

Nama : Aldias Ibnu Habib

NIM : L200170166

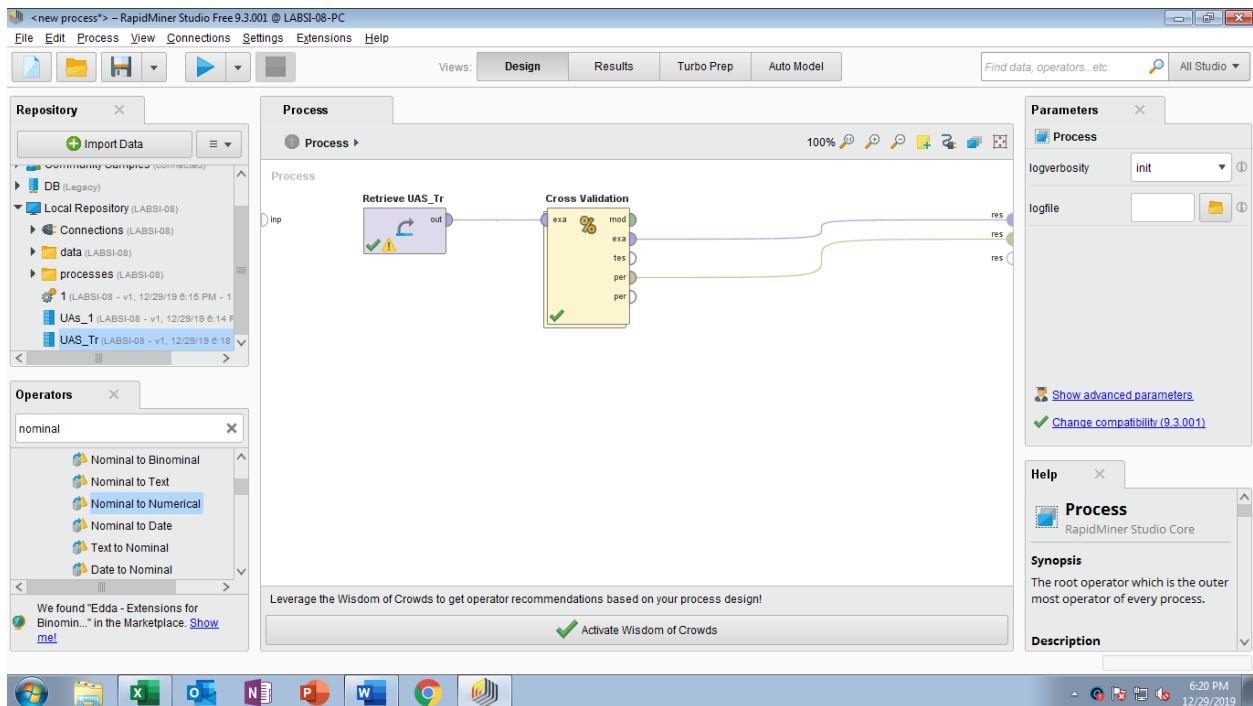
Kelas : ?

Ujian Akhir Semester Praktikum Data Warehouse Data Mining

---

## NOMOR 1

Role model daripada rangkaian naïve bayes, decision tree, dan neural nets operator pada cross validation daiganti dengan masing2 operator yang diinginkan naïve bayes/decision tree/neural net.



<new process> - RapidMiner Studio Free 9.3.001 @ LABSI-08-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

**Repository**

- Community Samples (connected)
- DB (Legacy)
- Local Repository (LABSI-08)
  - Connections (LABSI-08)
  - data (LABSI-08)
  - processes (LABSI-08)
    - 1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1)
    - UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - 1)
    - UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - 1)

**Operators**

nominal

- Nominal to Binomial
- Nominal to Text
- Nominal to Numerical
- Nominal to Date
- Text to Nominal
- Date to Nominal

We found "Edda - Extensions for Binomin..." in the Marketplace. [Show me!](#)

**Process**

Process > Cross Validation

100%

Training

Nominal to Numerical

Decision Tree

Testing

Nominal to Numerical

Apply Model

Performance

Parameters

Cross Validation

☐ leave one out

number of folds: 10

sampling type: automatic

[Show advanced parameters](#)

[Change compatibility \(9.3.001\)](#)

**Help**

Cross Validation

Concurrency

Tags: Cross-Validations, Cross-validation, Folds, K-Folds, K-folds, Validations, Estimations, Evaluations, Performances, Splitting, X-Validation, X-Prediction, Validation

Synopsis

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

<new process> - RapidMiner Studio Free 9.3.001 @ LABSI-08-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

ExampleSet (/Local Repository/UAs\_Tr) ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_1)

