

## Tutorial Instalasi Rapidminer



RapidMiner adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. RapidMiner dikembangkan dengan model open core. Terdapat 2 versi edisi dari Rapidminer, yang pertama adalah RapidMiner Basic Edition (gratis), yang dibatasi kemampuannya pada 1 prosesor logis dan maksimal 10.000 baris data, tersedia dengan lisensi AGPL, yang kedua adalah Rapidminer versi komersial dengan harga dimulai dari \$ 2.500.

Di kalangan dunia akademik, Rapidminer digunakan oleh Mahasiswa, Dosen atau peneliti dengan latar belakang Sistem Informasi untuk permodelan sistem berbasis kecerdasan buatan (optimasi, pengenalan pola gambar/ teks/grafik, peramalan). Tool yang sejenis dengan Rapidminer adalah Matlab Simulink, WEKA, R, Orange, KNIME serta TANAGRA. Rapidminer dibangun menggunakan bahasa Java sehingga dapat dijalankan diberbagai macam sistem operasi seperti Windows, Linux, UNIX serta Ma cOS.

Cara instalasi :

1. Rapidminer dapat diunduh melalui tautan :

<https://my.rapidminer.com/nexus/account/index.html#downloads>



My Account Profile Downloads Licenses Cloud

## Downloads

Click on a RapidMiner product of your choice to download it.

### RapidMiner Studio 7.5

Click on your operating system to start the download:



32bit



64bit



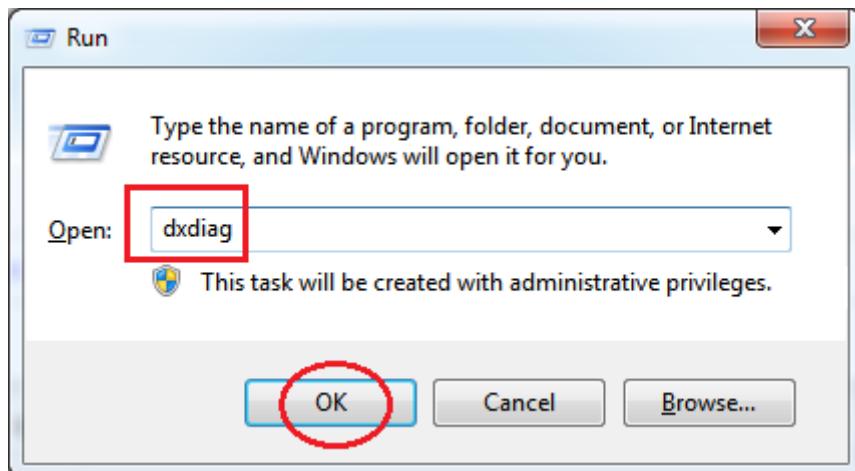
Requires: Mac OS 10.8+



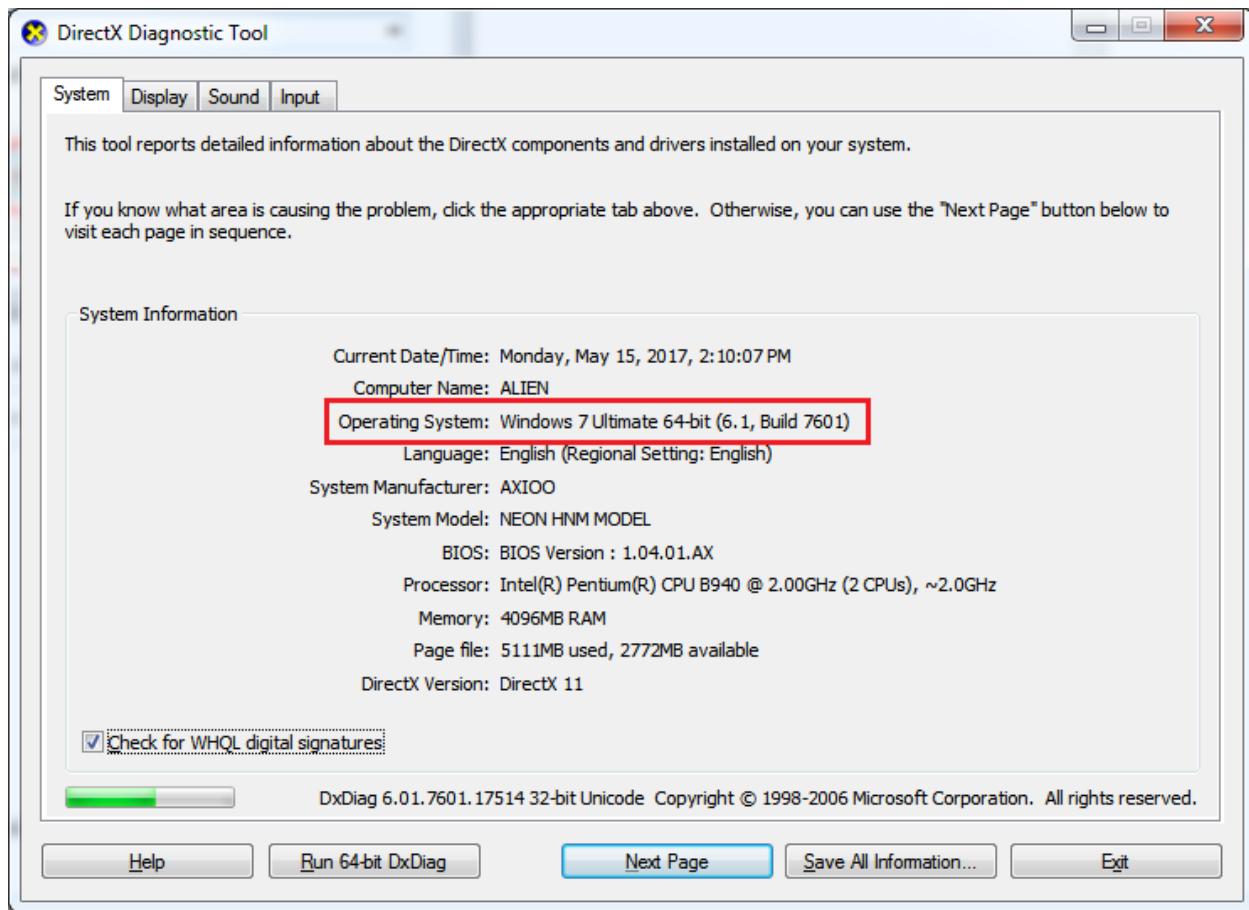
Requires: Java 8

- Installation Guide
- Getting Started Tutorials
- Support
- Download Source

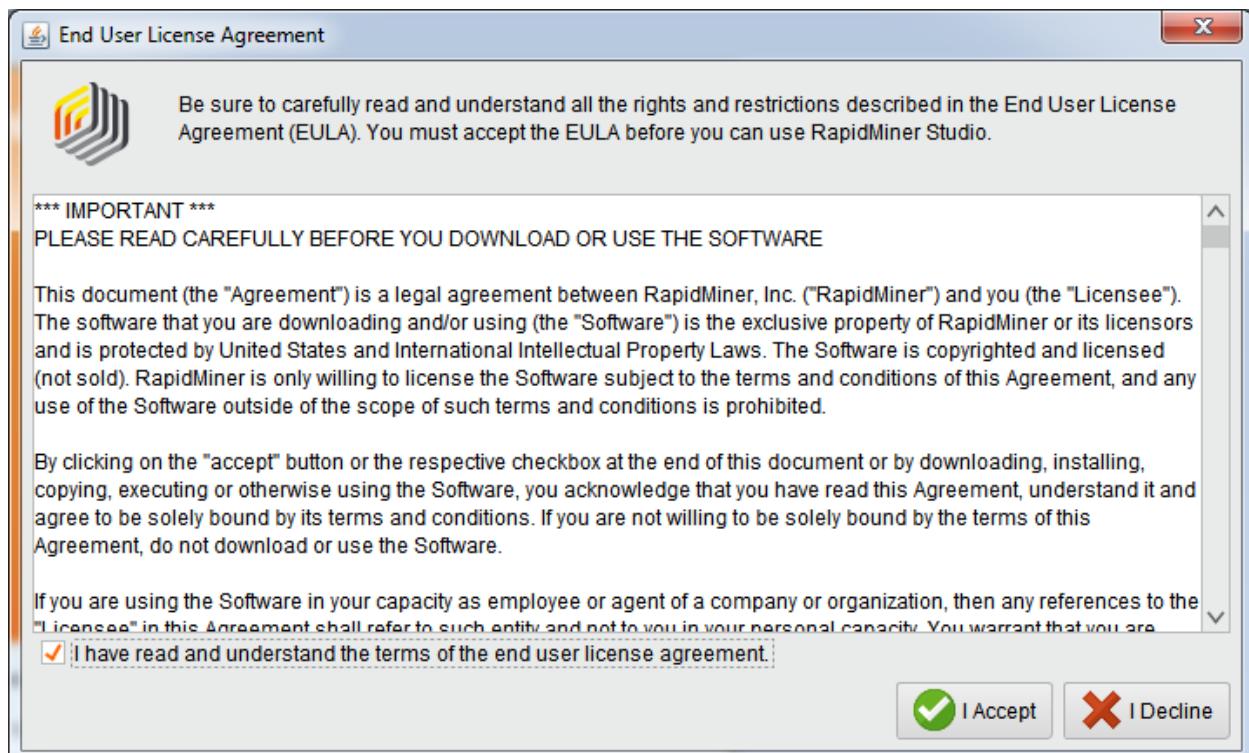
2. Terdapat 2 versi untuk sistem operasi windows, yaitu 32 Bit serta 64 Bit. Untuk mengatahui jenis sistem komputer Anda 32 Bit atau 64 Bit, jalankan perintah dxdiag dengan menekan 2 tombol bersamaan yaitu logo windows dan huruf R, ketikkan dxdiag kemudian klik tombol **OK**



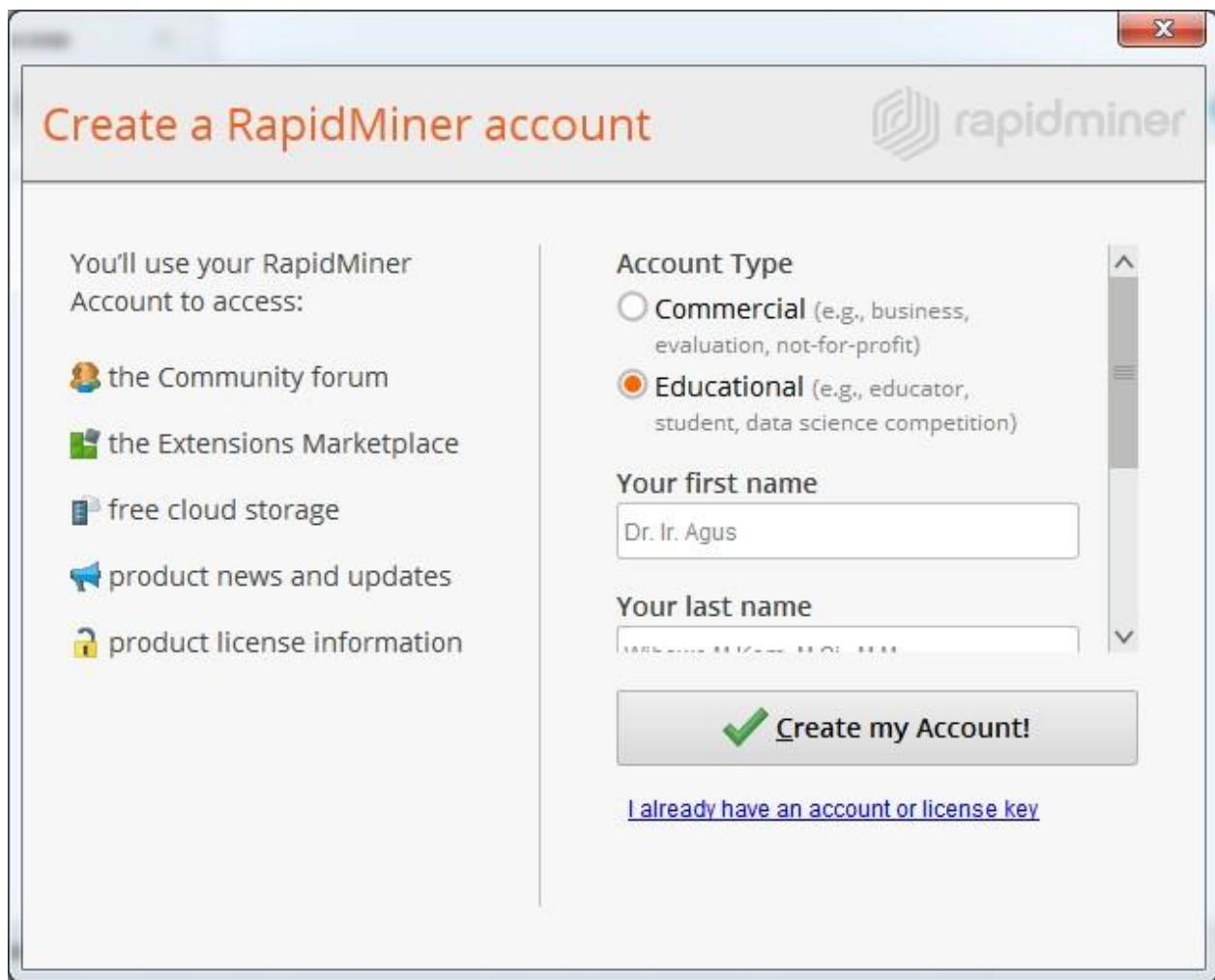
3. Akan terlihat jenis sistem operasi yang dipakai, pada contoh di bawah adalah sistem operasi Windows 7 dengan jenis 64 Bit. Jadi harus kita unduh yang versi 64 Bit



4. Lakukan instalasi secara biasa, kemudian jalankan, ketika keluar jendela dialog End User License Agreement, beri tanda centang pada opsi **i have read and understand the terms .....**. Kemudian tekan tombol **I Accept**



5. Pada kotak dialog Create a RapidMiner account, pilih opsi **Educational**



6. Kemudian isikan

- Your First Name
- Your lastname
- Your Email address
- Password

Kemudian tekan tombol **Create my Account** (komputer Anda **harus terkoneksi ke jaringan internet**)

## Create a RapidMiner account

You'll use your RapidMiner Account to access:

- the Community forum
- the Extensions Marketplace
- free cloud storage
- product news and updates
- product license information

Your last name  
Wibowo M.Kom, M.Si, M.M

Your email address  
as.kuliah.karyawan.semarang@gmail.com

Password  
\*\*\*\*\*

Remember my password

[Create my Account!](#)

[I already have an account or license key](#)

7. Buka alamat email anda, lakukan konfirmasi dengan menekan tautan yang ada ada email

Verify Your Email Inbox X Related Google+ Page

RapidMiner <noreply@rapidminer.com> to me 1:40 PM (0 minutes ago) ☆ ↶ ↷

 rapidminer

Hello and welcome to the RapidMiner community.

Please [confirm your email address](#) to get started. If you have any issues with your account, don't hesitate to [contact us](#).

Thanks and happy mining!

– the RapidMiner Team

8. Selanjutnya sistem akan membuka browser secara otomatis, isikan email dan password saat registrasi pertama kali

Sign in to your account:

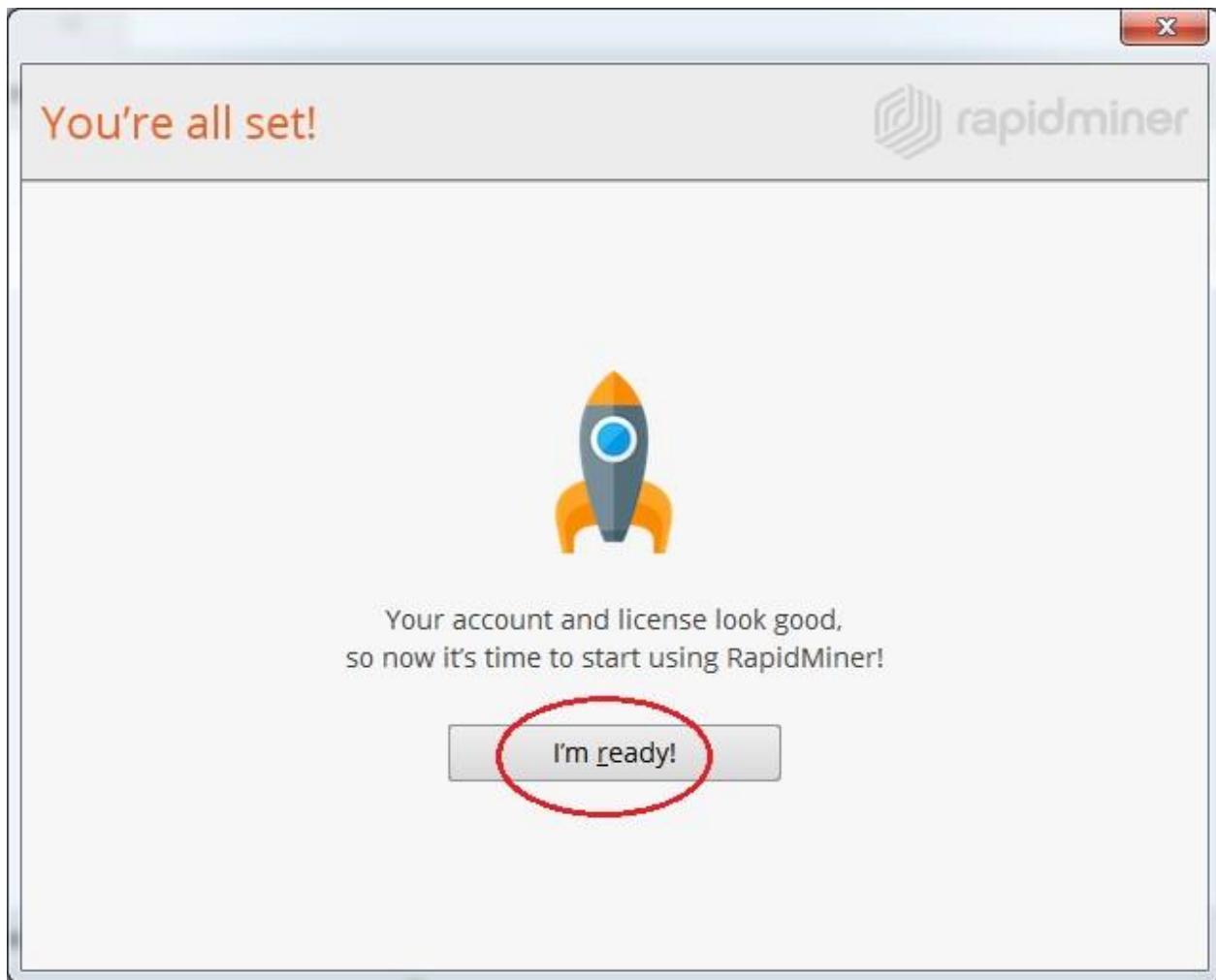
What's your email?

What's your password? [Forgot your password?](#)

[Sign in](#)

Don't have an account yet? [Register.](#)

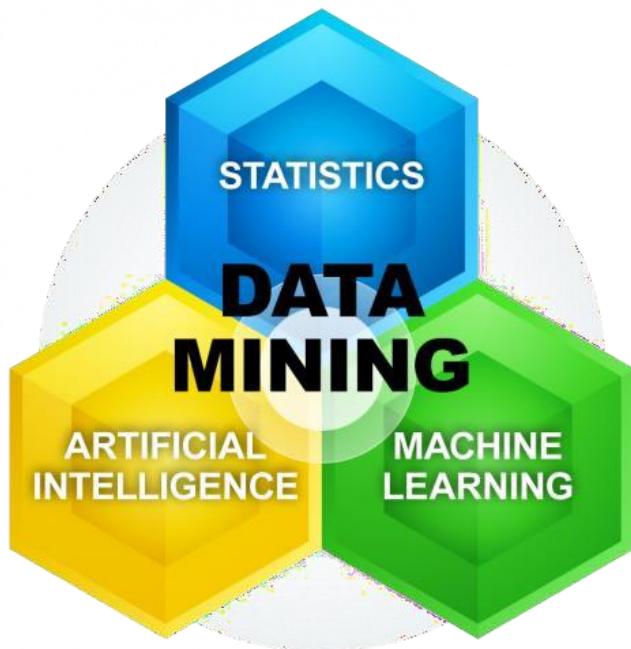
9. Kembali ke perangkat lunak Rapidminer, tekan tombol I'm Ready



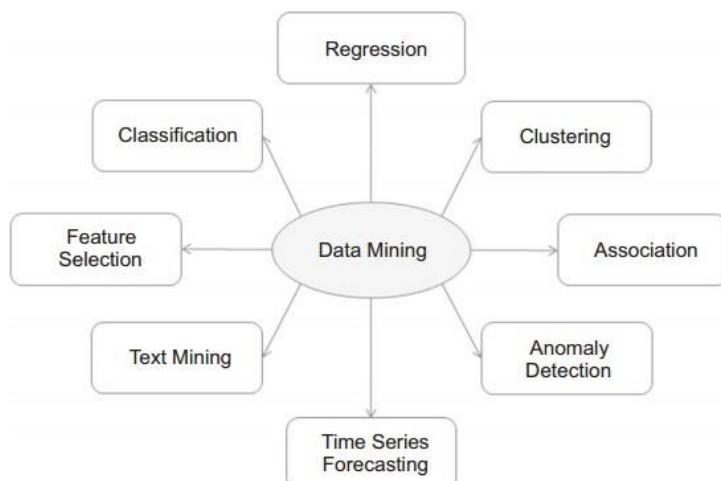
## 10. Akan keluar tampilan lingkungan terpadu(IDE) dari Rapidminer

A screenshot of the RapidMiner Studio Free 7.5.001 interface. The top menu bar includes File, Edit, Process, View, Connections, Cloud, Settings, and Extensions. The main workspace shows a process flow with an operator named "Retrieve Titanic Training". To the left, there is a "Tutorials" panel with a "Welcome to RapidMiner" section and an "EXPLANATION" section. The "Operators and Processes" section shows "1/5 Welcome to RapidMiner.". The "Repository" panel lists "Samples", "DB", "Local Repository (Zela)", and "Cloud Repository (disabled)". The "Process" panel shows the current process structure. The "Parameters" panel contains parameters for the "Process" operator, such as "logverbosity" set to "init" and "logfile". The "Help" panel provides information about the "Process" operator, including its synopsis and description. The bottom of the screen has a status bar with the text "Checking account activation".

## Tutorial Belajar Rapidminer (Bagian 2) - Pengantar Data Mining



Penambangan data (bahasa Inggris: *data mining*) adalah proses pencarian pengetahuan dari data dalam jumlah besar, menggunakan pengenalan pola, statistik dan teknik matematika. Suatu pola dikatakan menarik apabila pola tersebut tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. Penambangan data memiliki beberapa nama alternatif, meskipun definisi eksaknya berbeda, seperti KDD (knowledge discovery in database), analisis pola, arkeologi data, pemanenan informasi, dan intelegensi bisnis. Penggalian data diperlukan saat data yang tersedia terlalu banyak (misalnya data yang diperoleh dari sistem basis data perusahaan, e-commerce, data saham, dan data bioinformatika), tetapi tidak tahu pola apa yang bisa didapatkan.



Masalah data mining juga bisa dikelompokkan menjadi tugas :

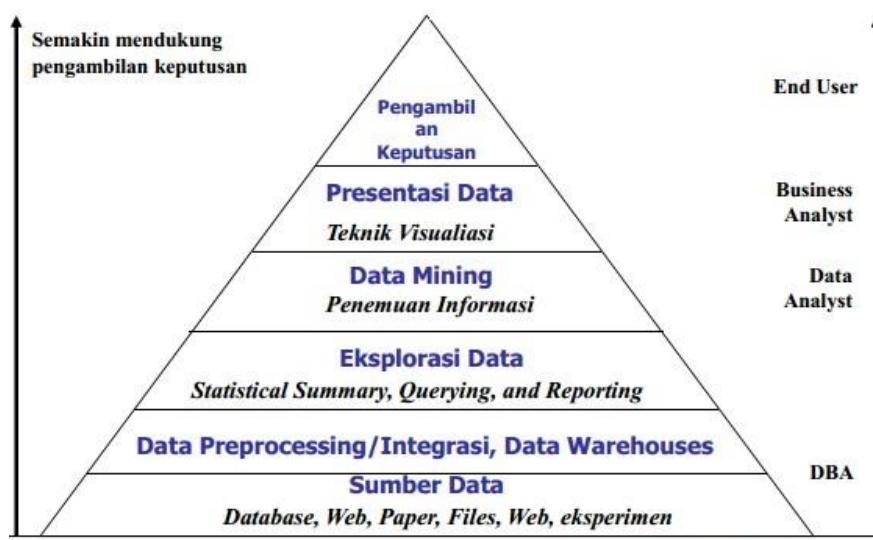
- Klasifikasi (*Classification*)
- Klusterisasi/Penggolongan (*Clustering*)
- Regresi (*Regression*)
- Analisis asosiasi (*Association*)
- deteksi anomali (*Anomaly Detection*)
- Peramalan deret waktu (*Time Series Forecasting*)
- penambangan teks (*Text Mining*)
- Pemilihan Fitur (*Feature Selection*)

Data mining memiliki beberapa nama alternatif yaitu :

- *Business Intelligence*
- *Knowledge discovery (mining) in databases (KDD)*
- *Knowledge extraction*
- *data/pattern analysis*
- *data archeology*
- *data dredging*
- *Information harvesting*

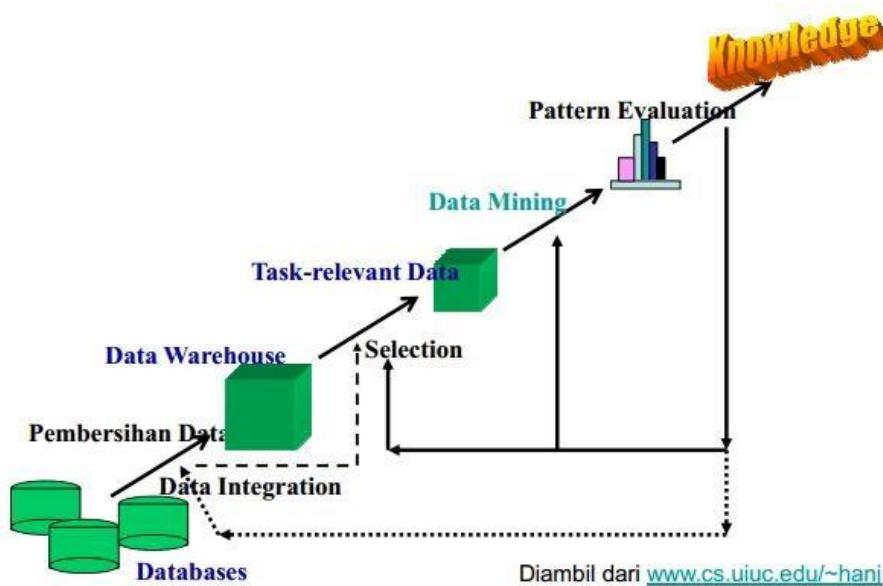
Keuntungan dari penerapan data mining (pengambilan keputusan) adalah :

- Perusahaan fokus ke informasi yg berharga di datawarehouse/databasenya.
- Meramalkan masa depan sehingga perusahaan dapat mempersiapkan diri



Diambil dari [www.cs.uiuc.edu/~hanj](http://www.cs.uiuc.edu/~hanj)

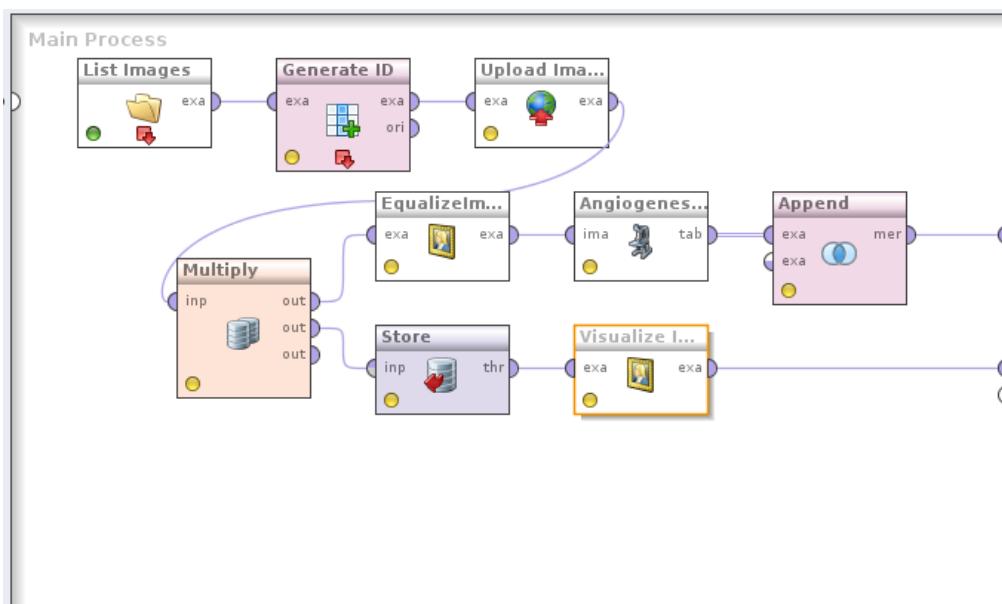
Proses dari data mining :



Setelah membaca sedikit pengantar tadi dapat diambil kesimpulan bahwa penerapan data mining berfungsi untuk mengoptimalkan kinerja lembaga atau perusahaan sehingga mampu beroperasi lebih baik dan lebih siap dalam persaingan.

Pada tutorial kali ini kita akan belajar dasar cara menjalankan proses serta menambahkan operator dalam data mining menggunakan Rapidminer Studio. Berikut adalah beberapa istilah yang akan digunakan dalam Rapidminer Studio :

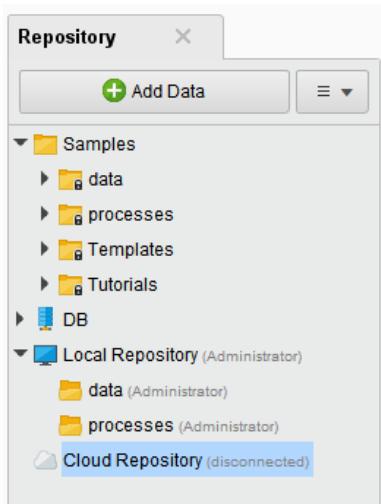
1. **Proses** adalah unit yang dapat dieksekusi yang berisi fungsionalitas yang dimilikinya. Pengguna menciptakan proses menggunakan operator dan menggabung secara bersama dengan cara apa pun yang diperlukan. Proses biasanya terdiri dari gabungan berbagai macam operator yang saling terhubung antara yang satu dengan yang lainnya.



**2. Operator** adalah satu blok fungsi yang disediakan RapidMiner Studio, yang bisa diatur dalam suatu proses dan terhubung menuju proses lainnya. Setiap operator memiliki parameter yang bisa dikonfigurasi menyesuaikan persyaratan spesifik dari proses. Contoh di bawah adalah tampilan operator dari **Decision Tree**.

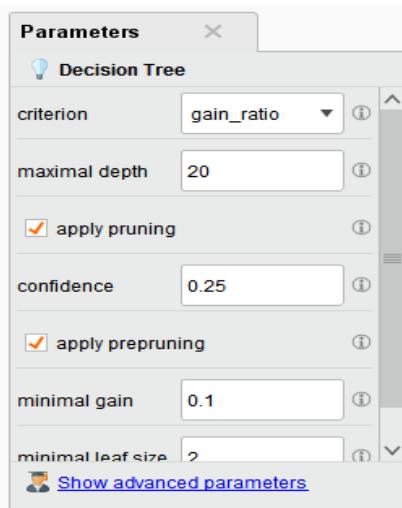


**3. Repository** adalah lokasi dimana proses, data, model, dan file dapat disimpan dan dibaca, baik dari RapidMiner Studio atau dari sebuah proses.



**4. Makro** adalah variabel global yang dapat diatur dan digunakan oleh sebagian besar operator dengan tujuan untuk memodifikasi perilaku operator.

**5. Role** adalah atribut mengidentifikasi bagaimana operator akan menggunakan atribut tersebut. Jenis standar lainnya dikenal sebagai Atribut khusus dan ini termasuk *label*, *id*, *cluster*, *prediction*, dan *outlier*. Role juga memungkinkan untuk mengatur peran atribut yang ada agar diabaikan oleh kebanyakan operator (misalnya karena ada pengecualian).

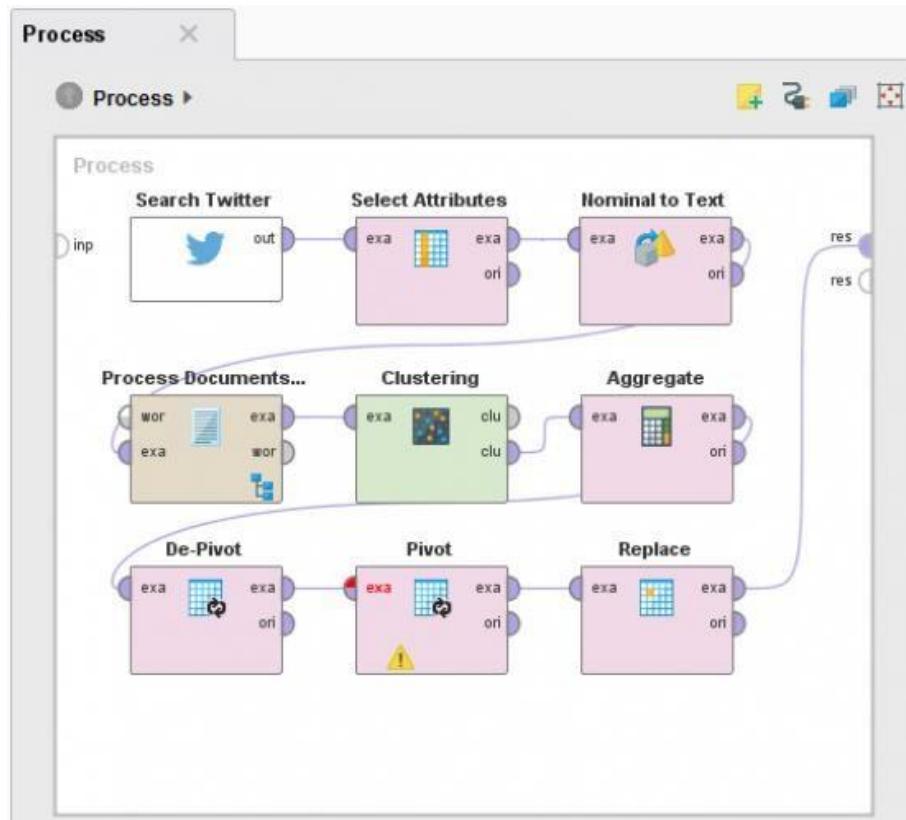


6. **Type** adalah jenis dari sebuah atribut. Type bisa berjenis nyata (*real*) contoh : 0.167, 0.25, bulat (*integer*) contoh : 349 dan 567, date\_time, Nominal (baik polinominal maupun binominal), atau teks.

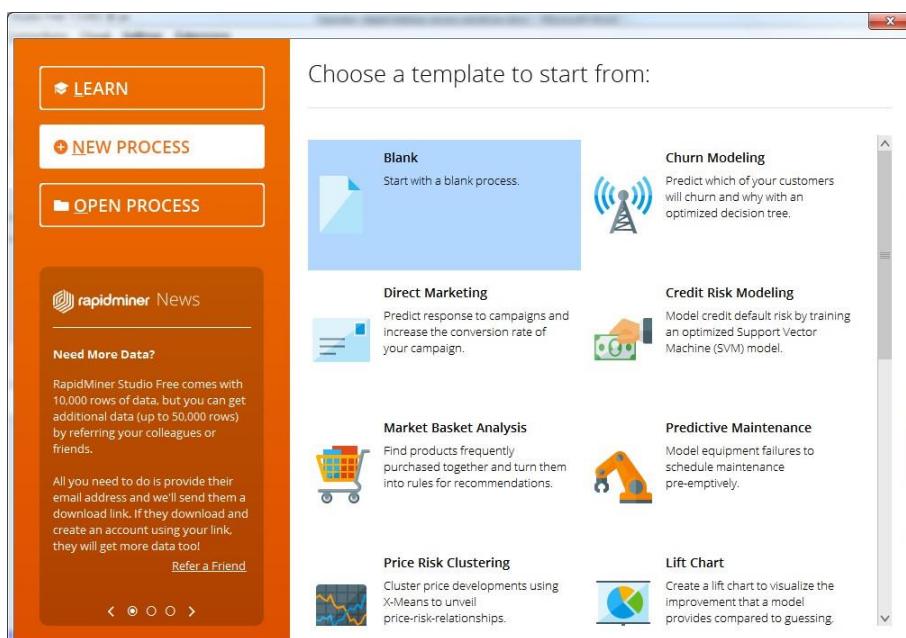
The screenshot shows the KNIME interface with the 'Data' tab selected in the left sidebar. The main area displays a table titled 'ExampleSet (Retrieve Titanic Training)'. The table has columns for 'Name', 'Type', 'Missing', and summary statistics ('Least', 'Most', 'Min', 'Max'). A red box highlights the first six rows of the table, which correspond to the attributes: Survived, Age, Passenger Class, Sex, No of Siblings or Spouses on Board, and No of Parents or Children on Board. The 'Survived' row is labeled 'Label' and has a green background. The other five rows have white backgrounds.

Name	Type	Missing	Least	Most
Label Survived	Binominal	0	Yes (349)	No (567)
Age	Real	0	0.167	80
Passenger Class	Polynomial	0	Second (184)	Third (491)
Sex	Binominal	0	Female (322)	Male (594)
No of Siblings or Spouses on Board	Integer	0	0	8
No of Parents or Children on Board	Integer	0	0	9
Passenger Fare	Numeric	0	0	512.329

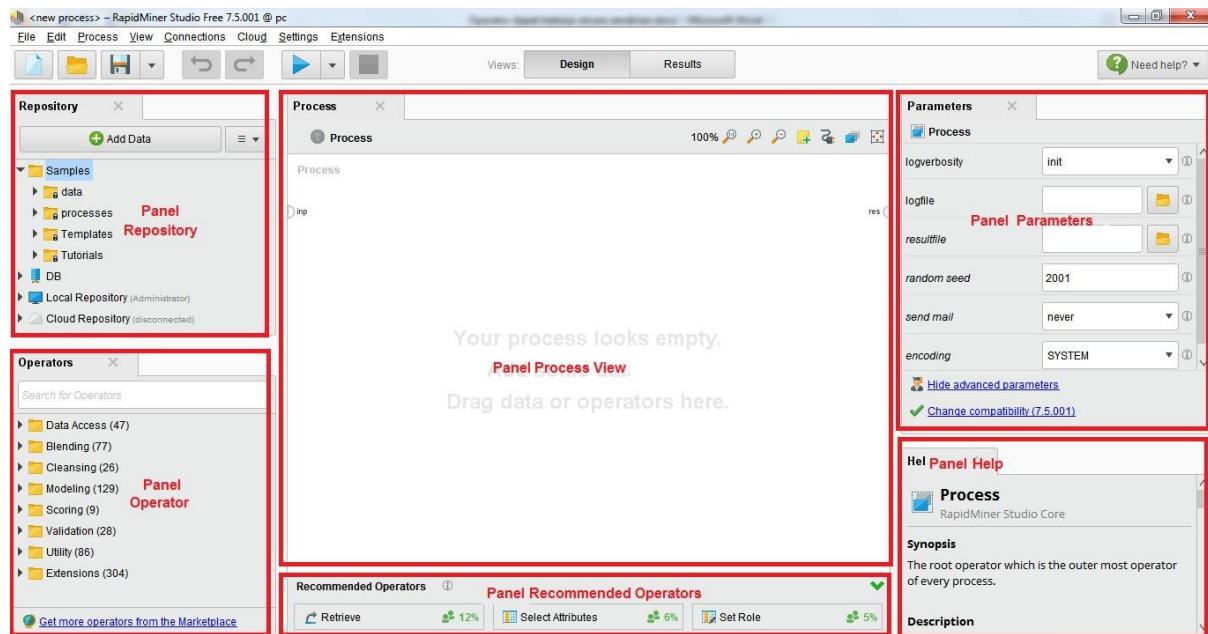
# Tutorial Belajar Rapidminer (Bagian 3) - Operator dan Proses



Rapidminer Studio memiliki tampilan antarmuka pengguna yang mudah untuk dioperasikan. Ketika dijalankan, Rapidminer Studio menampilkan jendela dialog dengan tiga opsi yaitu : *learn* yang berisi tutorial dasar penggunaan Rapidminer Studio, *new process* untuk menjalankan proses yang baru serta *open process* untuk membuka proses yang sebelumnya pernah dibuat. Untuk dasar pembelajaran kali ini maka pilih opsi *new process*, kemudian pilih **Blank**.



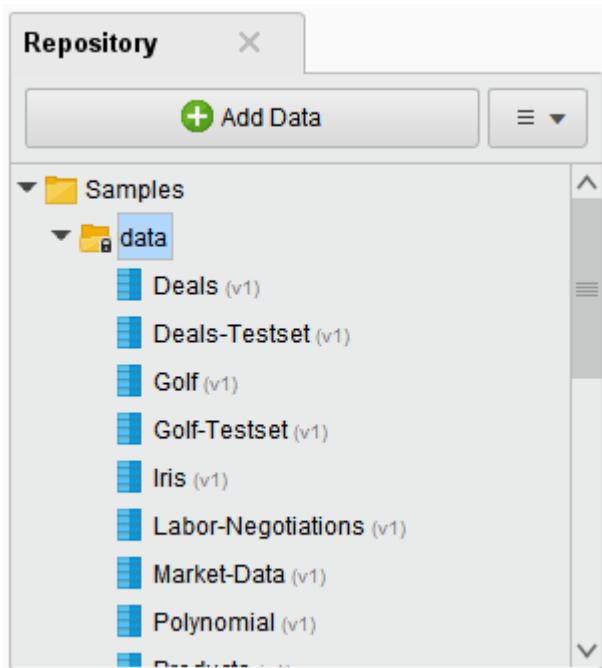
Setelah itu akan ditampilkan design perspektif



Berikut penjelasan antarmuka di Rapidminer Studio versi 7.5

#### Panel Repositori :

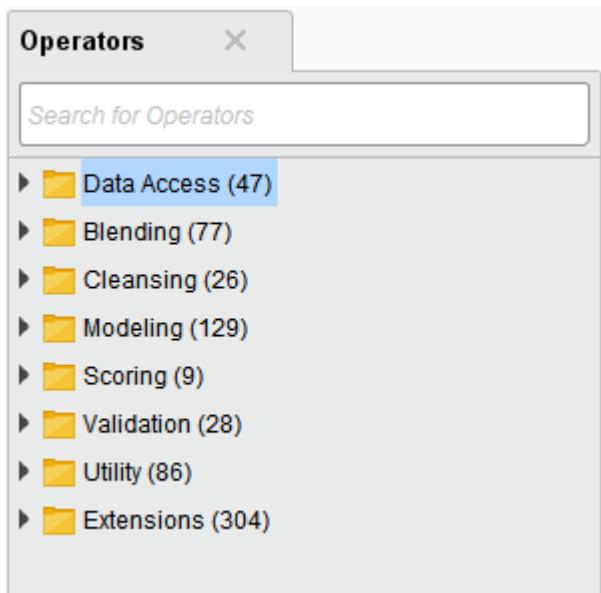
Repositori digunakan untuk mengelola dan menata proses analisis. Anda menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan metadata.



#### Panel Operator :

Berisi operator-operator yang dapat digunakan pada proses analisis.

