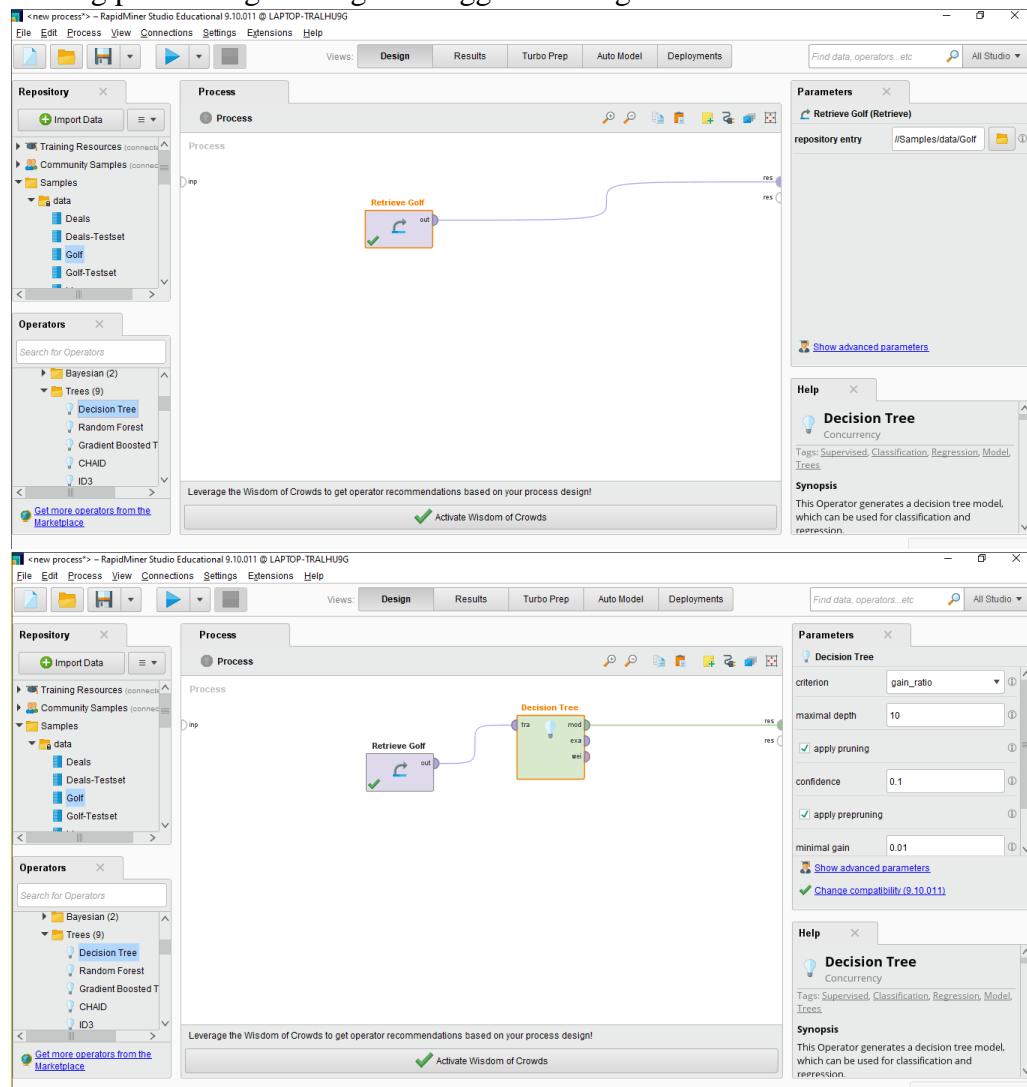


Nama : Yogaraksa Amjad Hernawan  
NIM : 3130020023  
Mata Kuliah : Data Warehouse dan Data Mining

## Data Mining Menggunakan Rapidminer

### 1. Rekomendasi Main Golf

- Trining pada data golf dengan menggunakan algoritma decision tree



**RapidMiner Studio Educational 9.10.011 @ LAPTOP-TRALHU9G**

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Data Statistics Visualizations Annotations

ExampleSet (/Samples/data/Golf) ExampleSet (/Local Repository/data/Golf) ExampleSet (Retrieve Golf)

Result History Filter (14 / 14 examples) all

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

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

Repository Import Data

- Training Resources (connected)
- Community Samples (connected)
- Samples
  - data
    - Deals
    - Deals-Testset
    - Golf
    - Golf-Testset
    - Iris
    - Labor-Negotiations
    - Market-Data
    - Polynomial
    - Products
    - Purchases
    - Ripley-Set
    - Sonar
    - Titanic
    - Titanic Training
    - Titanic Unlabeled
    - Transactions
    - Weighting

- Tampilkan Himpunan data (dataset) dan pengetahuan (model tree) yang terbentuk

**RapidMiner Studio Educational 9.10.011 @ LAPTOP-TRALHU9G**

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Data Statistics Visualizations Annotations

ExampleSet (/Samples/data/Golf) ExampleSet (/Local Repository/data/Golf) Tree (Decision Tree)

Result History

Zoom Tree Node Labels Edge Labels

```

graph TD
    Outlook[Outlook] -- overcast --> yes1[yes]
    Outlook -- rain --> Wind[Wind]
    Outlook -- sunny --> Humidity[Humidity]
    Wind -- false --> yes2[yes]
    Wind -- true --> no1[no]
    Humidity -- >77.500 --> no2[no]
    Humidity -- ≤ 77.500 --> yes3[yes]
  
```

Repository Import Data

- Training Resources (connected)
- Community Samples (connected)
- Samples
  - data
    - Deals
    - Deals-Testset
    - Golf
    - Golf-Testset
    - Iris
    - Labor-Negotiations
    - Market-Data
    - Polynomial
    - Products
    - Purchases
    - Ripley-Set
    - Sonar
    - Titanic
    - Titanic Training
    - Titanic Unlabeled
    - Transactions
    - Weighting

**RapidMiner Studio Educational 9.10.011 @ LAPTOP-TRALHU9G**

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Deployments

Data Statistics Visualizations Annotations

ExampleSet (/Samples/data/Golf) ExampleSet (/Local Repository/data/Golf) Tree (Decision Tree)

Result History

Tree

```

Outlook = overcast: yes (no=0, yes=4)
Outlook = rain
|   Wind = false: yes (no=0, yes=3)
|   Wind = true: no (no=2, yes=0)
Outlook = sunny
|   Humidity > 77.500: no (no=3, yes=0)
|   Humidity ≤ 77.500: yes (no=0, yes=2)
  
```

Repository Import Data

- Training Resources (connected)
- Community Samples (connected)
- Samples
  - data
    - Deals
    - Deals-Testset
    - Golf
    - Golf-Testset
    - Iris
    - Labor-Negotiations
    - Market-Data
    - Polynomial
    - Products
    - Purchases
    - Ripley-Set
    - Sonar
    - Titanic
    - Titanic Training
    - Titanic Unlabeled
    - Transactions
    - Weighting

## 2. Latihan Penentuan Jenis Bunga Iris

- Lakukan Trining pada data Bungan iris dengan menggunakan algoritma decision tree

The screenshot shows the RapidMiner Studio interface with the following components:

- Repository:** Contains datasets like Deals-Testset, Golf, Golf-Testset, Iris, Labor-Negotiations, Market-Data, Polynomial, and Products.
- Process View:** Displays a process flow starting with "Retrieve Iris" (operator) connected to a "Decision Tree" operator. The "Decision Tree" operator has four outputs labeled "res", "mod", "exa", and "wei".
- Parameters View:** Shows configuration for the Decision Tree operator, including criterion (gain\_ratio), maximal depth (10), apply pruning (checked), confidence (0.1), apply prepruning (checked), minimal gain (0.01), and checkboxes for "Show advanced parameters" and "Change compatibility (9.10.011)".
- Help View:** Provides a synopsis for the Decision Tree operator, stating it generates a decision tree model for classification and regression.
- Result History View:** Shows an "ExampleSet (Retrieve Iris)" table with 150 examples. The columns are Row No., id, label, a1, a2, a3, and a4. The data shows 150 entries of Iris-setosa.
- Repository View:** Shows a detailed list of training resources and samples, including sub-folders like data, processes, templates, time series, tutorials, and community samples.

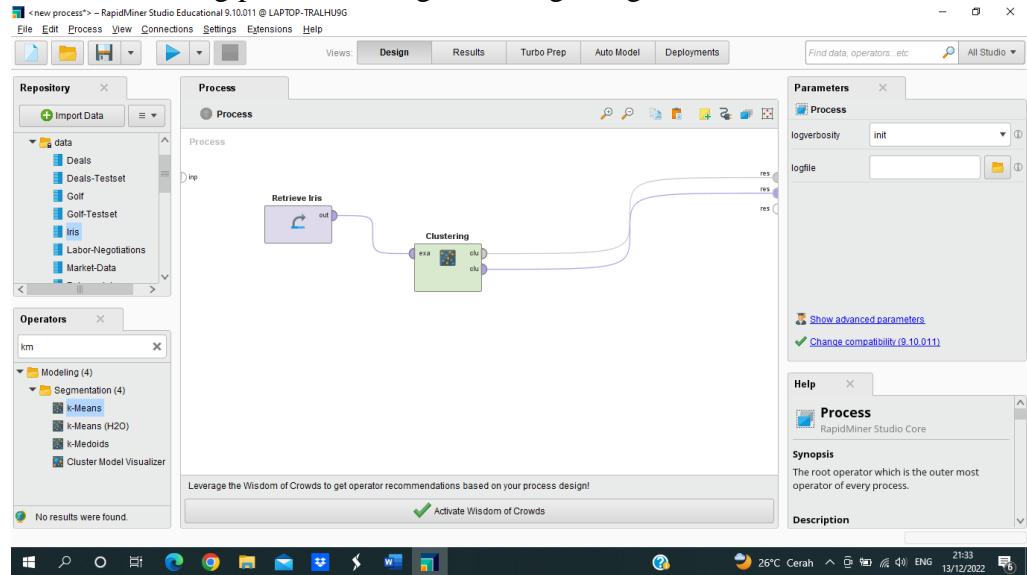
- Tampilkan himpunan data (dataset) dan pengetahuan (model tree) yang terbentuk

The screenshot shows the RapidMiner Studio interface with the following components:

- Result History:** Shows a "Tree (Decision Tree)" view.
- Repository:** Shows a detailed list of training resources and samples, including sub-folders like data, processes, templates, time series, tutorials, and community samples.
- Diagram View:** Displays the generated decision tree structure. The root node is "Iris-setosa". The tree splits based on attribute a3:
  - If  $a3 > 2.450$ , then "Iris-virginica".
  - If  $a3 \leq 2.450$ , then "Iris-versicolor".
 The "Iris-virginica" node further splits based on attribute a4:
  - If  $a4 > 1.750$ , then "Iris-versicolor".
  - If  $a4 \leq 1.750$ , then "Iris-virginica".
 The "Iris-virginica" node at this level further splits based on attribute a3:
  - If  $a3 > 5.350$ , then "Iris-versicolor".
  - If  $a3 \leq 5.350$ , then "Iris-virginica".
 The "Iris-virginica" node at this level further splits based on attribute a4:
  - If  $a4 > 4.950$ , then "Iris-versicolor".
  - If  $a4 \leq 4.950$ , then "Iris-virginica".