Result History

Performance

Criterion

- accuracy
- precision
- recall
- AUC (optimistic)
- AUC
- AUC (pessimistic)

Description

Annotations

**PerformanceVector (Performance)**

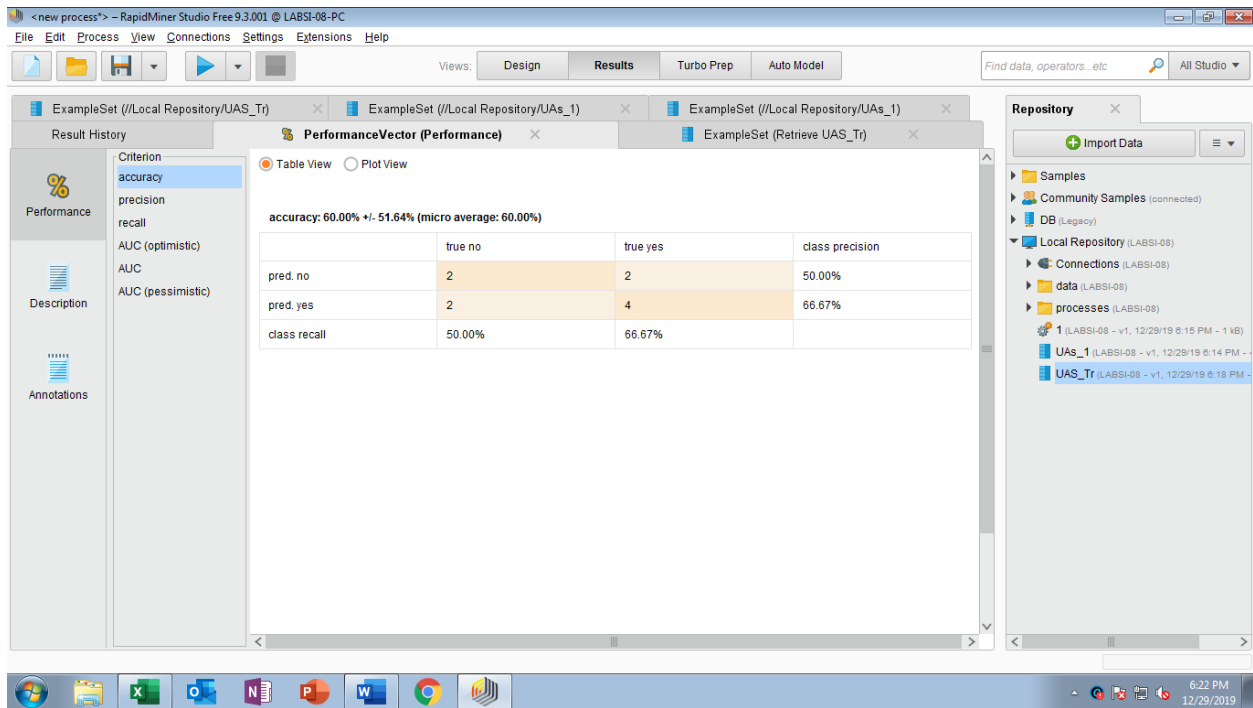
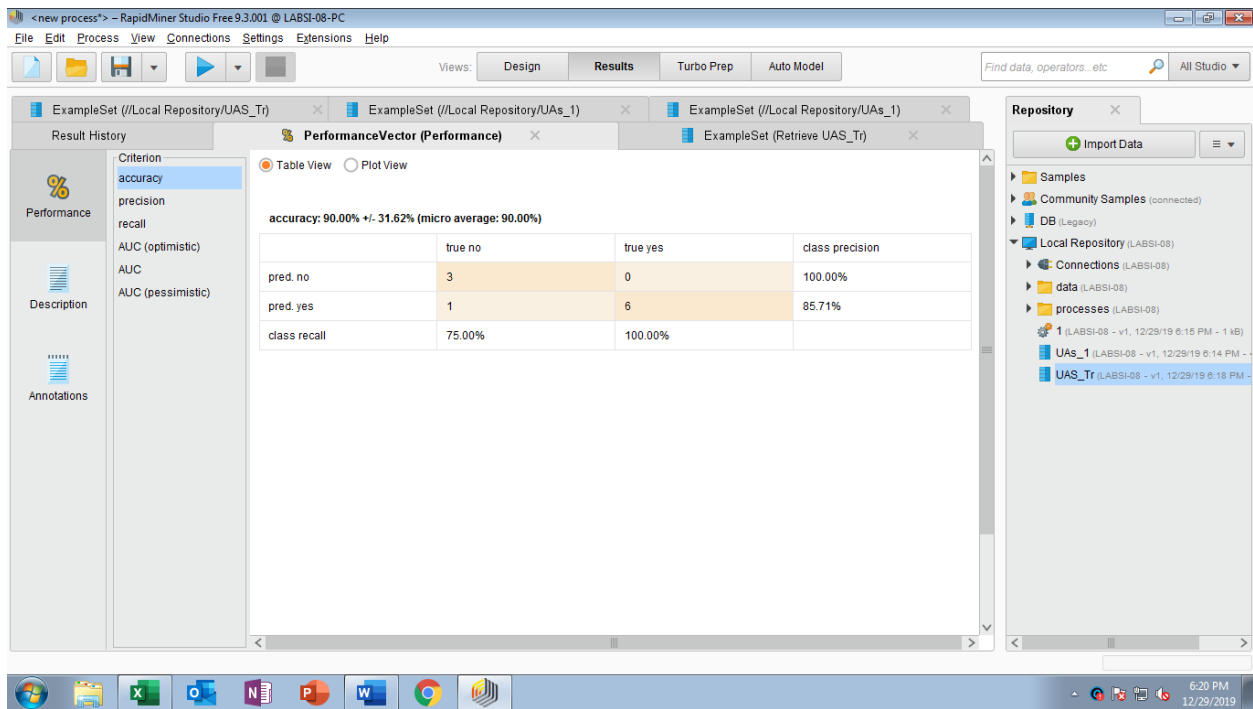
Table View Plot View

