MODULE 11

INDUCTION AND ASSOCIATION RULE

Sort by :

Hamzah L200154013

\

Informatics Study Program

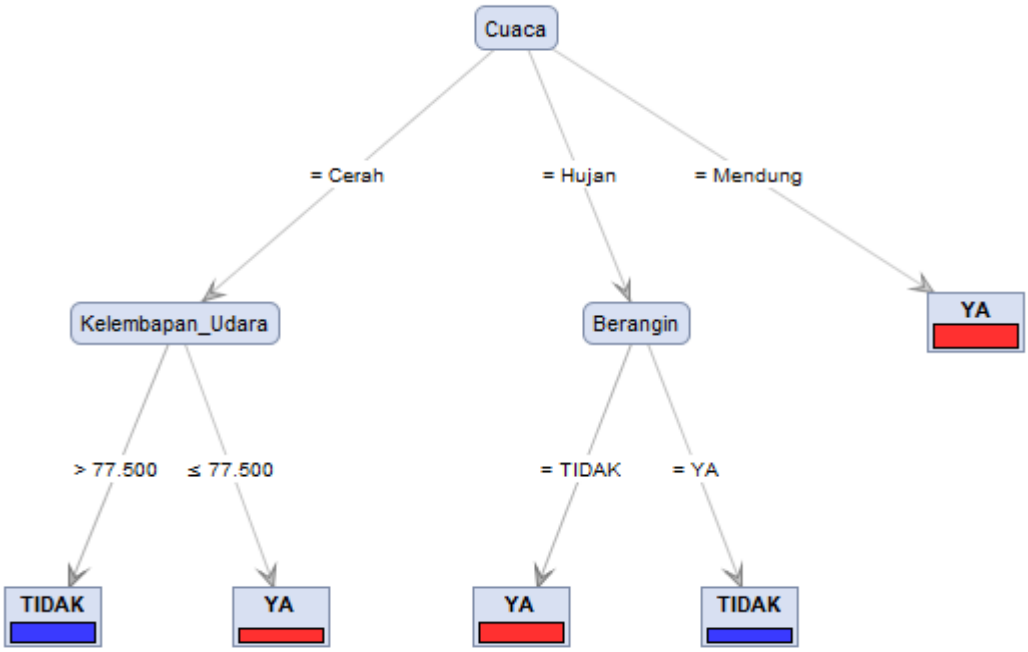
Faculty of Communication and Informatics

Muhammadiyah University of Surakarta

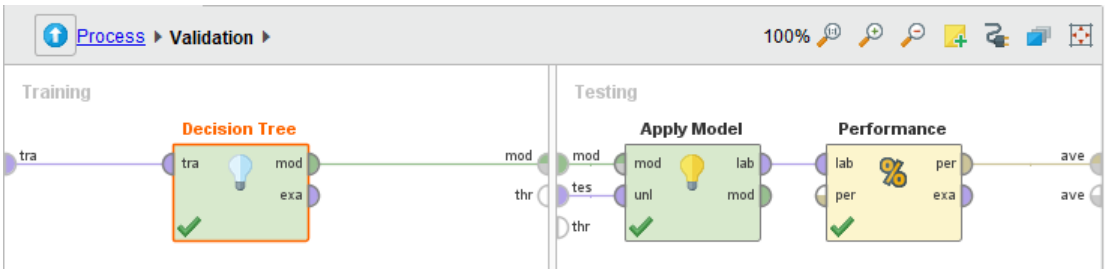
Praktikum Steps

Induction Rules of Data Cuaca

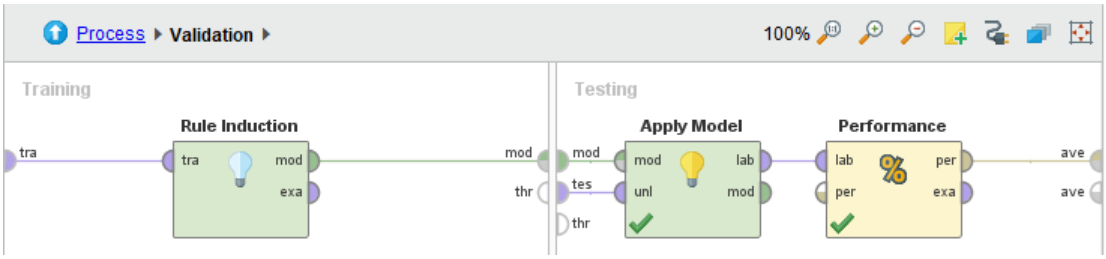
1. Use process model praktikum data mining from module 9 activity 9.4.2 (if not save, redo the activity again).
2. In module 9 activity will produce the following decision tree.