Pada Operator View ini terdapat beberapa kelompok operator sebagai berikut:



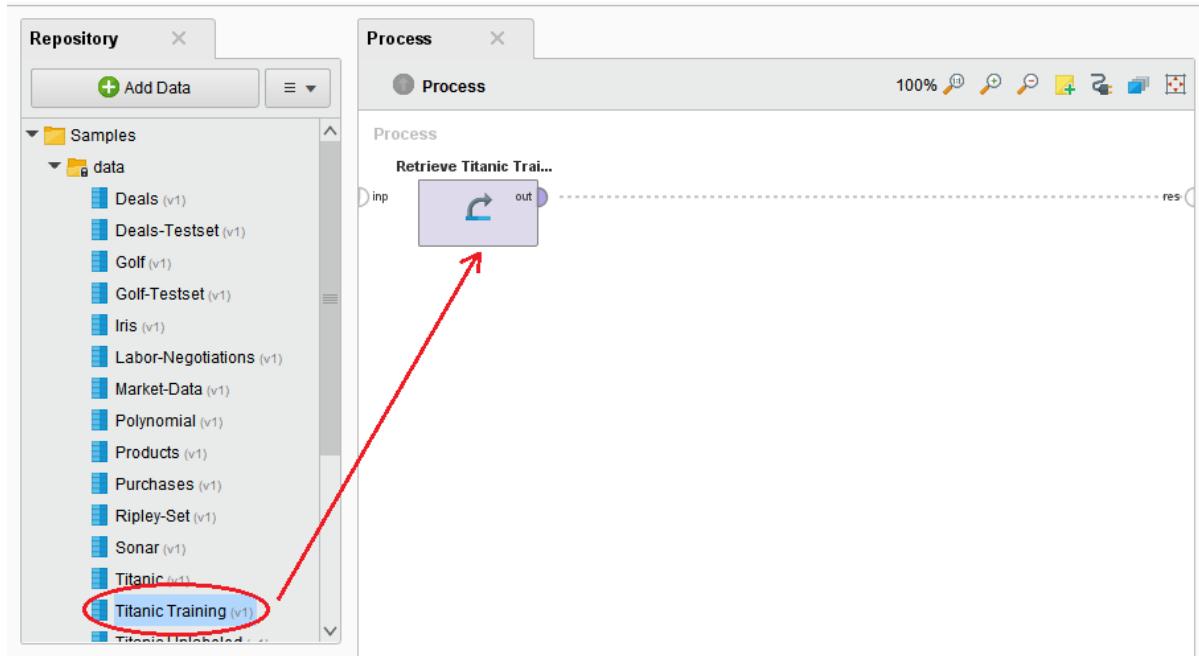
- **Data Access :**  
kelompok ini terdiri dari semua operator yang berguna untuk membaca dan menulis file dari repositori lokal atau cloud, mengakses file dijejaring social seperti twitter . Kelompok import digunakan untuk membaca data dan objek dari format tertentu seperti file, database, dan lain-lain. Kelompok export digunakan untuk menulis data dan objek menjadi format tertentu.
- **Blending :**  
kelompok ini terdiri dari semua operator yang berguna untuk transformasi data pada table yang berkaitan dengan atribut (membuat, menghilangkan, menyalin, agregasi, pengubahan bobot, join, union, interseksi dan lainnya).
- **Cleansing**  
kelompok ini terdiri dari operator untuk melakukan normalisasi/denormalisasi data, memperbaiki data yang hilang atau terduplikasi, pengubahan dimensi serta operasi diskrit (binning, size, dimensi)
- **Utility :**  
kelompok ini terdiri dari Operator bantuan, seperti operator macros, logging, subproses, dan operator perulangan dan percabangan yang dapat mengatur aliran proses.
- **Modeling :**  
kolompok ini berisi proses data mining untuk menerapkan model yang dihasilkan menjadi set data yang baru untuk keperluan prediksi, segmentasi, asosiasi, korelasi, kesamaan, optimasi serta pembobotan
- **Scoring :**  
kelompok ini terdiri dari semua operator yang berguna untuk penerapan model, mencari, membuat dan menerapkan ambang batas, menghasilkan prediksi serta pemeringkatan.
- **Validation :**  
kelompok ini berisi operator yang dapat digunakan untuk menghitung kinerja pemodelan serta untuk data baru.

- **Extension :**

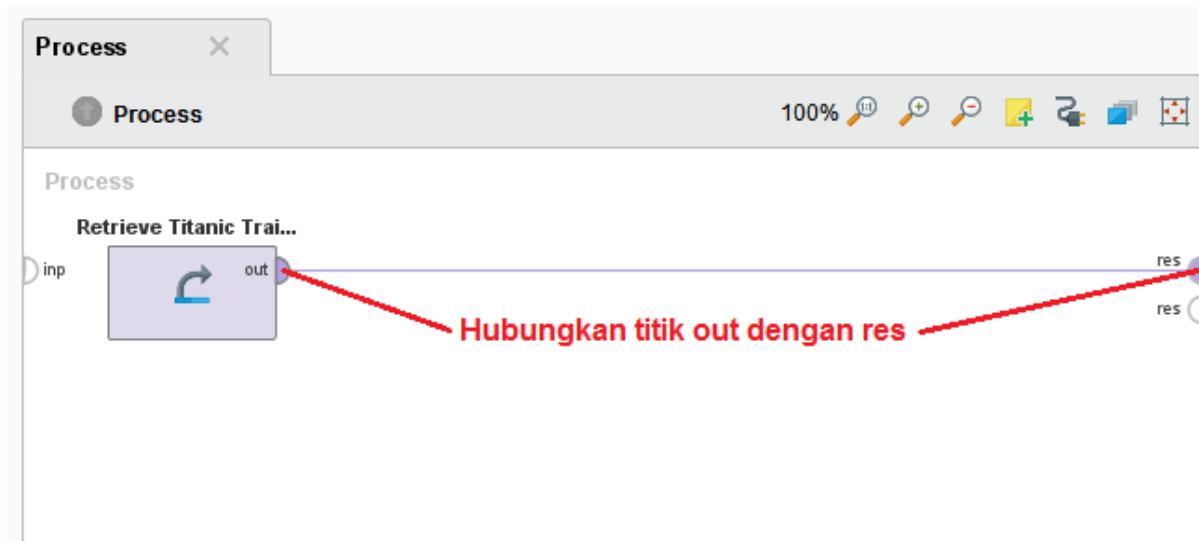
Berisi operator yang dikembangkan oleh pihak ketiga untuk keperluan khusus seperti pemrosesan teks, akses data pasar keuangan, transformasi deret waktu, transformasi data menggunakan integral, diferensial, moving average, eksponensial

Latihan cara menjalankan proses sederhana:

1. Cari panel **Repository** yang terletak pada sebelah kiri layar
2. Buka folder **samples**, kemudian buka **data**
3. Drag dan drop dataset **Titanic Training** pada panel process



4. Hubungkan port ("out") dari operator **Retrieve Titanic Training** dengan port ("res") pada panel proses yang terletak disebelah kanan dengan cara melakukan klik kiri pada port ("out") dilanjutkan dengan melakukan klik kiri pada port ("res").



5. Tekan tombol **Run** untuk menjalankan proses.



6. Pada tabulasi **Data**, Anda bisa melihat data mentah tentang penumpang Titanic seperti ukuran keluarga atau usianya.

Result History    ExampleSet (Retrieve Titanic Training)

Filter (916 / 916 examples): all

Row No.	Survived	Age	Passenger ...	Sex	No of Sibling...	No of Parent...	Passenger F...
1	Yes	29	First	Female	0	0	211.338
2	No	2	First	Female	1	2	151.550
3	No	30	First	Male	1	2	151.550
4	No	25	First	Female	1	2	151.550
5	Yes	48	First	Male	0	0	26.550
6	Yes	63	First	Female	1	0	77.958
7	No	39	First	Male	0	0	0
8	Yes	18	First	Female	1	0	227.525
9	Yes	26	First	Female	0	0	78.850
10	Yes	80	First	Male	0	0	30
11	No	29.881	First	Male	0	0	25.925
12	No	24	First	Male	0	1	247.521
13	Yes	50	First	Female	0	1	247.521
14	Yes	32	First	Female	0	0	76.292
15	No	36	First	Male	0	0	75.242
16	Yes	37	First	Male	1	1	52.554
17	Yes	26	First	Male	0	0	30

7. Pada tabulasi **Statistics**, kita mendapatkan ringkasan statistik, yang memberikan informasi yang berguna yaitu terdapat 349 penumpang yang selamat dari kecelakaan Titanic.

Result History    ExampleSet (Retrieve Titanic Training)

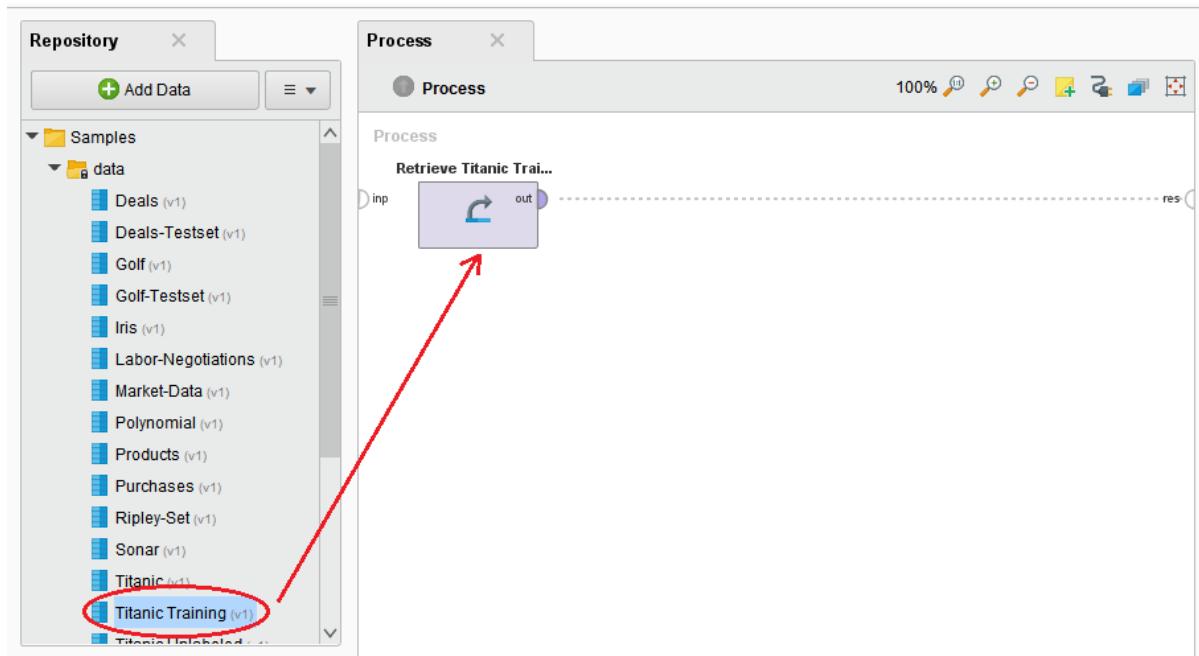
Filter (7 / 7 attributes): Search for Attributes

Name	Type	Missing	Statistics	Values
Survived	Binominal	0	Least Yes (349) Most No (567)	No (567), Yes (349)
Age	Real	0	Min 0.167 Max 80 Average 29.932	
Passenger Class	Polynomial	0	Least Second (184) Most Third (491)	Third (491), First (241), ...[1 more]
Sex	Binominal	0	Least Female (322) Most Male (594)	Male (594), Female (322)
No of Siblings or Spouses on Board	Integer	0	Min 0 Max 8 Average 0.514	
No of Parents or Children on Board	Integer	0	Min 0 Max 9 Average 0.386	
Passenger Fare	Numeric	0	Min 0 Max 512.329 Average 33.564	

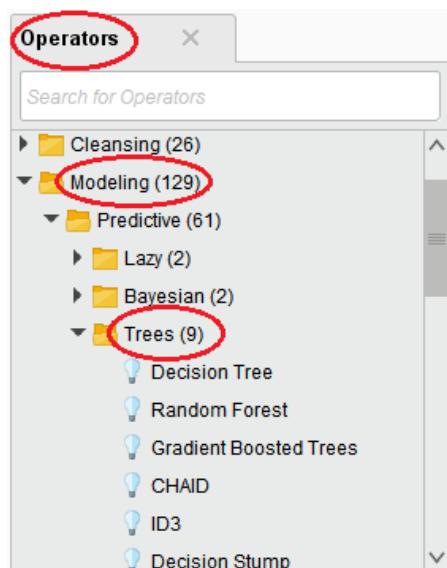
## Tutorial Belajar Rapidminer (Bagian 4) - Mengenal Permodelan Sederhana

Pada tutorial sebelumnya, kita mengimpor sekumpulan data (*dataset*) dan menjalankan sebuah proses. Kita menemukan bahwa 349 penumpang selamat dari kecelakaan Titanic. Pada bagian ini kita akan membangun sebuah proses untuk melihat apakah orang-orang yang selamat memiliki kesamaan atau tidak. Berikut langkah membuat permodelan sederhana menggunakan **Decission Tree** (Pohon Keputusan) :

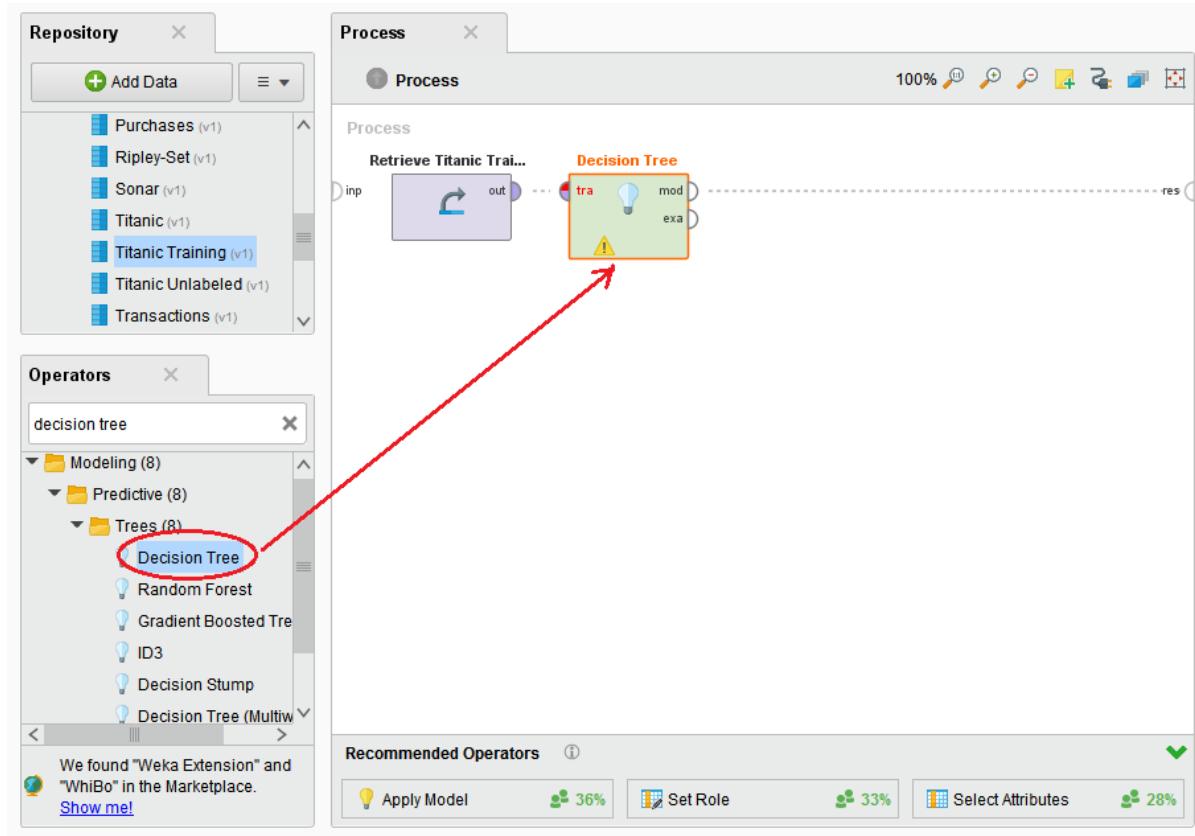
1. Ambil data Pelatihan Titanic dengan cara melakukan *drag and drop* dataset *Titanic Training* pada panel process.



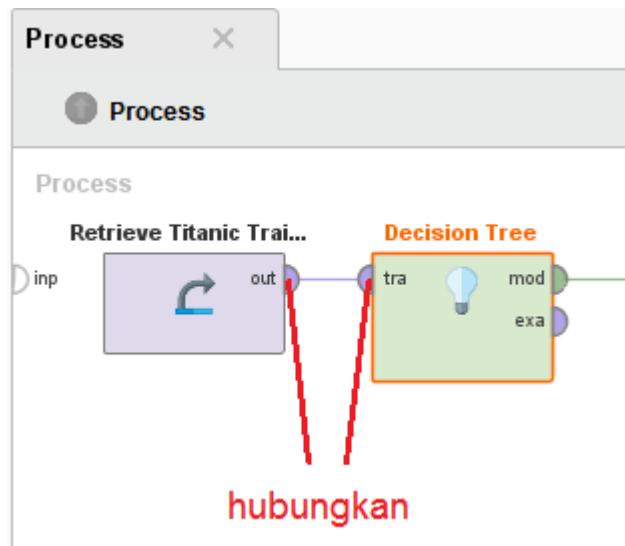
2. Kita akan membuat pohon keputusan (*decision tree*). Pohon Keputusan adalah teknik pemodelan statistik populer, yang digunakan untuk menemukan pola tersembunyi dalam data.
3. Pada bagian panel **Operators** di sebelah kiri, buka folder *Modeling*, *Predictive*, dan akhirnya *Trees*.



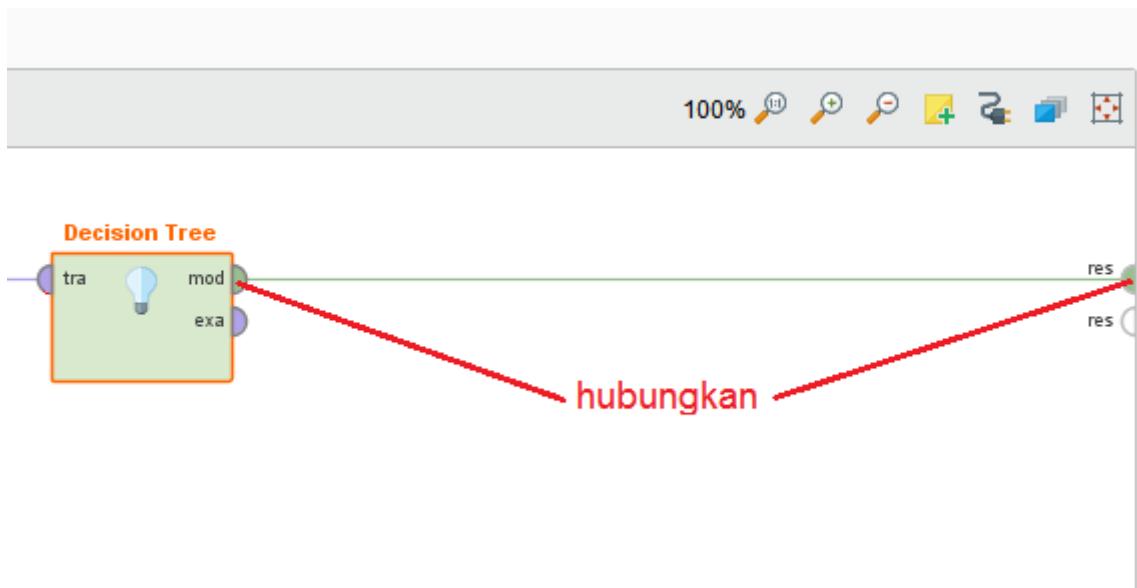
4. Seret pada operator *Decision Tree*, kemudian letakkan dengan posisi setelah *Retrieve Titanic Training*.



5. Hubungkan port **output ("out")** dari *Retrieve Titanic Training* dengan port **input ("tra"** untuk "training") *Decision Tree*. Anda dapat mengklik pada port, atau Anda dapat menarik port output ("out") menuju port input ("tra")



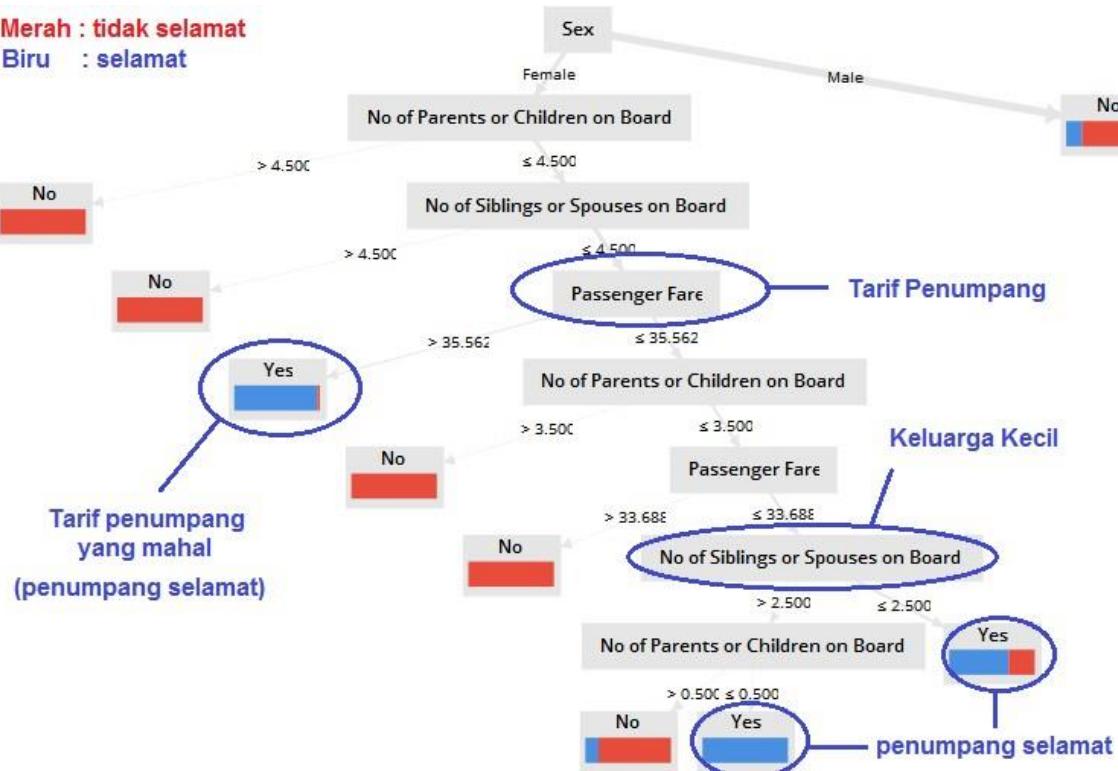
6. Hubungkan output pertama ("mod") dari *Decision Tree* dengan port hasil ("res") di sisi kanan panel Proses.



7. Tekan tombol **Run** pada toolbar untuk menjalankan proses.



8. Dari permodelan menggunakan pohon keputusan yang telah kita buat, dapat dengan jelas terlihat bahwa faktor keselamatan nyawa sama sekali bukan suatu kebetulan. Dari analisa didapatkan kesimpulan bahwa penumpang wanita dengan **keluarga kecil**, benar-benar beruntung - setidaknya jika mereka memegang **tiket yang mahal** (tarif penumpang/*passenger fare*)



# Tutorial Belajar Rapidminer (Bagian 5) - Mengimpor Data ke Dalam RapidMiner.

Mengimpor data Anda ke dalam RapidMiner seringkali merupakan tugas pertama yang harus kita lakukan dalam rangka proses menyelesaikan analisis. Dalam tutorial ini, kita akan belajar cara mengimpor file ke dalam penyimpanan pusat RapidMiner, yang disebut Repository. Panel Repository, secara baku terletak pada sudut kiri atas, di mana merupakan tempat menyimpan semua data, proses, dan hasil. Kita harus selalu mengimpor data ke dalam repositori, terutama bila berasal dari file jenis excell seperti XLS atau CSV. Repositori akan menyederhanakan desain proses analisis pada RapidMiner yang menggambarkan metadata beserta datanya. Sebagai contoh kita akan menggunakan data yang menggambarkan kecelakaan Titanic. Pada tutorial ini, file yang akan diimpor merupakan jenis dari format Excel.

Berikut langkah mengimpor data ke dalam rapidminer :

1. Unduh contoh file data Titanic dari tautan berikut :

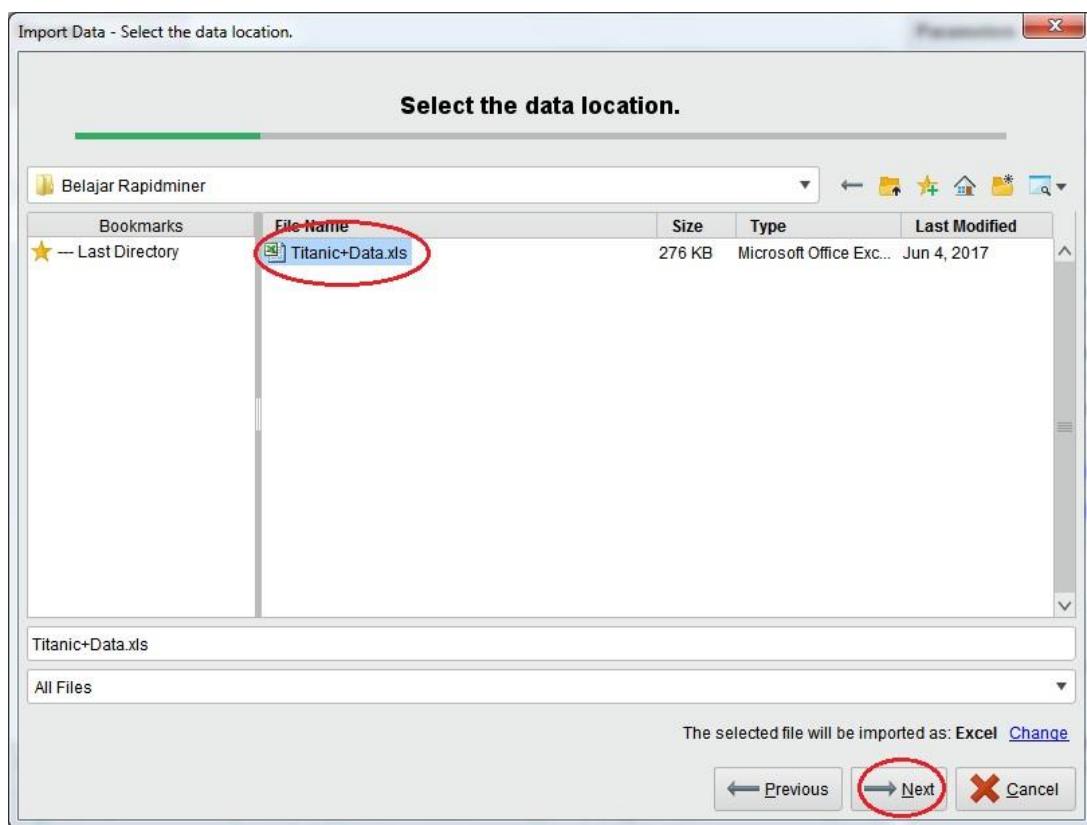
<https://s3.amazonaws.com/static.rapidminer.com/demo/titanic/Titanic+Data.xls>

Kemudian simpan.

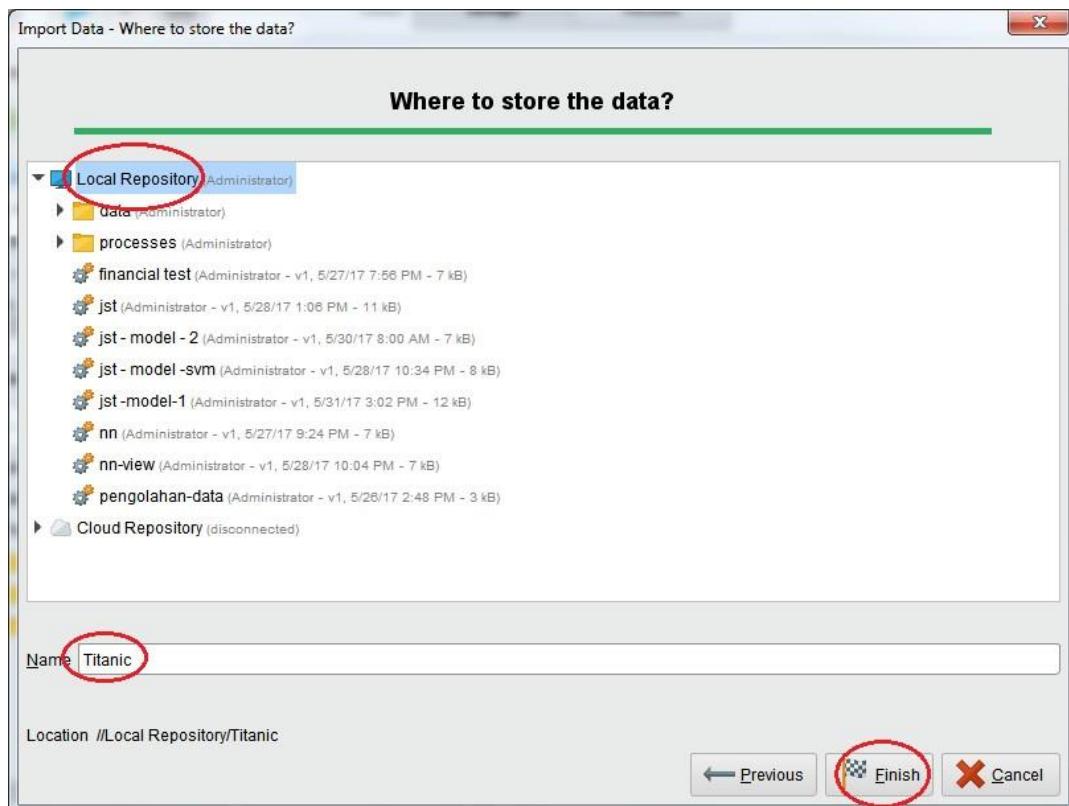
2. Untuk mengimpor data yang diunduh ke RapidMiner, klik **Add Data** pada panel Repository dan pilih **My Computer**



3. Arahkan ke letak file **Titanic+Data.xls** berada, pilih file **Titanic+Data.xls** kemudian klik **Next**.



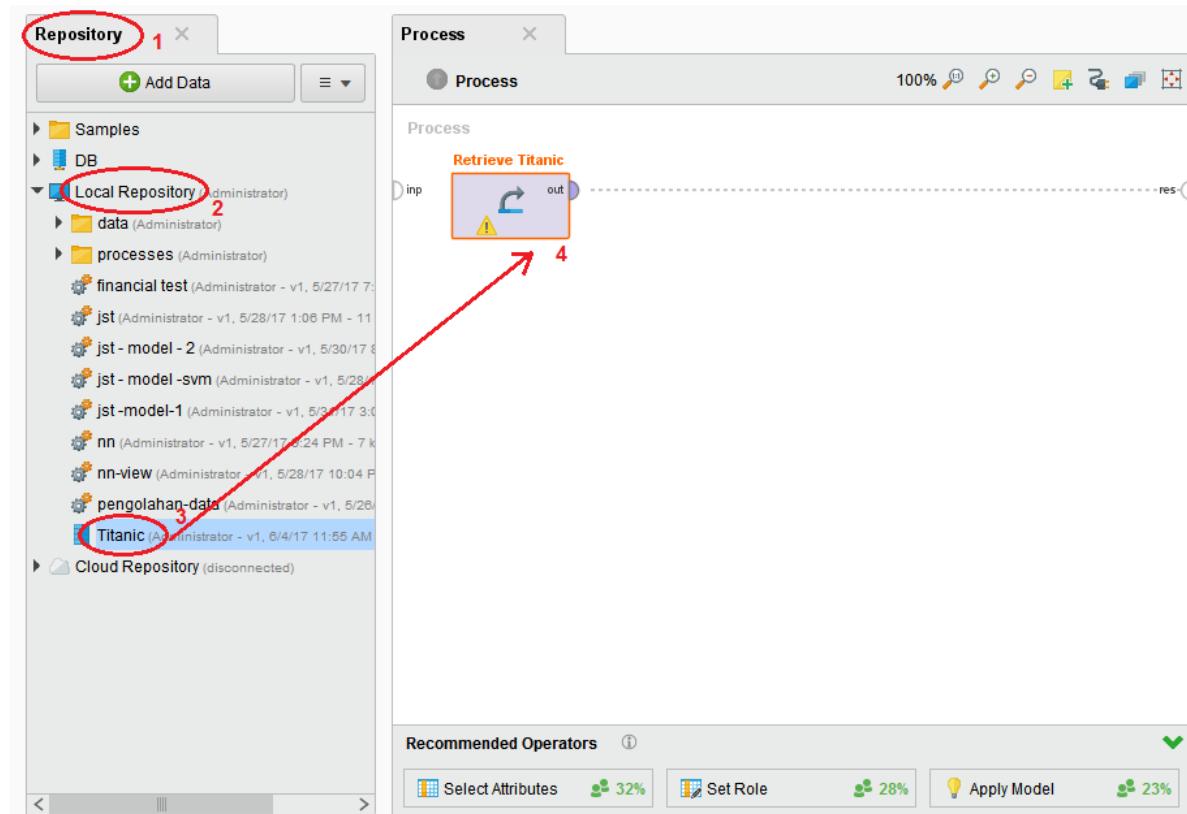
4. Klik **Next** beberapa kali sampai muncul jendela **Where to store the data**. Saat Anda melengkapi impor, simpan data sebagai **Titanic** di **Repositori Lokal** Anda. Beri nama **Titanic** kemudian tekan tombol **Finish**.



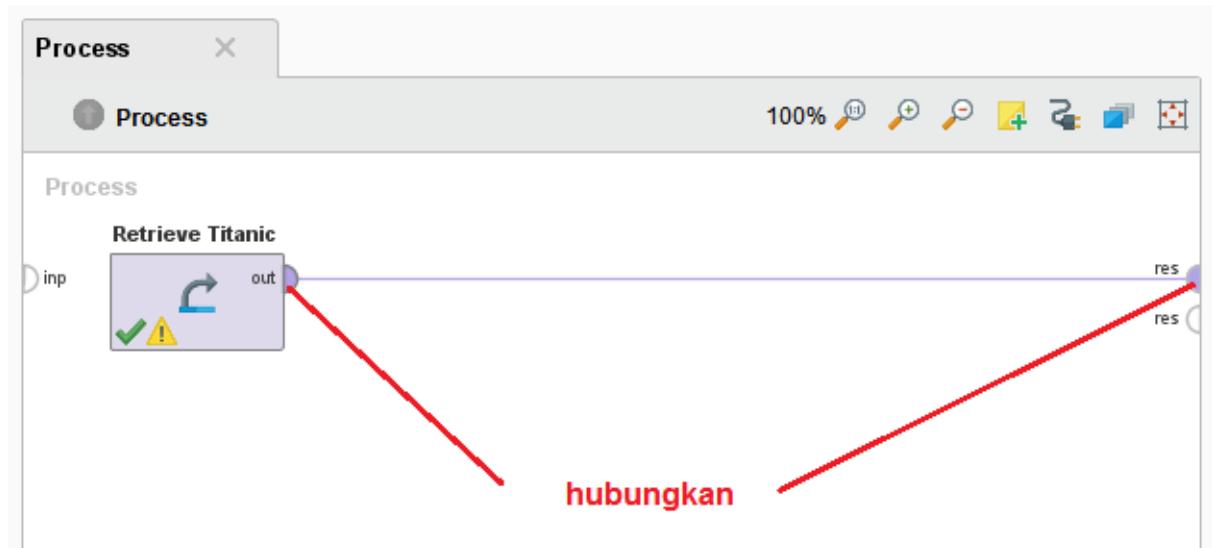
5. Akan ditampilkan kolom dan baris hasil impor dataset Titanic yang berhasil ditambahkan pada repositori lokal. Tekan tombol **Design** untuk menampilkan jendela panel **Process**.

The screenshot shows the KNIME interface with the 'Design' tab highlighted. The main area displays the 'ExampleSet (//Local Repository/Titanic)' with 1309 examples. The table includes columns: Row No., Survived, Passenger ..., Name, Sex, Age, No of Sibling..., No of Parent..., Ticket Numb..., and Passenger F... . The data preview shows various entries from the Titanic dataset.

6. Arahkan panel **Repository** (1), pilih **Local Repository** (2), pilih **Titanic** (3), kemudian drag and drop ke panel **Process** (4).



7. Hubungkan port **out** ("output") dari **Retrieve Titanic** dengan port **hasil/results** ("res") pada panel **Process**.



8. Tekan tombol **Run** pada toolbar untuk menjalankan proses



9. Sekarang kita dapat melihat hasil tampilan dari data yang berhasil ditambahkan ke repositori lokal. Pada baris ke 16 kita bisa melihat terdapat data yang hilang, yaitu diwakili dengan isian berupa **?** (tanda tanya).

ExampleSet (/Local Repository/Titanic)									
Data	ExampleSet (1309 examples, 0 special attributes, 12 regular attributes)								
	Row No.	Survived	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...
1	Yes	First	Allen, Miss. E...	Female	29	0	0	24160	211.338
2	Yes	First	Allison, Mast...	Male	0.917	1	2	113781	151.550
3	No	First	Allison, Miss. ...	Female	2	1	2	113781	151.550
4	No	First	Allison, Mr. H...	Male	30	1	2	113781	151.550
5	No	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550
6	Yes	First	Anderson, Mr....	Male	48	0	0	19952	26.550
7	Yes	First	Andrews, Mis...	Female	63	1	0	13502	77.958
8	No	First	Andrews, Mr. ...	Male	39	0	0	112050	0
9	Yes	First	Appleton, Mrs...	Female	53	2	0	11769	51.479
10	No	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504
11	No	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525
12	Yes	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525
13	Yes	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300
14	Yes	First	Barber, Miss. ...	Female	26	0	0	19877	78.850
15	Yes	First	Barkworth, Mr...	Male	80	0	0	27042	30
16	No	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925
17	No	First	Baxter, Mr. Ou...	Male	24	0	1	PC 17559	247.524

10. Untuk mencari data yang hilang (tidak lengkap) pada kolom **Age** (umur), klik kolom **Age** sampai muncul deretan tanda tanya

Row No.	Survived	Passenger ...	Name	Sex	Klik Age ↑	No of Sibling...	No of Parent...	Ticket Numb...	Passenger F...
16	No	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925
38	Yes	First	Bradley, Mr. G...	Male	?	0	0	111427	26.550
41	No	First	Brewe, Dr. Art...	Male	?	0	0	112379	39.600
47	No	First	Cairns, Mr. Al...	Male	?	0	0	113798	31
60	Yes	First	Cassebeer, ...	Female	?	0	0	17770	27.721
70	Yes	First	Chibnall, Mrs....	Female	?	0	1	113505	55
71	No	First	Chisholm, Mr...	Male	?	0	0	112051	0
75	No	First	Clifford, Mr. G...	Male	?	0	0	110465	52
81	No	First	Crafton, Mr. J...	Male	?	0	0	113791	26.550
107	No	First	Farthing, Mr. ...	Male	?	0	0	PC 17483	221.779
108	Yes	First	Flegenheim, ...	Female	?	0	0	PC 17598	31.683
109	Yes	First	Fleming, Mis...	Female	?	0	0	17421	110.883
119	No	First	Franklin, Mr. ...	Male	?	0	0	113778	26.550
122	Yes	First	Frauenthal, M...	Female	?	1	0	PC 17611	133.650
126	No	First	Fry, Mr. Richard	Male	?	0	0	112058	0
135	Yes	First	Goldenberg, ...	Female	?	1	0	17453	89.104

Kolom yang lain yang dapat kita coba (berisi data yang hilang/tidak lengkap) adalah kolom **Cabin** dan **Life Boat**

Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Numb...	Passenger F...	Klik Cabin ↑	Port of Emb...	Life Boat
Artagaveytia, ...	Male	71	0	0	PC 17609	49.504	?	Cherbourg	?
Barber, Miss. ...	Female	26	0	0	19877	78.850	?	Southampton	6
Baumann, Mr...	Male	?	0	0	PC 17318	25.925	?	Southampton	?
Bidois, Miss. ...	Female	42	0	0	PC 17757	227.525	?	Cherbourg	4
Birnbaum, Mr...	Male	25	0	0	13905	26	?	Cherbourg	?
Bowen, Miss....	Female	45	0	0	PC 17608	262.375	?	Cherbourg	4
Bradley, Mr. G...	Male	?	0	0	111427	26.550	?	Southampton	9
Brewe, Dr. Art...	Male	?	0	0	112379	39.600	?	Cherbourg	?
Cairns, Mr. Al...	Male	?	0	0	113798	31	?	Southampton	?
Candee, Mrs....	Female	53	0	0	PC 17606	27.446	?	Cherbourg	6
Carrau, Mr. Fr...	Male	28	0	0	113059	47.100	?	Southampton	?
Carrau, Mr. J...	Male	17	0	0	113059	47.100	?	Southampton	?
Case, Mr. Ho...	Male	49	0	0	19924	26	?	Southampton	?
Cassebeer, ...	Female	?	0	0	17770	27.721	?	Cherbourg	5
Chisholm, Mr...	Male	?	0	0	112051	0	?	Southampton	?
Cleaver, Miss...	Female	22	0	0	113781	151.550	?	Southampton	11

11. Tab **Statistik** (1) menampilkan ringkasan data pada setiap kolom. Sebagai contoh pertanyaan yang pertama : berapa banyak orang yang melakukan perjalanan di kelas satu? Jawabannya adalah 323 orang. Cara mendapatkan jawabannya : klik pada tanda panah **Passenger Class** (2), akan didapatkan informasi penumpang kelas 1 pada values **First (323)**.

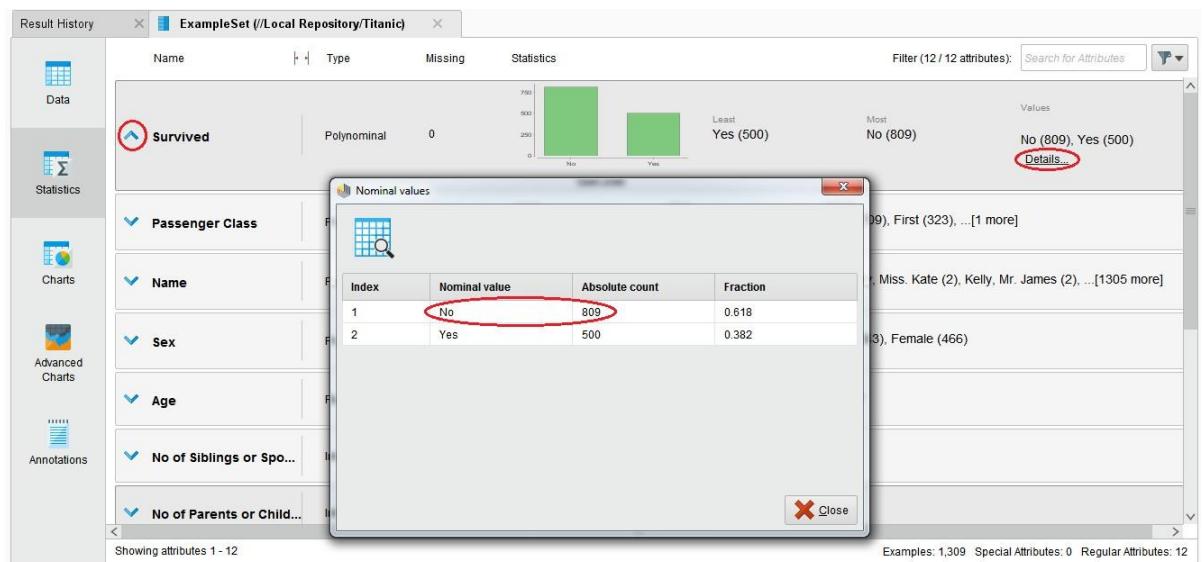
Result History | ExampleSet (/Local Repository/Titanic) | X

Data Statistics Charts Advanced Charts Annotations

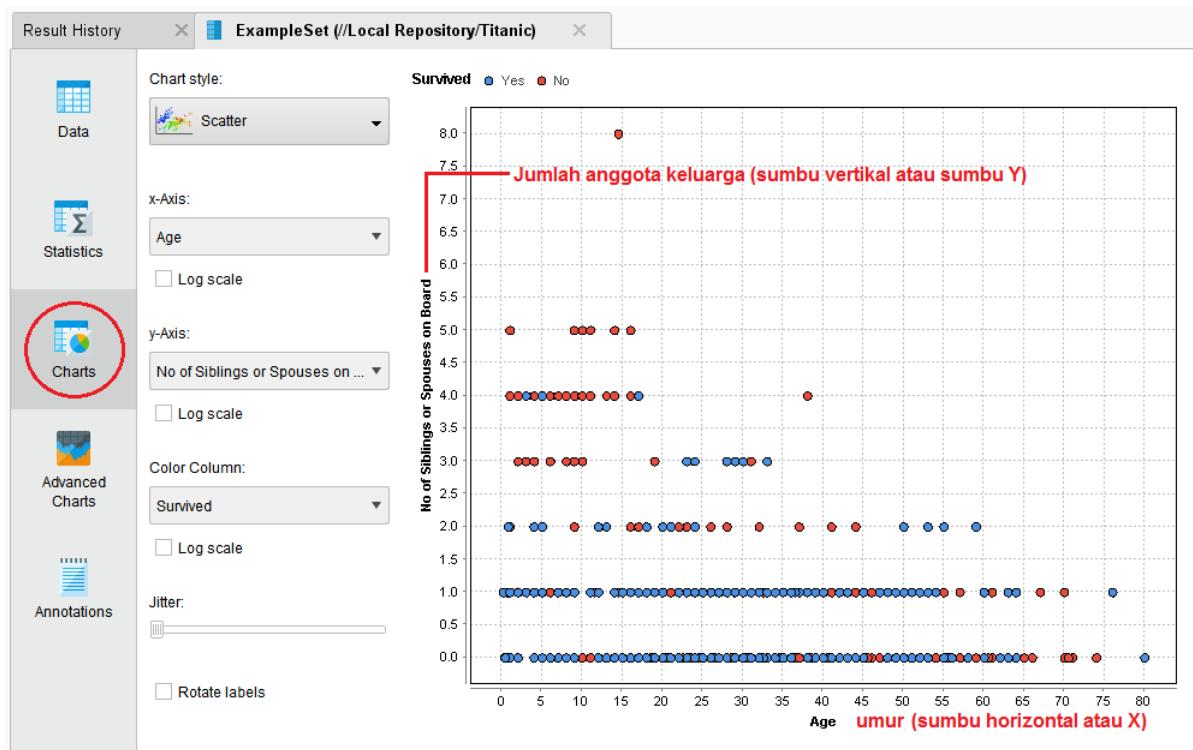
Showing attributes 1 - 12 Examples: 1,309 Special Attributes: 0 Regular Attributes: 12

Name	Type	Missing	Statistics	Filter (12 / 12 attributes): Search for Attributes
<b>Survived</b>	Polynomial	0	Least Yes (500)      Most No (809)      Values No (809), Yes (500)	
<b>Passenger Class</b>	Polynomial	0	Least Second (277)      Most Third (709)      Values Third (709), First (323), Second (277)      Details...	2
<b>Name</b>	Polynomial	0	Least van Melk [...] lemon (1)      Most Connolly, Miss. Kate (2)      Values Connolly, Miss. Kate (2), Kelly, Mr. James (2), ...[1305 more]	
<b>Sex</b>	Polynomial	0	Least Female (466)      Most Male (843)      Values Male (843), Female (466)	
<b>Age</b>	Real	263	Min 0.167      Max 80      Average 29.881	
<b>No of Siblings or Spouses on Board</b>	Integer	0	Min 0      Max 8      Average 0.499	
<b>No of Parents or Children on Board</b>	Integer	0	Min 0      Max 9      Average 0.385	

Sebagai contoh pertanyaan yang kedua : berapa banyak orang yang tidak selamat dari kecelakaan Titanic? Jawabannya adalah 809 orang. Cara mendapatkan jawabannya : klik pada tanda panah **Survived**, akan didapatkan informasi penumpang selamat (**Survived**) pada values Yes sebanyak 500 penumpang, serta yang tidak selamat pada value **No** sebanyak **809** penumpang. Klik tombol **Details** untuk melihat lebih rinci.



