
Experiment No - 01

Author Name : Vinni Fengade
Roll No. : 67
Sem & Sec : 7 CSE [B]
Date Compiled : 24-Jul-2023
File Name : Prac1.txt

AIM: (A) Introduction to Weka tool.
(B) Performing data understanding and preprocessing on the given data set in Weka.

EXECUTION

A. 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?

=> there are 14 instances in dataset

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

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

Outlook:nominal, Temperature:nominal, Humidity:Nominal, Windy:nominal, Play:nominal.

3.What is the class attribute?

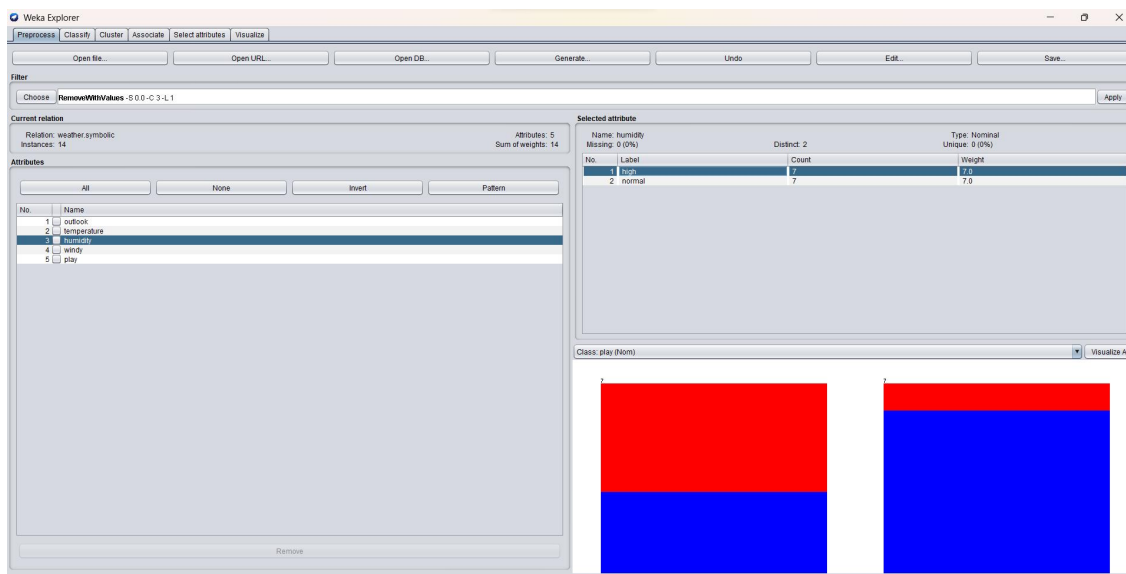
The class attribute is Play which can take 2 values Yes , No

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

By selecting the attribute from Attribute list and filtering for which value of attribute will have yes and no

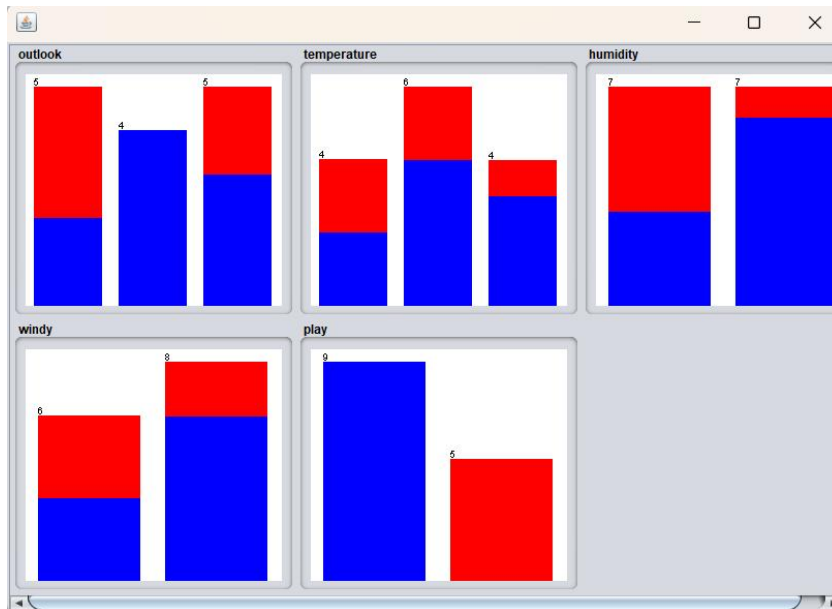
Like for example here yes-blue,no-red

So in case of humidity: for high humidity we have 7 instances in which 3 are yes and 4 are no
Normal humidity we have 7 instances in which 1 is no 6 is yes



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

It will give the all the different classification of Play attribute on basis of different classes



