Name: Bhavesh Kewalramani

Roll No.: A-25

Section: A

Batch: A1

Practical – 1

Questions:

1. Press the Explorer button on the main panel and load the **weather dataset** and answer the following questions

1. How many instances are there in the dataset?

Ans: 14

2. State the names of the attributes along with their types and values.

Ans:

S.No.	Name of the Attribute	Type of the Attribute
1.	Outlook	Nominal
2.	Temperature	Nominal
3.	Humidity	Nominal
4.	Windy	Nominal
5.	Play	Nominal

3. What is the class attribute?

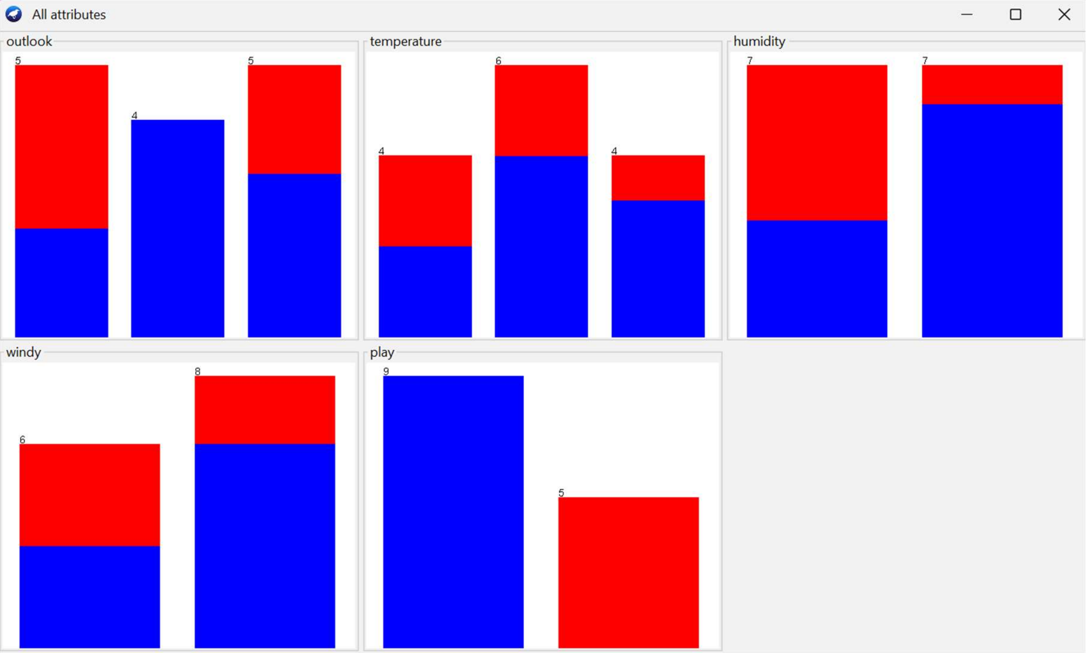
Ans: Play

4. How will you determine how many instances of each class are present in the data

Ans: By selecting the attribute in the left side panel. In the right side panel we can see all the information.

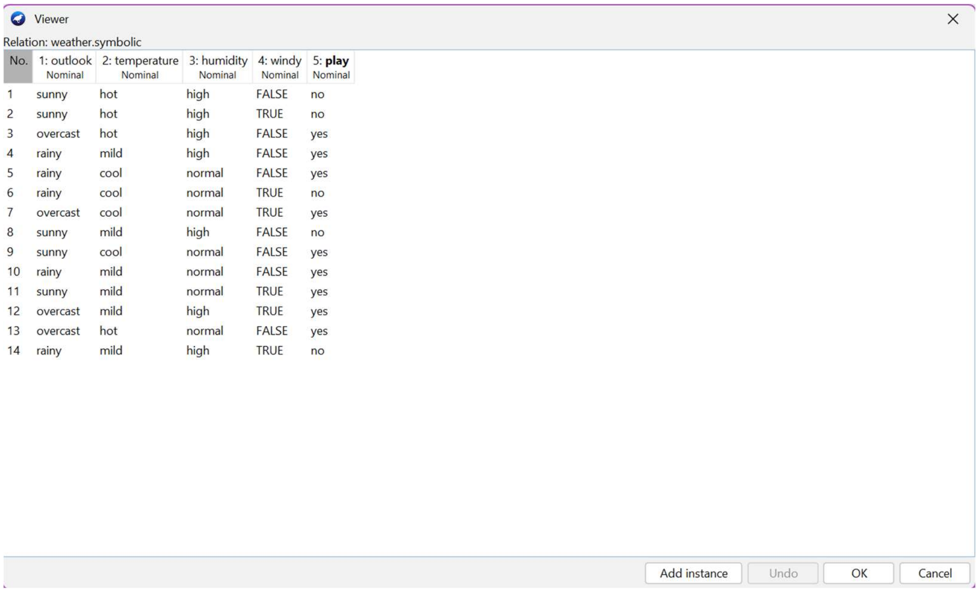
5. What happens with the Visualize All button is pressed?

Ans:



6. How will you view the instances in the dataset? How will you save the changes?

Ans: Using the Edit button in the top

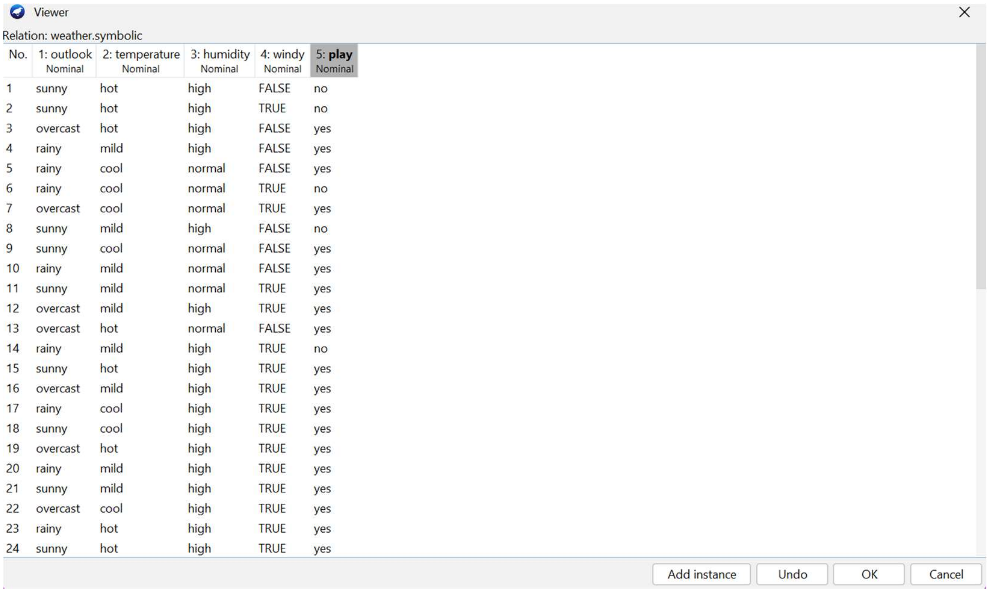


Relation: weather.symbolic

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

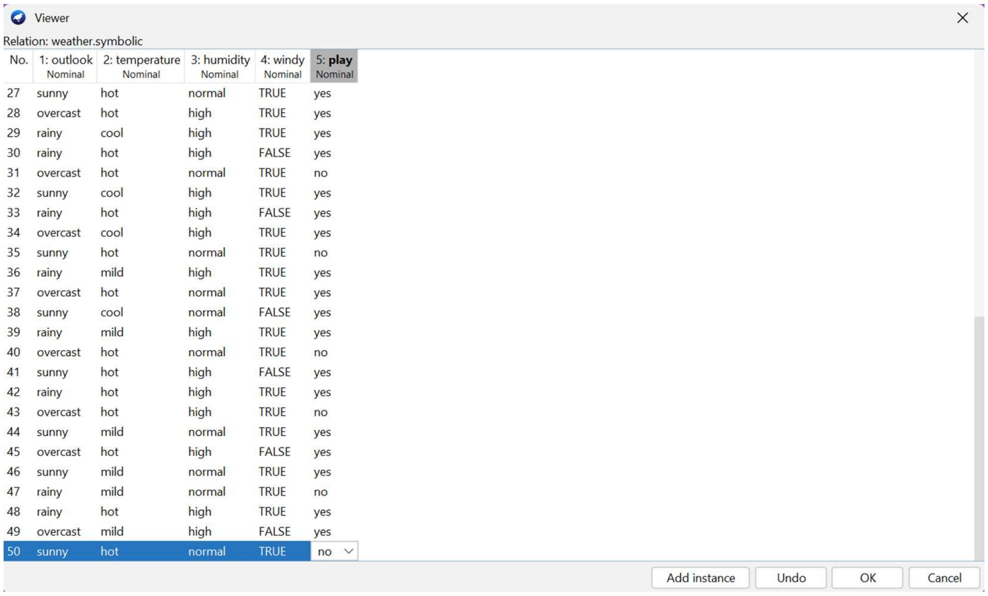
7. Now, extend the dataset to include 50 instances in total.

Ans:



Relation: weather.symbolic

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no
15	sunny	hot	high	TRUE	yes
16	overcast	mild	high	TRUE	yes
17	rainy	cool	high	TRUE	yes
18	sunny	cool	high	TRUE	yes
19	overcast	hot	high	TRUE	yes
20	rainy	mild	high	TRUE	yes
21	sunny	mild	high	TRUE	yes
22	overcast	cool	high	TRUE	yes
23	rainy	hot	high	TRUE	yes
24	sunny	hot	high	TRUE	yes



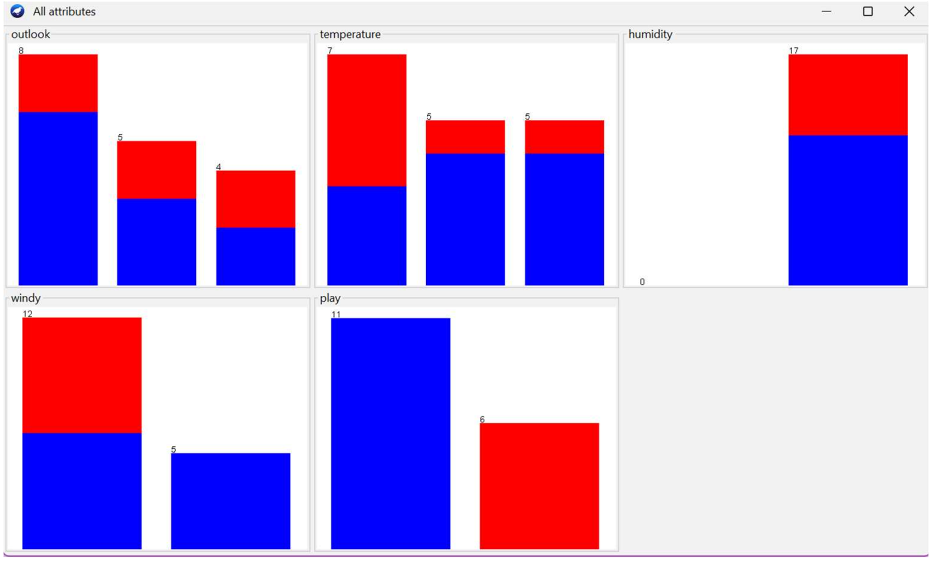
Relation: weather.symbolic

No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal	5: play Nominal
27	sunny	hot	normal	TRUE	yes
28	overcast	hot	high	TRUE	yes
29	rainy	cool	high	TRUE	yes
30	rainy	hot	high	FALSE	yes
31	overcast	hot	normal	TRUE	no
32	sunny	cool	high	TRUE	yes
33	rainy	hot	high	FALSE	yes
34	overcast	cool	high	TRUE	yes
35	sunny	hot	normal	TRUE	no
36	rainy	mild	high	TRUE	yes
37	overcast	hot	normal	TRUE	yes
38	sunny	cool	normal	FALSE	yes
39	rainy	mild	high	TRUE	yes
40	overcast	hot	normal	TRUE	no
41	sunny	hot	high	FALSE	yes
42	rainy	hot	high	TRUE	yes
43	overcast	hot	high	TRUE	no
44	sunny	mild	normal	TRUE	yes
45	overcast	hot	high	FALSE	yes
46	sunny	mild	normal	TRUE	yes
47	rainy	mild	normal	TRUE	no
48	rainy	hot	high	TRUE	yes
49	overcast	mild	high	FALSE	yes
50	sunny	hot	normal	TRUE	no

2. Do as directed to apply Filter

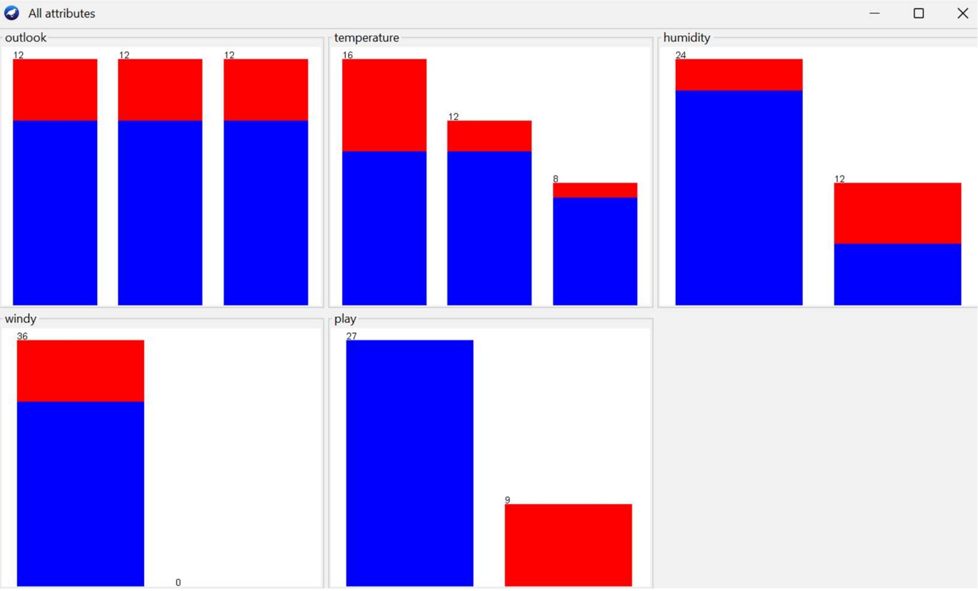
1. Use the unsupervised filter RemoveWithValues to remove all instances where the attribute ‘humidity’ has the value ‘high’? Undo the effect of the filter.

Ans:



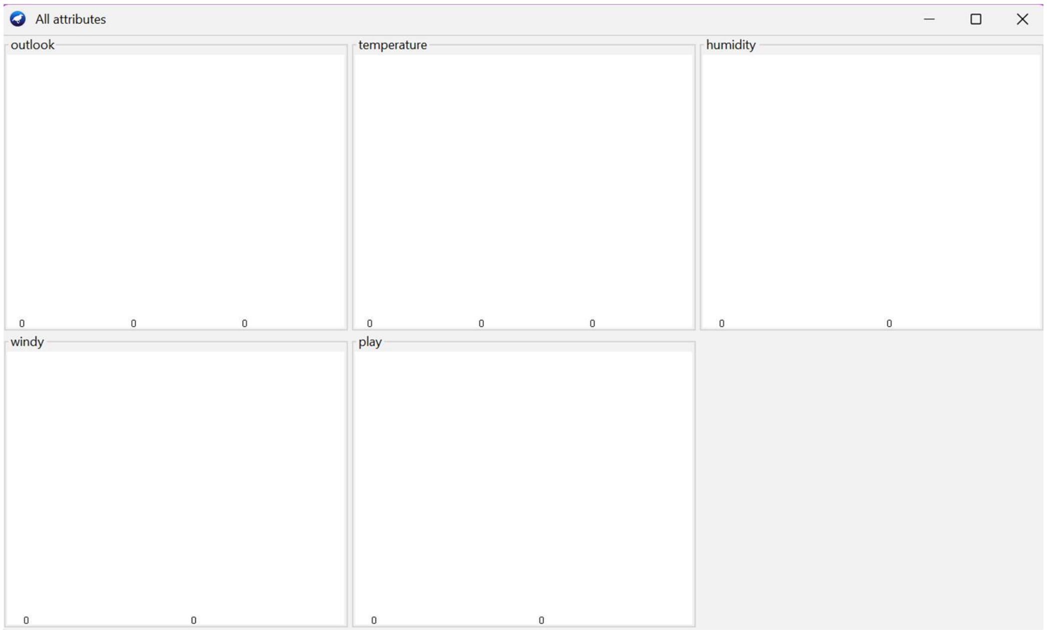
2. Remove the ‘FALSE’ instances of windy attribute and undo the effect.

Ans:



3. Remove the attribute outlook and undo the effect.

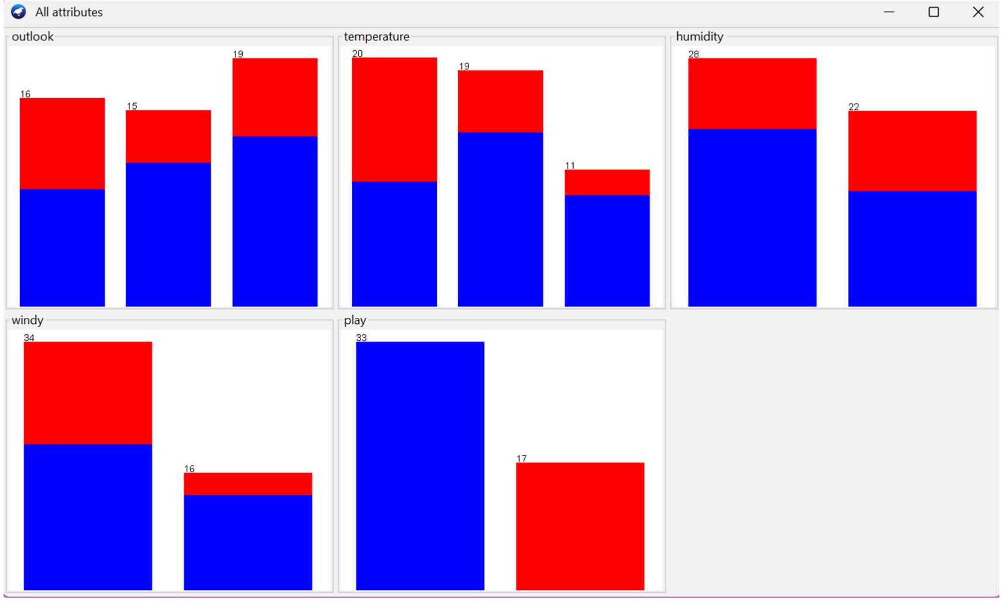
Ans:



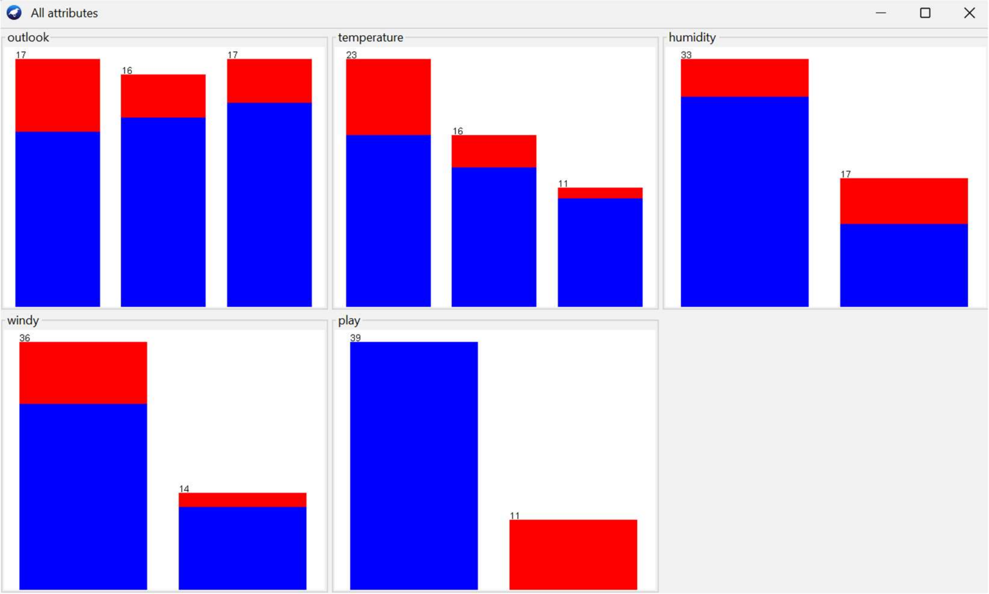
4. Experiment with different filters and report their effects.

Ans:

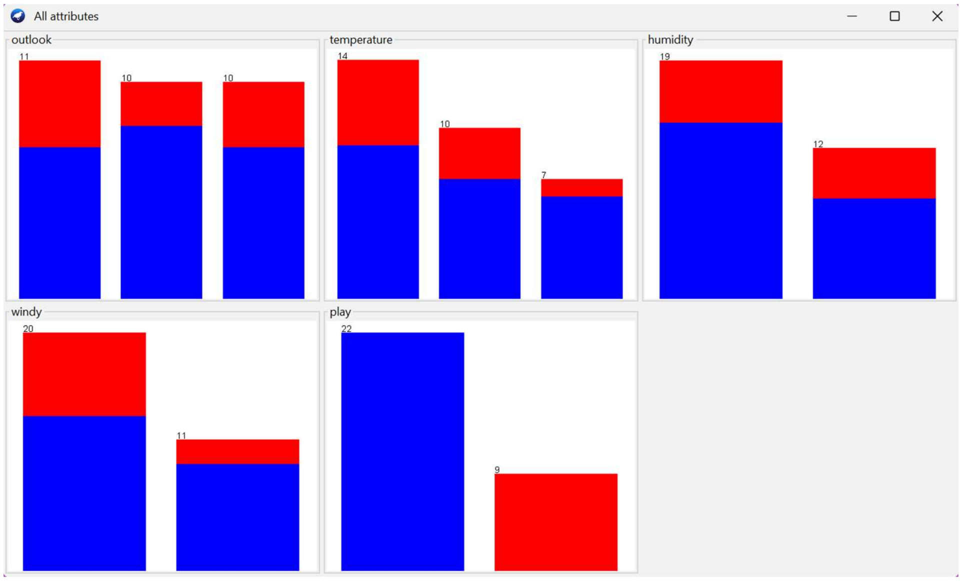
Resample Filter



ReservoirSample Filter



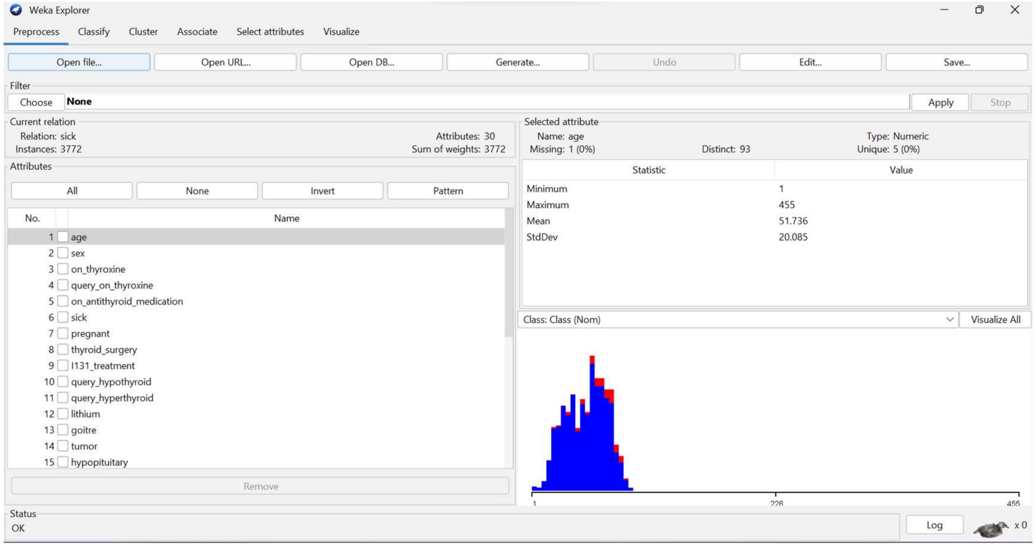
RemoveDuplicates Filter



3. Application of Discretization Filters [use sick.arff dataset]

1. Load the 'sick.arff' dataset.

Ans:



Filter
None

Current relation
Relation: sick
Instances: 3772
Attributes: 30
Sum of weights: 3772

Selected attribute
Name: age
Missing: 1 (0%)
Distinct: 93
Type: Numeric
Unique: 5 (0%)

Statistic	Value
Minimum	1
Maximum	455
Mean	51.736
StdDev	20.085

Attributes

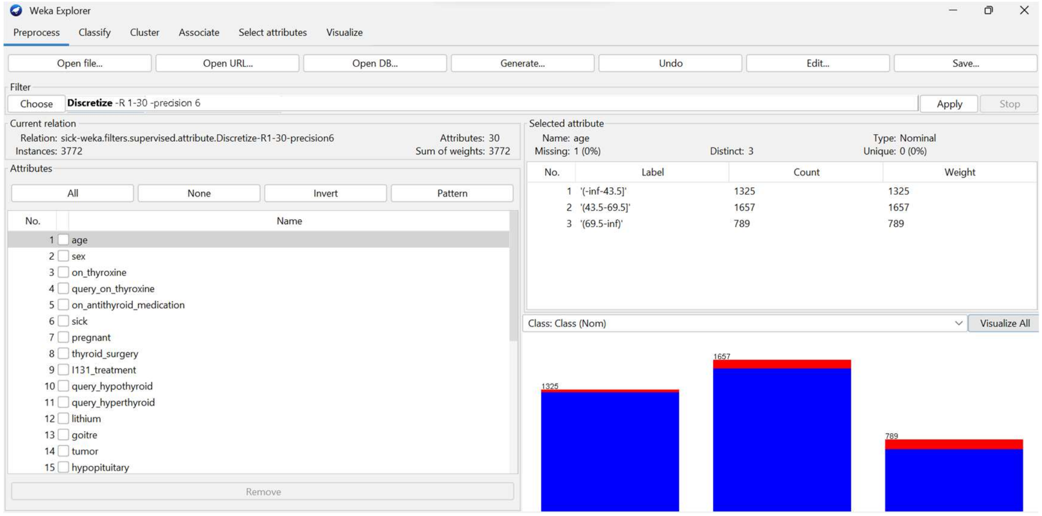
No.	Name
1	age
2	sex
3	on_thyroxine
4	query_on_thyroxine
5	on_antithyroid_medication
6	sick
7	pregnant
8	thyroid_surgery
9	I131_treatment
10	query_hypothyroid
11	query_hyperthyroid
12	lithium
13	goitre
14	tumor
15	hypopituitary

Class: Class (Nom)

Status
OK

2. Apply the supervised discretization filter on different attributes.

Ans:



Filter: **Discretize** -R 1-30 -precision 6

Current relation
Relation: sick-weka.filters.supervised.attribute.Discretize-R1-30-precision6 Attributes: 30
Instances: 3772 Sum of weights: 3772

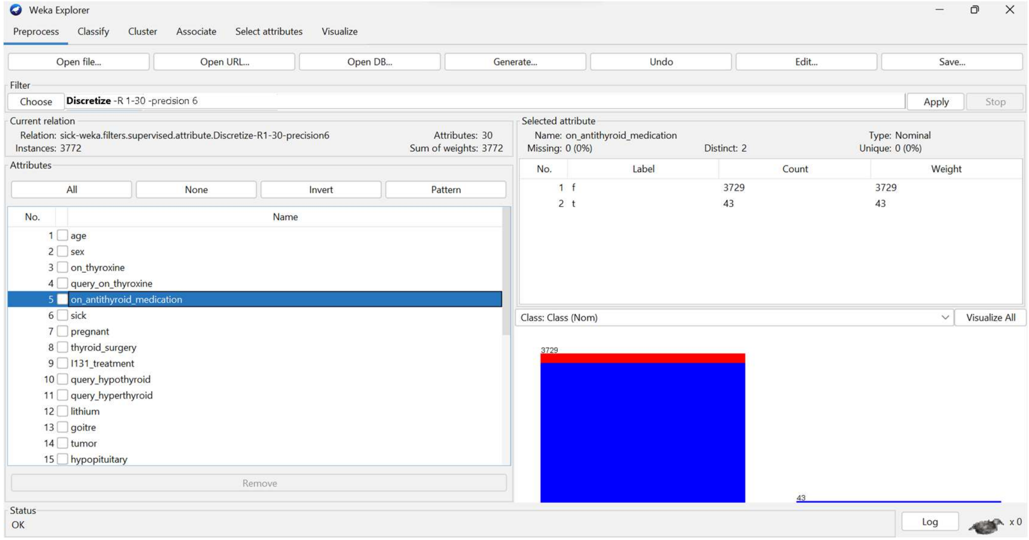
Selected attribute
Name: age Type: Nominal
Missing: 1 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-43.5]'	1325	1325
2	'(43.5-69.5]'	1657	1657
3	'(69.5-inf)'	789	789

Class: Class (Nom)

Attributes

No.	Name
1	□ age
2	□ sex
3	□ on_thyroxine
4	□ query_on_thyroxine
5	□ on_antithyroid_medication
6	□ sick
7	□ pregnant
8	□ thyroid_surgery
9	□ I131_treatment
10	□ query_hypothyroid
11	□ query_hyperthyroid
12	□ lithium
13	□ goitre
14	□ tumor
15	□ hypopituitary



Filter: **Discretize** -R 1-30 -precision 6

Current relation
Relation: sick-weka.filters.supervised.attribute.Discretize-R1-30-precision6 Attributes: 30
Instances: 3772 Sum of weights: 3772

Selected attribute
Name: on_antithyroid_medication Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

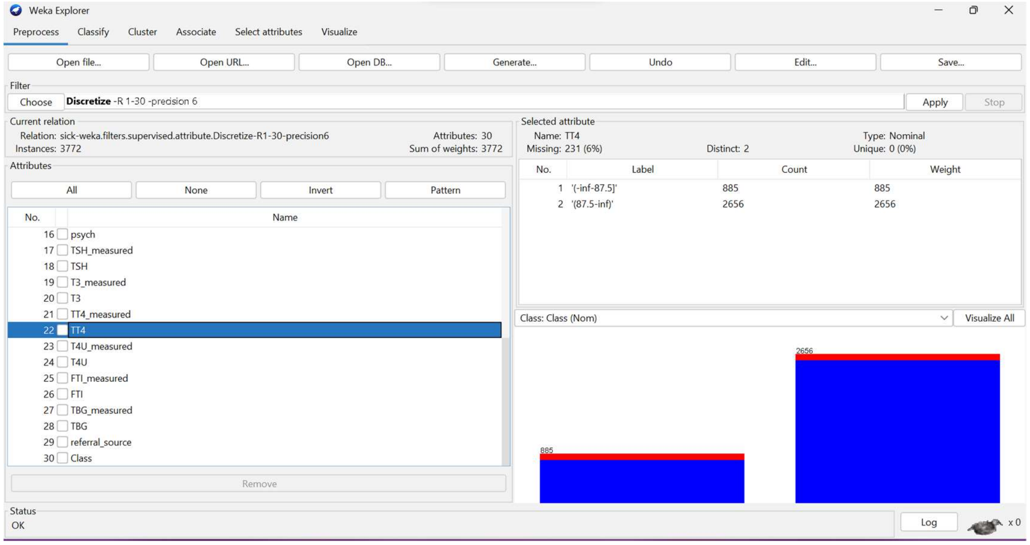
No.	Label	Count	Weight
1	f	3729	3729
2	t	43	43

Class: Class (Nom)

Attributes

No.	Name
1	□ age
2	□ sex
3	□ on_thyroxine
4	□ query_on_thyroxine
5	□ on_antithyroid_medication
6	□ sick
7	□ pregnant
8	□ thyroid_surgery
9	□ I131_treatment
10	□ query_hypothyroid
11	□ query_hyperthyroid
12	□ lithium
13	□ goitre
14	□ tumor
15	□ hypopituitary

Status
OK



Filter: **Discretize** -R 1-30 -precision 6

Current relation
Relation: sick-weka.filters.supervised.attribute.Discretize-R1-30-precision6 Attributes: 30
Instances: 3772 Sum of weights: 3772

Selected attribute
Name: TT4 Type: Nominal
Missing: 231 (6%) Distinct: 2 Unique: 0 (0%)

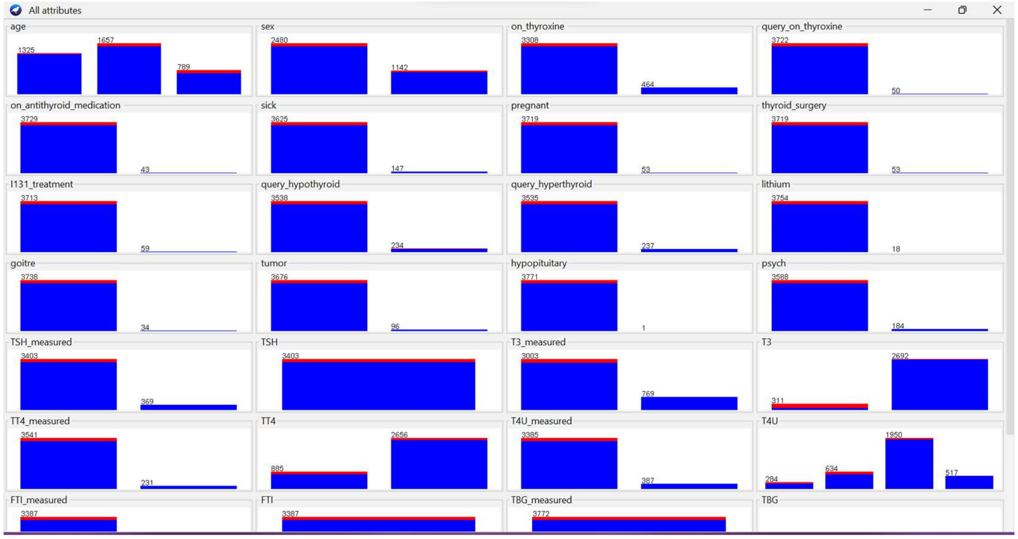
No.	Label	Count	Weight
1	'(-inf-87.5]'	885	885
2	'(87.5-inf)'	2656	2656

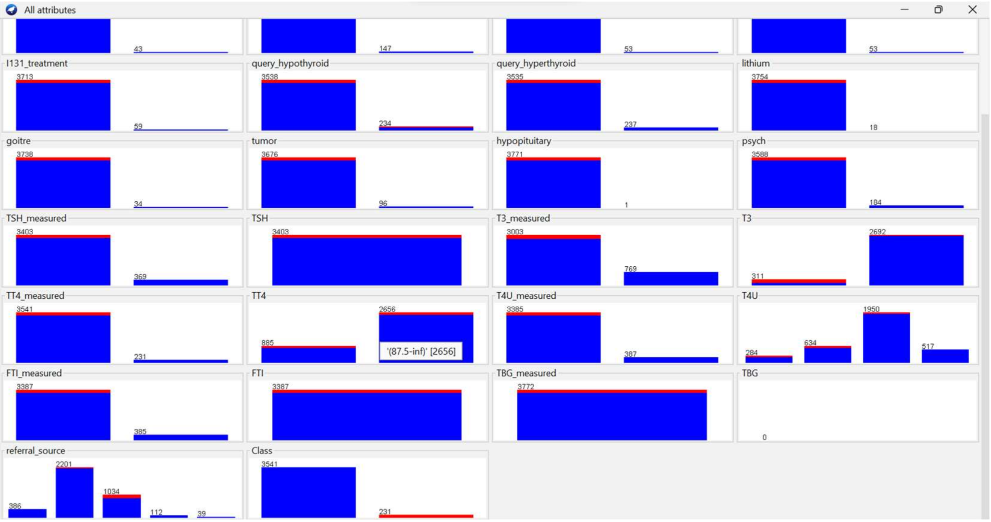
Class: Class (Nom)

Attributes

No.	Name
16	□ psych
17	□ TSH_measured
18	□ TSH
19	□ T3_measured
20	□ T3
21	□ TT4_measured
22	□ TT4
23	□ T4U_measured
24	□ T4U
25	□ FTI_measured
26	□ FTI
27	□ TBG_measured
28	□ TBG
29	□ referral_source
30	□ Class

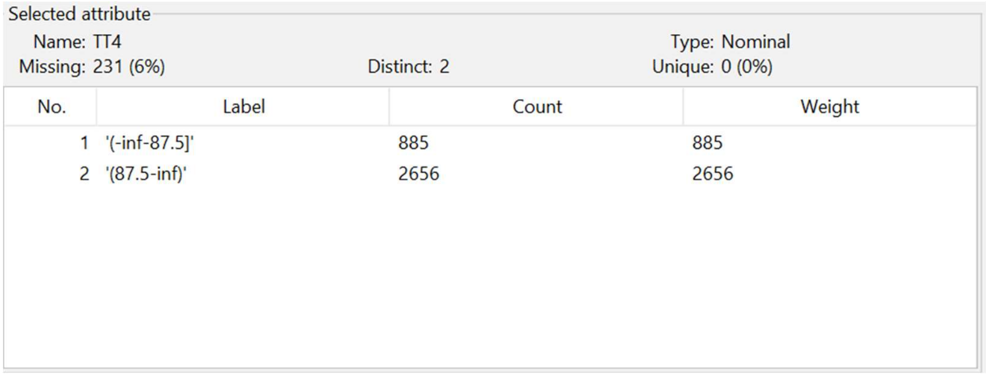
Status
OK





3. What is the effect of this filter on the attributes?

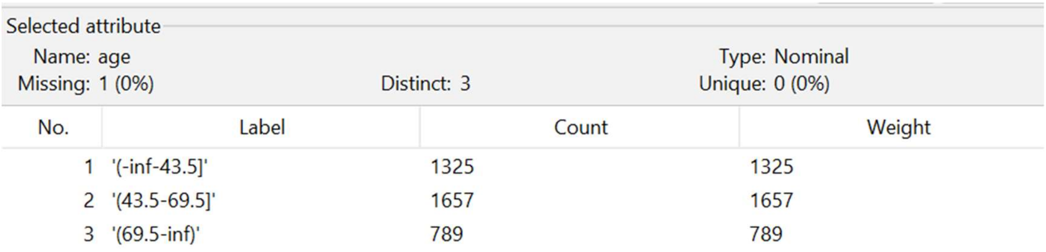
Ans: The discrete class intervals are formed and the frequency is calculated.



Selected attribute

Name: TT4 Type: Nominal
Missing: 231 (6%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-87.5]'	885	885
2	'(87.5-inf)'	2656	2656



Selected attribute

Name: age Type: Nominal
Missing: 1 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-43.5]'	1325	1325
2	'(43.5-69.5]'	1657	1657
3	'(69.5-inf)'	789	789

4. How many distinct ranges have been created for each attribute?

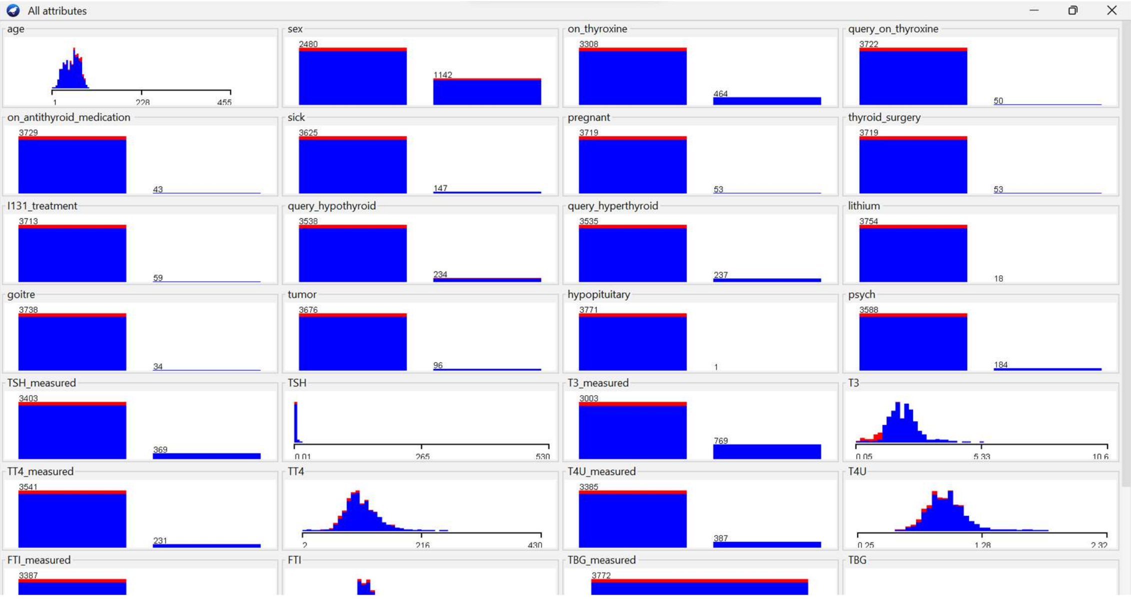
Ans:

age	3
sex	2
on_thyroxine	2
query_on_thyroxine	2
on_antithyroid_medication	2
sick	2
pregnant	2
thyroid_surgery	2
I131_treatment	2
query_hypothyroid	2

query_hyperthyroid	2
Lithium	2
Goitre	2
tumor	2
hypopituitary	2
psych	2
TSH_measured	2
TSH	1
T3_measured	2
T3	1
TT4_measured	2
TT4	1
T4U_measured	2
T4U	4
FTI_measured	2
FTI	1
TBG_measured	1
TBG	1
referral_source	5
Class	2

5. Undo the filter applied in the previous step.

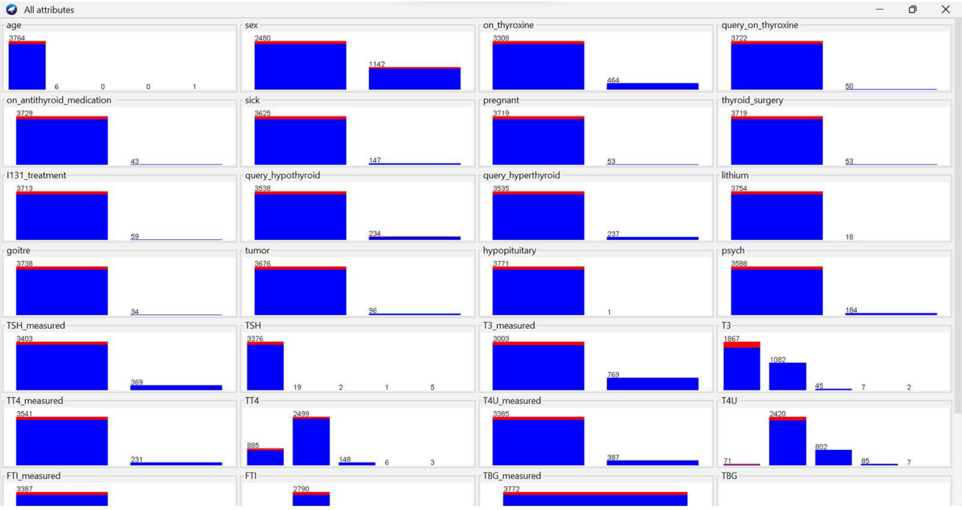
Ans:

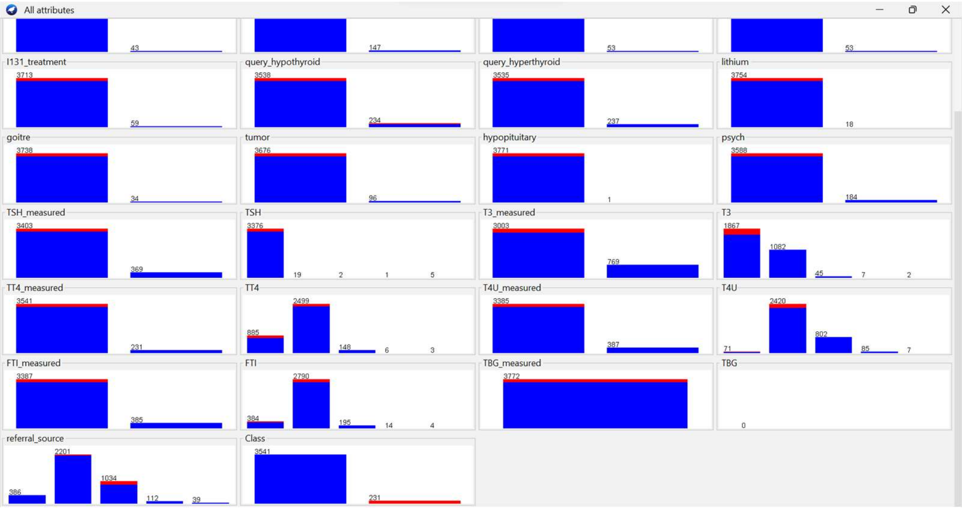


6. Apply the unsupervised discretization filter. Do this twice:

1. In this step, set 'bins'=5

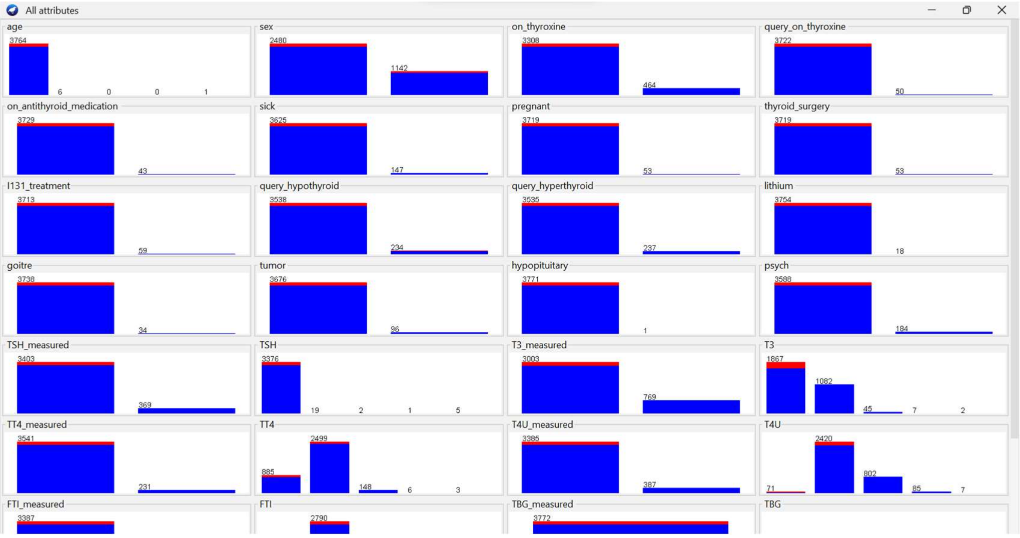
Ans:

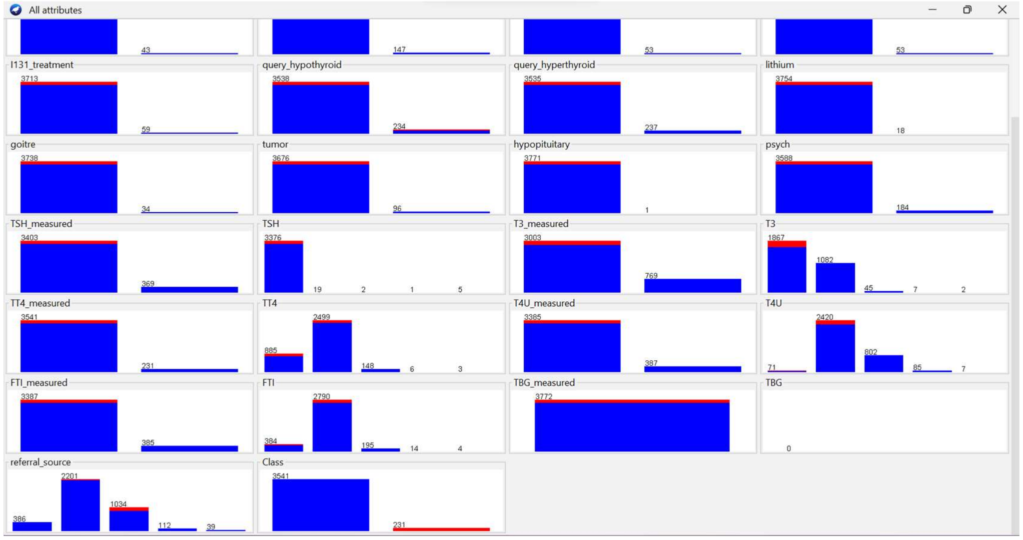




2. In this step, set 'bins'=10

Ans:





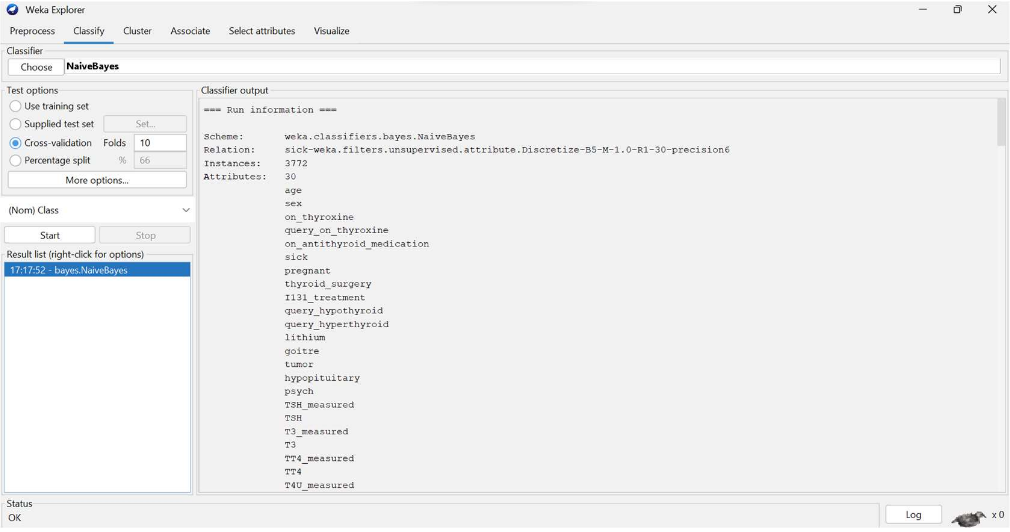
3. What is the effect of the unsupervised filter on the dataset?

Ans: Unsupervised filter work without taking any class distributions into account. The unsupervised *discretize* filter only considers the attribute being discretized. While it can 'optimize' the number of bins, it does so only with respect to self-encoding.

7. Run the Naive Bayes classifier after apply the following filters

1. Unsupervised discretized with 'bins'=5

Ans:



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10

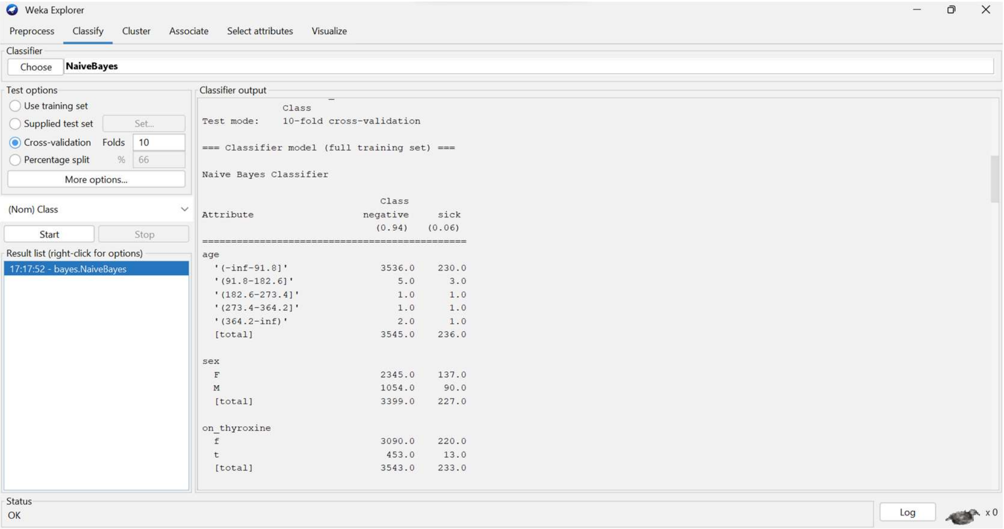
(Nom) Class

17:17:52 - bayes.NaiveBayes

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     sick-weka.filters.unsupervised.attribute.Discretize-B5-M-1.0-R1-30-precision6
Instances:    3772
Attributes:   30
              age
              sex
              on_thyroxine
              query_on_thyroxine
              on_antithyroid_medication
              sick
              pregnant
              thyroid_surgery
              I131_treatment
              query_hypothyroid
              query_hyperthyroid
              lithium
              goitre
              tumor
              hypopituitary
              psych
              TSH_measured
              TSH
              T3_measured
              T3
              TT4_measured
              TT4
              T4U_measured
```

Status: OK



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10

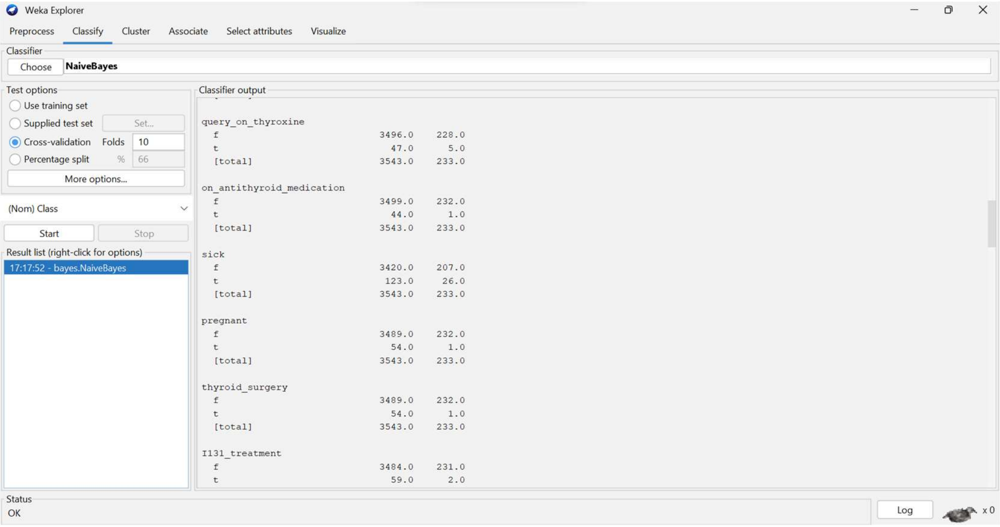
(Nom) Class

17:17:52 - bayes.NaiveBayes

Classifier output

```
              Class
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

                               Class
Attribute                   negative     sick
                              (0.94)   (0.06)
==============================================
age
  '(-inf-91.8]'               3536.0    230.0
  '(91.8-182.6]'                 5.0      3.0
  '(182.6-273.4]'                1.0      1.0
  '(273.4-364.2]'                1.0      1.0
  '(364.2-inf)'                  2.0      1.0
  [total]                     3545.0    236.0

sex
  F                           2345.0    137.0
  M                           1054.0     90.0
  [total]                     3399.0    227.0

on_thyroxine
  f                           3090.0    220.0
  t                            453.0     13.0
  [total]                     3543.0    233.0
```

Status: OK



Classifier: NaiveBayes

Test options

- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) Class

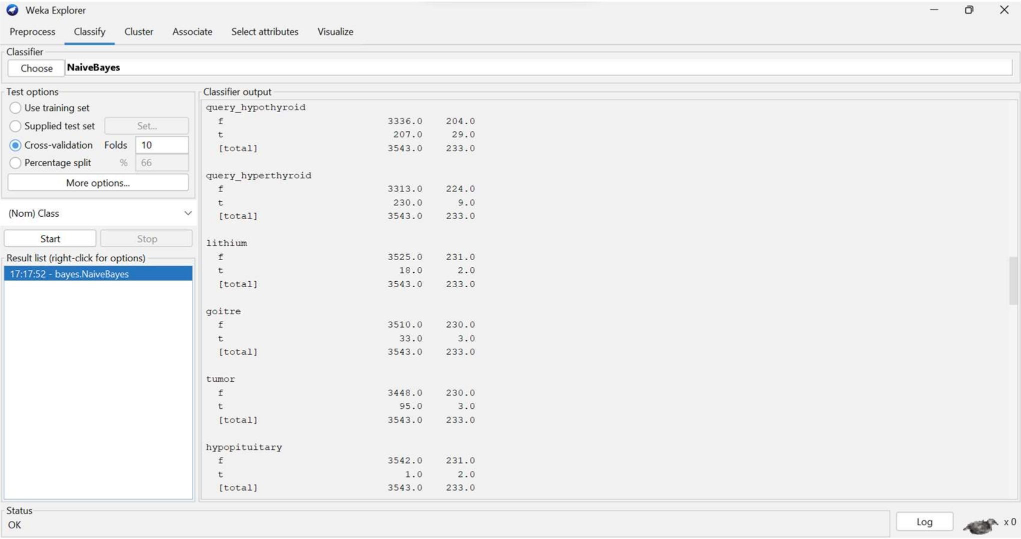
Result list (right-click for options)

17:17:52 - bayes.NaiveBayes

Classifier output

```
query_on_thyroxine
  f                             3496.0    228.0
  t                               47.0      5.0
  [total]                       3543.0    233.0

on_antithyroid_medication
  f                             3499.0    232.0
  t                               44.0      1.0
  [total]                       3543.0    233.0

sick
  f                             3420.0    207.0
  t                              123.0     26.0
  [total]                       3543.0    233.0

pregnant
  f                             3489.0    232.0
  t                               54.0      1.0
  [total]                       3543.0    233.0

thyroid_surgery
  f                             3489.0    232.0
  t                               54.0      1.0
  [total]                       3543.0    233.0

I131_treatment
  f                             3484.0    231.0
  t                               59.0      2.0
```

Status
OK



Classifier: NaiveBayes

Test options

- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) Class

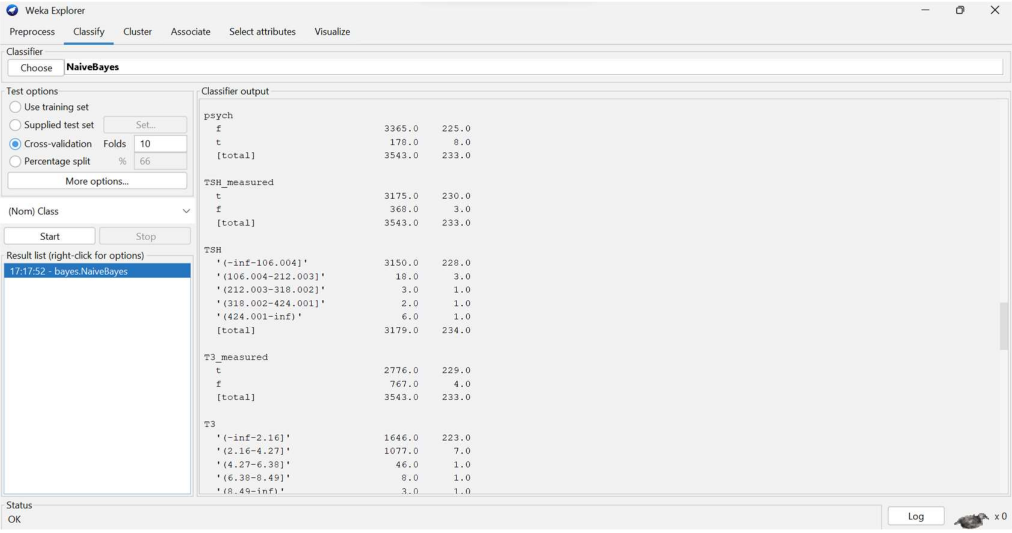
Result list (right-click for options)

17:17:52 - bayes.NaiveBayes

Classifier output

```
query_hypothyroid
  f                             3336.0    204.0
  t                              207.0     29.0
  [total]                       3543.0    233.0

query_hyperthyroid
  f                             3313.0    224.0
  t                              230.0      9.0
  [total]                       3543.0    233.0

lithium
  f                             3525.0    231.0
  t                               18.0      2.0
  [total]                       3543.0    233.0

goitre
  f                             3510.0    230.0
  t                               33.0      3.0
  [total]                       3543.0    233.0

tumor
  f                             3448.0    230.0
  t                               95.0      3.0
  [total]                       3543.0    233.0

hypopituitary
  f                             3542.0    231.0
  t                                1.0      2.0
  [total]                       3543.0    233.0
```

Status
OK



Classifier: NaiveBayes

Test options

- ○ Use training set
- ○ Supplied test set
- ● Cross-validation Folds 10
- ○ Percentage split % 66

(Nom) Class

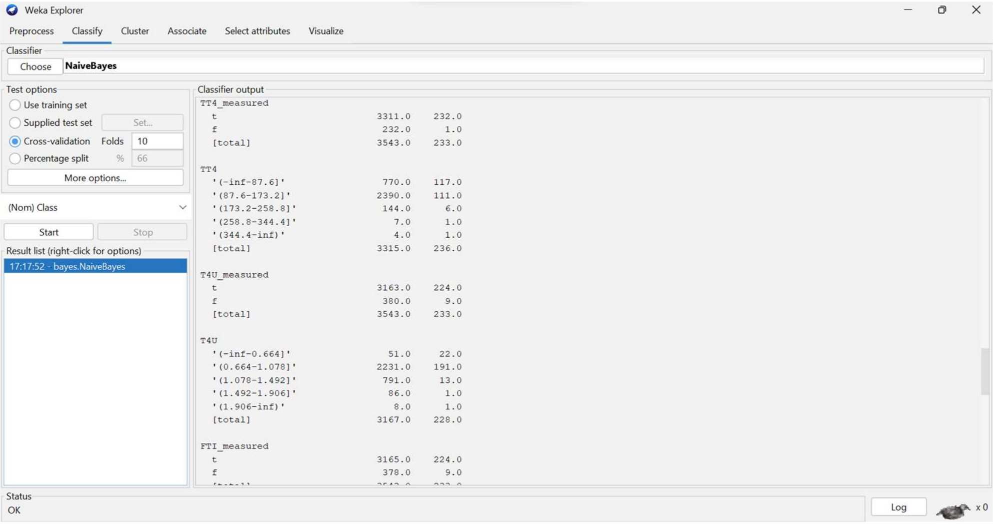
Result list (right-click for options)

17:17:52 - bayes.NaiveBayes

Classifier output

```
psych
  f                             3365.0    225.0
  t                              178.0      8.0
  [total]                       3543.0    233.0

TSH_measured
  t                             3175.0    230.0
  f                              368.0      3.0
  [total]                       3543.0    233.0

TSH
  '(-inf-106.004]'              3150.0    228.0
  '(106.004-212.003]'             18.0      3.0
  '(212.003-318.002]'              3.0      1.0
  '(318.002-424.001]'              2.0      1.0
  '(424.001-inf)'                  6.0      1.0
  [total]                       3179.0    234.0

T3_measured
  t                             2776.0    229.0
  f                              767.0      4.0
  [total]                       3543.0    233.0

T3
  '(-inf-2.16]'                 1646.0    223.0
  '(2.16-4.27]'                 1077.0      7.0
  '(4.27-6.38]'                   46.0      1.0
  '(6.38-8.49]'                    8.0      1.0
  '(8.49-inf)'                     3.0      1.0
```

Status
OK



Classifier: NaiveBayes

Test options: ○ Use training set ○ Supplied test set ● Cross-validation Folds 10 ○ Percentage split % 66

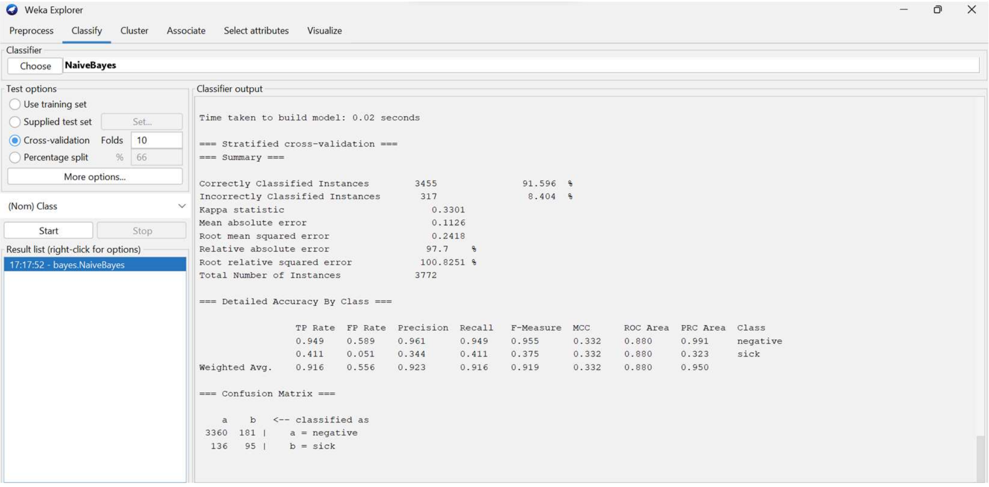
(Nom) Class

17:17:52 - bayes.NaiveBayes

Status: OK

Classifier output

```
TT4_measured
  t                              3311.0    232.0
  f                               232.0      1.0
  [total]                        3543.0    233.0

TT4
  '(-inf-87.6]'                   770.0    117.0
  '(87.6-173.2]'                 2390.0    111.0
  '(173.2-258.8]'                 144.0      6.0
  '(258.8-344.4]'                   7.0      1.0
  '(344.4-inf)'                     4.0      1.0
  [total]                        3315.0    236.0

T4U_measured
  t                              3163.0    224.0
  f                               380.0      9.0
  [total]                        3543.0    233.0

T4U
  '(-inf-0.664]'                   51.0     22.0
  '(0.664-1.078]'                2231.0    191.0
  '(1.078-1.492]'                 791.0     13.0
  '(1.492-1.906]'                  86.0      1.0
  '(1.906-inf)'                     8.0      1.0
  [total]                        3167.0    228.0

FTI_measured
  t                              3165.0    224.0
  f                               378.0      9.0
```



Classifier: NaiveBayes

Test options: ○ Use training set ○ Supplied test set ● Cross-validation Folds 10 ○ Percentage split % 66

(Nom) Class

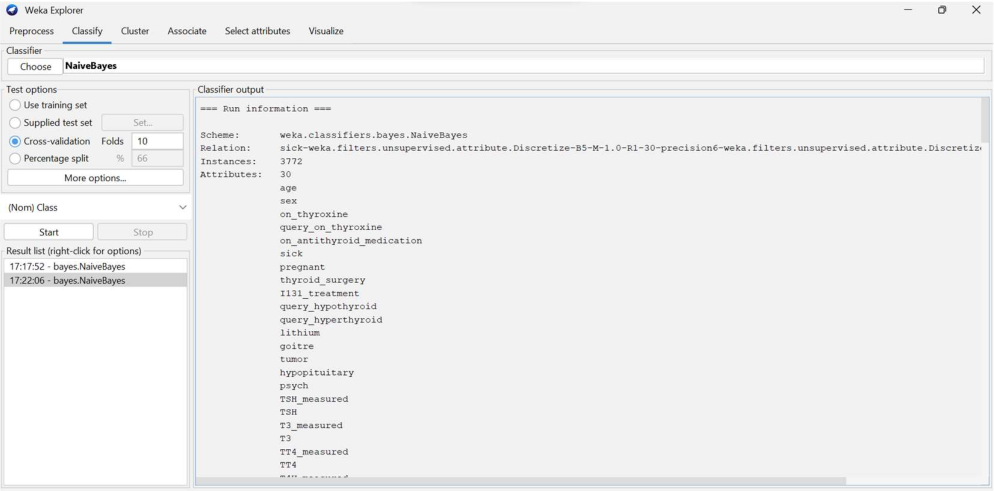
17:17:52 - bayes.NaiveBayes

Classifier output

```
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```

2. Unsupervised discretized with 'bins'=10

Ans:



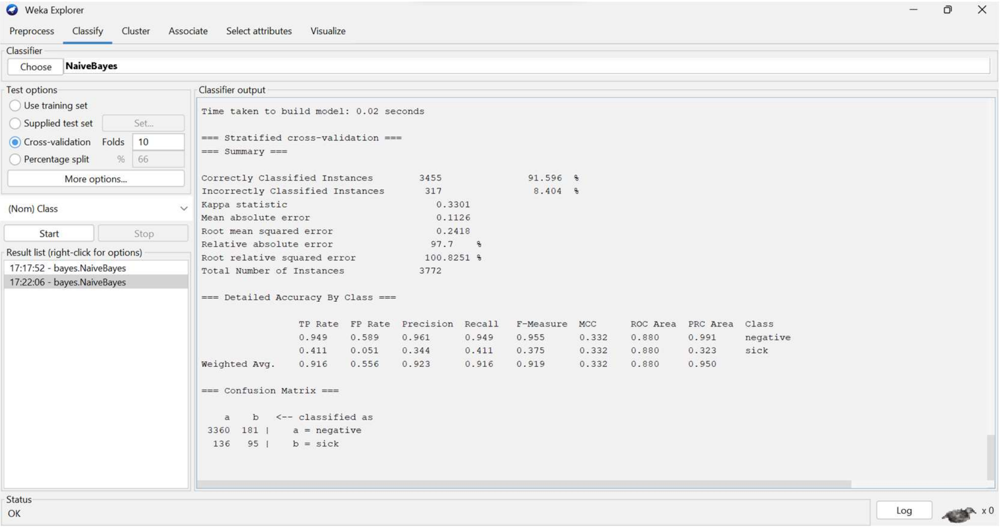
Classifier: NaiveBayes

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

17:17:52 - bayes.NaiveBayes
17:22:06 - bayes.NaiveBayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     sick-weka.filters.unsupervised.attribute.Discretize-B5-M-1.0-R1-30-precision6-weka.filters.unsupervised.attribute.Discretiz
Instances:    3772
Attributes:   30
              age
              sex
              on_thyroxine
              query_on_thyroxine
              on_antithyroid_medication
              sick
              pregnant
              thyroid_surgery
              I131_treatment
              query_hypothyroid
              query_hyperthyroid
              lithium
              goitre
              tumor
              hypopituitary
              psych
              TSH_measured
              TSH
              T3_measured
              T3
              TT4_measured
              TT4
```



Classifier: NaiveBayes

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

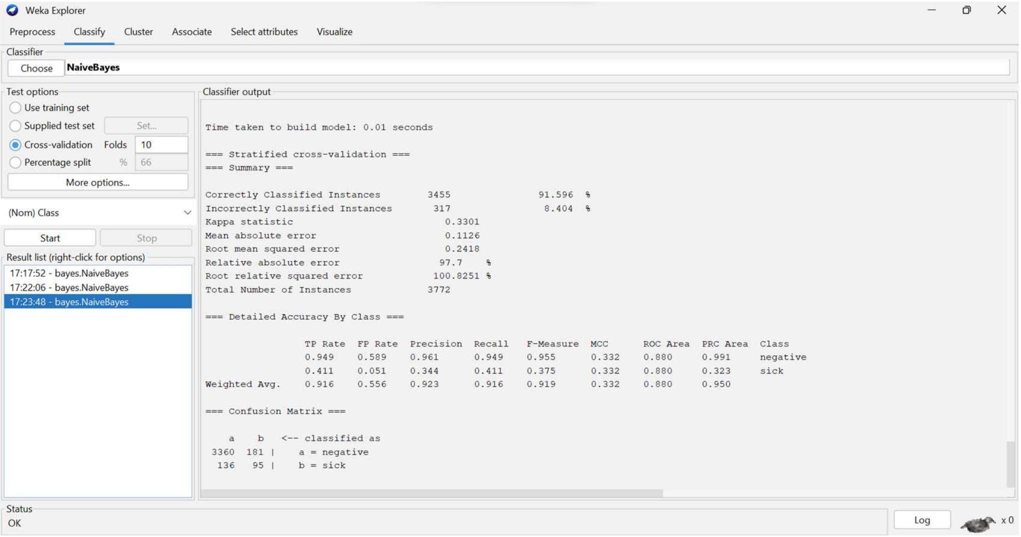
17:17:52 - bayes.NaiveBayes
17:22:06 - bayes.NaiveBayes

```
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```

Status: OK

3. Unsupervised discretized with 'bins''=20.

Ans:



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10 ○ Percentage split % 66

(Nom) Class

Result list:
17:17:52 - bayes.NaiveBayes
17:22:06 - bayes.NaiveBayes
17:23:48 - bayes.NaiveBayes