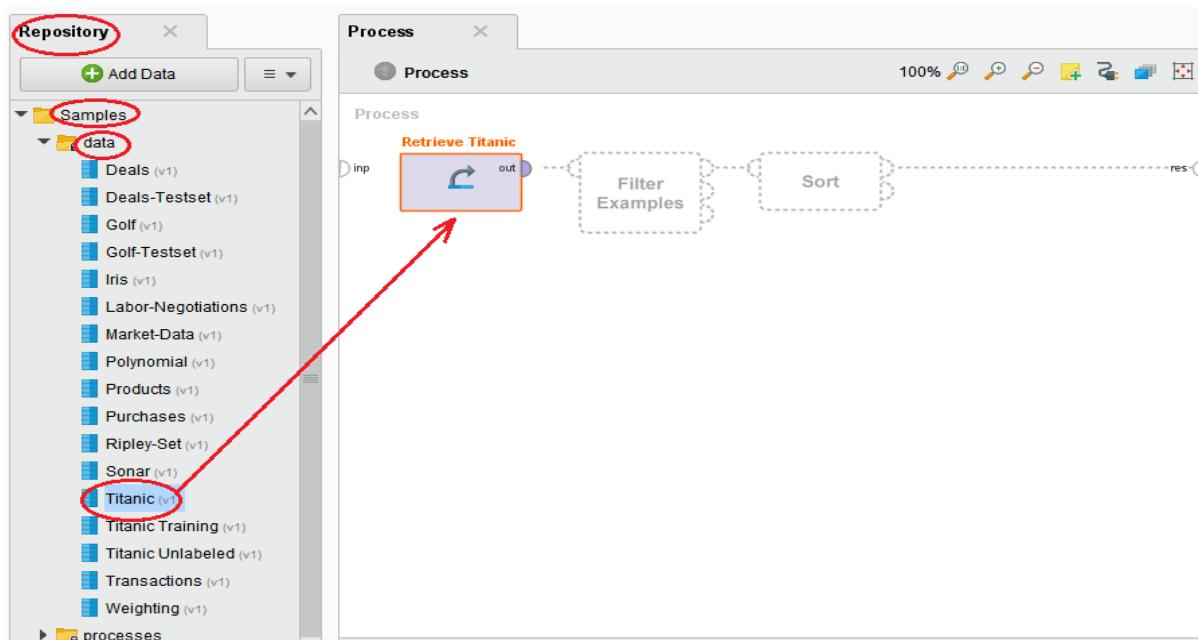
12. Pada tab **Charts**, dapat digambarkan hubungan antara umur (pada sumbu X) serta jumlah anggota keluarga (pada sumbu Y) di mana penumpang dengan umur muda serta jumlah anggota keluarga yang sedikit memiliki tingkat peluang selamat jauh lebih tinggi dibanding dengan yang memiliki anggota keluarga lebih banyak.



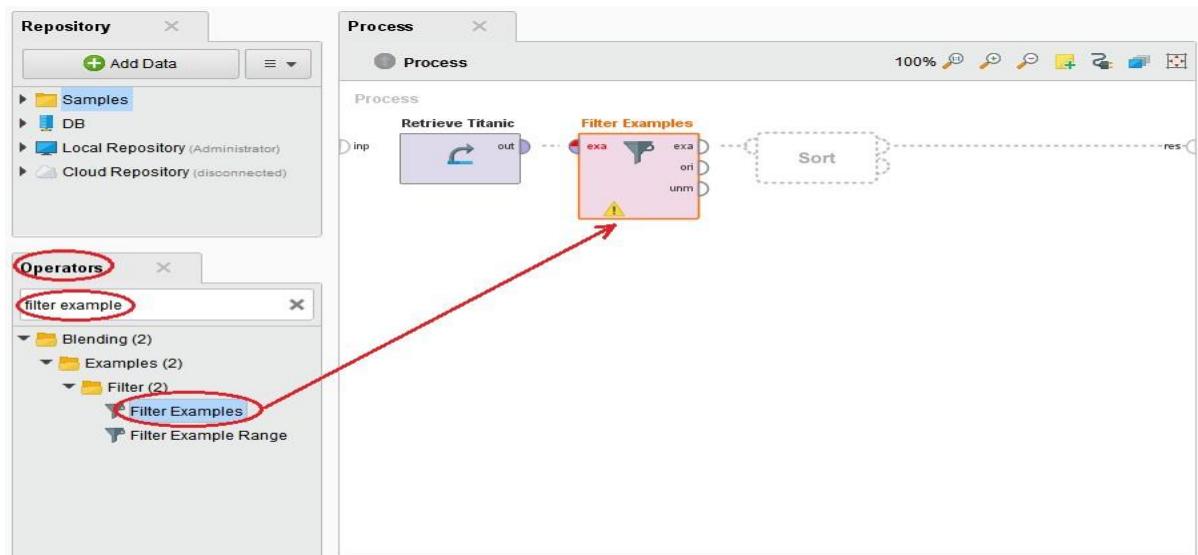
# Tutorial Belajar Rapidminer (Bagian 6) - Menyaring dan Mengurutkan Data

Pada tutorial terakhir, kita telah belajar bagaimana cara mengimpor data ke dalam repository dari RapidMiner, di mana merupakan penyimpanan utama untuk semua data, model, dan proses. Dalam tutorial ini, kita akan menerapkan filter ke himpunan data Titanic untuk melihat hanya penumpang wanita saja. Kemudian, kita dapat dengan mudah menyortir data untuk menemukan tarif tiket tertinggi yang dibayarkan oleh penumpang wanita. Kita kemudian dapat menerapkan keterampilan baru yang telah dipelajari barusan untuk menjawab pertanyaan yang sama untuk penumpang pria. Kita akan mencari cara mengetahui tarif tertinggi yang dibayarkan oleh penumpang wanita. Berikut langkah-langkahnya :

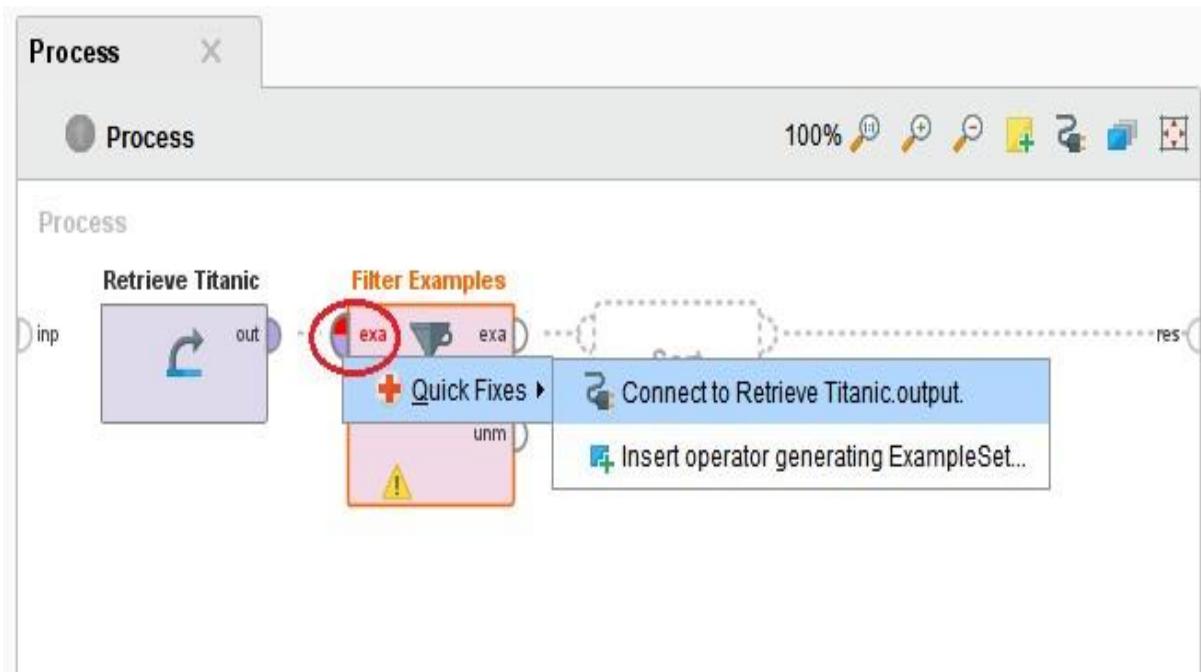
1. Seret data Titanic dari Repotori Sampel ke dalam Panel Process.



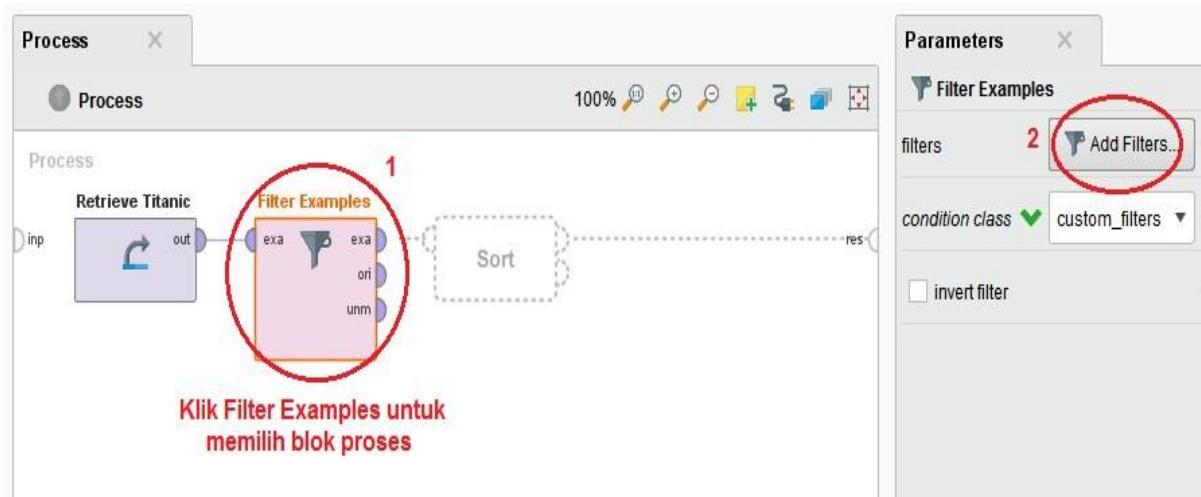
2. Cari operator **Filter Example** dengan menggunakan kotak pencarian di bagian atas panel **Operator**. Seret **Filter Example** ke dalam panel **Process**.



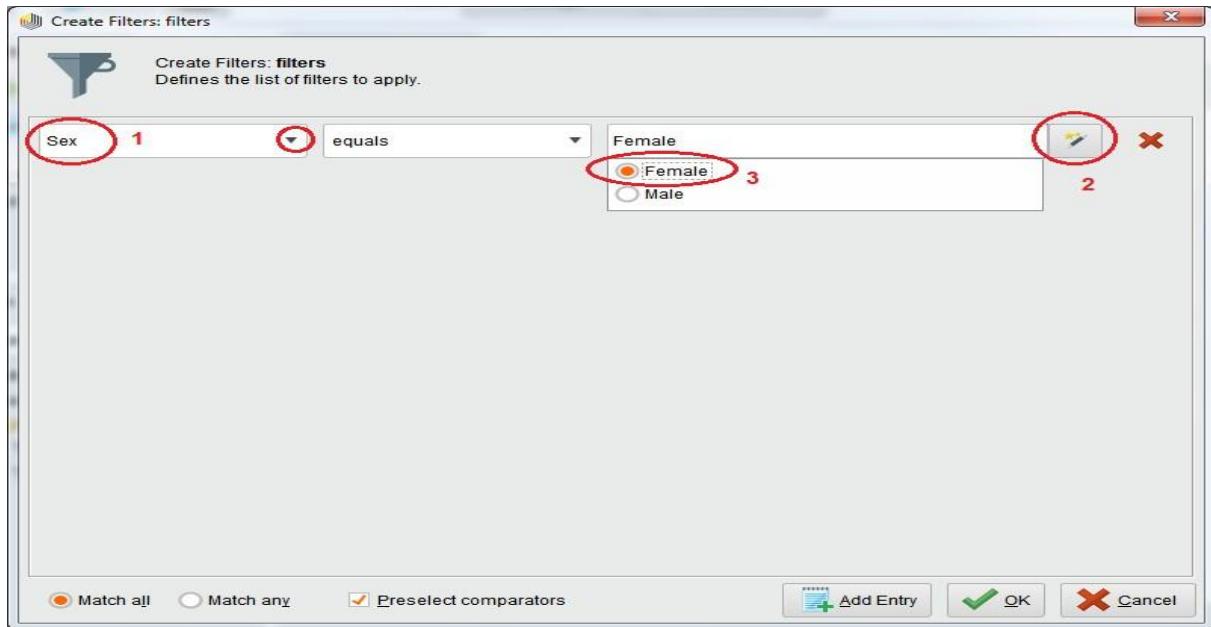
3. Hubungkan port output **Retrieve Titanic** dengan port input dari **Filter Example** dengan cara melakukan klik kanan port **exa** ("example"), pilih **Quick Fixes** kemudian pilih **Connect to Retrieve Titanic output**.



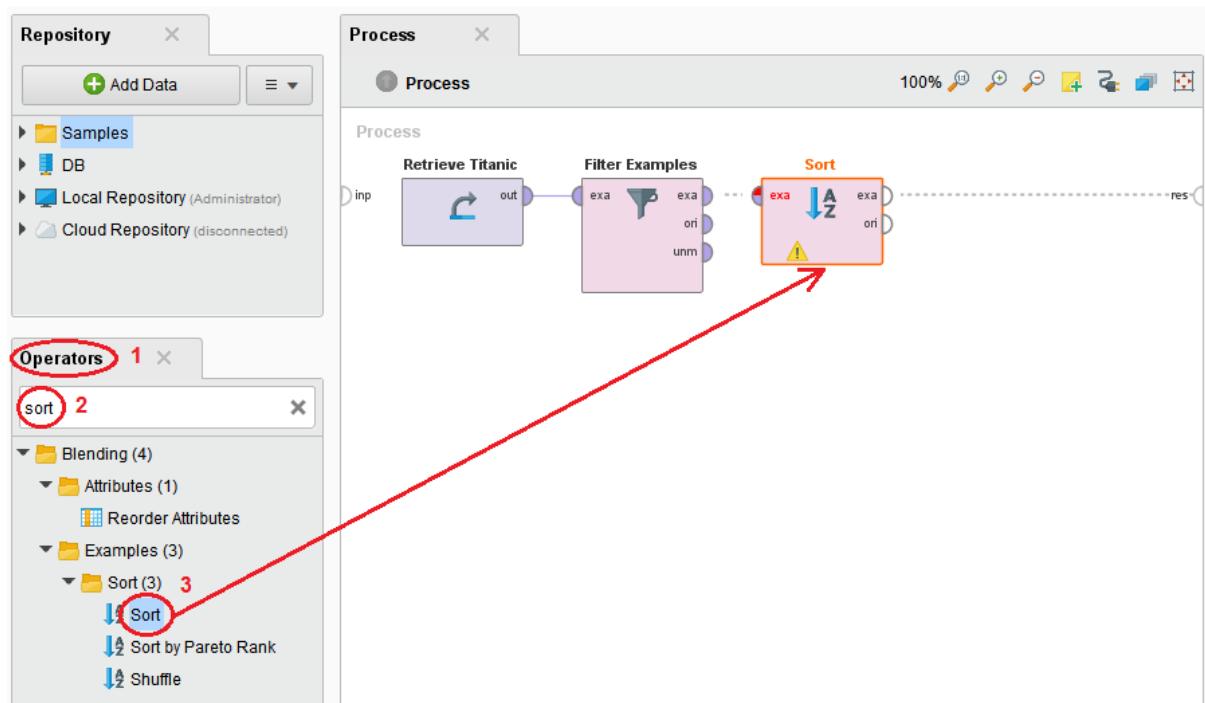
4. Klik **Filter Example** untuk memilihnya, lalu klik **Add Filters** pada panel **Parameters** untuk menentukan filter.



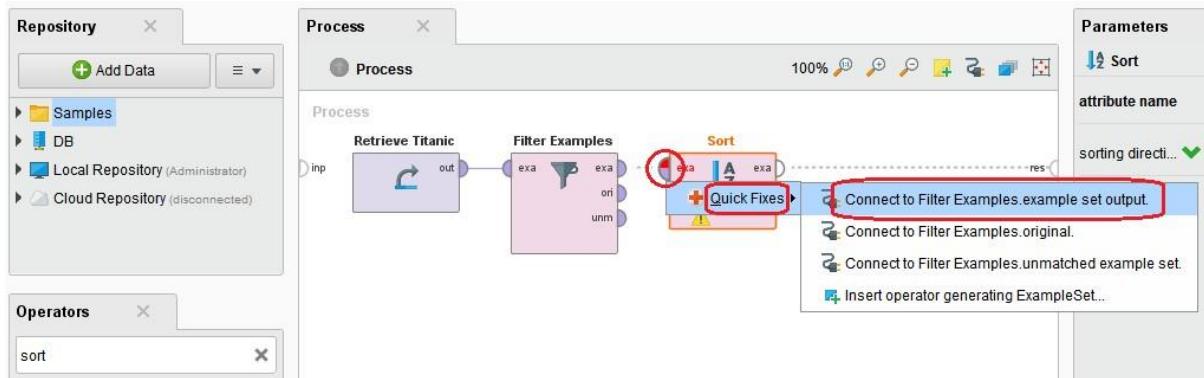
5. Pilih **Sex(1)** di sebelah kiri, sama dengan di tengah, dan ketik **Female** di kotak di sebelah kanan. Selain dengan cara mengetikkan, Anda bisa mengklik **tongkat ajaib(2)** dan memilih opsi **Female(3)** dari daftar.



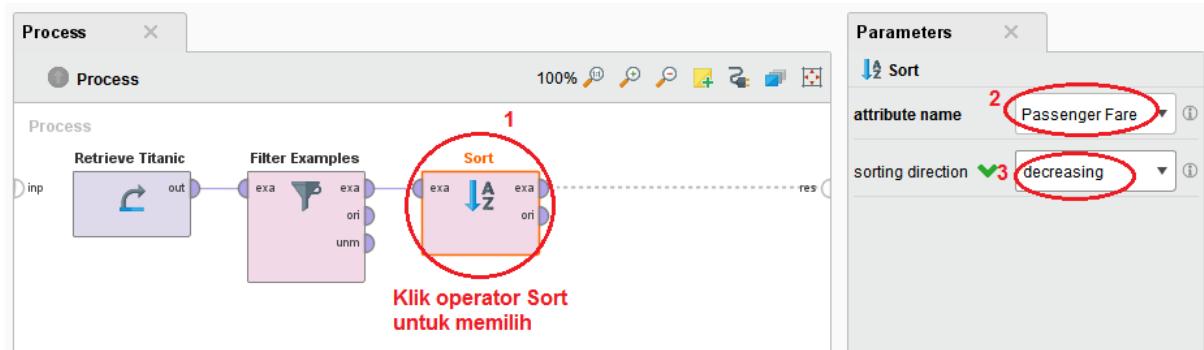
6. Cari dan kemudian seret operator **Sort** ke dalam panel **Process**.



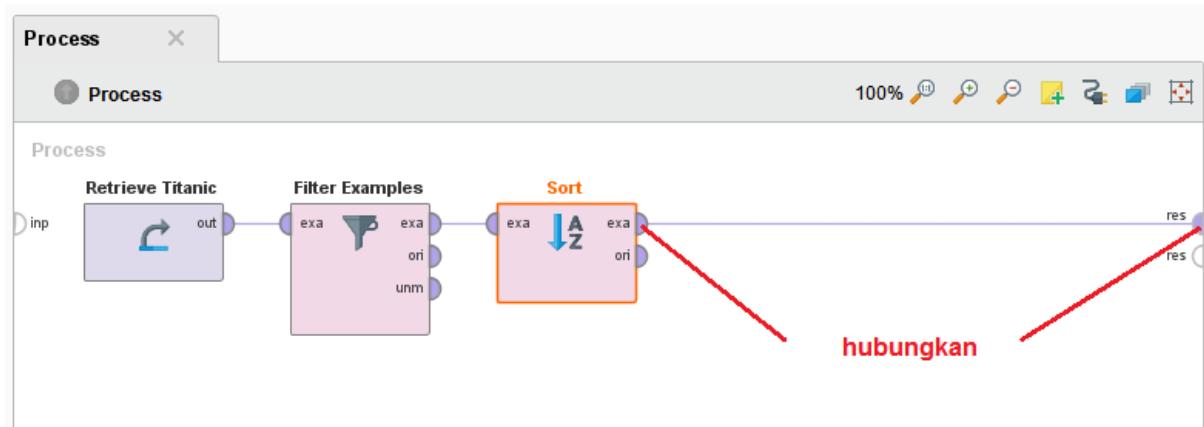
7. Hubungkan output Filter dengan masukan Sort dengan cara meng klik kanan **port exa** ("example"), pilih **Quick Fixes** kemudian pilih **Connect to Filter Examples.example set output**.



8. Klik pada Sort untuk memilihnya. Buat perubahan berikut pada panel **Parameters**, tetapkan attribute name (nama atribut) ke **Passenger Fare**, ubah sorting direction (arah pengurutan) menjadi **decreasing** (menurun).



9. Sambungkan port output **Sort** ke port **res** (hasil) di sebelah kanan panel Sort dengan cara mengklik kiri **port exa** ("example") kemudian dilanjutkan dengan cara mengklik kiri **port res** ("results").



10. Jalankan proses dan periksa hasilnya.



11. Kita sekarang dapat melihat tarif tertinggi yang dibayarkan untuk seorang wanita di kapal Titanic. Tarif tertinggi yang diperoleh, ditemukan pada **baris pertama** dan **kedua** pada kolom **Passenger Fare**.

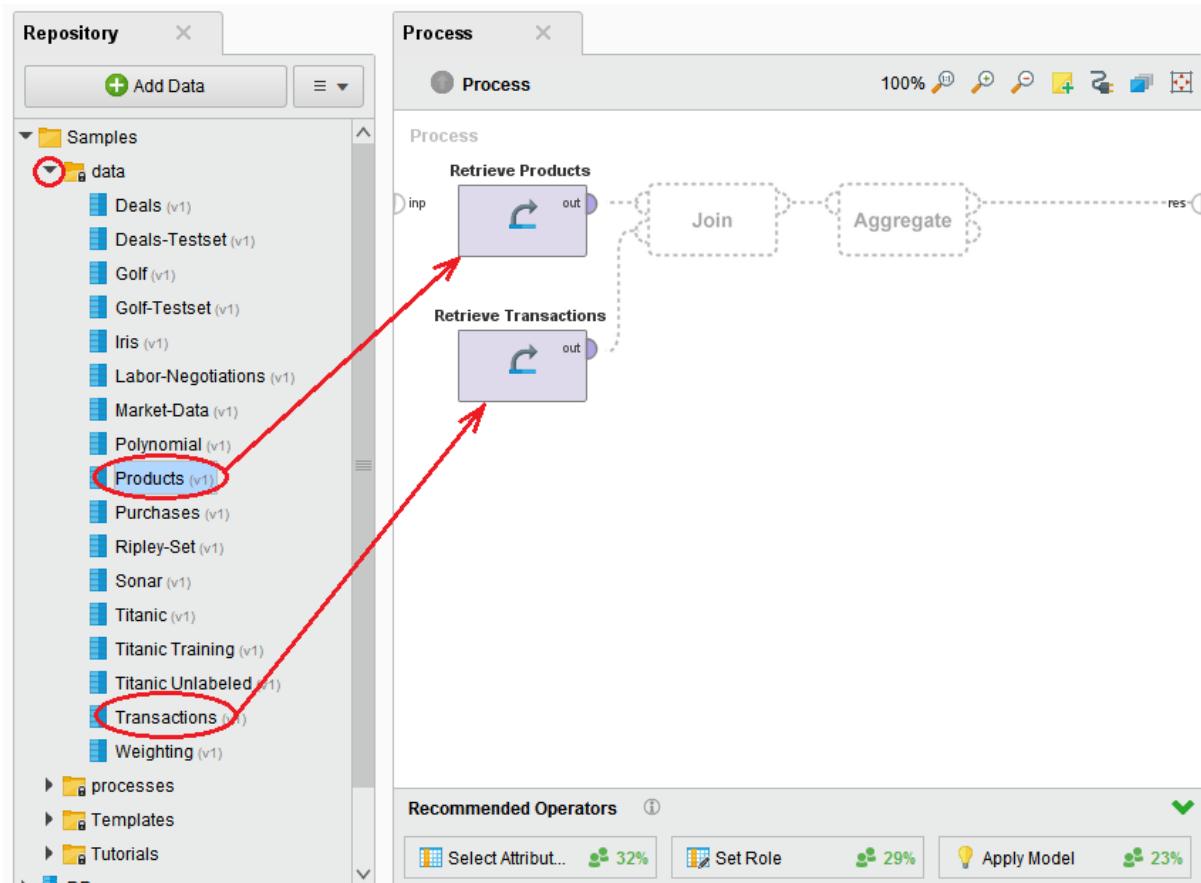
Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger Fare	Cabin
1	First	Cardeza, Mrs....	Female	58	0	1	PC 17755	512.329	B51 B53 B55
2	First	Ward, Miss. A...	Female	35	0	0	PC 17755	512.329	?
3	First	Fortune, Miss...	Female	24	3	2	19950	263	C23 C25 C2
4	First	Fortune, Miss...	Female	28	3	2	19950	263	C23 C25 C2
5	First	Fortune, Miss...	Female	23	3	2	19950	263	C23 C25 C2
6	First	Fortune, Mrs. ...	Female	60	1	4	19950	263	C23 C25 C2
7	First	Bowen, Miss....	Female	45	0	0	PC 17608	262.375	?
8	First	Chaudanson,...	Female	36	0	0	PC 17608	262.375	B61
9	First	Ryerson, Mis...	Female	18	2	2	PC 17608	262.375	B57 B59 B61
10	First	Ryerson, Mis...	Female	21	2	2	PC 17608	262.375	B57 B59 B61
11	First	Ryerson, Mrs....	Female	48	1	3	PC 17608	262.375	B57 B59 B61
12	First	Baxter, Mrs. J...	Female	50	0	1	PC 17558	247.521	B58 B60
13	First	Douglas, Mrs...	Female	27	1	1	PC 17558	247.521	B58 B60
14	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	C62 C64

# Tutorial Belajar Rapidminer (Bagian 7) - Menggabung dan Mengelompokkan

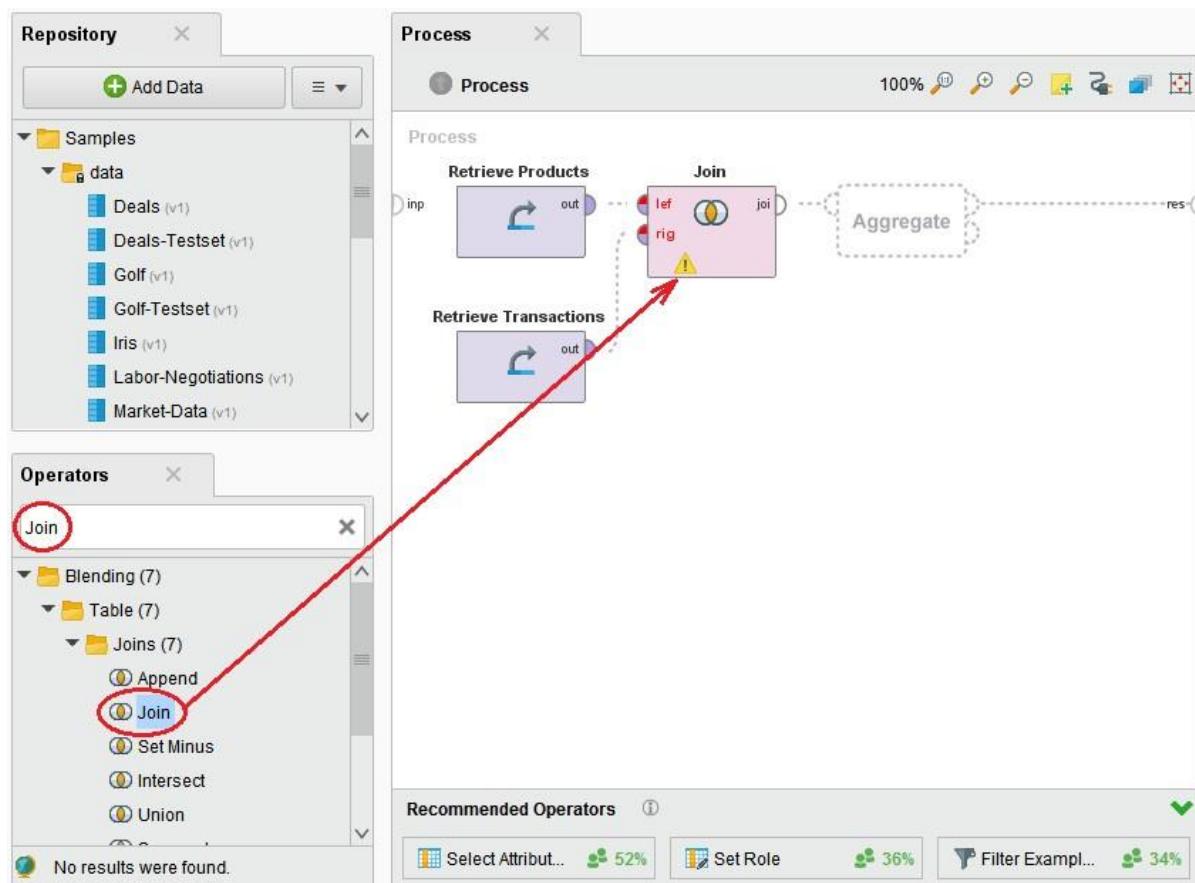
Kita akan berhenti sejenak dari topic Titanic dan mempelajari tentang beberapa tugas lain yang sering digunakan dalam persiapan data, khususnya penggabungan dan pengelompokan data bersama-sama. Kita akan menangani dua kumpulan data: satu berisi produk yang dijual oleh organisasi dan satu dengan transaksi (informasi tentang pelanggan mana yang membeli produk itu). Setelah menggabungkan rangkaian ini, kita dapat menjawab pertanyaan tentang produk yang paling sering dibeli atau siapa pelanggan kita yang paling setia.

Berikut Langkah cara menggabungkan dan mengelompokkan data :

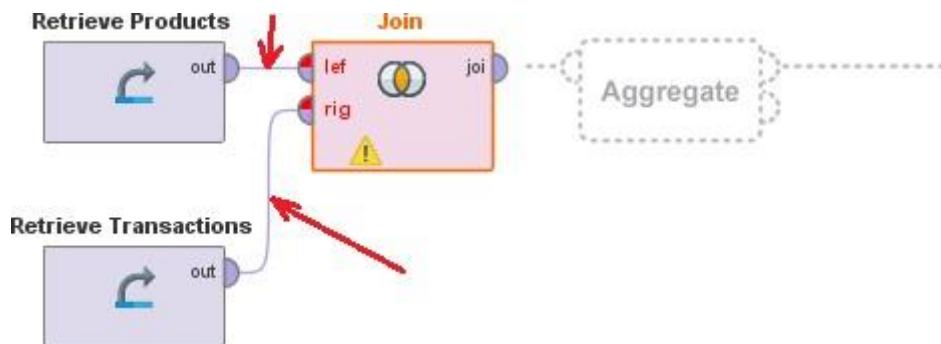
1. Perluas Sampel repositori di panel **Repositori**. Selanjutnya, buka folder **data** dalam contoh repositori untuk mengambil data **Produk** dan **Transaksi**.



2. Seret data **Products** dan data **Transactions** dari folder **Samples - Data** ke dalam panel Proses
3. Cari operator **Join** di kotak pencarian di bagian atas panel **Operator**. Seret operator **Join** menuju panel Proses.

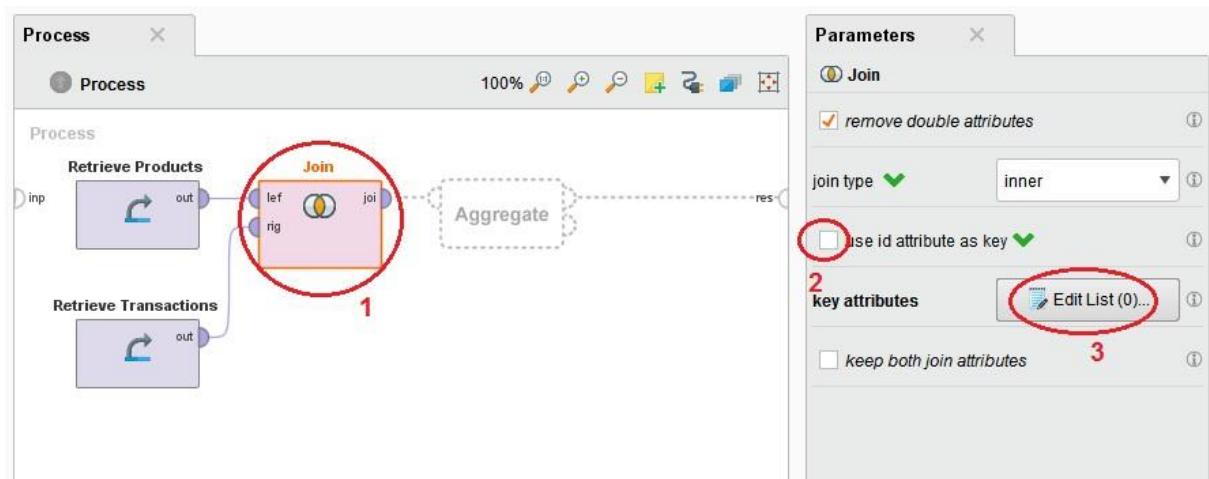


4. Sambungkan port output dari **Retrieve Products** ke port input **Join** (tidak masalah yang mana).

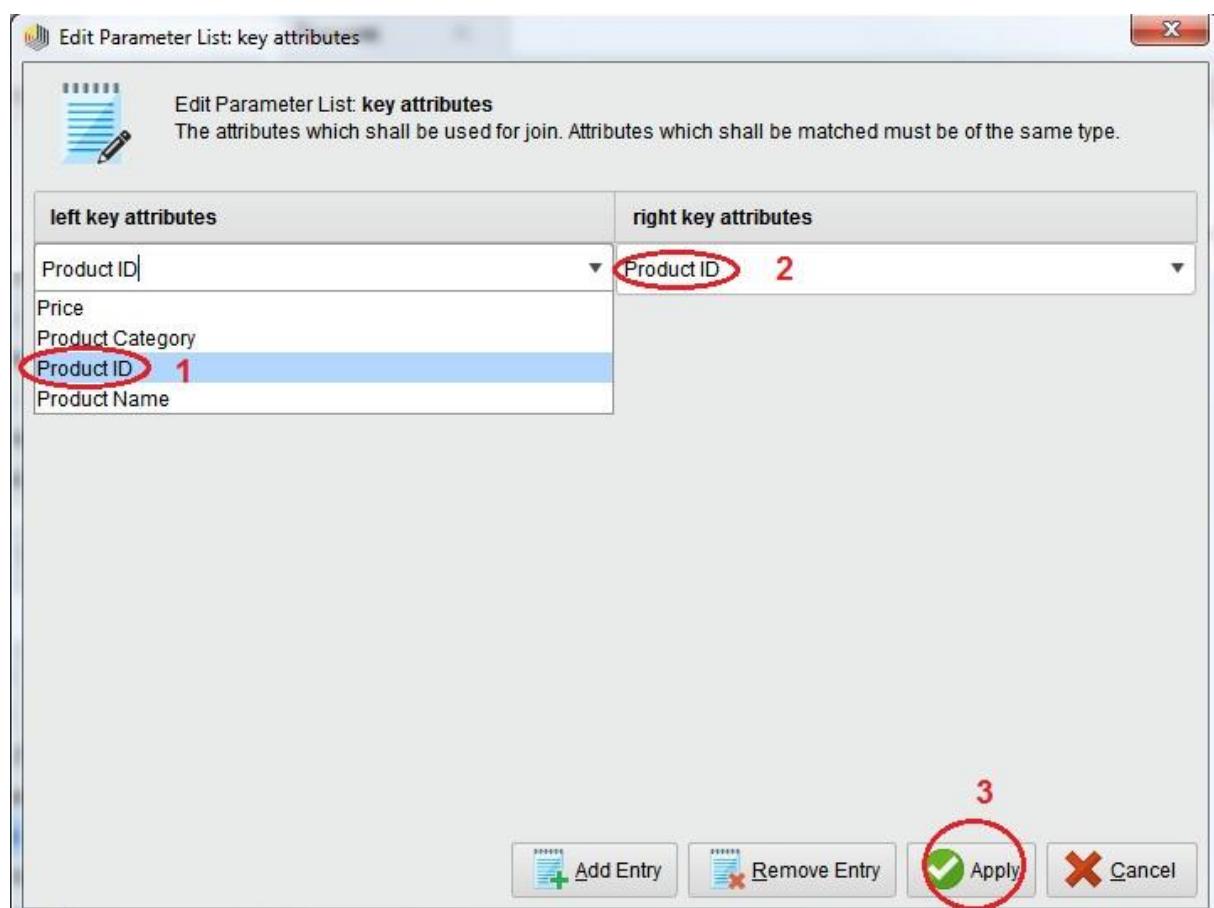


5. Hubungkan **Retrieve Transactions** ke port input Join lainnya.

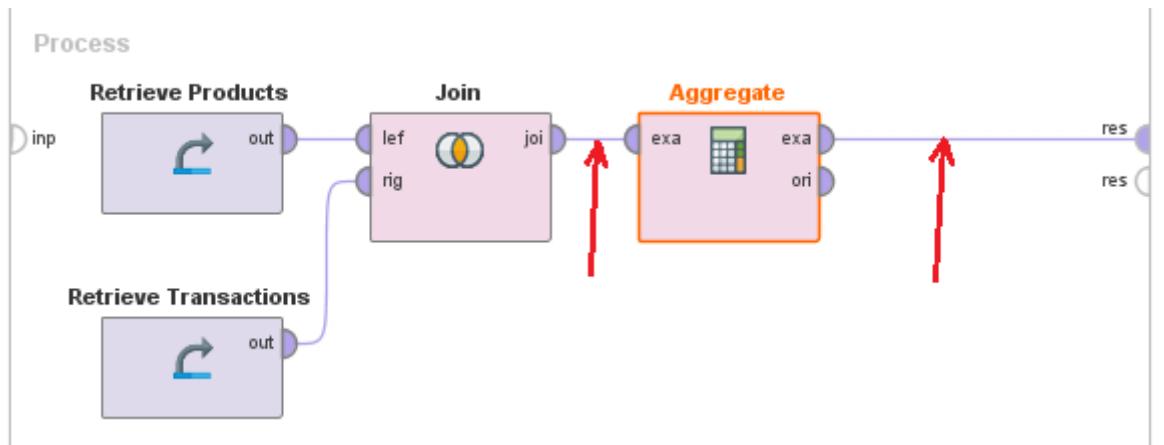
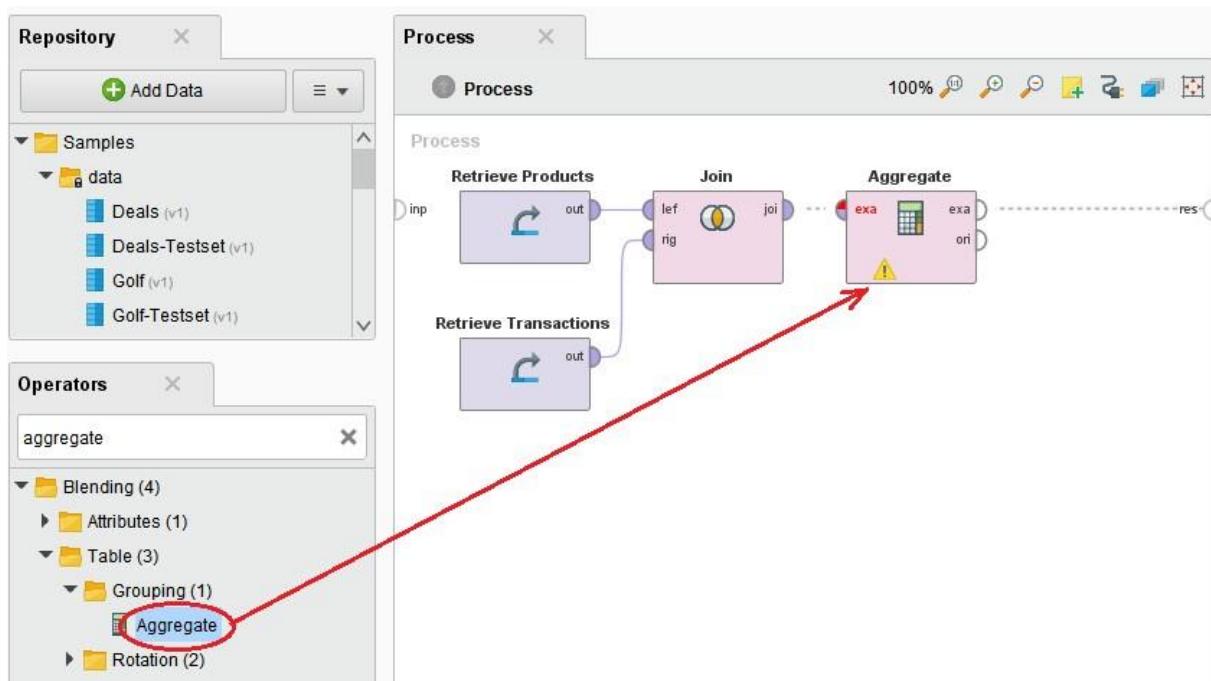
6. Klik **Join** untuk memilihnya. Pada panel **Parameters**, hilangkan tanda centang pada opsi **use id attribute as key**. Bidang atribut utama muncul.