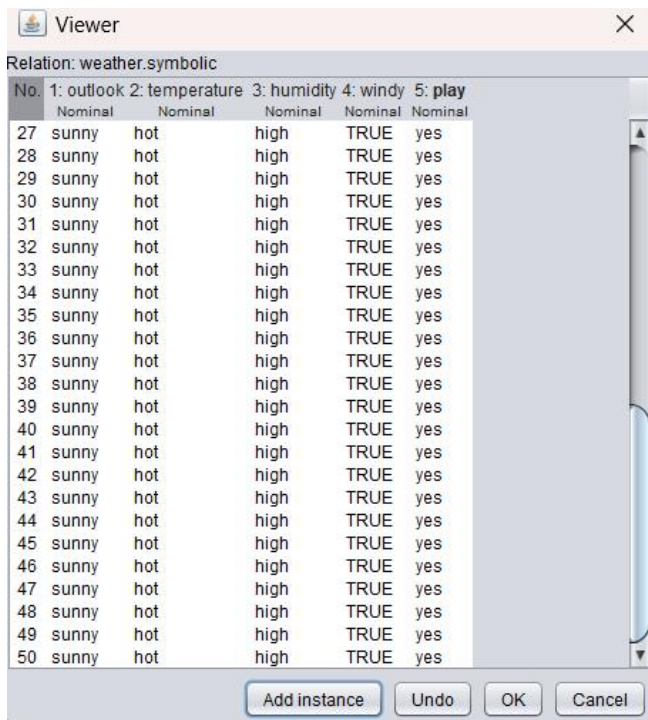
6. How will you view the instances in the dataset? How will you save the changes?

Click on edit button and then change any attribute you want and click on ok the changes will be saved:

The figure shows a 'Viewer' window with a table of instances from the 'weather.symbolic' relation. The table has columns for instance number, outlook, temperature, humidity, windy, and play. A dropdown menu is open for the 'play' column of instance 13, showing 'TRUE' and 'FALSE' options.

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	TRUE	yes
14	rainy	mild	high	TRUE	no

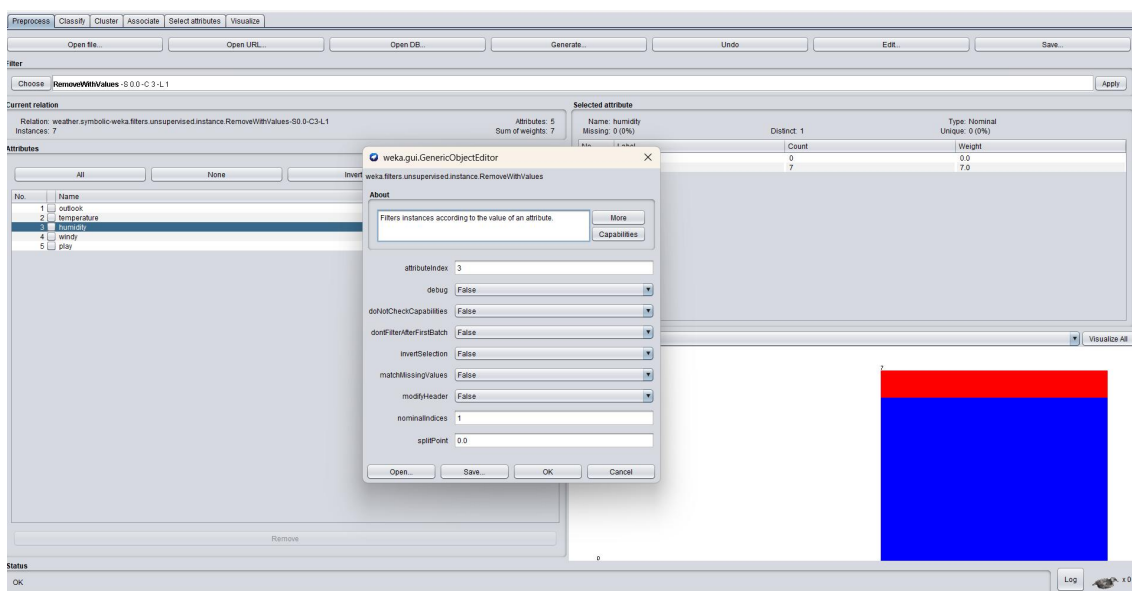
7. Now, extend the dataset to include 50 instances in total.



