MODULE 7

DATA PREPROCESSING

Sort by :

Hamzah L200154013

\

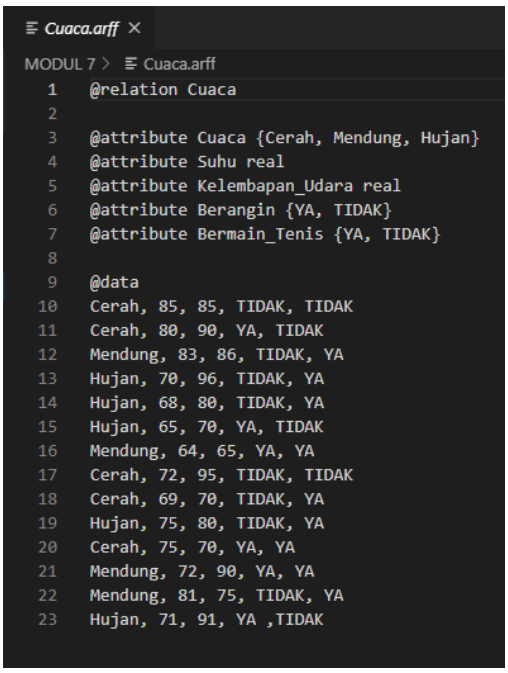
Informatics Study Program

Faculty of Communication and Informatics

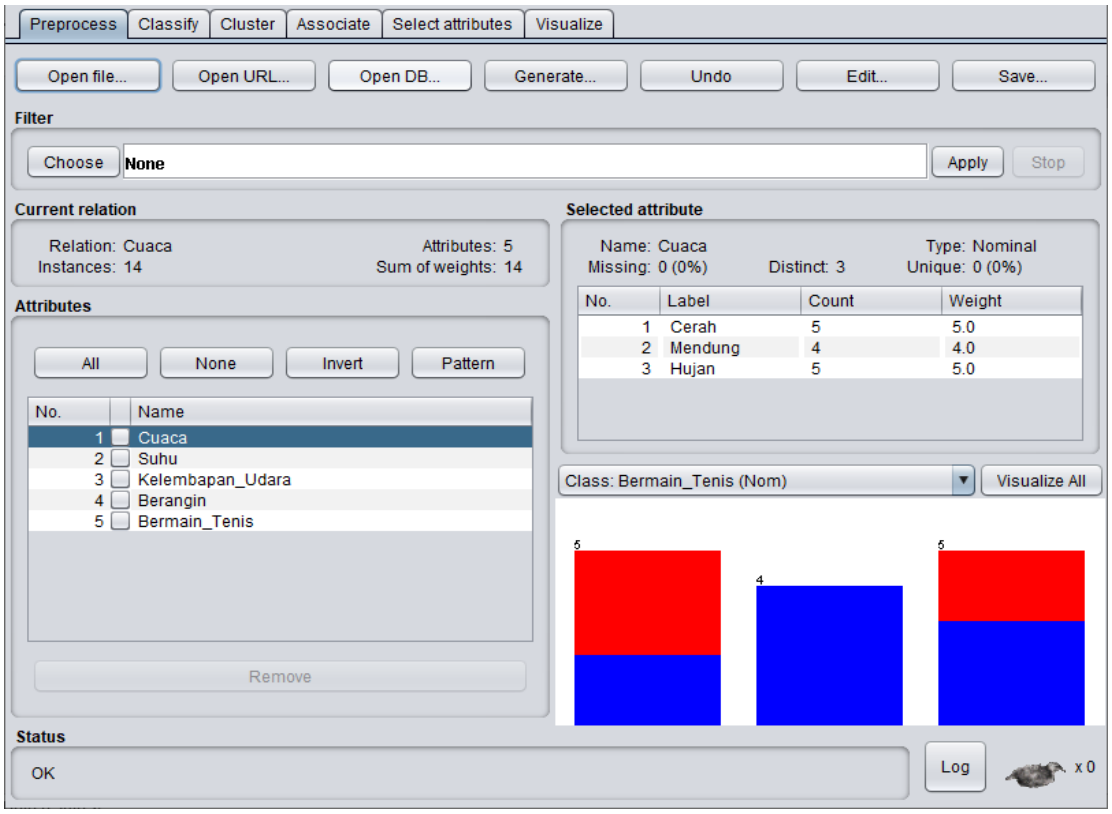
Muhammadiyah University of Surakarta

Praktikum Steps

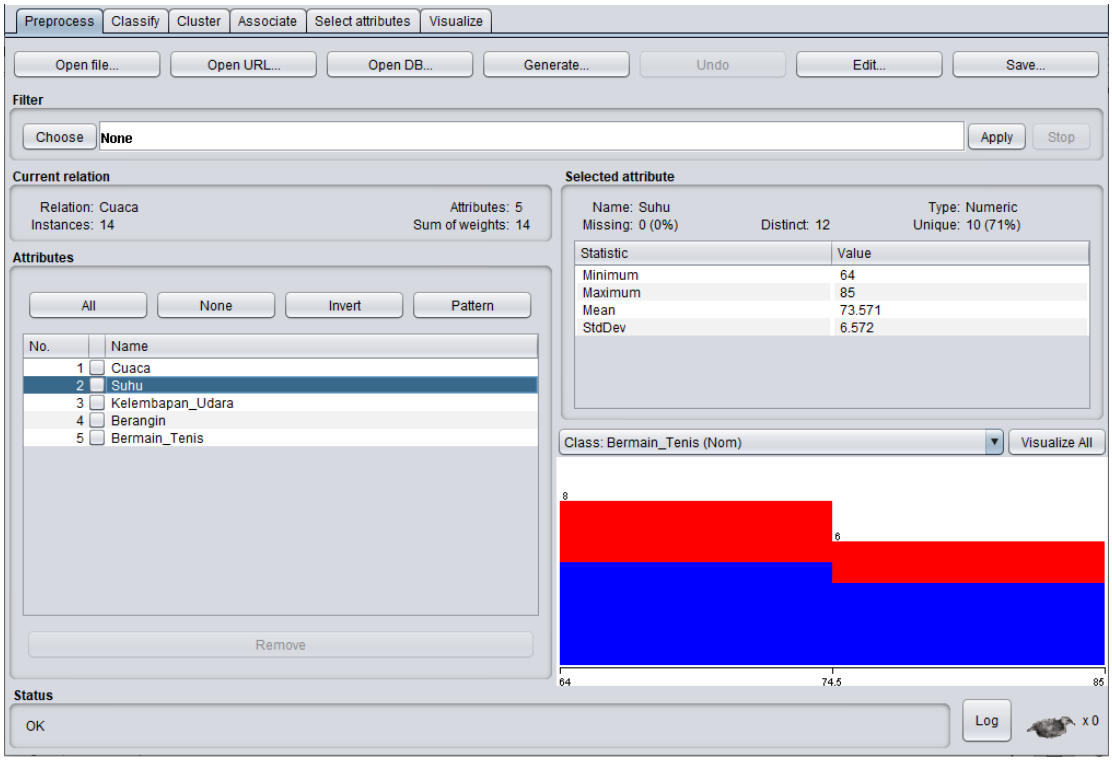
1. In this module 7 we will use the Weka application to process the data that has been prepared, we must first prepare the file with the extension arff.
2. There are 3 important components that must be contained in the file, namely:
   a. @relation, to name the relation of an ARFF file.
   b. @attribute, for naming each attribute either numeric, nominal, string and date.
   c. @data, there are each line to represent an instance.
3. Create a Cuaca.arff file containing data as below.



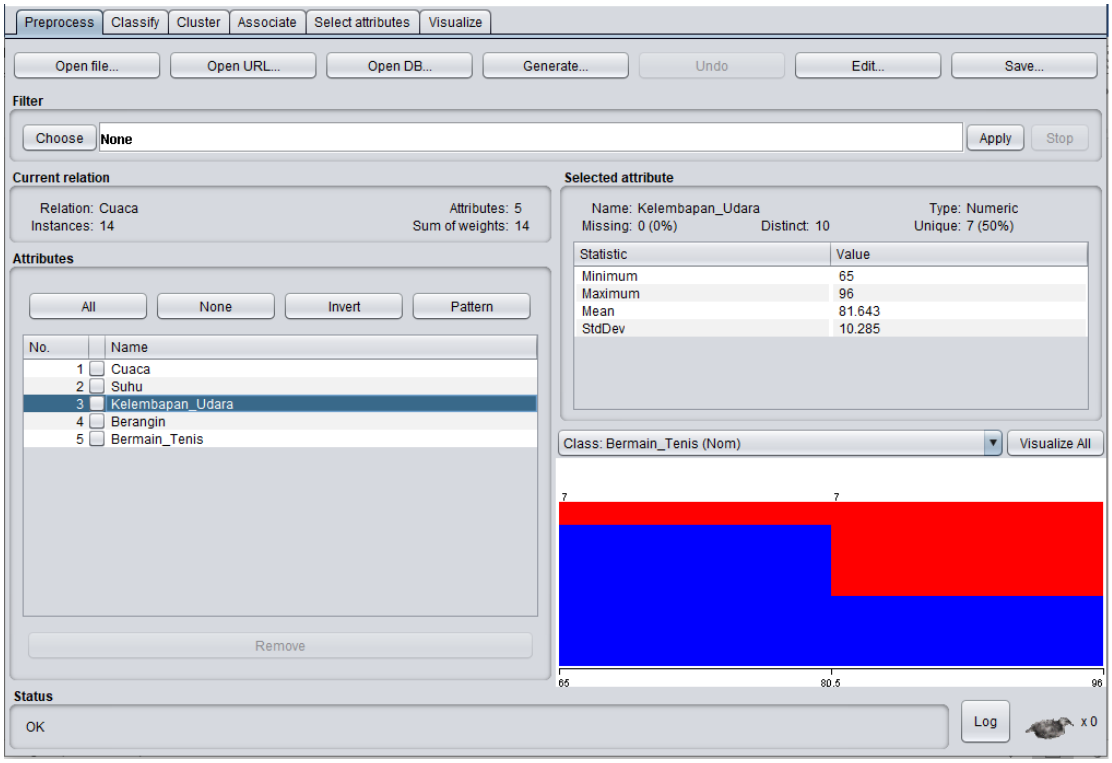
4. Open the Weka application on the Applications menu, select "Explorer", then open the Cuaca.arff file created earlier using the open file. Weather data looks like in the picture below.