### 3. Latihan Klastering jenis bunga iris

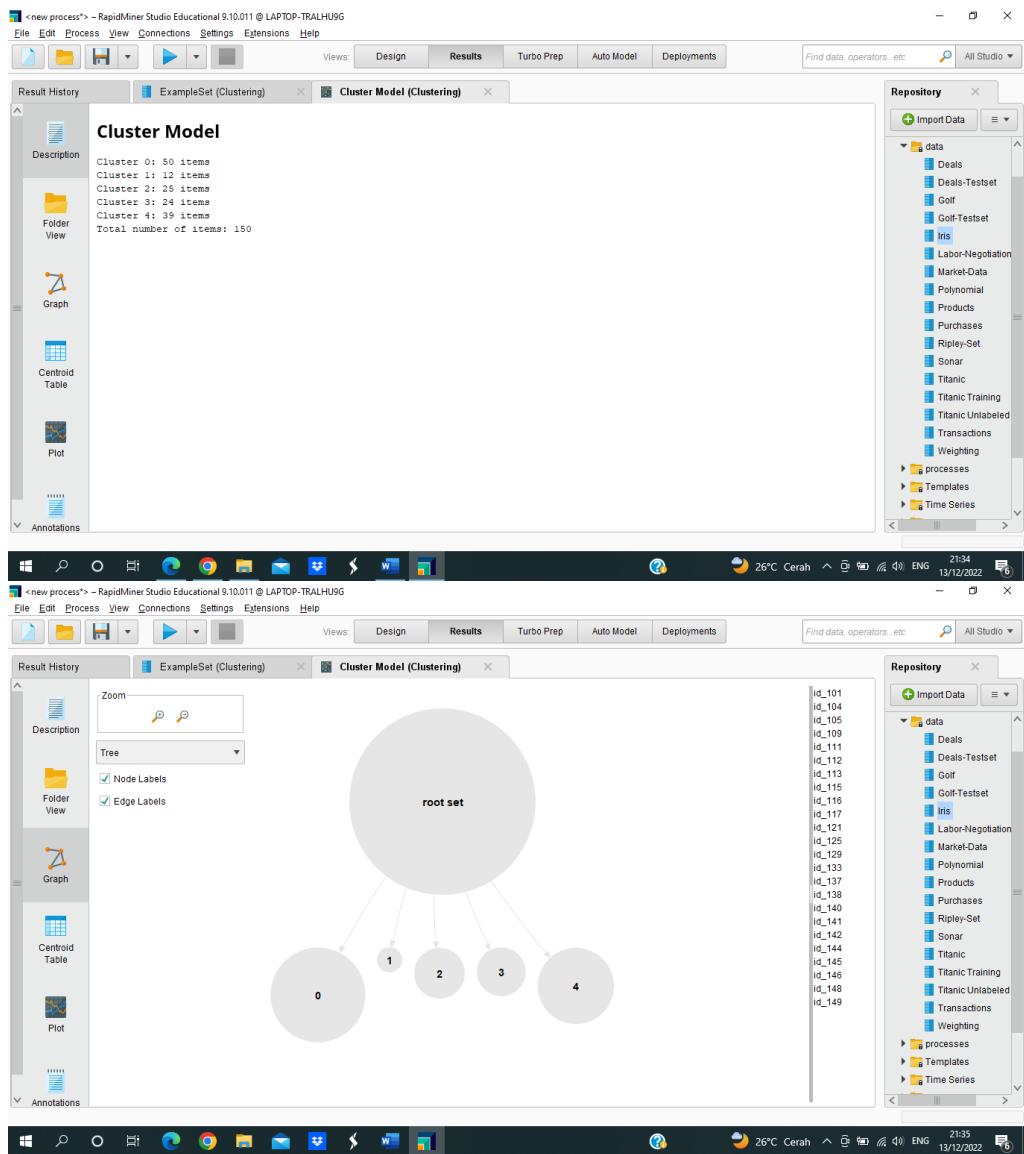
- Lakukan training pada data bunga iris dengan algoritma k – means



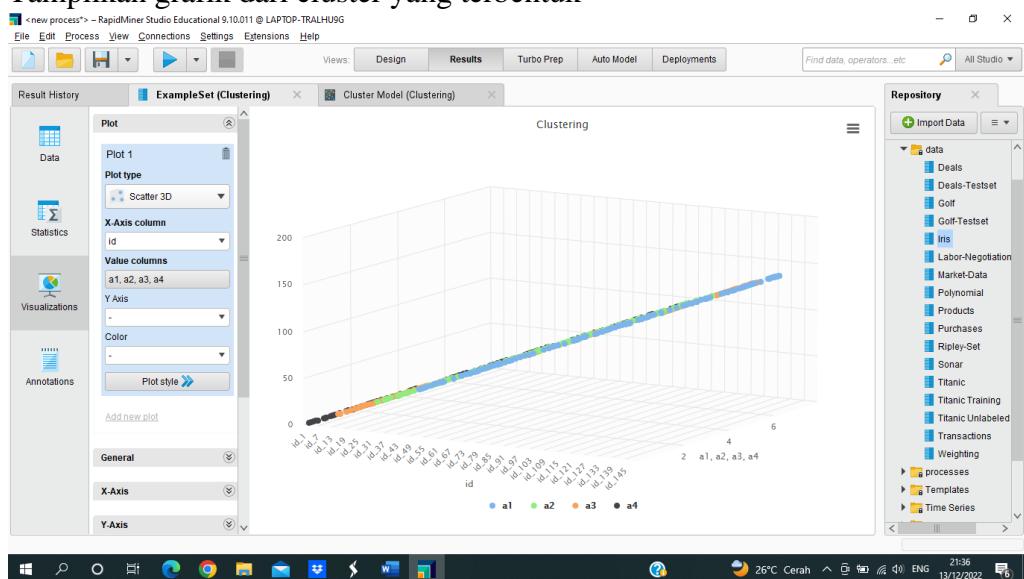
- Tampilkan impunan data dan model tree yang terbentuk

The screenshot shows the RapidMiner Studio interface in the 'Results' tab. It displays the 'ExampleSet (Clustering)' and 'Cluster Model (Clustering)' results. The 'ExampleSet' table has columns: Row No., id, label, cluster, a1, a2, a3, a4. The 'Cluster Model' table shows 15 rows of data. The 'Repository' sidebar lists various datasets like Deals, Golf, Iris, etc.

Row No.	id	label	cluster	a1	a2	a3	a4
1	id_1	Iris-setosa	cluster_0	5.100	3.500	1.400	0.200
2	id_2	Iris-setosa	cluster_0	4.900	3	1.400	0.200
3	id_3	Iris-setosa	cluster_0	4.700	3.200	1.300	0.200
4	id_4	Iris-setosa	cluster_0	4.600	3.100	1.500	0.200
5	id_5	Iris-setosa	cluster_0	5	3.600	1.400	0.200
6	id_6	Iris-setosa	cluster_0	5.400	3.900	1.700	0.400
7	id_7	Iris-setosa	cluster_0	4.600	3.400	1.400	0.300
8	id_8	Iris-setosa	cluster_0	5	3.400	1.500	0.200
9	id_9	Iris-setosa	cluster_0	4.400	2.900	1.400	0.200
10	id_10	Iris-setosa	cluster_0	4.900	3.100	1.500	0.100
11	id_11	Iris-setosa	cluster_0	5.400	3.700	1.500	0.200
12	id_12	Iris-setosa	cluster_0	4.800	3.400	1.600	0.200
13	id_13	Iris-setosa	cluster_0	4.800	3	1.400	0.100
14	id_14	Iris-setosa	cluster_0	4.300	3	1.100	0.100
15	id_15	Iris-setosa	cluster_0	5.800	4	1.200	0.200

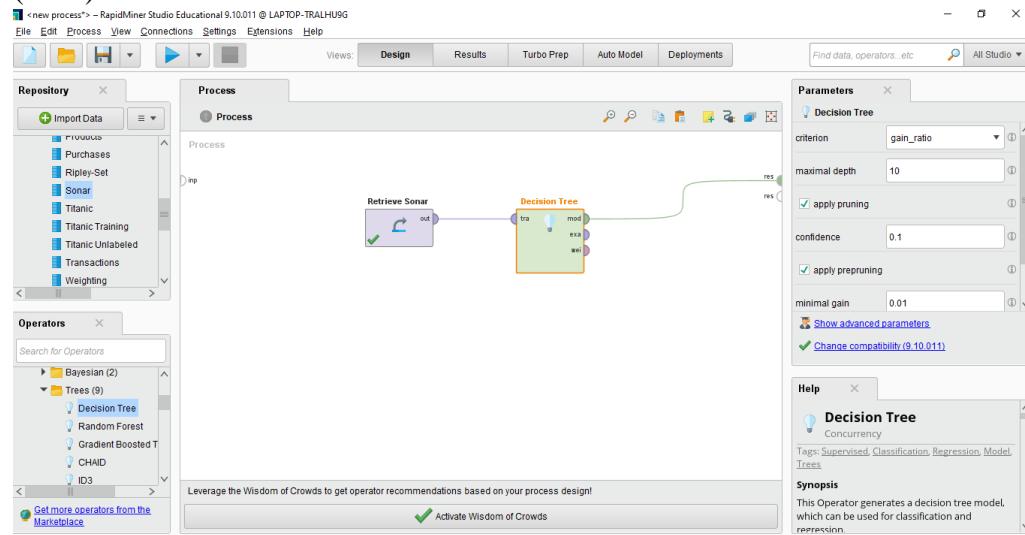


- Tampilkan grafik dari cluster yang terbentuk



#### 4. Latihan penentuan mine / rock

- Lakukan training pada data Sonar dengan menggunakan algoritma decision tree (C4.5)

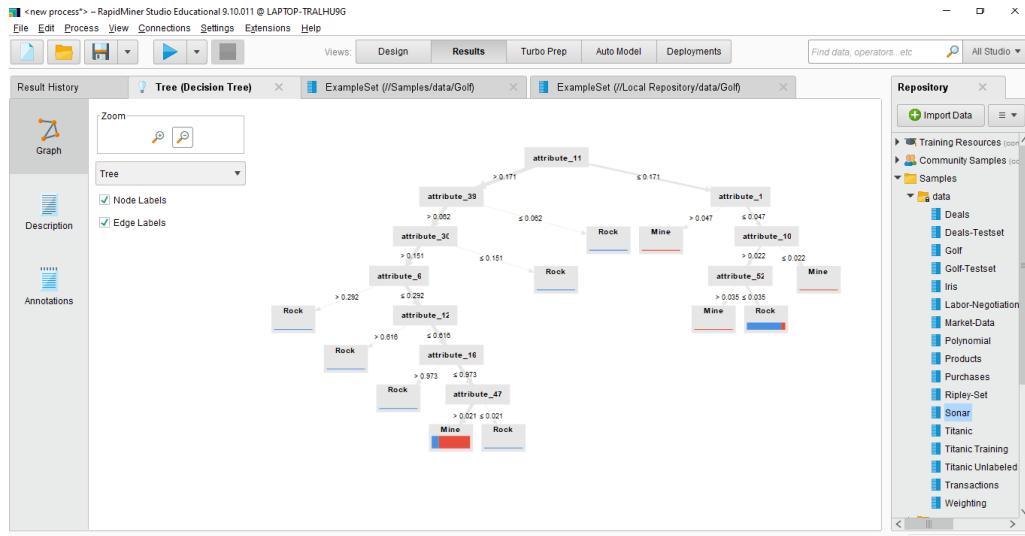


The screenshot shows the RapidMiner Studio interface with the following components:

- Result History:** Displays "AttributeWeights (Decision Tree)", "ExampleSet (Retrieve Sonar)", and "Tree (Decision Tree)".
- Data:** A table titled "AttributeWeights (Decision Tree)" showing attribute weights for the Sonar dataset. The columns are Row No., class, attribute\_1 through attribute\_8. The table has 20 rows.
- Annotations:** A table titled "ExampleSet (Retrieve Sonar)" showing the data instances. The columns are Row No., class, attribute\_1 through attribute\_8. The table has 208 rows.
- Repository:** Shows the "Sonar" dataset in the "Training Resources" section.