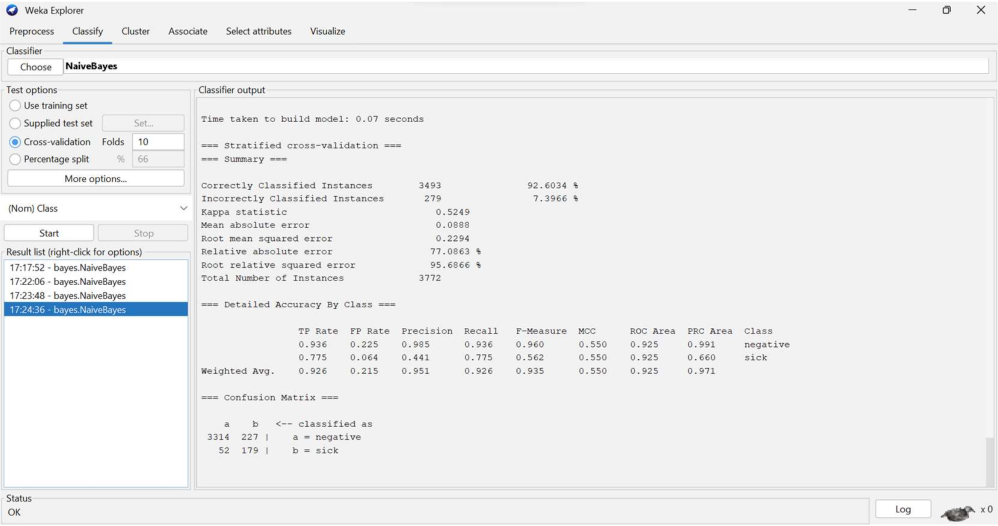
```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```

Status: OK

8. Compare the accuracy of the following cases

1. Naive Bayes without discretization filters

Ans:



Classifier: NaiveBayes

Test options: ● Cross-validation Folds 10 ○ Percentage split % 66

(Nom) Class

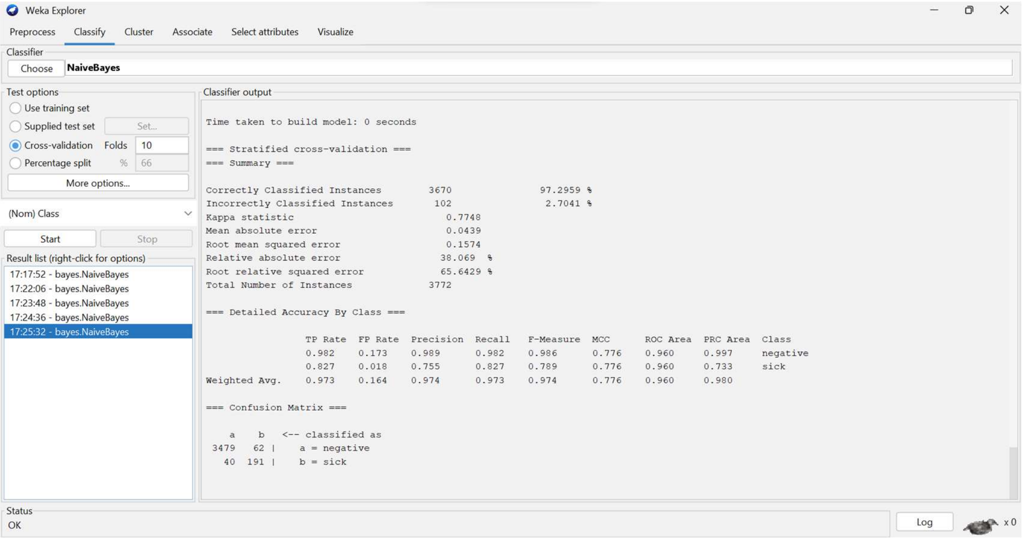
Result list:
17:17:52 - bayes.NaiveBayes
17:22:06 - bayes.NaiveBayes
17:23:48 - bayes.NaiveBayes
17:24:36 - bayes.NaiveBayes

```
Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3493               92.6034 %
Incorrectly Classified Instances       279                7.3966 %
Kappa statistic                          0.5249
Mean absolute error                      0.0888
Root mean squared error                  0.2294
Relative absolute error                 77.0863 %
Root relative squared error             95.6866 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.936    0.225    0.985      0.936    0.960      0.550    0.925     0.991     negative
                 0.775    0.064    0.441      0.775    0.562      0.550    0.925     0.660     sick
Weighted Avg.    0.926    0.215    0.951      0.926    0.935      0.550    0.925     0.971

=== Confusion Matrix ===

    a    b   <-- classified as
 3314  227 |    a = negative
   52  179 |    b = sick
```

Status: OK

2. Naive Bayes with a supervised discretization filter

Ans:



Classifier: NaiveBayes

Test options
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

Result list (right-click for options)
17:17:52 - bayes.NaiveBayes
17:22:06 - bayes.NaiveBayes
17:23:48 - bayes.NaiveBayes
17:24:36 - bayes.NaiveBayes
17:25:32 - bayes.NaiveBayes

Classifier output

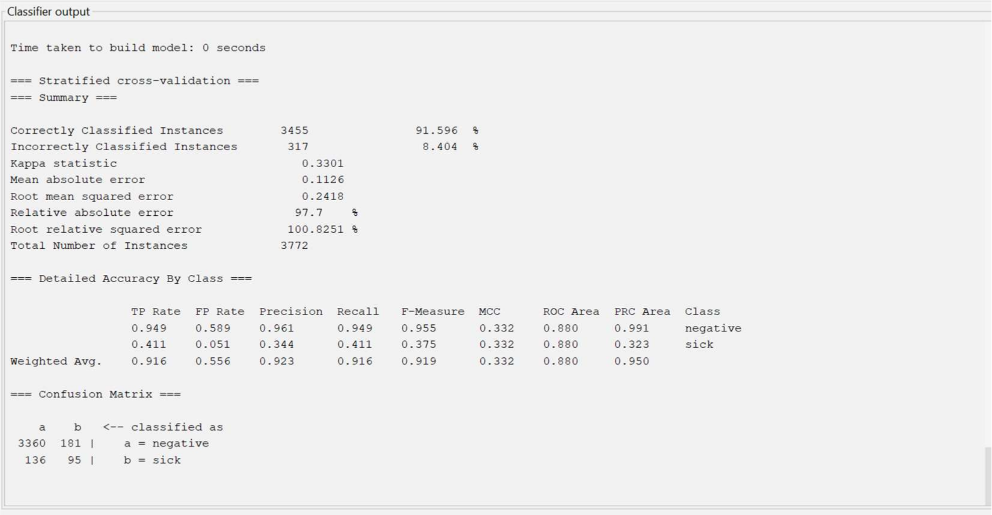
```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3670               97.2959 %
Incorrectly Classified Instances       102                2.7041 %
Kappa statistic                          0.7748
Mean absolute error                      0.0439
Root mean squared error                  0.1574
Relative absolute error                 38.069  %
Root relative squared error             65.6429 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.982    0.173    0.989      0.982    0.986      0.776    0.960     0.997     negative
                 0.827    0.018    0.755      0.827    0.789      0.776    0.960     0.733     sick
Weighted Avg.    0.973    0.164    0.974      0.973    0.974      0.776    0.960     0.980

=== Confusion Matrix ===

    a    b   <-- classified as
 3479   62 |    a = negative
   40  191 |    b = sick
```

Status
OK

3. Naive Bayes with an unsupervised discretization filter with different values for the 'bins attributes.

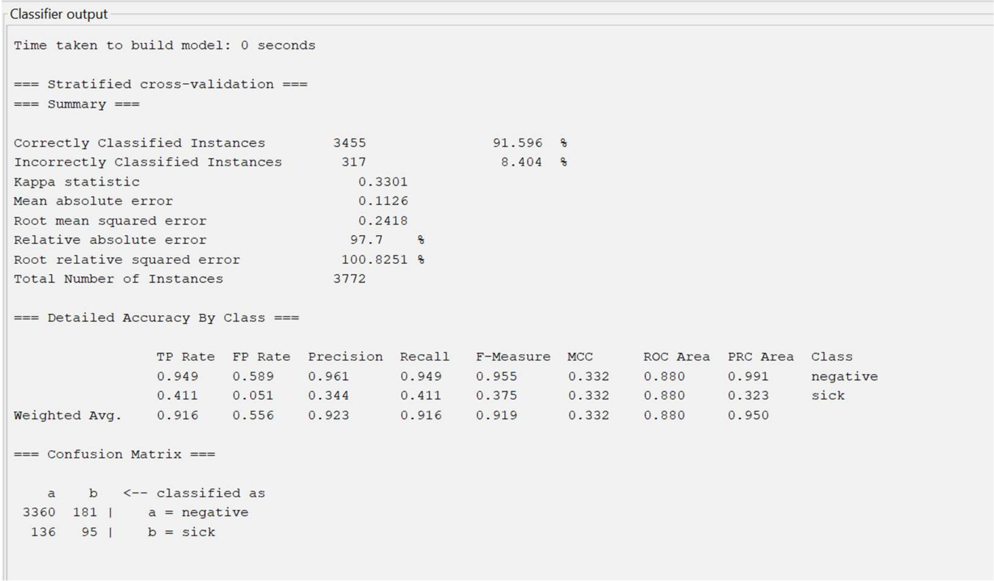
Ans:

Bins = 5


Classifier output

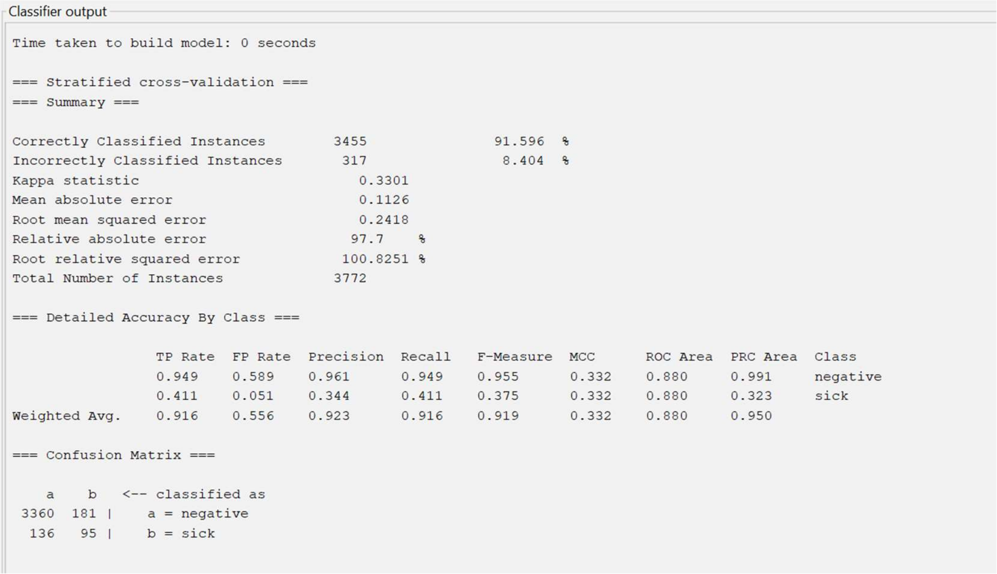
```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```

Bins = 10



```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```

Bins = 20



```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        3455               91.596  %
Incorrectly Classified Instances       317                8.404  %
Kappa statistic                          0.3301
Mean absolute error                      0.1126
Root mean squared error                  0.2418
Relative absolute error                 97.7    %
Root relative squared error            100.8251 %
Total Number of Instances             3772

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.589    0.961      0.949    0.955      0.332    0.880     0.991     negative
                 0.411    0.051    0.344      0.411    0.375      0.332    0.880     0.323     sick
Weighted Avg.    0.916    0.556    0.923      0.916    0.919      0.332    0.880     0.950

=== Confusion Matrix ===

    a    b   <-- classified as
 3360  181 |    a = negative
  136   95 |    b = sick
```