

# Pengenalan RapidMiner dan Data Preparation

Pertemuan 2

Data Mining

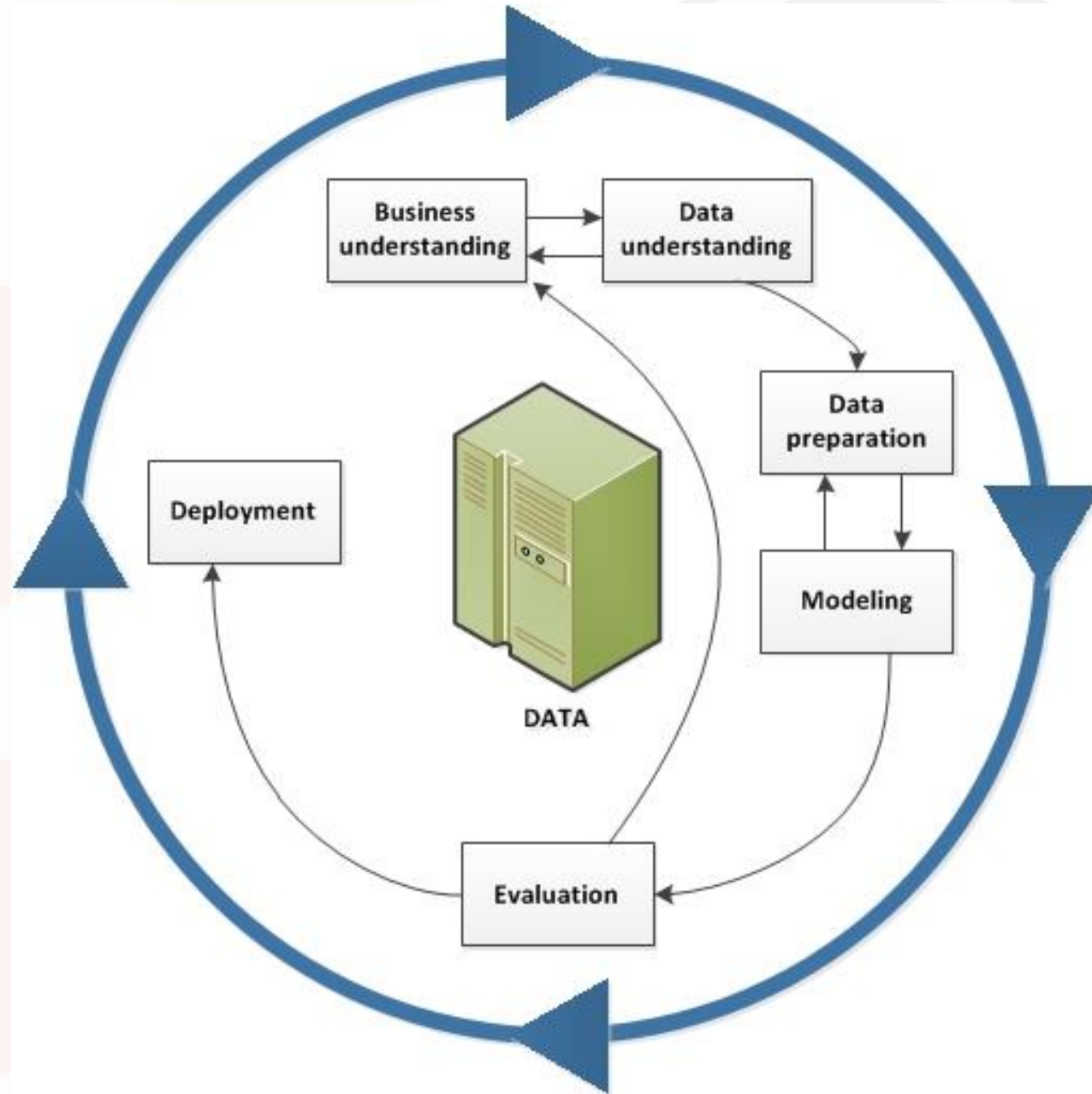
Tim Teaching

Jurusan Teknologi Informasi

2024

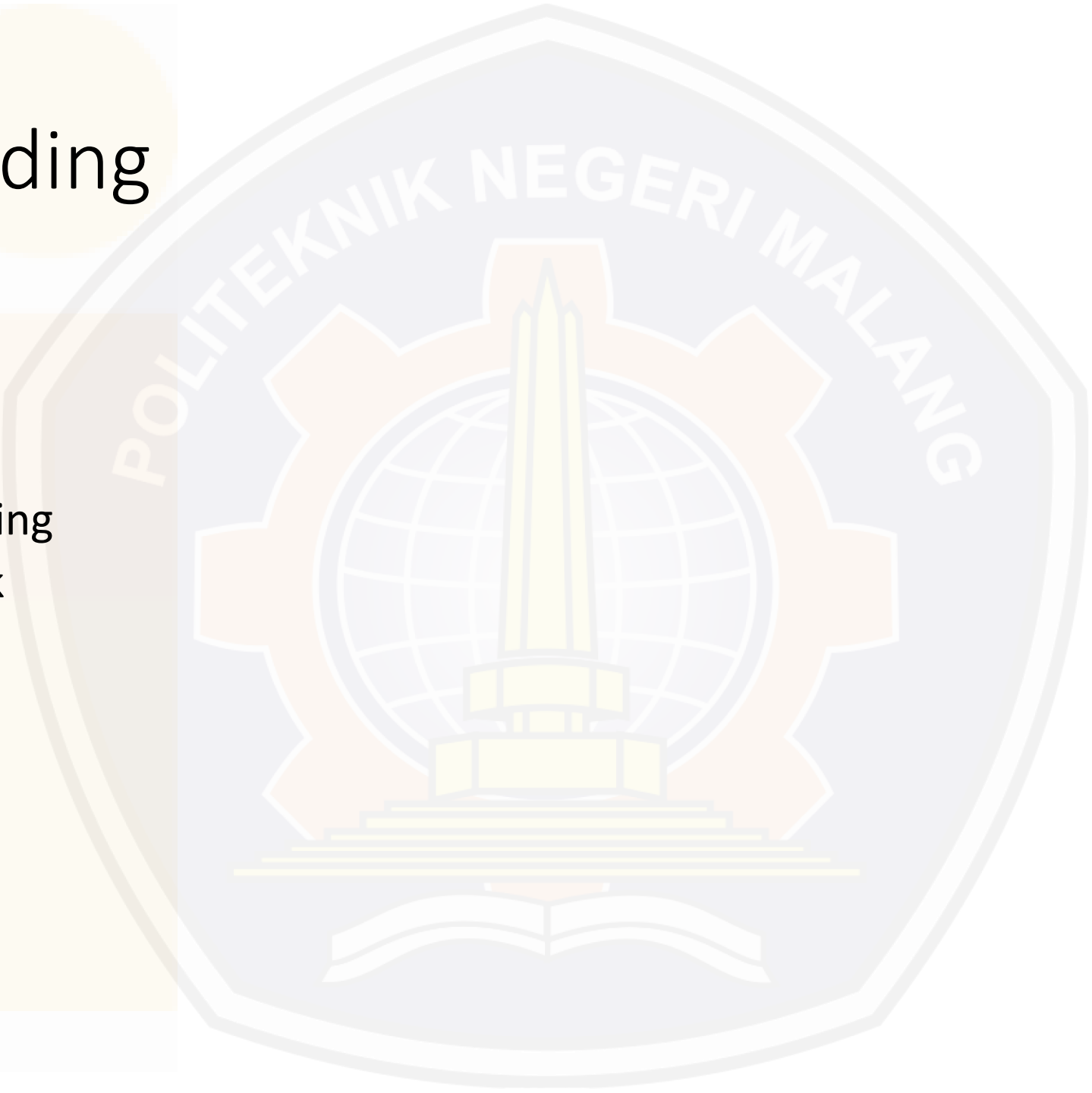
# CRISP-DM

<https://www.ibm.com/docs/zh/spss-modeler/18.0.0?topic=guide-business-understanding>

