

# JOBSHEET 1

## PENGENALAN RAPIDMINER DAN DATA PREPARATION

### 1.1. Tujuan Praktikum

Setelah menyelesaikan praktikum ini, mahasiswa mampu:

- Menggunakan perangkat RapidMiner dengan baik
- Mengimplementasikan tahapan Data Preparation menggunakan RapidMiner

### 1.2. Peralatan yang dibutuhkan

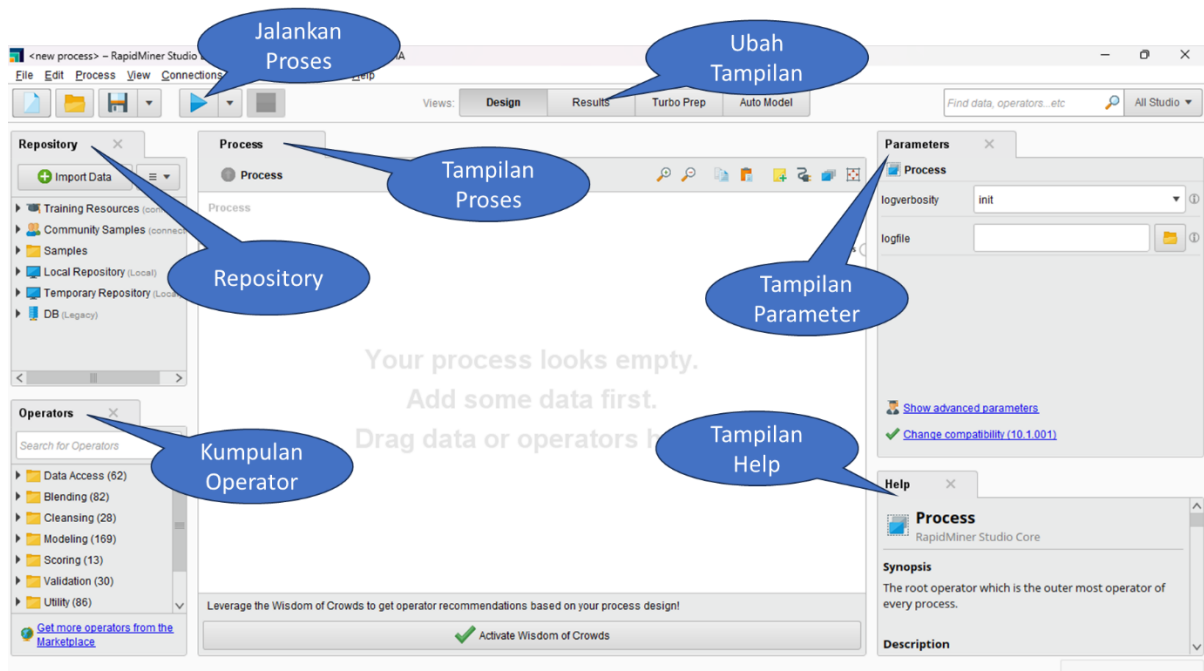
Beberapa peralatan yang dibutuhkan dalam menyelesaikan praktikum ini adalah:

- Aplikasi Notepad atau sejenisnya
- Aplikasi RapidMiner versi 7 ke atas
- Aplikasi Microsoft Excel