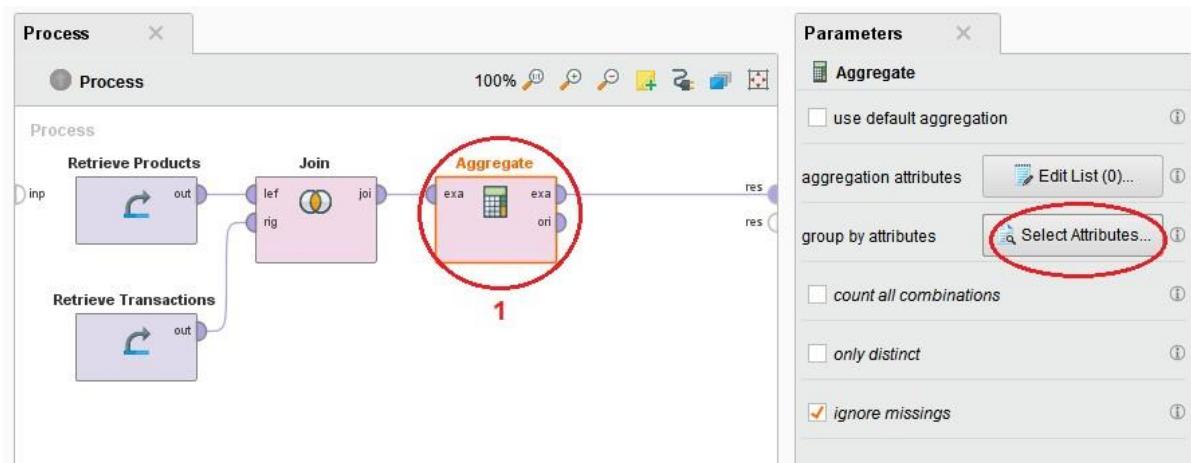
7. Klik **Edit List**. Pilih **Product ID** untuk **left** dan **right key attributes**. Kemudian, klik **Apply**. Hasil operasi Join akan membentuk tabel yang menunjukkan setiap transaksi dan rincian produknya. Dua kolom ID yang kita definisikan sebagai atribut utama untuk join menentukan pemetaan antara baris dari dua tabel asli.



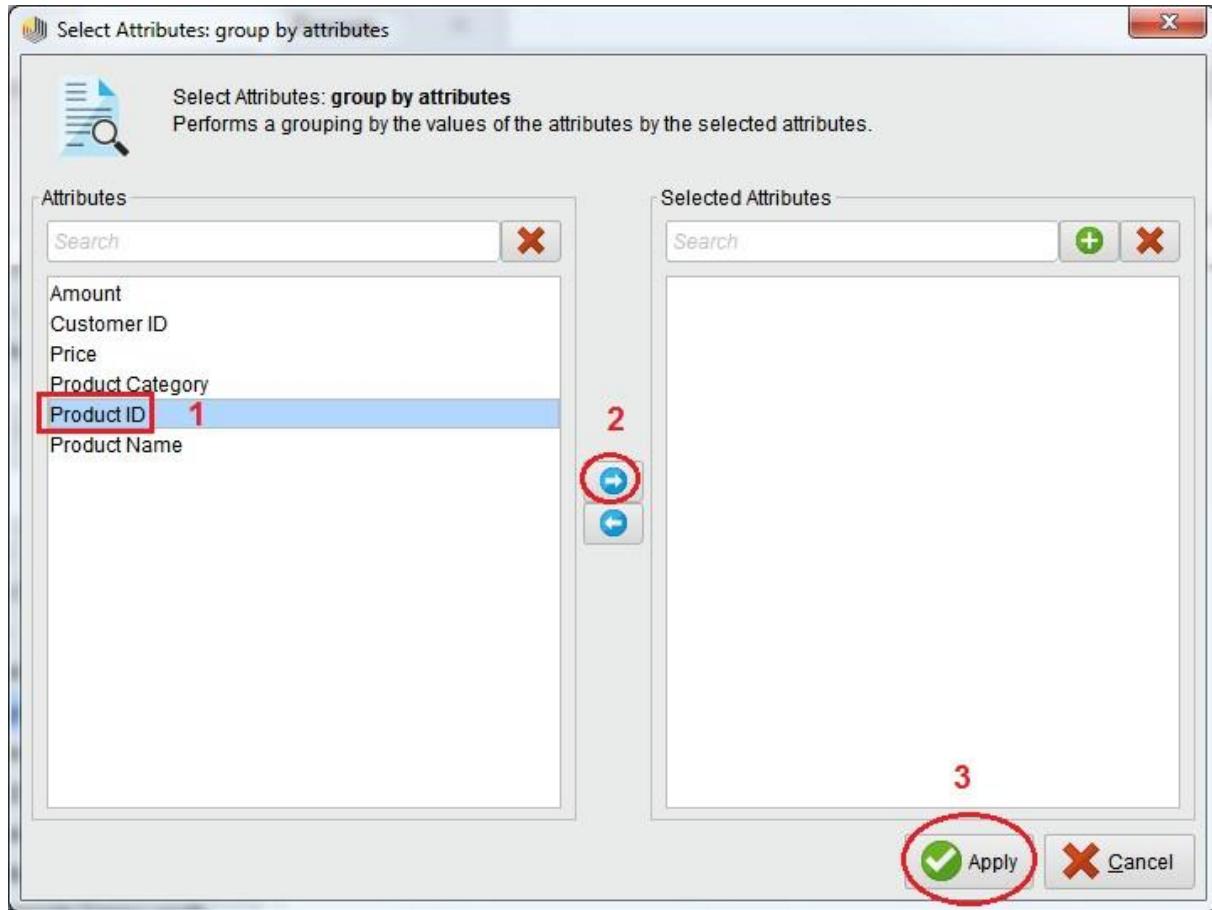
8. Seret operator **Agregat** ke dalam proses. Hubungkan ke **output** dari **Join**.



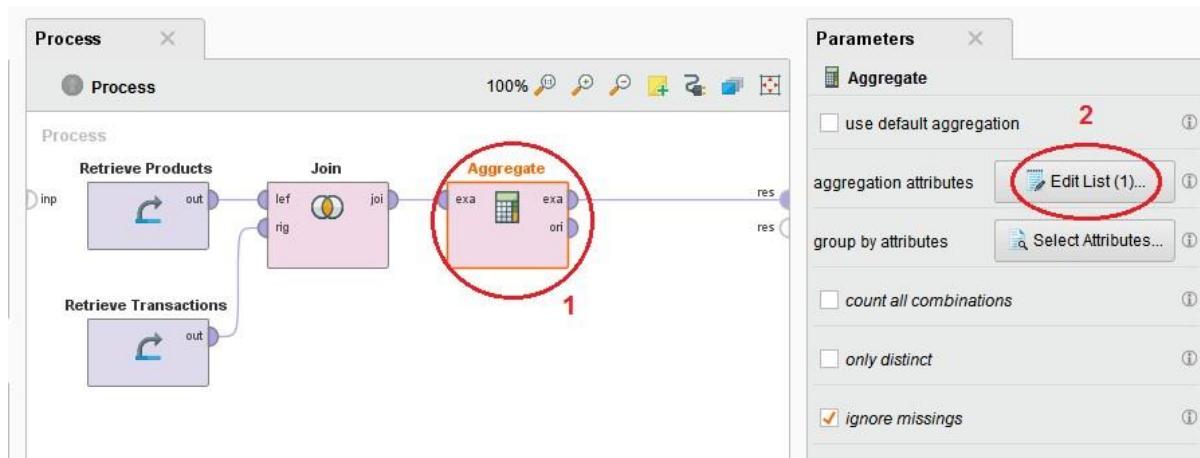
9. Klik Agregat untuk memilih. Buat perubahan berikut pada panel Parameter:



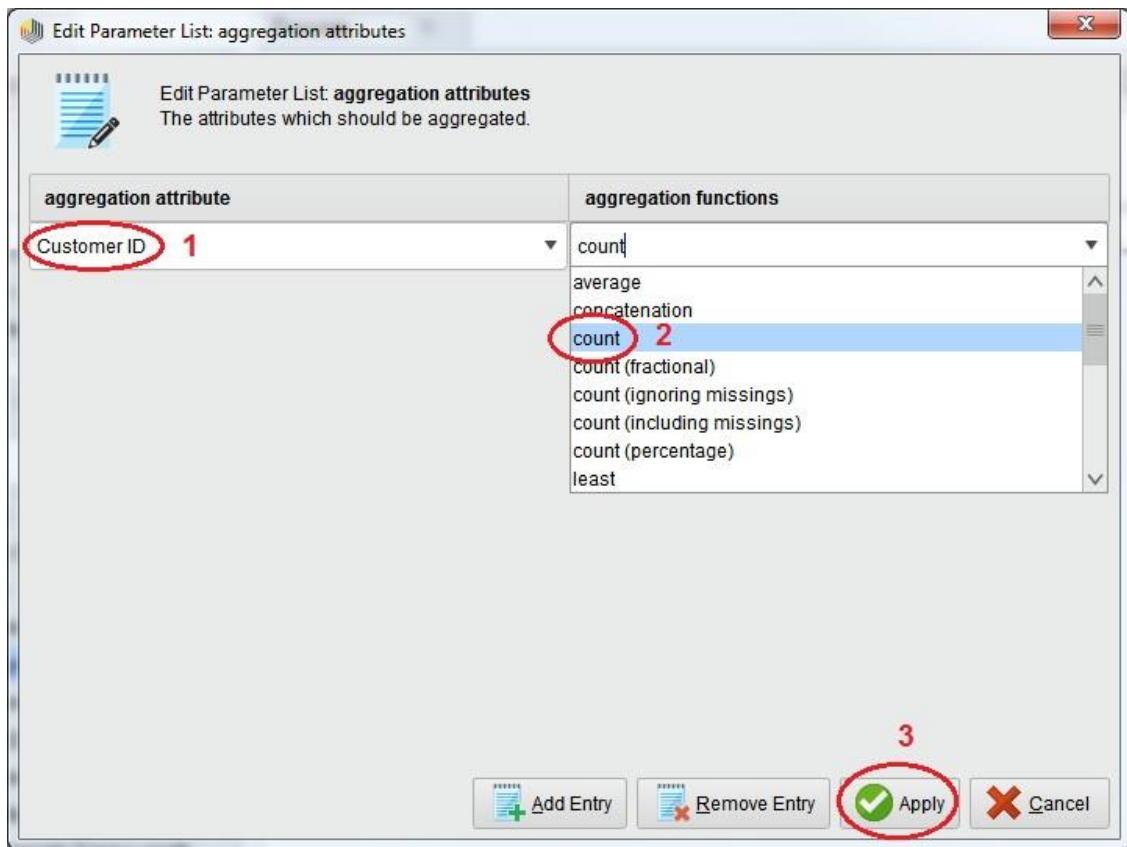
10. Klik pada **group by attributes**. Klik **select Attributes**
11. Kemudian, pilih **Product ID** dengan memindahkannya ke kanan. Klik **Apply**.



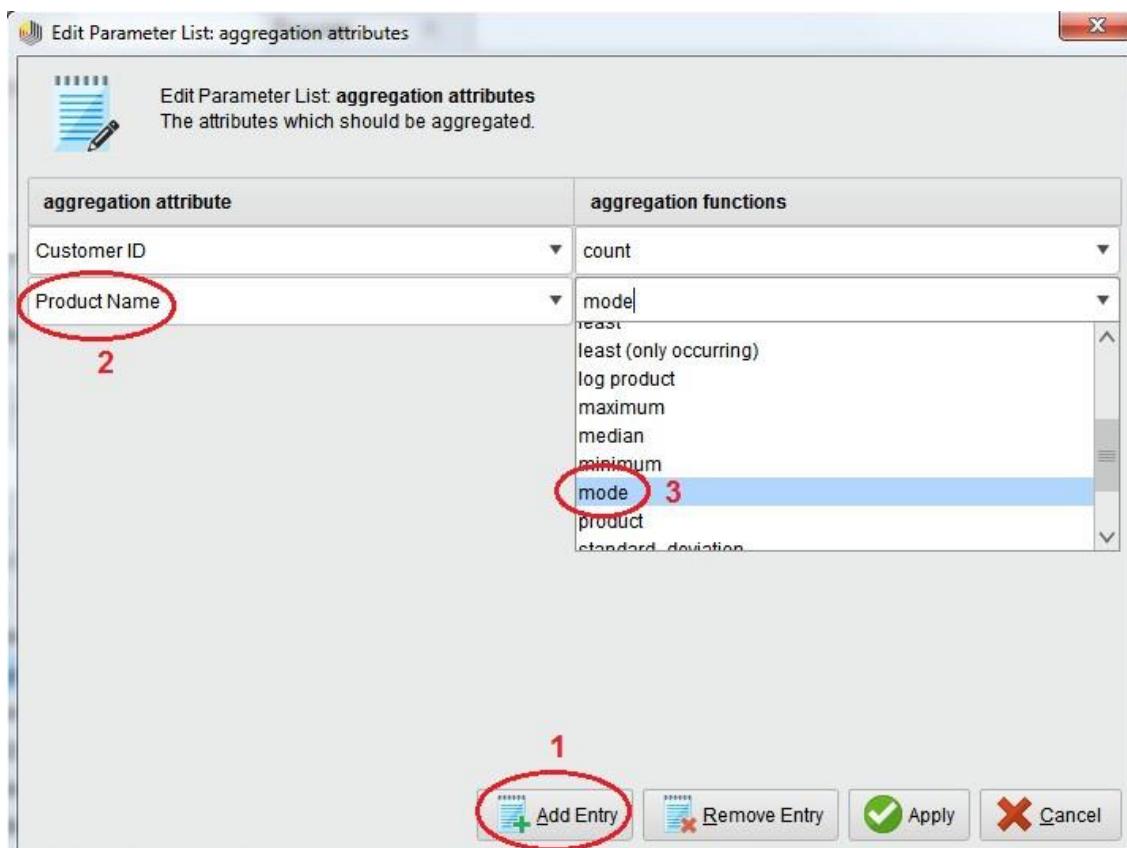
12. Klik **aggregation attributes**, kemudian klik **Edit list**



13. Pilih **Customer ID** di kotak kiri dan atur fungsi **count** untuk menghitung di kotak kanan.



14. Masih di kotak dialog ini dan tambahkan entri lain **Product Name** dengan mengatur fungsi ke mode. Klik **Apply**



15. Jalankan dengan mengklik tombol process

RapidMiner Studio Free 7.5.001 @ pc

File Edit Process View Connections Cloud Settings Extensions

Result History ExampleSet (Aggregate)

ExampleSet (178 examples, 0 special attributes, 3 regular attributes)

Data Statistics Charts Advanced Charts

Row No.	Product ID	count(Customer ...)	mode(Produ...
138	138	22	Athsat
42	42	21	Turbolax
175	175	21	Phalanx
34	34	20	Zombrex
82	82	20	Qualex
103	103	20	Safsprin
156	156	20	Jammitin
1	1	19	Repressitol
17	17	19	Doloxyan
41	41	19	Senzu Bean
67	67	19	Formula 977
79	79	19	Quinium

klik beberapa kali untuk mengurutkan

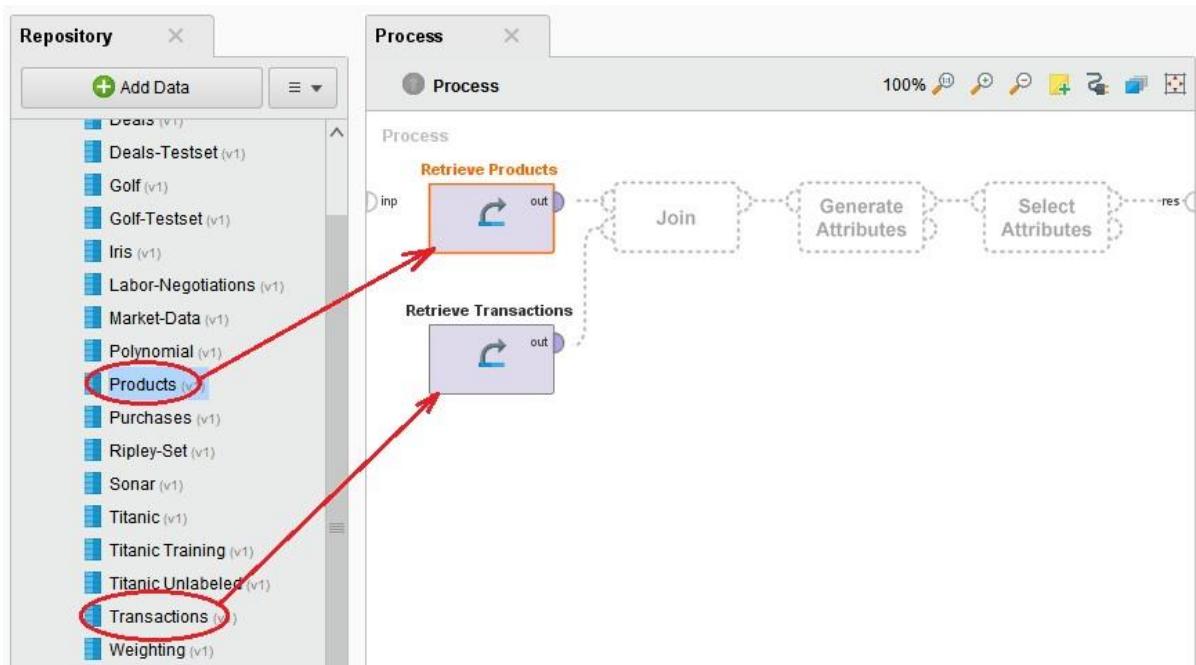
Agregat melakukan "group-by-function" yang mungkin kita ketahui dari database. Di samping menggabungkan dan menyaring, fungsi agregat adalah salah satu operator yang paling penting untuk pencampuran data. Dalam kasus ini, agregat mengelompokkan data berdasarkan produk, menghitung jumlah pembelian untuk setiap produk, dan menggunakan nama produk untuk mendeskripsikan produk ini. Hasilnya adalah tabel semua produk dengan atribut berikut: ID produk, nama produk, dan jumlah pelanggan yang telah membeli produk.

# Tutorial Belajar Rapidminer (Bagian 8) - Membuat dan Menghilangkan Atribut

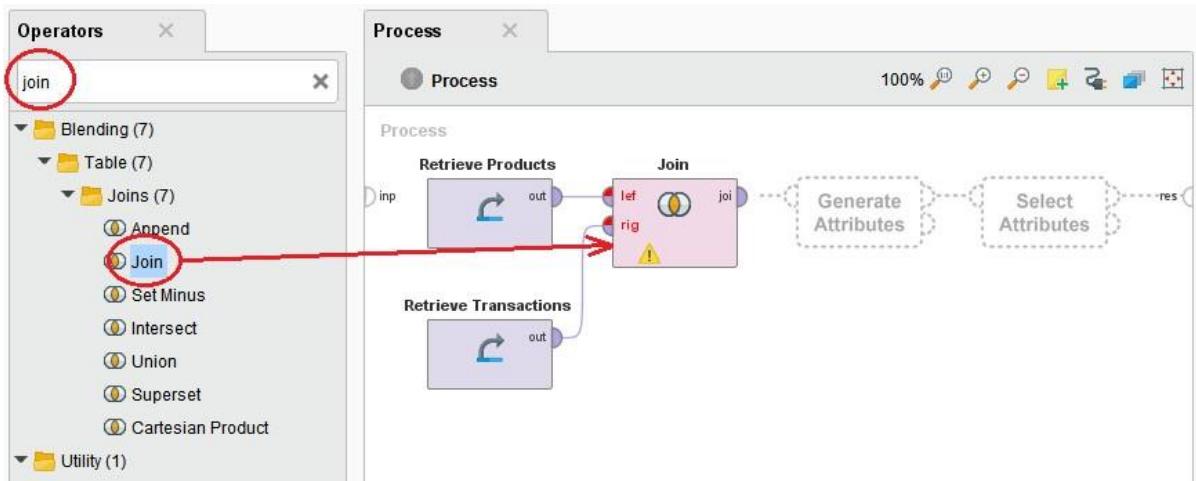
Kita sudah hampir siap untuk membangun model prediksi pertama menggunakan RapidMiner! Tapi pertama-tama, kita perlu terlebih dahulu mempelajari dua operasi yang sangat penting untuk mengubah dataset (kumpulan data) kita menjadi format yang lebih sesuai untuk pembelajaran. Awal proses dari ini sama dengan yang sebelumnya. Ini adalah kesempatan besar bagi kita untuk menerapkan apa yang telah kita pelajari sejauh ini. Kita kemudian akan membuat kolom data baru dan menghapus beberapa kolom dari data yang tidak terpakai / tidak diperlukan.

Berikut langkah untuk bekerja dengan atribut :

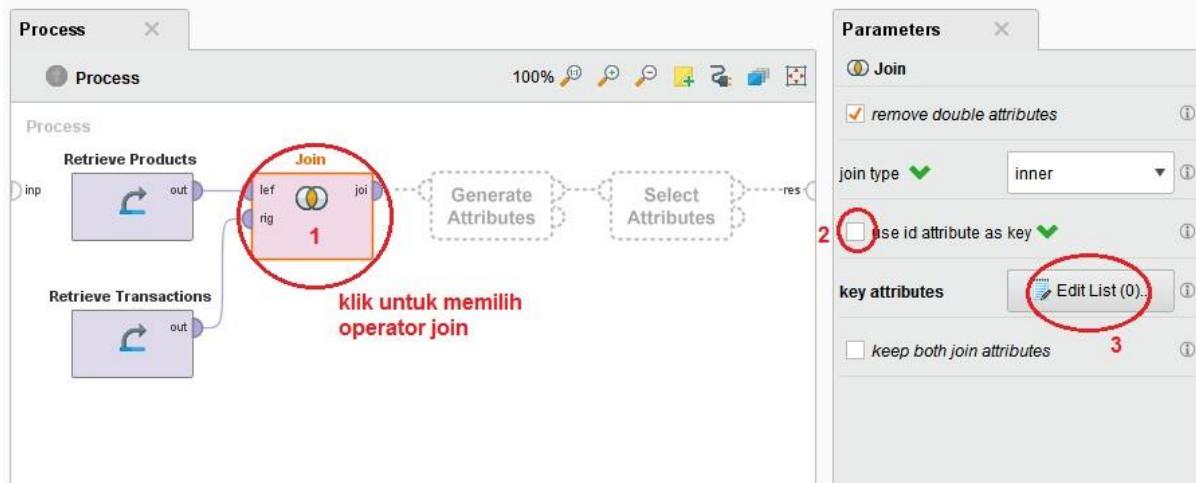
1. Seret dataset **Transactions**(Transaksi) dan **Products** (Produk) ke dalam Proses.



2. Tambahkan operator **Join** (Gabung).
3. Hubungkan semua operator.

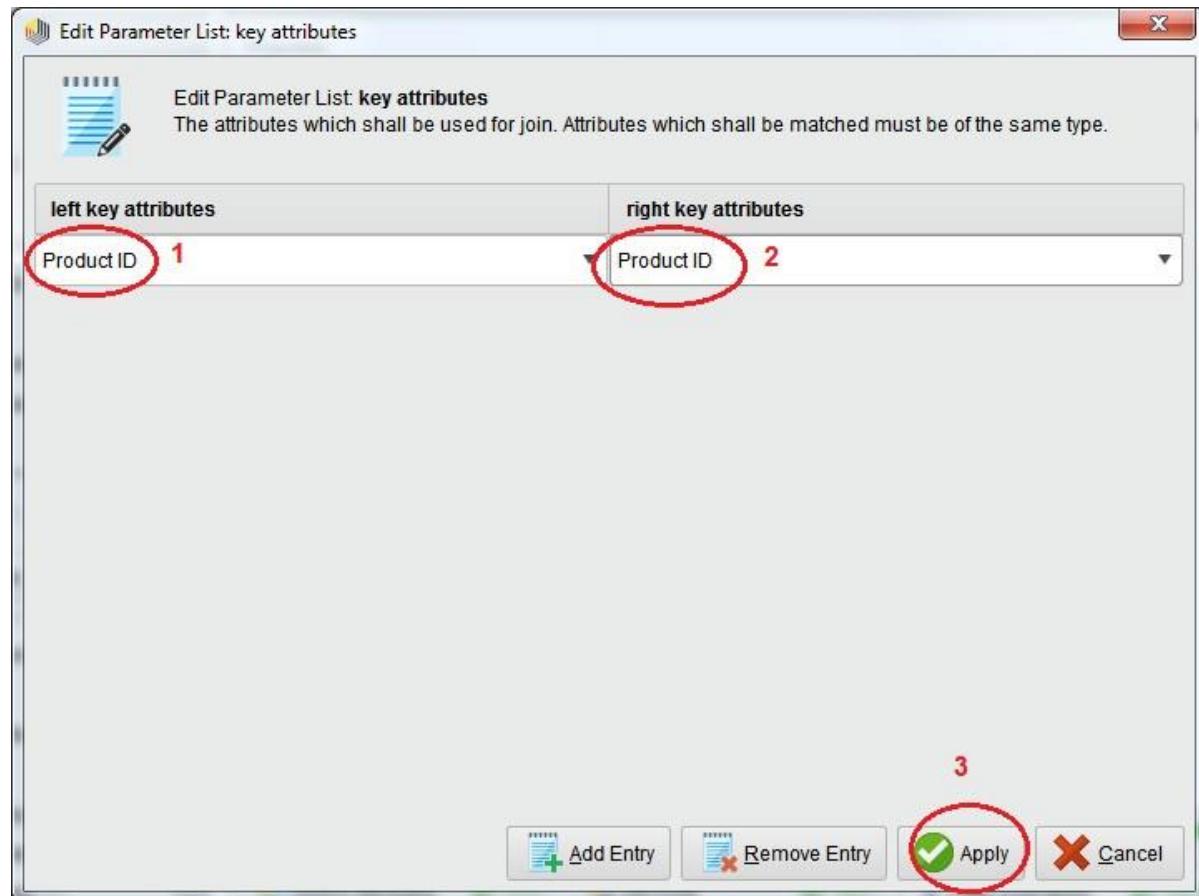


4. Tentukan kolom yang akan digunakan di Parameters for Join dengan cara meng klik operator **join**(1). Kemudian hilangkan tanda centang pada opsi **use id attributes as key** (2),



klik **Edit List** (3)

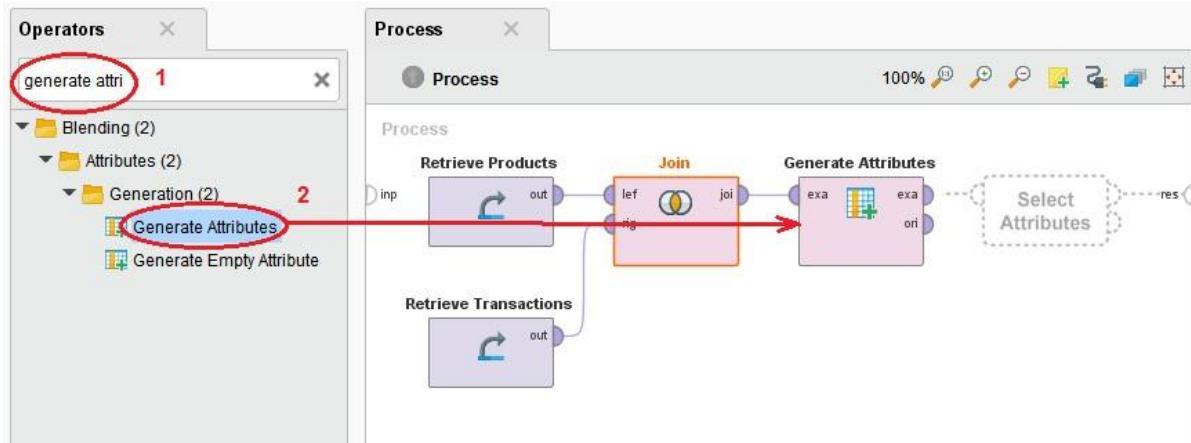
5. Gunakan **Product ID** untuk atribut kunci kiri(1) dan kanan(2) kemudian tekan tombol **Apply** (3).



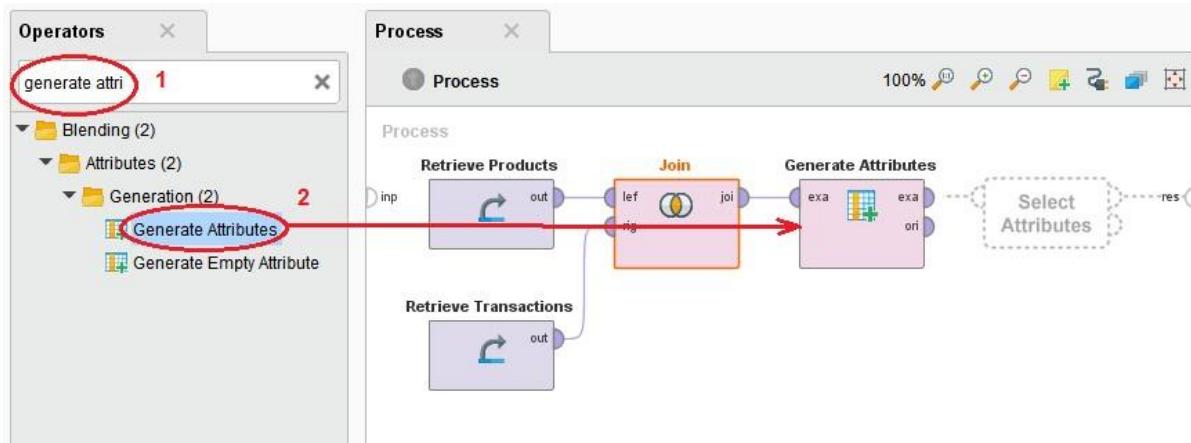
Langkah di atas akan menghasilkan keluaran berupa semua transaksi dengan rincian produk untuk setiap transaksi.

Proses berikutnya adalah mendefinisikan atribut baru, berikut langkahnya :

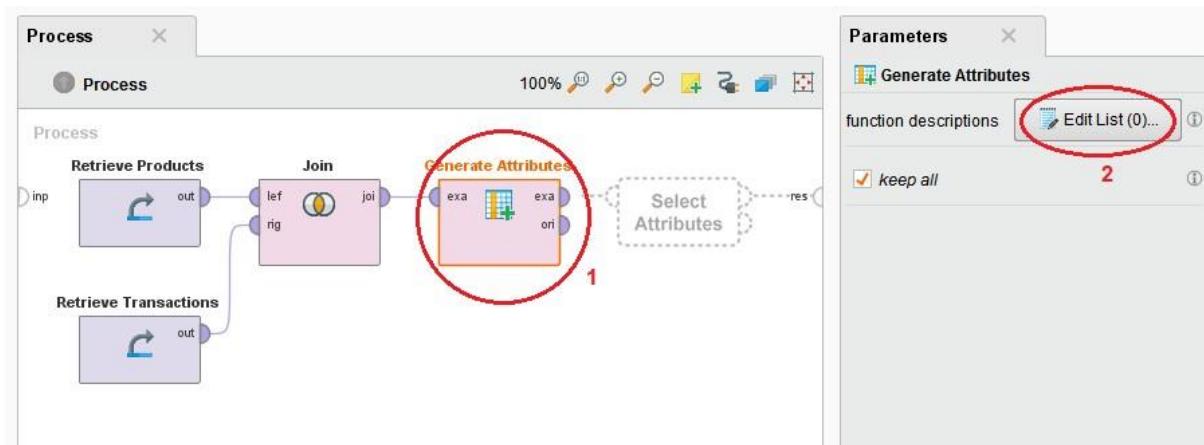
1. Tambahkan operator **Generate Attributes**.



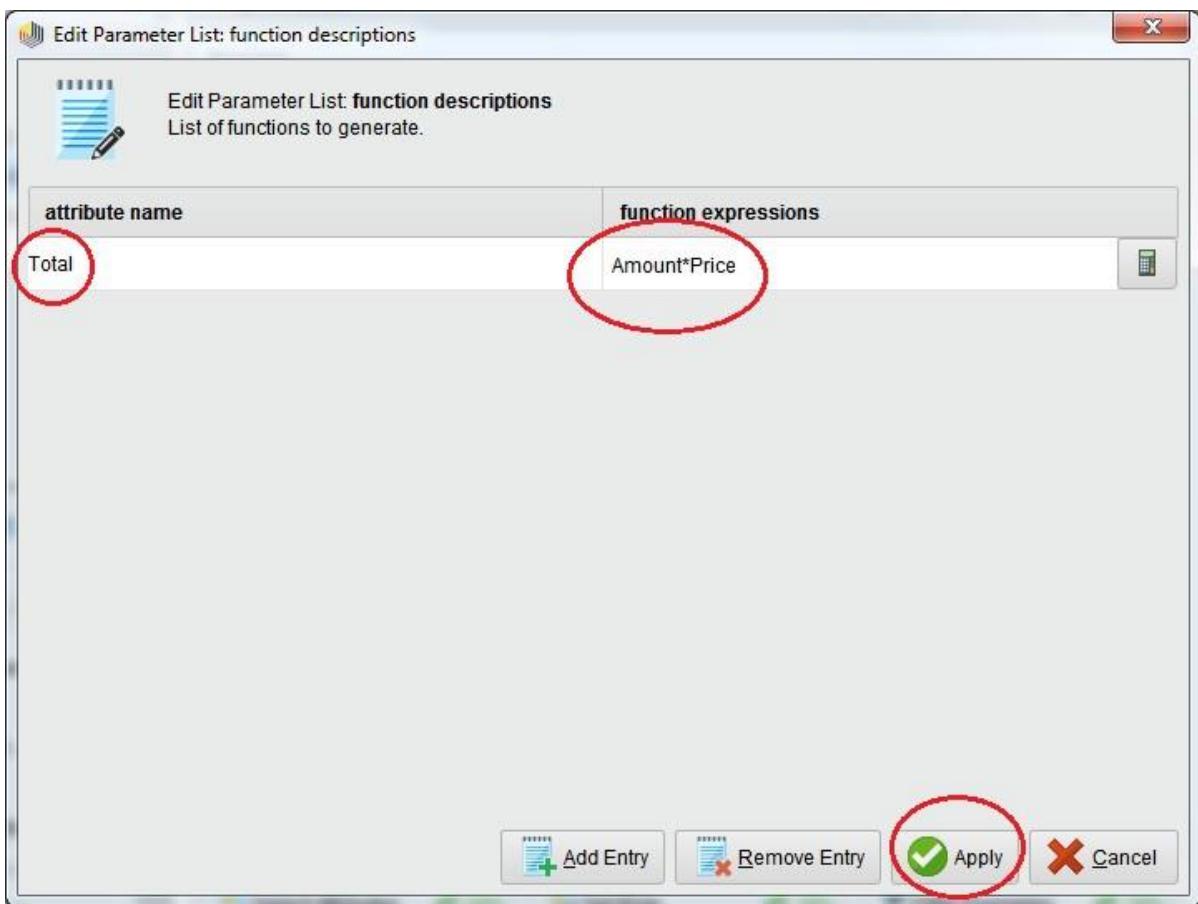
2. Hubungkan operator dengan **Join**.



3. Klik operator **Join(1)**, kemudian klik **Edit List** di Parameters (2) untuk Menghasilkan Atribut untuk menentukan atribut baru (kolom). Dialog akan muncul.

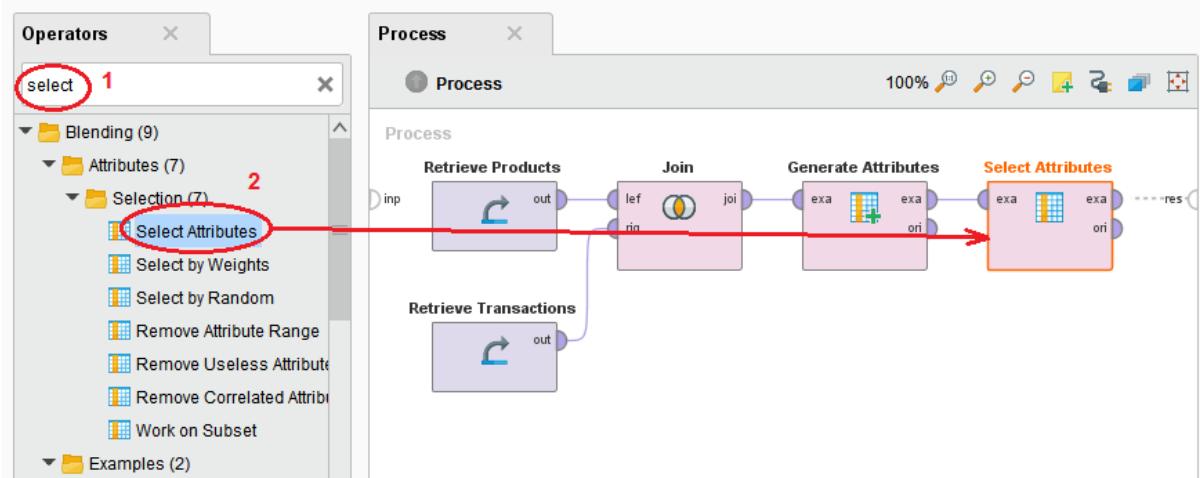


4. Di kolom kiri dialog, masukkan **Total** untuk nama atribut.
5. Di kolom kanan, ketik **Amount \* Price** untuk ekspresi fungsi.

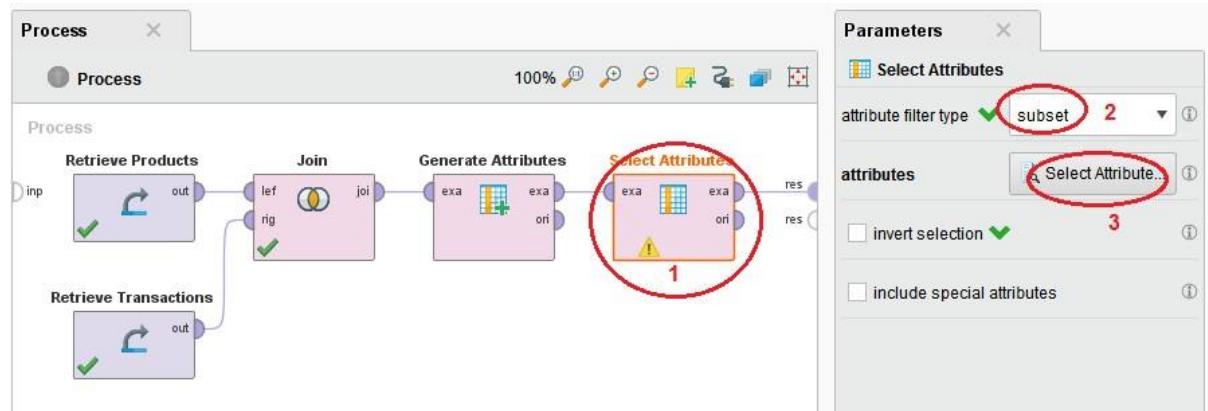


Berikutnya adalah membuang atribut yang tidak diperlukan. Langkahnya adalah sebagai berikut :

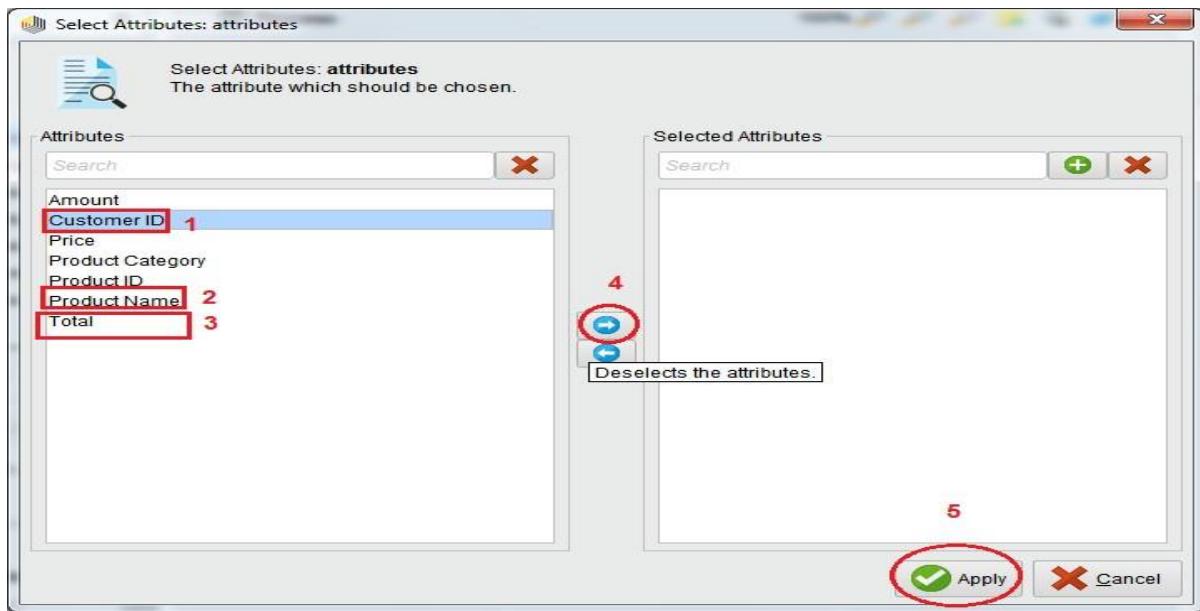
1. Tambahkan operator **Select Attributes** ke proses dan hubungkan.



2. Buat perubahan berikut pada Parameter: Atur **attribute filter type** ke **subset**. Secara umum, operator ini hanya akan diterapkan pada atribut (kolom) yang sudah kita tentukan. Hal ini akan memungkinkan kita memilih subkumpulan kolom untuk disimpan dalam data - semua kolom lainnya akan dihapus.



4. Kemudian Klik opsi **Select Atribut**. Pada dialog yang dihasilkan, pilih atribut **Customer ID** (ID Pelanggan), **Product Name** (Nama Produk), dan **Total**. Jika daftarnya masih kosong, mungkin kita lupa menghubungkan operator terlebih dahulu



## 5. Jalankan prosesnya

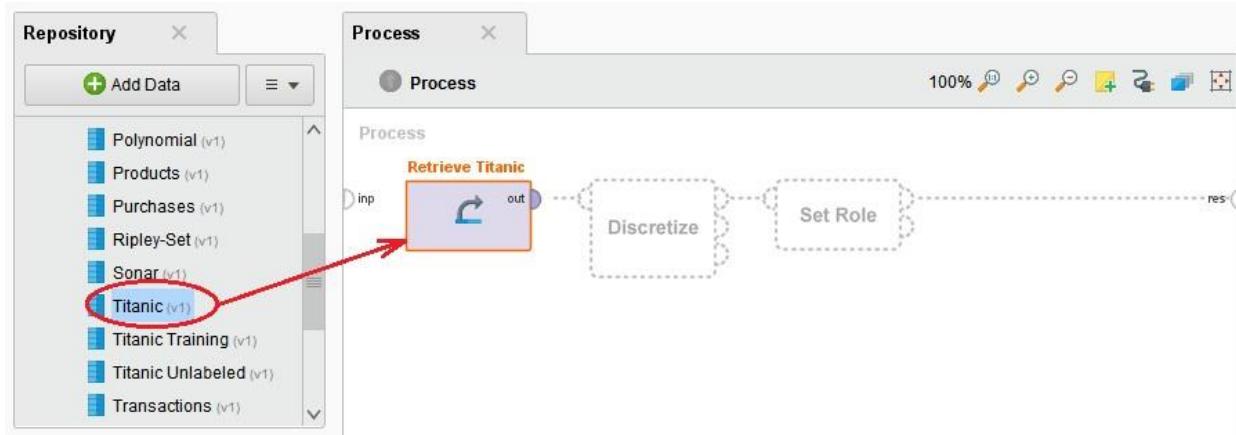
Row No.	Product Na...	Customer ID	Total
1	Repressitol	220	142.160
2	Repressitol	125	106.620
3	Repressitol	352	71.080
4	Repressitol	408	106.620
5	Repressitol	334	35.540
6	Repressitol	31	71.080
7	Repressitol	62	106.620
8	Repressitol	79	106.620
9	Repressitol	64	35.540
10	Repressitol	191	106.620
11	Repressitol	365	142.160
12	Repressitol	231	142.160
13	Repressitol	566	142.160
14	Repressitol	99	71.080

# Tutorial Belajar Rapidminer (Bagian 9) - Mengubah Tipe dan Role dari Atribut

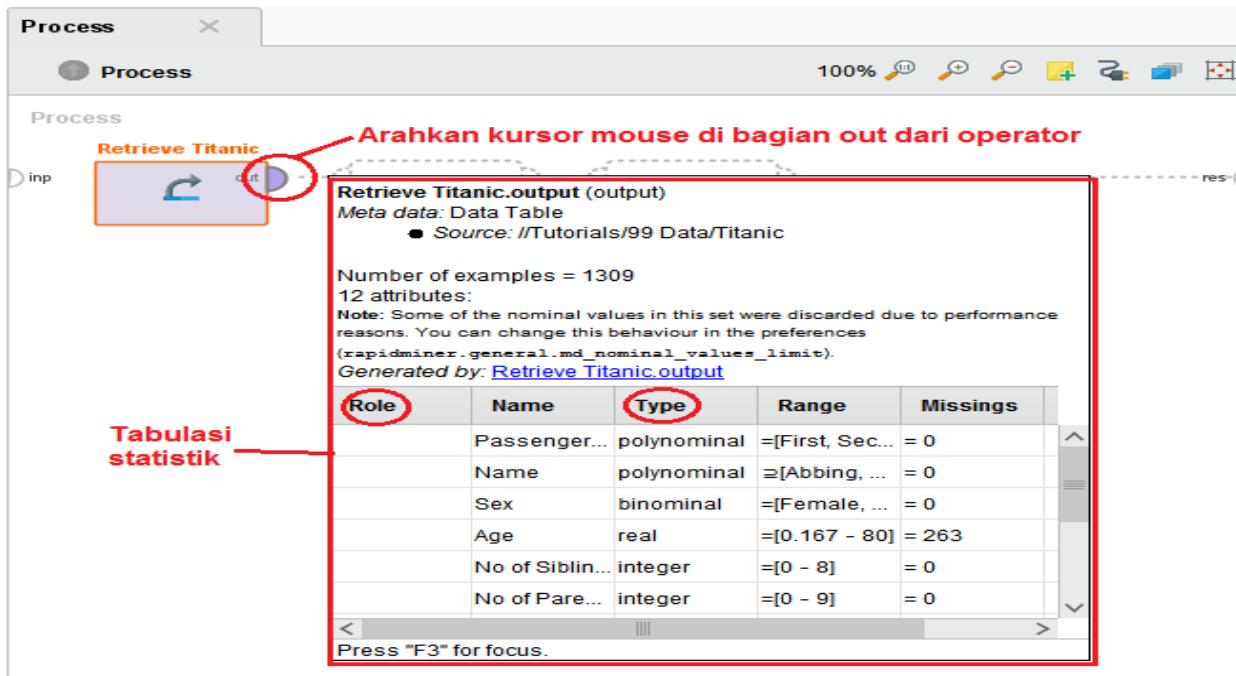
Untuk belajar membangun model prediktif, mari kembali ke data Titanic dari sebelumnya. Data ini agak sederhana dan tidak perlu banyak persiapan, tapi kita masih perlu menentukan kolom mana yang ingin kita prediksi! Berikut tahapan yang harus kita kerjakan.