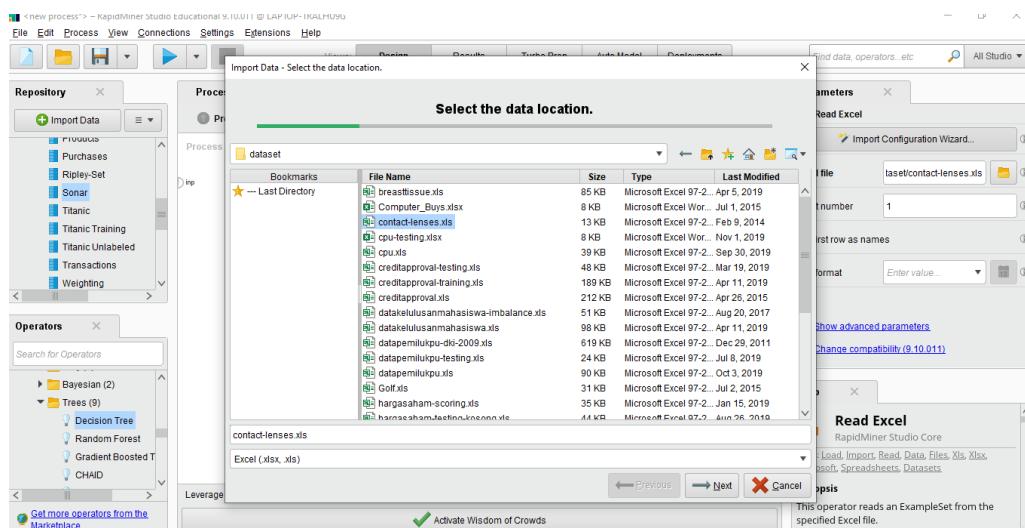
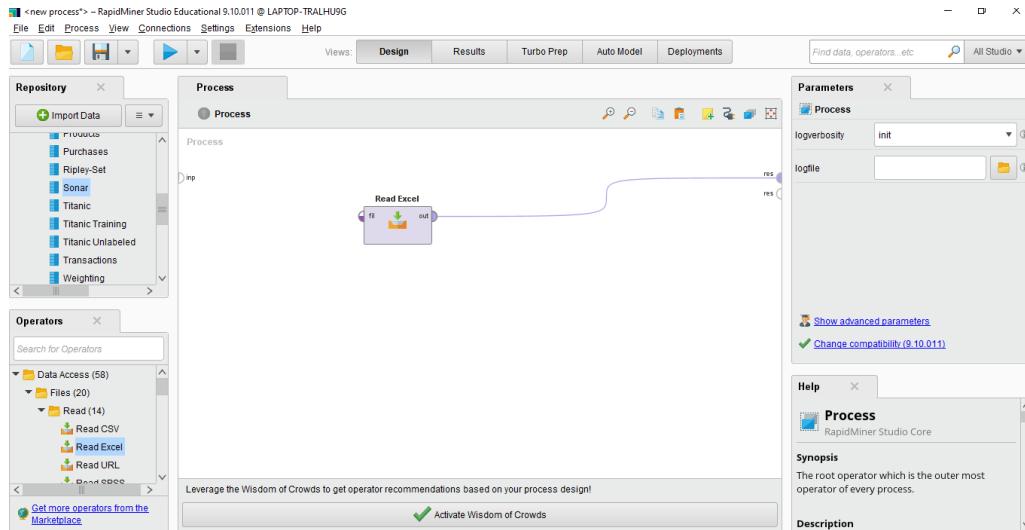
Row No.	class	attribute_1	attribute_2	attribute_3	attribute_4	attribute_5	attribute_6	attribute_7	attribute_8
1	Rock	0.020	0.037	0.043	0.021	0.095	0.099	0.154	0.160
2	Rock	0.045	0.052	0.084	0.069	0.118	0.258	0.216	0.348
3	Rock	0.026	0.058	0.110	0.108	0.097	0.228	0.243	0.377
4	Rock	0.010	0.017	0.062	0.021	0.021	0.037	0.110	0.128
5	Rock	0.076	0.067	0.048	0.039	0.059	0.065	0.121	0.247
6	Rock	0.029	0.045	0.028	0.017	0.038	0.099	0.120	0.183
7	Rock	0.032	0.096	0.132	0.141	0.167	0.171	0.073	0.140
8	Rock	0.052	0.055	0.084	0.032	0.115	0.092	0.103	0.061
9	Rock	0.022	0.037	0.048	0.048	0.065	0.059	0.075	0.010
10	Rock	0.016	0.017	0.035	0.007	0.019	0.067	0.105	0.070
11	Rock	0.004	0.006	0.015	0.034	0.031	0.028	0.040	0.027
12	Rock	0.012	0.031	0.017	0.031	0.036	0.010	0.018	0.058
13	Rock	0.008	0.009	0.005	0.025	0.034	0.055	0.053	0.098
14	Rock	0.009	0.006	0.025	0.049	0.120	0.159	0.139	0.099
15	Rock	0.012	0.043	0.060	0.045	0.060	0.035	0.053	0.034
16	Rock	0.030	0.061	0.065	0.092	0.162	0.229	0.218	0.203
17	Rock	0.035	0.012	0.019	0.047	0.074	0.118	0.168	0.154
18	Rock	0.019	0.061	0.038	0.077	0.139	0.081	0.057	0.022
19	Rock	0.027	0.009	0.015	0.028	0.041	0.076	0.103	0.114

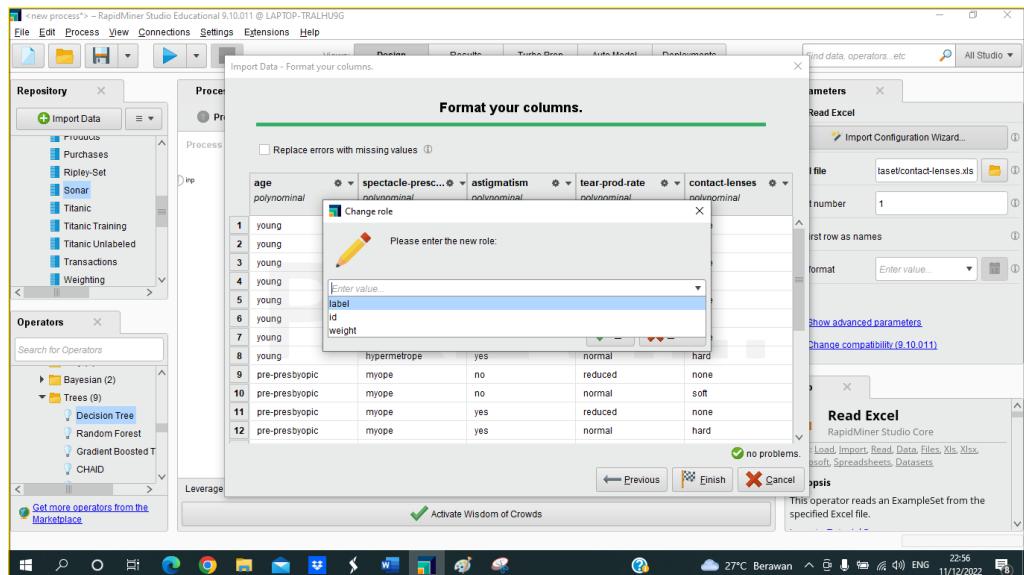
- Tampilkan Himpunan data model tree yang terbentuk



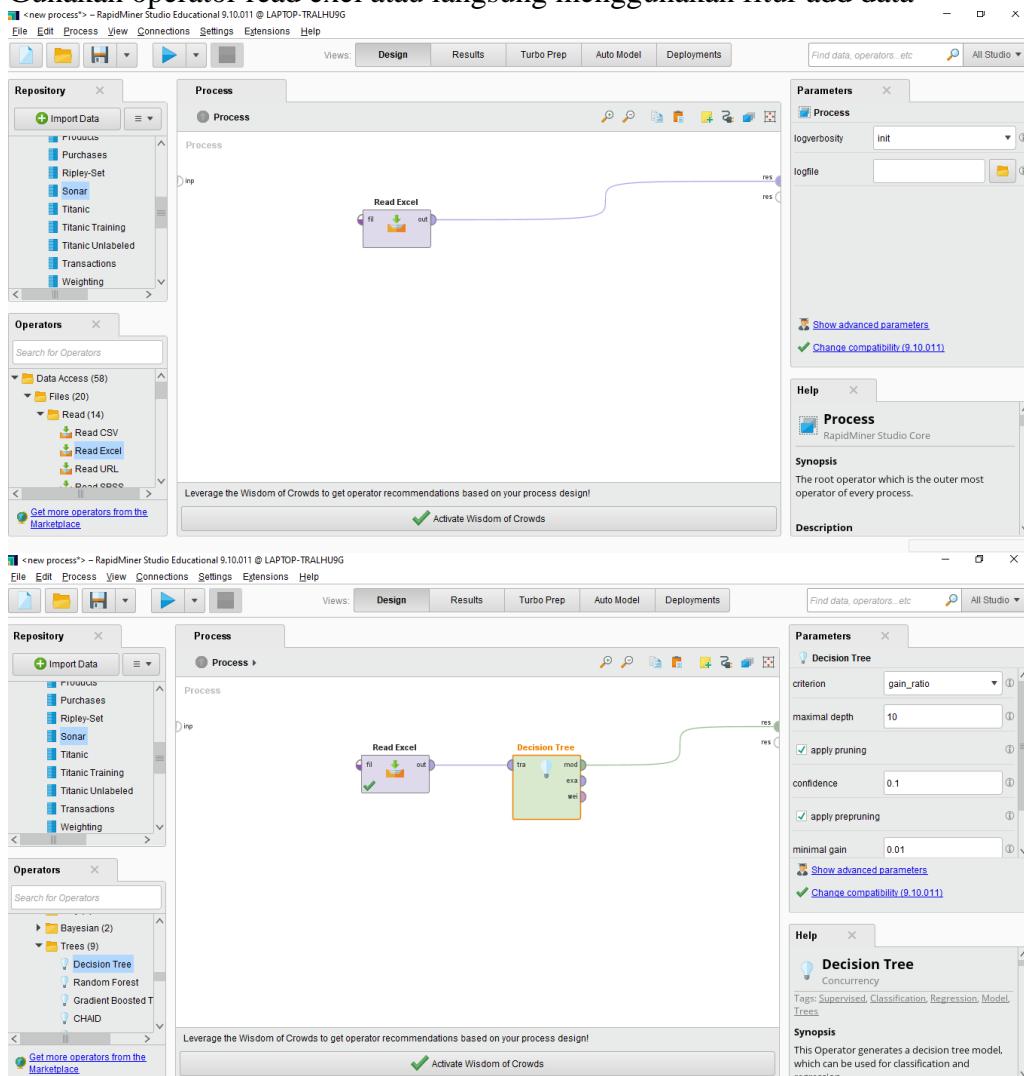
## 5. Lakukan Rekomendasi contact lenses

- Lakukan training pada data Contact lenses (contact-lenses.xls) dengan menggunakan algoritma decision tree