### 1.3. Dasar Teori

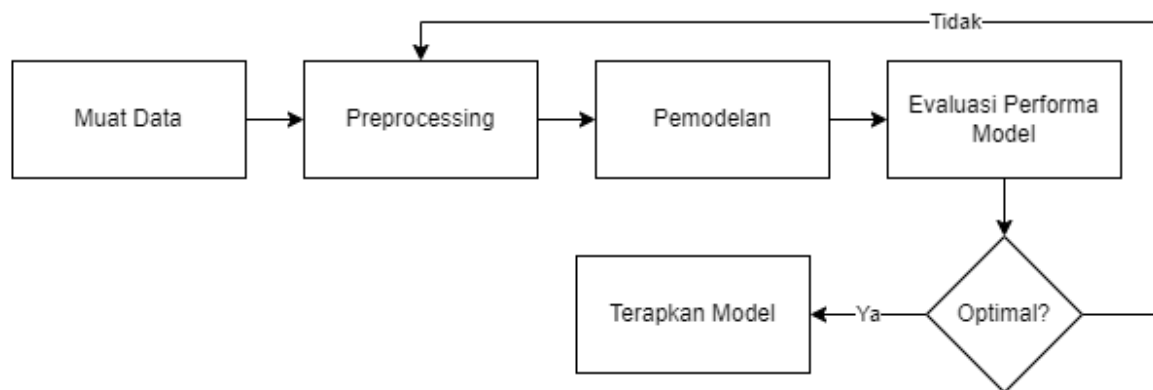
#### 1.3.1. Pengenalan RapidMiner

RapidMiner adalah platform analisis data yang kuat dan intuitif yang memungkinkan pengguna untuk dengan cepat mempersiapkan data, menerapkan berbagai teknik analisis, dan membangun model prediktif kompleks tanpa perlu menulis kode. Dengan antarmuka pengguna yang ramah, RapidMiner menyediakan berbagai alat visual untuk menjalankan langkah-langkah analisis data, termasuk pembersihan data, transformasi, pengelompokan, dan pemodelan. Selain itu, platform ini mendukung integrasi dengan berbagai sumber data dan menyediakan katalog ekstensi yang luas untuk memperluas fungsionalitasnya. Dengan kombinasi kekuatan analisis yang mendalam dan kemudahan penggunaan, RapidMiner memungkinkan pengguna dari berbagai latar belakang untuk mengeksplorasi dan memahami data mereka dengan efisien, membantu mereka mengambil keputusan yang lebih baik dan mendapatkan wawasan berharga dari data mereka.

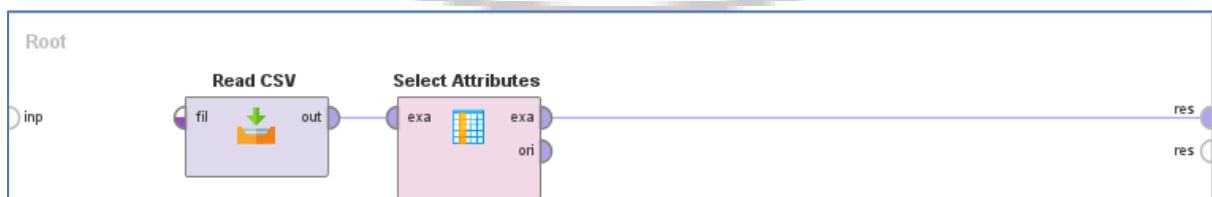


Gambar 1 Tampilan RapidMiner

Ketika menggunakan RapidMiner, perlu diingat bahwa algoritma yang selalu digunakan oleh RapidMiner untuk setiap proses yang dilakukan adalah seperti pada Gambar 2. Tahapan yang diperlukan adalah Pemuatan data ke RapidMiner, melakukan pra-proses data agar sesuai dengan kebutuhan masukan dari model yang akan digunakan, mengevaluasi hasil performa mode, dan menerapkan model.



Gambar 2 Algoritma Proses dari RapidMiner



Gambar 3 Operator dan Port pada RapidMiner

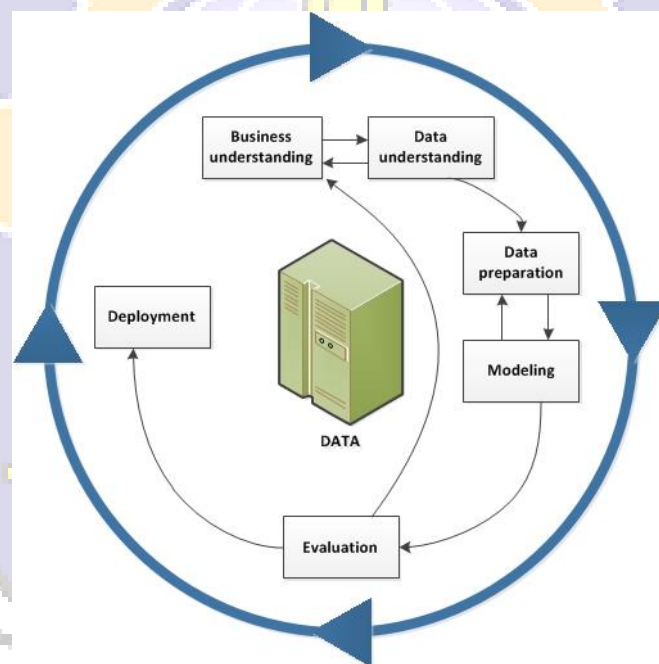
Gambar 3 menunjukkan penggunaan operator pada RapidMiner dalam membaca file data berupa csv dengan menggunakan Operator Read CSV dan memilih atribut yang dibutuhkan dengan menggunakan Operator Select Attribute. Setiap operator yang digunakan dalam RapidMiner, terdapat port pada bagian kiri dan kanan yang disertai dengan penamaan yang sesuai dengan hasil keluaran operator yang digunakan. Pada Tabel 1 menunjukkan beberapa port yang sering muncul. Untuk keterangan lebih lanjut dari deskripsi port yang ada di RapidMiner dapat dibaca pada <https://docs.rapidminer.com/9.4/studio/getting-started/important-terms.html#port-info>