3. Open area **Process – Validation.**



4. Replace Decision Tree operator to Rule Induction then run the process without changing any parameters.



5. An Induction Rule will be obtained from the training data provided which is called a **RuleModel (Rule Model)**.

RuleModel

```
if Kelembapan_Udara ≤ 82.500 then YA  (1 / 6)
if Cuaca = Cerah then TIDAK  (3 / 0)
if Cuaca = Mendung then YA  (0 / 2)
if Suhu ≤ 70.500 then YA  (0 / 1)
else TIDAK  (0 / 0)

correct: 12 out of 13 training examples.
```

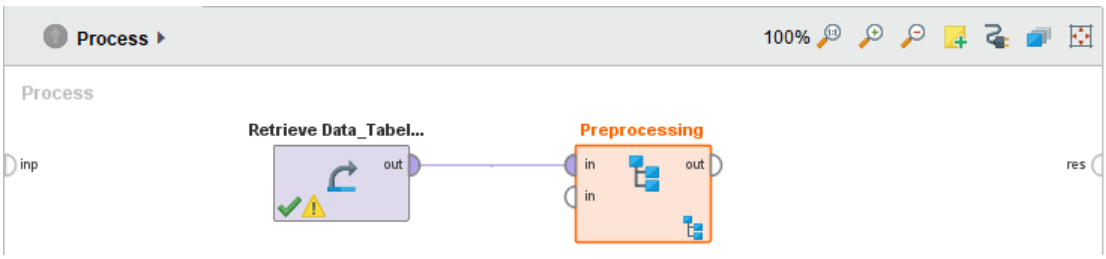
6. From the result of RuleModel we can see causality (using if… then…). The rule can be use as decision making whether someone will be playing tennis or not depends on the weather which occur.
7. Correct parameter showing the number of corrected data in the induction rules on the total data used as training data.
8. Performance of Induction Rule Model

accuracy: 65.00% +/- 47.43% (micro average: 71.43%)

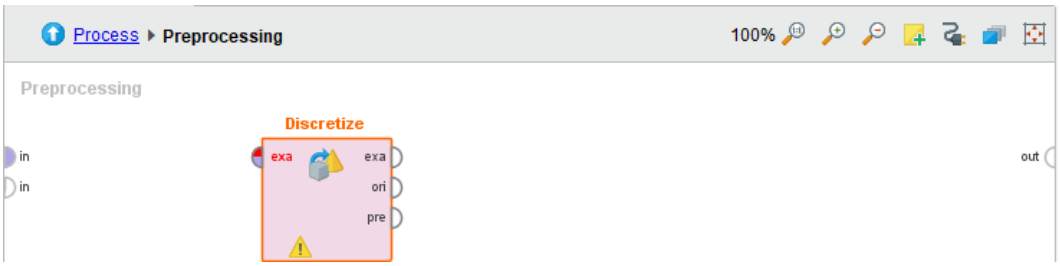
	true TIDAK	true YA	class precision
pred. TIDAK	2	1	66.67%
pred. YA	3	8	72.73%
class recall	40.00%	88.89%	

Data Cuaca Association Rule

1. Create **New Process** in RapidMiner.
2. Drag **DataCuaca_Training** into process area.
3. Add **Subprocess** operator into area then rename it into **Preprocessing.** Connect retrieve output port into preprocessing input port.

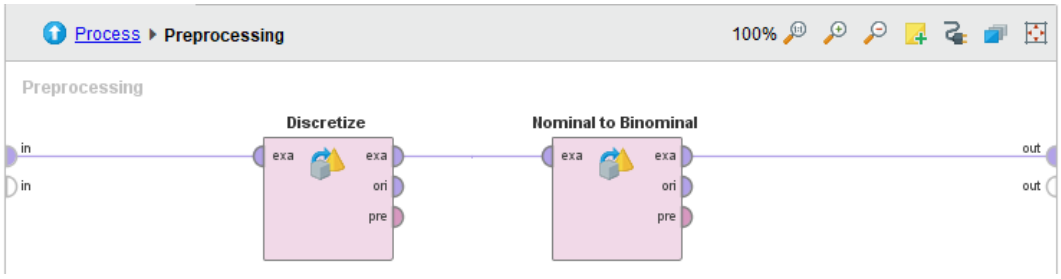


4. Double click on **Preprocessing** operator to open *Nested Chain* area then add the following operators:
   a) **Discretize by Frequency**