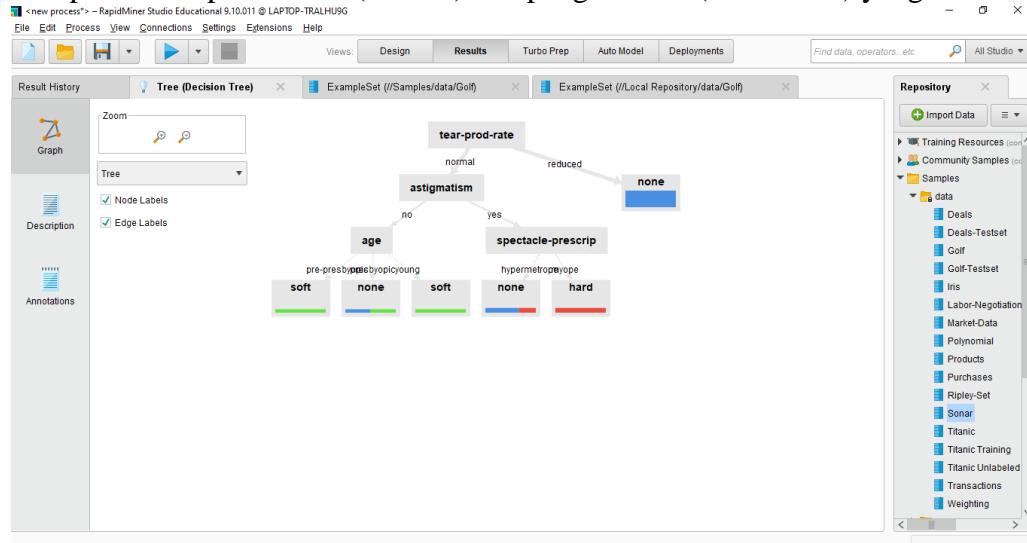




- Gunakan operator read excel atau langsung menggunakan fitur add data

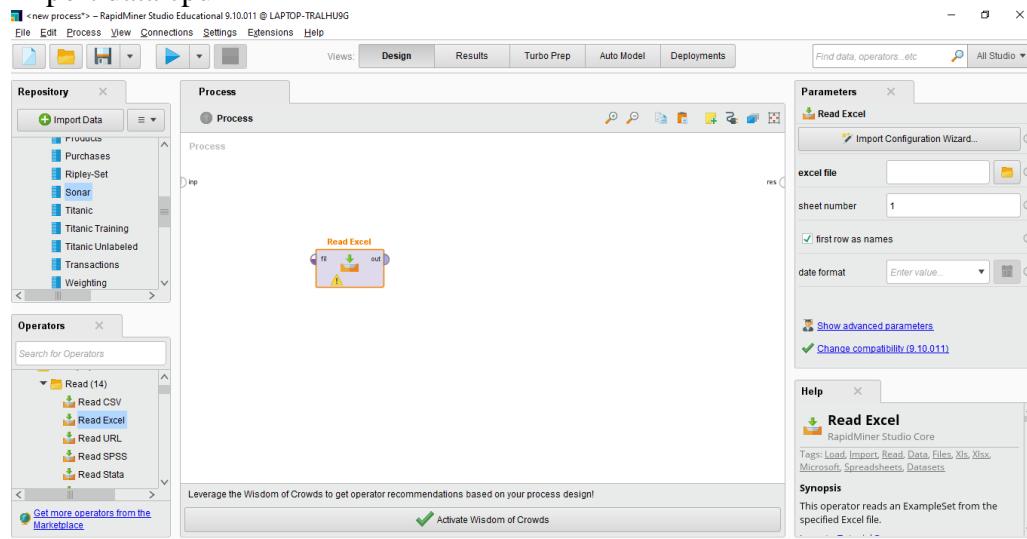


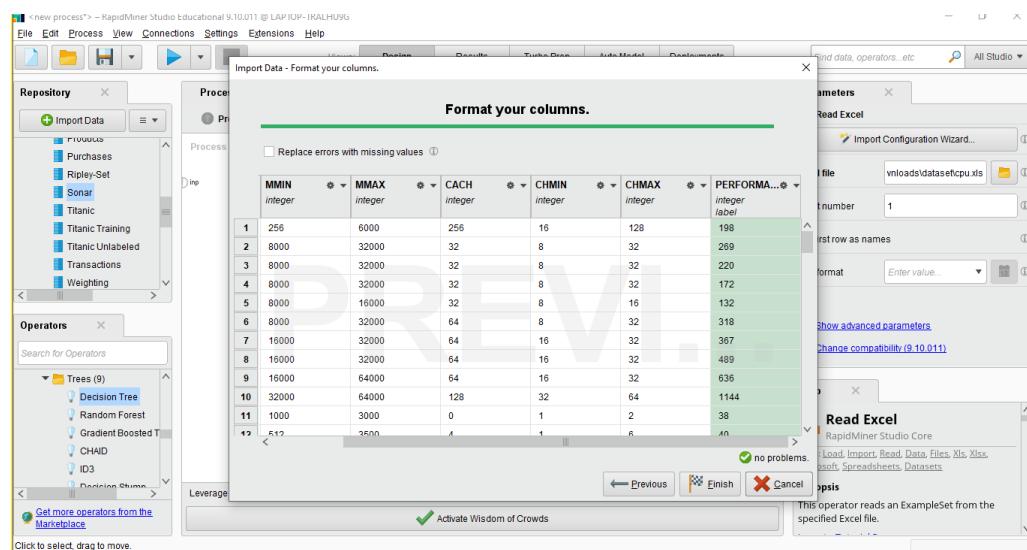
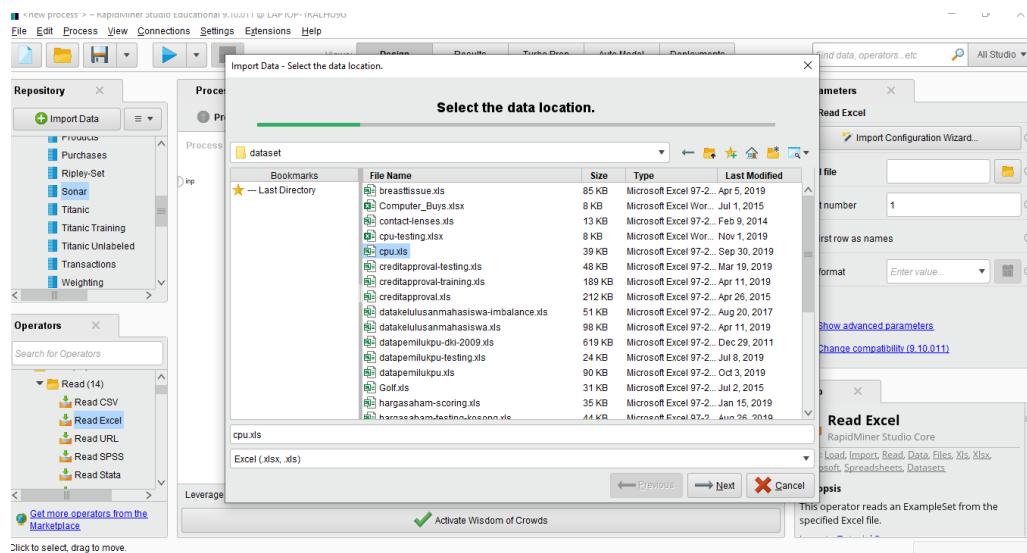
- Tampilkan himpunan data (dataset) dan pengetahuan (model tree) yang terbentuk



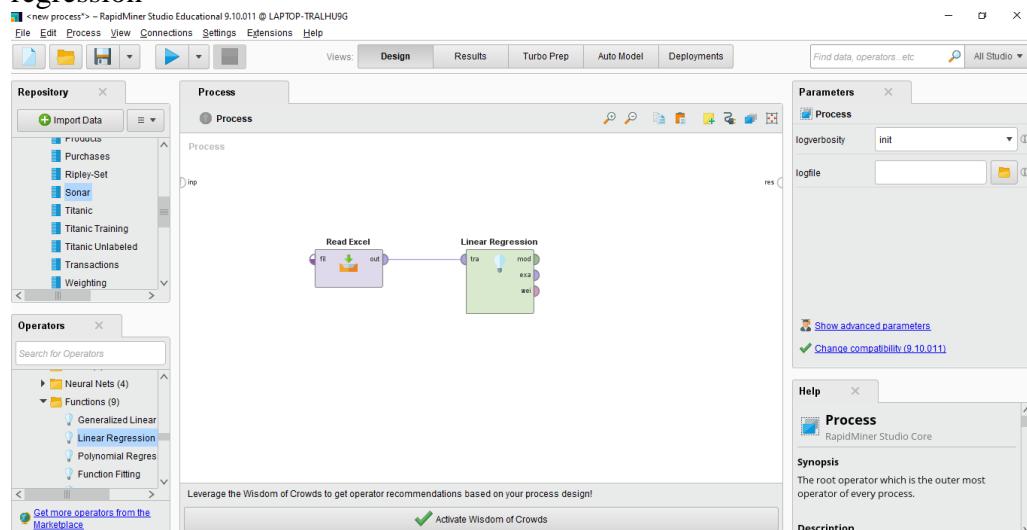
## 6. Latihan Estimasi Performance CPU

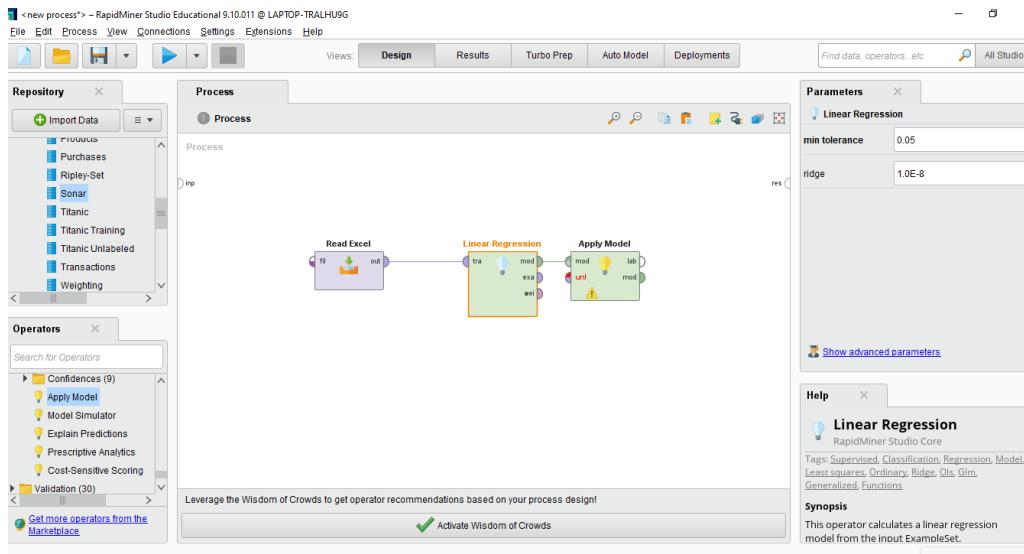
- Import data cpu xlx



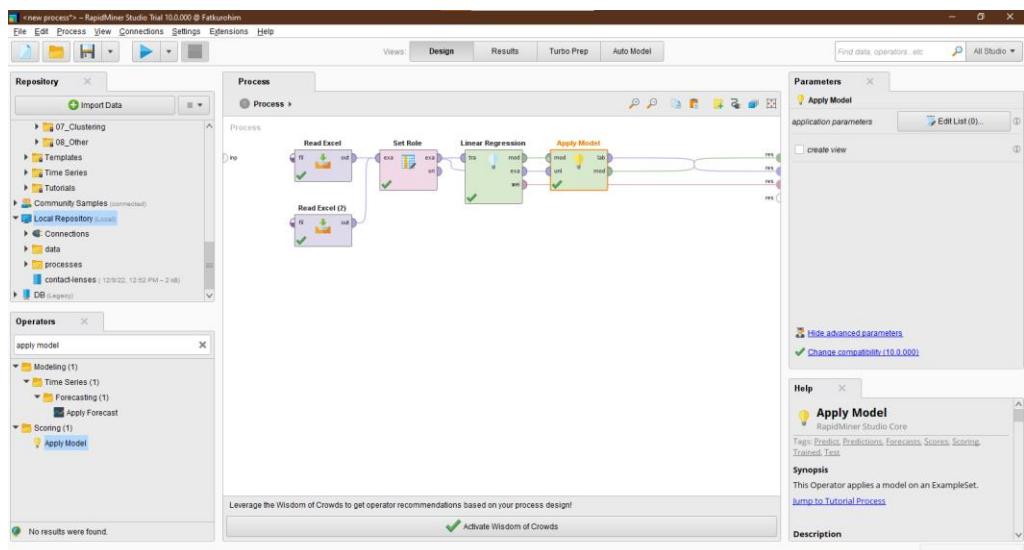


- Lakukan training pada data CPU (cpu.xls) dengan menggunakan algoritma linear regression





