

LAPORAN UAS
PRAKTIKUM DATA WAREHOUSING DAN DATA MINING



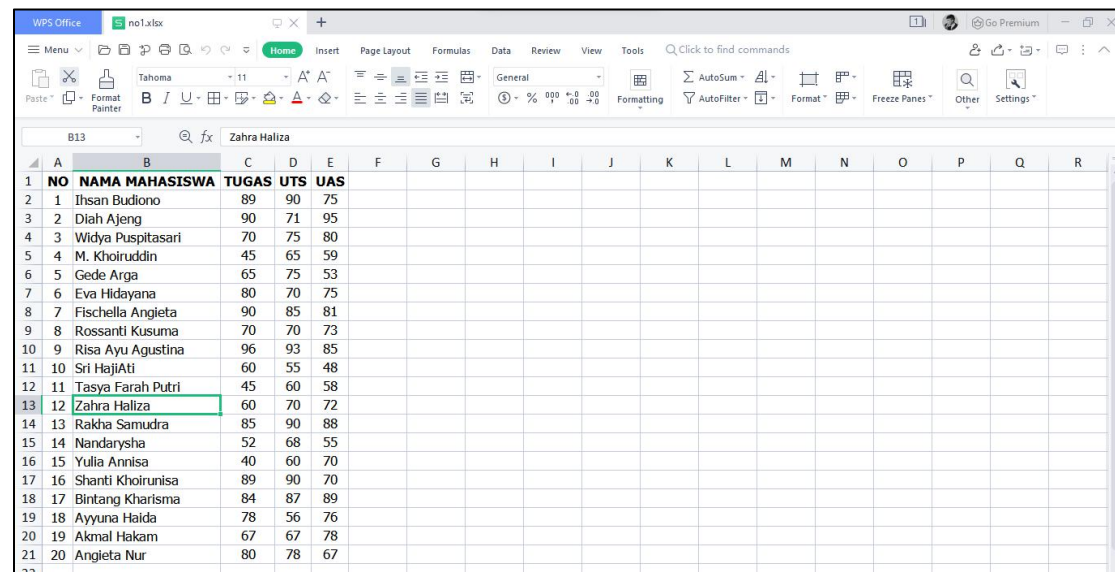
Oleh:

NAMA : Daffa Putra Alwansyah
NIM : L200190031
KELAS : B
PRODI : INFORMATIKA

Fakultas Komunikasi dan Informatika
Universitas Muhammadiyah Surakarta

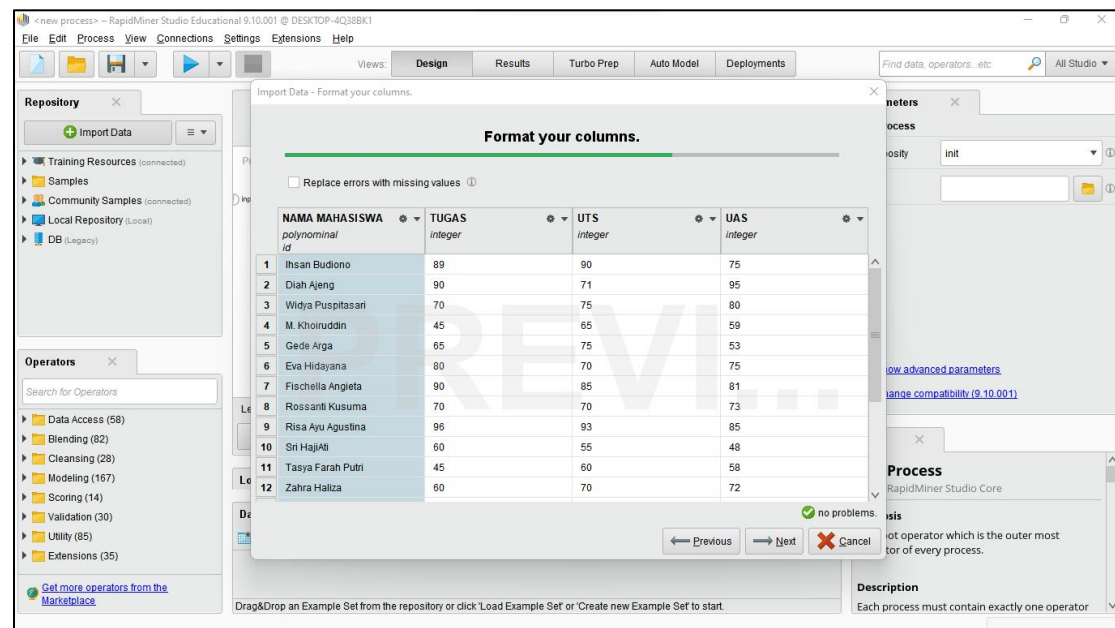
1. Tabel di bawah merupakan data nilai dari 20 mahasiswa yang mengambil mata kuliah Data Warehouse dan Data Mining.