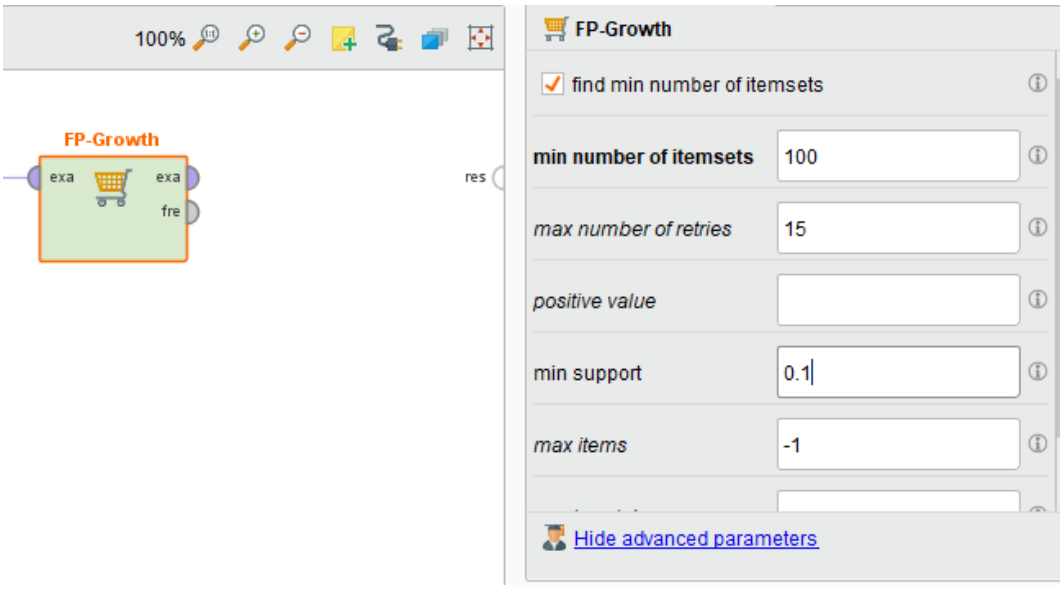
Rename this operator into **DiscretizebyFrequency** and left the parameter **number of bins = 2** then connect the input port into panel **in.**

   b) **Nominal to Binomial**



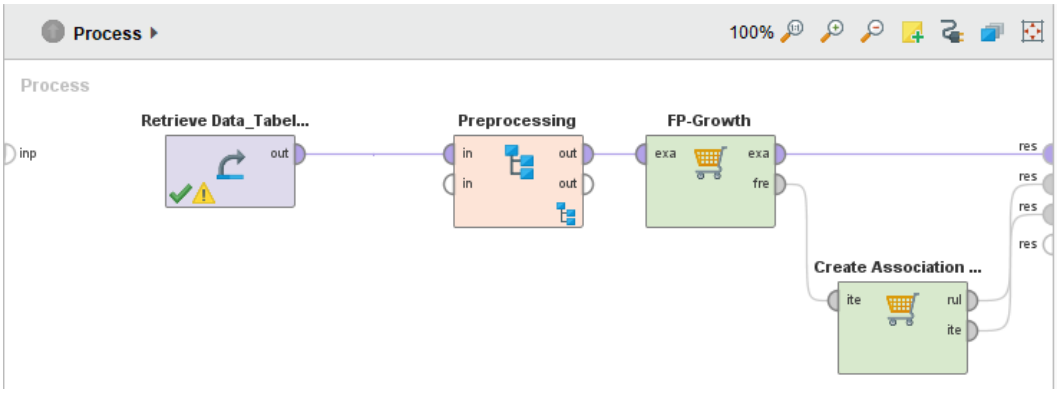
Rename this operator into **Nominal2Binomial**, connect input port with output port of **DiscretizebyFrequency** and output port with out panel.