Tabel 1 Nama Port dan Deskripsinya

Nama Port	Deskripsi
Fil	File
Out	Output
Exa	Example Set
Ori	Original
Res	Result

### 1.3.2. Data Preparation

Tahapan CRISP-DM seperti yang terdapat pada Gambar 1, terdiri dari 6 tahapan:



Gambar 4 Tahapan CRISP-DM

Pada praktikum ini, berfokus pada tahapan Data Preparation dimana untuk bisa melakukan pada tahap ini, maka tahapan Business Understanding dan Data Understanding telah dilakukan sebelumnya. Untuk masing-masing tahapan dari Business Understanding sampai dengan Data Preparation, yang perlu dilakukan adalah:

a) **Business Understanding:**

Tujuan dari tahapan ini adalah mendapatkan informasi sebanyak mungkin tentang Proses Bisnis, tahapan proses bisnis yang mengalami permasalahan, dan tahapan proses bisnis apa saja yang terpengaruh. Beberapa hal yang perlu dilakukan adalah:

- i) Menentukan Tujuan Bisnis
- ii) Menilai Situasi
- iii) Menentukan Tujuan Data Mining
- iv) Menghasilkan Rencana Proyek

**b) Data Understanding**

Fase pemahaman data CRISP-DM melibatkan melihat lebih dekat data yang tersedia untuk penambangan. Langkah ini penting untuk menghindari masalah yang tidak terduga pada fase berikutnya--persiapan data--yang biasanya merupakan bagian terpanjang dari sebuah proyek. Beberapa hal yang perlu dilakukan adalah:

- i) Mengumpulkan Data Awal
- ii) Menjelaskan Data
- iii) Menjelajahi Data
- iv) Memverifikasi Kualitas Data

**c) Data Preparation**

Mempersiapkan data mentah menjadi dataset yang bersih, terstruktur, dan siap untuk diolah lebih lanjut dalam proses analisis data dan pemodelan. Beberapa hal yang dilakukan dalam tahapan ini adalah:

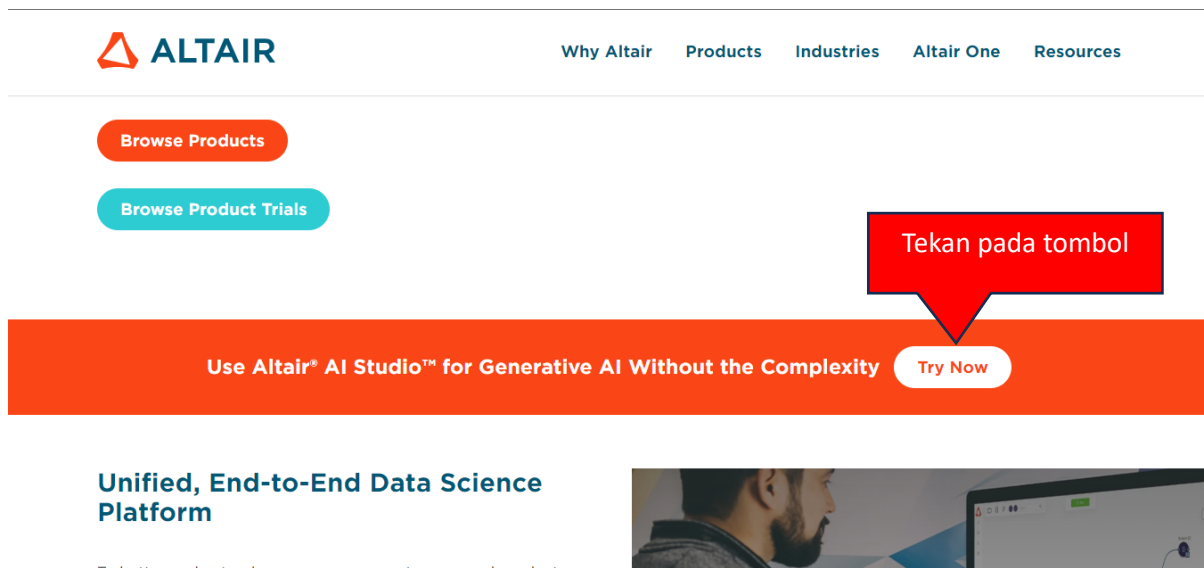
- i) Menggabungkan kumpulan data dan/atau catatan
- ii) Memilih sampel subset data
- iii) Mengumpulkan catatan
- iv) Mendapatkan atribut baru
- v) Menyortir data untuk pemodelan
- vi) Menghapus atau mengganti nilai yang kosong atau hilang
- vii) Membagi menjadi kumpulan data pelatihan dan pengujian