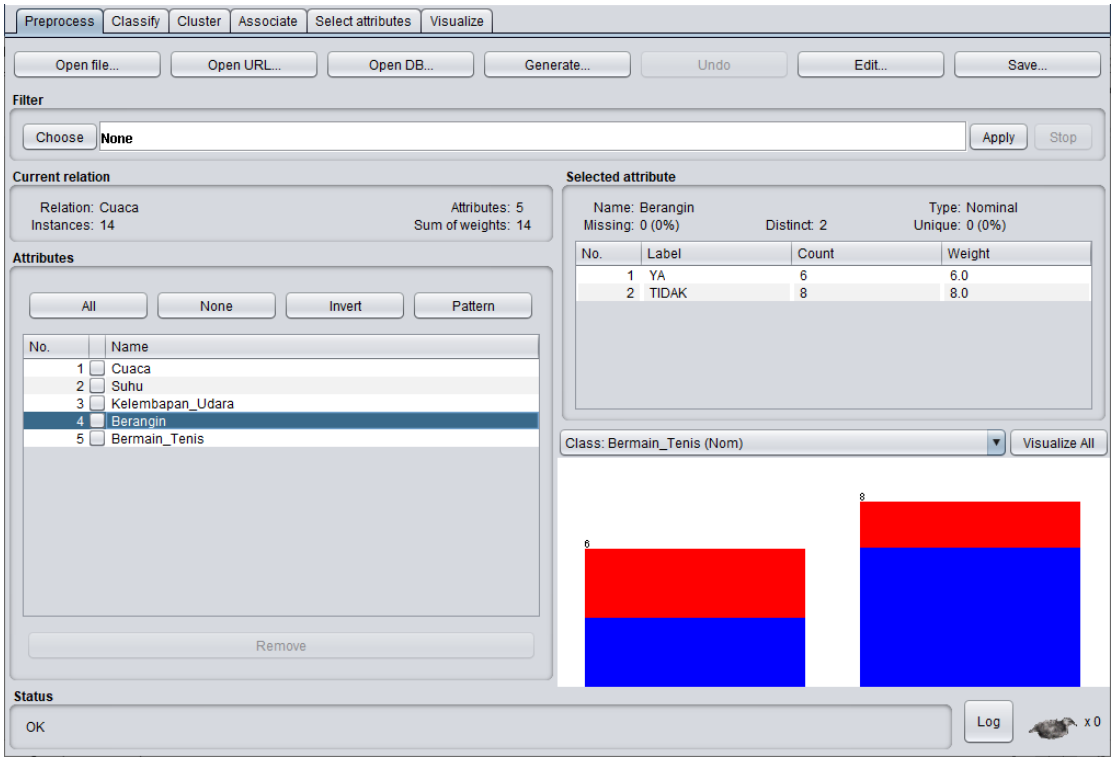
5. Temperature data



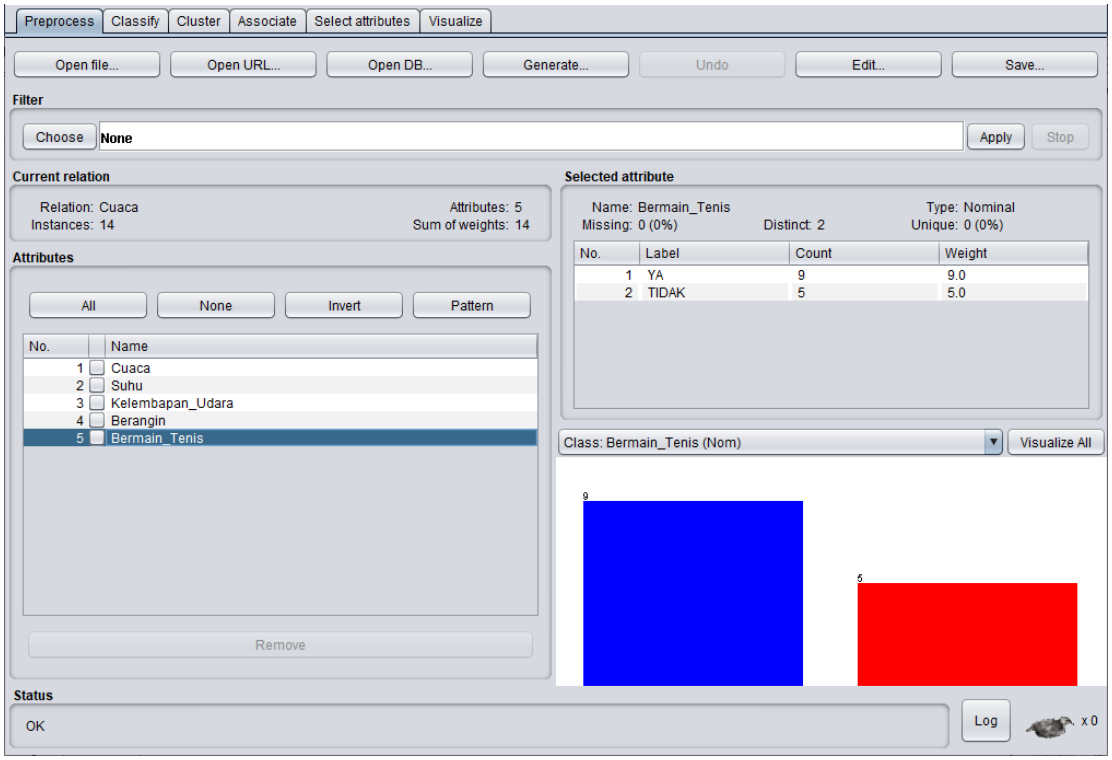
6. Humidity data



7. Wind data



8. Playing Tennis data

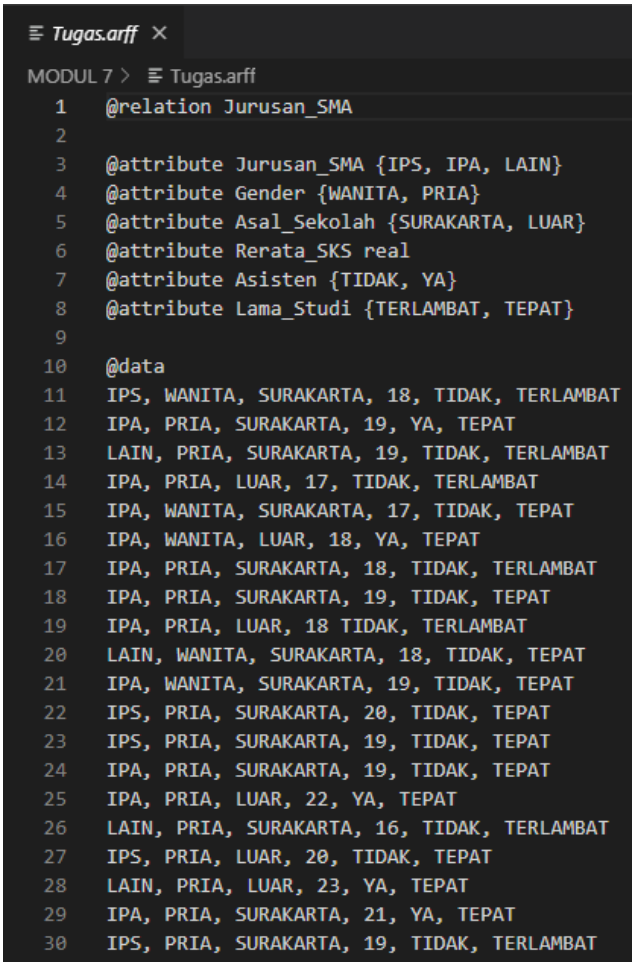


Task

1. Create an ARFF file based on the assignment in Module 6, question number 1.
2. Display the ARFF file and also the graphic image for each data displayed in Weka based on the ARFF file.
3. How many binomial and polynomial attributes?
4. How many attributes are of the real type?
5. In the rerata_SKS attribute, what is the value of the Maximum, Mean, and StdDev (Standard Deviation)?

Answers

1. File arff in module 6.

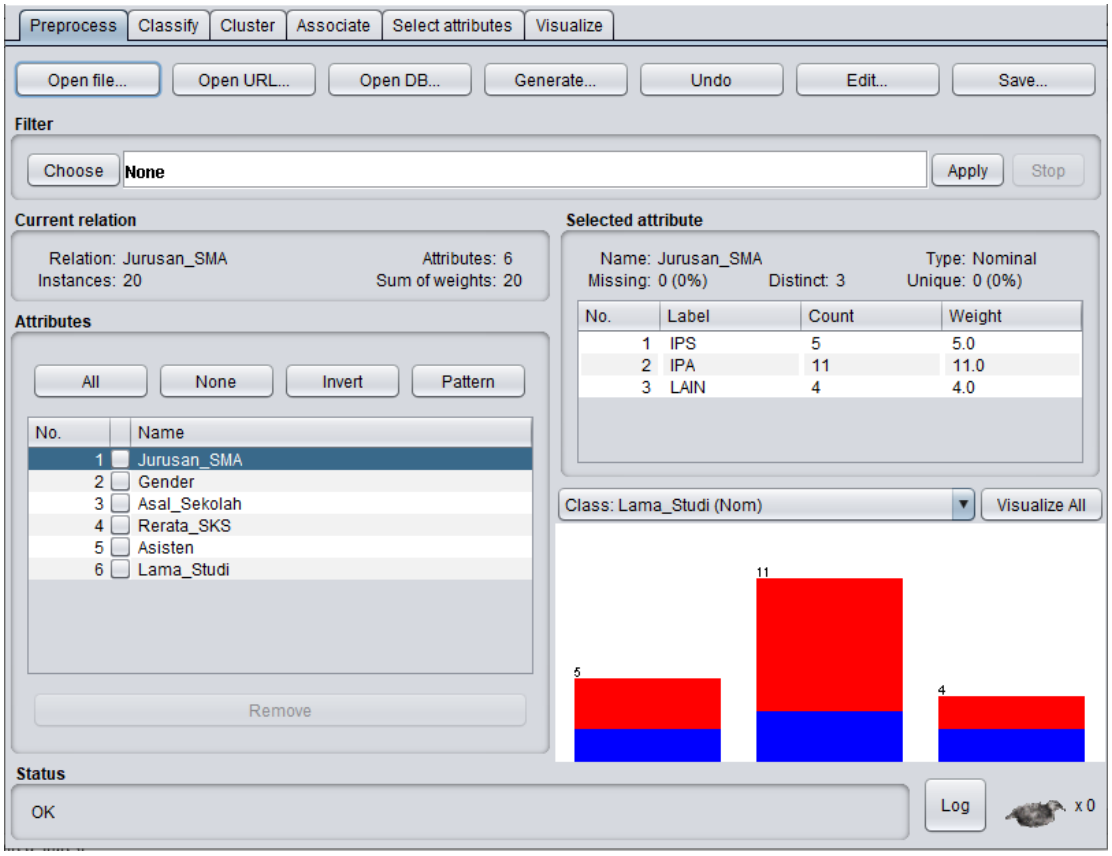


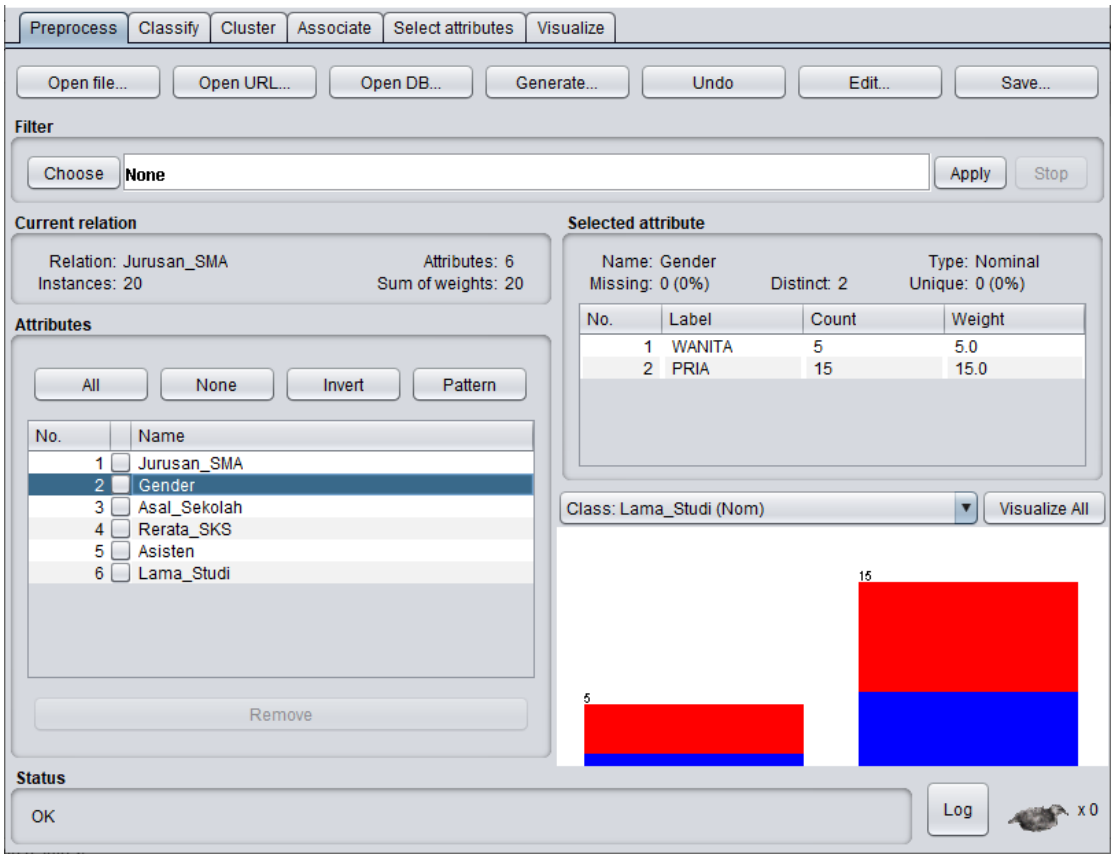
Tugas.arff

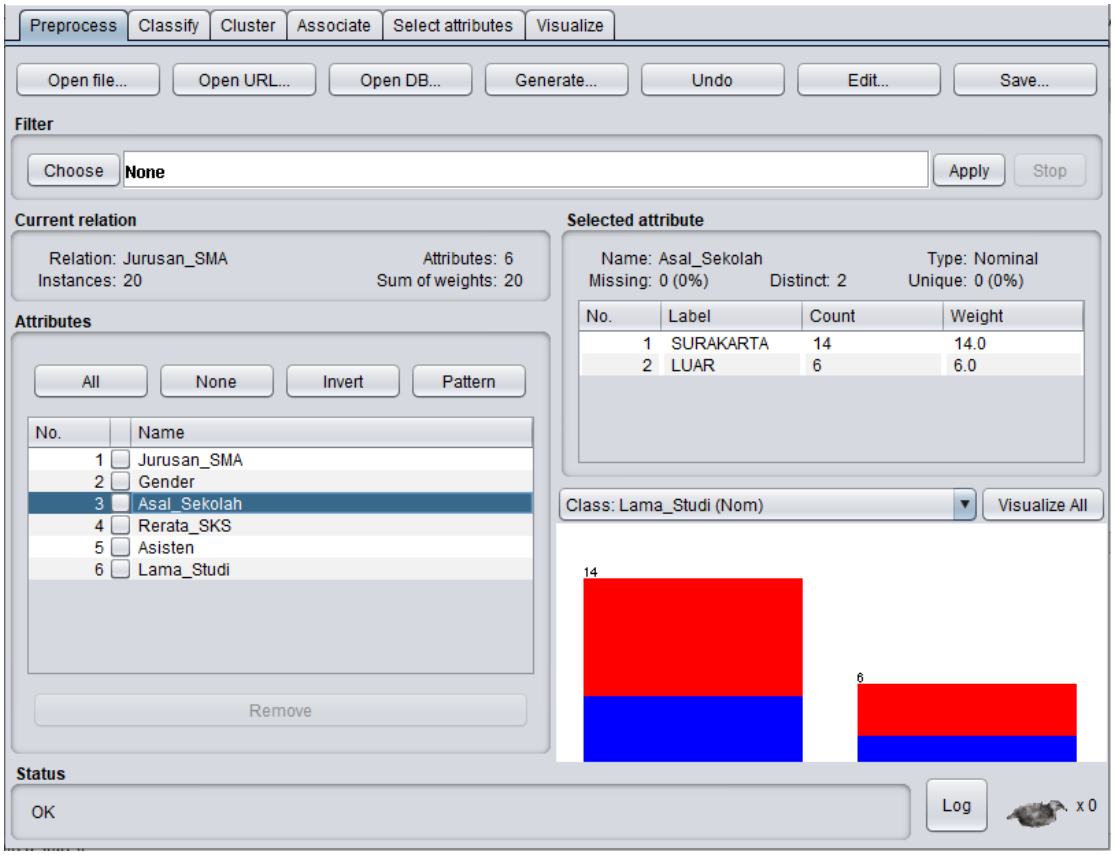
MODUL 7 > Tugas.arff

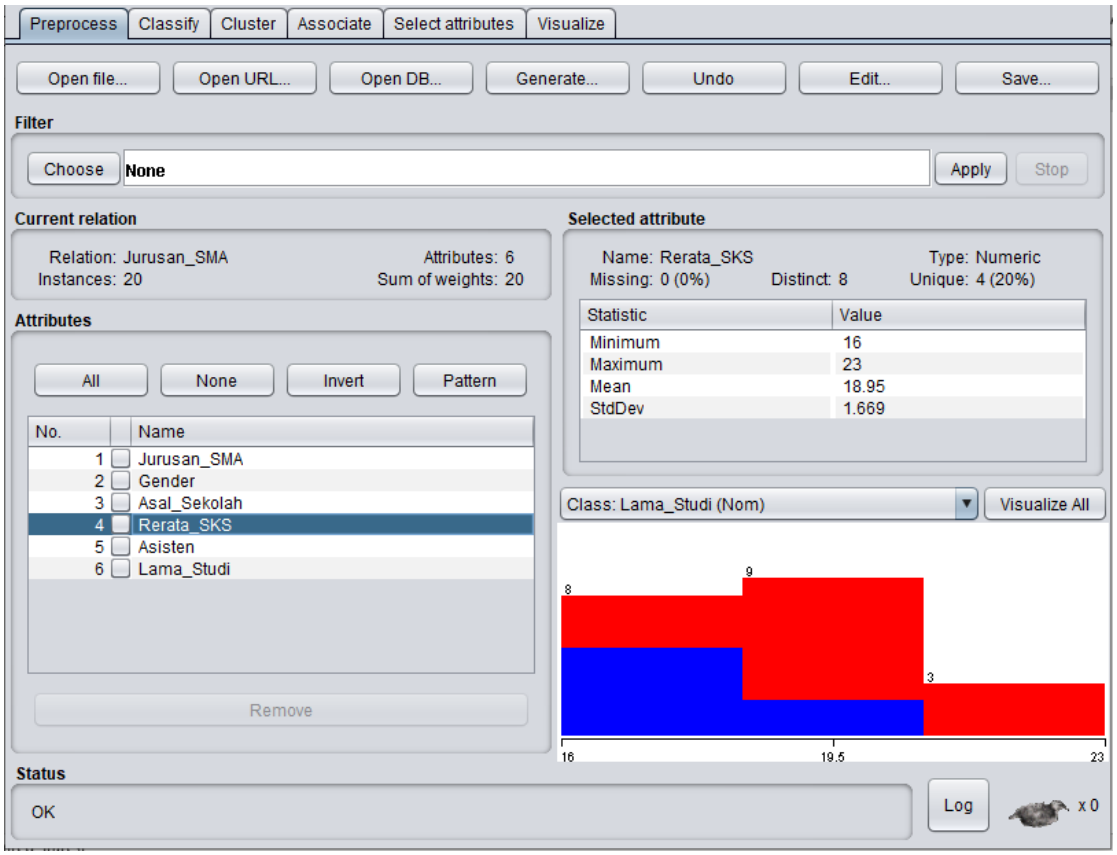
```
@relation Jurusan_SMA

@attribute Jurusan_SMA {IPS, IPA, LAIN}
@attribute Gender {WANITA, PRIA}
@attribute Asal_Sekolah {SURAKARTA, LUAR}
@attribute Rerata_SKS real
@attribute Asisten {TIDAK, YA}
@attribute Lama_Studi {TERLAMBAT, TEPAT}

@data
IPS, WANITA, SURAKARTA, 18, TIDAK, TERLAMBAT
IPA, PRIA, SURAKARTA, 19, YA, TEPAT
LAIN, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
IPA, PRIA, LUAR, 17, TIDAK, TERLAMBAT
IPA, WANITA, SURAKARTA, 17, TIDAK, TEPAT
IPA, WANITA, LUAR, 18, YA, TEPAT
IPA, PRIA, SURAKARTA, 18, TIDAK, TERLAMBAT
IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, LUAR, 18 TIDAK, TERLAMBAT
LAIN, WANITA, SURAKARTA, 18, TIDAK, TEPAT
IPA, WANITA, SURAKARTA, 19, TIDAK, TEPAT
IPS, PRIA, SURAKARTA, 20, TIDAK, TEPAT
IPS, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, SURAKARTA, 19, TIDAK, TEPAT
IPA, PRIA, LUAR, 22, YA, TEPAT
LAIN, PRIA, SURAKARTA, 16, TIDAK, TERLAMBAT
IPS, PRIA, LUAR, 20, TIDAK, TEPAT
LAIN, PRIA, LUAR, 23, YA, TEPAT
IPA, PRIA, SURAKARTA, 21, YA, TEPAT
IPS, PRIA, SURAKARTA, 19, TIDAK, TERLAMBAT
```

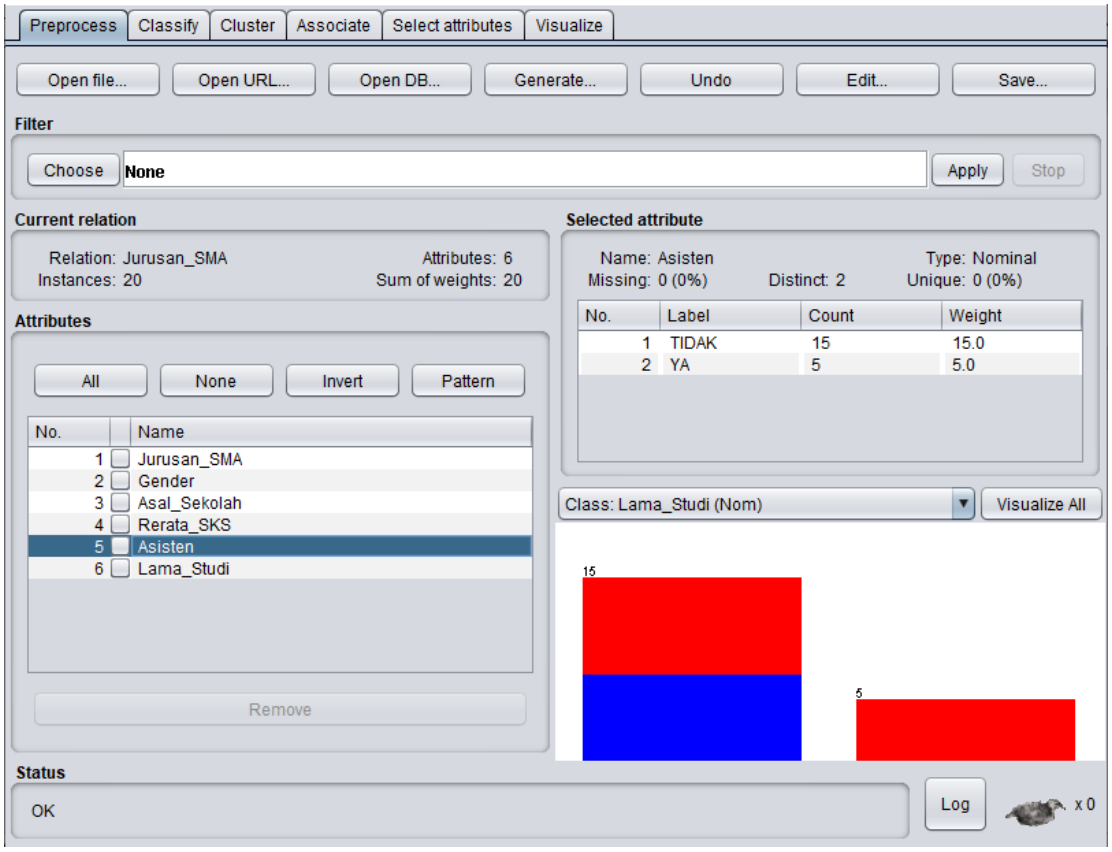
2. The result data from module 6 arff.

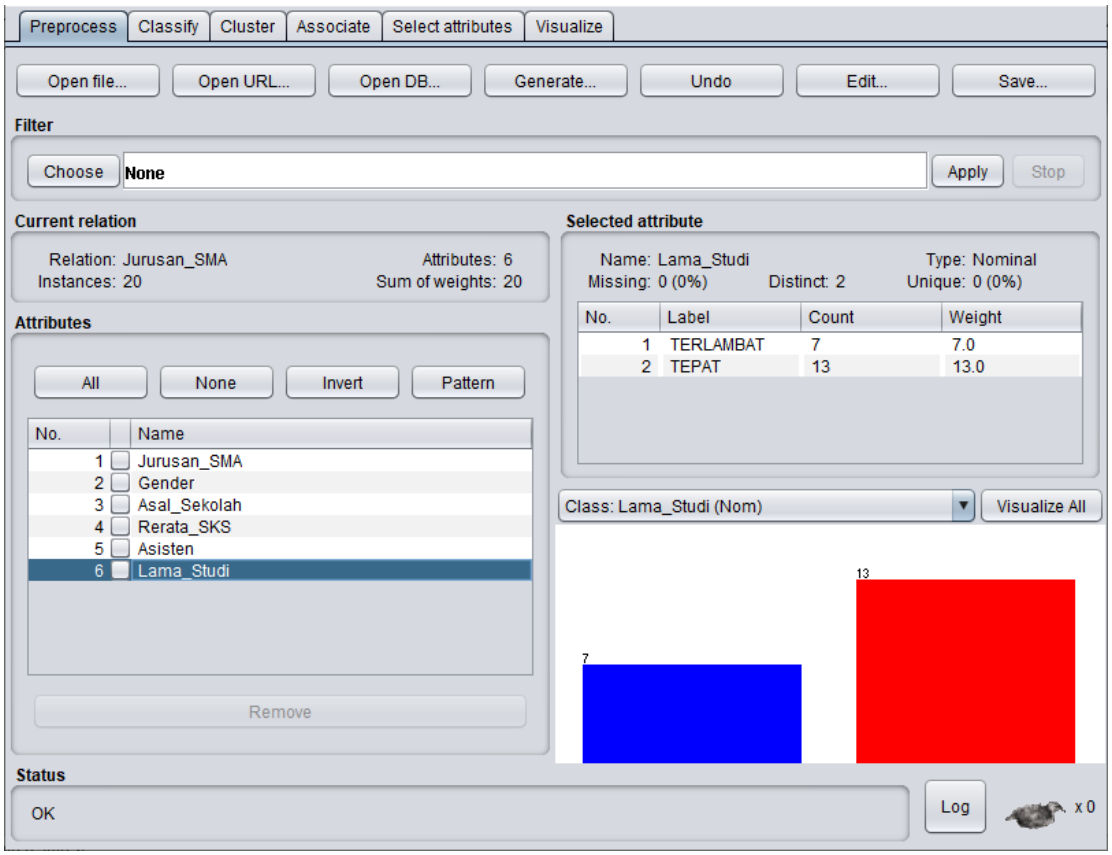




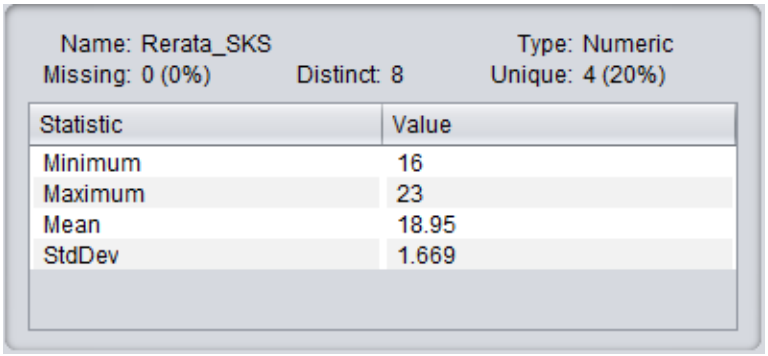

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: Jurusan_SMA
Instances: 20
Attributes: 6
Sum of weights: 20
Selected attribute
Name: Asal_Sekolah
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 SURAKARTA 14 14.0
2 LUAR 6 6.0
Attributes
All
None
Invert
Pattern
No. Name
1 Jurusan_SMA
2 Gender
3 Asal_Sekolah
4 Rerata_SKS
5 Asisten
6 Lama_Studi
Remove
Class: Lama_Studi (Nom)
Visualize All
14
6
Status
OK
Log
x 0


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: Jurusan_SMA
Instances: 20
Attributes: 6
Sum of weights: 20
Selected attribute
Name: Rerata_SKS
Missing: 0 (0%)
Distinct: 8
Type: Numeric
Unique: 4 (20%)
Statistic Value
Minimum 16
Maximum 23
Mean 18.95
StdDev 1.669
Attributes
All
None
Invert
Pattern
No. Name
1 Jurusan_SMA
2 Gender
3 Asal_Sekolah
4 Rerata_SKS
5 Asisten
6 Lama_Studi
Remove
Class: Lama_Studi (Nom)
Visualize All
8
9
3
16
19.5
23
Status
OK
Log
x 0


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: Jurusan_SMA
Instances: 20
Attributes: 6
Sum of weights: 20
Attributes
All
None
Invert
Pattern
No.
Name
1 Jurusan_SMA
2 Gender
3 Asal_Sekolah
4 Rerata_SKS
5 Asisten
6 Lama_Studi
Remove
Selected attribute
Name: Asisten
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No.
Label
Count
Weight
1 TIDAK 15 15.0
2 YA 5 5.0
Class: Lama_Studi (Nom)
Visualize All
15
5
Status
OK
Log
x 0


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: Jurusan_SMA
Instances: 20
Attributes: 6
Sum of weights: 20
Attributes
All
None
Invert
Pattern
No.
Name
1 Jurusan_SMA
2 Gender
3 Asal_Sekolah
4 Rerata_SKS
5 Asisten
6 Lama_Studi
Remove
Selected attribute
Name: Lama_Studi
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No.
Label
Count
Weight
1 TERLAMBAT 7 7.0
2 TEPAT 13 13.0
Class: Lama_Studi (Nom)
Visualize All
13
7
Status
OK
Log
x 0

3. There are 4 binomial attributes, namely gender, asal_sekolah, asisten and lama_studi. While there are 1 polynomial attributes namely jurusan_sma.
4. Attributes that are of the real type in the data are only the average credits, because only those attributes have a lot of data values.
5. The values of the maximum, minimum, mean and standard deviation are.



Name: Rerata_SKS Type: Numeric
Missing: 0 (0%) Distinct: 8 Unique: 4 (20%)

Statistic	Value
Minimum	16
Maximum	23
Mean	18.95
StdDev	1.669