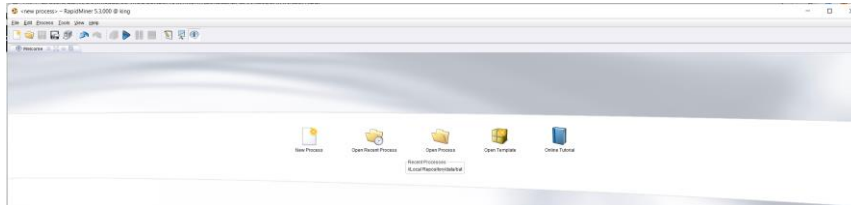


Nama: Gumilar Ichsnaulhaq

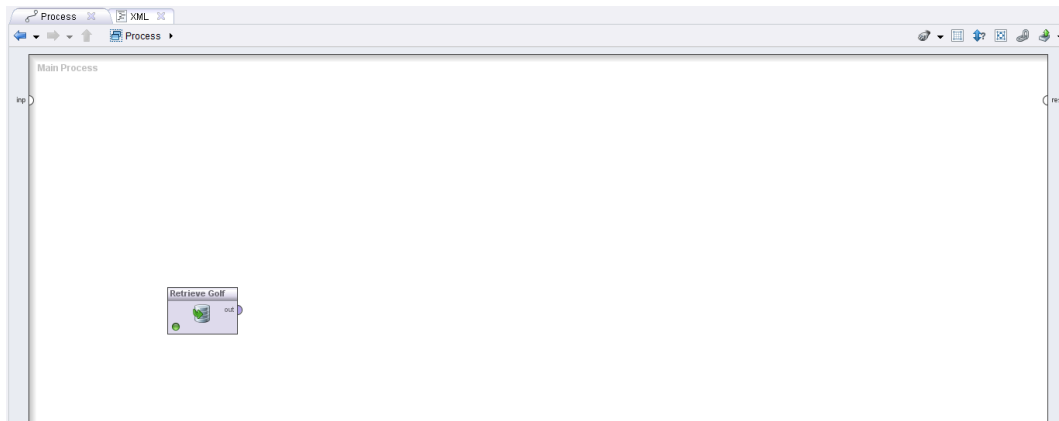
NIM: 2106028

Kelas: A

Buka RapidMiner Studio dan buatlah sebuah proses baru. Pilih New Process.



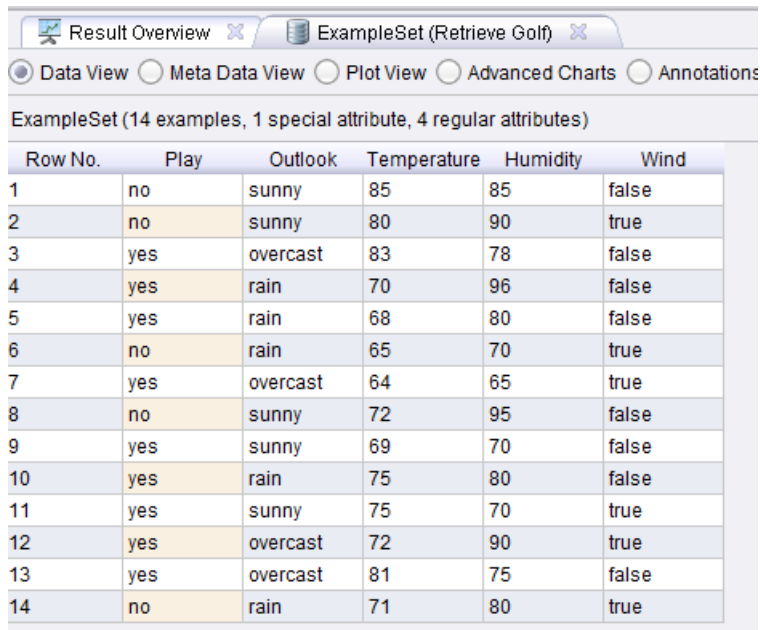
Selanjutnya, ke menu repositories yang ada di sebelah kiri bawah, buka Samples -> data -> golf lalu tarik golf ke workspace sehingga tampil seperti berikut.



Setelah dataset terimpor, 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.



Jika sudah, maka akan tampil seperti berikut.