## **1.4. Praktikum**

### **1.4.1. Instalasi RapidMiner Studio**

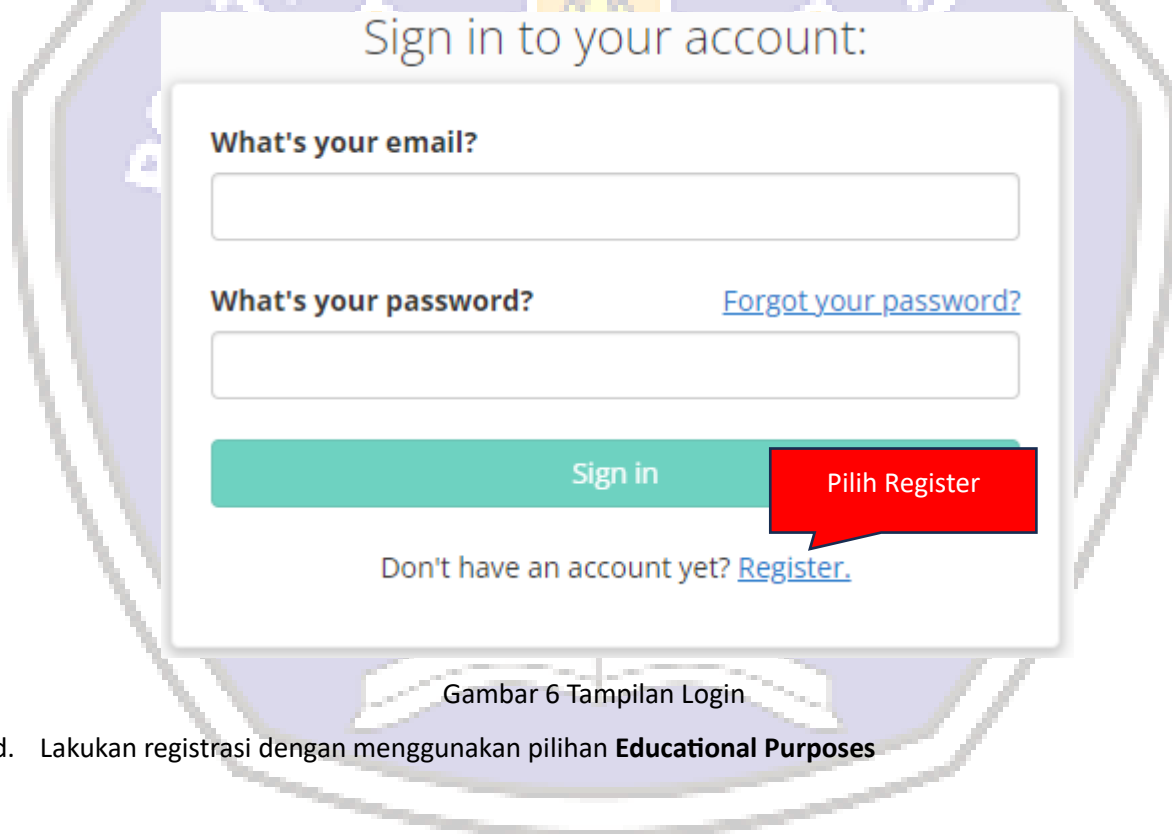
Beberapa kegiatan yang perlu dilakukan dalam melakukan instalasi RapidMiner

- a. Download installer RapidMiner studio pada <https://altair.com/altair-rapidminer>
- b. Tekan pada tombol “Try Now” seperti pada Gambar 5.



Gambar 5 Tombol untuk download

- c. Setelah menekan tombol, akan muncul halaman Log In, pilih pada bagian Register



Gambar 6 Tampilan Login

- d. Lakukan registrasi dengan menggunakan pilihan **Educational Purposes**

What are you using Rapidminer for?