5. Back into main process, add 2 more operators:
   a) **FP – Growth**



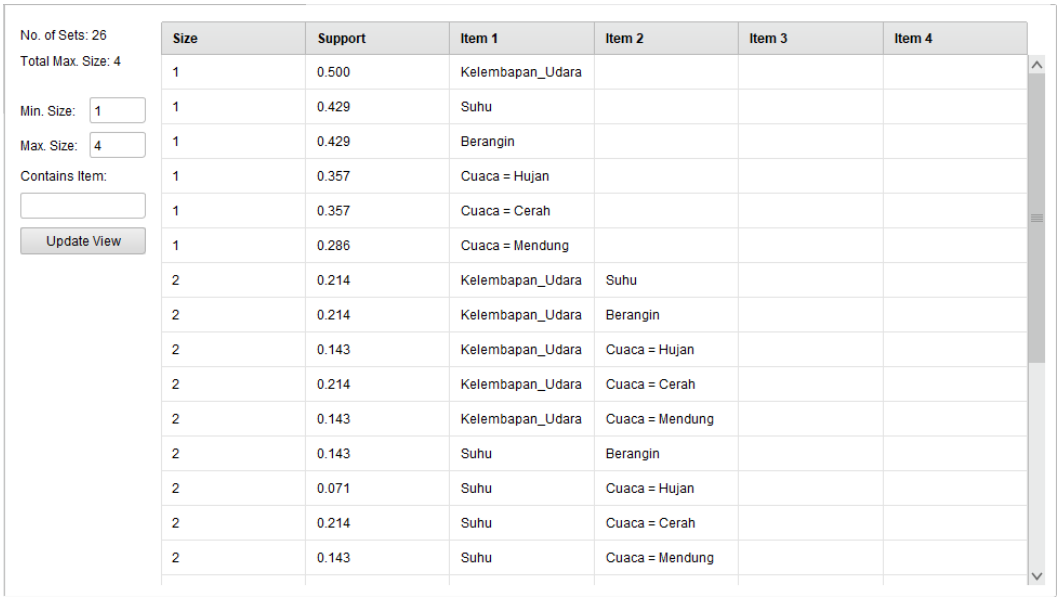
Change parameter **min support into 0.1**, connect output port into with res connector in work area and input port into output port Preprocessing operator.

b) **Create Association Rules**



Connect input port into FP – Growth fre ouput (frequent sets) and both output port into res connector in work area.

6. Run the process
7. The results of the association rules can be seen as follows:
   a) Frequent Item Set (FP – Growth)



No. of Sets: 26
Total Max. Size: 4
Min. Size: 1
Max. Size: 4
Contains Item:
Update View

Size	Support	Item 1	Item 2	Item 3	Item 4
1	0.500	Kelembapan_Udara			
1	0.429	Suhu			
1	0.429	Berangin			
1	0.357	Cuaca = Hujan			
1	0.357	Cuaca = Cerah			
1	0.286	Cuaca = Mendung			
2	0.214	Kelembapan_Udara	Suhu		
2	0.214	Kelembapan_Udara	Berangin		
2	0.143	Kelembapan_Udara	Cuaca = Hujan		
2	0.214	Kelembapan_Udara	Cuaca = Cerah		
2	0.143	Kelembapan_Udara	Cuaca = Mendung		
2	0.143	Suhu	Berangin		
2	0.071	Suhu	Cuaca = Hujan		
2	0.214	Suhu	Cuaca = Cerah		
2	0.143	Suhu	Cuaca = Mendung		

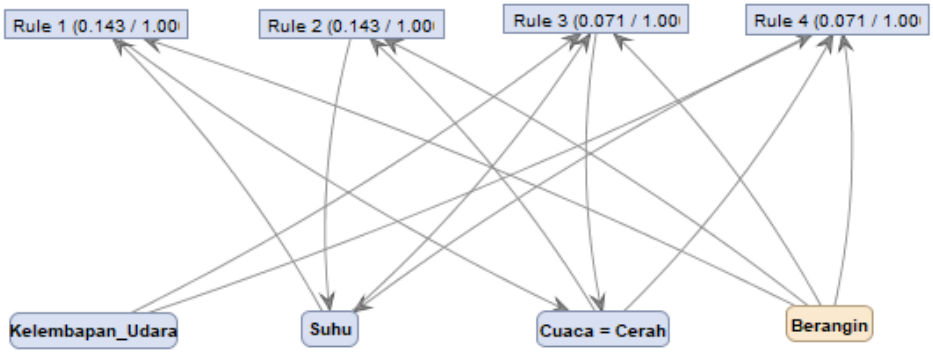
It can be seen that the number of association rules formed is 26 sets, and the total number of max sizes = 4, which consists of 4 itemset.

