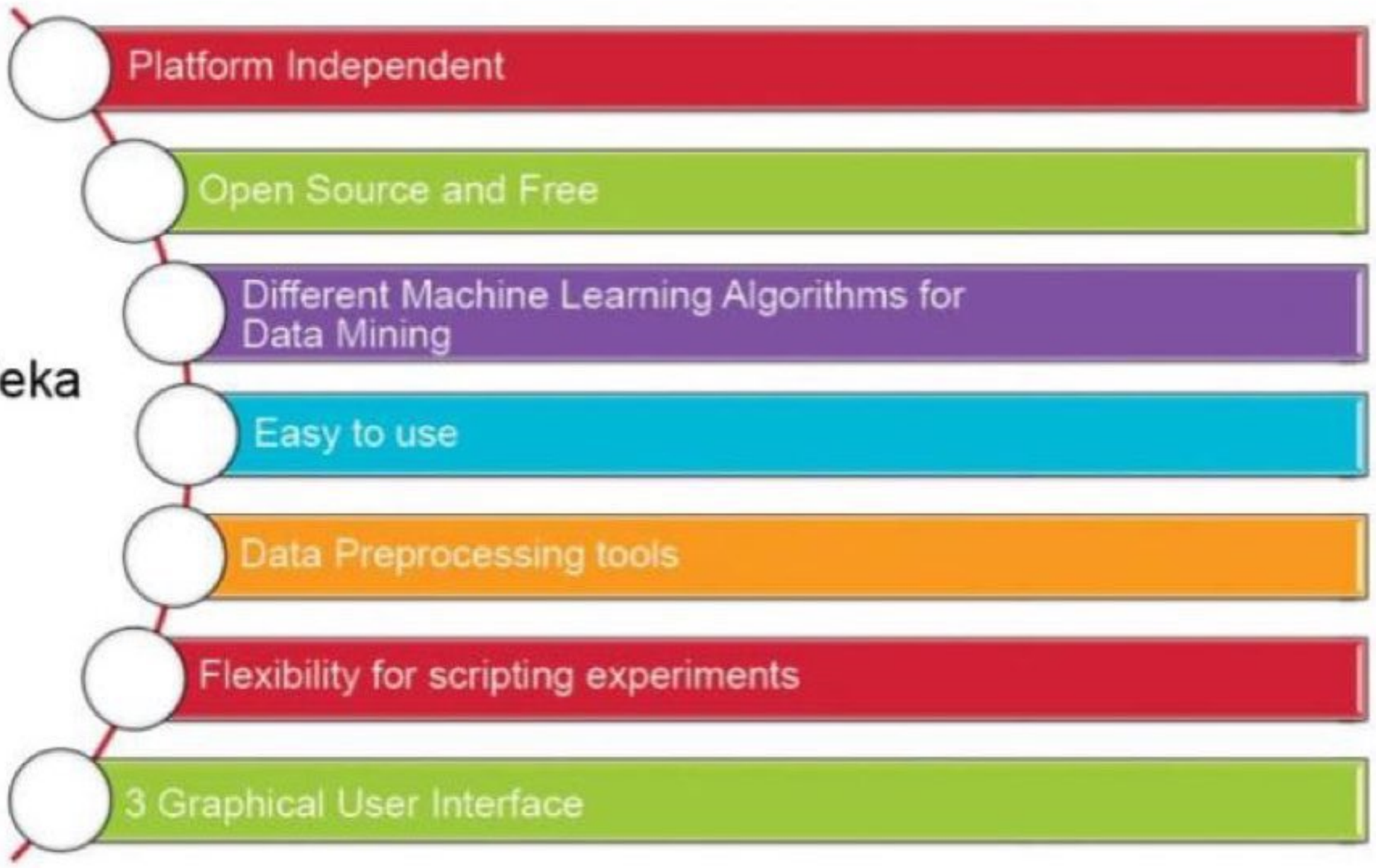