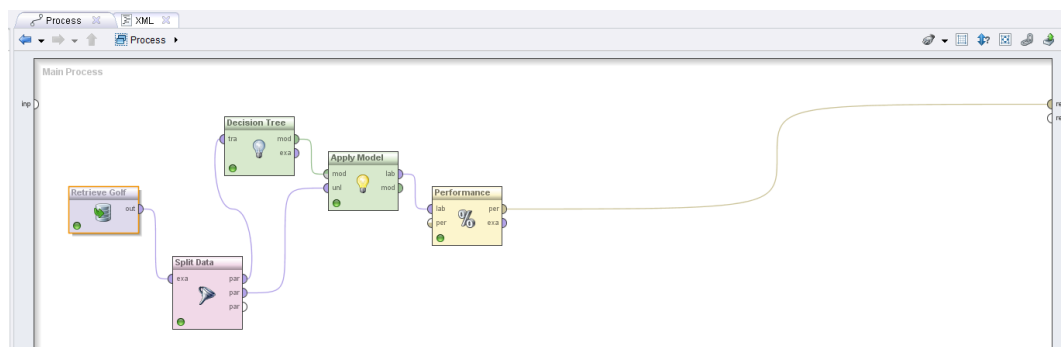


ExampleSet (14 examples, 1 special attribute, 4 regular attributes)

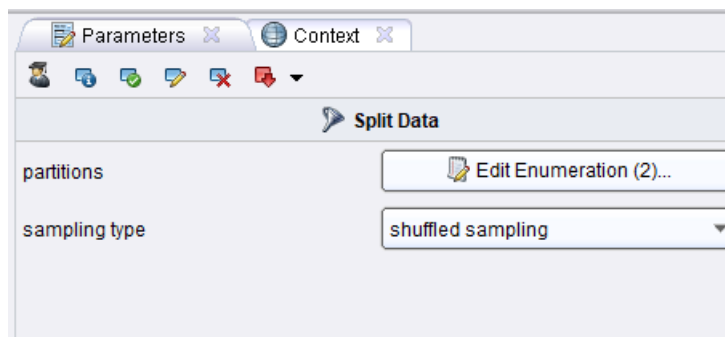
Row No.	Play	Outlook	Temperature	Humidity	Wind
1	no	sunny	85	85	false
2	no	sunny	80	90	true
3	yes	overcast	83	78	false
4	yes	rain	70	96	false
5	yes	rain	68	80	false
6	no	rain	65	70	true
7	yes	overcast	64	65	true
8	no	sunny	72	95	false
9	yes	sunny	69	70	false
10	yes	rain	75	80	false
11	yes	sunny	75	70	true
12	yes	overcast	72	90	true
13	yes	overcast	81	75	false
14	no	rain	71	80	true

Decision Tree (Classifiaction)

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



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



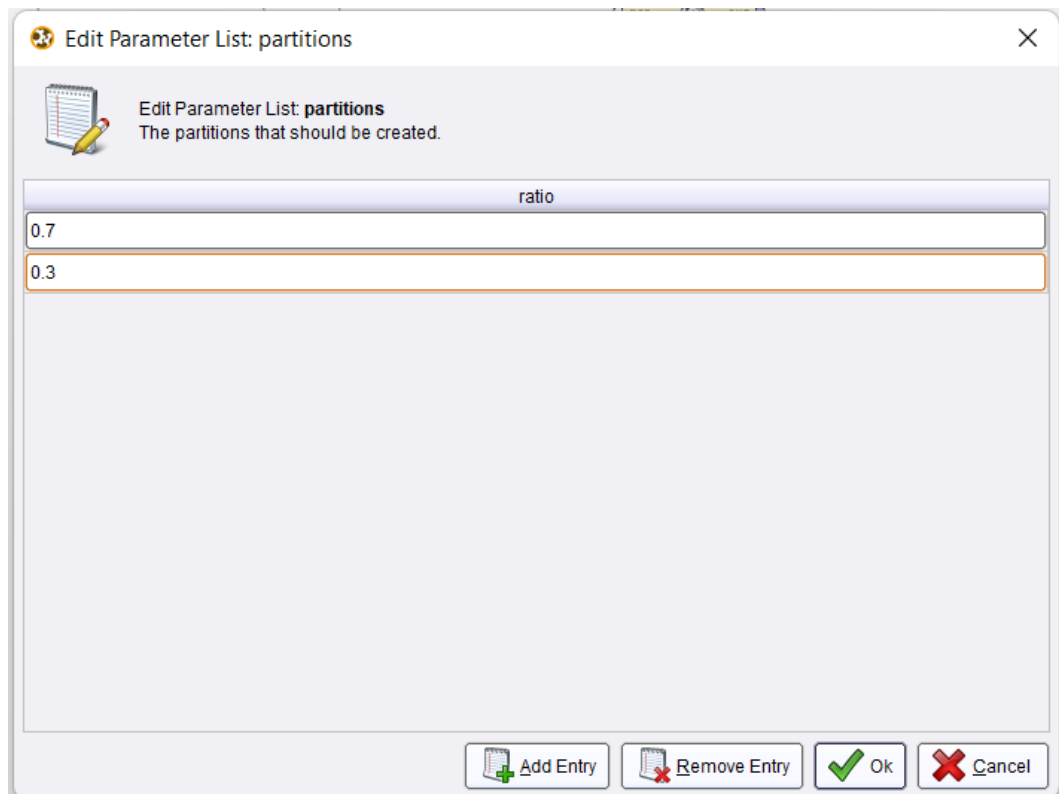
Parameters

Split Data

partitions Edit Enumeration (2)...

sampling type shuffled sampling

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.