Tabel Nomor 1



NO	NAMA MAHASISWA	TUGAS	UTS	UAS
1	Ihsan Budiono	89	90	75
2	Diah Ajeng	90	71	95
3	Widya Puspitasari	70	75	80
4	M. Khoiruddin	45	65	59
5	Gede Arga	65	75	53
6	Eva Hidayana	80	70	75
7	Fischella Angieta	90	85	81
8	Rossanti Kusuma	70	70	73
9	Risa Ayu Agustina	96	93	85
10	Sri HajiAti	60	55	48
11	Tasya Farah Putri	45	60	58
12	Zahra Haliza	60	70	72
13	Rakha Samudra	85	90	88
14	Nandarysha	52	68	55
15	Yulia Annisa	40	60	70
16	Shanti Khoirunisa	89	90	70
17	Bintang Kharisma	84	87	89
18	Ayyuna Haida	78	56	76
19	Akmal Hakam	67	67	78
20	Angieta Nur	80	78	67

Menggunakan 4 data, dengan ketentuan Nama Mahasiswa (polynomial, id), Tugas (integer), UTS (integer), dan UAS (integer).



Format your columns.

☐ Replace errors with missing values ⓘ

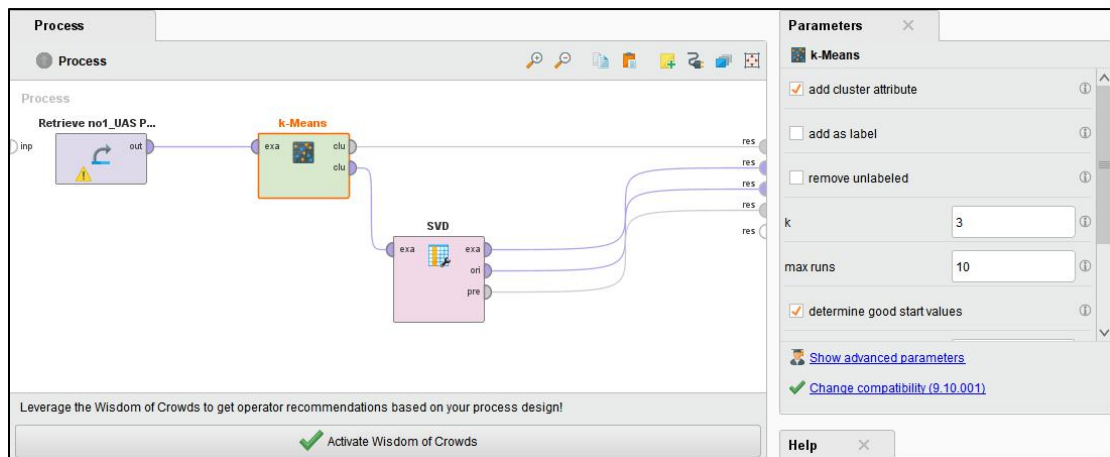
	NAMA MAHASISWA polynomial id	TUGAS integer	UTS integer	UAS integer
1	Ihsan Budiono	89	90	75
2	Diah Ajeng	90	71	95
3	Widya Puspitasari	70	75	80
4	M. Khoiruddin	45	65	59
5	Gede Arga	65	75	53
6	Eva Hidayana	80	70	75
7	Fischella Angieta	90	85	81
8	Rossanti Kusuma	70	70	73
9	Risa Ayu Agustina	96	93	85
10	Sri HajiAti	60	55	48
11	Tasya Farah Putri	45	60	58
12	Zahra Haliza	60	70	72

no problems.

Previous Next Cancel

Drag&Drop an Example Set from the repository or click 'Load Example Set' or 'Create new Example Set' to start.

Drag tabel nomor 1 tadi lalu masukan operator k-Means (Ubah $k = 3$), dan SVD.



a. Tentukan *Singular Value Decomposition*nya!