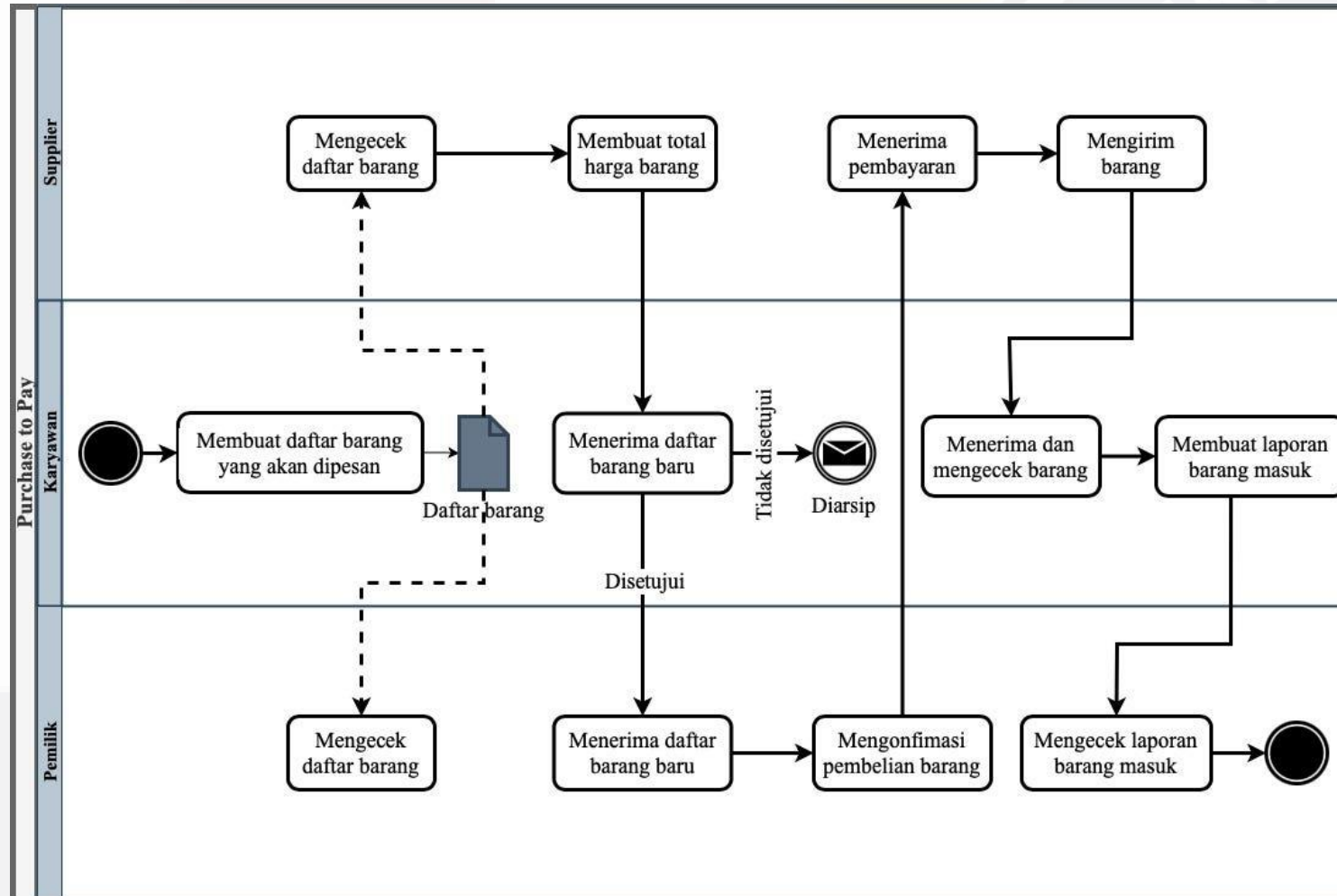


# Business Understanding

- Yang perlu dilakukan:
  - Menentukan Tujuan Bisnis
  - Menilai Situasi
  - Menentukan Tujuan Data Mining
  - Menghasilkan Rencana Proyek



# Business Understanding (Contoh)



create by. CP157-Khairunisa Maharsanti-UPY-Akuntansi

Proses Bisnis Purchase to Pay  
hasil wawancara dengan pihak  
Industri

Beberapa Pertanyaan yang perlu  
dijawab:

- Pada tahap mana yang mengalami permasalahan?
- Data apa saja yang berhubungan dengan tahap tersebut?
- Data Mining digunakan untuk memberikan Solusi seperti apa?
- Perencanaan apa saja yang diperlukan untuk menerapkan Data Mining?

# Data Understanding

- Yang perlu dilakukan:
  - Mengumpulkan Data Awal
  - Menjelaskan Data
  - Menjelajahi Data
  - Memverifikasi Kualitas Data



# Data Understanding (Contoh)

```
ID;Nama Barang;Jumlah;Harga;Total
1;Buku Mewarnai;9;5000;45000
2;Pensil;8;3000;24000
3;Penggaris;8;1000;8000
4;Penghapus;5;1000;5000
5;Buku Tulis;4;6000;24000
6;Tas;7;25000;175000
7;Kotak Pensil;5;15000;75000
8;Buku Gambar;1;5000;5000
9;Stabilo;2;4000;8000
10;Bolpoin;9;3000;27000
```

- Data Awal yang dapat dikumpulkan adalah data penjualan dalam bentuk CSV
- Penjelasan data:
  - Data terdiri dari 5 kolom:
    - ID : Kode Barang → Arti Kode: ...
    - Nama Barang : Nama Barang
    - Jumlah : Jumlah barang keluar
    - Harga: Harga Jual
    - Total : Jumlah x harga

# Data Preparation

- Yang Perlu dilakukan:
  - Memilih Data
  - Pembersihan Data
  - Membangun Data Baru
  - Mengintegrasikan Data
  - Memformat Data
- Bisa menggunakan RapidMiner