**Tahap 1 :** Menambah detil produk ke transaksi :

1. Tarik data **Titanic** ke dalam proses.



2. Sekarang, arahkan mouse ke port **output** operator **Retrieve** dan tunggu sampai jendela kecil muncul dan menampilkan beberapa meta data tentang dataset **Titanic**. Beberapa informasi yang Anda tahu dari tab **Statistik**.

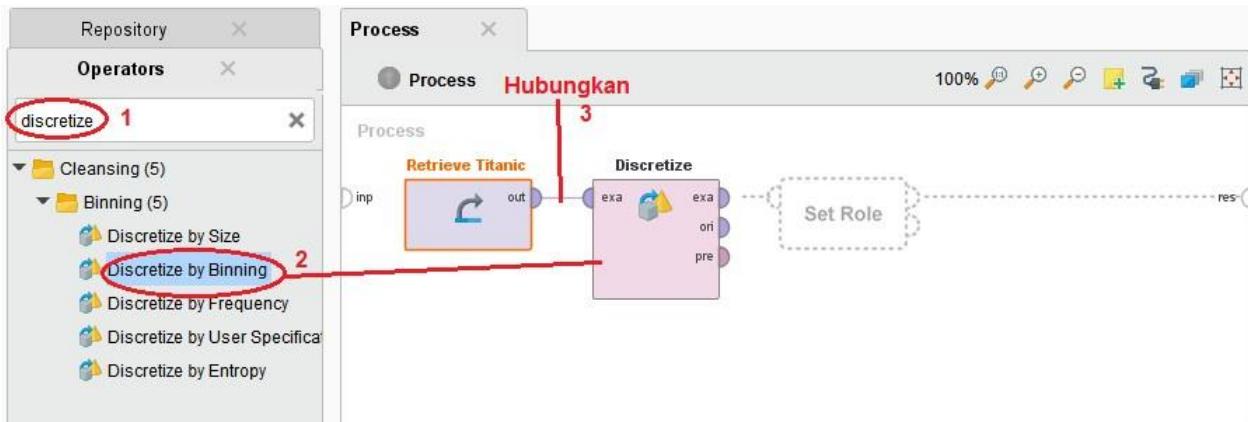


3. Perhatikan dua kolom **Role** dan **Type** di tabel di bagian bawah.

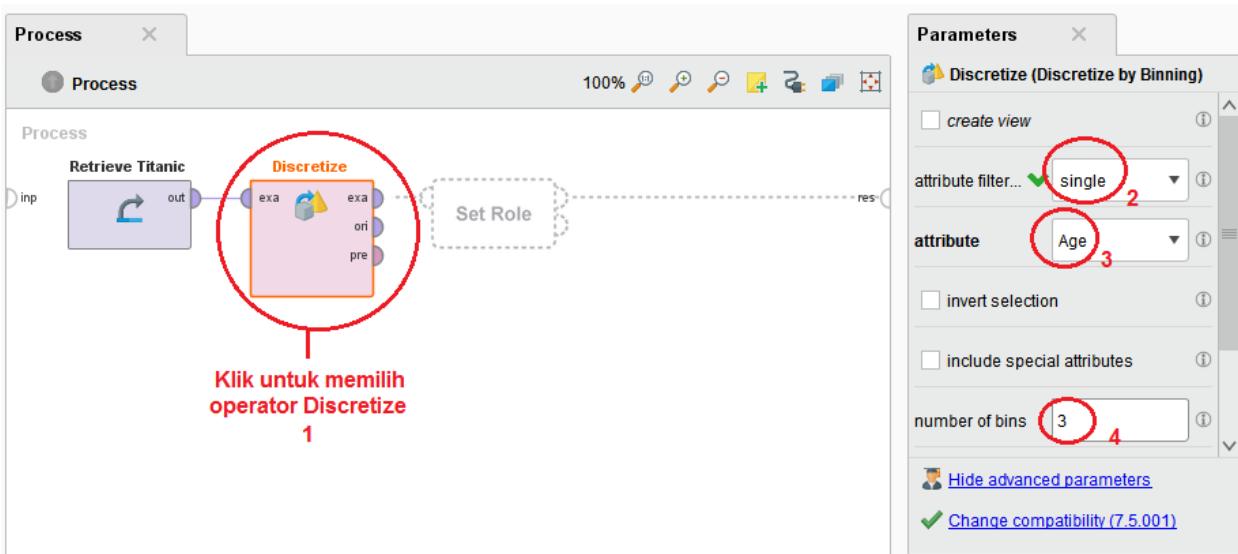
Setiap atribut memiliki tipe yang mendefinisikan nilai yang mungkin untuk kolom (misalnya, jika nilai bisa nominal atau numerik).

#### Tahap 2 : Mendefinisikan Tipe dan Peran Atribut :

1. Tambahkan **Discretize by Binning** operator dan hubungkan



2. Tetapkan jenis filter atribut menjadi single (yaitu Anda hanya mengerjakan salah satu atribut).



3. Ubah opsi **attribut** ke **Age** (Umur).

4. Ubah opsi **jumlah binning** menjadi **3**.

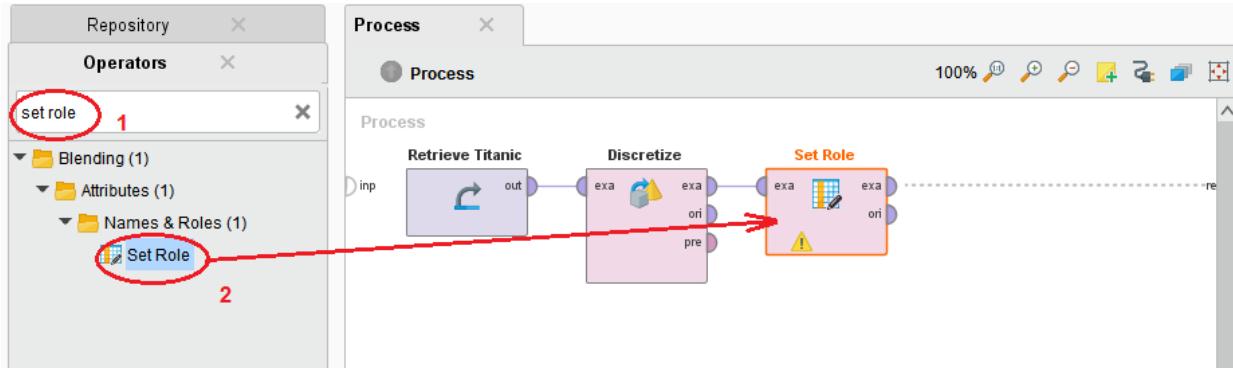
"Binning" adalah teknik umum untuk mengubah tipe dari numerik menjadi polinomial (nominal dengan lebih dari 2 nilai). Di sini, kita membuat tiga "binning" yang mencakup rangkaian nilai lengkap. Operator kemudian mengganti nilai numerik asli dengan nama bin yang dimiliknya.

Tahap 3 : Mendefinisikan kolom mana yang akan diprediksi :

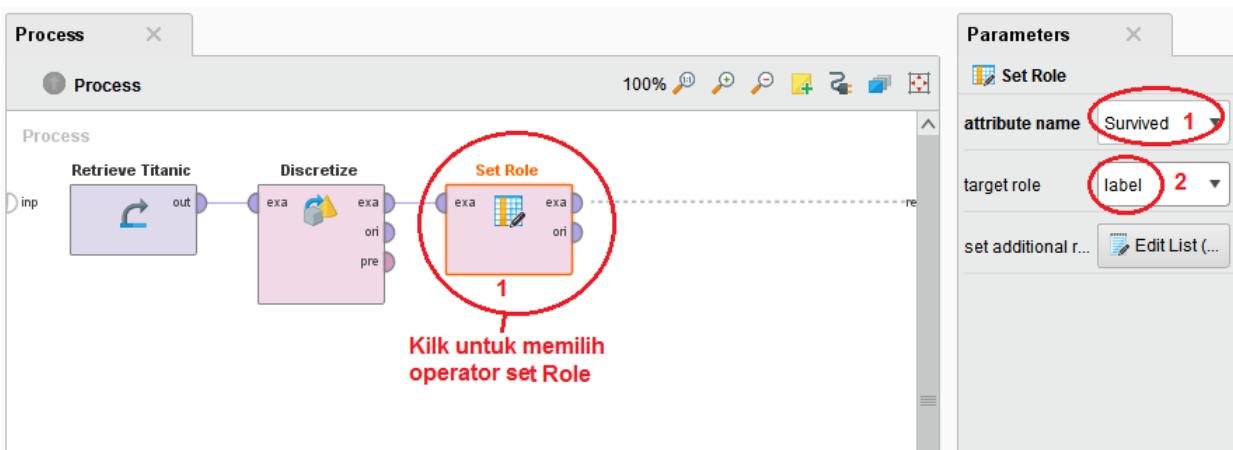
Pada RapidMiner, kita menyebut kolom yang harus diprediksi dengan nama label. Terkadang juga disebut sebagai target atau kelas.

Langkah :

1. Tambahkan operator **Set Role** dan hubungkan.



2. Pada parameter, untuk **nama atribut** ubah ke **Survived**, kemudian pada opsi **target role** ubah menjadi **label**.



### 3. Jalankan proses dan periksa hasilnya.

The screenshot shows the RapidMiner Studio Free interface with the 'Results' tab selected. A red circle highlights the play button in the toolbar. On the left, there's a 'Tutorials' panel with steps for changing types and roles, and an 'EXPLANATION' section about statistics. The main area displays a table titled 'ExampleSet (1309 examples, 1 special attribute, 11 regular attributes)'. The columns are Row No., Survived, Age, Passenger ..., Name, Sex, No of Sibling..., No of Parent..., Ticket Num..., and P... (partially visible). The 'Survived' column is highlighted with a red border. The 'Age' column is also highlighted with a red border. The data rows show various passenger details like name, sex, and ticket number.

Row No.	Survived	Age	Passenger ...	Name	Sex	No of Sibling...	No of Parent...	Ticket Num...	P...
1	Yes	range2 [26.778 - 53.389]	First	Alien, Miss. E...	Female	0	0	24160	21
2	Yes	range1 [-∞ - 26.778]	First	Allison, Mast...	Male	1	2	113781	15
3	No	range1 [-∞ - 26.778]	First	Allison, Miss. ....	Female	1	2	113781	15
4	No	range2 [26.778 - 53.389]	First	Allison, Mr. H...	Male	1	2	113781	15
5	No	range1 [-∞ - 26.778]	First	Allison, Mrs. ....	Female	1	2	113781	15
6	Yes	range2 [26.778 - 53.389]	First	Anderson, Mr....	Male	0	0	19952	26
7	Yes	range3 [53.389 - ∞]	First	Andrews, Mis....	Female	1	0	13502	77
8	No	range2 [26.778 - 53.389]	First	Andrews, Mr. ....	Male	0	0	112050	0
9	Yes	range2 [26.778 - 53.389]	First	Appleton, Mrs....	Female	2	0	11769	51
10	No	range3 [53.389 - ∞]	First	Artagaveitia, ...	Male	0	0	PC 17609	49
11	No	range2 [26.778 - 53.389]	First	Astor, Col. Jo...	Male	1	0	PC 17757	22
12	Yes	range1 [-∞ - 26.778]	First	Astor, Mrs. Jo...	Female	1	0	PC 17757	22
13	Yes	range1 [-∞ - 26.778]	First	Aubart, Mme. ....	Female	0	0	PC 17477	69
14	Yes	range1 [-∞ - 26.778]	First	Barber, Miss. ....	Female	0	0	19877	78
15	Yes	range3 [53.389 - ∞]	First	Barkworth, Mr....	Male	0	0	27042	30
16	No	?	First	Baumann, Mr....	Male	0	0	PC 17318	25
17			First	Brown, Mr. Goo...	Male	0	1	PC 17560	34

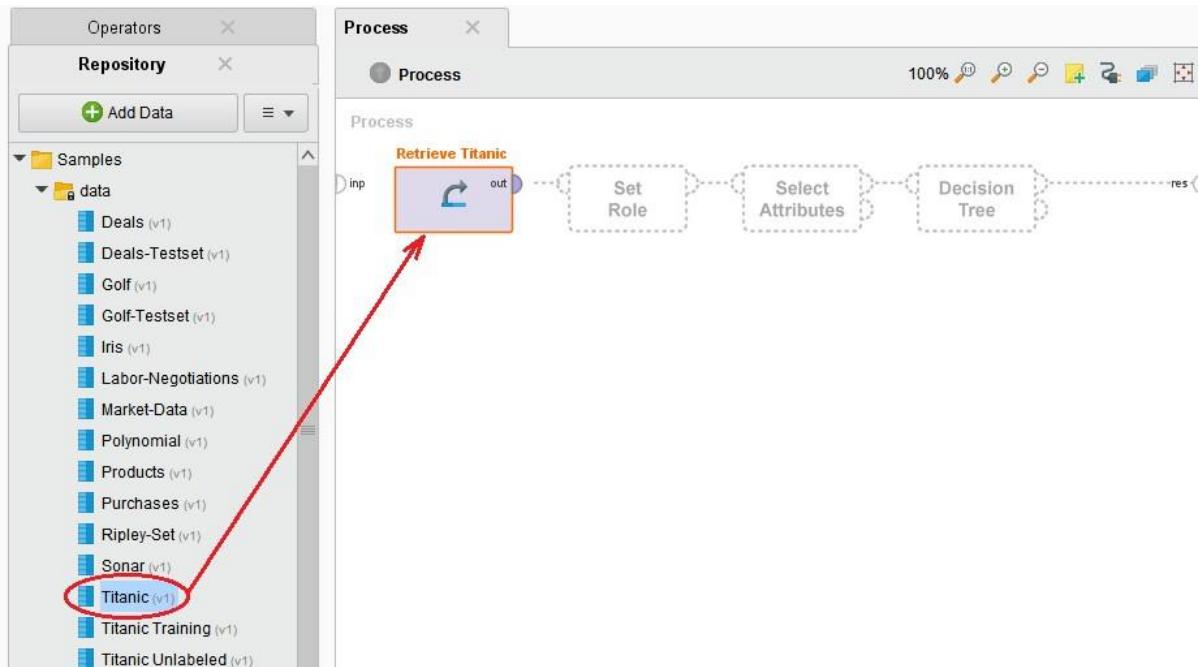
Lihat tab Statistik di tampilan **Results**. Perhatikan bahwa role **Survived** telah berubah menjadi label. Atribut **Age** juga memiliki tipe baru sekarang: nilai numerik untuk Age telah diganti dengan nama baru.

# Tutorial Belajar Rapidminer (Bagian 10) - Membuat Model Prediksi dengan Decision Tree

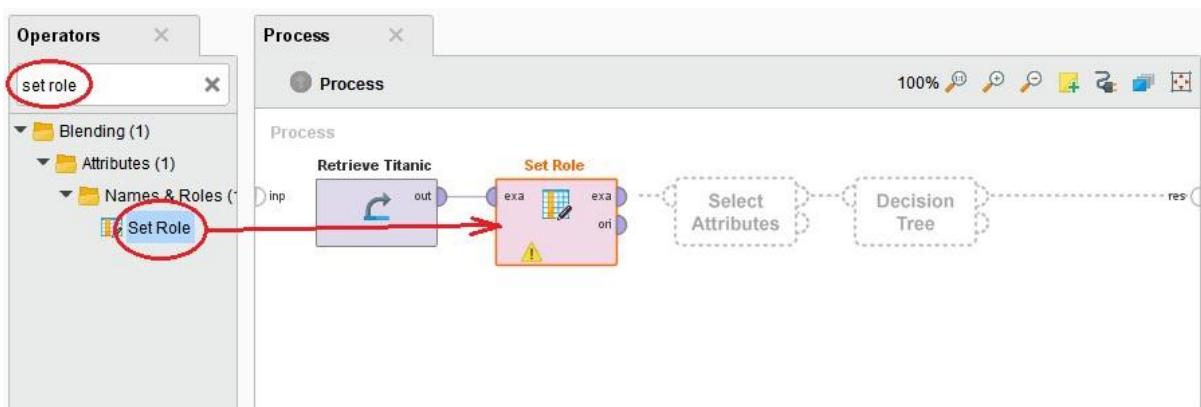
Dengan operator preprocessing yang telah kita bahas sejauh ini, kita dapat memadukan dan menyiapkan sebagian besar data set untuk membangun model prediktif. Dalam tutorial ini, kita akan menggunakan salah satu metode pembelajaran mesin yang paling banyak digunakan, yaitu Decision Tree, untuk memprediksi siapa yang akan bertahan dalam kecelakaan Titanic. Tentu saja kapal Titanic sudah tenggelam, jadi tidak ada yang dapat kita lakukan untuk menyelamatkan penumpangnya, namun kita tetap dapat menggunakan model ini untuk situasi yang sama dan membuat prediksi saat itu. Jadi perlukah kita membeli tiket kelas tiga saat bepergian dengan keluarga? Model akan menunjukkannya!

**Langkah 1.** Mengambil data Titanic.

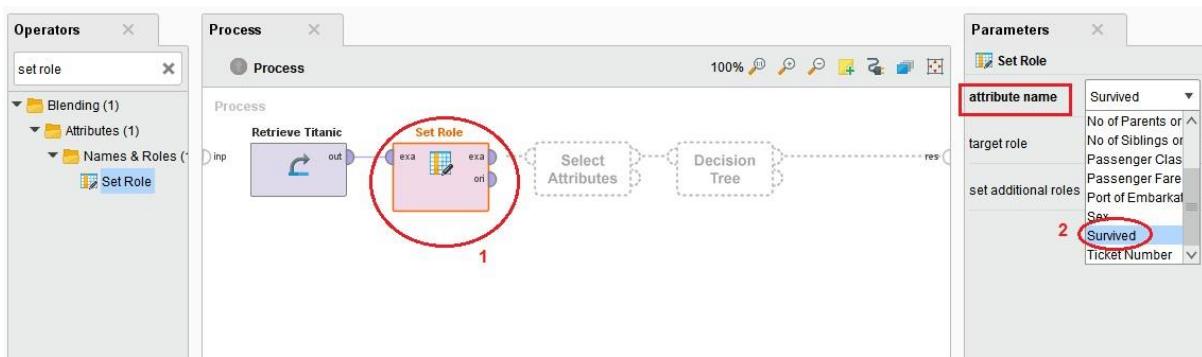
1. Seret data **Titanic** ke dalam jendela process.



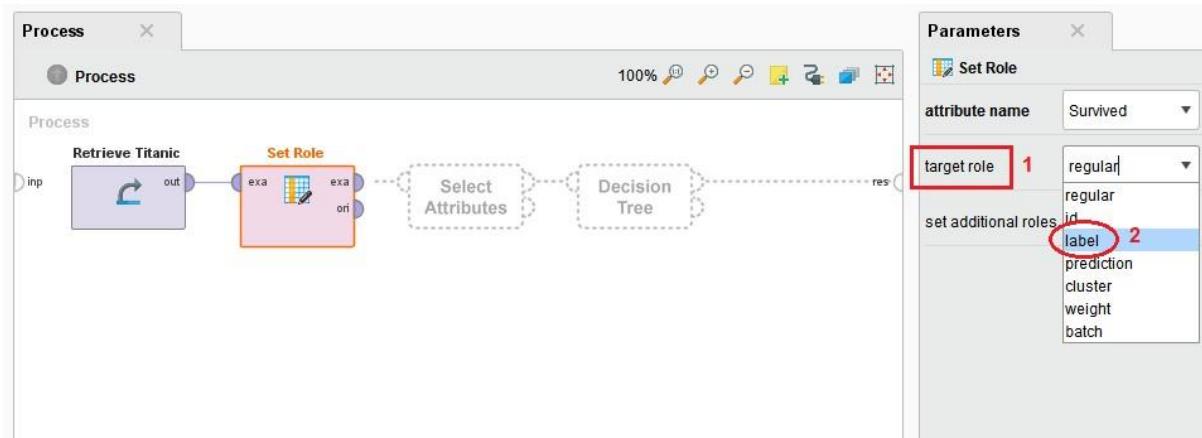
2. Tambahkan **Set Role**, hubungkan, dan konfigurasikan seperti yang sudah kita lakukan pada tutorial sebelumnya.



Ubah opsi attribut name menjadi **Survived**



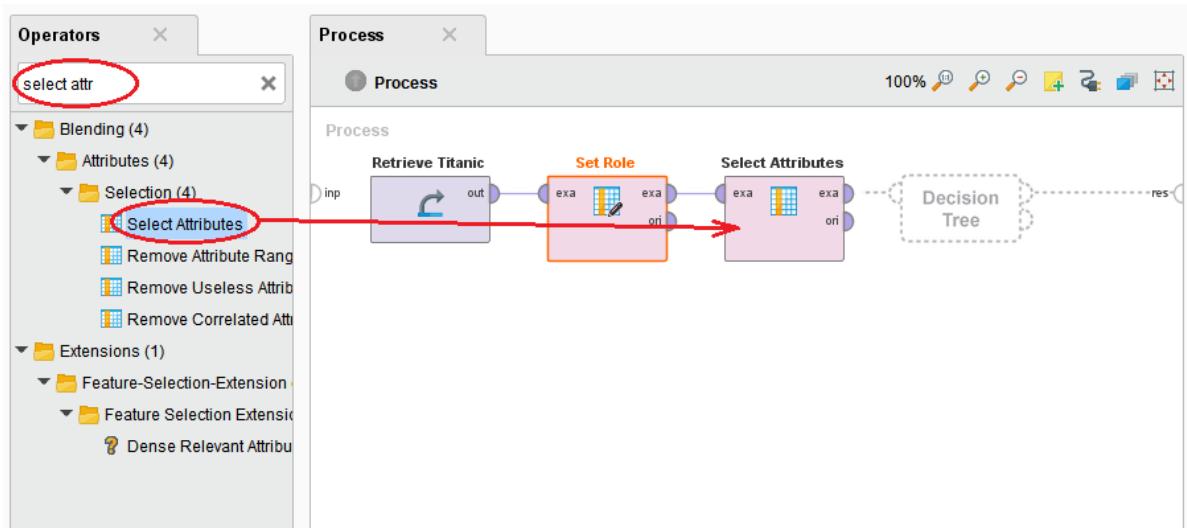
Ubah opsi **target role** menjadi **label**.



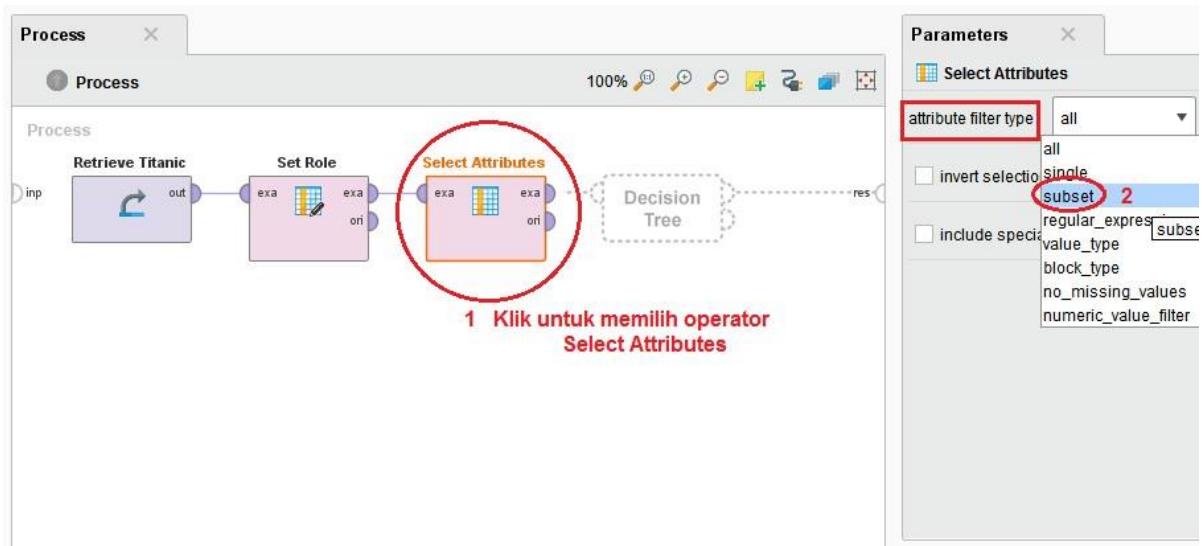
Ingatlah bahwa atribut dengan **target role**, label adalah yang ingin kita prediksi. Penting untuk mengatur label, karena apda metode pembelajaran mesin, seperti algoritma decision tree, menggunakan data yang ada dengan nilai label yang diketahui (data set pelatihan) untuk menemukan pola tersembunyi. Decision tree kemudian menciptakan prediksi dari pola tersebut dan menerapkannya ke data baru tanpa label yang dikenal (data set pengujian).

**Langkah 2.** Menghapus atribut yang tidak diperlukan.

Tambahkan operator **Select Atribut** ke proses dan hubungkan.

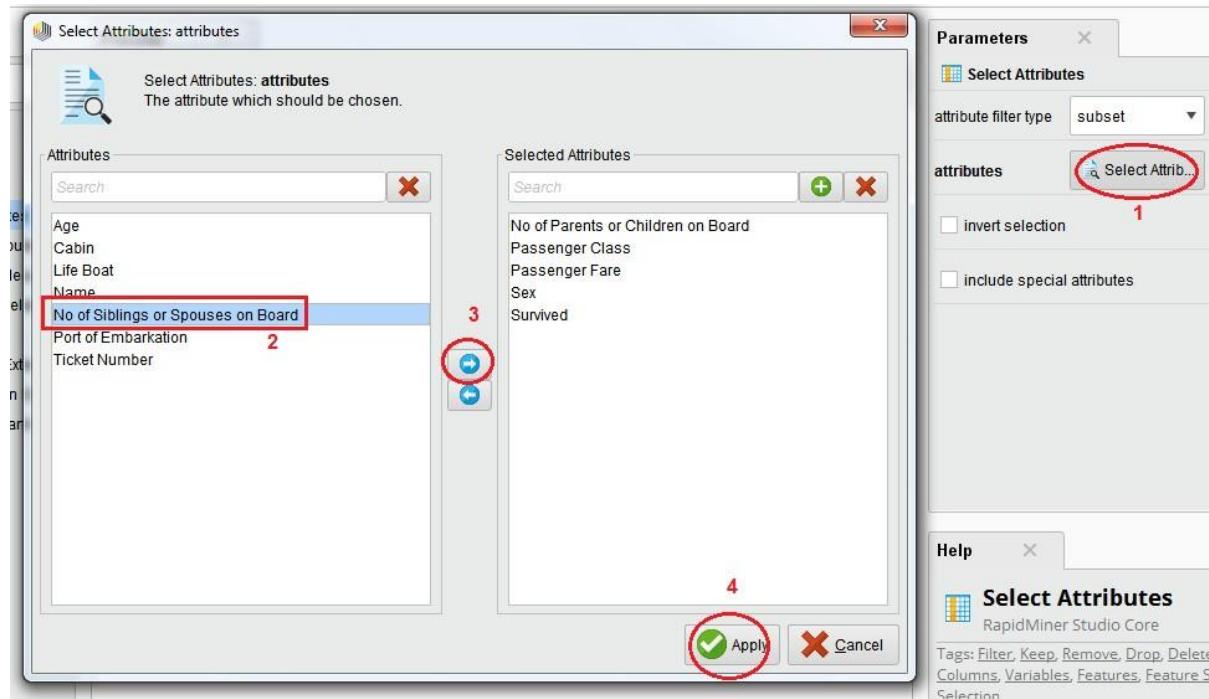


## Atur atribut filter type ke subset



klik **Select Atributes**.

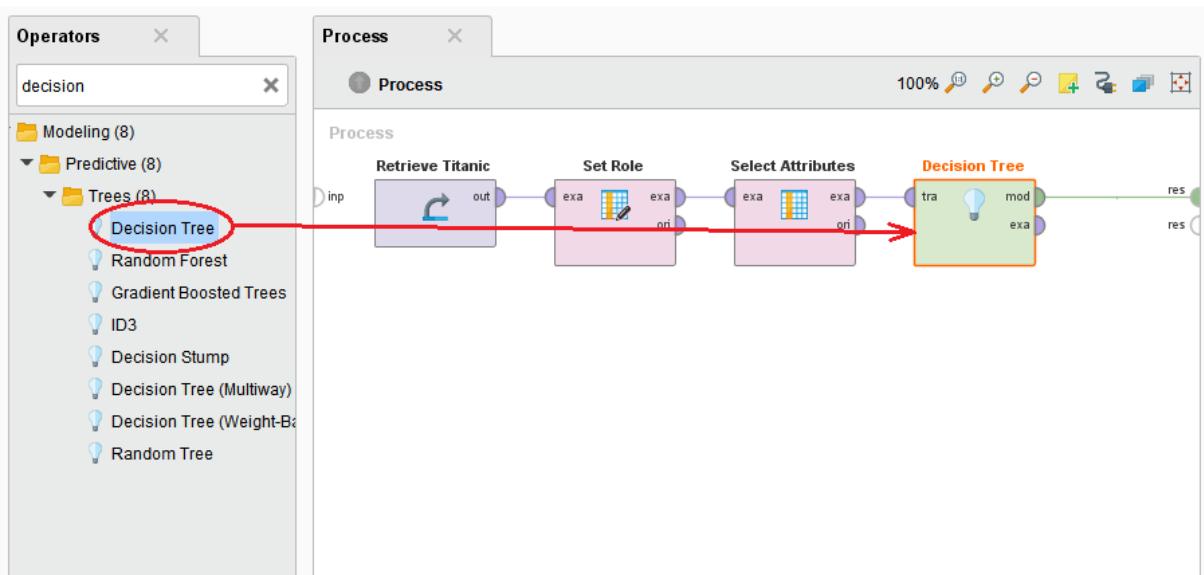
Dalam dialog yang dihasilkan, pilih *Survived, Sex, Passenger Class, Passenger Fare, dan No of ... parents, children, siblings, and spouses.*



Kita menghilangkan (tidak memilih) **Life Boat** karena penumpang yang naik pada Life Boat adalah orang yang kemungkinannya selamat. Menambahkan informasi Life Boat akan menyebabkan model praktis hanya bergantung pada informasi ini. Jadi pertanyaan yang tepat adalah : siapa yang berhasil naik Life Boat? Nama dan nomor tiket adalah jenis ID yang berbeda, jadi kita juga akan membuangnya.

**Langkah 3.** Membangun model pohon keputusan.

Seret pada operator **Decision Tree**, hubungkan inputnya, dan hubungkan port output "mod" ke port "res".

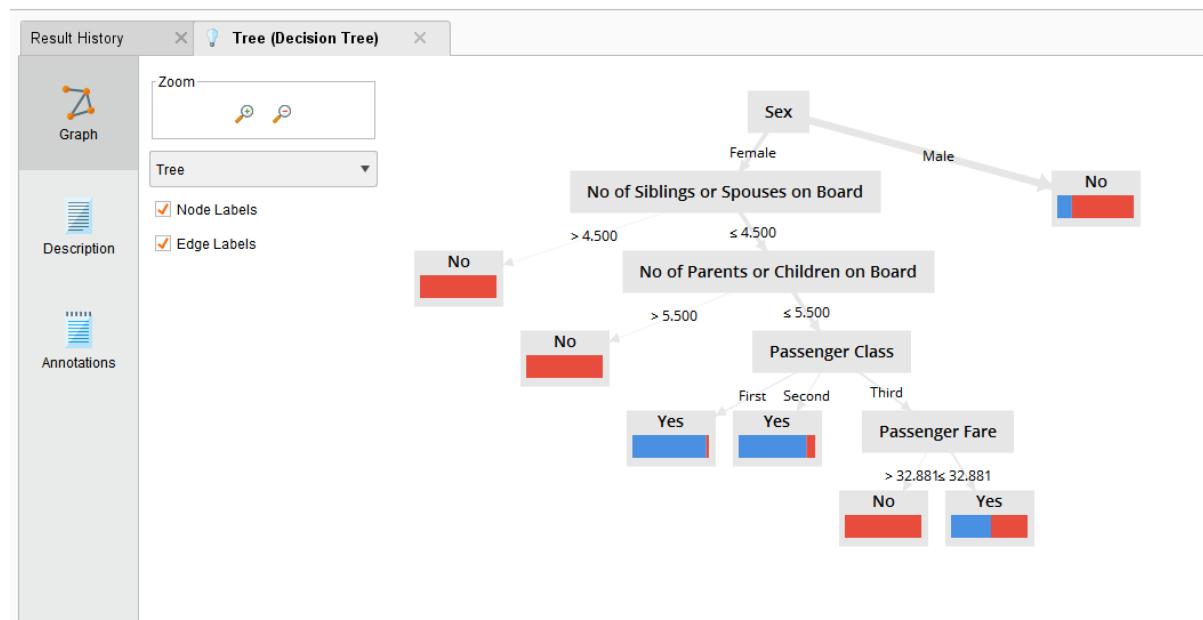


Perhatikan bahwa koneksi data berwarna biru sedangkan koneksi modelnya berwarna hijau. Ini membantu untuk menemukan dan memverifikasi port koneksi yang benar dengan mudah.

Jalankan prosesnya.



Periksa hasil prediksi yang dihasilkan oleh model pohon keputusan.



Hal yang menarik untuk dilihat bahwa bagi penumpang wanita, serta besar kecilnya ukuran keluarga memegang peranan lebih penting dibandingkan faktor kelas penumpang. Pola perilaku ini tidak bisa dideteksi untuk penumpang pria. Secara umum, pria memiliki kemungkinan lebih rendah untuk bertahan hidup ("wanita dan anak-anak lebih diutamakan!").

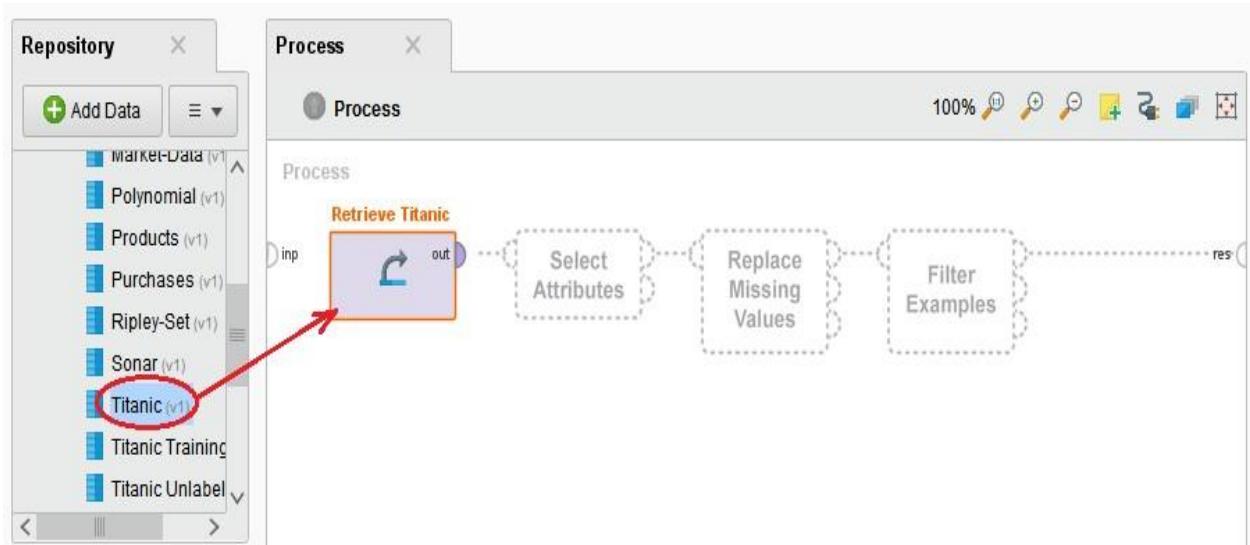
# Tutorial Belajar Rapidminer (Bagian 11) - Mengatasi Atribut yang Hilang

Dalam melakukan kegiatan analis, data preprocessing kemungkinan akan menghabiskan sebagian besar waktu kita. Ada dua kelompok umum dalam penanganan data: pencampuran dan pembersihan. Pada tutorial berikut ini selanjutnya akan dibahas beberapa operasi terpenting untuk pembersihan data.

Pencampuran adalah mengubah kumpulan data dari satu keadaan menuju keadaan lain atau menggabungkan beberapa kumpulan data. Pembersihan adalah segala kegiatan tentang memperbaiki data, sehingga pemodelan akan memberikan hasil yang lebih baik. Kita kembali melihat data Titanic lagi. Kita masih ingat sebelumnya bahwa data Titanic berisi beberapa nilai yang hilang, yang digambarkan dalam bentuk tanda tanya pada tampilan data. Anda juga dapat melihat jumlah nilai yang hilang untuk setiap atribut dalam tampilan Statistik. Nilai yang hilang merupakan masalah bagi banyak pemrosesan preprocessing atau pemodelan data. Tutorial ini akan menunjukkan kepada kita tentang pendekatan yang paling umum untuk menghapus atau mengganti nilai yang hilang. Berikut langkahnya :

Langkah 1 : Mempersiapkan data.

1. Tarik data **Titanic** ke dalam proses.



2. Arahkan cursor mouse ke port **out** dan tunggu sampai tooltip menampilkan data meta.

Process

**Retrieve Titanic**

Retrieves the Titanic dataset from the local repository.

**Retrieval Information:**

- Source: /Tutorials/99 Data/Titanic

**Number of examples = 1309**

**12 attributes:**

Note: Some of the nominal values in this set were discarded due to performance reasons. You can change this behaviour in the preferences (rapidminer.general.md\_nominal\_values\_limit).

Generated by: [Retrieve Titanic.output](#)

Role	Name	Type	Range	Missing
	Passenger...	polynominal	= [First, Sec...]	= 0
	Name	polynominal	= [Abbing, ...]	= 0
	Sex	binominal	= [Female, ...]	= 0
	Age	real	= [0.167 - 80]	= 263
	No of Siblin...	integer	= [0 - 8]	= 0
	No of Pare...	integer	= [0 - 9]	= 0

Press "F3" for focus.

3. Tekan F3 saat tooltip ditampilkan. Sekarang berubah menjadi jendela dan Anda dapat menggulir ke bawah untuk melihat informasi tentang semua kolom. Periksa kolom dengan nilai yang hilang.

Retrieve Titanic.output (output)

Meta data: Data Table

Source: /Tutorials/99 Data/Titanic

Number of examples = 1309

12 attributes:

Note: Some of the nominal values in this set were discarded due to performance reasons. You can change this behaviour in the preferences (rapidminer.general.md\_nominal\_values\_limit).

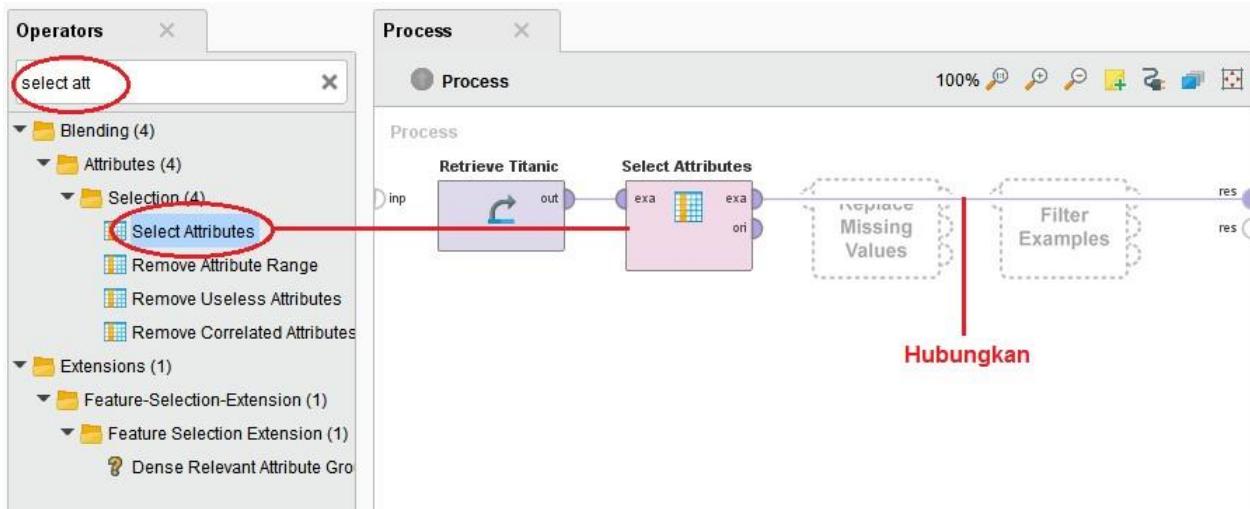
Generated by: [Retrieve Titanic.output](#)

Role	Name	Type	Range	Missing	Comm
	Ticket Num...	polynominal	= [110152, ...]	= 0	
	Passenger...	numeric	= [0 - 512.3...]	= 1	
	Cabin	polynominal	= [A10, A11...]	= 1014	
	Port of Em...	polynominal	= [Cherbour...]	= 2	
	Life Boat	polynominal	= [1, 10, 11, ...]	= 823	
	Survived	binominal	= [No, Yes]	= 0	

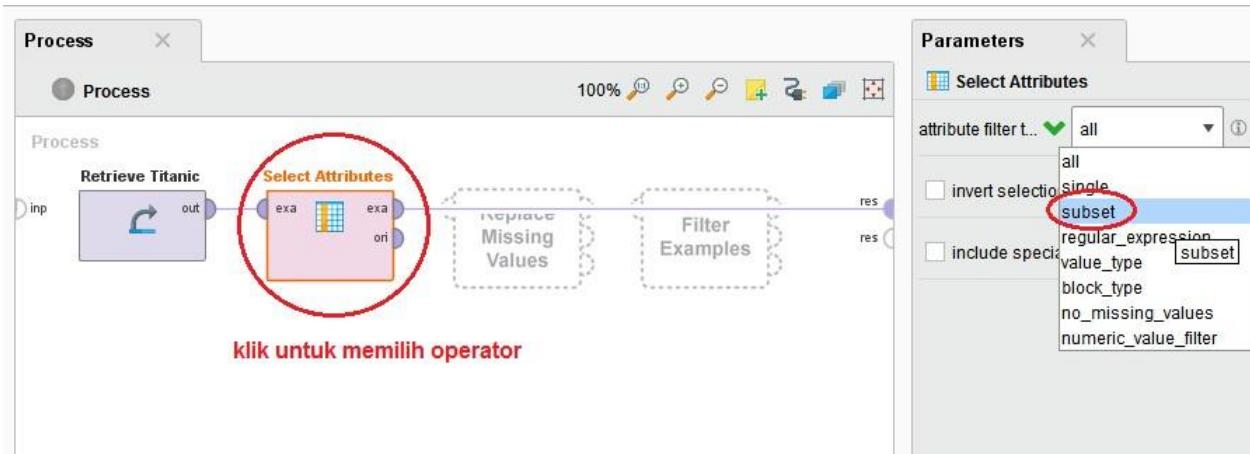
Ada lima atribut dengan nilai yang hilang: **Age**, **Passenger Fare**, **Cabin**, **Port of Embarkation**, dan **Life Boat**. **Cabin** dan **Life Boat** memiliki nilai yang paling hilang

Langkah 2. Hapus atribut yang terlalu banyak hilang.

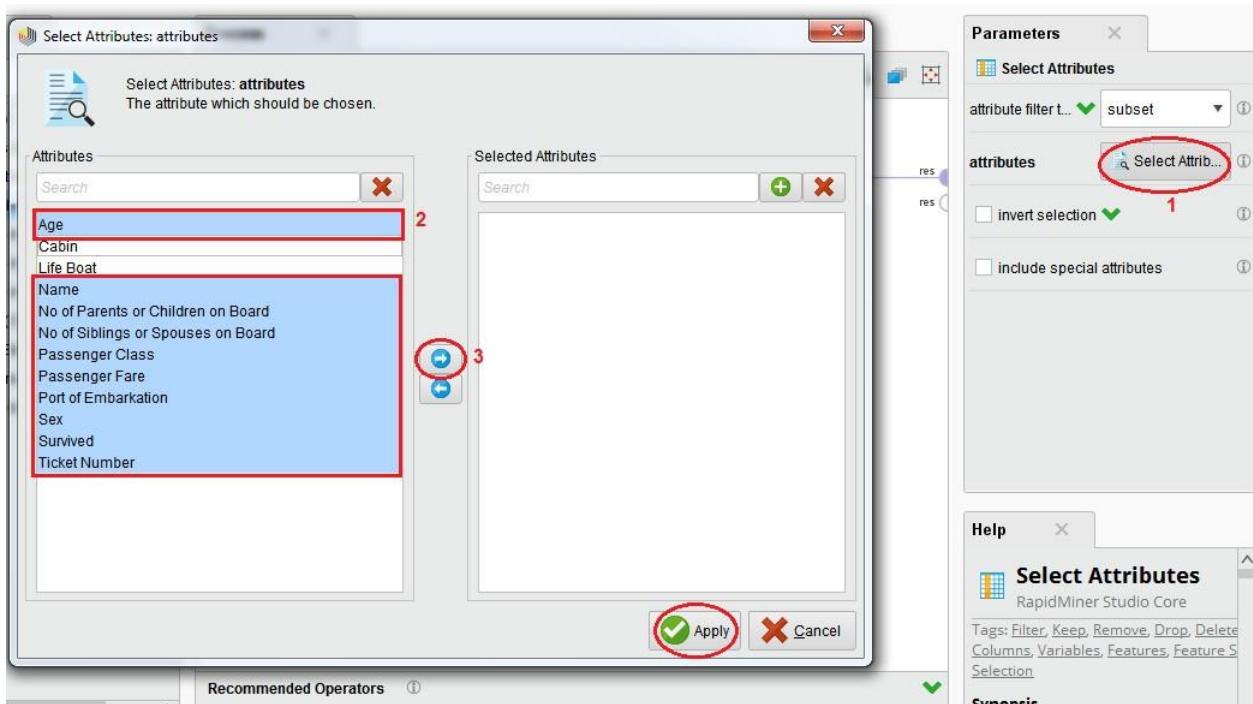
1. Tambahkan operator baru **Select Attributes**.



2. Sambungkan operator baru ke operator Retrieve dan output ke port hasil "res" di sebelah kanan.



3. Di Parameters, ubah tipe filter atribut menjadi **Subset** dan pilih semua atribut terkecuali **Cabin** dan **Life Boat**. Ini berarti keduanya akan dilepas oleh operator.



4. Jalankan prosesnya. Klik pada tab Statistik dan periksa atribut mana dengan nilai yang hilang yang tersisa.

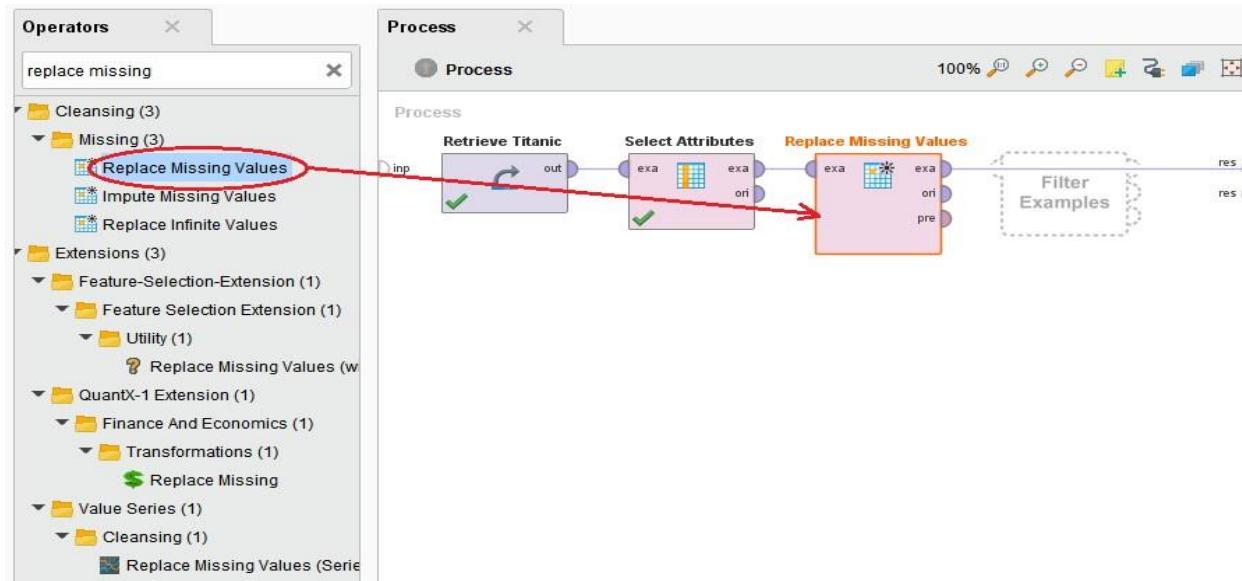
The screenshot shows the 'Results' view in RapidMiner. The top navigation bar includes icons for back, forward, and play (circled with a red number '1'), views (Design, Results), and help. The main area displays the 'ExampleSet (Select Attributes)' table. The table has 13 columns: Row No., Passenger..., Name, Sex, Age, No of Sibling..., No of Parent..., Ticket Num..., Passenger F..., and Port of Emb.... The table contains 1309 rows of data. The left sidebar features tabs for Data, Statistics, Charts, Advanced Charts, and Annotations. The 'Statistics' tab is selected. A filter at the top right of the table shows 'Filter (1,309 / 1,309 examples): all'.

Row No.	Passenger...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Port of Emb...
1	First	Allen, Miss. E...	Female	29	0	0	24160	211.338	Southampton
2	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	Southampton
3	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	Southampton
4	First	Allison, Mr. H...	Male	30	1	2	113781	151.550	Southampton
5	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	Southampton
6	First	Anderson, Mr....	Male	48	0	0	19952	26.550	Southampton
7	First	Andrews, Mis...	Female	63	1	0	13502	77.958	Southampton
8	First	Andrews, Mr. ...	Male	39	0	0	112050	0	Southampton
9	First	Appleton, Mrs...	Female	53	2	0	11769	51.479	Southampton
10	First	Artagaveitia, ...	Male	71	0	0	PC 17609	49.504	Cherbourg
11	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525	Cherbourg
12	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	Cherbourg
13	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	Cherbourg
14	First	Barber, Miss. ...	Female	26	0	0	19877	78.850	Southampton
15	First	Barkworth, Mr...	Male	80	0	0	27042	30	Southampton
16	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925	Southampton

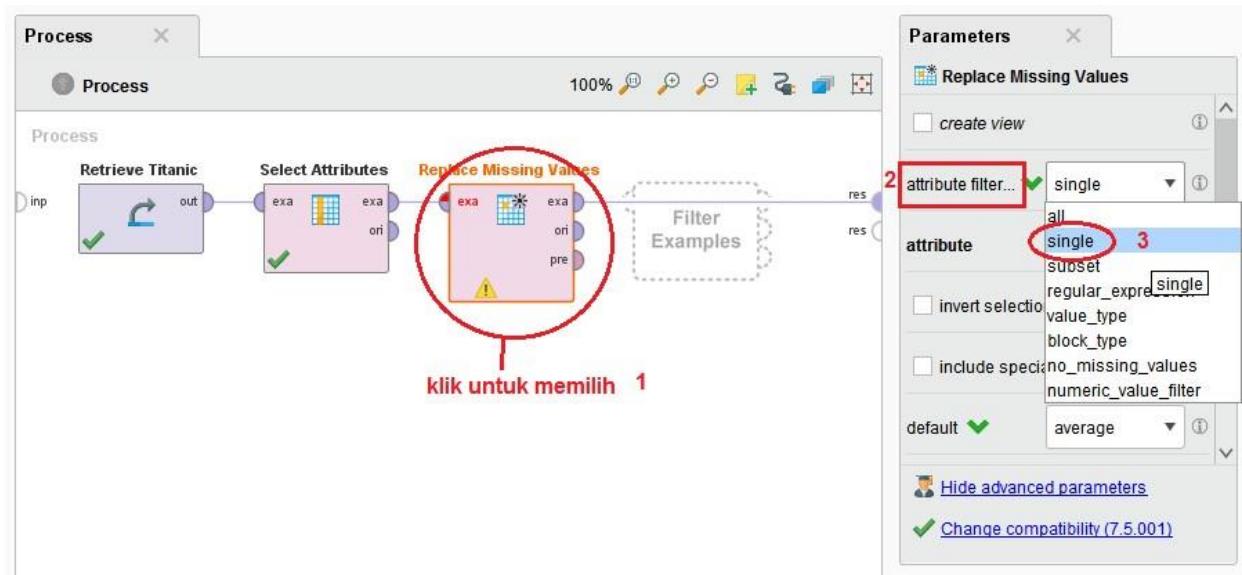
Kita menghapus Cabin karena hampir semua nilai di kolom ini hilang dan nilai yang tersisa mungkin tidak mengandung banyak informasi bermanfaat. Kita menghilangkan Life boat dengan alasan yang sama dan karena atribut ini sangat berkorelasi dengan label. Atribut Umur juga memiliki sejumlah nilai hilang yang signifikan, namun kita akan menangani hal ini dengan menggunakan metode yang berbeda pada tutorial berikutnya.

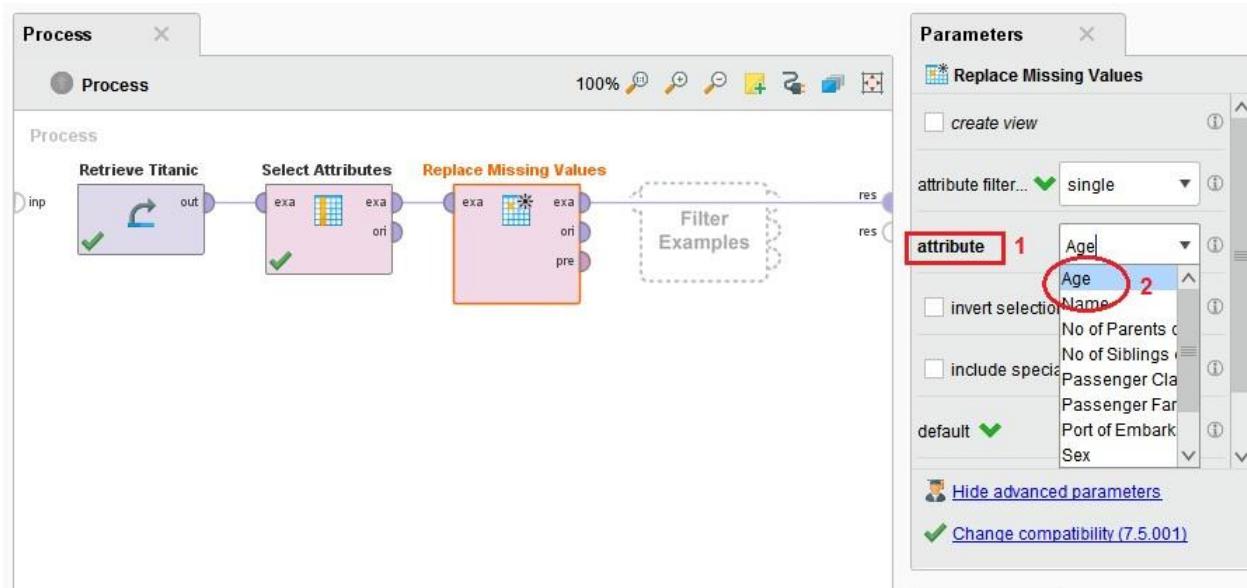
### Langkah 3 Mengganti nilai yang hilang

1. Cari operator **Replace Missing Values** dan menambahkannya ke proses. Kita dapat menjatuhkannya pada jalur koneksi antara Select Attributes dan port res (gerakkan pointer mouse sampai sambungan di highlight (sorot) sebelum kita menambahkan operator baru). Dengan cara ini kita tidak perlu secara manual menghubungkan kembali operator.



2. Pada Parameter dari operator ini, gunakan satu untuk jenis **filter atribut** dan pilih **Age for attribute**.





3. Jalankan prosesnya lagi.

The screenshot shows the KNIME results viewer. The 'ExampleSet (Replace Missing Values)' tab is active. On the left, there are tabs for 'Result History', 'Data' (highlighted with a red box), 'Statistics', 'Charts', 'Advanced Charts', and 'Annotations'. The 'Data' tab shows a table with 1309 rows of passenger data. The 'Age' column is highlighted with a red box. The table includes columns: Row No., Age, Passenger Class, Name, Sex, No of Sibling..., No of Parent..., Ticket Number, Passenger F..., and Port of Embark. A red circle highlights the 'Age' column in the table header.

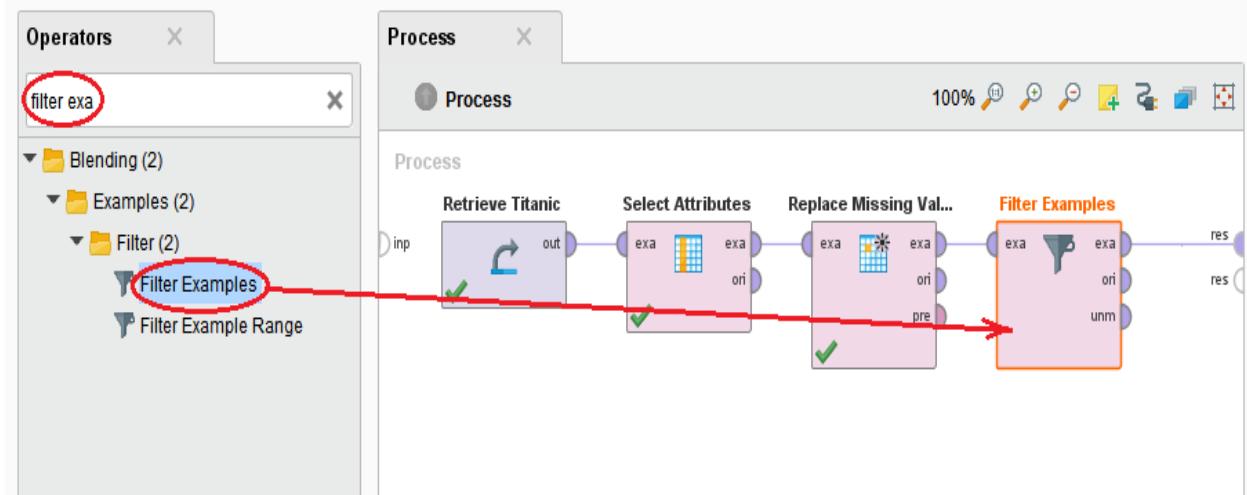
Row No.	Age	Passenger ...	Name	Sex	No of Sibling...	No of Parent...	Ticket Numbr...	Passenger F...	Port of Emb...
1	29	First	Allen, Miss. E...	Female	0	0	24160	211.338	Southampton
2	0.917	First	Allison, Mast...	Male	1	2	113781	151.550	Southampton
3	2	First	Allison, Miss. ...	Female	1	2	113781	151.550	Southampton
4	30	First	Allison, Mr. H...	Male	1	2	113781	151.550	Southampton
5	25	First	Allison, Mrs. ...	Female	1	2	113781	151.550	Southampton
6	48	First	Anderson, Mr....	Male	0	0	19952	26.550	Southampton
7	63	First	Andrews, Mis...	Female	1	0	13502	77.958	Southampton
8	39	First	Andrews, Mr. ...	Male	0	0	112050	0	Southampton
9	53	First	Appleton, Mrs...	Female	2	0	11769	51.479	Southampton
10	71	First	Artagaveytlia, ...	Male	0	0	PC 17609	49.504	Cherbourg
11	47	First	Astor, Col. Jo...	Male	1	0	PC 17757	227.525	Cherbourg
12	18	First	Astor, Mrs. Jo...	Female	1	0	PC 17757	227.525	Cherbourg
13	24	First	Aubart, Mme. ...	Female	0	0	PC 17477	69.300	Cherbourg
14	26	First	Barber, Miss. ...	Female	0	0	19877	78.850	Southampton
15	80	First	Barkworth, Mr...	Male	0	0	27042	30	Southampton
16	29.881	First	Baumann, Mr...	Male	0	0	PC 17318	25.925	Southampton

Anda akan melihat di tab Statistik bahwa hanya beberapa sampel dengan nilai yang hilang yang tersisa. Ketika kita menjalankan proses ini, nilai pada atribut **Age** yang hilang telah diganti dengan nilai rata-rata dari **Age**. Cara ini adalah teknik umum untuk menangani banyak nilai yang hilang pada sebuah atribut. Karena hanya ada sedikit nilai yang hilang, kita dapat menyaring sampel ini dengan aman dari dataset.

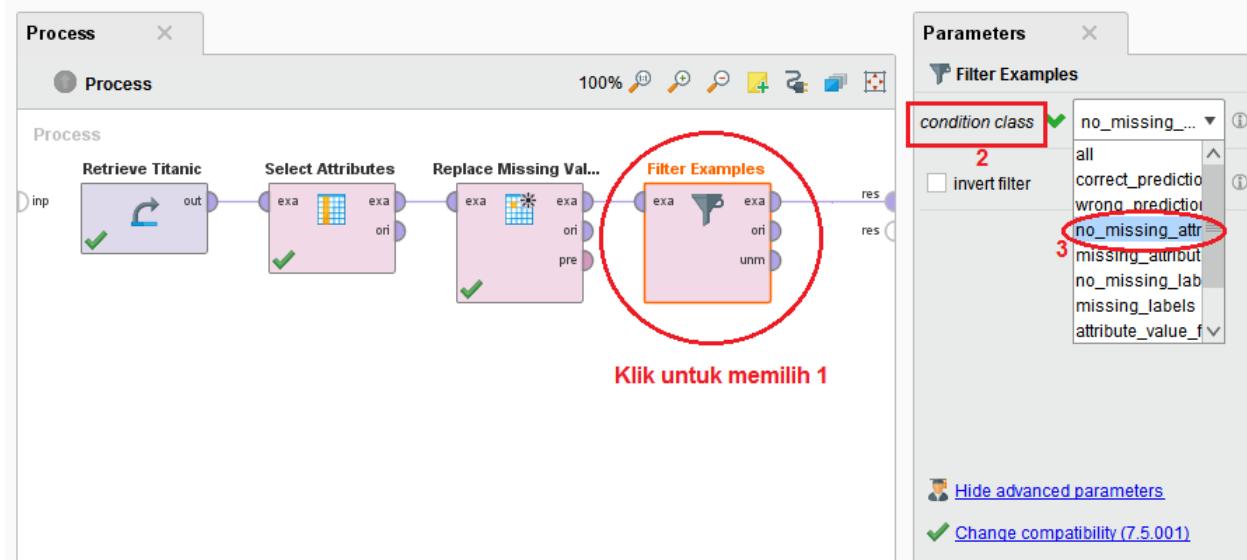
Langkah 4 .Menghapus sampel dengan nilai yang hilang.

Kita telah menggunakan operator Filter Example sebelumnya, tapi kali ini kita akan menggunakan beberapa pengaturan lanjutan untuk menghapus sampel di mana memiliki nilai atribut yang hilang.

1. Cari **Filter Example** dan drop pada jalur koneksi ke port **result**. Atau kita bisa menghubungkan operator secara manual.



2. Perhatikan link di bagian bawah panel Parameter yang menunjukkan / menyembunyikan parameter lanjutan. Klik pada **Show advanced parameters** untuk menampilkan semua parameter operator.



3. Parameter baru akan muncul. Atur kondisi class ke **no\_missing\_attributes**.
4. Jalankan prosesnya lagi.

Result History    ExampleSet (Filter Examples)

Data      Statistics      Charts      Advanced Charts      Annotations

ExampleSet (1306 examples, 0 special attributes, 10 regular attributes)      Filter (1,306 / 1,306 examples): all

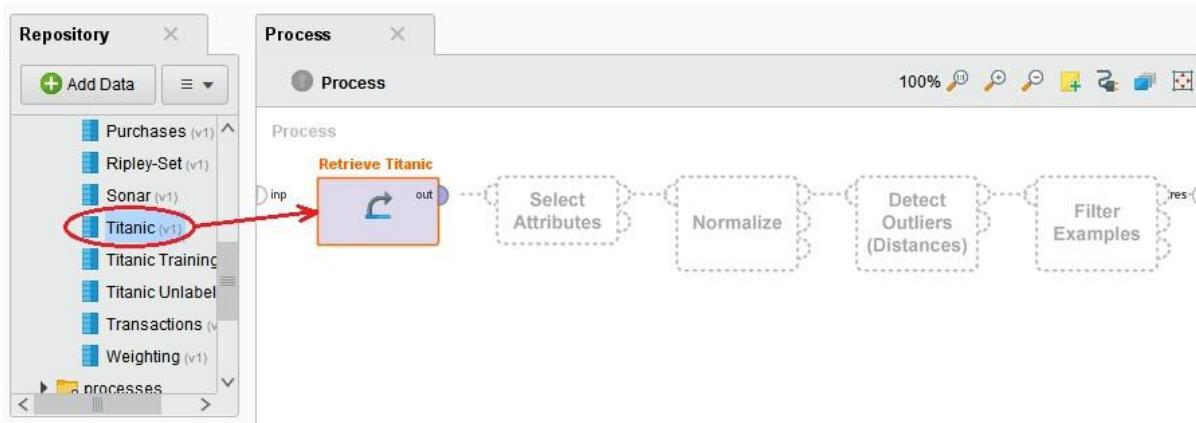
Row No.	Age	Passenger ...	Name	Sex	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Port of Emb...
1	29	First	Allen, Miss. E...	Female	0	0	24160	211.338	Southampton
2	0.917	First	Allison, Mast...	Male	1	2	113781	151.550	Southampton
3	2	First	Allison, Miss. ...	Female	1	2	113781	151.550	Southampton
4	30	First	Allison, Mr. H...	Male	1	2	113781	151.550	Southampton
5	25	First	Allison, Mrs. ...	Female	1	2	113781	151.550	Southampton
6	48	First	Anderson, Mr...	Male	0	0	19952	26.550	Southampton
7	63	First	Andrews, Mis...	Female	1	0	13502	77.958	Southampton
8	39	First	Andrews, Mr. ...	Male	0	0	112050	0	Southampton
9	53	First	Appleton, Mrs...	Female	2	0	11769	51.479	Southampton
10	71	First	Artagaveytia, ...	Male	0	0	PC 17609	49.504	Cherbourg
11	47	First	Astor, Col. Jo...	Male	1	0	PC 17757	227.525	Cherbourg
12	18	First	Astor, Mrs. Jo...	Female	1	0	PC 17757	227.525	Cherbourg
13	24	First	Aubart, Mme. ...	Female	0	0	PC 17477	69.300	Cherbourg
14	26	First	Barber, Miss. ...	Female	0	0	19877	78.850	Southampton

# Tutorial Belajar Rapidminer (Bagian 12) - Melakukan Normalisasi dan Deteksi Outlier

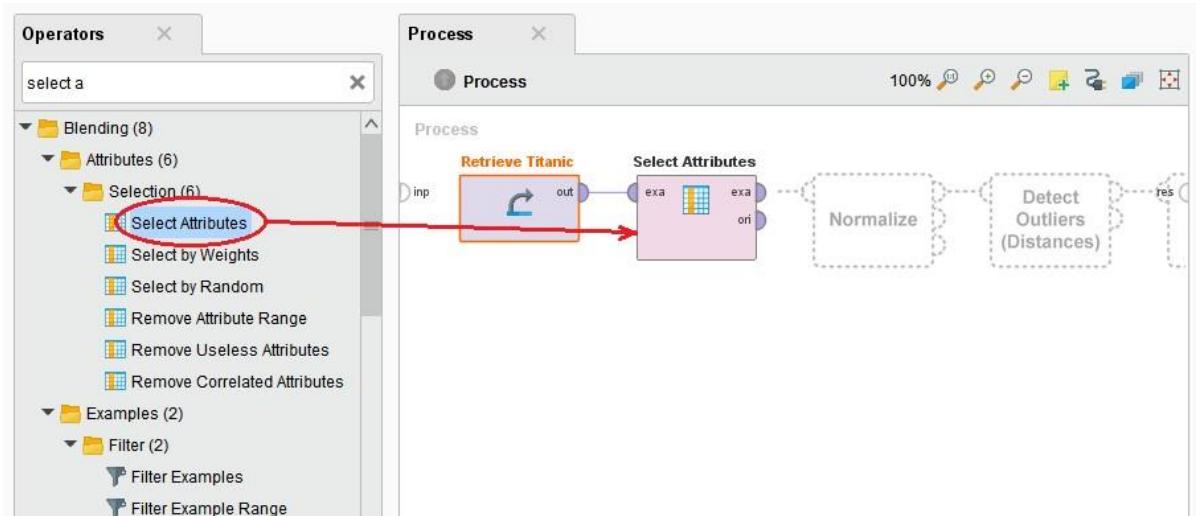
Langkah penting lain pada pembersihan data adalah untuk mengidentifikasi kasus yang tidak biasa dan menghapusnya dari dataset. Dalam beberapa situasi, outlier sendiri mungkin merupakan kasus yang paling menarik (misalnya mendeteksi penyimpangan transaksi kartu kredit), namun dalam kebanyakan kasus, outlier hanyalah hasil dari pengukuran yang salah dan harus dibuang dari dataset. Inilah yang akan kita bahas dalam tutorial ini.

**Langkah 1** : menyiapkan data.

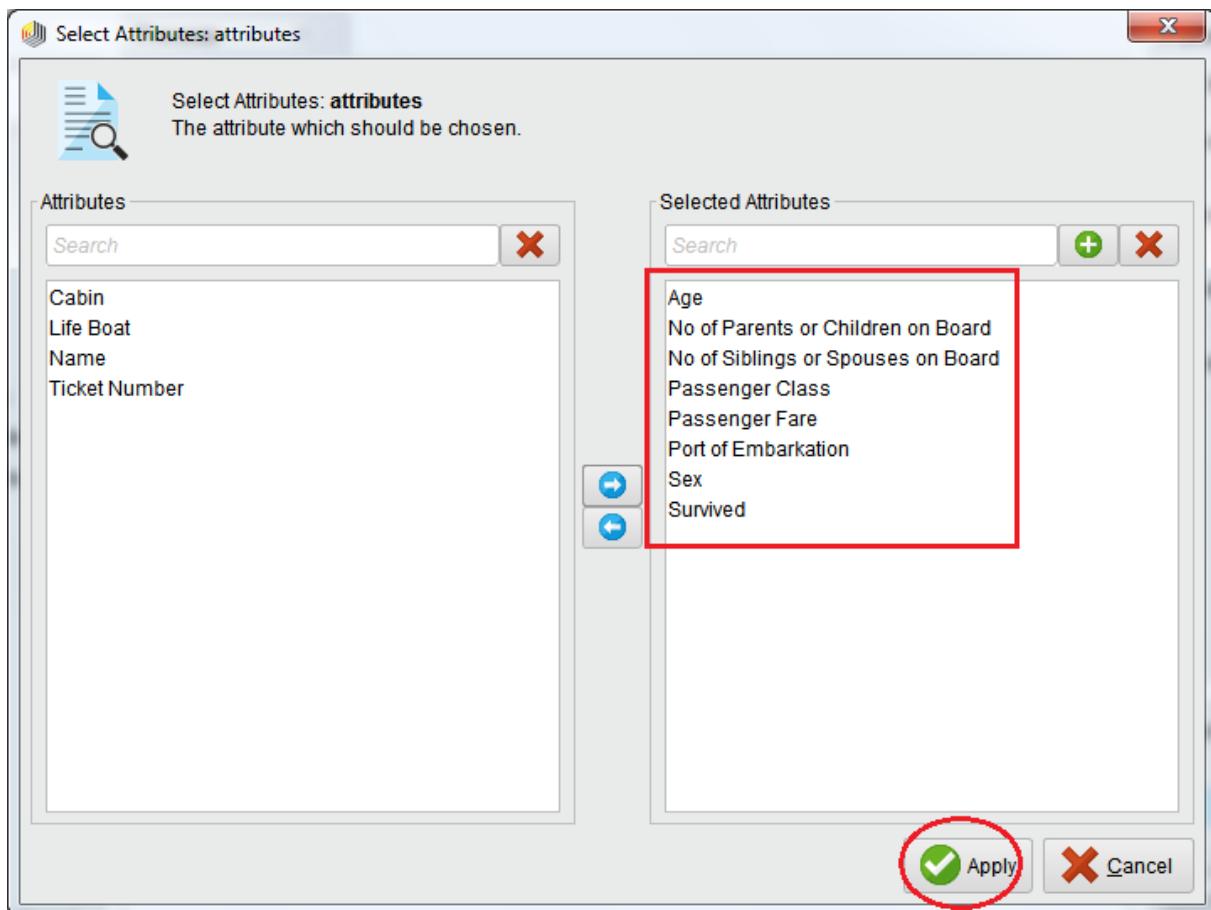
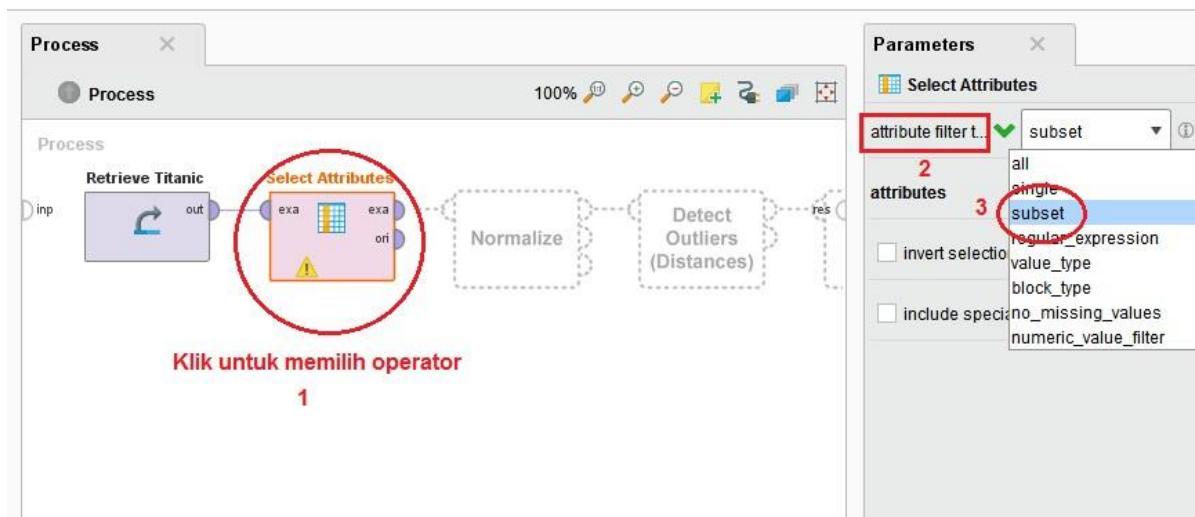
1. Tarik data **Titanic** ke dalam proses.



2. Menambahkan operator **Select Attributes** dan hubungkan.



2. Ubah Parameters sehingga Kita membuat **Cabin**, **Life Boat**, **Name**, dan **Ticket Number**.



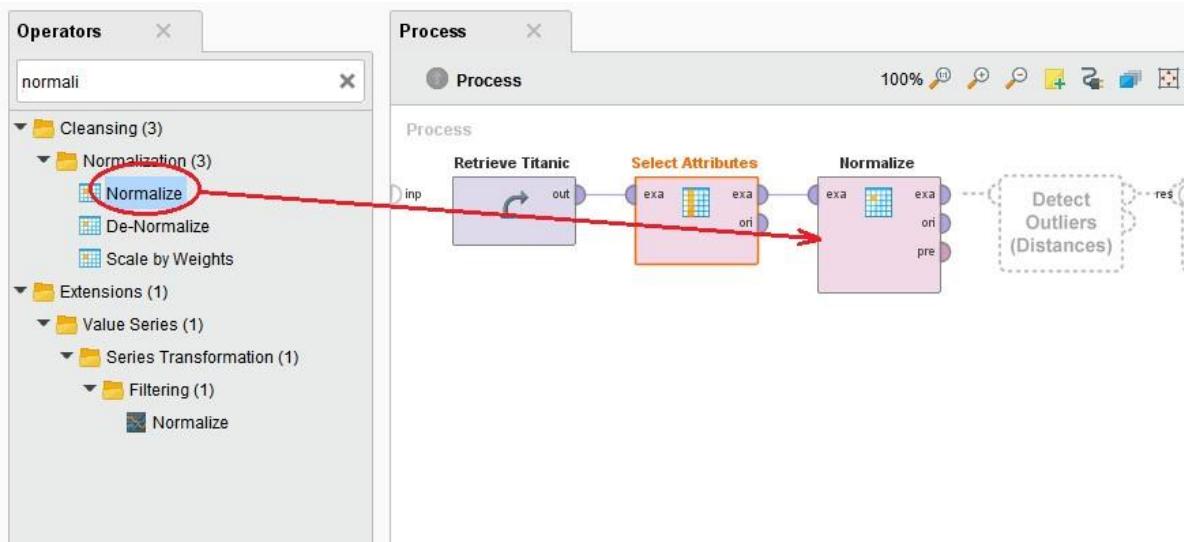
Hasilnya akan menjadi kumpulan data yang hanya berisi kolom yang kita percaya akan berkontribusi baik terhadap deteksi outlier. Kita akan menggunakan algoritma deteksi outlier berbasis jarak jauh yang menghitung jarak Euclidean antara titik data dan menandai titik-titik yang paling jauh dari titik data lainnya sebagai outlier. Jarak Euclidean menggunakan jarak antara dua titik data untuk masing-masing atribut.

Apa efeknya pada jarak jika atribut memiliki rentang nilai yang berbeda (satu atribut antara 0 dan 5 dan atribut lainnya antara 1 dan 1000)? Atribut dengan nilai lebih besar akan memberi kontribusi

lebih banyak daripada nilai yang lebih kecil. Untuk alasan ini, kita harus memastikan bahwa semua atribut menggunakan rentang nilai yang serupa. Transformasi ini disebut Normalisasi.

### Langkah 2. Menormalisasi rentang nilai atribut.

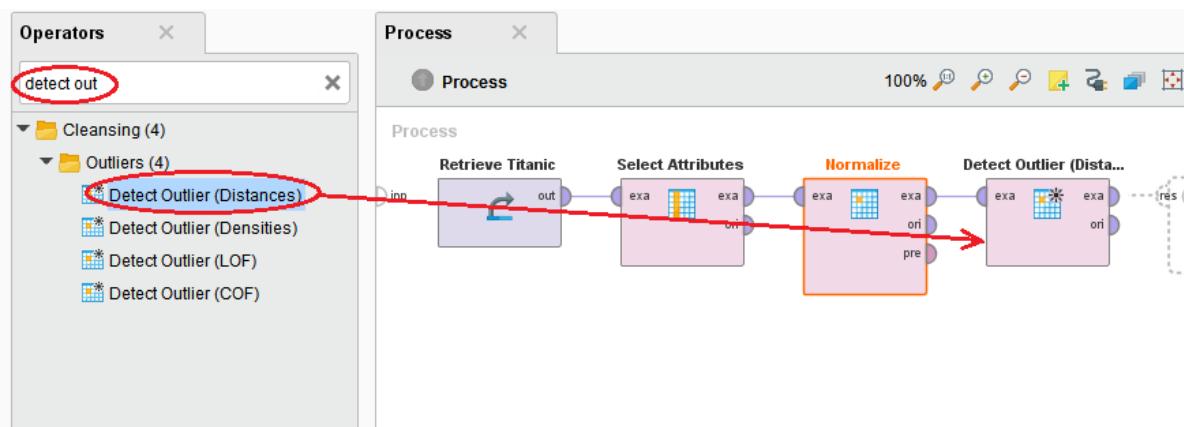
Tambahkan operator **Normalize** dan hubungkan.



Secara umum, kita harus selalu menormalkan data sebelum menerapkan algoritma berbasis jarak seperti deteksi outlier atau k-Means clustering. Dengan menggunakan parameter default, operator **Normalize** akan melakukan Transformasi z (juga dikenal sebagai Standardisasi) yang menghasilkan nilai rata-rata 0 dan deviasi standar 1 untuk setiap atribut. Dengan kata lain, semua atribut berada pada skala yang sama setelah normalisasi dan dapat dibandingkan satu sama lain.

### Langkah 3 . Menormalisasi rentang nilai atribut.

Tambahkan operator Normalize dan hubungkan.

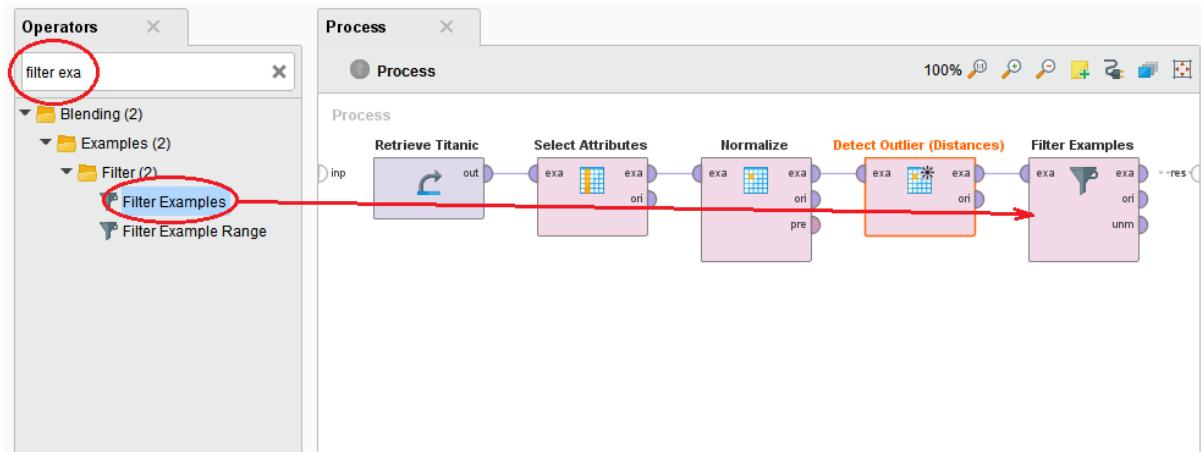


Secara umum, kita harus selalu menormalkan data sebelum menerapkan algoritma berbasis jarak seperti deteksi outlier atau k-Means clustering. Dengan menggunakan parameter default, operator **Normalize** akan melakukan Transformasi z (juga dikenal sebagai Standardisasi) yang

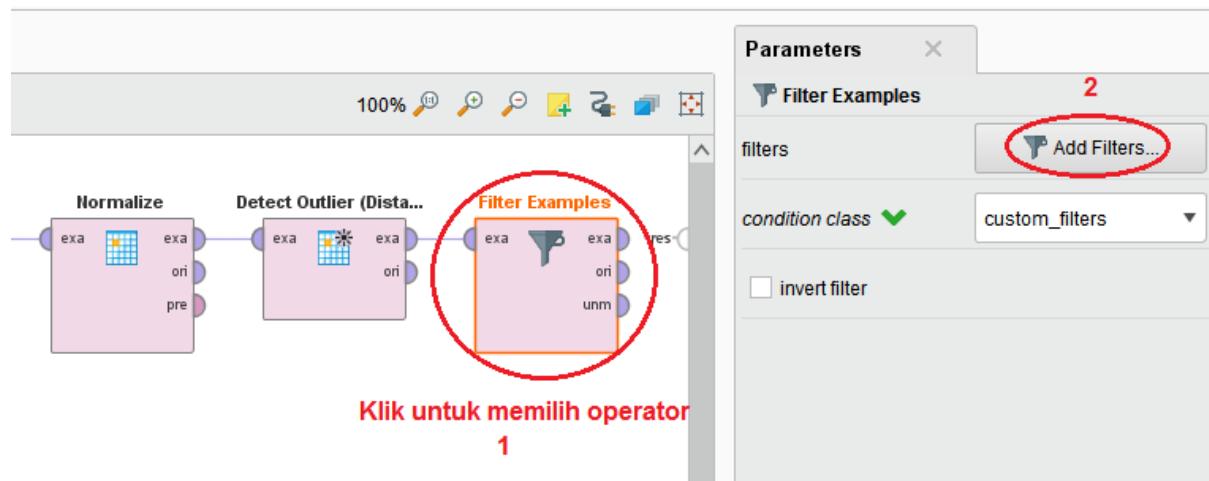
menghasilkan nilai rata-rata 0 dan deviasi standar 1 untuk setiap atribut. Dengan kata lain, semua atribut berada pada skala yang sama setelah normalisasi dan dapat dibandingkan satu sama lain.

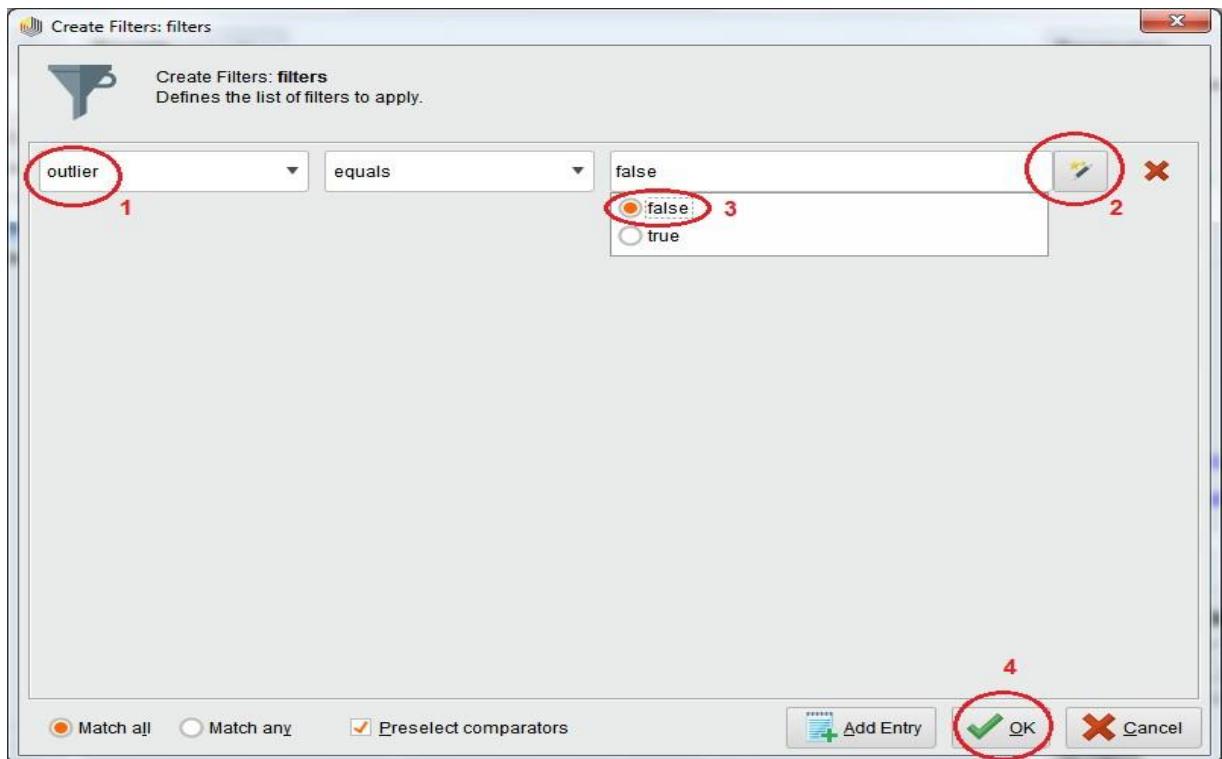
#### Langkah 4. Menghapus outlier dari contoh.

1. Tambahkan **Filter Examples** ke proses dan hubungkan ke operator sebelumnya dan juga ke port hasil di sebelah kanan.



2. Di Parameternya, tambahkan filter baru dengan Outlier, equal, dan false sebagai nilai.





### 3. Jalankan prosesnya.

The screenshot shows the KNIME interface with the 'Results' tab selected. On the left, there is a sidebar with icons for Data, Statistics, Charts, Advanced Charts, and Annotations. The main area displays a table titled 'ExampleSet (Filter Examples)' with 17 rows. The table includes columns for Row No., outlier, Age, No of Sibling..., No of Parent..., Passenger F..., Passenger ... (partially visible), Sex, Port of Emb..., and Survived. All rows in the 'outlier' column are marked as 'false'. The first few columns show numerical values for Age, Siblings, Parents, and so on.