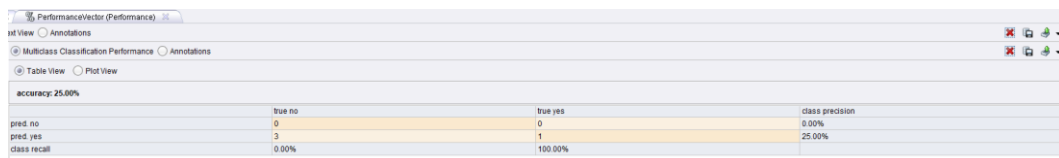
Edit Parameter List: partitions

Edit Parameter List: **partitions**
The partitions that should be created.

ratio
0.7
0.3

Buttons: Add Entry, Remove Entry, Ok, Cancel

Setelah itu jalankan dan hasilnya akan tampil seperti berikut



PerformanceVector (Performance)

nt View Annotations

● Multiclass Classification Performance ○ Annotations

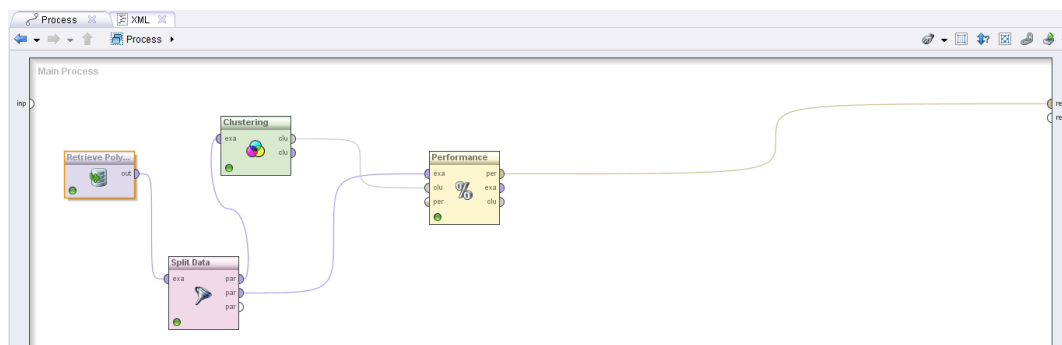
● Table View ○ Plot View

accuracy: 25.00%

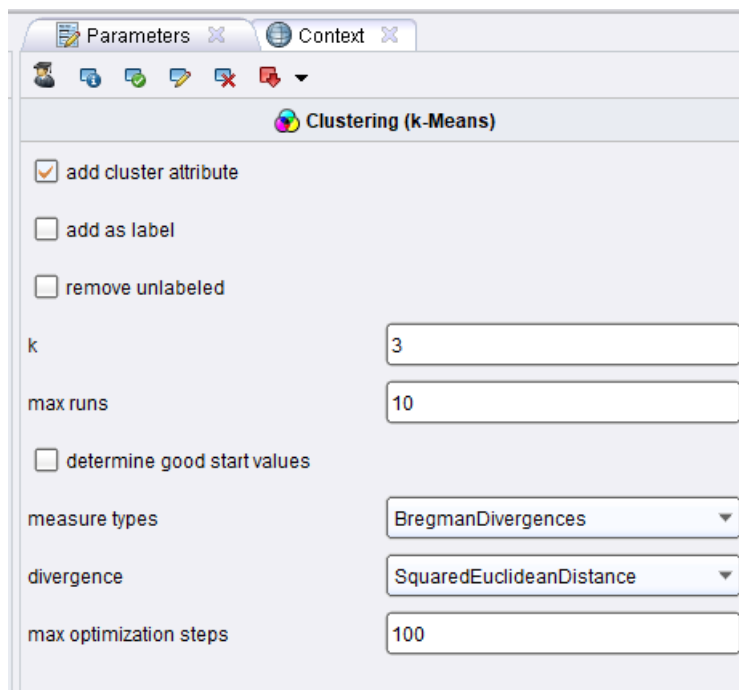
	true no	true yes	class precision
pred no	0	0	0.00%
pred yes	3	1	25.00%
class recall	0.00%	100.00%	

K-Means (Clustering)

Untuk kasus clustering, kita akan menggunakan dataset polynomial dengan yang sebelumnya yaitu dataset polynomial. 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.



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



Setelah itu, jalankan dan akan tampil seperti berikut.

Result Overview

PerformanceVector (Performance)

☒ Table / Plot View

☐ Text View

☐ Annotations

Criterion Selector

Avg. within centroid distance

Avg. within centroid distance_cluster_0

Avg. within centroid distance_cluster_1

Avg. within centroid distance_cluster_2

Davies Bouldin

☒ Text View

☐ Annotations

Avg. within centroid distance

Avg. within centroid distance: -30.688