# Pengenalan RapidMiner

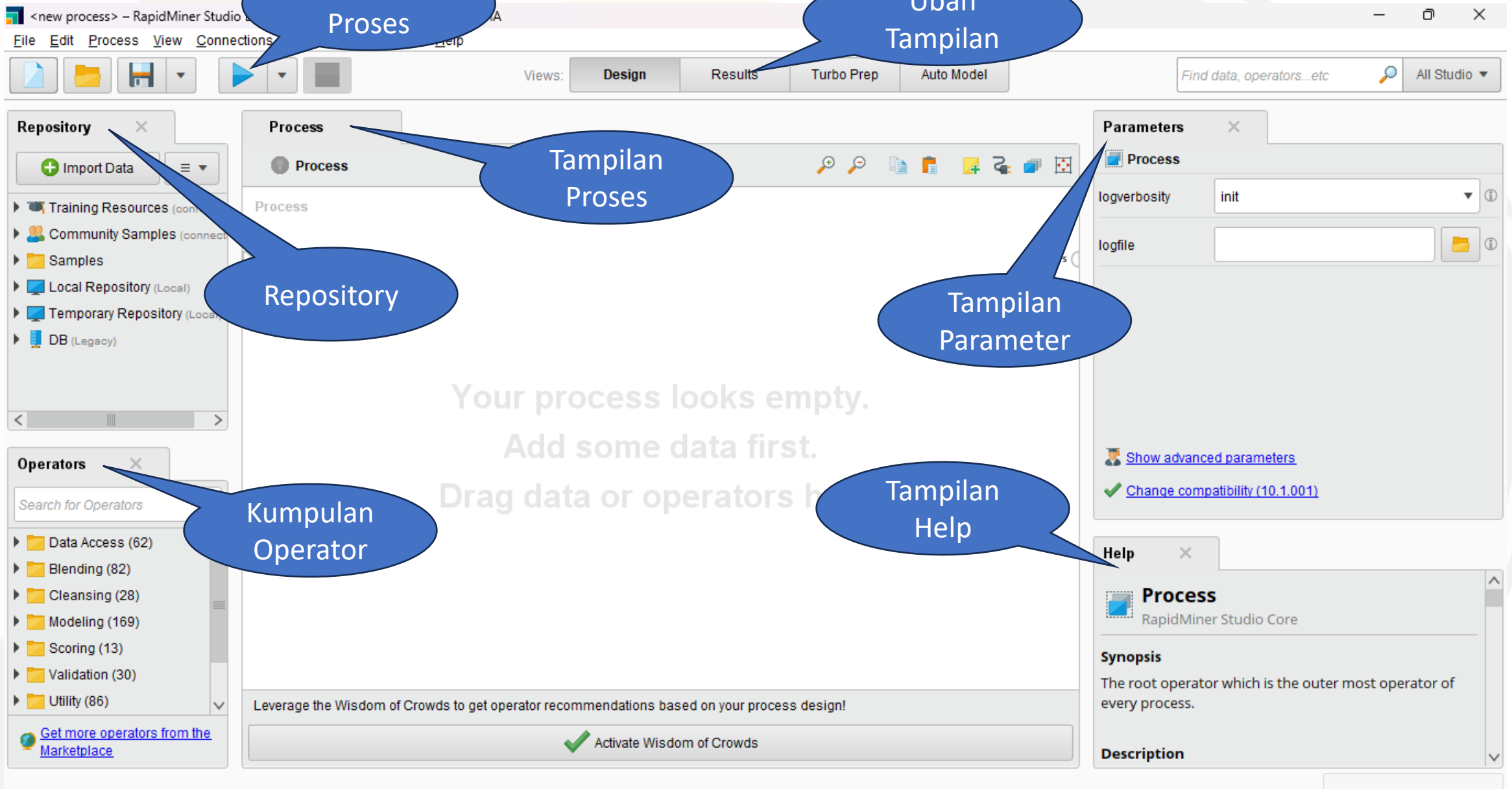




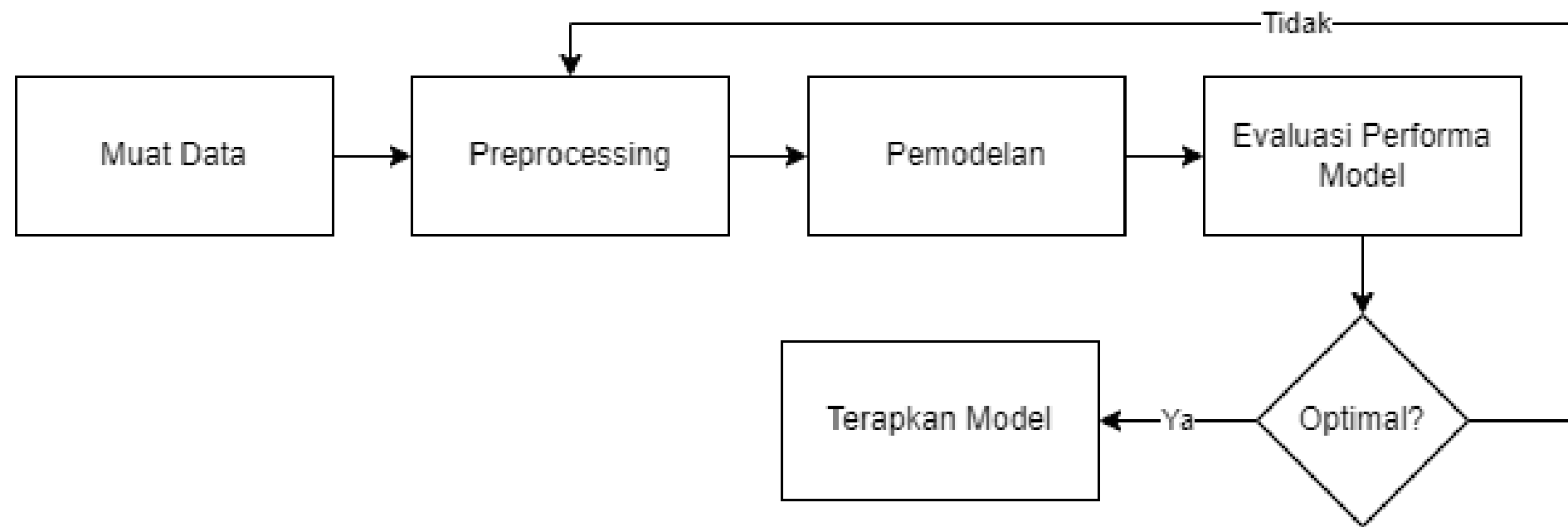
# RapidMiner



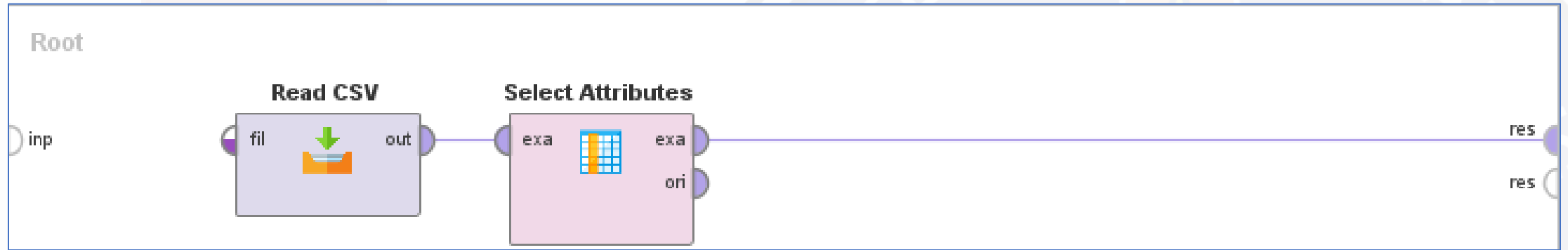
- Platform analitik yang kuat dan serbaguna yang digunakan untuk melakukan analisis data, pemodelan prediktif, dan penambangan data
- Menyediakan lingkungan visual untuk mendefinisikan proses analisis data tanpa perlu menulis kode secara manual
- Pengguna dapat mengimpor data dari berbagai sumber, menyiapkan data, membangun dan mengevaluasi model prediktif, serta menggali informasi yang berharga dari data secara efisien
- Platform ini banyak digunakan di berbagai industri untuk berbagai tujuan, mulai dari analisis keuangan hingga analisis Kesehatan.