Row No.	outlier	Age	No of Sibling...	No of Parent...	Passenger F...	Passenger ...	Sex	Port of Emb...	Survived
1	false	-0.061	-0.479	-0.445	3.440	First	Female	Southampton	Yes
2	false	-2.010	0.481	1.866	2.285	First	Male	Southampton	Yes
3	false	-1.934	0.481	1.866	2.285	First	Female	Southampton	No
4	false	0.008	0.481	1.866	2.285	First	Male	Southampton	No
5	false	-0.339	0.481	1.866	2.285	First	Female	Southampton	No
6	false	1.257	-0.479	-0.445	-0.130	First	Male	Southampton	Yes
7	false	2.298	0.481	-0.445	0.863	First	Female	Southampton	Yes
8	false	0.633	-0.479	-0.445	-0.643	First	Male	Southampton	No
9	false	1.604	1.441	-0.445	0.351	First	Female	Southampton	Yes
10	false	2.853	-0.479	-0.445	0.313	First	Male	Cherbourg	No
11	false	1.188	0.481	-0.445	3.753	First	Male	Cherbourg	No
12	false	-0.824	0.481	-0.445	3.753	First	Female	Cherbourg	Yes
13	false	-0.408	-0.479	-0.445	0.696	First	Female	Cherbourg	Yes
14	false	-0.269	-0.479	-0.445	0.880	First	Female	Southampton	Yes
15	false	3.477	-0.479	-0.445	-0.064	First	Male	Southampton	Yes
16	false	?	-0.479	-0.445	-0.142	First	Male	Southampton	No
17	false	-0.408	-0.479	0.710	4.139	First	Male	Cherbourg	No

Prosesnya mungkin berjalan beberapa lama tapi akan beralih ke tampilan **Results** secara otomatis saat selesai. Kita akan melihat bahwa hasilnya adalah kumpulan data dengan 1.299 contoh serta 10 outlier yang telah berhasil dihapus.

# Tutorial Belajar Rapidminer (Bagian 13) - Melakukan Pivot dan Mengganti nama Atribut

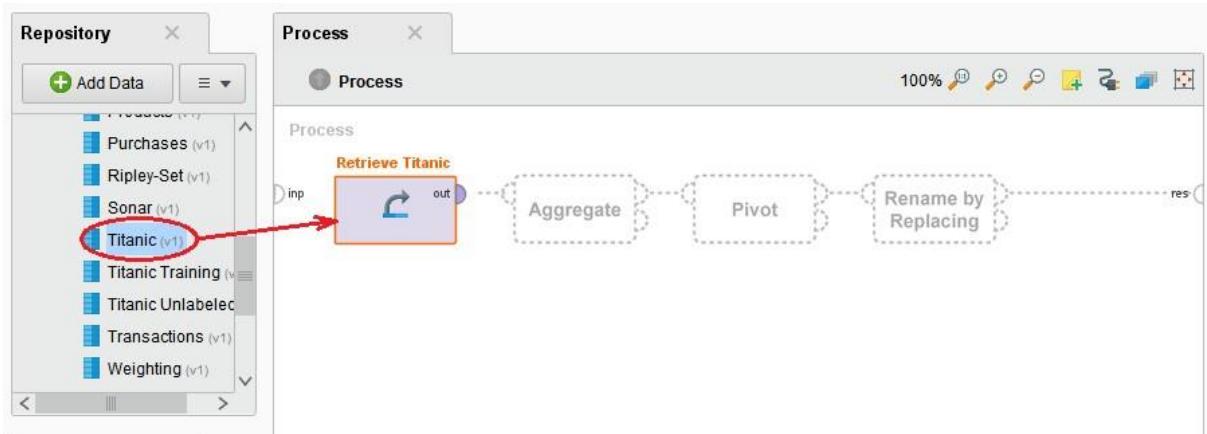
Pivot : Mengubah format tabel yang panjang menjadi lebar.

Dalam tutorial ini, kita akan belajar tentang teknik pencampuran data umum lainnya, yaitu melakukan pivot pada data. Kita mungkin sudah familiar dengan konsep pivot dari perangkat BI atau Excel: mengubah data dari format tabel panjang (satu atribut dengan banyak contoh) ke dalam format tabel yang lebar (banyak atribut dengan satu contoh saja). Transformasi ini sangat berguna setelah agregasi karena informasi agregat selalu disimpan dalam format yang panjang. Model pembelajaran mesin, bagaimanapun juga memerlukan data untuk disimpan ke dalam format kolom yang luas, sehingga kita akan sering menghadapi tahap preprocessing ini, sebelum kita memulai dengan pemodelan yang sebenarnya.

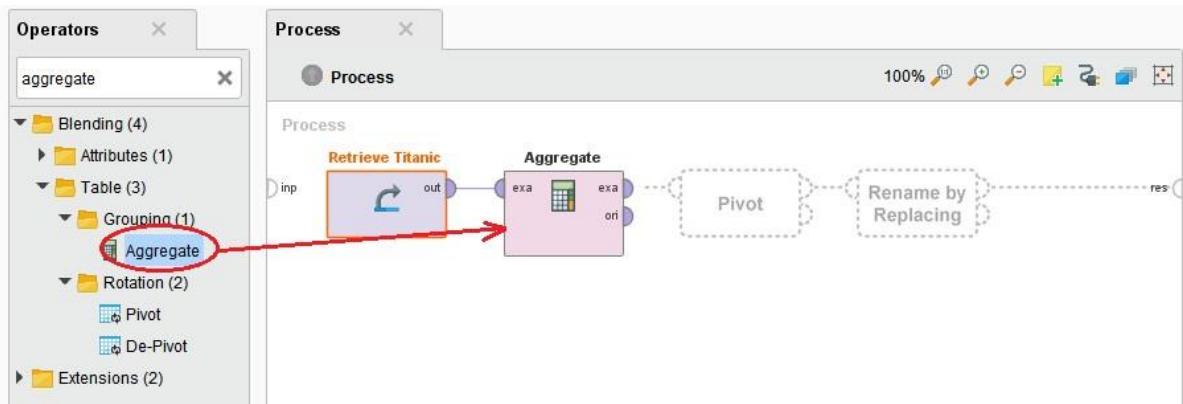
## Langkah 1. Mempersiapkan Data

Mari kita mulai dengan membuat tabel yang menunjukkan berapa banyak penumpang di setiap kelas, dikelompokkan menurut jenis kelamin.

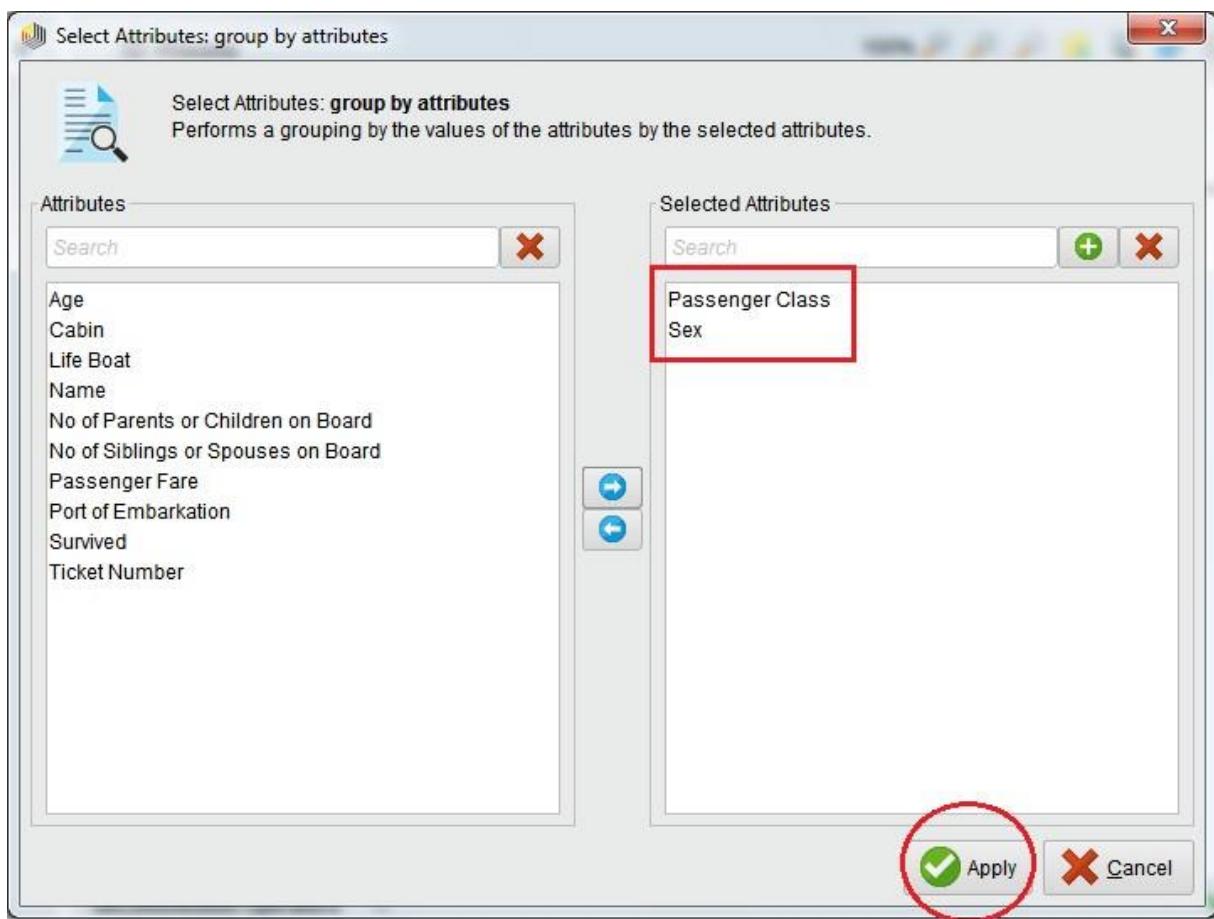
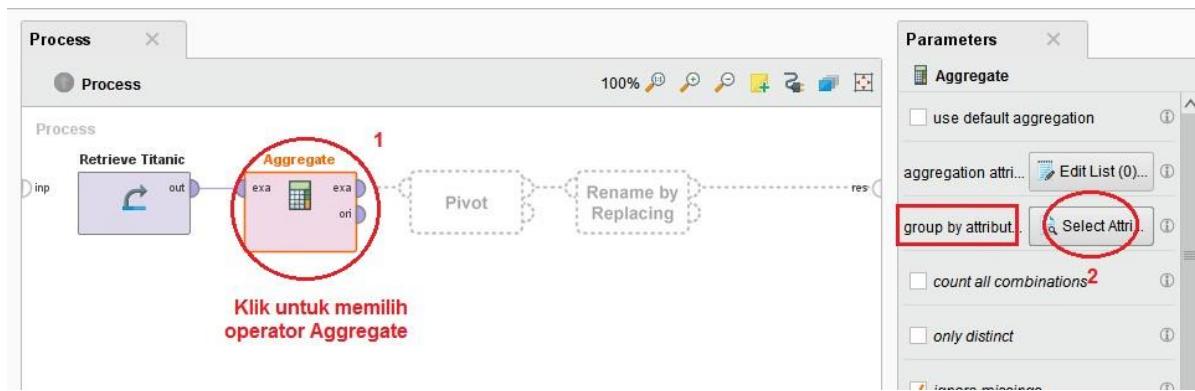
### 1. Tarik data **Titanic** ke dalam jendela **Process**.



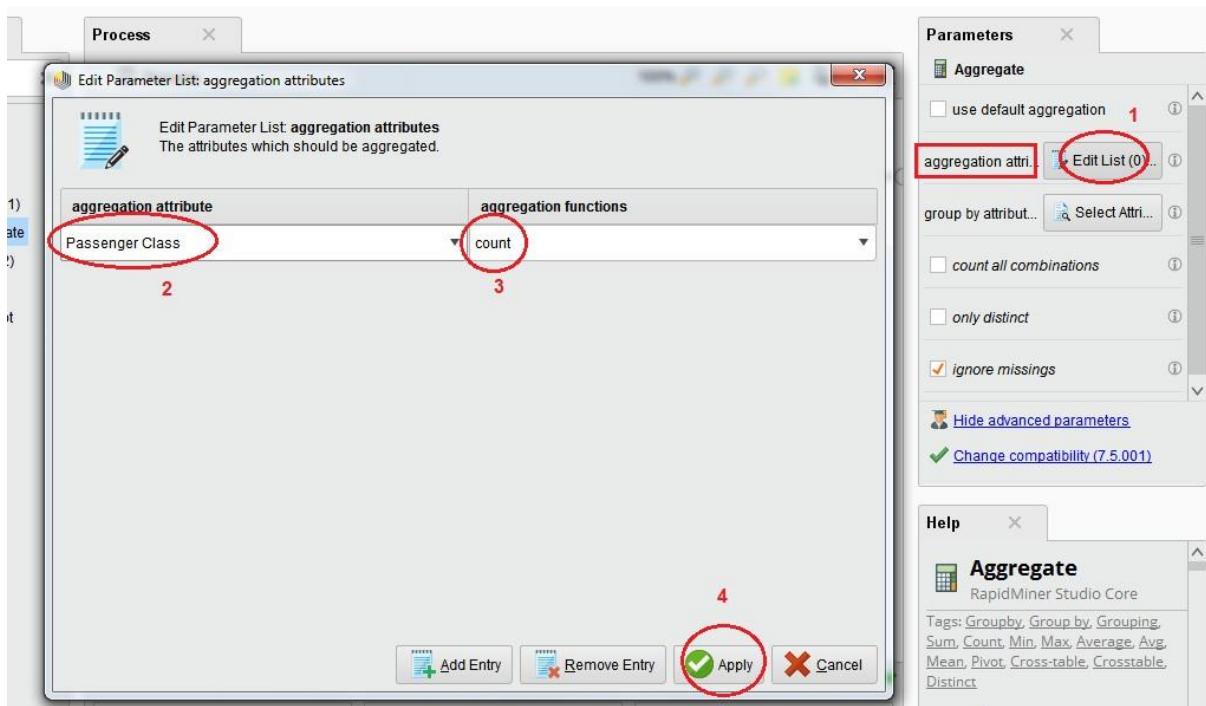
### 2. Tambahkan operator **Aggregate** dan hubungkan.



### 3. Pada **Parameters**, pilih **Passenger Class** dan **Sex** sebagai **group by attributes**.



4. Juga gunakan **Passenger Class** dengan fungsi count sebagai entri baru untuk aggregation attributes.

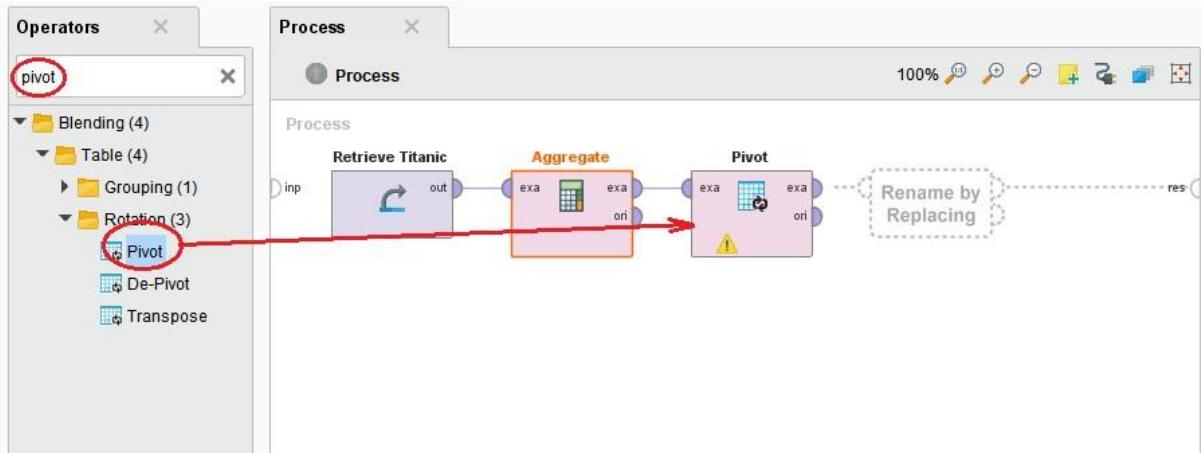


Hasilnya akan berupa tabel dengan tiga kolom: Kelas Penumpang, Jenis Kelamin, dan hitungan untuk masing-masing kombinasi tersebut. Baris pertama adalah First, Female dengan jumlah 144, baris kedua adalah First, Male dengan jumlah 179 dan seterusnya. Semua kombinasi kelas penumpang dan jenis kelamin diwakili sebagai baris, karena itulah kita menyebutnya dengan format tabel panjang.

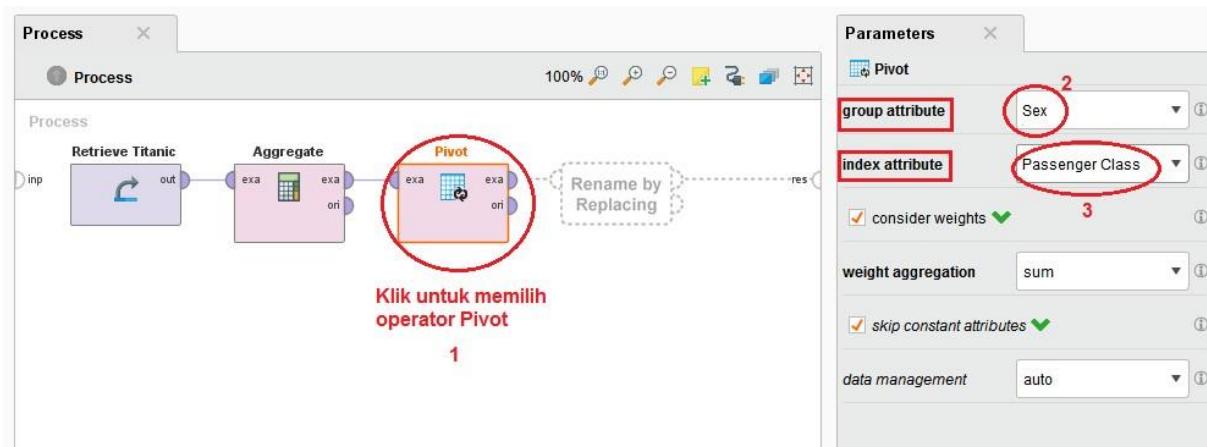
## Langkah 2 . Melakukan pivot data

Mari kita ubah format tabel panjang menjadi format lebar yang akan membuat jumlah pria dan wanita di kelas masing-masing menjadi sedikit lebih jelas.

### 1. Cari operator **Pivot**, tambahkan, dan hubungkan ke **Aggregate**.



### 2. Pilih operator **Pivot** untuk mengatur parameteranya. Gunakan **Sex** sebagai grup attribute.



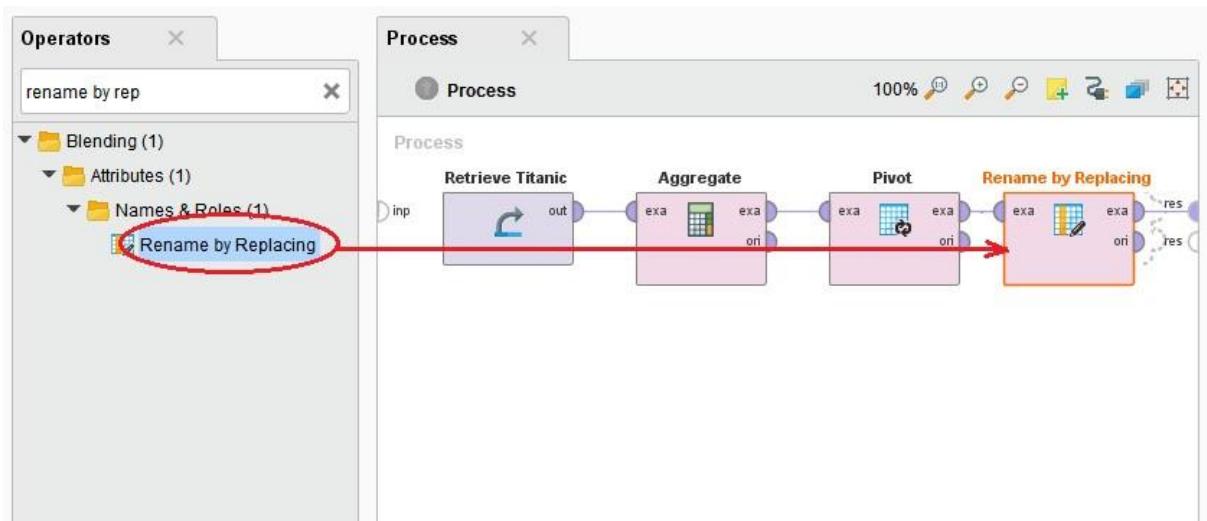
### 3. Gunakan Passenger Class sebagai index attribute.

Kumpulan data yang dihasilkan memiliki empat kolom dan dua baris. Setiap baris mewakili salah satu nilai dari kolom Sex (parameter atribut grup). Tiga nilai atribut indeks (kelas penumpang) yang berbeda menjadi tiga kolom baru. Nilai aktual dalam tabel mewakili jumlah dari gabungan kombinasi kelompok (gender) dengan nilai indeks (atau subkelompok), yang merupakan kelas penumpang dalam kasus ini. Sebagai contoh kita mendapatkan 144 wanita yang memesan kelas satu.

#### Langkah 3. Mengganti nama atribut dengan ekspresi reguler.

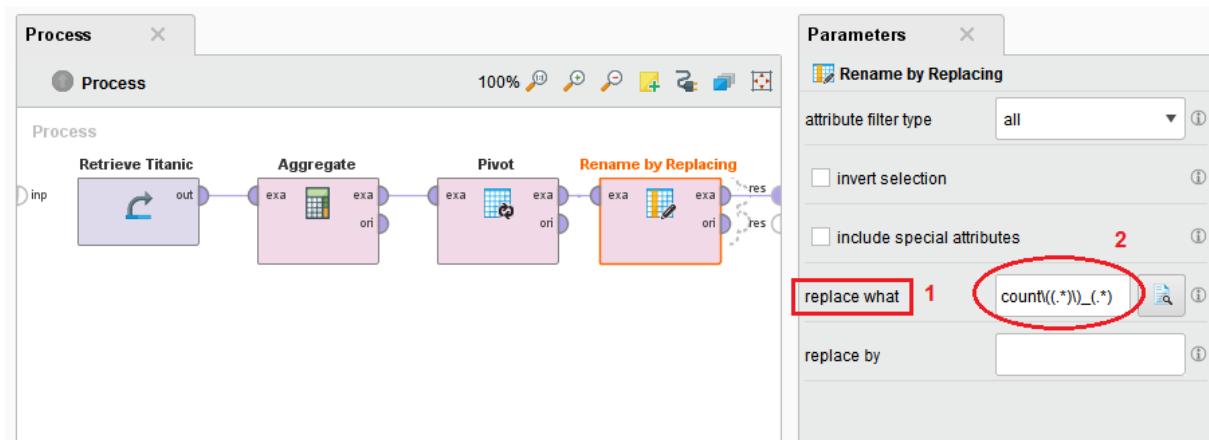
Nama kolom baru mencerminkan bagaimana kolom tersebut diciptakan, tapi ini tidak selalu mudah untuk dibaca. Kita bisa menggunakan operator Rename untuk mengganti nama tiga atribut secara manual dengan sesuatu yang lebih bagus, seperti "Passenger Class First". Mengganti nama adalah sebuah cara jika kita memiliki beberapa atribut yang harus diganti namanya, namun kita akan mengambil pendekatan yang lebih baik yang memungkinkan kita mengganti nama ratusan atribut sekaligus.

#### 1. Cari operator **Rename by Replacing**, tambahkan dan hubungkannya ke **Pivot**.

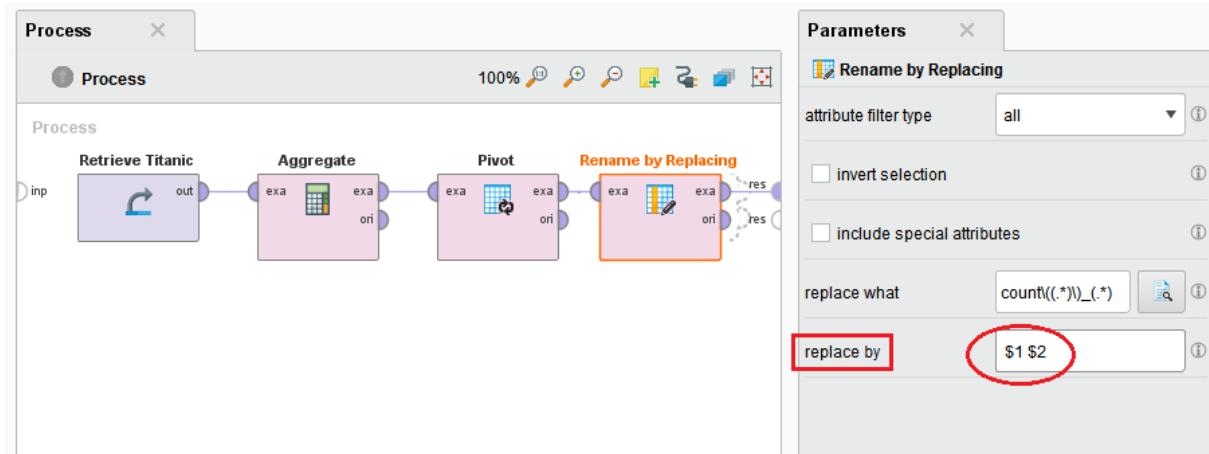


#### 2. Hubungkan operator ke port result di sebelah kanan.

#### 3. Salin **count ((\*)) ) \_ (. \*)** menuju kolom **replace what**. Pastikan bahwa kita menuliskan dengan lengkap semua tanda kurung dengan benar!



4. Salin \$1 \$2 ke parameter replace by.



5. Jalankan prosesnya

KNIME Free 7.5.001 @ pc

Actions Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Rename by Replacing)

Data

Row No.	Sex	Passenger Class First	Passenger Class Second	Passenger Class Third
1	Female	144	106	216
2	Male	179	171	493

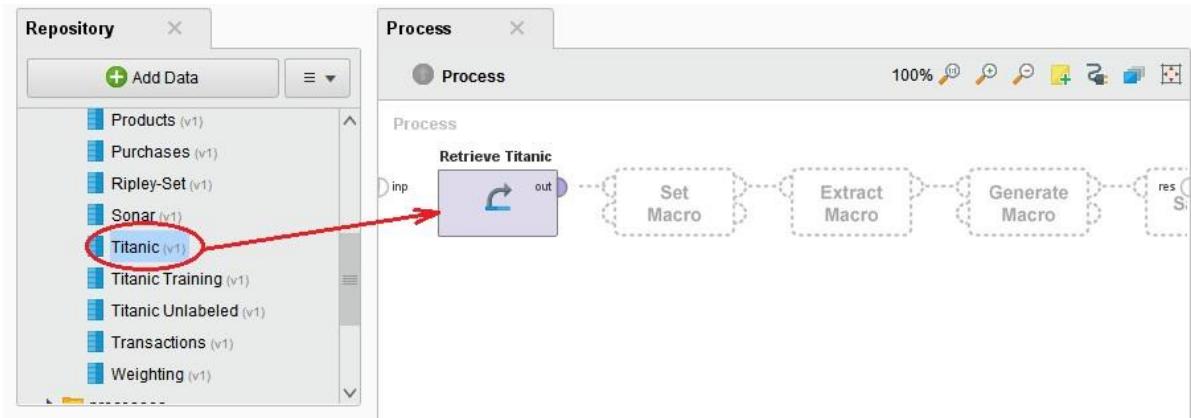
Statistics

# Tutorial Belajar Rapidminer (Bagian 14) - Macros sebagai Variabel Proses dan Sampling

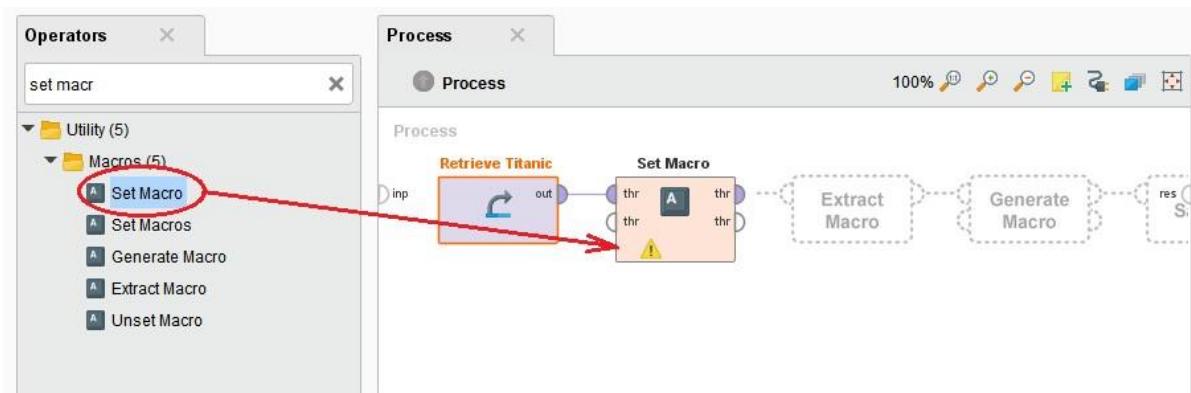
Dalam tutorial ini, Anda akan belajar tentang konsep RapidMiner yang sangat berguna yaitu macro. Makro adalah variabel yang dapat kita gunakan untuk menyimpan dan memuat nilai secara dinamis dalam proses kita. Dalam tutorial ini, kita akan menggunakan macro untuk mengurangi ukuran data sebesar 50% untuk melakukan perhitungan menggunakan ukuran yang baru.

**Langkah 1.** menyiapkan data.

1. Tarik data **Titanic** ke dalam jendela **process**.

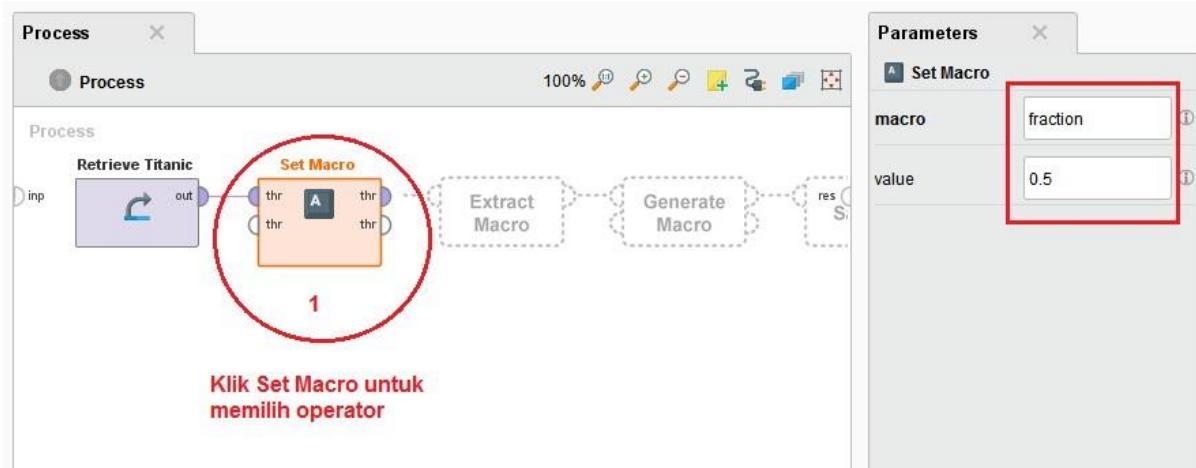


2. Cari operator **Set Macro** dan tambahkan ke prosesnya.



3. Hubungkan **Set Macro** dengan **Retrieve Titanic**.

4. Klik **Set Macro** dan buat perubahan berikut pada Parameternya:



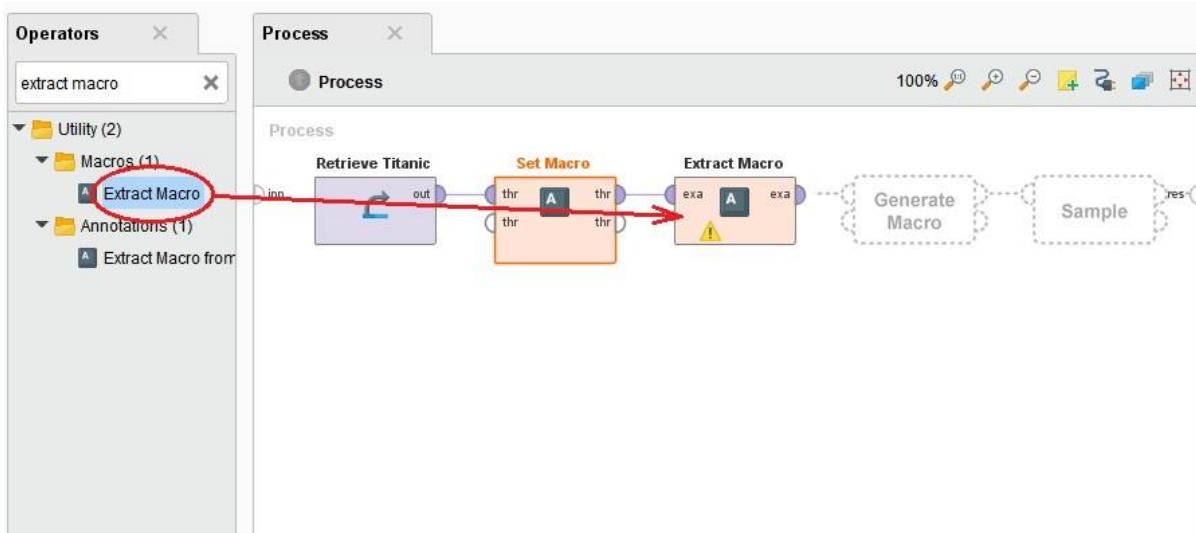
Atur **macro** ke fraction, atur **value** ke 0.5.

Makro dapat digunakan untuk menyimpan sembarang nilai alfanumerik. Setiap makro memiliki nama dan nilai. Set Macro hanyalah salah satu cara untuk mengatur makro, kita akan melihat beberapa cara lain nantinya.

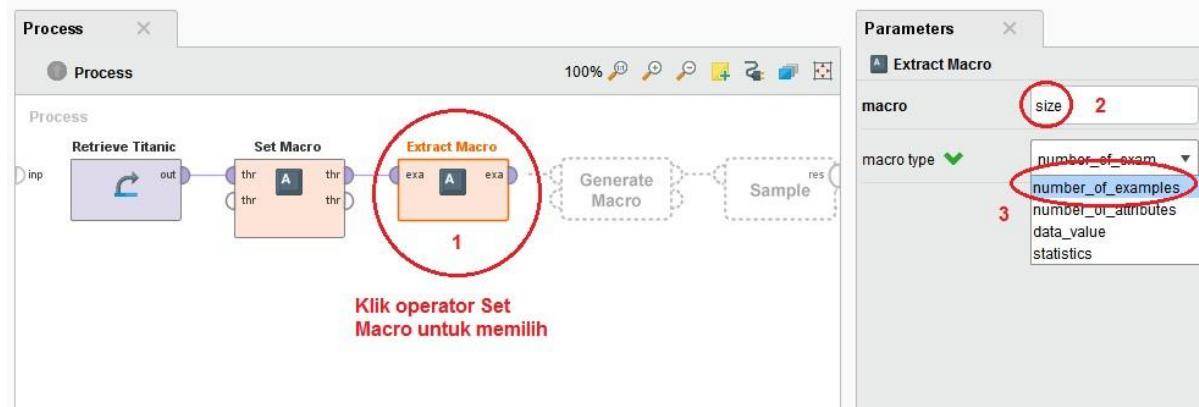
**Langkah 2 :** Mengekstrak makro dari data.

Dengan **Set Macro**, kita akan membuat makro dan menentukan nilainya secara manual. Kita juga dapat membuat makro di mana nilainya mengambil dari bagian proses yang lain, seperti jumlah contoh dalam dataset. Berikut langkahnya :

1. Tambahkan operator **Extract Macro** ke dalam proses dan hubungkan.



2. Dalam parameternya, atur opsi **macro** ke **size**, atur jenis macro ke **number\_of\_examples**.

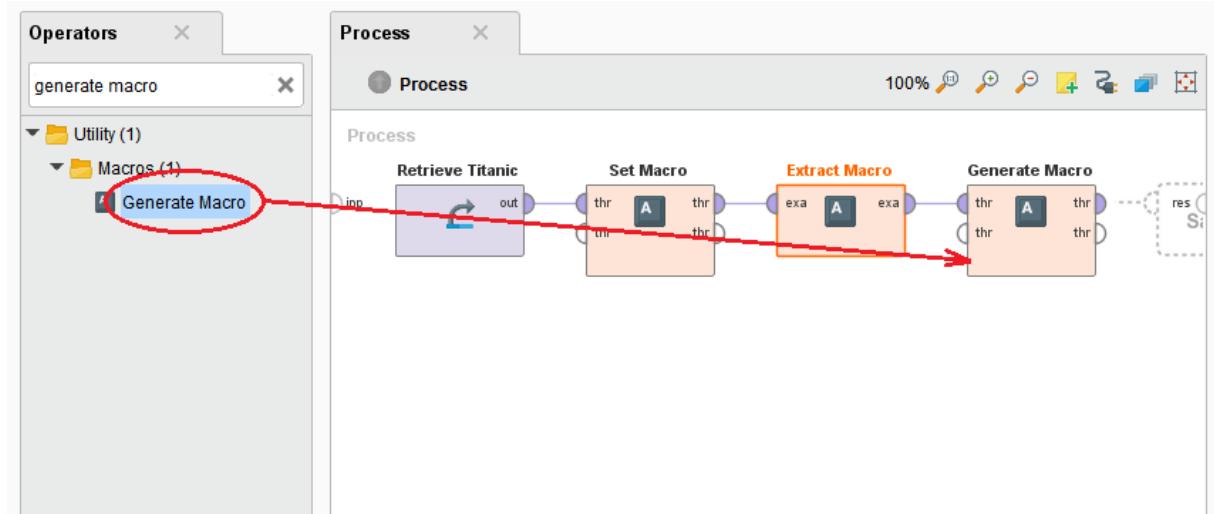


### Langkah 3 : Menghitung makro yang baru

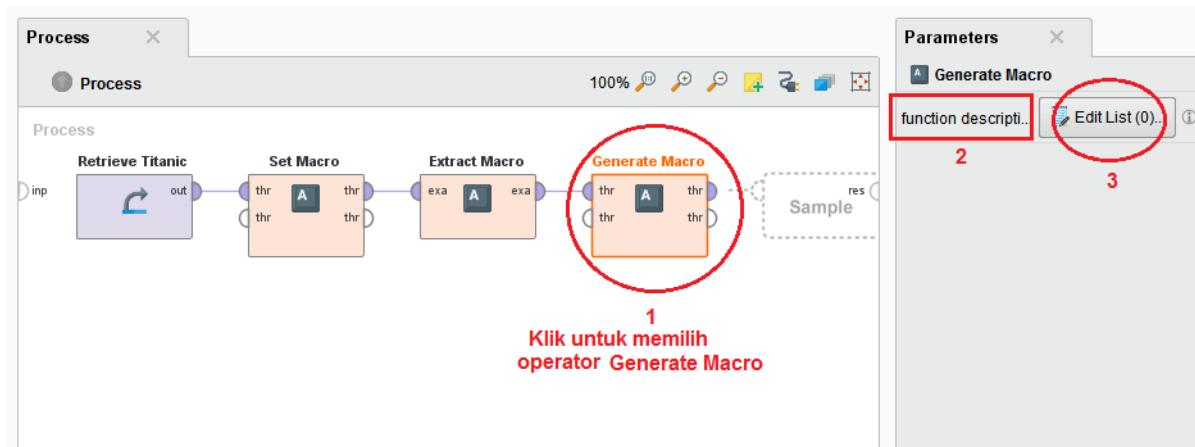
Kita dapat menghasilkan nilai makro menggunakan ekspresi sewenang-wenang dengan cara yang sama seperti membuat kolom baru dengan operator **Generate Attributes**. Di sini kita akan menghitung ukuran contoh himpunan baru dengan mengalikan pecahan yang telah kita definisikan sebelumnya (**Set Macro**) dengan ukuran lama (**Extract Macro**) dan membulatkan hasilnya ke bilangan bulat.

Berikut langkahnya :

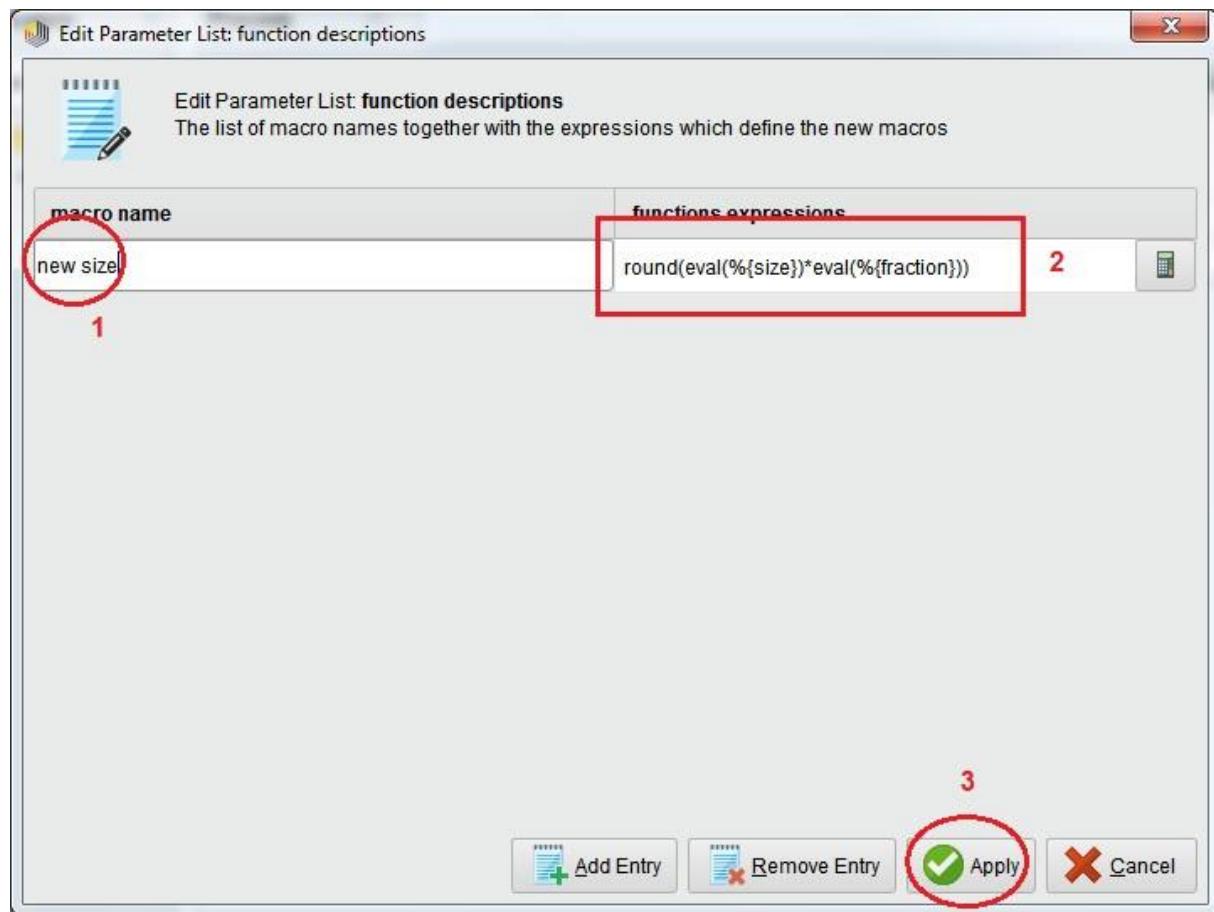
1. Mencari operator **Generate Macro** dan masukan dalam proses, kemudian hubungkan.



2. Di Parameternya, klik pada **function description**.



3. Tambahkan entri baru dengan **new size** sebagai **macro name** dan **round (eval (% {size}) \* eval (% {fraction}))** sebagai **function expressions**.



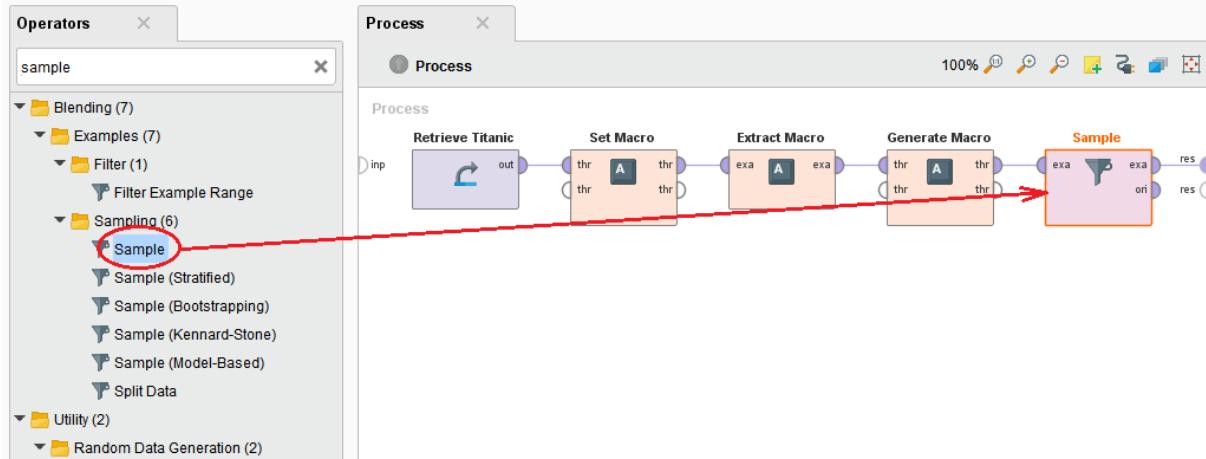
Harap diperhatikan: Bila kita mendefinisikan makro, kita bisa menggunakan sembarang teks sebagai nama, misal. "nama makro". Tapi bila kita menggunakan parameter makro pada operator lain, kita harus menggunakan format **% {nama makro}**.