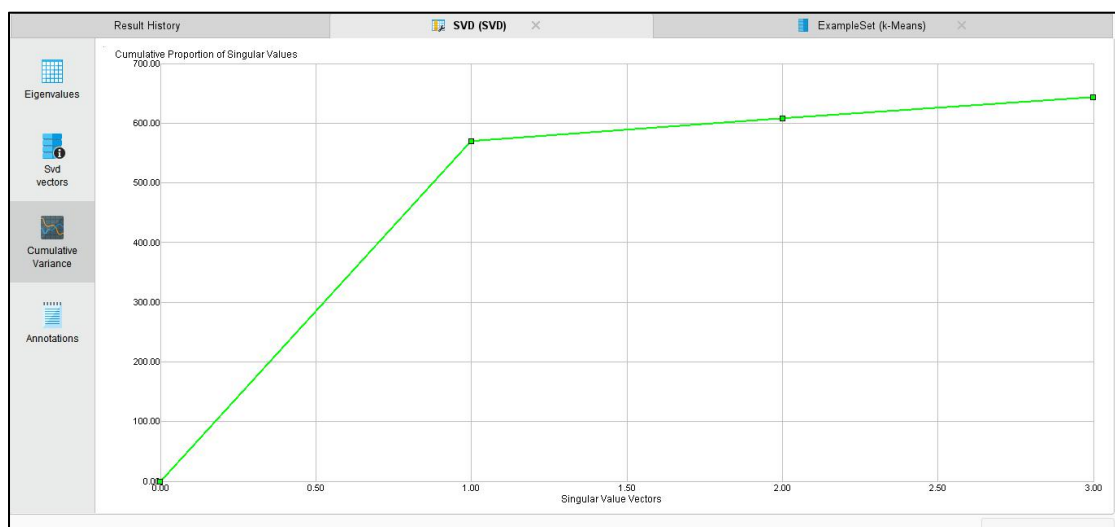
- Nilai Eigenvalue

Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Singular Val...
SVD 1	570.104	0.885	570.104	0.885
SVD 2	38.776	0.060	608.880	0.945
SVD 3	35.445	0.055	644.326	1.000

- Nilai SVD Vector

Attribute	SVD Vector 1	SVD Vector 2
TUGAS	0.575	0.525
UTS	0.584	0.274
UAS	0.573	-0.806