accuracy: 70.00% +/- 48.30% (micro average: 70.00%)

	true no	true yes	class precision
pred. no	3	2	60.00%
pred. yes	1	4	80.00%
class recall	75.00%	66.67%	

**Repository**

- Import Data
- Samples
- Community Samples (connected)
- DB (Legacy)
- Local Repository (LABSI-08)
  - Connections (LABSI-08)
  - data (LABSI-08)
  - processes (LABSI-08)
    - 1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 kB)
    - UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - 1)
    - UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - 1)



Dari komparasi menggunakan algoritma klasifikasi (Naïve Bates VS Decission Tree VS Neural Nets) didapatkan hasil seperti gambar diatas, dengan begitu didapatkan Decission tree yang memiliki nilai akurasi lebih baik dibandingkan algoritma klasifikasi yang lain.

**Naïve Bayes = 70%**

**Decission Tree = 90%**

**Neural Nets = 60%**

## NOMOR 2

Pengelompokkan menjadi tiga kelompok pada dataset ini dengan **K-means**

The screenshot shows the RapidMiner Studio interface with a process design in the 'Design' view. The process flow is as follows:

- Retrieve UAs\_2**: Retrieves data from the 'UAs\_2' dataset.
- Clustering**: Performs K-means clustering. The 'add cluster attribute' checkbox is checked. The 'k' parameter is set to 3, and 'max runs' is set to 10.
- SVD**: Performs Singular Value Decomposition on the output of the clustering process.

The 'Parameters' panel for the 'Clustering (k-Means)' operator is visible on the right, showing the following settings:

- ☒ add cluster attribute
- ☐ add as label
- ☐ remove unlabeled
- k: 3
- max runs: 10
- ☒ determine good start values
- [Show advanced parameters](#)
- ☒ Change compatibility (9.3.001)

The 'Help' panel for 'k-Means' is also visible, showing tags: Unsupervised, Clustering, Segmentation, Similarity, Similarities, Euclidean, Distance, K means, Kmeans.

The screenshot shows the 'Results' view of the RapidMiner Studio interface. The 'Result History' panel displays the following data:

Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Sin...
SVD 1	694.512	0.982	694.512	0.982
SVD 2	12.135	0.017	706.647	0.999
SVD 3	0.922	0.001	707.569	1.000

The 'Repository' panel on the right shows the 'Local Repository' with the following data:

- Connections (LABSI-08)
- data (LABSI-08)
- processes (LABSI-08)
- 1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 KB)
- nomor 1 (LABSI-08 - v1, 12/29/19 6:26 PM - 1 KB)
- UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - 1 KB)
- UAs\_2 (LABSI-08 - v1, 12/29/19 6:28 PM - 1 KB)
- UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - 1 KB)

<new process> - RapidMiner Studio Free 9.3.001 @ LABSI-08-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_2) ExampleSet (/Local Repository/UAs\_Tr)

Result History SVD (SVD) ExampleSet (Clustering) ExampleSet (SVD) Cluster Model (Clustering)

Open in Turbo Prep Auto Model Filter (10 / 10 examples): all

Row No.	id	cluster	weight	cholesterol	gender
2	2	cluster_0	156	192	0
4	4	cluster_0	178	213	0
6	6	cluster_0	168	204	1
7	7	cluster_0	152	189	0
9	9	cluster_0	153	189	0
1	1	cluster_1	96	102	0
5	5	cluster_1	109	125	0
10	10	cluster_1	107	122	0
3	3	cluster_2	125	152	0
8	8	cluster_2	133	163	0