b) Association Rules (Create Association Rules)
   i. Table View

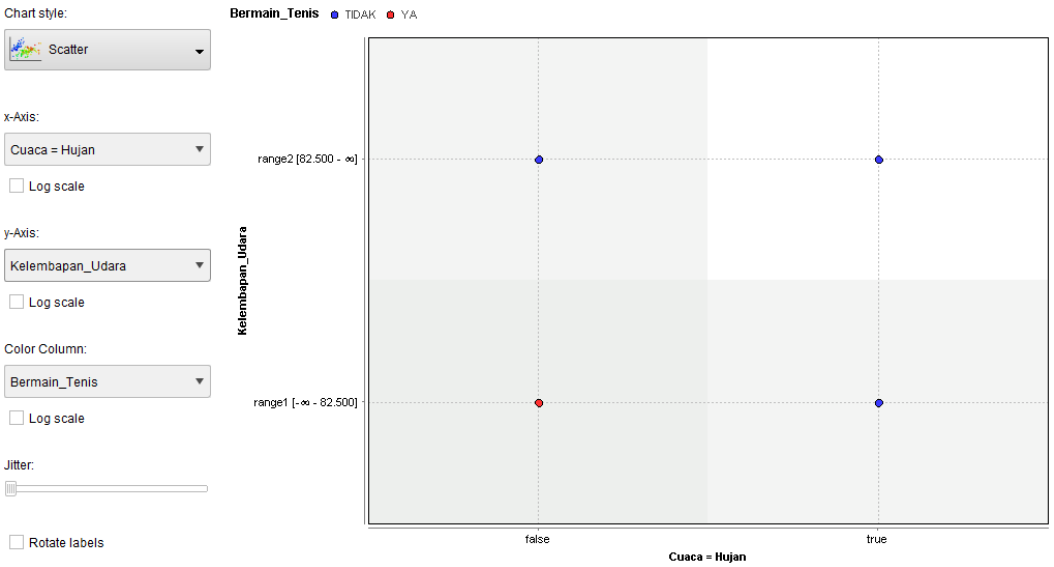
No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
1	Suhu, Berangin	Cuaca = Cerah	0.143	1	1	-0.143	0.092
2	Berangin, Cuaca = Cerah	Suhu	0.143	1	1	-0.143	0.082
3	Kelembapan_Udara, Suhu, Berangin	Cuaca = Cerah	0.071	1	1	-0.071	0.046
4	Kelembapan_Udara, Berangin, Cuaca = Cerah	Suhu	0.071	1	1	-0.071	0.041

In this table, we can see that there are 4 data pairs of premises and conclusions that show the relationship of implications based on association rules.

ii. Graph View



c) ExampleSet (Nominal2Binomial) Charts View



In this result it can be shown the distribution pattern of each weather data which is processed using association rules. We can change the variable in x-Axis and y-Axis and Jitter value. However, make sure that the Color Column is variable dependent, that is Bermain_Tenis.

Task (Using RapidMiner 7.1.001)

1. Induction Rule (Rule Model) and Performance Vector Value using Lama Studi Data in Task Module 6 Question No. 1

RuleModel

```
if Asisten = YA then TEPAT  (0 / 5)
if Rerata_SKS > 19.500 then TEPAT  (0 / 2)
if Gender = PRIA and Asal_Sekolah = LUAR then TERLAMBAT  (2 / 0)
else TEPAT  (4 / 5)

correct: 14 out of 18 training examples.
```

Criterion: accuracy

Table View / Plot View

accuracy: 65.00% +/- 32.02% (mikro: 65.00%)

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	1	1	50.00%
pred. TEPAT	6	12	66.67%
class recall	14.29%	92.31%	

2. With the same data training in no 1, with the provisions of the value operator Discretize by Frequency:
   a. Number of Bins = 2
      a) FP Growth (Table View)