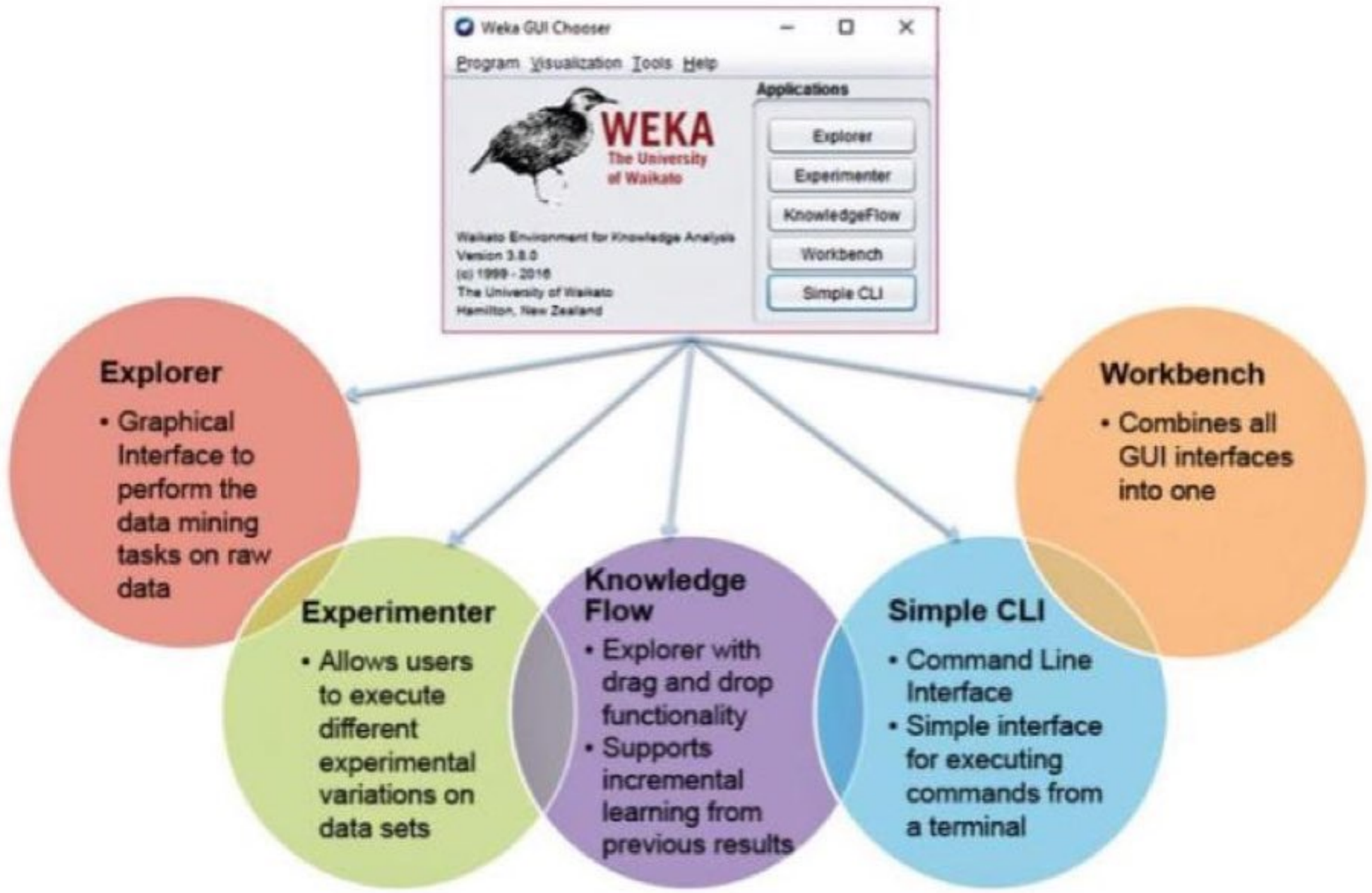