ExampleSet (10 examples, 2 special attributes, 3 regular attributes)

Repository

- Import Data
- Samples
- Community Samples (connected)
- DB (Legacy)
- Local Repository (LABSI-08)
  - Connections (LABSI-08)
  - data (LABSI-08)
  - processes (LABSI-08)
    - 1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 kB)
    - nomor 1 (LABSI-08 - v1, 12/29/19 6:26 PM)
    - UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - )
    - UAs\_2 (LABSI-08 - v1, 12/29/19 6:28 PM - )
    - UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - )

<new process> - RapidMiner Studio Free 9.3.001 @ LABSI-08-PC

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators...etc All Studio

ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_2) ExampleSet (/Local Repository/UAs\_Tr)

Result History SVD (SVD) ExampleSet (Clustering) ExampleSet (SVD) Cluster Model (Clustering)

Open in Turbo Prep Auto Model Filter (10 / 10 examples): all

Row No.	id	cluster	svd_1
1	1	cluster_1	0.201
2	2	cluster_0	0.356
3	3	cluster_2	0.283
4	4	cluster_0	0.400
5	5	cluster_1	0.239
6	6	cluster_0	0.381
7	7	cluster_0	0.349
8	8	cluster_2	0.303
9	9	cluster_0	0.350
10	10	cluster_1	0.234

ExampleSet (10 examples, 2 special attributes, 1 regular attribute)

Repository

- Import Data
- Samples
- Community Samples (connected)
- DB (Legacy)
- Local Repository (LABSI-08)
  - Connections (LABSI-08)
  - data (LABSI-08)
  - processes (LABSI-08)
    - 1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 kB)
    - nomor 1 (LABSI-08 - v1, 12/29/19 6:26 PM)
    - UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - )
    - UAs\_2 (LABSI-08 - v1, 12/29/19 6:28 PM - )
    - UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - )

# Cluster Model

Cluster 0: 5 items  
 Cluster 1: 3 items  
 Cluster 2: 2 items  
 Total number of items: 10

Pengelompokkan menjadi tiga kelompok pada dataset ini dengan **Fuzzy K-means**.

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

- Repository:** Lists data sources including 'UAs\_2'.
- Process:** A workflow diagram showing the sequence: 'Retrieve UAs\_2' (input) → 'Fuzzy C-Means' (clustering) → 'SVD' (dimensionality reduction).
- Parameters:** Configured for 'Fuzzy C-Means' with:
  - add cluster attribute: ☒
  - add as label: ☐
  - Clusters: 3
  - Iterations: 50
  - measure types: MixedMeasur...
  - mixed measure: MixedEuclide...
- Help:** Provides information about 'Fuzzy C-Means' (FCM), including tags like 'Clustering' and a synopsis.

The screenshot displays the 'Results' view of the RapidMiner Studio, showing the output of the 'SVD' process. The 'ExampleSet (SVD)' is expanded, revealing a table of singular values and proportions.

Component	Singular Value	Proportion of Singular Values	Cumulative Singular Values	Cumulative Proportion of Sin...
SVD 1	362.100	0.983	362.100	0.983
SVD 2	6.088	0.017	368.187	1.000
SVD 3	0.136	0.000	368.323	1.000

On the left, a sidebar shows 'Eigenvalues', 'Svd vectors', 'Cumulative Variance', and 'Annotations'.

ExampleSet (Fuzzy C-Means) → SVD.original

Row No.	cluster	weight	cholesterol	gender
1	cluster_0	104.219	116.717	0.000
2	cluster_1	129.364	157.914	0.000
3	cluster_2	161.374	197.371	0.221

ExampleSet (3 examples, 1 special attribute, 3 regular attributes)

ExampleSet (SVD)