Fungsi **eval** memberitahu ekspresi parser pada RapidMiner bahwa nilai makro harus dievaluasi. Dalam kasus ini, ini mengubah dua makro menjadi angka sehingga dapat digunakan dalam perkalian

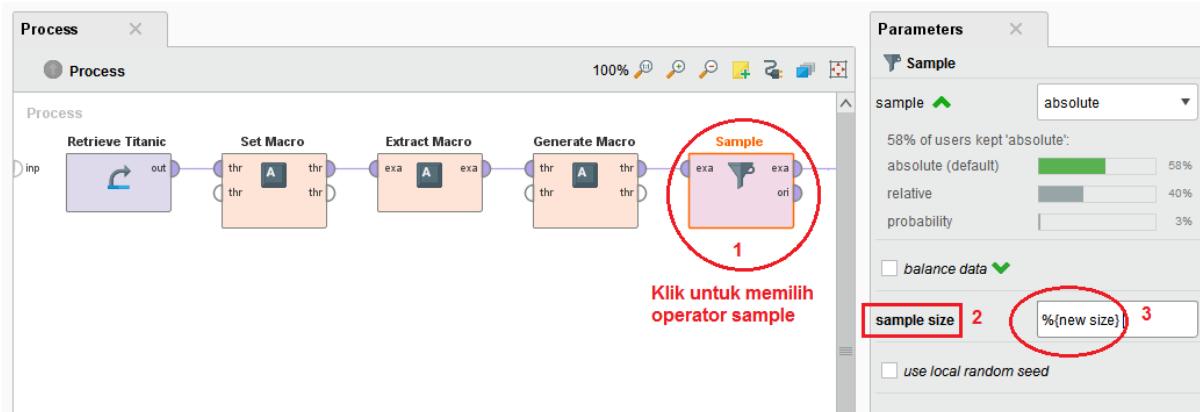
(contohnya bagaimana kita dapat memperbanyak dua buah teks?). Fungsi eval sering diperlukan jika kita ingin menggunakan macro sebagai bagian dari ekspresi fungsi namun kita tidak perlu menggunakannya jika kita hanya menggunakan makro sebagai parameter nilai pada operator lain.

#### Langkah 4 : Melakukan Sampel data.

1. Seret operator **sample** ke dalam proses dan hubungkan.

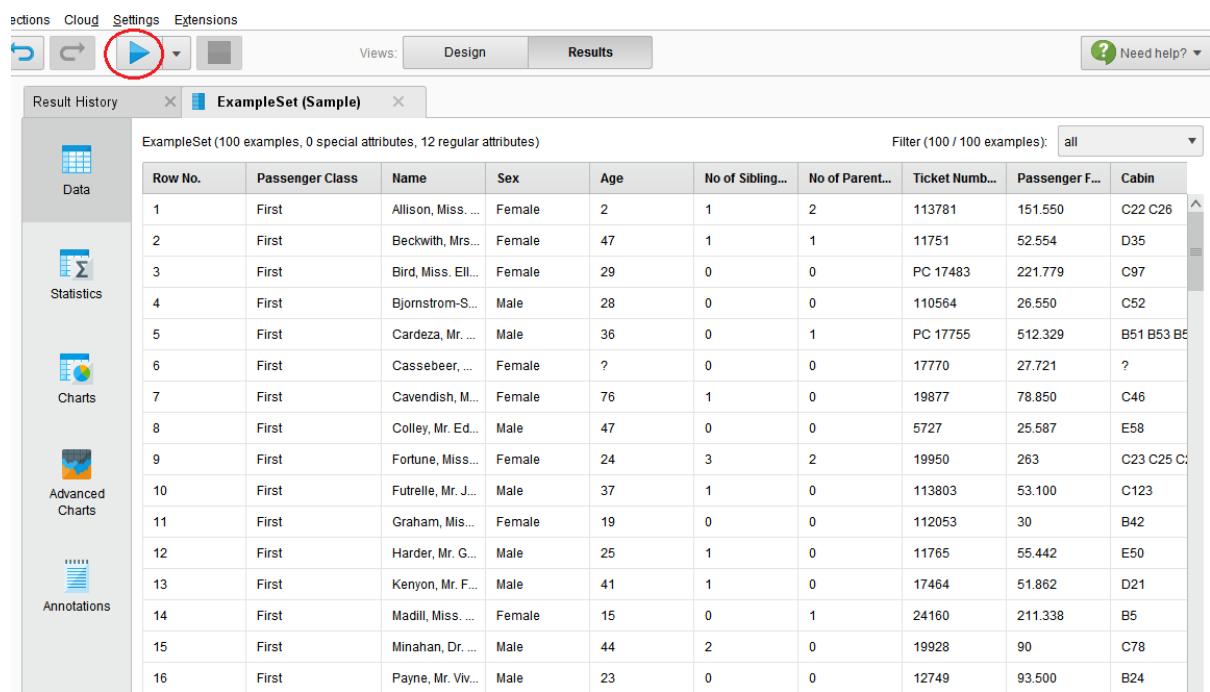


2. Atur parameter **sample size** ke **% {new size}**.



3. Hubungkan **output of Sample** ke port result di sebelah kanan.

#### 4. Jalankan prosesnya.



The screenshot shows the KNIME interface with the 'Data' tab selected. A red circle highlights the play button icon in the toolbar. The main area displays a table titled 'ExampleSet (100 examples, 0 special attributes, 12 regular attributes)'. The table has columns: Row No., Passenger Class, Name, Sex, Age, No of Sibling..., No of Parent..., Ticket Num..., Passenger F..., and Cabin. The data includes entries for passengers like Allison, Beckwith, Bird, Bjornstrom-S..., Cardeza, Cassebeer, Cavendish, Colley, Fortune, Futrelle, Graham, Harder, Kenyon, Madill, Minahan, and Payne. The 'Annotations' column on the far right contains labels such as C22 C26, D35, C97, C52, B51 B53 B54, ?, C46, E58, C23 C25 C26, C123, B42, E50, D21, B5, C78, and B24.

Row No.	Passenger Class	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Cabin
1	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26
2	First	Beckwith, Mrs. ...	Female	47	1	1	11751	52.554	D35
3	First	Bird, Miss. Ell... ...	Female	29	0	0	PC 17483	221.779	C97
4	First	Bjornstrom-S...	Male	28	0	0	110564	26.550	C52
5	First	Cardeza, Mr. ...	Male	36	0	1	PC 17755	512.329	B51 B53 B54
6	First	Cassebeer, ...	Female	?	0	0	17770	27.721	?
7	First	Cavendish, M... ...	Female	76	1	0	19877	78.850	C46
8	First	Colley, Mr. Ed...	Male	47	0	0	5727	25.587	E58
9	First	Fortune, Miss... ...	Female	24	3	2	19950	263	C23 C25 C26
10	First	Futrelle, Mr. J... ...	Male	37	1	0	113803	53.100	C123
11	First	Graham, Mis... ...	Female	19	0	0	112053	30	B42
12	First	Harder, Mr. G... ...	Male	25	1	0	11765	55.442	E50
13	First	Kenyon, Mr. F... ...	Male	41	1	0	17464	51.862	D21
14	First	Madill, Miss. ...	Female	15	0	1	24160	211.338	B5
15	First	Minahan, Dr. ...	Male	44	2	0	19928	90	C78
16	First	Payne, Mr. Viv... ...	Male	23	0	0	12749	93.500	B24

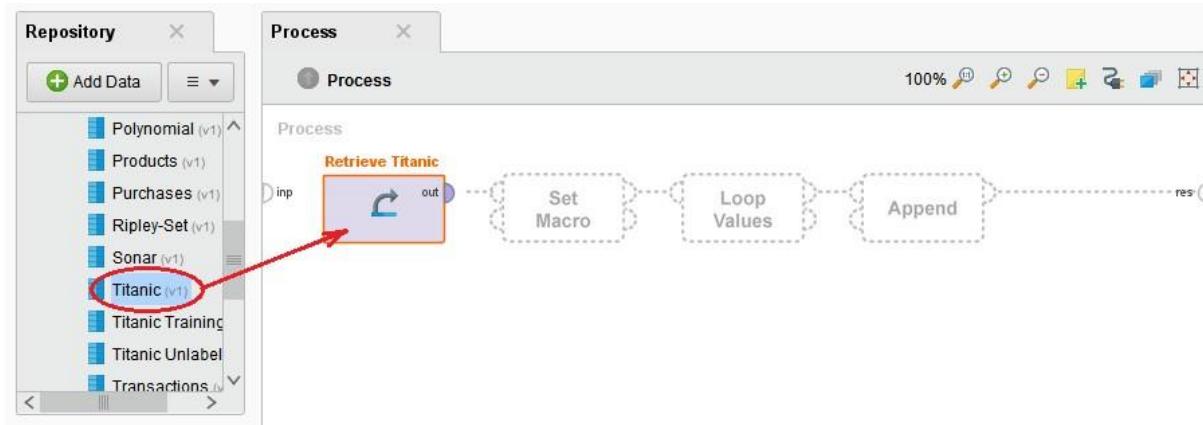
Perlu diketahui lagi bahwa kita harus menggunakan format % {new size} untuk mengakses nilai makro ini dalam pengaturan parameter.

# Tutorial Belajar Rapidminer (Bagian 15) - Perulangan dan Percabangan

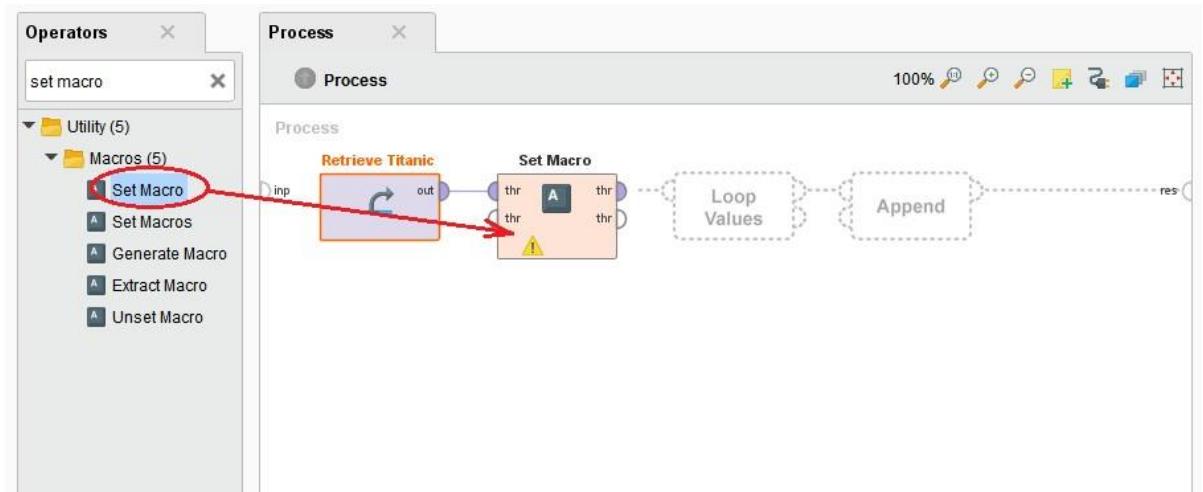
Perulangan dan percabangan adalah konsep yang sangat berguna yang dikenal di hampir semua bahasa pemrograman. RapidMiner menawarkan puluhan operator untuk perulangan contoh, atribut, atau nilai atribut tertentu. Dalam tutorial ini, kita akan membangun sebuah proses yang membahas tiga kelas penumpang yang berbeda dari data Titanic dan memeriksa apakah jumlah sampelnya lebih tinggi dari angka yang telah ditentukan sebelumnya. Jika lebih tinggi, kita menurunkan subkelompok ini ke jumlah sebelumnya, jika tidak kita akan menyimpan semua contohnya saja. Proses seperti ini misalnya, dapat digunakan untuk menyeimbangkan kelas dalam kasus di mana beberapa kelas yang sangat besar mendominasi beberapa kelas minoritas.

**Langkah 1** Menyiapkan data dan makro.

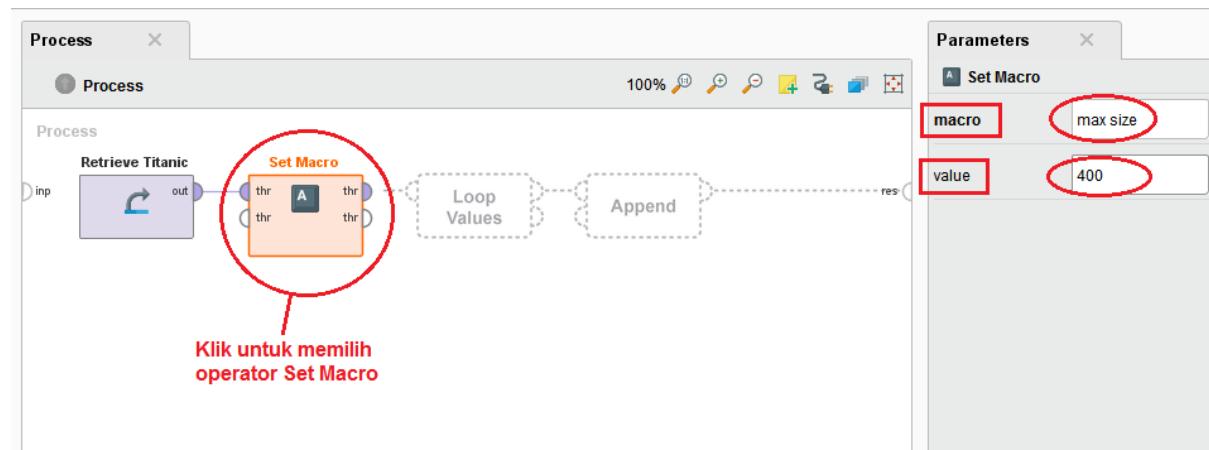
1. Seret data **Titanic** ke dalam proses.



2. Tambahkan operator **Set Macro** ke proses dan hubungkan.



3. Gunakan ukuran **max size** sebagai makro dan **400** sebagai value pada **Parameters**.

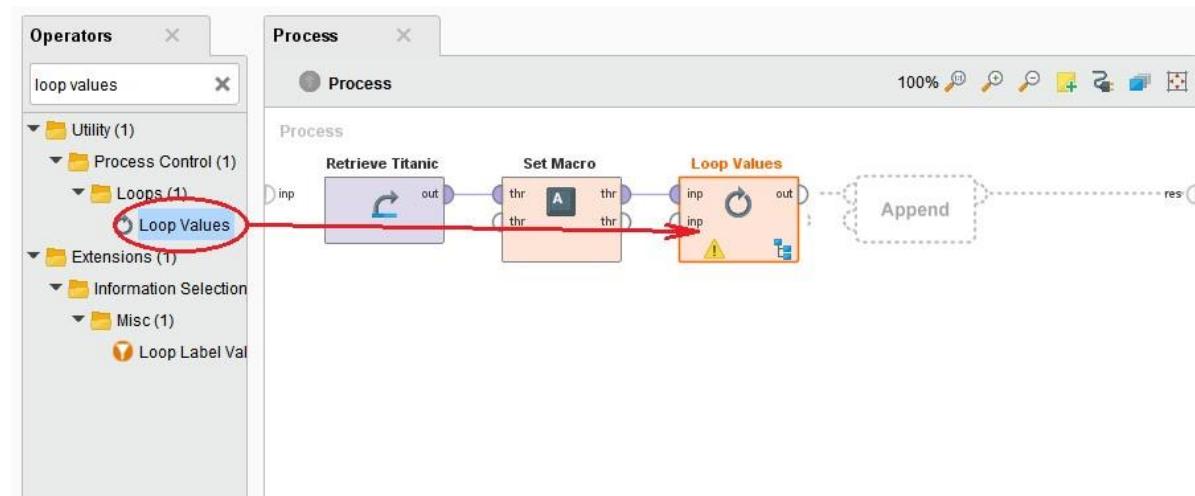


Kita akan menggunakan 400 contoh sebagai jumlah maksimum untuk masing-masing dari tiga kelas penumpang. Dengan mendefinisikan ukuran max size dengan makro pada awal proses, akan memudahkan untuk membuat perubahan pada pengaturannya nanti. Hal ini sangat berguna jika kita menggunakan nilai di banyak bagian proses.

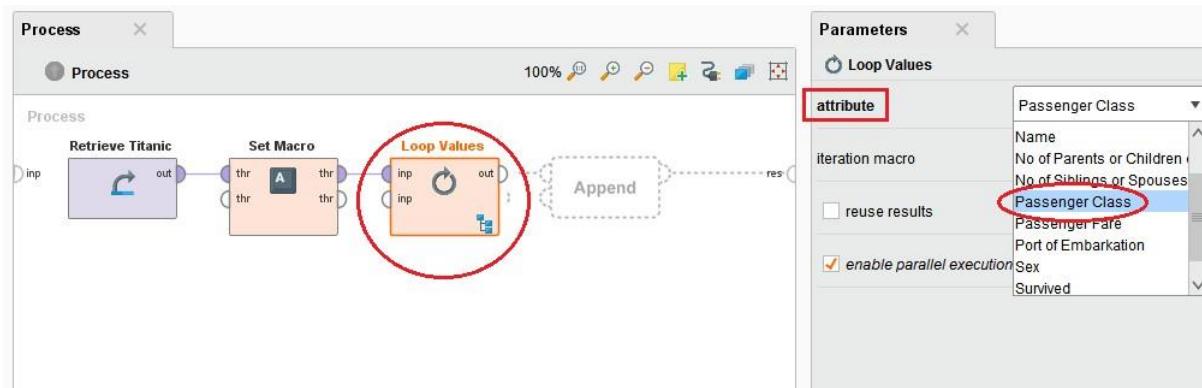
**Langkah 2 .** Melakukan perulangan nilai atribut.

Kita akan melakukan perulangan pada tiga kelas penumpang dan memeriksa sebuah kondisi untuk masing-masing kelas. Dalam kasus ini, kondisinya adalah bahwa kelas harus memiliki nilai lebih kecil dari max size sampel. Berikut langkahnya :

1. Cari operator **Loop Values** dan seret ke dalam proses. Kemudian hubungkan operator.



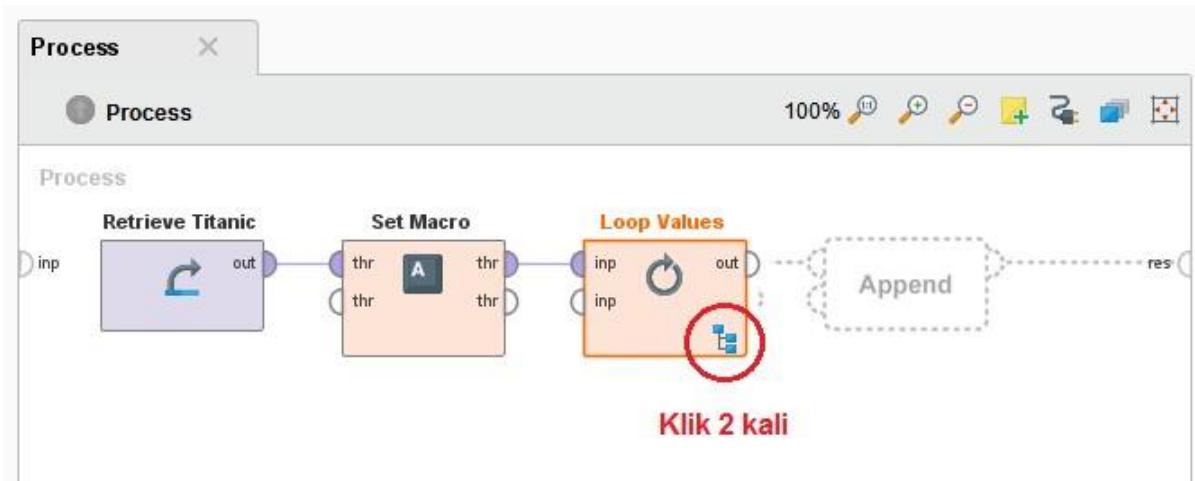
2. Pada **Parameter**, atur attribut ke **Passenger Class**.



**Langkah 3.** Pengaturan di dalam perulangan

Apakah Anda memperhatikan ikon kecil di sudut kanan bawah Nilai Loop operator? Ikon ini menunjukkan bahwa operator ini dapat memiliki operator lain yang bersarang di dalamnya. Klik dua kali operator Loop Values untuk melihat isinya.

1. Klik dua kali **Loop Values**.

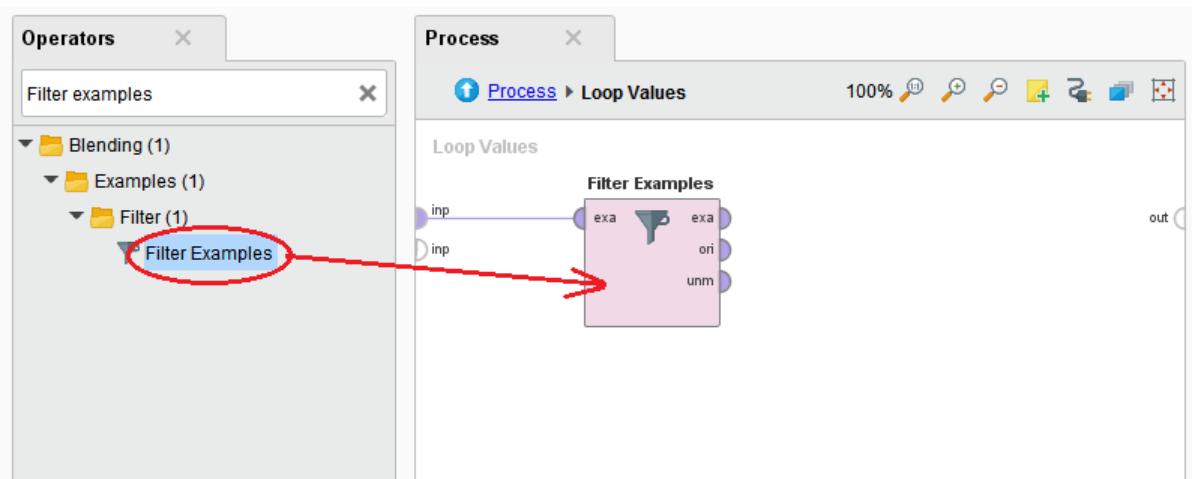


Di dalam perulangan kita dapat menentukan sub-proses yang akan dijalankan untuk masing-masing dari tiga kelas penumpang yang sedang dilakukan proses perulangan. Kapan pun kita berada dalam sub-proses nested, kita bisa menemukan nama sub-proses di pojok kiri atas. Di bagian atas panel Process, kita juga akan menemukan jalur yang dapat digunakan untuk navigasi kembali ke proses utama.

**Langkah 4.** Menyimpan sampel untuk kelas .

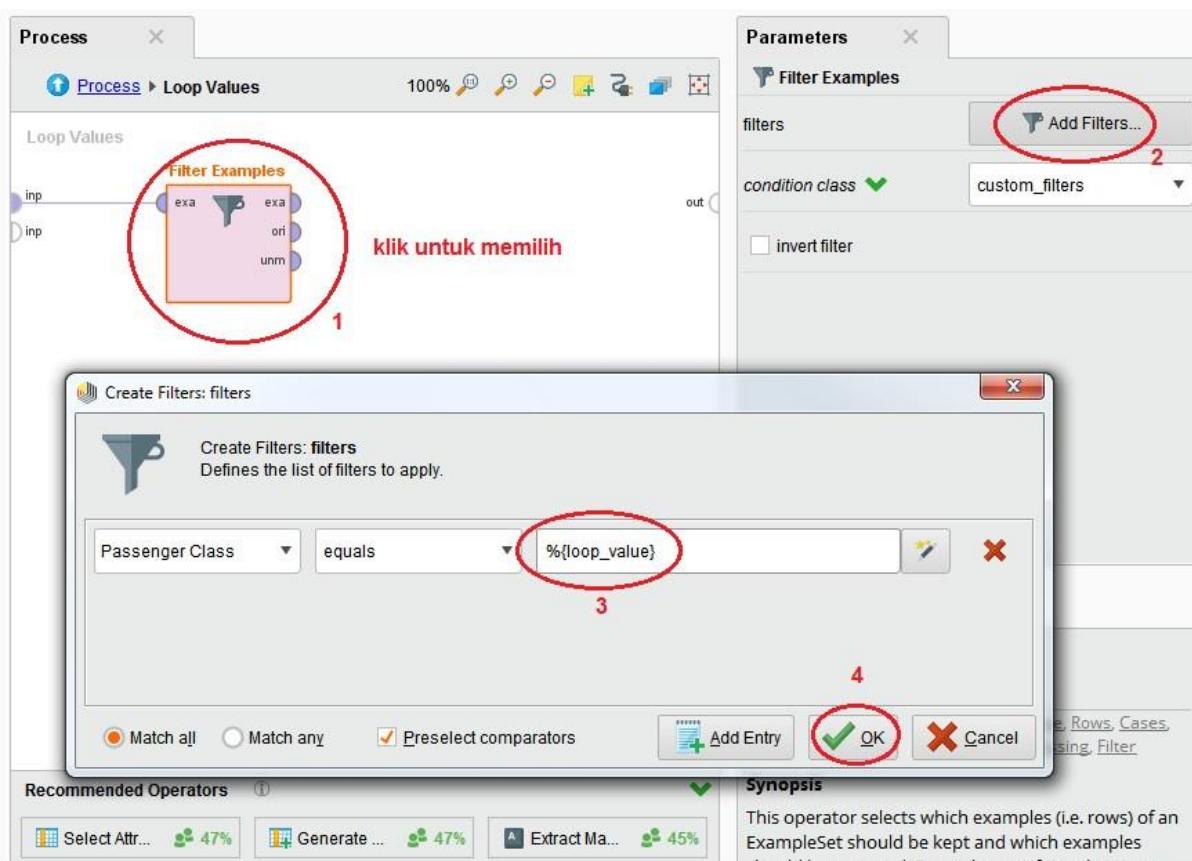
Konsep operator bersarang memungkinkan proses yang sangat berguna. Kita akan sering melihat ini dalam proses selanjutnya saat kita memvalidasi model atau mengoptimalkan pengaturan parameter. Untuk saat ini, kita mendefinisikan apa yang terjadi di dalam perulangan.

1. Pastikan kita berada di dalam operator Loop dengan mengklik dua kali,
2. Di dalam operator Loop Values, lakukan hal berikut:



### 3. Tambahkan Operator **Filter Examples**.

4. Hubungkan input filter dengan port di sebelah kiri sub-proses. Ini akan memungkinkan data set dikirimkan ke Operator Loop untuk mengalir menuju sub-proses. Pada Parameters dari Filter Example, tambahkan filter baru dengan setting **Passenger Class**, **equals**, dan **%{loop\_value}**



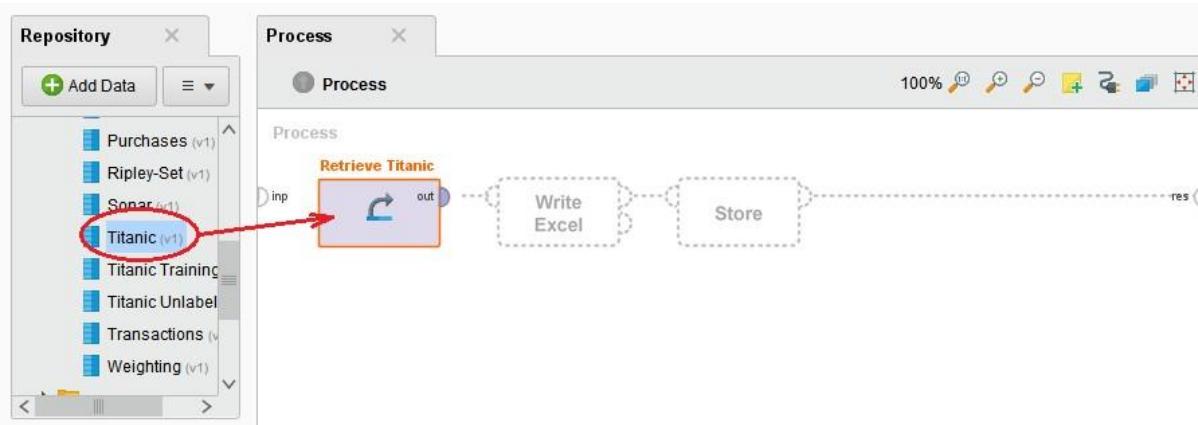
Macro **loop\_value** adalah makro yang telah ditentukan dari operator Loop Values yang dapat digunakan di dalam operator. Dalam setiap iterasi, makro diatur ke nilai atribut loop saat ini. Kita bisa menggunakan makro pada operator lain seperti misalnya, Filter Example. Kita hanya menyimpan sekumpulan sampel asli yang memiliki nilai atribut dari iterasi loop saat ini.

# Tutorial Belajar Rapidminer (Bagian 16) - Mengekspor Data dari Rapidminer

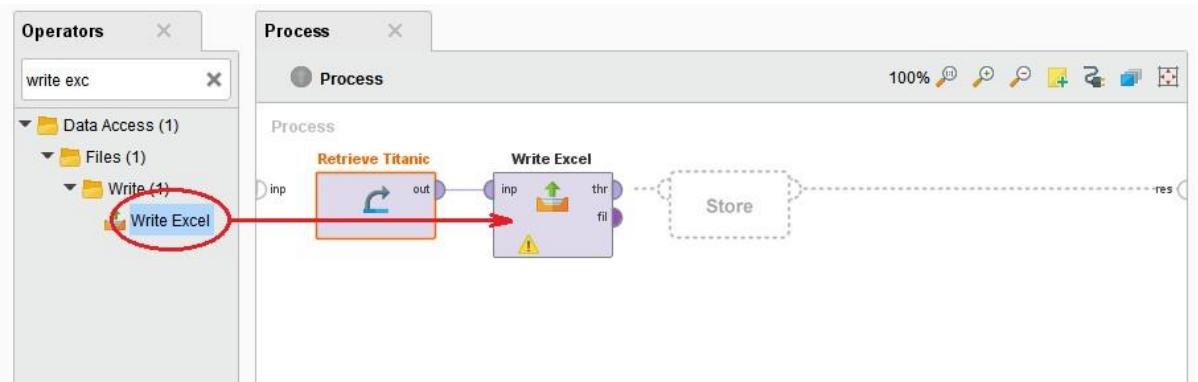
Sampai sejauh ini, kita hanya berfokus untuk memanipulasi data di RapidMiner, namun kita mungkin bertanya: bagaimana cara kita mendapatkan data dari RapidMiner? Misalnya untuk kolaborasi, analisis lanjutan, atau redundansi, di mana mengekspor data merupakan tugas penting untuk dapat dilakukan. Tutorial ini akan menunjukkan bagaimana dengan mudah mengekspor data kita pada RapidMiner.

**Langkah 1.** Menuliskan data ke file.

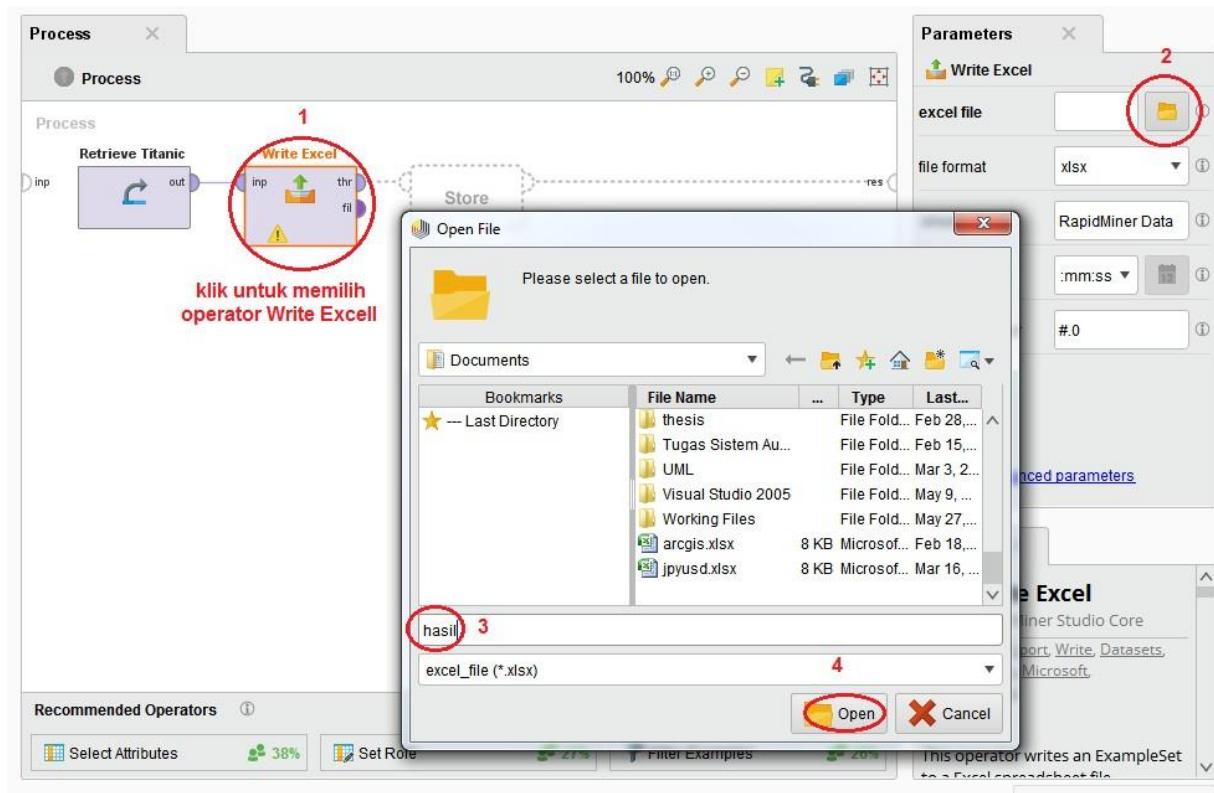
1. Tarik data **Titanic** ke dalam proses.



2. Cari operator **Write Excel** dan seret ke dalam proses, hubungkan operator.



3. Di **Parameters** dari **Write Excel**, tentukan lokasi file untuk file parameter excel. Pastikan bahwa Anda memiliki ijin untuk menulis file ke lokasi ini.



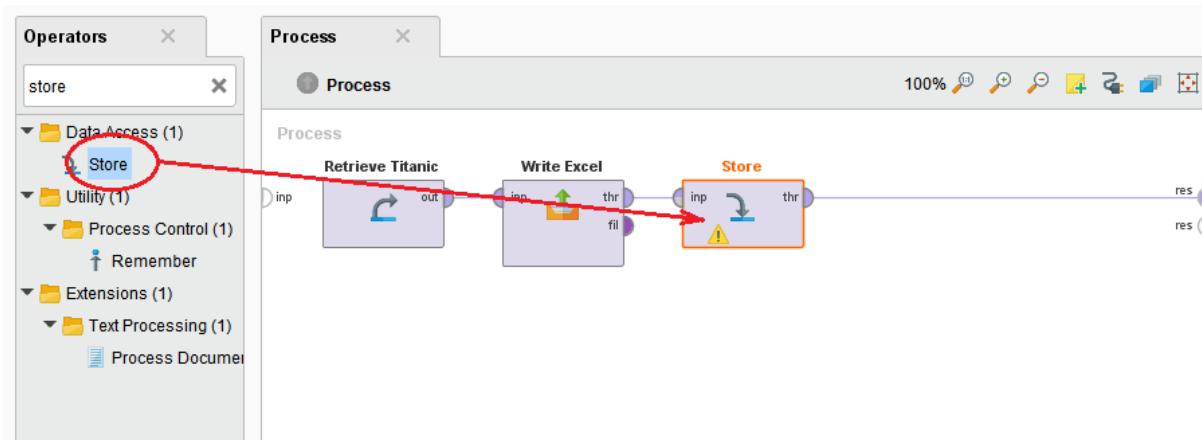
Kita dapat menentukan path secara bebas untuk lokasi file Excel. Operator akan mengambil data Titanic dan menulis dataset lengkap ke dalam file excel. Ada puluhan operator yang tersedia untuk format file yang berbeda, serta format untuk penulisan ke database. Kita juga dapat menemukan operator untuk menulis hasilnya kembali ke aplikasi untuk bisnis, seperti sistem CRM atau ERP. Jenis koneksi ke aplikasi bisnis lainnya sangat penting. Kapan pun kita mengambil keputusan bisnis dengan menggunakan model prediktif, kita dapat mengoperasionalkan permodelannya.

Jika kita membutuhkan lebih banyak operator, misalnya untuk penulisan ke dalam format yang berbeda, cobalah tautan Get More Operators di bagian bawah panel Operator. Tautan ini akan membuka RapidMiner Marketplace yang menawarkan ekstensi ratusan operator tambahan.

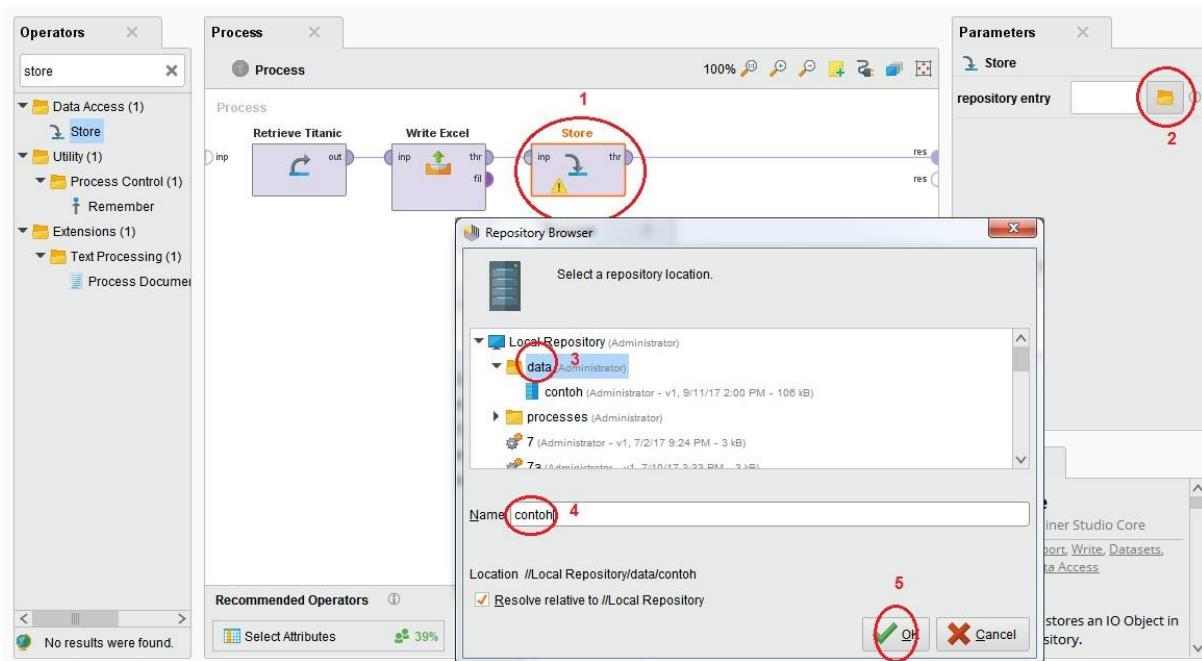
## Langkah 2. Menyimpan ke dalam repositori.

Bila memungkinkan, kita sangat merekomendasikan penggunaan repositori atau database, bukan file, karena RapidMiner juga menyimpan data meta di repositori. Data meta ini kemudian digunakan selama proses desain. Misalnya, RapidMiner dapat menggunakannya untuk menampilkan nama atribut yang berbeda pada dialog pilihan. Kita telah melihat sebelumnya bahwa kita dapat menyimpan data dalam repositori dengan mengimportnya. Sekarang, kita akan belajar bagaimana cara menggunakan operator untuk menyimpan data, model, atau hasil lainnya ke dalam repositori.

1. Tambahkan operator **Store** ke proses dan hubungkan.

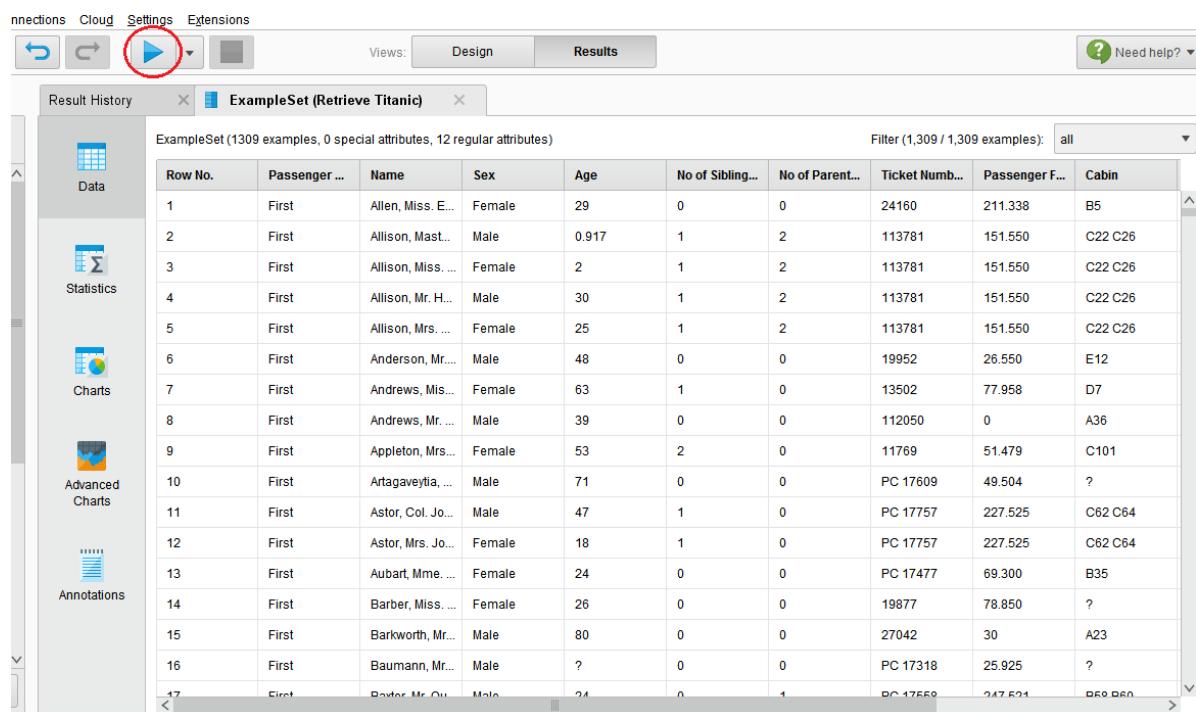


2. Tentukan parameter entri repositori dan pilih lokasi di **Repositori Lokal** Anda.



Hubungkan output **Store** juga dengan port "res" di sebelah kanan.

### 3. Jalankan prosesnya.



The screenshot shows the KNIME interface with the 'Results' tab selected. The toolbar at the top has a play button icon circled in red. The main area displays a table titled 'ExampleSet (Retrieve Titanic)' with 1309 rows. The columns include Row No., Passenger Class, Name, Sex, Age, No of Sibling..., No of Parent..., Ticket Number, Passenger F..., and Cabin. The table shows various passengers from the Titanic dataset, such as Allen, Miss. E., Allison, Mast., and Astor, Col. J. O.

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Numbr...	Passenger F...	Cabin
1	First	Allen, Miss. E...	Female	29	0	0	24160	211.338	B5
2	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26
3	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26
4	First	Allison, Mr. H...	Male	30	1	2	113781	151.550	C22 C26
5	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26
6	First	Anderson, Mr...	Male	48	0	0	19952	26.550	E12
7	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7
8	First	Andrews, Mr. ...	Male	39	0	0	112050	0	A36
9	First	Appleton, Mrs...	Female	53	2	0	11769	51.479	C101
10	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504	?
11	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525	C62 C64
12	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	C62 C64
13	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35
14	First	Barber, Miss. ...	Female	26	0	0	19877	78.850	?
15	First	Barkworth, Mr...	Male	80	0	0	27042	30	A23
16	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925	?
17	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17550	217.521	D60 D60