# Penggunaan RapidMiner



# Pengenalan Port



No	Nama Port	Deskripsi
1	Fil	File
2	Out	Output
3	Exa	Example Set
4	Ori	Original
5	Res	Result

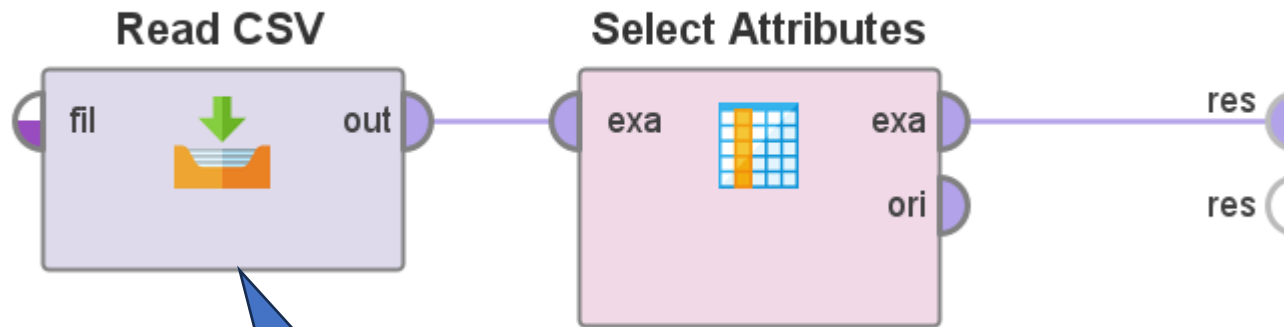
<https://docs.rapidminer.com/9.4/studio/getting-started/important-terms.html#port-info>

Port Output menghasilkan **Example Set**

Sebuah **Example Set** berisikan data

Data pada RapidMiner disebut **Example**

# Pengenalan Operator



Operator

Parameter  
dari Operator  
Read CSV

**Parameters** [X]

**Read CSV**

Import Configuration Wizard...

**csv file** nents\DataPraktek01.csv

**column separators** ;

☒ **use quotes**

**quotes character** "

☒ skip comments

[Show advanced parameters](#)

[Change compatibility \(10.1.001\)](#)

# Data pada RapidMiner

Import Data - Format your columns.