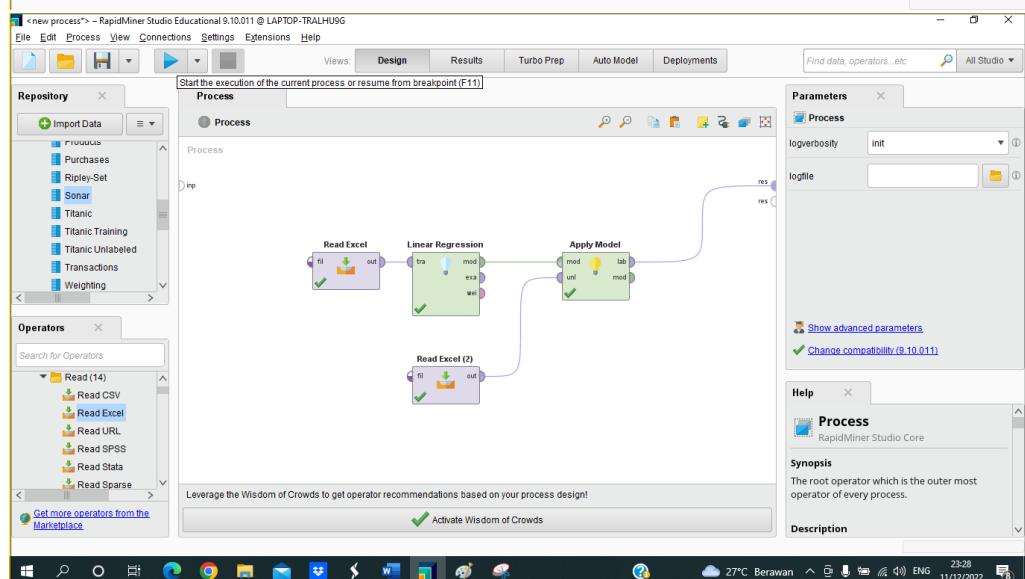
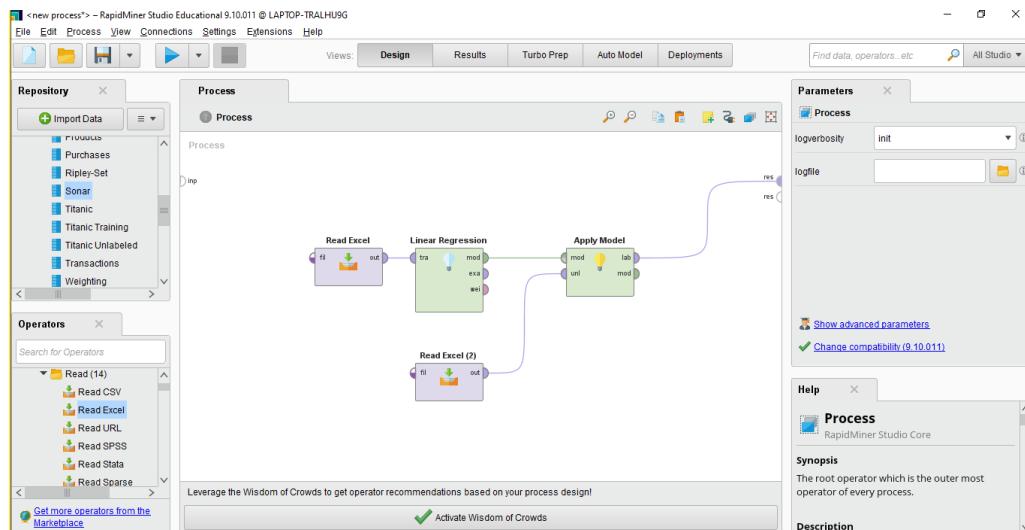
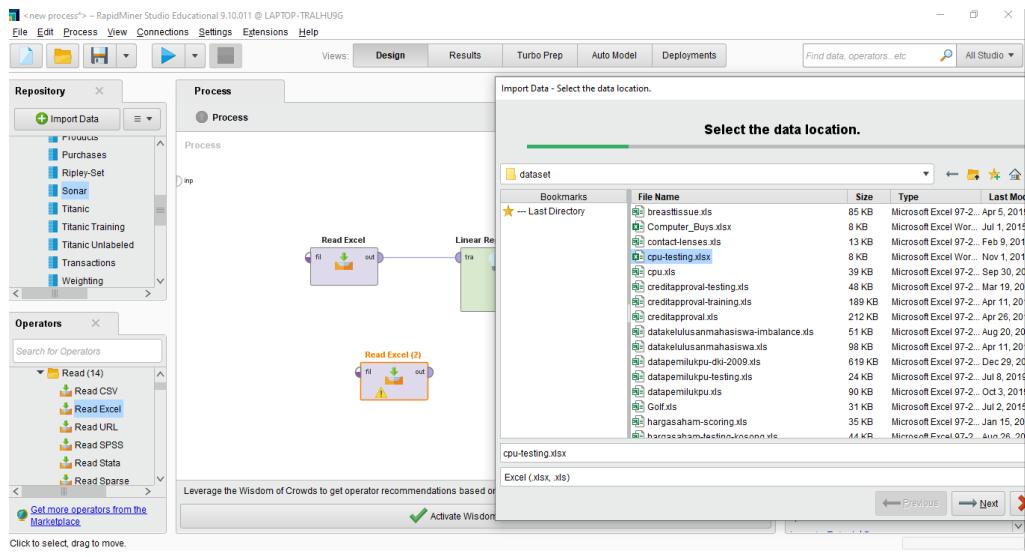
- Tampilkan himpunan data (dataset) dan pengetahuan (model regresi) yang terbentuk



This screenshot shows the 'Result History' tab in RapidMiner Studio. It displays two results: 'AttributeWeights (Linear Regression)' and 'LinearRegression (Linear Regression)'. The 'AttributeWeights' table shows the coefficient for each attribute. The 'LinearRegression' table shows the model's performance metrics. The Repository pane on the right lists various training resources and samples.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
MMIN	-0.018	0.009	-0.269	0.588	-1.984	0.049	**
MMAX	-0.009	0.003	-0.381	0.487	-3.007	0.003	***
CACH	-1.381	0.575	-0.212	0.748	-2.402	0.017	**
CHMAX	-1.563	0.929	-0.142	0.845	-1.682	0.094	*
PERFORMANCE	0.797	0.325	0.440	0.167	2.448	0.015	**
(Intercept)	342.880	27.213	?	?	12.594	0	****

- Lakukan pengujian terhadap data baru (cpu-testing.xls), untuk model yang dihasilkan dari tahapan 1



Result History    ExampleSet (Apply Model)    ExampleSet (/Samples/data/Golf)    ExampleSet (/Local Repository/data/Golf)

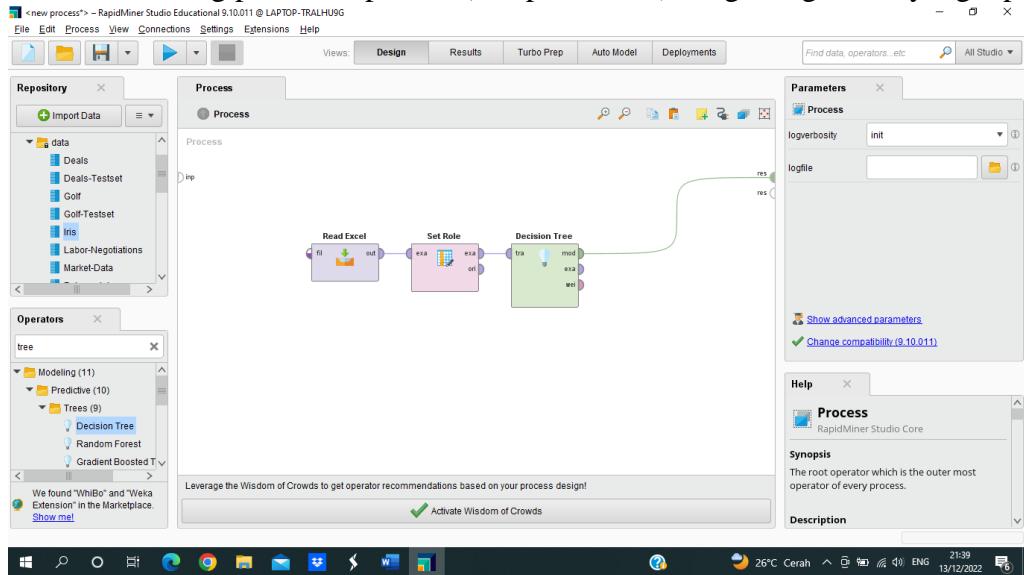
Data    Statistics    Visualizations    Annotations

Row No.	prediction(P...)	MYCT	MMIN	MMAX	CACH	CHMIN	CHMAX
1	9.025	480	1000	4000	0	0	0
2	628.047	30	8000	64000	128	12	176
3	-10.854	180	252	4000	0	1	3
4	-6.669	180	512	4000	0	1	3
5	-10.854	180	262	4000	0	1	3
6	-6.669	180	512	4000	0	1	3
7	21.681	124	1000	8000	0	1	8
8	41.254	98	1000	8000	32	2	8
9	45.186	125	2000	8000	0	2	14
10	37.941	480	512	8000	32	0	0

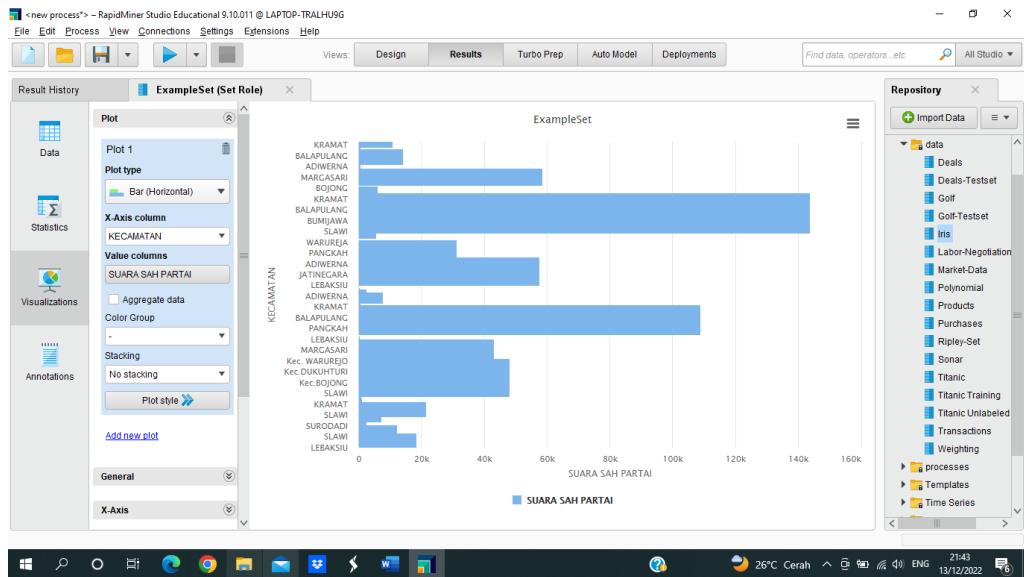
ExampleSet (10 examples, 1 special attribute, 6 regular attributes)

## 7. Latihan prediksi elektabilitas caleg

- Lakukan training pada data pemilu (datapemilu.xls) dengan algoritma yang tepat

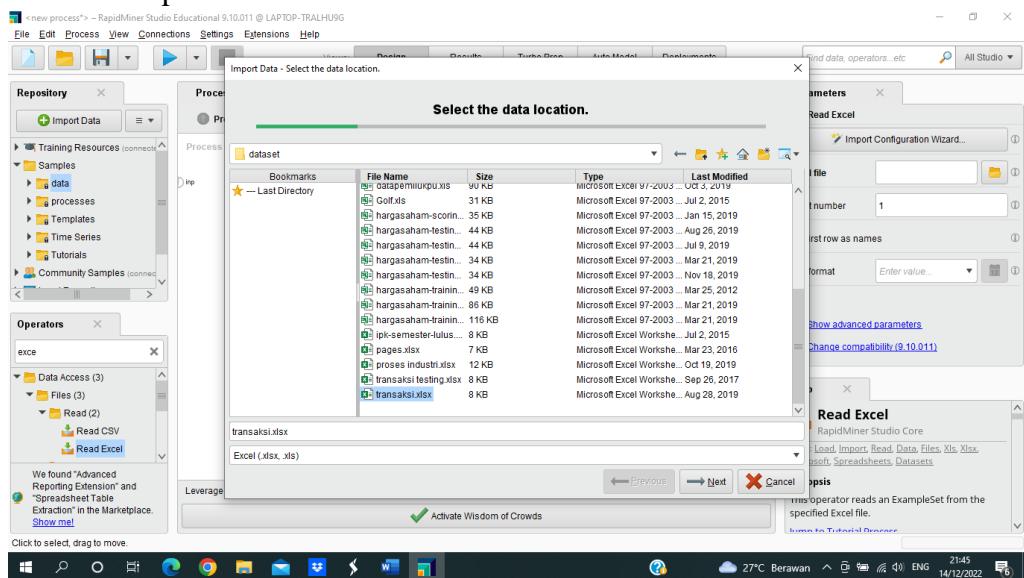


- Tentukan tipe data atribut dan class dari “import configuration wizard”
- Tampilkan himpunan data (dataset) dan pengetahuan (pola/model) yang terbentuk



8. Latihan Aturan Asosiasi Data Transaksi
  - Lakukan import data transaksi

- Lakukan import data transaksi



- Kolom customer ubah type menjadi polynominal dan gula dan seterusnya ubah type menjadi binomial