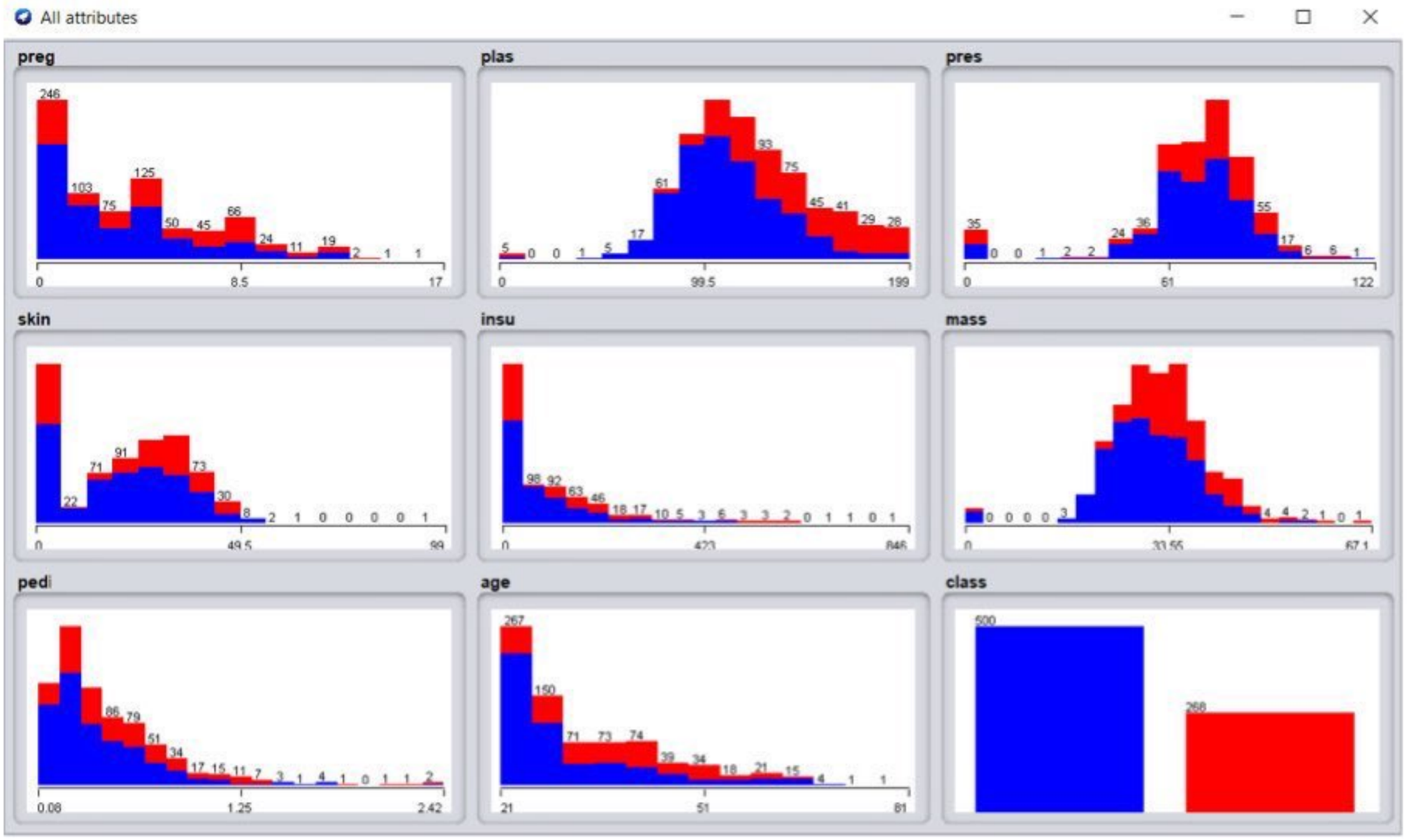
Features of Weka



Weka's application interfaces





After Discretization

Output :

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 4. outlook=sunny play=no 3 ==> humidity=high 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 5. outlook=sunny humidity=high 3 ==> play=no 3    <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
 6. outlook=rainy play=yes 3 ==> windy=FALSE 3    <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
 7. outlook=rainy windy=FALSE 3 ==> play=yes 3    <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
 8. temperature=cool play=yes 3 ==> humidity=normal 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 9. outlook=sunny temperature=hot 2 ==> humidity=high 2    <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2    <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)
```

Output

```
=== Run information ===

Scheme:       weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:     supermarket
Instances:    4627
Attributes:   217
              [list of attributes omitted]
=== Associator model (full training set) ===

FPGrowth found 16 rules (displaying top 10)

 1. [fruit=t, frozen foods=t, biscuits=t, total=high]: 788 ==> [bread and cake=t]: 723   <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.35)
 2. [fruit=t, baking needs=t, biscuits=t, total=high]: 760 ==> [bread and cake=t]: 696   <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.28)
 3. [fruit=t, baking needs=t, frozen foods=t, total=high]: 770 ==> [bread and cake=t]: 705   <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.27)
 4. [fruit=t, vegetables=t, biscuits=t, total=high]: 815 ==> [bread and cake=t]: 746   <conf:(0.92)> lift:(1.27) lev:(0.03) conv:(3.26)
 5. [fruit=t, party snack foods=t, total=high]: 854 ==> [bread and cake=t]: 779   <conf:(0.91)> lift:(1.27) lev:(0.04) conv:(3.15)
 6. [vegetables=t, frozen foods=t, biscuits=t, total=high]: 797 ==> [bread and cake=t]: 725   <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.06)
 7. [vegetables=t, baking needs=t, biscuits=t, total=high]: 772 ==> [bread and cake=t]: 701   <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3.01)
 8. [fruit=t, biscuits=t, total=high]: 954 ==> [bread and cake=t]: 866   <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(3)
 9. [fruit=t, vegetables=t, frozen foods=t, total=high]: 834 ==> [bread and cake=t]: 757   <conf:(0.91)> lift:(1.26) lev:(0.03) conv:(3)
10. [fruit=t, frozen foods=t, total=high]: 969 ==> [bread and cake=t]: 877   <conf:(0.91)> lift:(1.26) lev:(0.04) conv:(2.92)
```

Output

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     iris
Instances:    150
Attributes:   5
              sepallength
              sepalwidth
              petallength
              petalwidth
              class
Test mode:    10-fold cross-validation

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

Naive Bayes Classifier

                         Class
Attribute          Iris-setosa Iris-versicolor  Iris-virginica
                        (0.33)          (0.33)          (0.33)
============================================================
sepallength
  mean                 4.9913          5.9379          6.5795
  std. dev.             0.355          0.5042          0.6353
  weight sum               50              50              50
  precision            0.1059          0.1059          0.1059

sepalwidth
  mean                 3.4015          2.7687          2.9629
  std. dev.            0.3925          0.3038          0.3088
  weight sum               50              50              50
  precision            0.1091          0.1091          0.1091

petallength
  mean                 1.4694          4.2452          5.5516
  std. dev.            0.1782          0.4712          0.5529
  weight sum               50              50              50
  precision            0.1405          0.1405          0.1405

petalwidth
  mean                 0.2743          1.3097          2.0343
  std. dev.            0.1096          0.1915          0.2646
  weight sum               50              50              50
  precision            0.1143          0.1143          0.1143
```

```
Time taken to build model: 0 seconds

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

Correctly Classified Instances         144               96      %
Incorrectly Classified Instances         6                4      %
Kappa statistic                          0.94
Mean absolute error                      0.0342
Root mean squared error                  0.155
Relative absolute error                  7.6997 %
Root relative squared error             32.8794 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     Iris-setosa
               0.960    0.040    0.923      0.960    0.941      0.911    0.992     0.983     Iris-versicolor
               0.920    0.020    0.958      0.920    0.939      0.910    0.992     0.986     Iris-virginica
Weighted Avg.  0.960    0.020    0.960      0.960    0.960      0.940    0.994     0.989

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 48  2 |  b = Iris-versicolor
  0  4 46 |  c = Iris-virginica
```

```
=== Run information ===

Scheme:       weka.classifiers.trees.DecisionStump
Relation:     soybean
Instances:    683
Attributes:   36
              date
              plant-stand
              precip
              temp
              hail
              crop-hist
              area-damaged
              severity
              seed-tmt
              germination
              plant-growth
              leaves
              leafspots-halo
              leafspots-marg
              leafspot-size
              leaf-shread
              leaf-malf
              leaf-mild
              stem
              lodging
              stem-cankers
              canker-lesion
              fruiting-bodies
              external-decay
              mycelium
              int-discolor
              sclerotia
              fruit-pods
              fruit-spots
              seed
              mold-growth
              seed-discolor
              seed-size
              shriveling
              roots
              class
Test mode:    10-fold cross-validation
```

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

Decision Stump

Classifications

leafspot-size = gt-1/8 : brown-spot
leafspot-size != gt-1/8 : anthracnose
leafspot-size is missing : phytophthora-rot
```

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
```

```
=== Summary ===

Correctly Classified Instances         191               27.9649 %
Incorrectly Classified Instances       492               72.0351 %
Kappa statistic                          0.1942
Mean absolute error                      0.0826
Root mean squared error                  0.2033
Relative absolute error                 85.9537 %
Root relative squared error             92.7899 %
Total Number of Instances              683
```

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     diaporthe-stem-canker
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     charcoal-rot
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     rhizoctonia-root-rot
               0.625     0.049     0.655       0.625   0.640      0.588    0.865     0.475     phytophthora-rot
               0.000     0.000     ?           0.000   ?          ?        0.680     0.105     brown-stem-rot
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     powdery-mildew
               0.000     0.000     ?           0.000   ?          ?        0.763     0.060     downy-mildew
               1.000     0.398     0.281       1.000   0.439      0.412    0.794     0.272     brown-spot
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     bacterial-blight
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     bacterial-pustule
               0.000     0.000     ?           0.000   ?          ?        0.804     0.071     purple-seed-stain
               1.000     0.357     0.162       1.000   0.278      0.323    0.808     0.150     anthracnose
               0.000     0.000     ?           0.000   ?          ?        0.763     0.060     phyllosticta-leaf-spot
               0.000     0.000     ?           0.000   ?          ?        0.793     0.269     alternarialeaf-spot
               0.000     0.000     ?           0.000   ?          ?        0.796     0.272     frog-eye-leaf-spot
               0.000     0.000     ?           0.000   ?          ?        0.938     0.149     diaporthe-pod-&-stem-blight
               0.000     0.000     ?           0.000   ?          ?        0.938     0.143     cyst-nematode
               0.000     0.000     ?           0.000   ?          ?        0.773     0.048     2-4-d-injury
               0.000     0.000     ?           0.000   ?          ?        0.305     0.010     herbicide-injury
Weighted Avg.  0.280     0.083     ?           0.280   ?          ?        0.797     0.212
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  r  s   <-- classified as
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  a = diaporthe-stem-canker
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  b = charcoal-rot
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  c = rhizoctonia-root-rot
  0  0  0 55  0  0  0  0  0  0  0 33  0  0  0  0  0  0  0 |  d = phytophthora-rot
  0  0  0  0  0  0  0  9  0  0  0 35  0  0  0  0  0  0  0 |  e = brown-stem-rot
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  f = powdery-mildew
  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0  0  0  0  0 |  g = downy-mildew
  0  0  0  0  0  0  0 92  0  0  0  0  0  0  0  0  0  0  0 |  h = brown-spot
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  i = bacterial-blight
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  j = bacterial-pustule
  0  0  0  0  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0 |  k = purple-seed-stain
  0  0  0  0  0  0  0  0  0  0  0 44  0  0  0  0  0  0  0 |  l = anthracnose
  0  0  0  0  0  0  0 20  0  0  0  0  0  0  0  0  0  0  0 |  m = phyllosticta-leaf-spot
  0  0  0  0  0  0  0 91  0  0  0  0  0  0  0  0  0  0  0 |  n = alternarialeaf-spot
  0  0  0  0  0  0  0 91  0  0  0  0  0  0  0  0  0  0  0 |  o = frog-eye-leaf-spot
  0  0  0 15  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  p = diaporthe-pod-&-stem-blight
  0  0  0 14  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  q = cyst-nematode
  0  0  0  0  0  0  0  0  0  0  0 16  0  0  0  0  0  0  0 |  r = 2-4-d-injury
  0  0  0  0  0  0  0  4  0  0  0  4  0  0  0  0  0  0  0 |  s = herbicide-injury
```

Procedure

Step1 : Open weka tool and select Explorer from the Applications segment.

Step2 : Click on open file and select the below file
Local Disk (C:) → Program files → Weka 3 8 → Data → diabetes.arff

Step3: Click on cluster tab.

Step 4 : In the classify tab select the following
Weka → clusters → EM

Step 5 : Click on start and wait for the weka tool to build model on the data.

Output

```
=== Run information ===

Scheme:       weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation:     pima_diabetes
Instances:    768
Attributes:   9
              preg
              plas
              pres
              skin
              insu
              mass
              pedi
              age
              class
Test mode:    evaluate on training data


=== Clustering model (full training set) ===


EM
==

Number of clusters selected by cross validation: 3
Number of iterations performed: 9
```

Attribute	Cluster 0 (0.32)	1 (0.33)	2 (0.35)
preg			
mean	4.7239	1.6393	5.1314
std. dev.	3.7151	1.3561	3.3295
plas			
mean	138.3297	103.3789	121.3019
std. dev.	30.6884	19.08	33.8355
pres			
mean	68.9333	62.2903	75.7883
std. dev.	22.716	19.9888	11.2993
skin			
mean	29.0645	22.1298	11.013
std. dev.	13.4995	12.5528	15.895
insu			
mean	173.816	71.039	0
std. dev.	145.668	62.7051	0.0008
mass			
mean	34.6052	29.458	31.9674
std. dev.	7.6535	8.1157	7.0137
pedi			
mean	0.642	0.4023	0.3789
std. dev.	0.4342	0.2257	0.2266
age			
mean	36.6656	23.9583	38.9112
std. dev.	10.7088	2.5897	12.6593
class			
tested_negative	108.1931	236.1879	158.619
tested_positive	142.6434	19.676	108.6807
[total]	250.8365	255.8639	267.2997

```
Time taken to build model (full training data) : 2.04 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       228 ( 30%)
1       203 ( 26%)
2       337 ( 44%)


Log likelihood: -24.97229
```

Step 3 : Assign each data point to their closest centroid, which will form the predefined K clusters

Step 4 : Calculate the variance and place a new centroid of each cluster.

Step 5 : Repeat the third step, which means reassign each datapoint to the new closest centroid of each cluster.

Step 6 : If any reassignment occurs, then go to step 4 else go to finish.

Step 7 : The model is ready.

Procedure

Step 1 : open Weka tool and select Explorer from the Applications segment

Step 2 : click on open file and select the below file
Local Disk (C:) → Program Files → Weka 3-8 → Data → vote.arff

Step 3 : click on cluster tab

Step 4 : In the classify tab select the following
weka → clusters → SimpleKMeans

Step 5 : click on start and wait for the weka tool to build model on the data.

Output

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation:     vote
Instances:    435
Attributes:   17
              handicapped-infants
              water-project-cost-sharing
              adoption-of-the-budget-resolution
              physician-fee-freeze
              el-salvador-aid
              religious-groups-in-schools
              anti-satellite-test-ban
              aid-to-nicaraguan-contras
              mx-missile
              immigration
              synfuels-corporation-cutback
              education-spending
              superfund-right-to-sue
              crime
              duty-free-exports
              export-administration-act-south-africa
              Class
Test mode:    evaluate on training data


=== Clustering model (full training set) ===


kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 1510.0

Initial starting points (random):

Cluster 0: n,n,y,y,y,y,n,n,y,n,n,n,y,y,y,y,democrat
Cluster 1: n,n,y,n,y,n,y,y,y,n,n,n,n,y,n,y,democrat

Missing values globally replaced with mean/mode
```

```
Final cluster centroids:
                                                Cluster#
Attribute                               Full Data          0          1
                                         (435.0)    (214.0)    (221.0)
=============================================================================
handicapped-infants                            n          n          y
water-project-cost-sharing                     y          y          n
adoption-of-the-budget-resolution              y          n          y
physician-fee-freeze                           n          y          n
el-salvador-aid                                y          y          n
religious-groups-in-schools                    y          y          n
anti-satellite-test-ban                        y          n          y
aid-to-nicaraguan-contras                      y          n          y
mx-missile                                     y          n          y
immigration                                    y          y          y
synfuels-corporation-cutback                   n          n          n
education-spending                             n          y          n
superfund-right-to-sue                         y          y          n
crime                                          y          y          n
duty-free-exports                              n          n          y
export-administration-act-south-africa         y          y          y
Class                                 democrat republican   democrat


Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       214 ( 49%)
1       221 ( 51%)
```