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

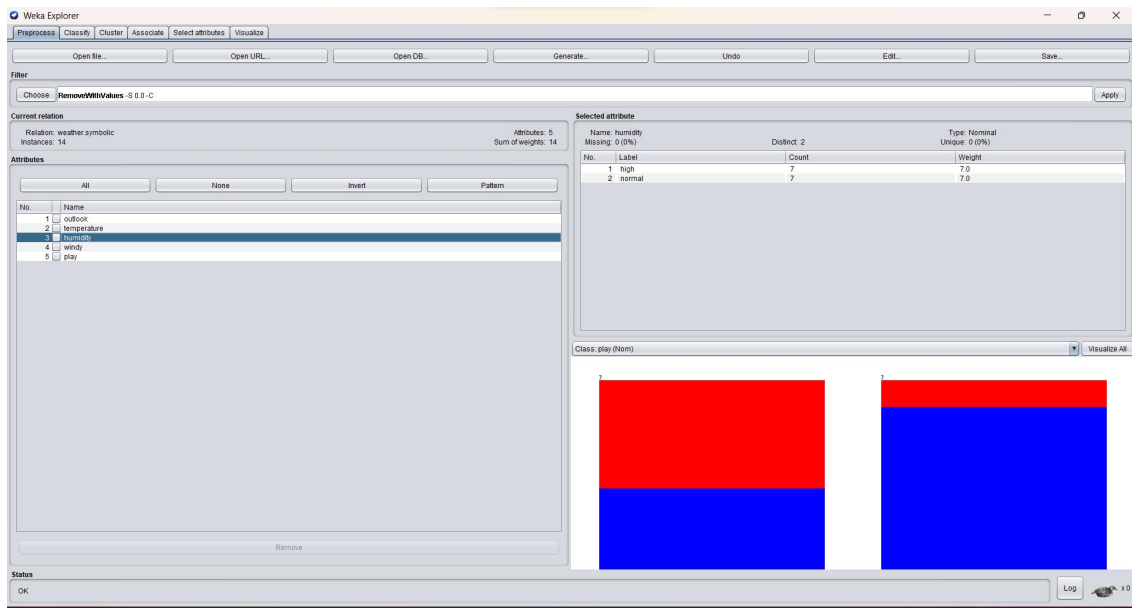
2. Do as directed to apply Filter

a) Use the unsupervised filter RemoveWithValues to remove all instances where the attribute 'humidity' has the value 'high'? Undo the effect of the filter.

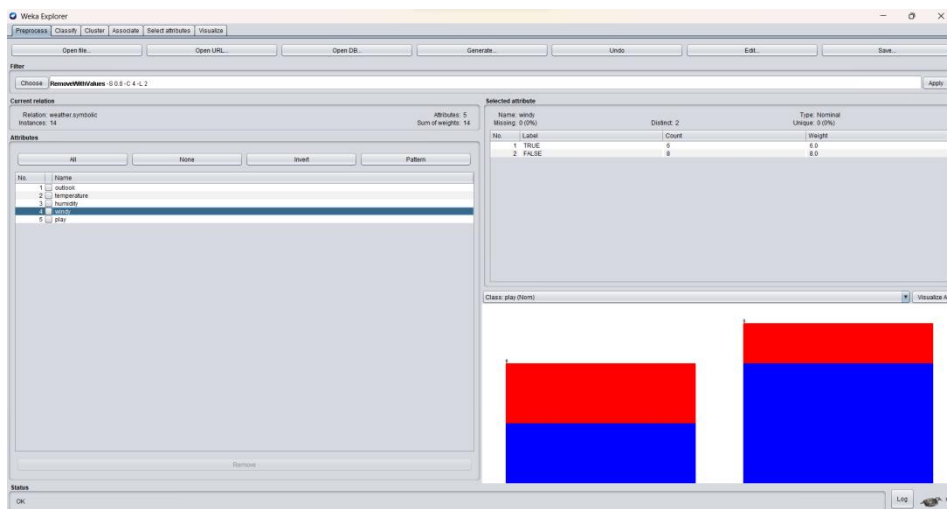
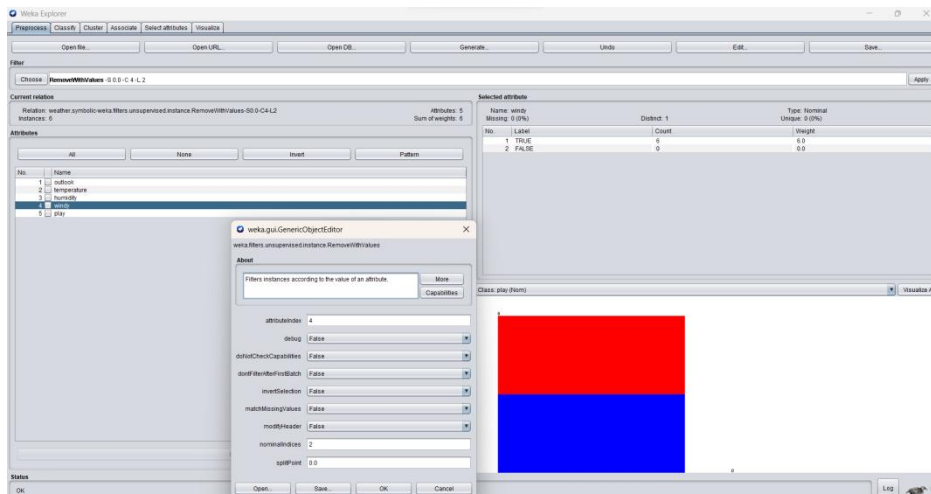