☒ Commercial purposes (e.g., business, evaluation, not-for-profit)

☐ Educational purposes (e.g., educator, student)

First name:

Last name:

Country:

Organization:

Your company name is required

Role:

Gambar 7 Tampilan Registrasi

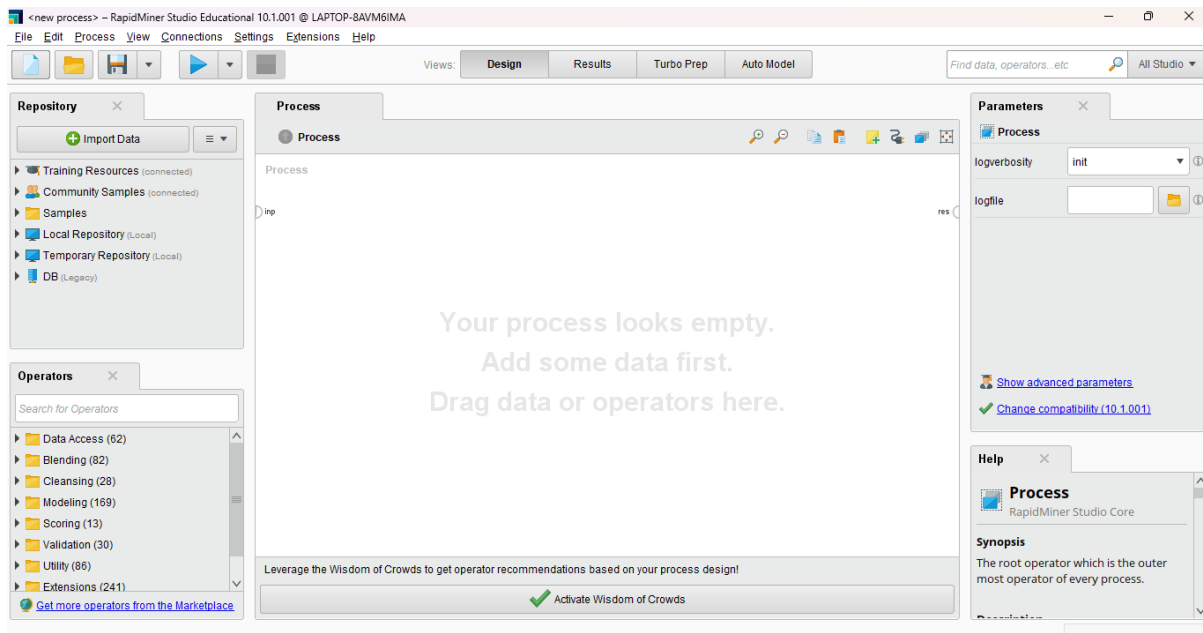
e. Selanjutnya adalah melakukan masuk pada proses download dan install hasil download

#### 1.4.2. Verifikasi Hasil Instalasi RapidMiner Studio

Pastikan hasil dari praktikum instalasi RapidMiner Studio seperti pada Gambar 8 dan siap untuk digunakan seperti pada Gambar 9.



Gambar 8 Proses Membuka RapidMiner Studio



Gambar 9 RapidMiner telah terbuka dan siap digunakan

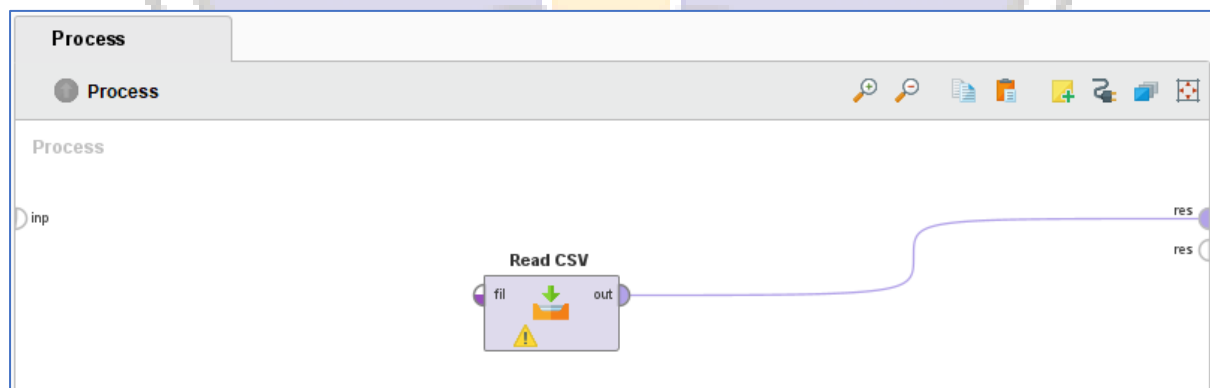
#### Pertanyaan 1

1. Berapakah besar file yang harus didownload dari untuk mendapatkan Installer dari RapidMiner Studio?
2. Berapa lama yang diperlukan dalam proses download jika dilakukan pada Laboratorium?
3. Berapa storage yang diperlukan ketika sudah melakukan instalasi?
4. Tuliskan versi RapidMiner dan Spesifikasi komputer / laptop yang digunakan (jika tidak dapat dijalankan dengan baik, tuliskan juga spesifikasi minimal yang diperlukan)!