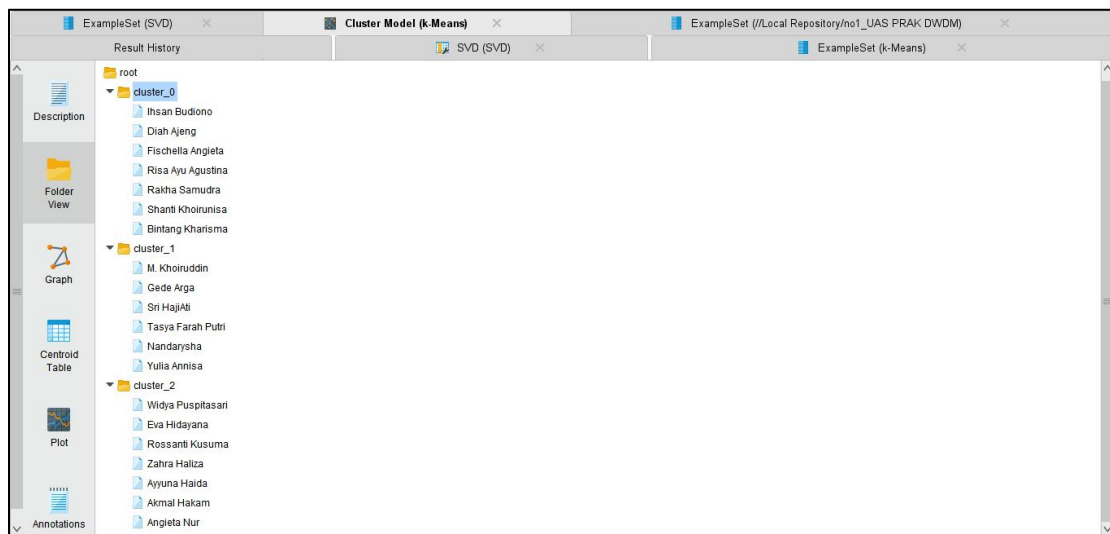
- Nilai Cumulative Variance



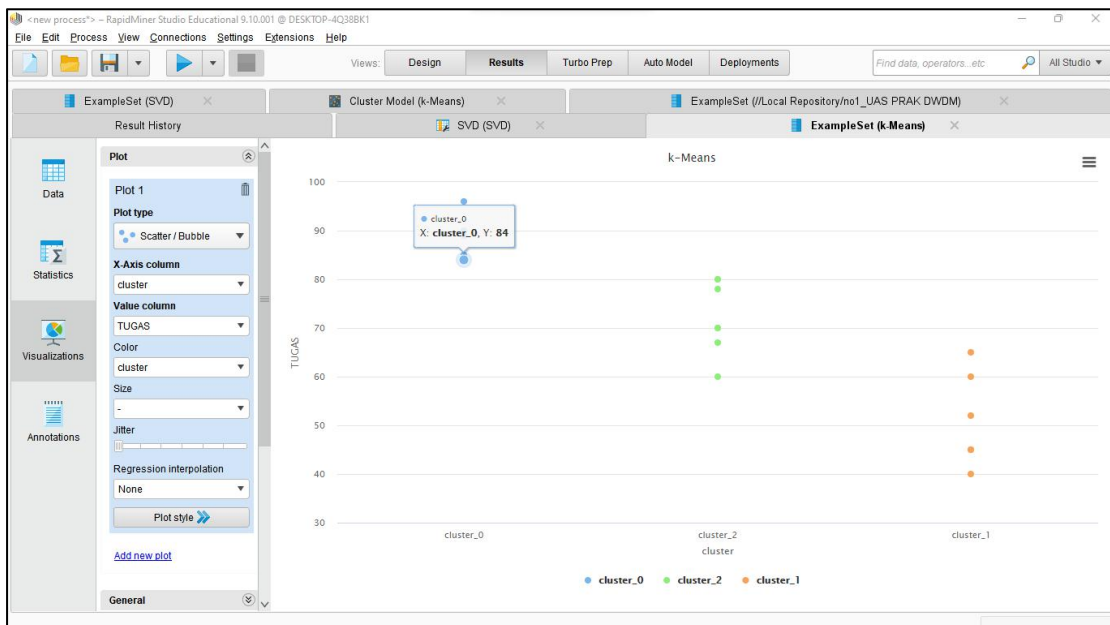
b. Tentukan nilai Example Set (K-Means) dengan memperhatikan:

- Plotter = Scatter
- X-Axis = Cluster
- Y-Axis = Pintar, Sedang, Kurang (Diubah-ubah)
- Color Colomn = Cluster

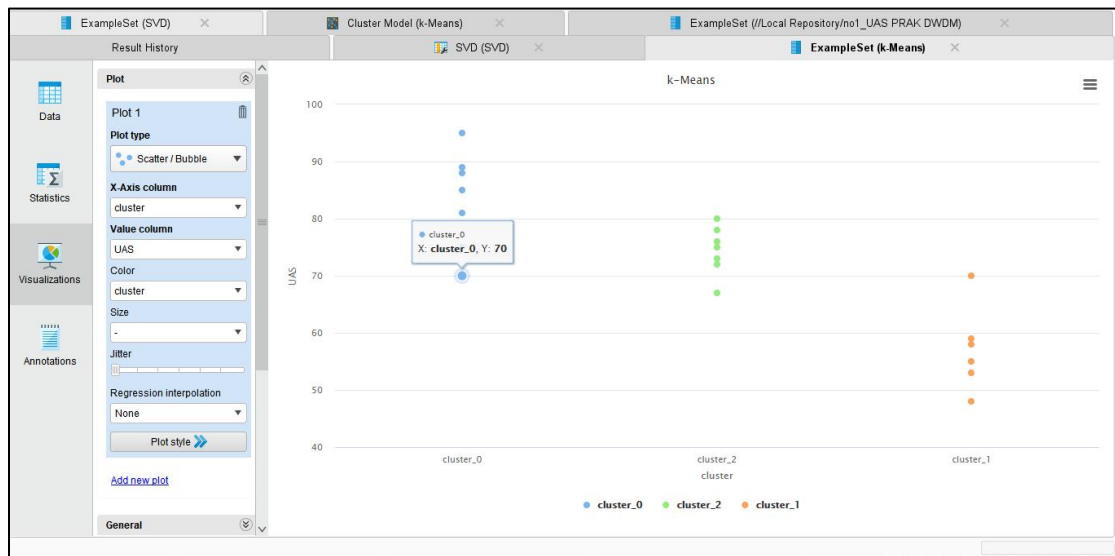
Untuk cluster model disini sudah ada kelompok berdasarkan clusternya Cluster 0 = pintar, Cluster 1 = Sedang, Cluster 2 = Kurang. Totalnya 7 Orang Pintar, 6 Orang sedang, dan 7 Orang kurang.



● Cluster berdasarkan Tugas.



- Cluster berdasarkan UAS



- Cluster berdasarkan UTS



c. Tentukan nilai Example Set (SVD) dan simpulkan hasilnya!