Weka GUI Screenshot:

- Filter: RemoveWithValues - S 0.0 - C 3-L1
- Current relation: weather.symbolic:weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C3-L1
- Instances: 7
- Attributes: 1 outlook, 2 temperature, 3 humidity, 4 windy, 5 play
- Selected attribute: humidity
- Attribute Index: 3
- Debug: False
- Output Check Capabilities: False
- Dont Filter After First Batch: False
- Invert Selection: False
- Match Missing Values: False
- Modify Header: False
- Nominal Indices: 1
- Split Point: 0.0



b) Remove the 'FALSE' instances of windy attribute and undo the effect.



c. Remove the attribute outlook and undo the effect.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose RemoveByName - E outlook Apply

Current relation: weather.symbolic-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.RemoveByName-E-outlook
Instances: 14 Attributes: 4 Sum of weights: 14

Attributes:

No.	Name
1	temperature
2	humidity
3	windy
4	play

Selected attribute:

No.	Label	Count	Weight
1	hot	4	4.0
2	mild	6	6.0
3	cool	4	4.0

Class: play (Nom)

Visualize All

Log x0

Status: OK

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.RemoveByName

About

Removes attributes based on a regular expression matched against their names but will not remove the class attribute.

debug: False

doNotCheckCapabilities: False

expression: outlook

invertSelection: False

Open... Save... OK Cancel

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose RemoveByName - E outlook Apply

Current relation: weather.symbolic
Instances: 14 Attributes: 5 Sum of weights: 14

Attributes:

No.	Name
1	outlook
2	temperature
3	humidity
4	windy
5	play

Selected attribute:

No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

Class: play (Nom)

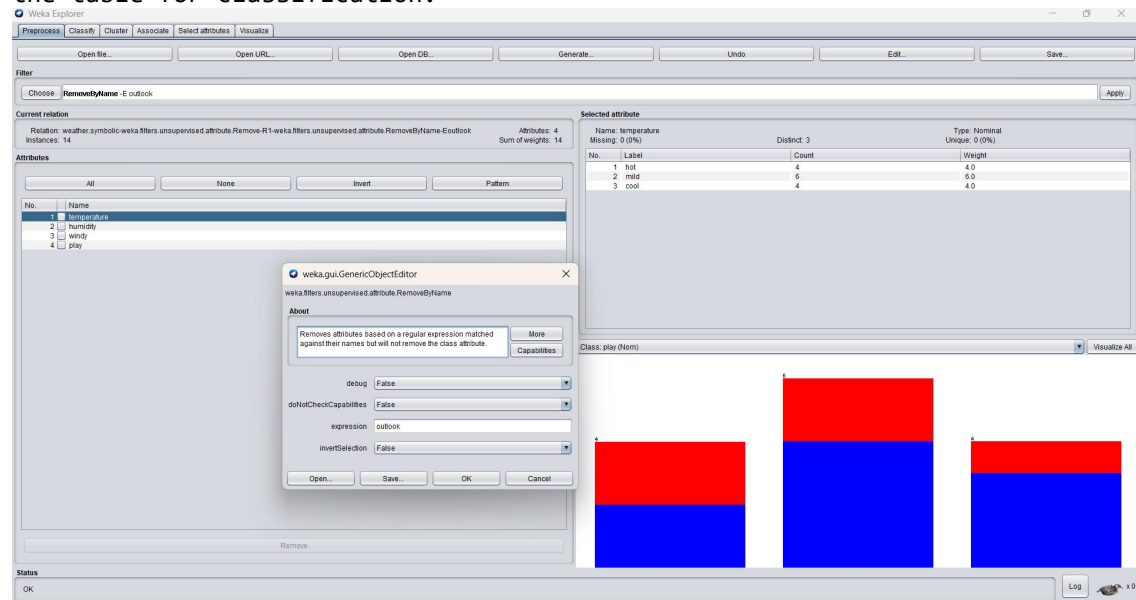
Visualize All

Log x0

Status: OK

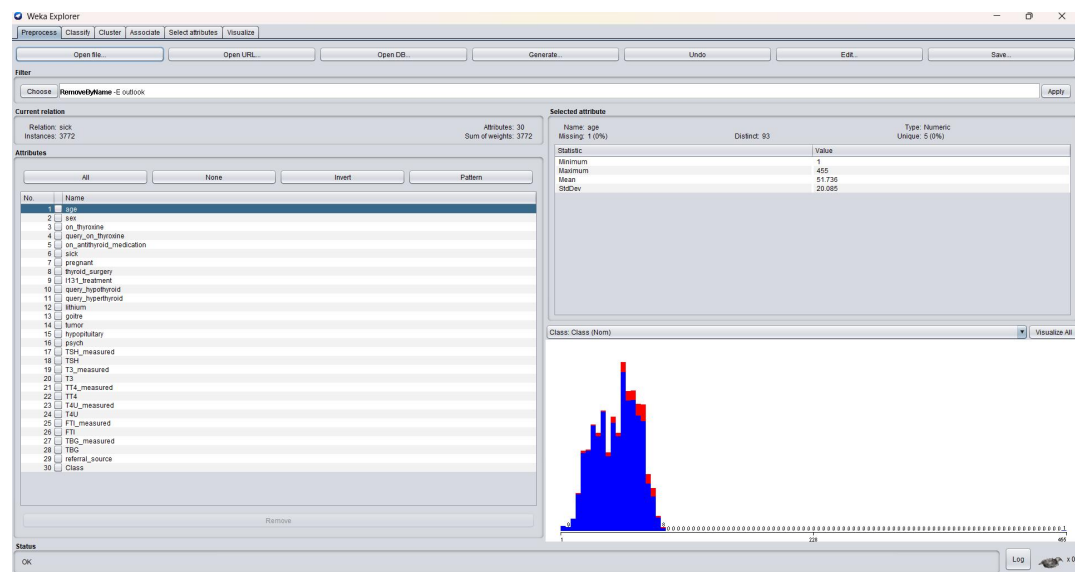
d. Experiment with different filters and report their effects.

The different filters have different effect which can be used to classify the data like with help of removebyvalue filter we can remove the instance according to its value and by using remove we can remove the attribute from the table for classification.



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

a. Load the 'sick.arff' dataset.



b. Apply the supervised discretization filter on different attributes.

Weka Explorer interface showing the Discretize filter applied to the 'age' attribute. The 'Attributes' list on the left shows 'age' selected. The 'Selected attribute' panel on the right shows the distribution of 'age' with three distinct values: 1, 2, and 3. The 'weka.gui.GenericObjectEditor' dialog is open, showing the 'Discretize' filter settings, including 'binRangePrecision' set to 100. The 'Status' bar at the bottom indicates 'OK'.

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

The numeric attribute is converted to nominal attribute after placing a filter on it

Two side-by-side screenshots of the Weka Viewer. The left screenshot shows the 'Viewer' window with the 'age' attribute as a numeric attribute. The right screenshot shows the 'Viewer' window after applying the Discretize filter, where the 'age' attribute is now a nominal attribute with discrete values like '(-inf...', '43...', and '69...'.

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

For age it created 3 different ranges, for TSH it created only one

Name: age		Type: Nominal	
Missing: 1 (0%)		Unique: 0 (0%)	
		Distinct: 3	
No.	Label	Count	Weight
1	'(-inf-43.5]'	1325	1325.0
2	'(43.5-69.5]'	1657	1657.0
3	'(69.5-inf)'	789	789.0

e. Undo the filter applied in the previous step.

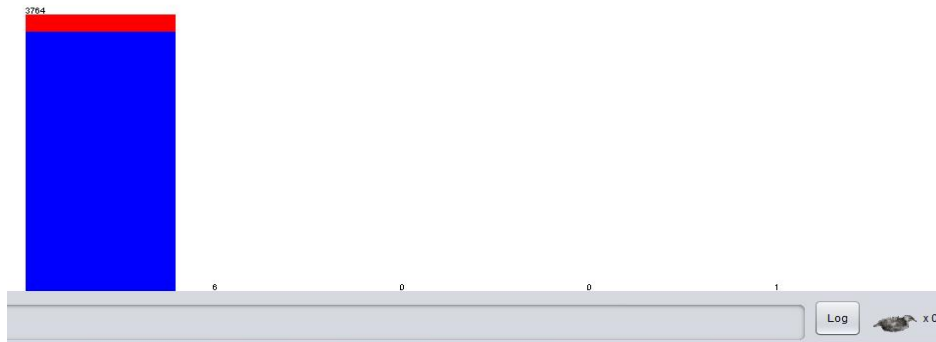
Name: age		Type: Numeric	
Missing: 1 (0%)		Unique: 5 (0%)	
		Distinct: 93	
Statistic	Value		
Minimum	1		
Maximum	455		
Mean	51.736		
StdDev	20.085		

f. Apply the unsupervised discretization filter. [Use equal-width binning approach]

1. In this step, set 'bins'=5
 2. In this step, set 'bins'=10
 3. What is the effect of the unsupervised filter on the dataset?
-

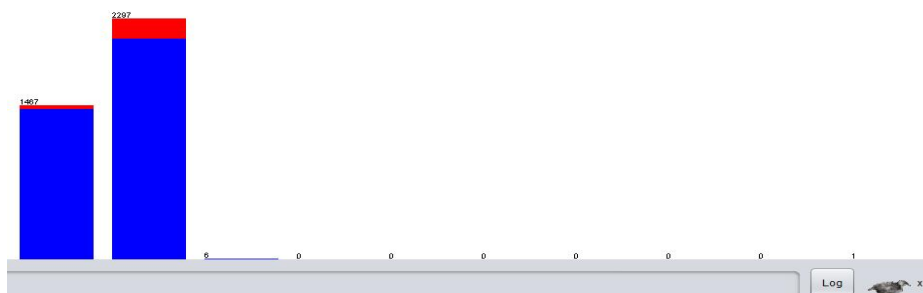
Bins=5

Name: age Missing: 1 (0%)		Distinct: 3		Type: Nominal Unique: 1 (0%)	
No.	Label	Count		Weight	
1	'(-inf-91.8]	3764		3764.0	
2	'(91.8-182.6]	6		6.0	
3	'(182.6-273.4]	0		0.0	
4	'(273.4-364.2]	0		0.0	
5	'(364.2-inf)	1		1.0	



For bin=10

Name: age Missing: 1 (0%)		Distinct: 4		Type: Nominal Unique: 1 (0%)	
No.	Label	Count		Weight	
1	'(-inf-46.4]	1467		1467.0	
2	'(46.4-91.8]	2297		2297.0	
3	'(91.8-137.2]	6		6.0	
4	'(137.2-182.6]	0		0.0	
5	'(182.6-228]	0		0.0	
6	'(228-273.4]	0		0.0	
7	'(273.4-318.8]	0		0.0	
8	'(318.8-364.2]	0		0.0	
9	'(364.2-409.6]	0		0.0	
10	'(409.6-inf)	1		1.0	



By using the unsupervised discretization filter we were able to convert the numeric data into nominal data for age etc. Also it help us to get the data in format of ranges as it will be better to form a conclusion out of the data as the ranged data is better in term for parameters like age.

- g. Run the the Naive Bayes classifier after apply the following filters
1. Unsupervised discretized with 'bins'=5
 2. Unsupervised discretized with 'bins'=10
 3. Unsupervised discretized with 'bins'='=20.

Bin =5

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

15:27:11 - bayes.NaiveBayes

Classifier output

```
=====  
TBG  
mean          0          0  
std. dev.     0.0017    0.0017  
weight sum    0          0  
precision     0.01      0.01  
  
referral_source  
SVBC          378.0     10.0  
other         2169.0    34.0  
SVI           849.0    187.0  
STMW          113.0     1.0  
SVHD           37.0     4.0  
[total]      3546.0    236.0  
  
Time taken to build model: 0.03 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
Correctly Classified Instances      3519      93.292 %  
Incorrectly Classified Instances    253      6.707 %  
Kappa statistic                    0.5447  
Mean absolute error                 0.0853  
Root mean squared error             0.2184  
Relative absolute error             74.0569 %  
Root relative squared error         91.0919 %  
Total Number of Instances          3772  
  
=== Detailed Accuracy By Class ===  
  
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class  
Weighted Avg.   0.933   0.235   0.952    0.933   0.940    0.563   0.926    0.972   sick  
  
=== Confusion Matrix ===  
      a  b  <-- classified as  
3345  196 |  a = negative  
  57  174 |  b = sick
```

Bin=10

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

15:27:11 - bayes.NaiveBayes
15:28:57 - bayes.NaiveBayes
15:29:06 - bayes.NaiveBayes
15:29:21 - bayes.NaiveBayes
15:32:25 - bayes.NaiveBayes

Classifier output

```
=====  
TBG  
mean          0          0  
std. dev.     0.0017    0.0017  
weight sum    0          0  
precision     0.01      0.01  
  
referral_source  
SVBC          378.0     10.0  
other         2169.0    34.0  
SVI           849.0    187.0  
STMW          113.0     1.0  
SVHD           37.0     4.0  
[total]      3546.0    236.0  
  
Time taken to build model: 0 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
Correctly Classified Instances      3510      93.0541 %  
Incorrectly Classified Instances    262      6.9459 %  
Kappa statistic                    0.538  
Mean absolute error                 0.0867  
Root mean squared error             0.2237  
Relative absolute error             75.2603 %  
Root relative squared error         93.3004 %  
Total Number of Instances          3772  
  
=== Detailed Accuracy By Class ===  
  
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class  
Weighted Avg.   0.931   0.227   0.952    0.931   0.938    0.558   0.926    0.971   negative  
              0.931   0.227   0.952    0.931   0.938    0.558   0.926    0.971   sick  
  
=== Confusion Matrix ===  
      a  b  <-- classified as  
3334  207 |  a = negative  
  55  176 |  b = sick
```

Bin =20

Choose **NaiveBayes**

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

More options...