No. of Sets: 55

Total Max. Size: 5

Min. Size: 1

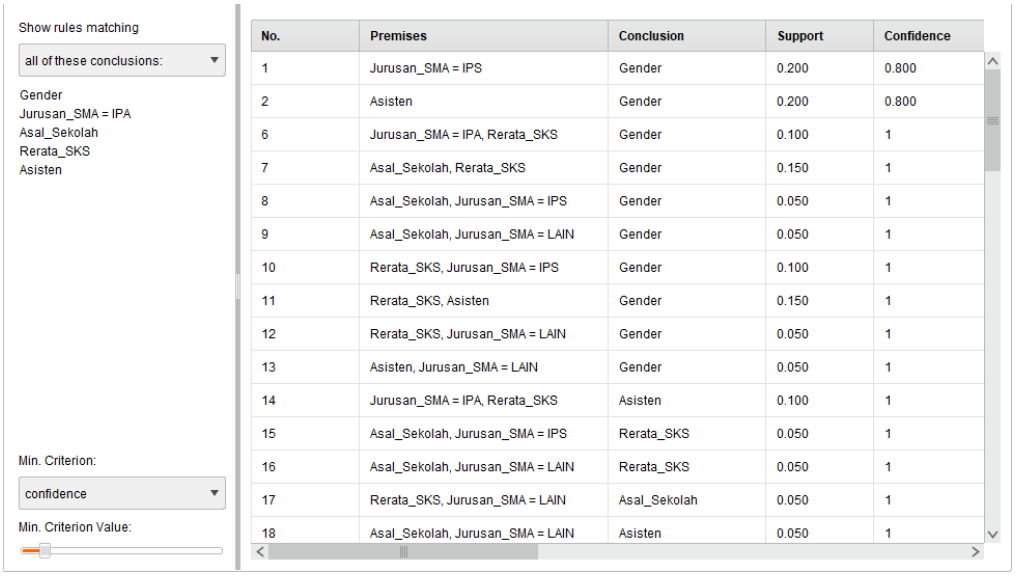
Max. Size: 5

Contains Item:

Update View

Size	Support	Item 1	Item 2
1	0.750	Gender	
1	0.550	Jurusan_SMA = IPA	
1	0.300	Asal_Sekolah	
1	0.250	Rerata_SKS	
1	0.250	Jurusan_SMA = IPS	
1	0.250	Asisten	
1	0.200	Jurusan_SMA = L...	
2	0.400	Gender	Jurusan_SMA = IPA
2	0.250	Gender	Asal_Sekolah

b) Association Rules



Show rules matching

all of these conclusions:

Gender
Jurusan_SMA = IPA
Asal_Sekolah
Rerata_SKS
Asisten

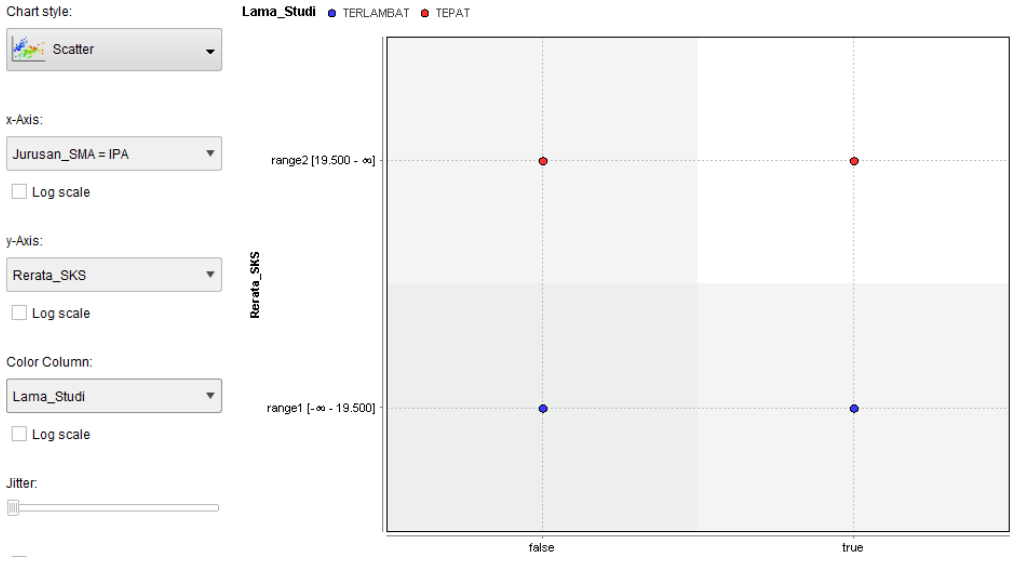
Min. Criterion:

confidence

Min. Criterion Value:

No.	Premises	Conclusion	Support	Confidence
1	Jurusan_SMA = IPS	Gender	0.200	0.800
2	Asisten	Gender	0.200	0.800
6	Jurusan_SMA = IPA, Rerata_SKS	Gender	0.100	1
7	Asal_Sekolah, Rerata_SKS	Gender	0.150	1
8	Asal_Sekolah, Jurusan_SMA = IPS	Gender	0.050	1
9	Asal_Sekolah, Jurusan_SMA = LAIN	Gender	0.050	1
10	Rerata_SKS, Jurusan_SMA = IPS	Gender	0.100	1
11	Rerata_SKS, Asisten	Gender	0.150	1
12	Rerata_SKS, Jurusan_SMA = LAIN	Gender	0.050	1
13	Asisten, Jurusan_SMA = LAIN	Gender	0.050	1
14	Jurusan_SMA = IPA, Rerata_SKS	Asisten	0.100	1
15	Asal_Sekolah, Jurusan_SMA = IPS	Rerata_SKS	0.050	1
16	Asal_Sekolah, Jurusan_SMA = LAIN	Rerata_SKS	0.050	1
17	Rerata_SKS, Jurusan_SMA = LAIN	Asal_Sekolah	0.050	1
18	Asal_Sekolah, Jurusan_SMA = LAIN	Asisten	0.050	1

c) Distribution Graphic in ExampleSet



b. Number of Bins = 3

a) FP Growth (Table View)

No. of Sets: 82
Total Max. Size: 5
Min. Size: 1
Max. Size: 5
Contains Item:
Update View

Size	Support	Item 1	Item 2	Item 3	Item 4	Item 5
1	0.750	Gender				
1	0.550	Jurusan_SMA = ...				
1	0.400	Rerata_SKS = r...				
1	0.350	Rerata_SKS = r...				
1	0.300	Asal_Sekolah				
1	0.250	Rerata_SKS = r...				
1	0.250	Jurusan_SMA = ...				
1	0.250	Asisten				
1	0.200	Jurusan_SMA = ...				
2	0.400	Gender	Jurusan_SMA = ...			
2	0.200	Gender	Rerata_SKS = r...			
2	0.300	Gender	Rerata_SKS = r...			
2	0.250	Gender	Asal_Sekolah			
2	0.250	Gender	Rerata_SKS = r...			

b) Association Rules

Show rules matching
all of these conclusions:
Gender
Jurusan_SMA = IPA
Asal_Sekolah
Rerata_SKS = range3 [19.500 - ∞]
Asisten
Min. Criterion:
confidence
Min. Criterion Value:

No.	Premises	Conclusion	Support
4	Asal_Sekolah	Gender	0.250
5	Rerata_SKS = range2 [18.500 - 19.500]	Gender	0.300
6	Rerata_SKS = range3 [19.500 - ∞]	Gender	0.250
7	Jurusan_SMA = IPA, Rerata_SKS = range3 [19.500 - ∞]	Gender	0.100
8	Rerata_SKS = range2 [18.500 - 19.500], Jurusan_SMA = IPS	Gender	0.100
9	Rerata_SKS = range2 [18.500 - 19.500], Asisten	Gender	0.050
10	Rerata_SKS = range2 [18.500 - 19.500], Jurusan_SMA = LAIN	Gender	0.050
11	Asal_Sekolah, Rerata_SKS = range3 [19.500 - ∞]	Gender	0.150
12	Asal_Sekolah, Jurusan_SMA = IPS	Gender	0.050
13	Asal_Sekolah, Jurusan_SMA = LAIN	Gender	0.050
14	Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA = IPS	Gender	0.100
15	Rerata_SKS = range3 [19.500 - ∞], Asisten	Gender	0.150
16	Rerata_SKS = range3 [19.500 - ∞], Jurusan_SMA = LAIN	Gender	0.050
17	Asisten, Jurusan_SMA = LAIN	Gender	0.050
18	Rerata_SKS = range1 [-∞ - 18.500], Asal_Sekolah	Jurusan_SMA = IPA	0.150

c) Distribution Graphic in ExampleSet