Row No.	NAMA MAHA...	cluster	svd_1
1	Ihsan Budiono	cluster_0	0.257
2	Diah Ajeng	cluster_0	0.259
7	Fischella Ang...	cluster_0	0.259
9	Risa Ayu Agu...	cluster_0	0.278
13	Rakha Samu...	cluster_0	0.266
16	Shanti Khoiru...	cluster_0	0.252
17	Bintang Khari...	cluster_0	0.263
4	M. Khoiruddin	cluster_1	0.171
5	Gede Arga	cluster_1	0.196
10	Sri HajiAli	cluster_1	0.165
11	Tasya Farah ...	cluster_1	0.165
14	Nandarysha	cluster_1	0.177
15	Yulia Annisa	cluster_1	0.172
3	Widya Puspit...	cluster_2	0.228
6	Eva Hidayana	cluster_2	0.228

ExampleSet (20 examples, 2 special attributes, 1 regular attribute)

Data Testing

NO	NAMA MAKANAN	JENIS MAKANAN	LAMA PEMBUATAN (MENIT)	HARGA
1	Soto Ayam	Makanan Berat	15	5.000
2	Pisang Goreng	Cemilan	20	17.000
3	Sosis Goreng	Cemilan	30	9.000
4	Steak	Makanan Berat	35	25.000
5	Ayam Geprek	Makanan Berat	25	15.000
6	Bakso	Makanan Berat	45	21.000
7	Kornet	Makanan Berat	20	20.000
8	Steak	Makanan Berat	40	45.000
9	Nugget Ayam	Cemilan	15	34.000
10	Cireng Ayam	Cemilan	25	38.000

Mengubah Nama Makanan (polynomial), Jenis Makanan (binominal), Lama Pembuatan (integer), Harga (integer), dan Terjual (polynomial, label).

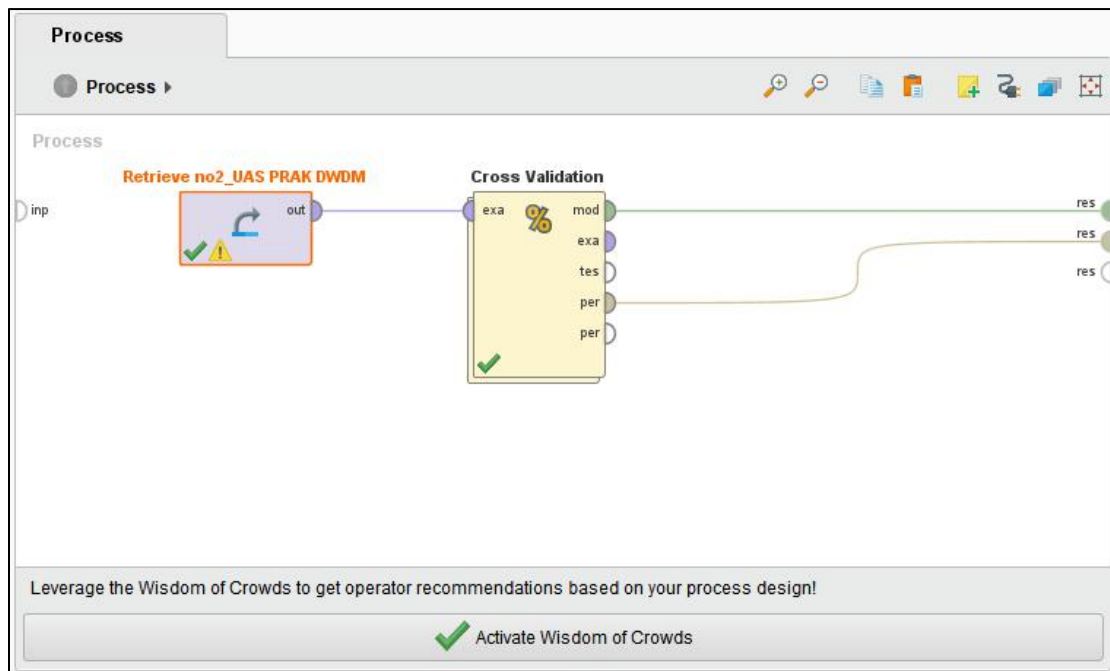
The screenshot shows the RapidMiner Studio interface with the 'Format your columns' dialog box open. The dialog lists the following columns and their data types:

- NAMA MAKANAN: polynomial
- JENIS MAKANAN: binominal
- LAMA PEMBU...: integer
- HARGA: integer
- TERJUAL: polynomial label