(Nom) Class

Start Stop

Result list (right-click for options)

- 15:27:11 - bayes NaiveBayes
- 15:28:57 - bayes NaiveBayes
- 15:29:06 - bayes NaiveBayes
- 15:29:21 - bayes NaiveBayes
- 15:32:25 - bayes NaiveBayes
- 15:33:35 - bayes NaiveBayes**

Classifier output

```

TBG
  mean      0      0
  std. dev. 0.0017 0.0017
  weight sum 0      0
  precision 0.01   0.01

referral_source
  SVHC      378.0   10.0
  other     2169.0  34.0
  SVI       849.0   187.0
  STMW      113.0   1.0
  SVHD      37.0    4.0
  [total]   3546.0  236.0

Time taken to build model: 0 seconds

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

Correctly Classified Instances      3505      92.9215 %
Incorrectly Classified Instances    267      7.0785 %
Kappa statistic                    0.9371
Mean absolute error                 0.0661
Root Mean squared error             0.225
Relative absolute error             74.7094 %
Root relative squared error        93.8478 %
Total Number of Instances          3772

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MDC  ROC Area  FBC Area  Class
Weighted Avg.  0.929  0.215  0.952  0.929  0.938  0.560  0.926  0.970  negative
                    0.775  0.061  0.454  0.775  0.573  0.560  0.526  0.644  sick

=== Confusion Matrix ===
  a  b  <-- classified as
3326 215 | a = negative
 52 179 | b = sick

```

Here after changing the size of bn the data classification of data become different so the values in confusion matrix also changes as the size changed

And the results were better as the no of bin were increasing to some extent

h. Compare the accuracy of the following cases

1. Naive Bayes without discretization filters
2. Naive Bayes with a supervised discretization filter
3. Naive Bayes with an unsupervised discretization filter with different values for the 'bins' attributes.

Without filter:

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

(Nom) Class

Result list (right-click for options)

- 15:27:11 - bayes.NaiveBayes
- 15:28:57 - bayes.NaiveBayes
- 15:29:06 - bayes.NaiveBayes
- 15:29:21 - bayes.NaiveBayes
- 15:32:25 - bayes.NaiveBayes
- 15:33:35 - bayes.NaiveBayes
- 15:37:26 - bayes.NaiveBayes**

Classifier output

TSG

mean	0	0
std. dev.	0.0017	0.0017
weight sum	0	0
precision	0.01	0.01

referral_source

SVBC	378.0	10.0
other	2169.0	34.0
SVI	849.0	187.0
STMW	113.0	1.0
SVBD	37.0	4.0
[total]	3546.0	236.0

Time taken to build model: 0 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
Weighted Avg.	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

=== Confusion Matrix ===

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

with a supervised discretization filter

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds **10**

☐ Percentage split % **66**

(Nom) Class

Result list (right-click for options)

- 15:27:11 - bayes.NaiveBayes
- 15:28:57 - bayes.NaiveBayes
- 15:29:06 - bayes.NaiveBayes
- 15:29:21 - bayes.NaiveBayes
- 15:32:25 - bayes.NaiveBayes
- 15:33:35 - bayes.NaiveBayes
- 15:37:26 - bayes.NaiveBayes
- 15:38:22 - bayes.NaiveBayes
- 15:38:46 - bayes.NaiveBayes
- 15:40:04 - bayes.NaiveBayes
- 15:40:31 - bayes.NaiveBayes**

Classifier output

TSG

mean	0	0
std. dev.	0.0017	0.0017
weight sum	0	0
precision	0.01	0.01

referral_source

SVBC	378.0	10.0
other	2169.0	34.0
SVI	849.0	187.0
STMW	113.0	1.0
SVBD	37.0	4.0
[total]	3546.0	236.0

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	3503	92.8685 %
Incorrectly Classified Instances	269	7.1315 %
Kappa statistic	0.542	
Mean absolute error	0.0876	
Root mean squared error	0.2282	
Relative absolute error	76.0614 %	
Root relative squared error	95.1602 %	
Total Number of Instances	3772	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.937	0.203	0.986	0.937	0.961	0.568	0.927	0.992	negative
	0.797	0.063	0.453	0.797	0.578	0.568	0.927	0.652	sick

=== Confusion Matrix ===

a	b	<-- classified as
3319	222	a = negative
47	184	b = sick

As we can see the value of class with respect to filter and non filtered data is different as is better for filtered data as it is easy to predict the reason more.

i. Repeat steps 6 to 8 using equal-frequency binning approach and present your conclusion.