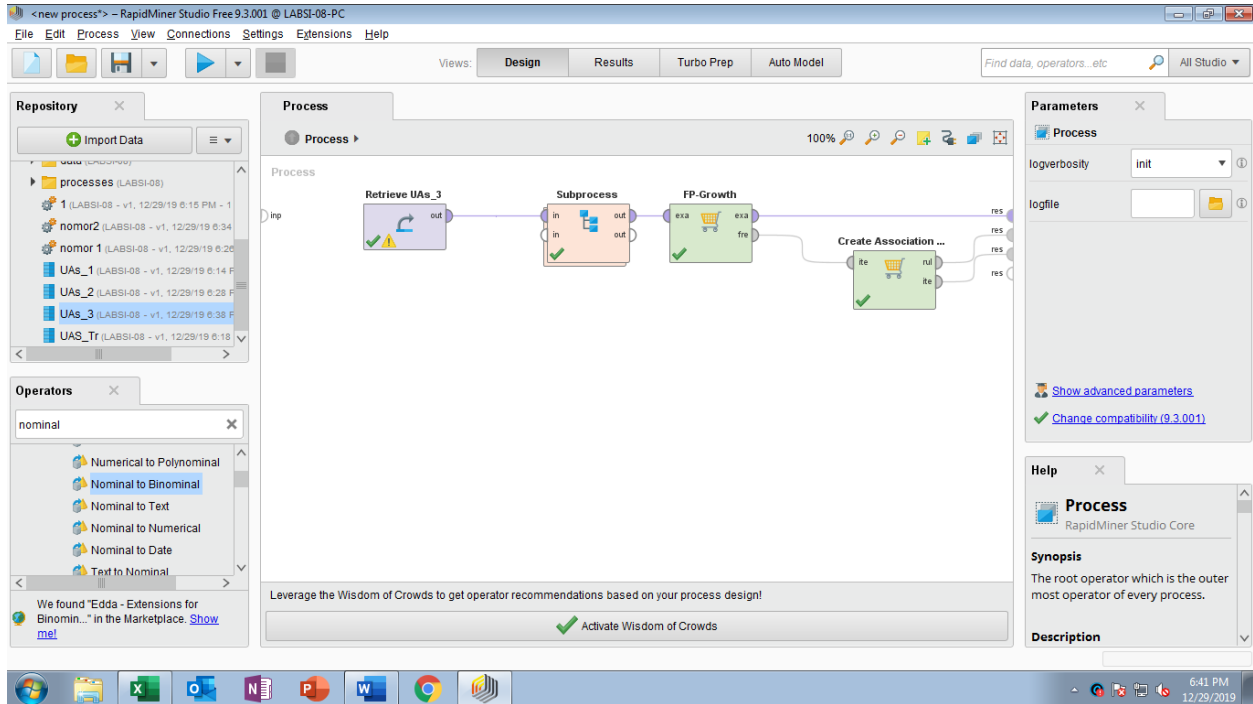
Row No.	cluster	svd_1
1	cluster_0	0.432
2	cluster_1	0.564
3	cluster_2	0.704

ExampleSet (3 examples, 1 special attribute, 1 regular attribute)

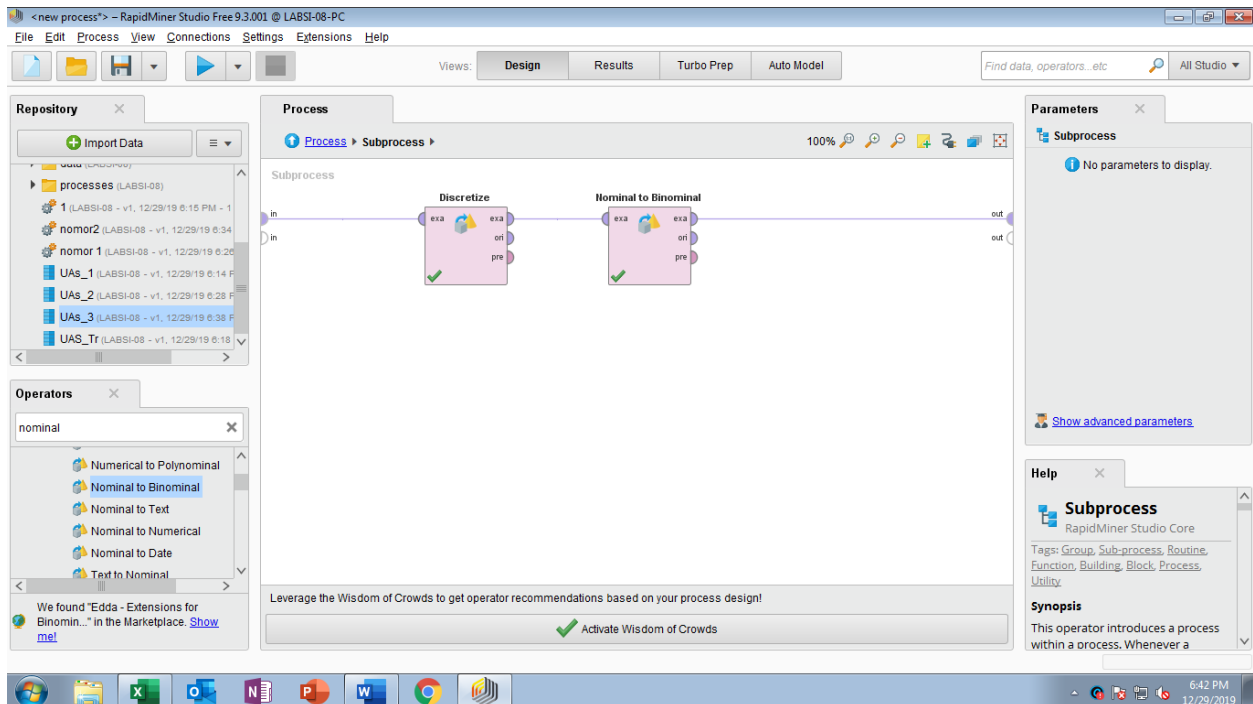
Dari gambar diatas dapat disimpulkan bahwa hasil dari kedua operasi tersebut k-means, dan fuzzy k-means mendapatkan hasil clustering lebih sedikit seperti gambar diatas dengan 3 cluster dan ketentuan dari svd seperti data masing masing model.

## NOMOR 3

Dengan algoritma FP-Growth untuk mendapatkan aturan asosiasi pada dataset berikut



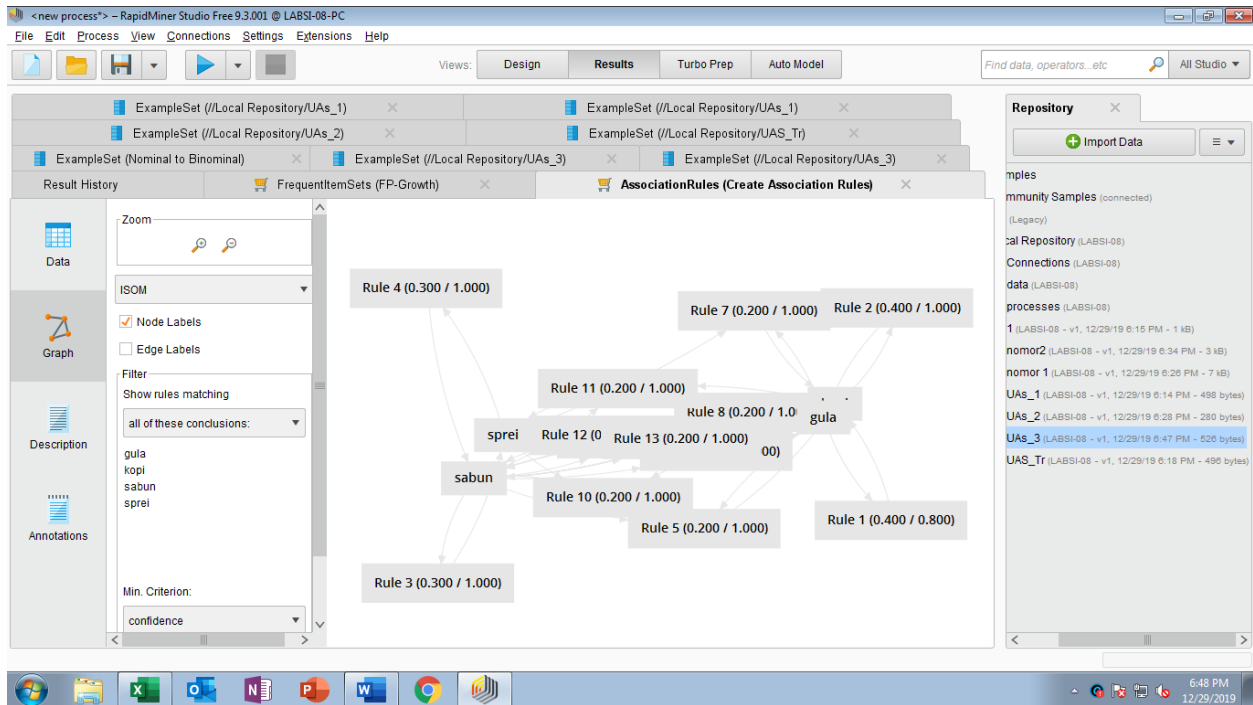
Klik ganda pada subprocess dan rule model seperti berikut



Hasil didapatkan frekuensiItemSet(FP-Growth) berikut



1	0.500	gula			
1	0.400	aqua			
1	0.400	kopi			
1	0.300	sabun			
1	0.300	sprei			
2	0.400	gula	kopi		
2	0.200	gula	sabun		
2	0.200	gula	sprei		
2	0.200	kopi	sabun		
2	0.200	kopi	sprei		
2	0.300	sabun	sprei		
3	0.200	gula	kopi	sabun	
3	0.200	gula	kopi	sprei	
3	0.200	gula	sabun	sprei	
3	0.200	kopi	sabun	sprei	
4	0.200	gula	kopi	sabun	sprei