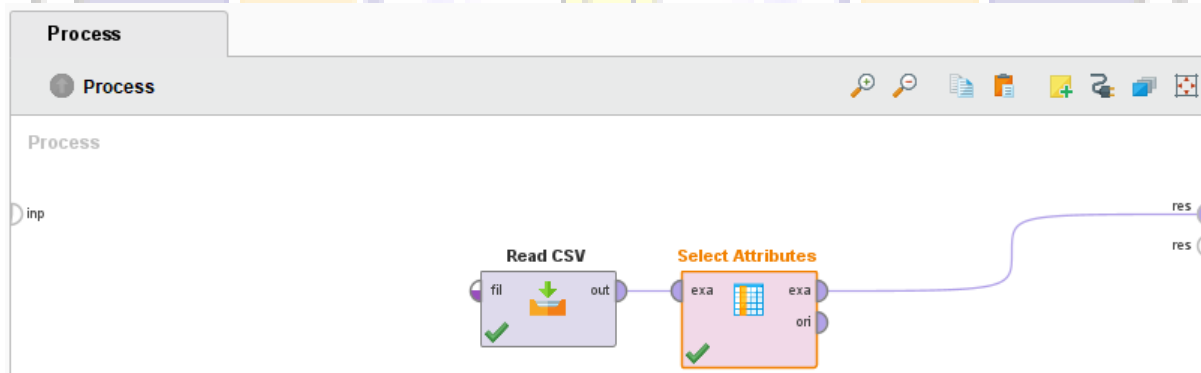
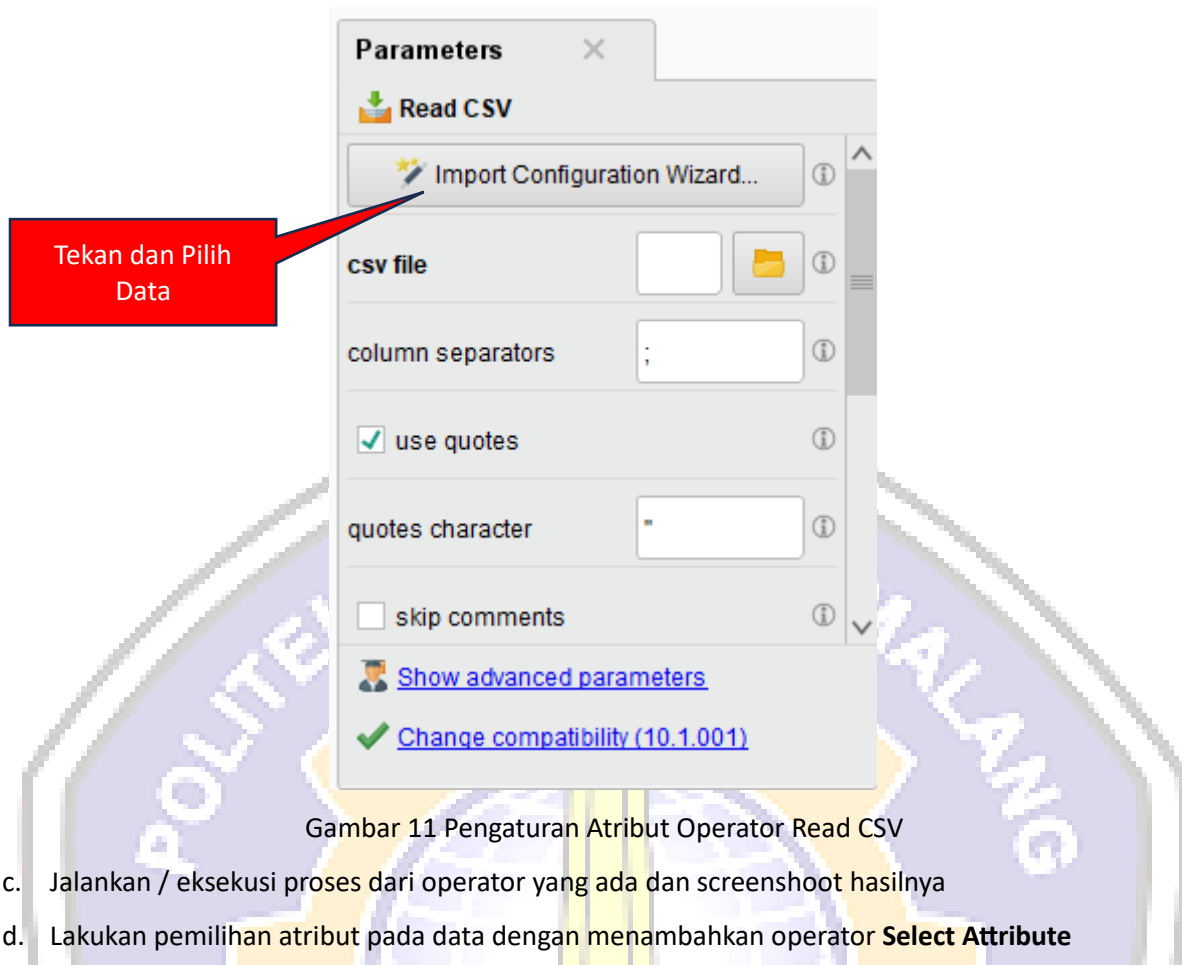
#### 1.4.3. Data Preparation

Lakukan tahapan Praktikum sesuai dengan urutannya di bawah ini:

- a. Download DataPraktek01.csv di LMS
- b. Load Data menggunakan operator Read CSV pada RapidMiner



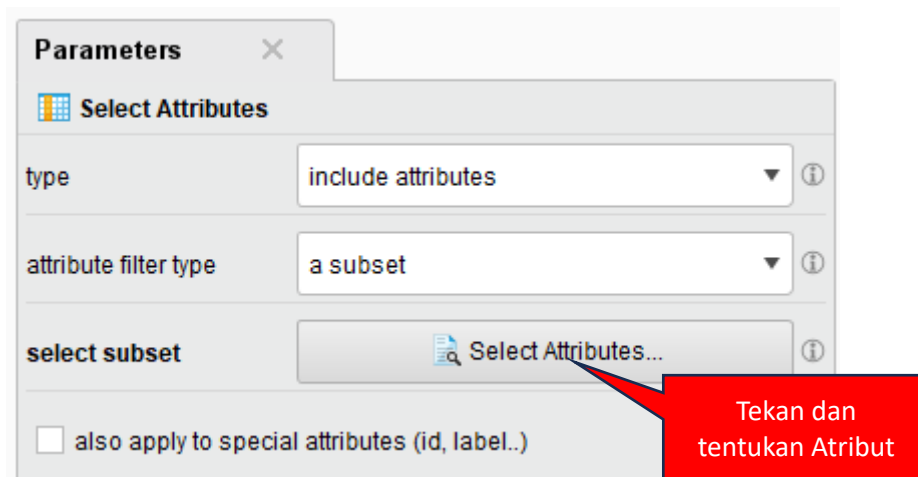
Gambar 10 Penggunaan Operator Read CSV



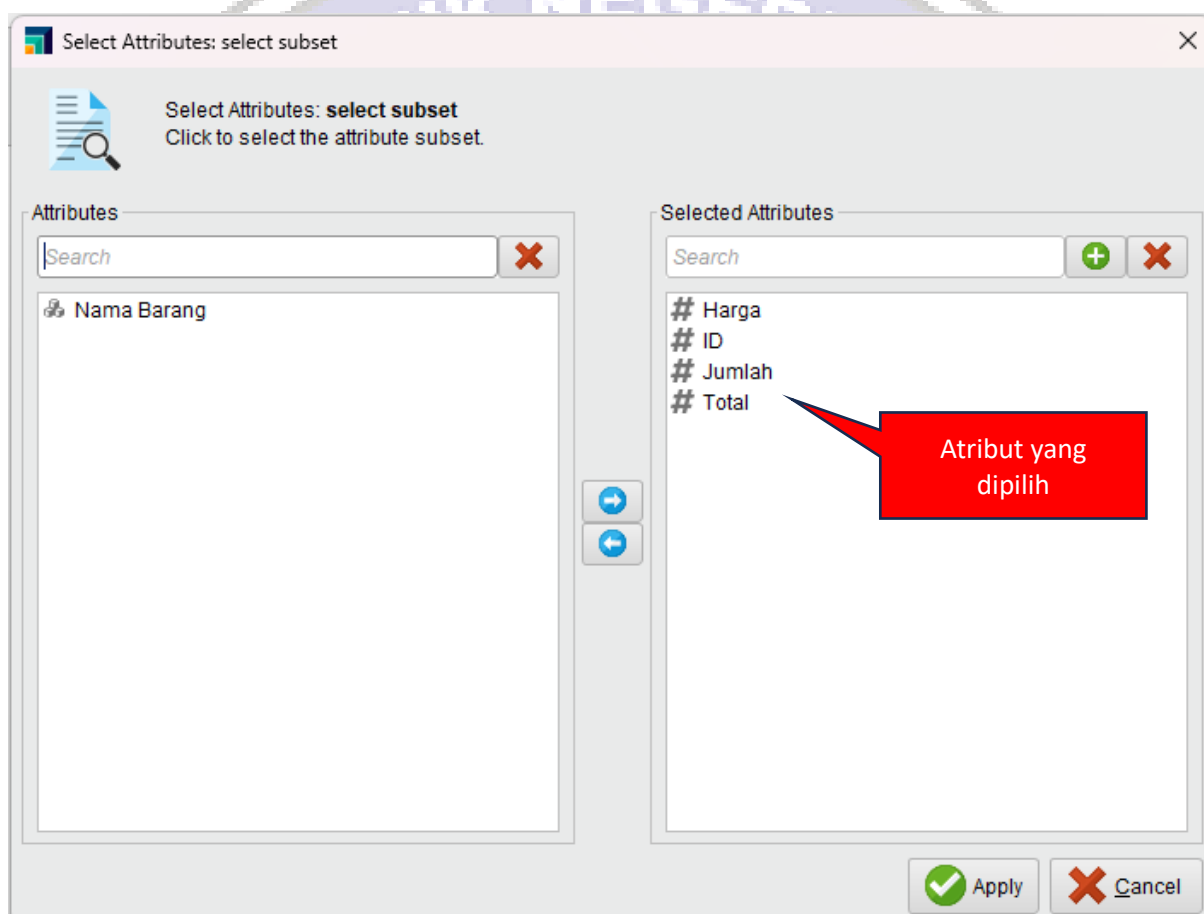
Gambar 12 Penambahan Operator Select Attribute

- e. Lakukan pengaturan atribut dari Operator **Select Attribute**





Gambar 13 Pengaturan Atribut dari Operator Select Attribute





Gambar 14 Pemilihan Atribut dari Data

f. Jalankan / eksekusi proses dari operator yang ada dan screenshot hasilnya.

#### 1.4.4. Verifikasi Hasil Data Preparation

Pastikan Luaran dari praktikum yang dilakukan sesuai dengan gambar yang ada di bawah ini.

ExampleSet (Read CSV)



Open in  Turbo Prep  Auto Model

Filter (10 / 10 examples): all

Row No.	ID	Nama Barang	Jumlah	Harga	Total
1	1	Buku Gambar	9	5000	45000
2	2	Pensil	8	3000	24000
3	3	Penggaris	8	1000	8000
4	4	Penghapus	5	1000	5000
5	5	Buku Tulis	4	6000	24000
6	6	Tas	7	25000	175000
7	7	Kotak Pensil	5	15000	75000
8	8	Buku Gambar	1	5000	5000
9	9	Pensil	2	3000	6000
10	10	Pensil	9	3000	27000

Gambar 15 Hasil Eksekusi Read CSV

ExampleSet (Select Attributes)

Open in  Turbo Prep  Auto Model

Filter (10 / 10 examples): all

Row No.	ID	Jumlah	Harga	Total
1	1	9	5000	45000
2	2	8	3000	24000
3	3	8	1000	8000
4	4	5	1000	5000
5	5	4	6000	24000
6	6	7	25000	175000
7	7	5	15000	75000
8	8	1	5000	5000
9	9	2	3000	6000
10	10	9	3000	27000

Gambar 16 Hasil Eksekusi Pemilihan Atribut

## Pertanyaan 2

1. Apakah fungsi dari Operator Read CSV?
2. Jelaskan perbedaan dari data asli dengan hasil dari penggunaan Operator Read CSV!
3. Apakah fungsi dari Select Attribute?
4. Jelaskan perbedaan hasil dari sebelum dan sesudah dilakukan Pemilihan Atribut!

## 1.4.5. Tugas Praktikum

1. Gunakan data **DataPraktek02.csv** yang ada di LMS!
2. Jelaskan dan tunjukkan hasil dari tahapan Load Data menggunakan Operator Read CSV dan simpan hasil eksekusi menggunakan tampilan **Visualizations**!
3. Jelaskan dan tunjukkan hasil dari tahapan pemilihan atribut dengan hanya menampilkan data **Kode, Jumlah Siswa, Range Nilai, dan Kehadiran**!
4. Tunjukkan hasil eksekusi dalam bentuk **Statistics** dan jelaskan maksudnya dari masing-masing atribut!