Without filter:

Classifier
Choose: NaiveBayes

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☐ Percentage split %: 66
 More options...

(Nom) Class
Start Stop

Result list (right-click for options)
 15:27:11 - bayes NaiveBayes
 15:28:57 - bayes NaiveBayes
 15:29:06 - bayes NaiveBayes
 15:29:21 - bayes NaiveBayes
 15:32:25 - bayes NaiveBayes
 15:33:35 - bayes NaiveBayes
 15:37:26 - bayes NaiveBayes
 15:40:04 - bayes NaiveBayes

Classifier output

```

TBG
mean                0      0
std. dev.          0.0017  0.0017
weight sum          0      0
precision           0.01   0.01

referral_source
SVHC                378.0   10.0
other               2169.0  34.0
SVI                 849.0   187.0
STHW               113.0    1.0
SVHD                37.0    4.0
[total]            3546.0  236.0

Time taken to build model: 0 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
Weighted Avg.   0.926   0.215   0.951     0.926   0.935     0.550   0.925     0.660    sick

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

with a supervised discretization filter

Classifier
Choose: NaiveBayes

Test options
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds: 10
☐ Percentage split %: 66
 More options...

(Nom) Class
Start Stop

Result list (right-click for options)
 15:27:11 - bayes NaiveBayes
 15:28:57 - bayes NaiveBayes
 15:29:06 - bayes NaiveBayes
 15:29:21 - bayes NaiveBayes
 15:32:25 - bayes NaiveBayes
 15:33:35 - bayes NaiveBayes
 15:37:26 - bayes NaiveBayes
 15:39:22 - bayes NaiveBayes
 15:39:46 - bayes NaiveBayes
 15:40:04 - bayes NaiveBayes
 15:40:31 - bayes NaiveBayes

Classifier output

```

TBG
mean                0      0
std. dev.          0.0017  0.0017
weight sum          0      0
precision           0.01   0.01

referral_source
SVHC                378.0   10.0
other               2169.0  34.0
SVI                 849.0   187.0
STHW               113.0    1.0
SVHD                37.0    4.0
[total]            3546.0  236.0

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      3503      92.6685 %
Incorrectly Classified Instances    269       7.1315 %
Kappa statistic                    0.542
Mean absolute error                 0.0876
Root mean squared error             0.2282
Relative absolute error             76.0614 %
Root relative squared error        95.1602 %
Total Number of Instances          3772

--- Detailed Accuracy By Class ---
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0.929   0.195   0.953     0.929   0.938     0.568   0.927     0.971    negative

--- Confusion Matrix ---
      a   b   <-- classified as
3119  212 |   a = negative
  47  184 |   b = sick
  
```

equal-frequency binning approach

Choose **NaiveBayes**

Test options

☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds
☐ Percentage split %

(Nom) Class

Result list (right-click for options)

- 15:27:11 - bayes NaiveBayes
- 15:28:57 - bayes NaiveBayes
- 15:29:06 - bayes NaiveBayes
- 15:29:21 - bayes NaiveBayes
- 15:32:25 - bayes NaiveBayes
- 15:33:35 - bayes NaiveBayes
- 15:37:26 - bayes NaiveBayes
- 15:39:22 - bayes NaiveBayes
- 15:39:46 - bayes NaiveBayes
- 15:40:04 - bayes NaiveBayes
- 15:40:31 - bayes NaiveBayes
- 15:48:32 - bayes NaiveBayes**

Classifier output

```

f                3542.0   232.0
[total]          3542.0   232.0

TSG
'All'            1.0      1.0
[total]          1.0      1.0

referral_source
SVHC              378.0    10.0
other             2169.0    34.0
SVI               849.0    187.0
STMW              113.0     1.0
SVSD              37.0      4.0
[total]          3546.0    236.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      3651      96.7922 %
Incorrectly Classified Instances    121      3.2078 %
Kappa statistic                    0.7404
Mean absolute error                 0.0511
Root mean squared error             0.1659
Relative absolute error              44.3476 %
Root relative squared error         70.437 %
Total Number of Instances          3772

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0.978    0.182    0.988    0.978    0.983    0.743    0.961    0.597    negative
      0.818    0.022    0.705    0.818    0.758    0.743    0.961    0.676    sick
Weighted Avg.   0.968    0.172    0.971    0.968    0.969    0.743    0.961    0.977

=== Confusion Matrix ===
      a   b   <-- Classified as
3462  79 |   a = negative
  42 189 |   b = sick
  
```

As compared to the non filtered data and supervised descritized filter the unsupervised descritized filter has better classification as only 3.6% of cases are wrongly classified which is half of supervised learning also the accuracy is incread by a large gap