**Format your columns.**

Date format  ☐ Replace errors with missing values ⓘ

	ID <i>integer</i>	Nama Barang <i>polynomial</i>	Jumlah <i>integer</i>	Harga	Total
1	1				
2	2				
3	3				
4	4	Penghapus	5	1000	
5	5	Buku Tulis		6000	
6	6	Tas		25000	
7	7	Kotak Pensil		15000	75000
8	8	Buku Gambar	1	5000	5000
9	9	Stabilo	2	4000	8000
10	10	Bolpoin	9	3000	27000

Annotations:

- Nama Atribut (points to the column header area)
- Tipe Data Atribut (points to the data type dropdown menu)
- Peran Data (points to the context menu options)

Context Menu Options:

- Change Type (selected)
- Change Role
- Rename column
- Exclude column

Change Type Options:

- polynomial
- binominal
- real
- integer (checked)
- date\_time
- date
- time

no problems.

Previous Finish Cancel

# Data Pada RapidMiner

- Tipe Data

<b>Tipe data</b>	<b>Deskripsi</b>	<b>Contoh Data</b>
<b>Binomial</b>	Hanya dua nilai berbeda yang diizinkan	True dan False
<b>Polynomial</b>	Lebih dari dua nilai berbeda diperbolehkan	Sunny, Cloudy, Rainy, True, False, High, Low, 0, 1
<b>Integer</b>	Bilangan bulat, positif dan negatif	1,2,3,-1,-2,-3
<b>Real</b>	Bilangan real, positif dan negatif	0.1, 0.2, 0.3, -0.1, -0.2, -0.3
<b>Date_time</b>	Tanggal dan juga waktu	2008-12-31 00:00:00
<b>Date</b>	Hanya Tanggal	2008-12-31
<b>Time</b>	Hanya Waktu	00:00:01

# Data Pada Rapidminer

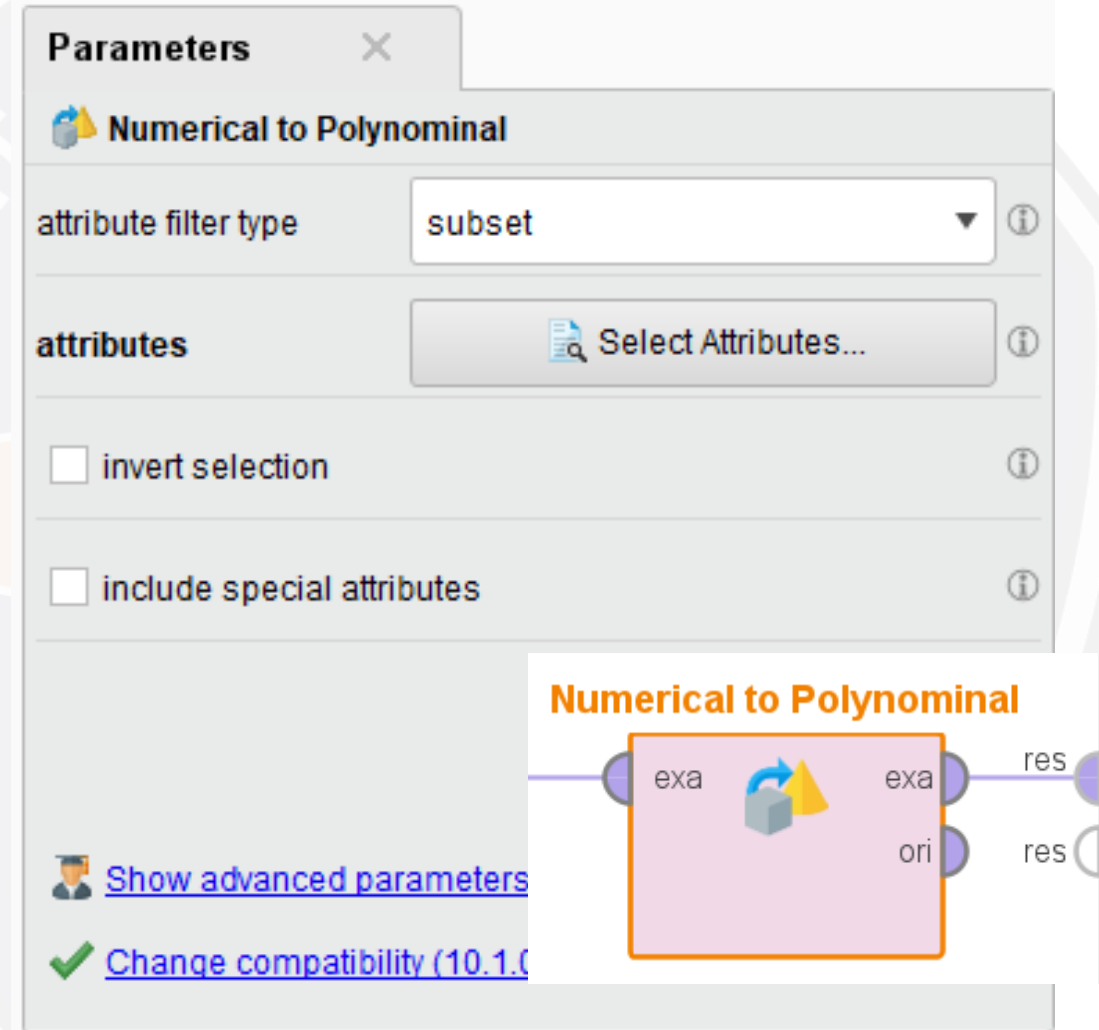
- Peran Data

Role	Description
ID	Pengidentifikasi unik, tidak ada dua Example dalam Example Set yang dapat memiliki nilai yang sama
Regular (default)	Atribut reguler yang berisi data
Label	Atribut target untuk tugas klasifikasi
Weight	Bobot Example berkaitan dengan label
Cluster	Dibuat oleh RapidMiner sebagai hasil tugas pengelompokan
Prediction	Dibuat oleh RapidMiner sebagai hasil tugas klasifikasi



# Jenis Operator Preprocessing

- Type and Role Conversions
  - **TypeA to TypeB**: Ubah tipe data
  - **Set Role**: Ubah peran data
- Attribute Set Transformation
  - **Select Attributes**: Memilih Atribut
  - **Generate Attributes**: Membuat Atribut baru
- Value Transformation
  - **Normalize** : mengubah semua nilai ke kisaran tertentu
- Filtering
  - **Filter Example**: Menyaring Example
- Agregation:
  - **Aggregate** : Agregasi mirip SQL (count, sum)



# Cara menemukan Operator





# Contoh Result View


Result History


ExampleSet (Read CSV (2))

ExampleSet (Read CSV)


Data


Statistics

Visualizations

Annotations

Open in

 Turbo Prep

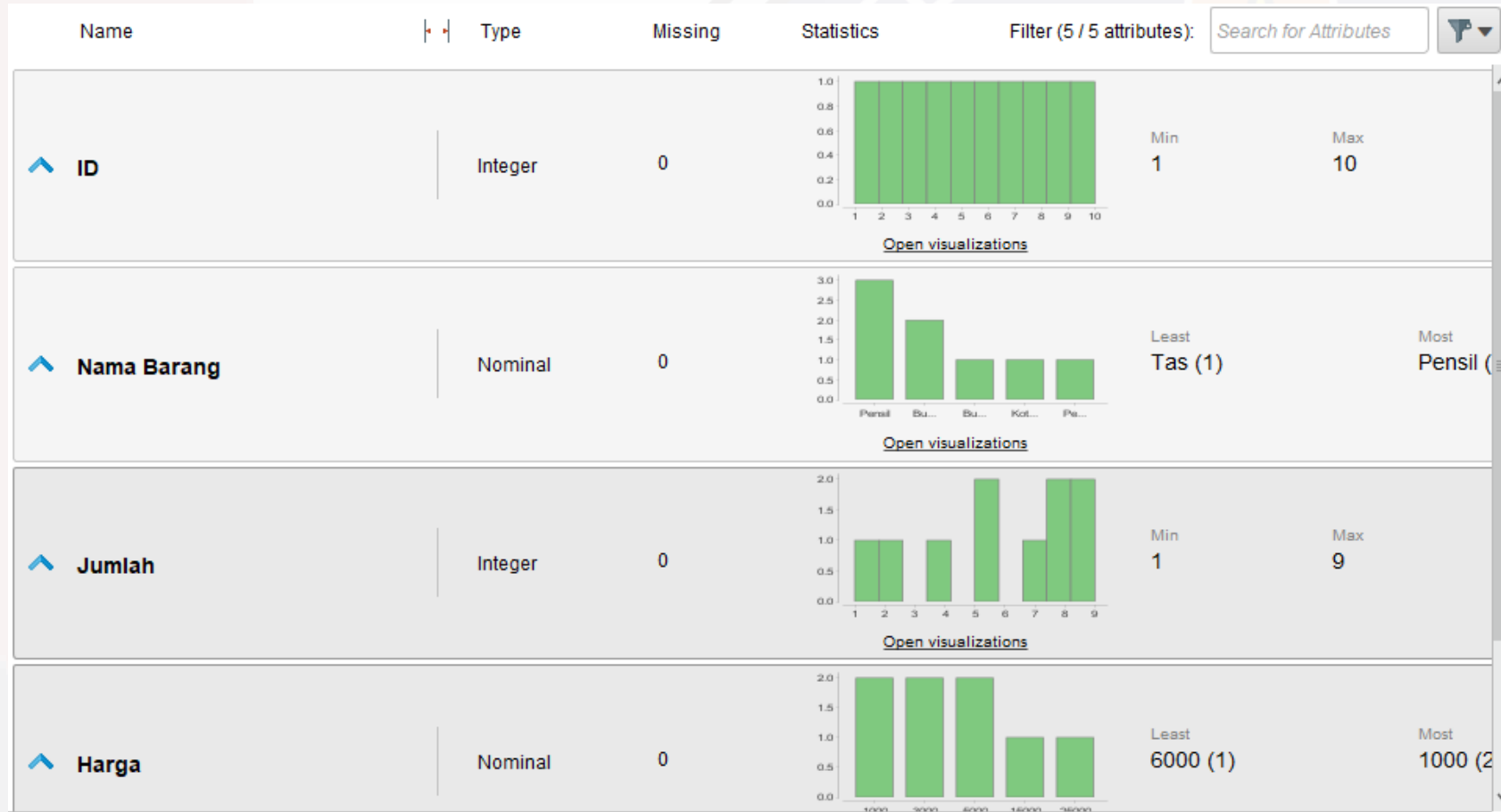
 Auto Model

Filter (10 / 10 examples):

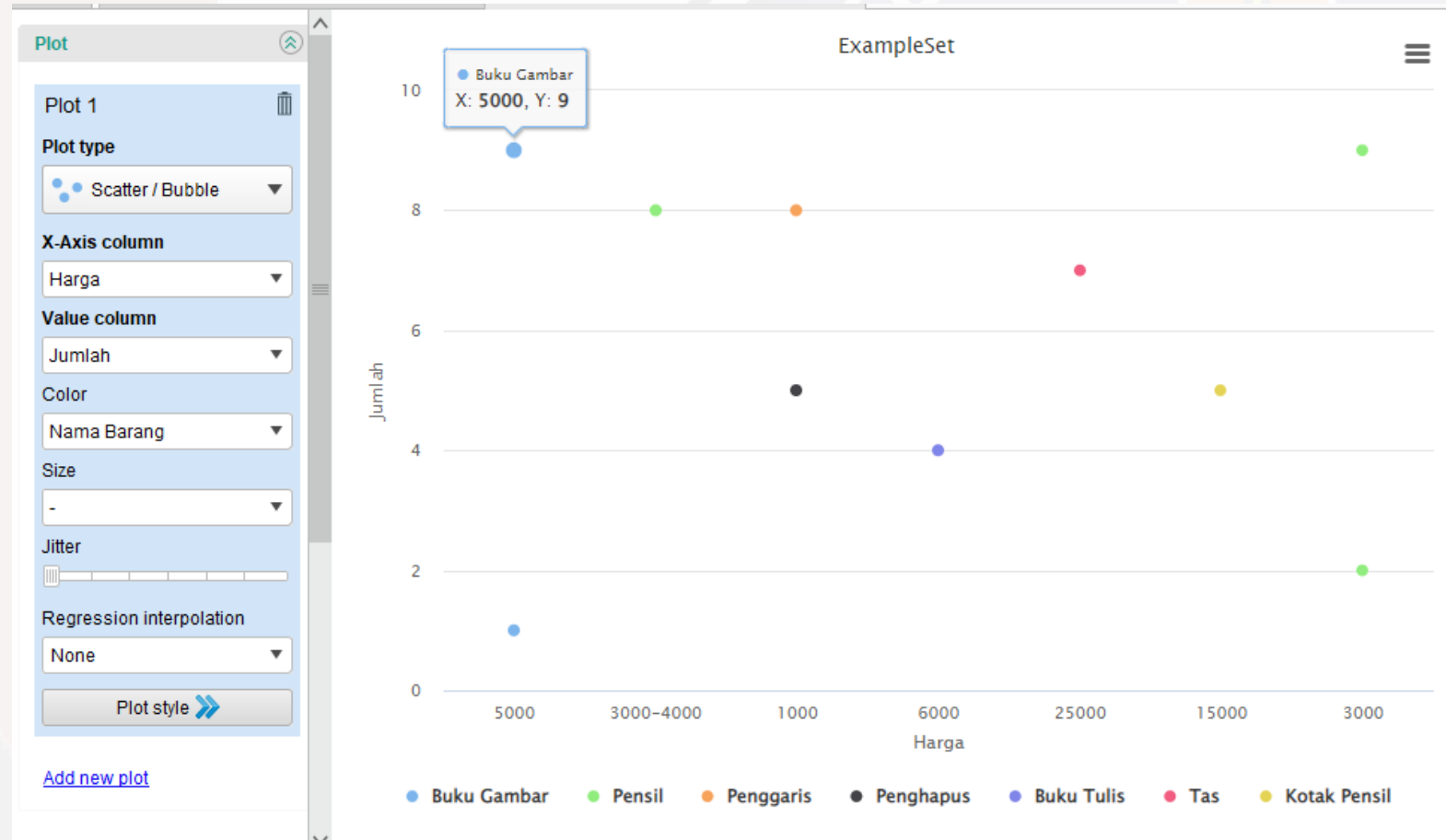
all

Row No.	ID	Nama Barang	Jumlah	Harga	Total
1	1	Buku Mewarnai	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	Stabilo	2	4000	8000
10	10	Bolpoin	9	3000	27000

# Contoh Statistic View



# Contoh Visualisasi Data



PRAKTIKUM