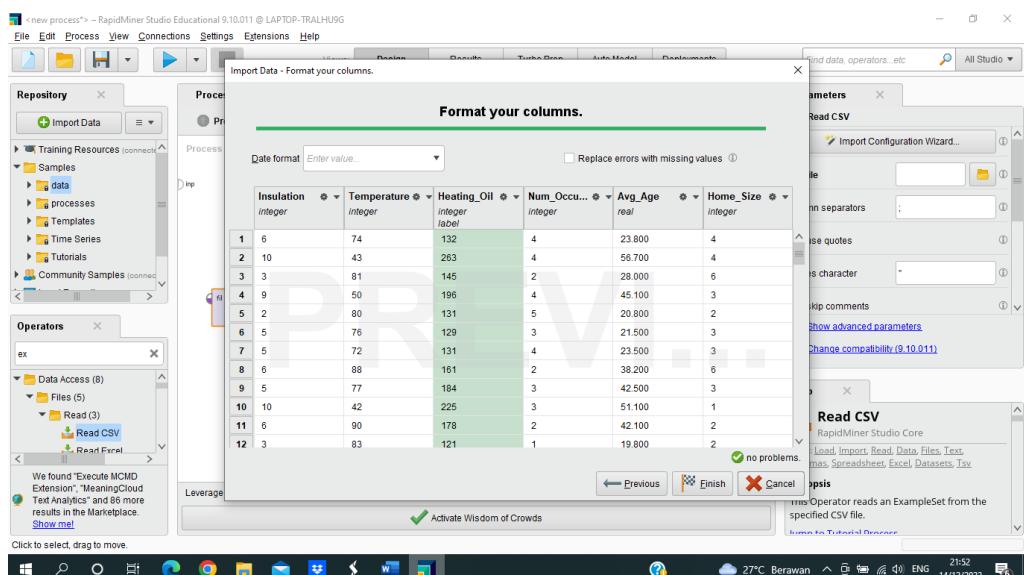
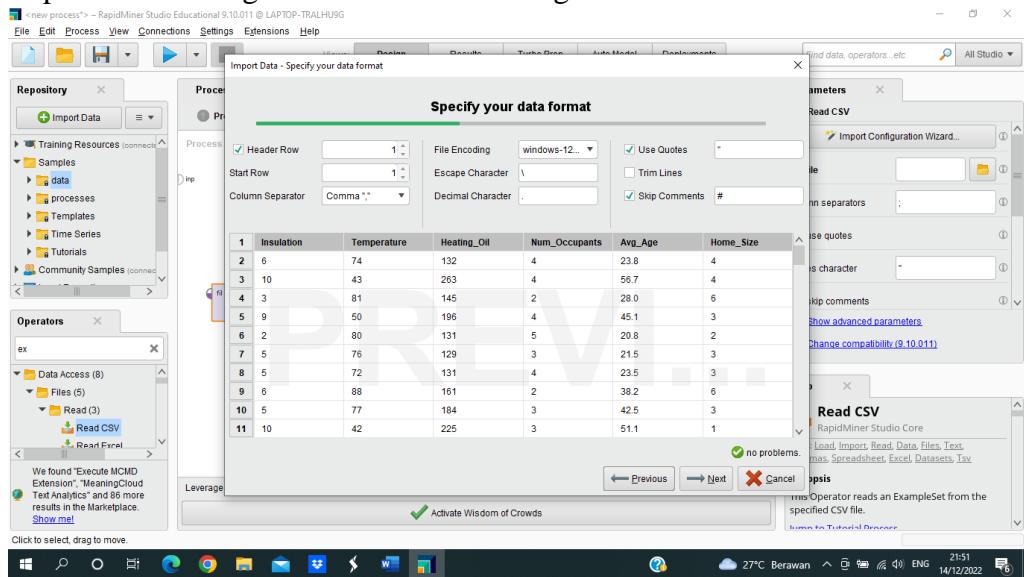
- Tambahkan operation FP-Growth dan create association rules

- Dan resultnya sebagai berikut

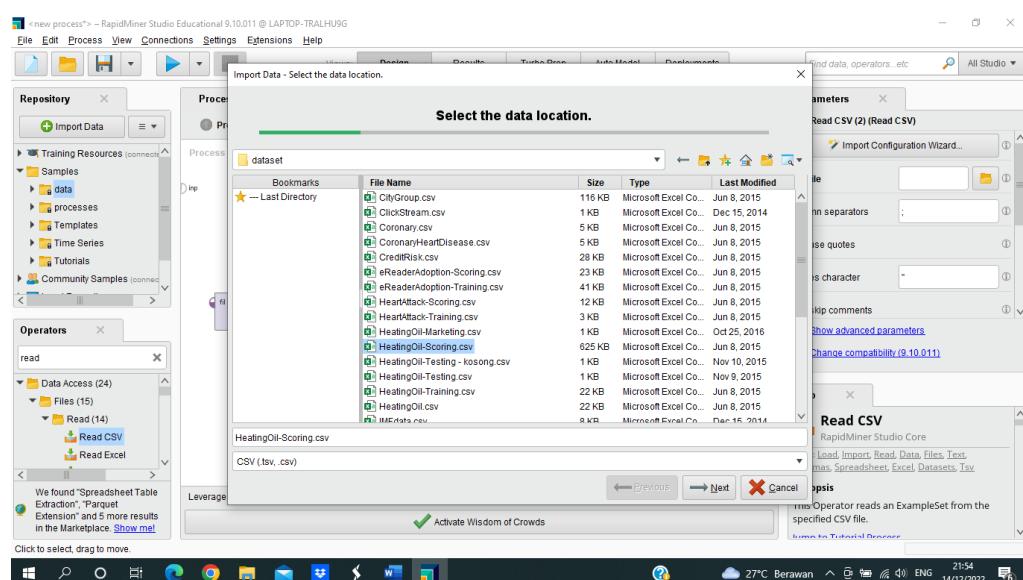
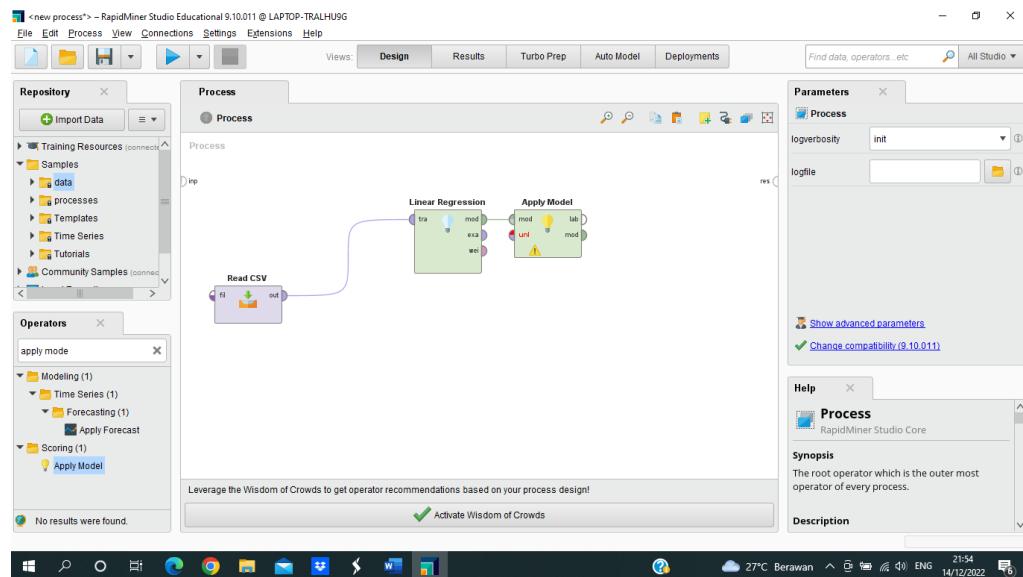
No.	Premises	Conclusion	Support	Confidence	LaPace	Gain
7	Sampo	Sabun	0.500	0.857	0.947	-0.667
8	Kopi	Gula	0.500	0.857	0.947	-0.667
9	Boneka	Sabun	0.250	1	1	-0.250
10	Celana	Sabun	0.250	1	1	-0.250
11	Gula	Kopi	0.500	1	1	-0.500
12	Boneka	Sampo	0.250	1	1	-0.250
13	Celana	Sampo	0.250	1	1	-0.250
14	Boneka	Sprei	0.250	1	1	-0.250
15	Kopi, Sampo	Sabun	0.250	1	1	-0.250
16	Sabun, Gula	Kopi	0.333	1	1	-0.333
17	Sabun, Sprei	Sampo	0.250	1	1	-0.250
18	Sampo, Sprei	Sabun	0.250	1	1	-0.250
19	Boneka	Sabun, Sampo	0.250	1	1	-0.250
20	Sabun, Boneka	Sampo	0.250	1	1	-0.250
21	Sampo, Boneka	Sabun	0.250	1	1	-0.250
22	Celana	Sabun, Sampo	0.250	1	1	-0.250

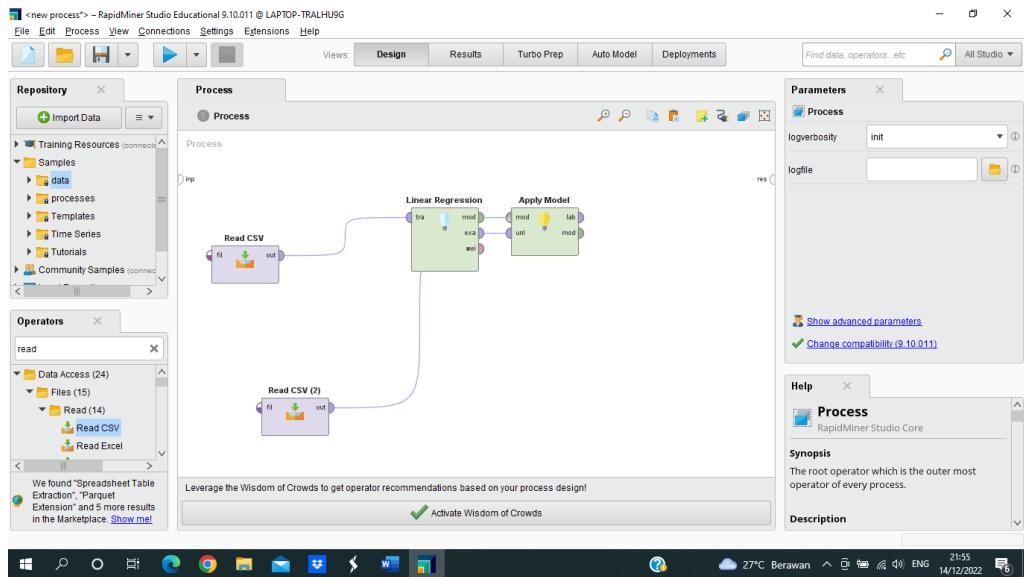
## 9. Latihan Estimasi Konsumsi Minyak

- Import data heatingoil.csv dan ubah heating oil ke label



- Tambahkan operation linear regression dan apply model serta tambahkan read CSV lalu import data heating oil scoring.csv





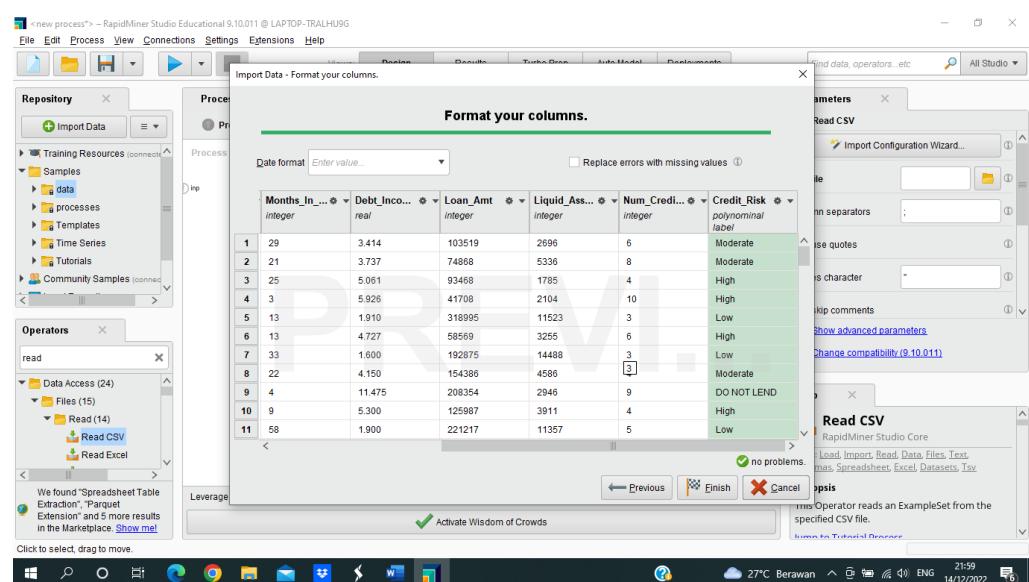
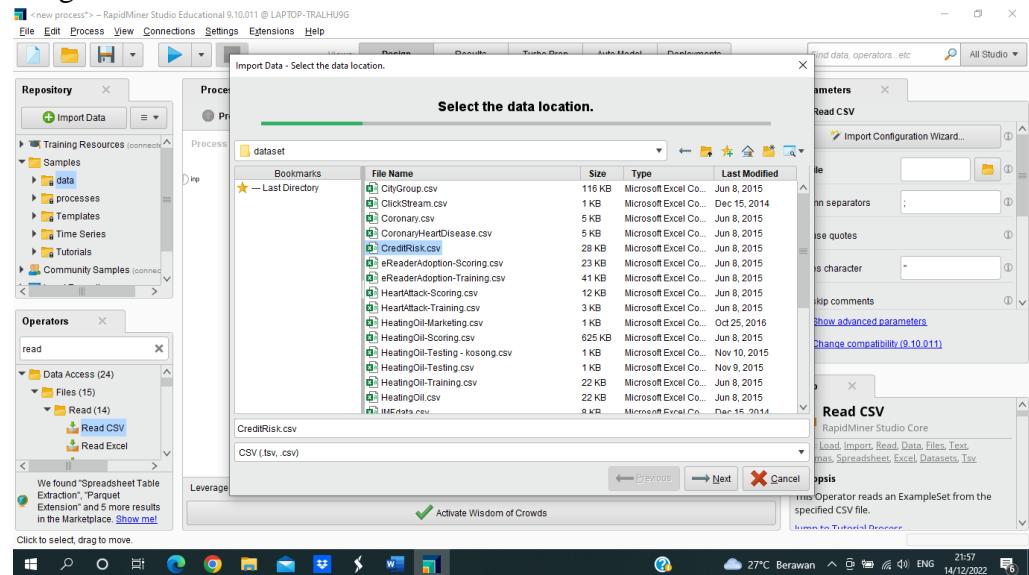
- Tampilkan hasil runningnya seperti dibawah