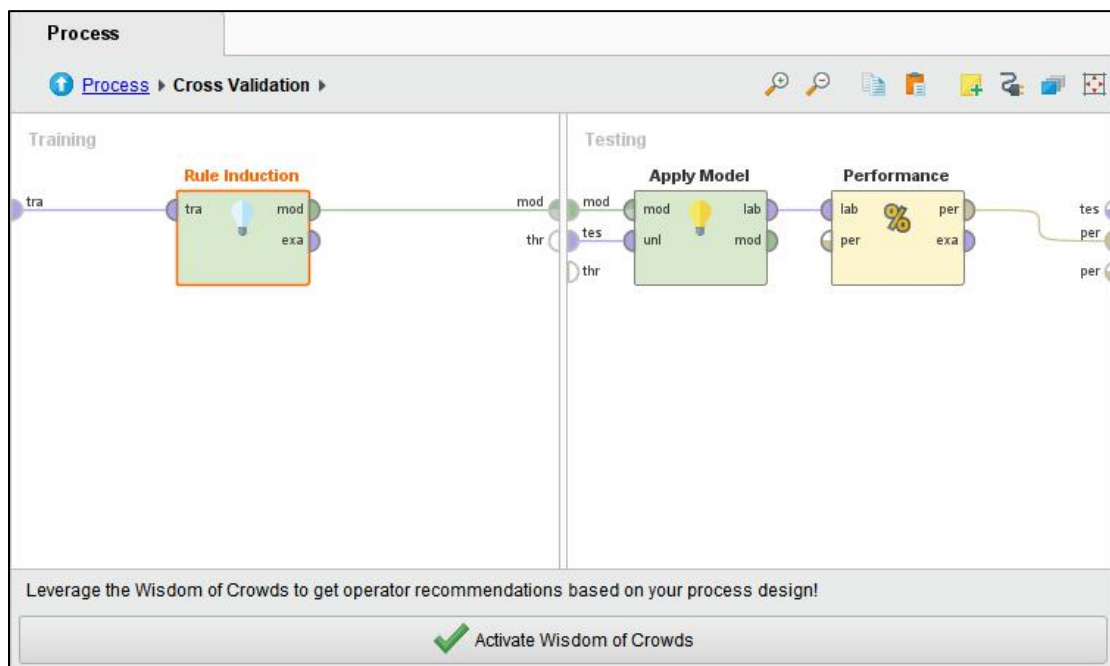
The data preview table shows 12 rows of food items with their respective prices and 'Terjual' status (Habis or Sisa).

NO	NAMA MAKANAN	JENIS MAKANAN	LAMA PEMBU... (MENIT)	HARGA	TERJUAL
1	Soto Ayam	Makanan Berat	15	5000	Habis
2	Nasi Goreng	Makanan Berat	40	21000	Sisa
3	Sosis Goreng	Cemilan	30	9000	Habis
4	Tempe Goreng	Cemilan	15	5000	Habis
5	Ayam Goreng	Makanan Berat	16	14000	Sisa
6	Bakso	Makanan Berat	23	21000	Habis
7	Mie Ayam	Makanan Berat	45	67000	Sisa
8	Steak	Makanan Berat	35	45000	Habis
9	Nugget Ayam	Cemilan	35	34000	Sisa
10	Cireng Ayam	Cemilan	67	38000	Sisa
11	Kornet	Makanan Berat	19	23000	Sisa
12	Bandeng	Makanan Berat	20	21000	Habis

Drag lalu masukan Cross Validation, lalu klik 2x.



Masukan operator Decision Tree, klik kanan operator Decision Tree lalu pilih Replace Operator - Modeling - Predictive - Rules - Rule Induction, lalu klik Run tanpa mengubah apapun.



Akan diperoleh hasil induksi aturan dari data training yang disebut Rule Model (Rule Induction).

The screenshot shows a software window titled "RuleModel (Rule Induction)". It has a sidebar with "Description" and "Annotations" sections. The main area displays the following rule and performance metrics:

RuleModel

if LAMA PEMBUATAN (MENIT) > 37.500 then Sisa (0 / 4)
else Habis (6 / 3)

correct: 10 out of 13 training examples.

Model Rule Induction juga bisa ditunjukan hasil Performance Vector

The screenshot shows a software window titled "PerformanceVector (Performance)". It has a sidebar with "Performance", "Description", and "Annotations" sections. The main area displays a table of performance metrics. The table has columns for "true Habis", "true Sisa", and "class precision". The rows are "pred. Habis", "pred. Sisa", and "class recall".

accuracy: 50.00% +/- 40.82% (micro average: 46.67%)

	true Habis	true Sisa	class precision
pred. Habis	3	4	42.86%
pred. Sisa	4	4	50.00%
class recall	42.86%	50.00%	

3. Sebuah perusahaan melakukan survei kepada 30 karyawannya terhadap lama bekerja seorang karyawan dan besar gajinya. Data karyawan tersebut akan digunakan sebagai dasar perhitungan untuk memprediksi gaji karyawan lainnya berdasarkan lamanya bekerja.

Tabel Daftar Gaji.

	A	B	C	D
1	No	Nama	Tahun_bekerja	Gaji
2	1	Fikri	1,1	39343
3	2	Fitri	1,3	46205
4	3	Fika	1,5	37731
5	4	Fani	2	43525
6	5	Fina	2,2	39891
7	6	Farid	2,9	56642
8	7	Fadel	3	60150
9	8	Galih	3,2	54445
10	9	Gading	3,2	64445
11	10	Guntur	3,7	57189
12	11	Gilang	3,9	63218
13	12	Geri	4	55794
14	13	Gibran	4	56957
15	14	Ita	4,1	57081
16	15	Ilham	4,5	61111
17	16	Indra	4,9	67938
18	17	Ikbal	5,1	66029
19	18	Irwan	5,3	83088
20	19	Ivan	5,9	81363
21	20	Irfan	6	93940
22	21	Karin	6,8	91738
23	22	Kirana	7,1	98273
24	23	Keisya	7,9	101302
25	24	Kevin	8,2	113812
26	25	Keyla	8,7	109431
27	26	Luna	9	105582
28	27	Lala	9,5	116969
29	28	Larisa	9,6	112635
30	29	Latif	10,3	122391
31	30	Lukman	10,5	121872

Tabel prediksi nomor 3

	A	B	C	D
1	No	Nama	Tahun Bekerja	Gaji
2	1	Mila	5.6	
3	2	Monika	4.8	
4	3	Maya	5.7	
5	4	Mira	3.8	
6	5	Malik	7	
7	6	Oki	4.2	
8	7	Okta	8	
9	8	Fadhila	7.5	
10	9	Oskar	8.2	
11	10	Olivia	3	

a. Gunakan RapidMiner untuk membuat model regresi liniernya.

Ubah tipe data No (integer, id), Nama (exclude coloum), Tahun Bekerja (real), dan Gaji (integer, label).

Import Data - Format your columns.

Format your columns.

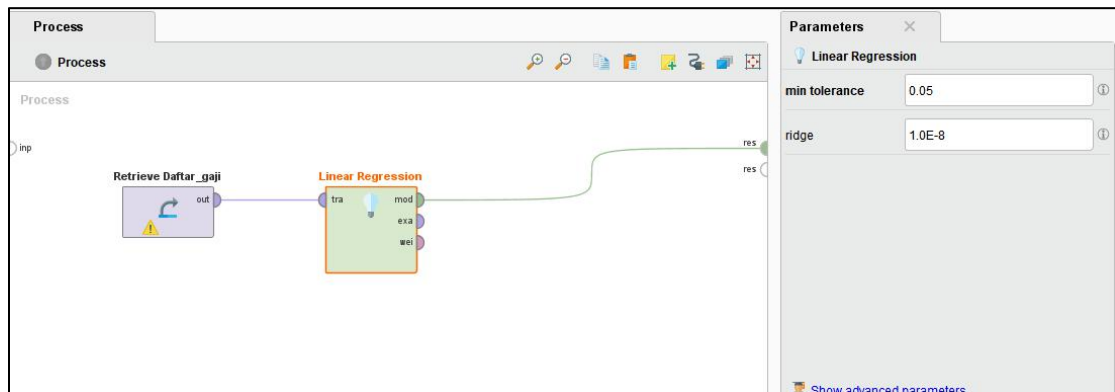
☐ Replace errors with missing values ⓘ

	No	Nama	Tahun_bekerja	Gaji
	<i>integer</i> <i>id</i>	<i>polynomial</i>	<i>real</i>	<i>integer</i> <i>label</i>
1	1	Fikri	1.100	39343
2	2	Fitri	1.300	46205
3	3	Fika	1.500	37731
4	4	Fani	2.000	43525
5	5	Fina	2.200	39891
6	6	Farid	2.900	56642
7	7	Fadel	3.000	60150
8	8	Galih	3.200	54445
9	9	Gading	3.200	64445
10	10	Guntur	3.700	57189
11	11	Gilang	3.900	63218
12	12	Geri	4.000	55794

☒ no problems.

Previous
Next
Cancel

Drag tabel daftar gaji dan masukan operator Linear Regression, pastikan tolerance 0.05 lalu run.



Hasil Proses Regresi Linear:

Tabel View

The screenshot shows the 'Result History' window for the 'LinearRegression (Linear Regression)' tool. The 'Table View' is selected, displaying the following data:

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Tahun_bekerja	9449.962	378.755	0.978	?	24.950	0	****
(Intercept)	25792.200	2273.053	?	?	11.347	0.000	****

Text View

The screenshot shows the 'Result History' window for the 'LinearRegression (Linear Regression)' tool. The 'Text View' is selected, displaying the following text:

LinearRegression

9449.962 * Tahun_bekerja
+ 25792.200

b. Prediksilah gaji dari 10 karyawan yang ada di tabel berikut

Ubah tipe No (integer, id), Nama (exclude coloum), dan Tahun Bekerja (real).

Import Data - Format your columns.

Format your columns.

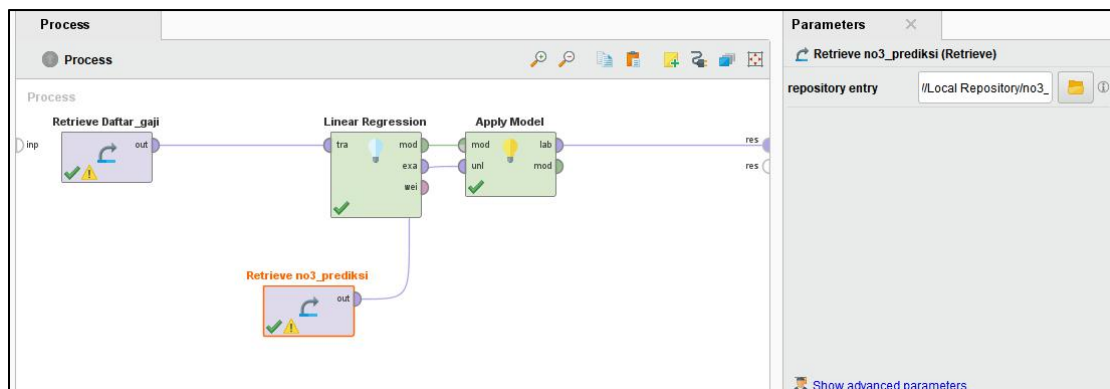
☐ Replace errors with missing values ⓘ

	No <i>integer</i> <i>id</i>	Nama <i>polynomial</i>	Tahun Bekerja <i>real</i>
1	1	Mila	5.600
2	2	Monika	4.800
3	3	Maya	5.700
4	4	Mira	3.800
5	5	Malik	7.000
6	6	Oki	4.200
7	7	Okta	8.000
8	8	Fadhila	7.500
9	9	Oskar	8.200
10	10	Olivia	3.000

no problems.

Previous Next Cancel

Masukan tabel daftar gaji dan prediksi lalu masukan operator Linear Regression dan Apply Model, lalu run.



Hasil prediksi gajinya pada setiap orang:

Result History		ExampleSet (Apply Model)	
<div>Data</div> <div>Statistics</div> <div>Visualizations</div> <div>Annotations</div>	Open in		<div>Turbo Prep</div> <div>Auto Model</div>
	Row No.	No	prediction(G... Tahun Beke...
	1	1	78711.989 5.600
	2	2	71152.019 4.800
	3	3	79656.985 5.700
	4	4	61702.057 3.800
	5	5	91941.936 7
	6	6	65482.042 4.200
	7	7	101391.899 8
	8	8	96666.918 7.500
	9	9	103281.891 8.200
	10	10	54142.087 3

No	Nama	Tahun Bekerja	Gaji
1	Mila	5.6	78711.989
2	Monika	4.8	71152.019
3	Maya	5.7	79656.985
4	Mira	3.8	61702.057
5	Malik	7	91941.936
6	Oki	4.2	65482.042
7	Okta	8	101391.899
8	Fadhila	7.5	96666.918
9	Oskar	8.2	103281.891
10	Olivia	3	54142.087

Pembuktian model Regresi.

	D2		$=(9449,962 * C2) + 25792,2$	
	A	B	C	D
1	No	Nama	Tahun Bekerja	Gaji
2	1	Mila	5.6	#VALUE!
3	2	Monika	4.8	#VALUE!
4	3	Maya	5.7	#VALUE!
5	4	Mira	3.8	#VALUE!
6	5	Malik	7	91941,934
7	6	Oki	4.2	#VALUE!
8	7	Okta	8	101391,896
9	8	Fadhila	7.5	#VALUE!
10	9	Oskar	8.2	#VALUE!
11	10	Olivia	3	54142,086