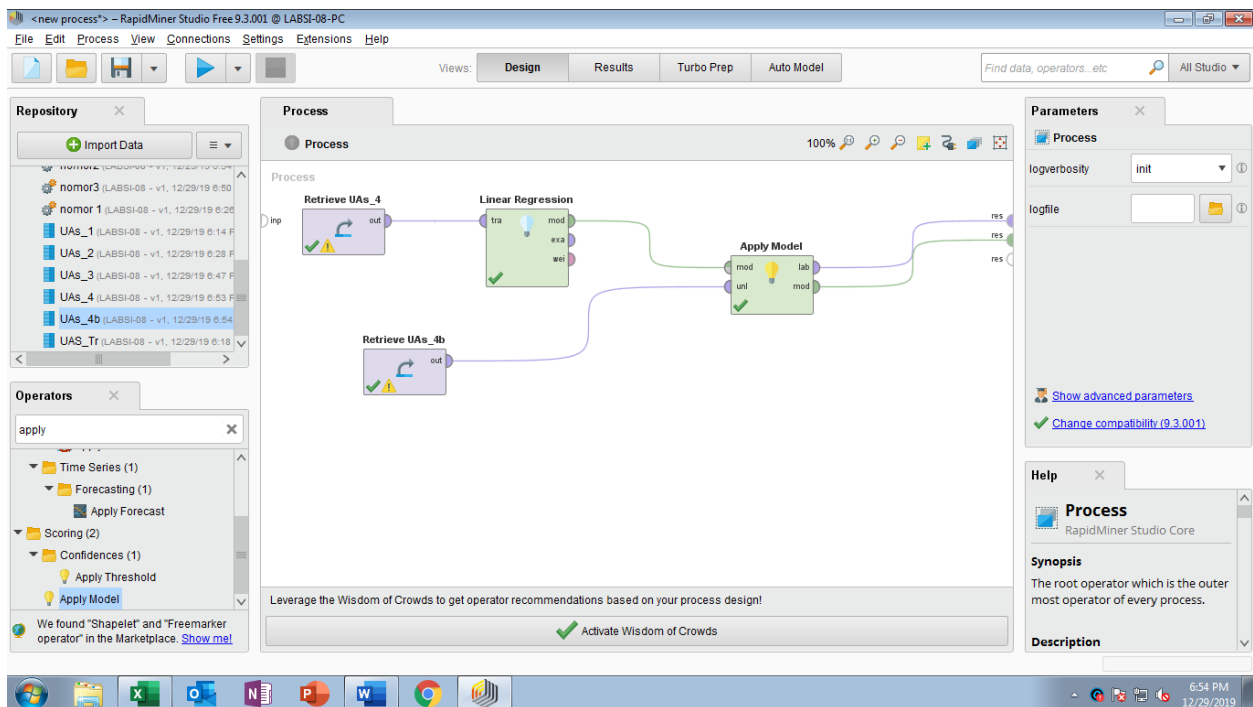
# Didapatkan Association Rules

## Association Rules

```
[gula] --> [kopi] (confidence: 0.800)
[kopi] --> [gula] (confidence: 1.000)
[sabun] --> [sprei] (confidence: 1.000)
[sprei] --> [sabun] (confidence: 1.000)
[gula, sabun] --> [kopi] (confidence: 1.000)
[kopi, sabun] --> [gula] (confidence: 1.000)
[gula, sprei] --> [kopi] (confidence: 1.000)
[kopi, sprei] --> [gula] (confidence: 1.000)
[gula, sabun] --> [sprei] (confidence: 1.000)
[gula, sprei] --> [sabun] (confidence: 1.000)
[kopi, sabun] --> [sprei] (confidence: 1.000)
[kopi, sprei] --> [sabun] (confidence: 1.000)
[gula, sabun] --> [kopi, sprei] (confidence: 1.000)
[kopi, sabun] --> [gula, sprei] (confidence: 1.000)
[gula, kopi, sabun] --> [sprei] (confidence: 1.000)
[gula, sprei] --> [kopi, sabun] (confidence: 1.000)
[kopi, sprei] --> [gula, sabun] (confidence: 1.000)
[gula, kopi, sprei] --> [sabun] (confidence: 1.000)
[gula, sabun, sprei] --> [kopi] (confidence: 1.000)
[kopi, sabun, sprei] --> [gula] (confidence: 1.000)
```

## NOMOR 4

Algoritma linear regression pada dataset ini dengan role model berikut



Didapatkan hasil seperti berikut

The screenshot displays the RapidMiner Studio interface. The main window shows the 'Results' tab with a 'LinearRegression (Linear Regression)' model. The 'Data' view on the left shows a table with the following data:

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
temperature	-0.623	0.283	-0.225	0.445	-2.200	0.064	*
avg_age	2.812	0.358	0.804	0.445	7.857	0.000	****
(Intercept)	113.534	30.064	?	?	3.776	0.007	***

The right sidebar shows the 'Repository' panel with a list of data sources, including 'ExampleSet (/Local Repository/UAs\_1)' through 'ExampleSet (/Local Repository/UAs\_Tr)'. The bottom status bar indicates the time is 6:55 PM on 12/29/2019.