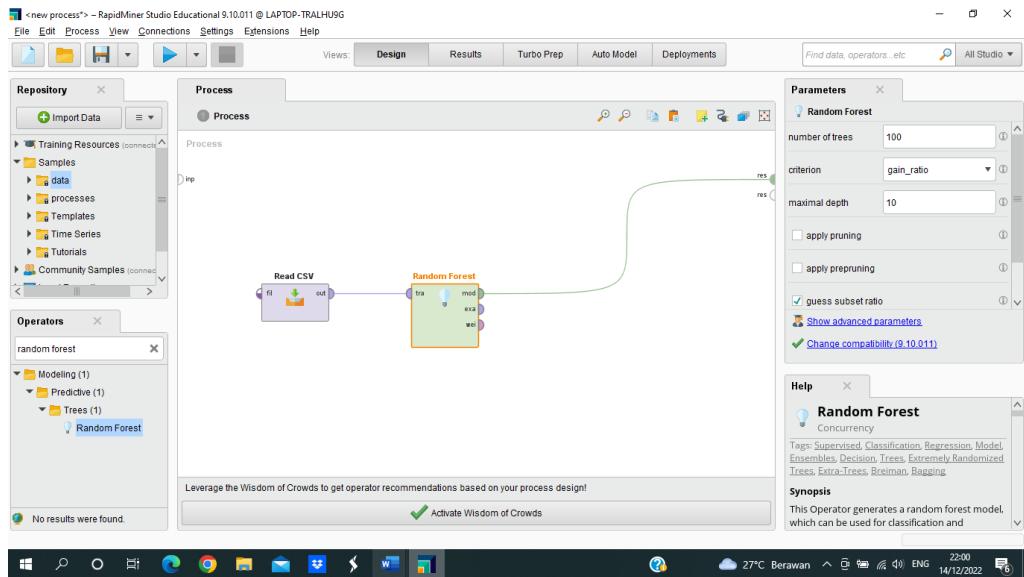
Row No.	prediction(HL...)	Insulation	Temperature	Num_Occup...	Avg_Age	Home_Size
1	251.321	5	69	10	70.100	7
2	216.028	5	80	1	66.700	1
3	226.087	4	89	9	67.800	7
4	209.529	7	81	9	52.400	6
5	164.669	4	58	8	22.900	7
6	180.512	4	58	6	37.400	3
7	221.188	6	51	2	51.600	3
8	164.001	2	73	5	37.400	4
9	264.712	9	39	1	56.900	7
10	221.364	8	84	5	64.500	2
11	221.328	10	74	6	58.300	1
12	262.580	5	49	6	68.600	6
13	214.082	8	45	2	33.900	8
14	212.392	3	49	4	49.700	4
15	253.199	9	66	6	66.200	5

ExampleSet (42,650 examples, 1 special attribute, 5 regular attributes)

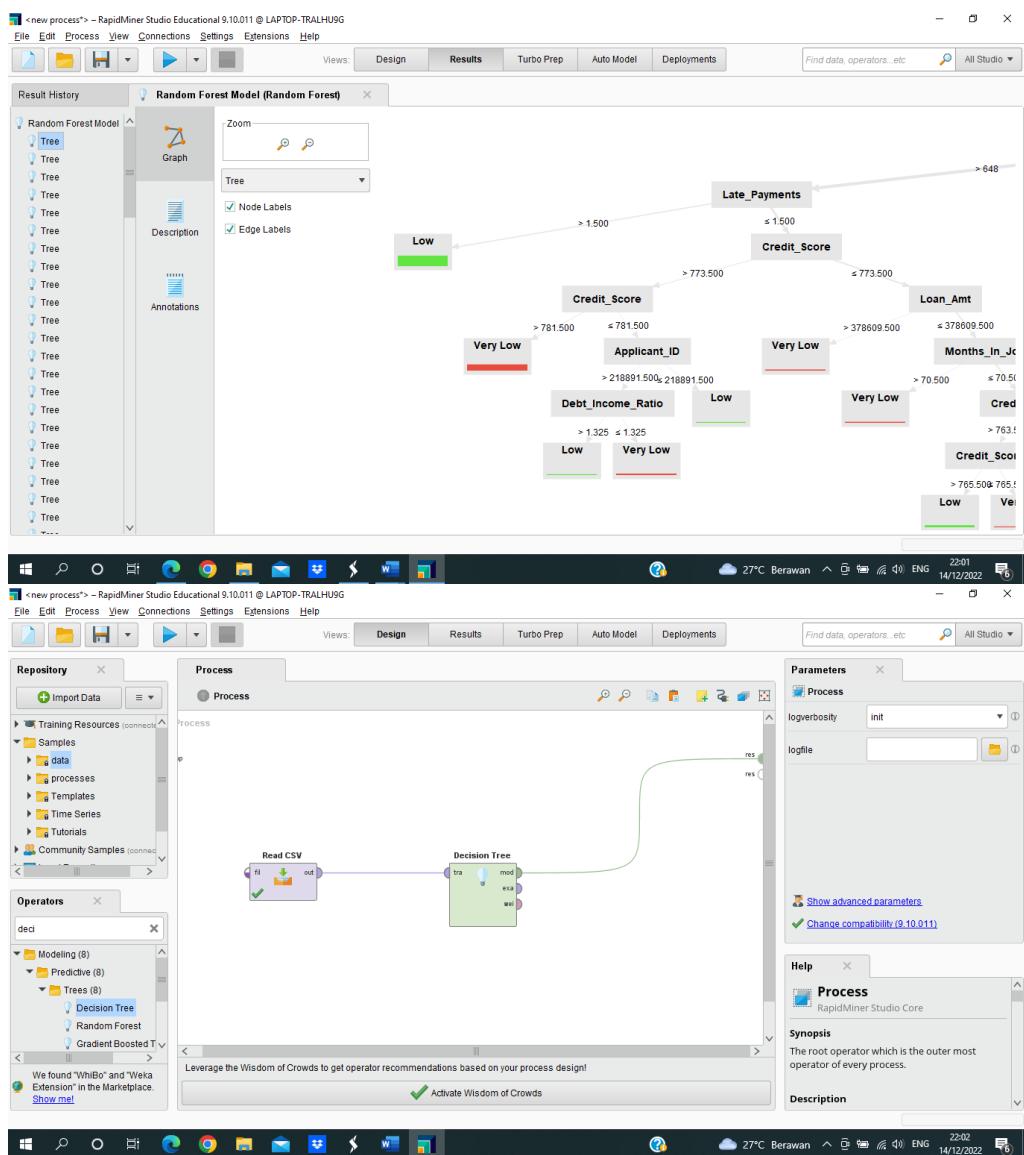
## 10. Latihan Klasifikasi Resiko Kredit

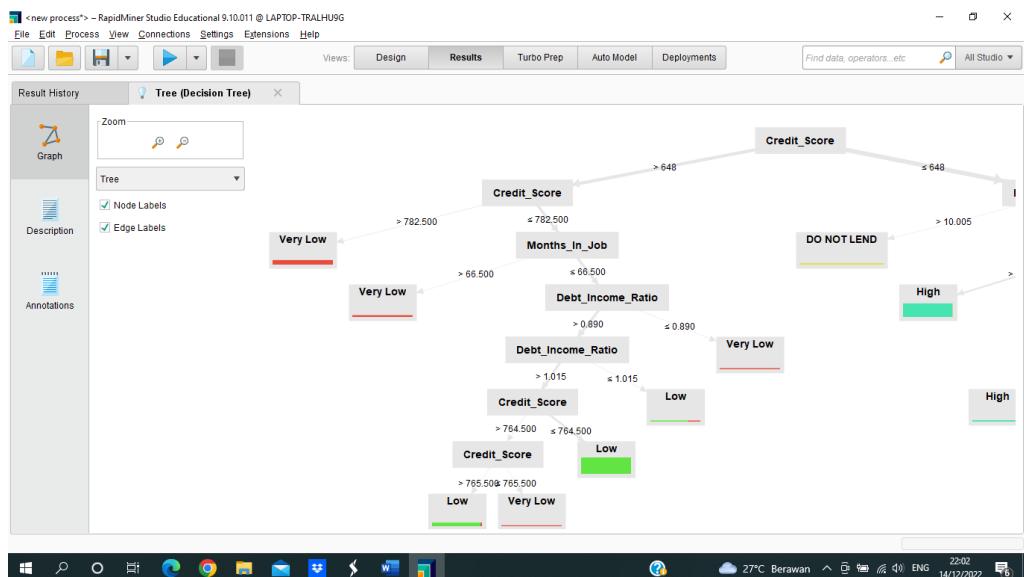
- Lakukan import data creditrisk.csv dan ubah menjadi label serta sambungkan dengan random forest





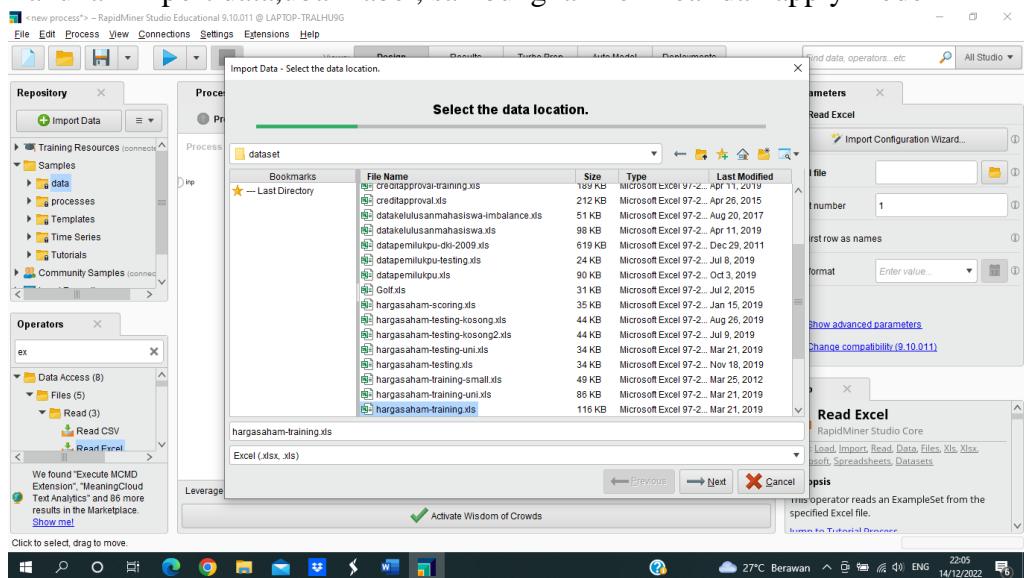
- Tampilan pada random forest dan decision tree





## 11. Latihan Forecasting Harga Saham

- Lakukan import data, ubah label, sambungkan ke linear dan apply model



**Import Data - Format your columns.**

**Format your columns.**

Replace errors with missing values

Date	Open	High	Low	Close	Volume
date	real	real	real	real	real/label
1 Apr 11, 2006	1296.600	1300.710	1282.960	1286.570	2232880000.000
2 Apr 12, 2006	1288.570	1290.930	1286.450	1288.120	1938100000.000
3 Apr 13, 2006	1288.120	1292.090	1283.370	1289.120	1891940000.000
4 Apr 17, 2006	1289.120	1292.450	1280.740	1285.330	1794650000.000
5 Apr 18, 2006	1288.330	1309.020	1285.330	1307.280	2595440000.000
6 Apr 19, 2006	1307.650	1310.390	1302.790	1309.930	2447310000.000
7 Apr 20, 2006	1309.930	1318.150	1308.380	1311.460	2512920000.000
8 Apr 21, 2006	1311.460	1317.670	1306.590	1311.280	2392330000.000
9 Apr 24, 2006	1311.280	1311.280	1303.790	1308.110	2117330000.000
10 Apr 25, 2006	1308.110	1310.790	1299.170	1301.740	2366380000.000
11 Apr 26, 2006	1301.740	1310.970	1301.740	1305.410	2502690000.000
12 Apr 27, 2006	1305.410	1315.000	1295.570	1309.720	2772010000.000

no problems.

