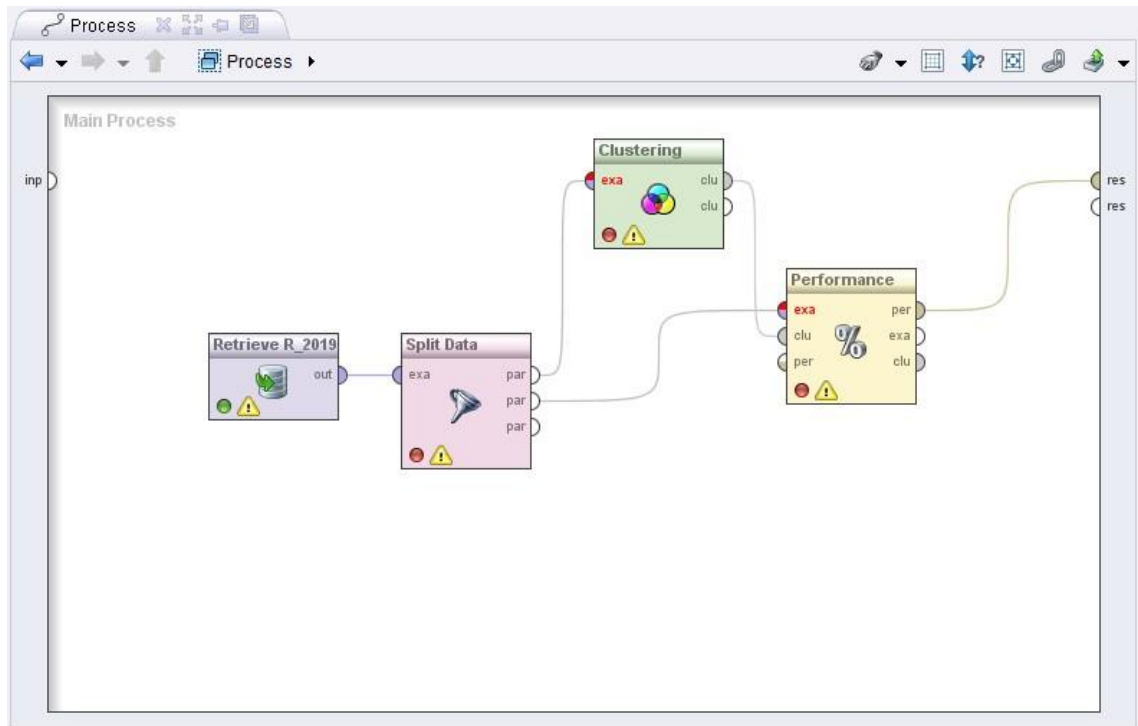


4. Setelah itu jalankan dan hasilnya akan tampil seperti berikut.



c. K-Means (Clustering)

1. Untuk kasus clustering, kita akan menggunakan dataset yang sama dengan yang sebelumnya yaitu dataset bunga Iris. Untuk melakukannya, hapus operator Decision Tree, Apply Model dan Performance. Setelah itu, tarik dan letakkan operator K-Means dan Cluster Distance Performance ke workspace dan sambungkan antar operatornya sehingga seperti berikut.



2. Lalu, klik operator clustering dan ubah nilai k pada parameter menjadi 3 seperti berikut.

3. Setelah itu, jalankan dan akan tampil seperti berikut.

PerformanceVector

PerformanceVector:

Avg. within centroid distance: -242.764

Avg. within centroid distance_cluster_0: -256.141

Avg. within centroid distance_cluster_1: -223.732

Avg. within centroid distance_cluster_2: -250.158

Davies Bouldin: -0.507