Dan hasil prediksi dari dataset testing sebagai berikut:

prediction:

ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_1)

ExampleSet (/Local Repository/UAs\_3) ExampleSet (/Local Repository/UAs\_2) ExampleSet (/Local Repository/UAs\_Tr)

ExampleSet (/Local Repository/UAs\_4b) ExampleSet (/Local Repository/UAs\_4) ExampleSet (/Local Repository/UAs\_3)

Result History LinearRegression (Linear Regression) ExampleSet (Apply Model)

Open in Turbo Prep Auto Model Filter (10 / 10 examples): all

Row No.	prediction(h...	insulation	temperature	num_occup...	avg_age	home_size
1	231.318	10	74	6	58.300	1
2	275.861	5	49	6	68.600	6
3	180.794	8	45	2	33.900	8
4	222.723	3	49	4	49.700	4
5	258.517	9	66	6	66.200	5
6	275.092	9	57	10	70.100	7
7	164.892	9	66	10	32.900	6
8	239.433	4	47	3	55.200	6
9	136.164	5	53	1	19.800	7
10	253.628	5	77	3	66.900	5

ExampleSet (10 examples, 1 special attribute, 5 regular attributes)

Repository

Import Data

mmunity Samples (connected)

(Legacy)

Local Repository (LABSI-08)

Connections (LABSI-08)

data (LABSI-08)

processes (LABSI-08)

1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 kB)

nomor2 (LABSI-08 - v1, 12/29/19 6:34 PM - 3 kB)

nomor3 (LABSI-08 - v1, 12/29/19 6:50 PM - 6 kB)

nomor 1 (LABSI-08 - v1, 12/29/19 6:26 PM - 7 kB)

UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - 498 bytes)

UAs\_2 (LABSI-08 - v1, 12/29/19 6:28 PM - 280 bytes)

UAs\_3 (LABSI-08 - v1, 12/29/19 6:47 PM - 526 bytes)

UAs\_4 (LABSI-08 - v1, 12/29/19 6:53 PM - 597 bytes)

UAs\_4b (LABSI-08 - v1, 12/29/19 6:54 PM - 502 bytes)

UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - 496 bytes)

ExampleSet (/Local Repository/UAs\_1) ExampleSet (/Local Repository/UAs\_1)

ExampleSet (/Local Repository/UAs\_3) ExampleSet (/Local Repository/UAs\_2) ExampleSet (/Local Repository/UAs\_Tr)

ExampleSet (/Local Repository/UAs\_4b) ExampleSet (/Local Repository/UAs\_4) ExampleSet (/Local Repository/UAs\_3)

Result History LinearRegression (Linear Regression) ExampleSet (Apply Model)

Name Type Missing Statistics Filter (6 / 6 attributes): Search for Attributes

Name	Type	Missing	Min	Max	Average
prediction(heating_oil)	Integer	0	136.164	275.861	223.842
insulation	Integer	0	3	10	6.700
temperature	Integer	0	45	77	58.300
num_occupants	Integer	0	1	10	5.100
avg_age	Real	0	19.800	70.100	52.160
home_size	Integer	0	1	8	5.500

Showing attributes 1 - 6 Examples: 10 Special Attributes: 1 Regular Attributes: 5

Repository

Import Data

mmunity Samples (connected)

(Legacy)

Local Repository (LABSI-08)

Connections (LABSI-08)

data (LABSI-08)

processes (LABSI-08)

1 (LABSI-08 - v1, 12/29/19 6:15 PM - 1 kB)

nomor2 (LABSI-08 - v1, 12/29/19 6:34 PM - 3 kB)

nomor3 (LABSI-08 - v1, 12/29/19 6:50 PM - 6 kB)

nomor 1 (LABSI-08 - v1, 12/29/19 6:26 PM - 7 kB)

UAs\_1 (LABSI-08 - v1, 12/29/19 6:14 PM - 498 bytes)

UAs\_2 (LABSI-08 - v1, 12/29/19 6:28 PM - 280 bytes)

UAs\_3 (LABSI-08 - v1, 12/29/19 6:47 PM - 526 bytes)

UAs\_4 (LABSI-08 - v1, 12/29/19 6:53 PM - 597 bytes)

UAs\_4b (LABSI-08 - v1, 12/29/19 6:54 PM - 502 bytes)

UAs\_Tr (LABSI-08 - v1, 12/29/19 6:18 PM - 496 bytes)

# LinearRegression

- 0.623 \* temperature  
+ 2.812 \* avg\_age  
+ 113.534