← Previous     

Activate Wisdom of Crowds

**Read Excel**  
Import Configuration Wizard...  
File: \_\_\_\_\_  
Sheet number: 1  
First row as names  
Format: Enter value...  
Show advanced parameters  
Change compatibility (9.10.011)

**Read Excel**  
RapidMiner Studio Core  
Load, Import, Read, Data, Files, Xls, Xlsx, Csv, Spreadsheets, Datasets  
Synopsis  
This operator reads an ExampleSet from the specified Excel file.  
[Jump to Tutorial Process](#)

**File Edit Process View Connections Settings Extensions Help**

**Process**

**Read Excel** → **Linear Regression** → **Apply Model**

**Apply Model** (mod, lab, unl, mod)

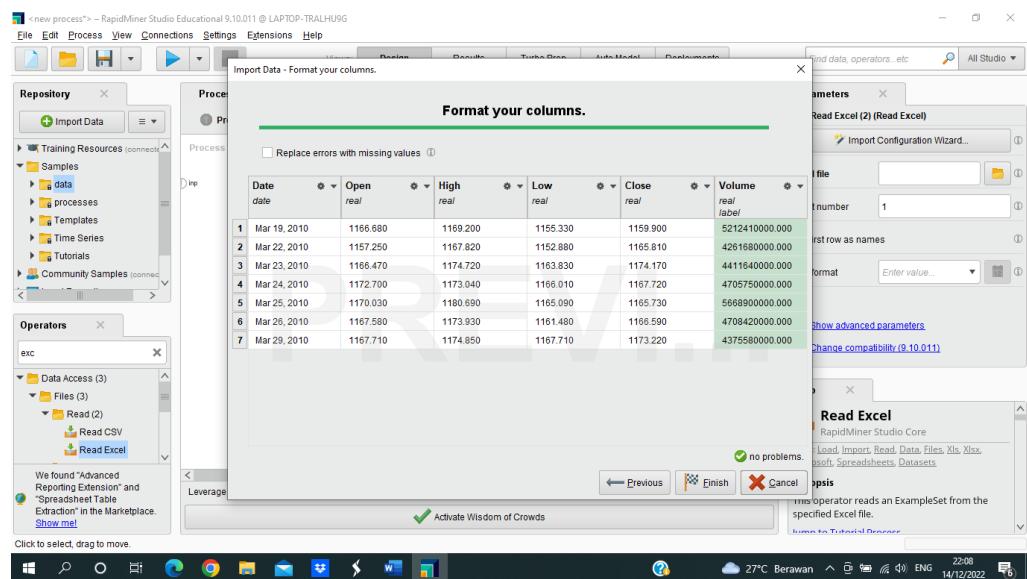
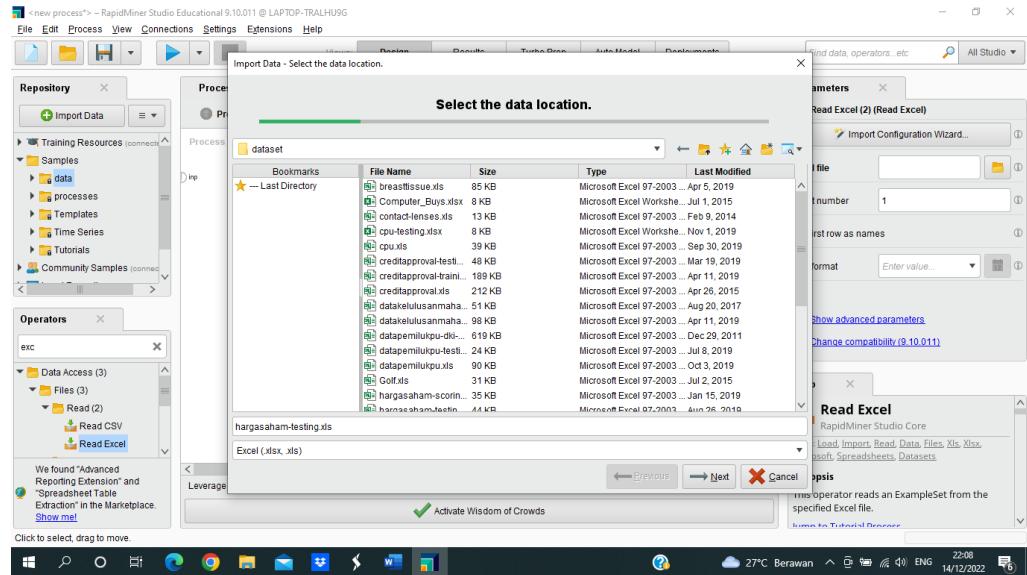
**Parameters**

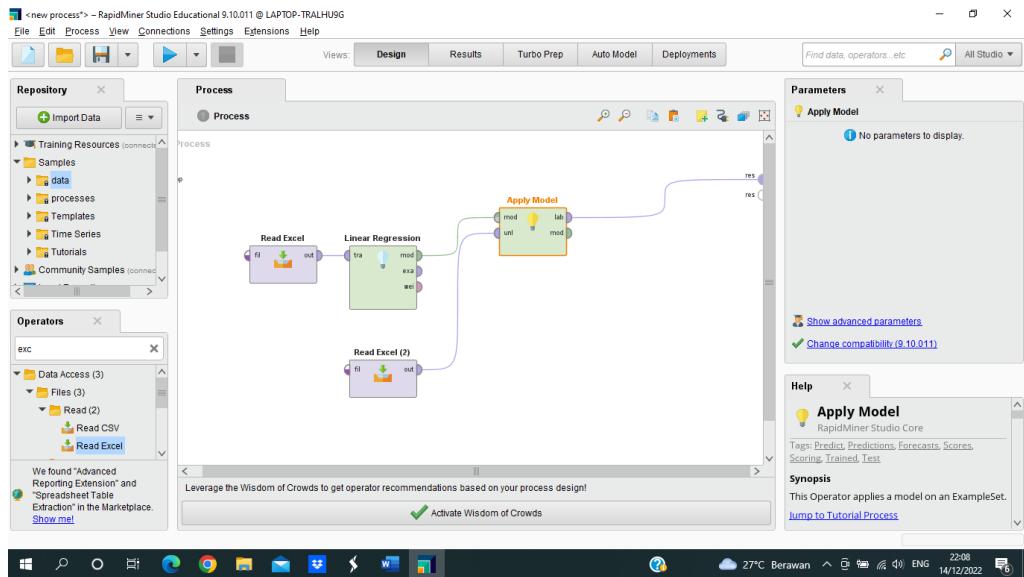
No parameters to display.

**Help**

**Apply Model**  
RapidMiner Studio Core  
Tags: Predict, Predictions, Forecasts, Scores, Scoring, Trained, Test  
Synopsis  
This Operator applies a model on an ExampleSet.  
[Jump to Tutorial Process](#)

- Tambahkan read excel harga saham testing lalu sambungkan dengan apply model dan run hasilnya

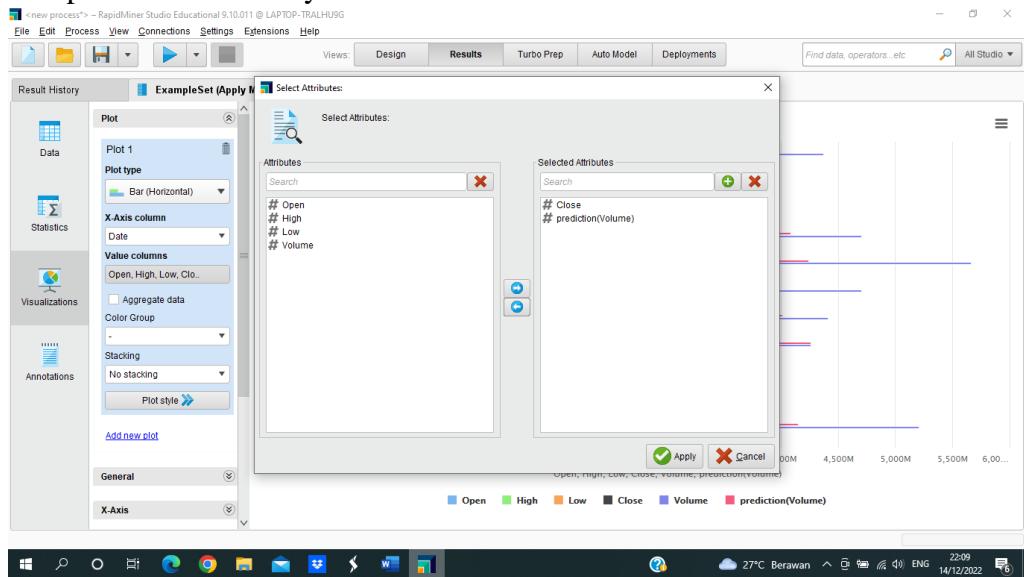


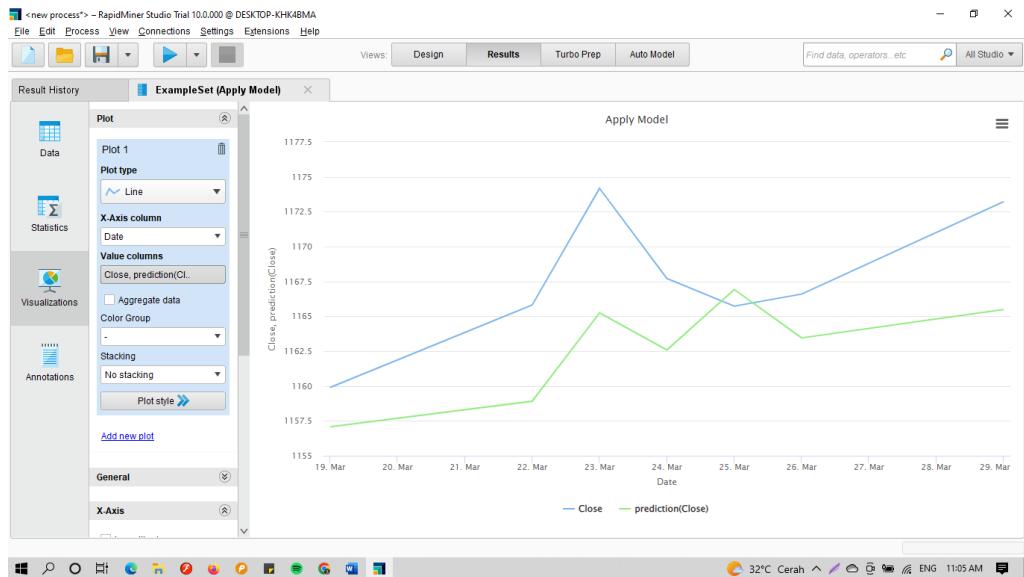


The screenshot shows the "Results" view for the "ExampleSet (Apply Model)". The table displays the following data:

Row No.	Volume	prediction(Volume)	Date	Open	High	Low	Close
1	5212410000	4152729826...	Mar 19, 2010	1166.680	1169.200	1155.330	1159.900
2	4261680000	4261464100...	Mar 22, 2010	1157.250	1167.820	1152.880	1165.810
3	4411640000	4013885278...	Mar 23, 2010	1168.470	1174.720	1163.830	1174.170
4	4705750000	3777122006...	Mar 24, 2010	1172.700	1173.040	1166.010	1167.720
5	5668900000	4242790454...	Mar 25, 2010	1170.030	1180.690	1165.090	1165.730
6	4708420000	4084649976...	Mar 26, 2010	1167.580	1173.930	1161.480	1166.590
7	4375580000	3816560536...	Mar 29, 2010	1167.710	1174.850	1167.710	1173.220

## - Tampilkan visualisasinya





## Kesimpulan

Pada saat melakukan latihan mengolah data mining menggunakan RapidMiner sangatlah efisien karena dapat menampilkan data numerik maupun visualisasi, RapidMiner sendiri adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platfrom dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran.