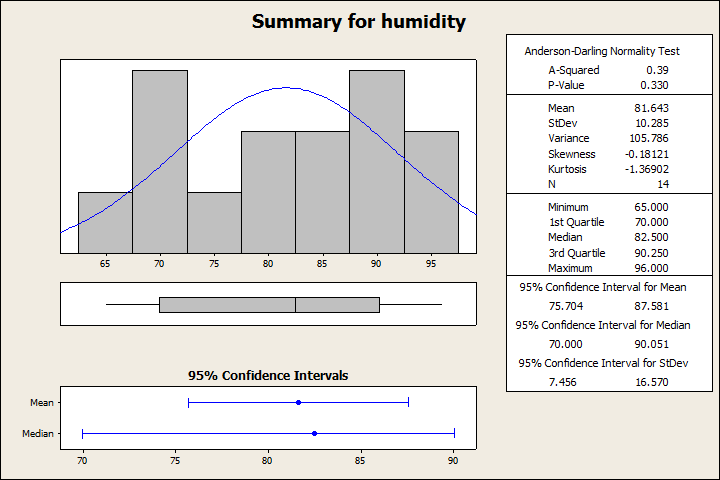
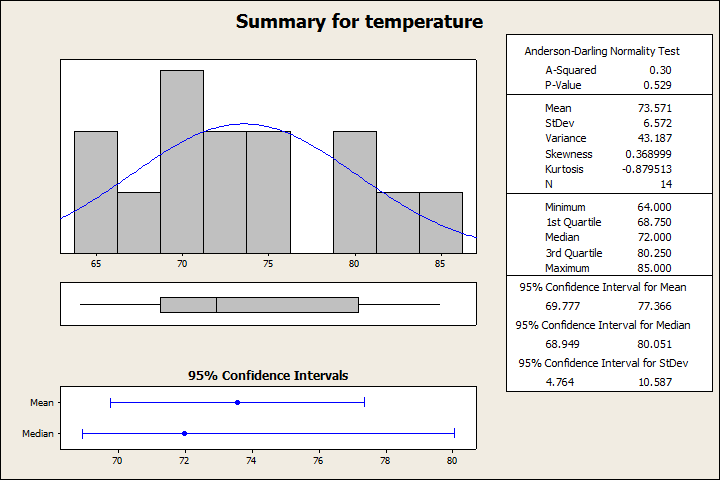
Project 2.1

Summary Stats for Weather Data Set

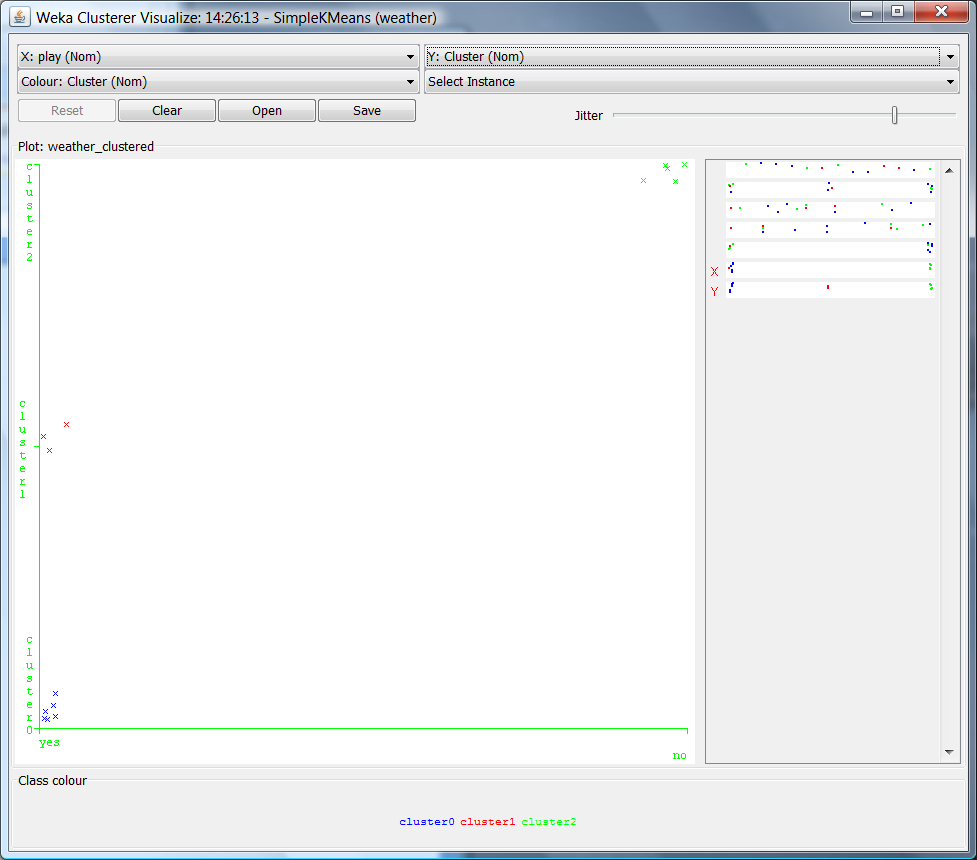
Summary stats where generated for numeric attributes Humidity and Temperature in Mini Tab 15 as given below. Humidity shows a normal distribution but is almost bimodal with prominent quartiles at 70 and 90 percent. Temperature is slightly skewed toward the left with a record number at 70 degrees.





Clusters

Clustering was run the numeric dataset using a SimpleKmeans algorithm with a k of 3. One of the cluster visualizations is given below displaying class "play" on the x axis and class "Cluster" on the Y axis. In this clustering run as seen below two the clusters are grouped according to play-yes and one is grouped according to play-no. This may suggest a division of class values that will both give an outcome of play-yes when classifying.



Classifier J48

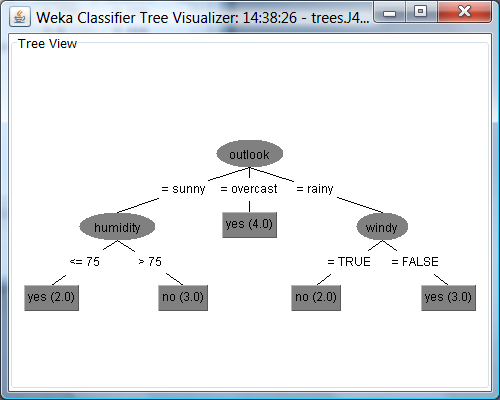
A J48 classifier was used to make a decision tree based on the numeric weather data set. Results from the J48 run are given below including a classification tree visualization. The classification tree resulted in 8 leaves with 9 out of 14 correctly classified instances.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

J48 pruned tree

------------------

outlook = sunny

| humidity <= 75: yes (2.0)

| humidity > 75: no (3.0)

outlook = overcast: yes (4.0)

outlook = rainy

| windy = TRUE: no (2.0)

| windy = FALSE: yes (3.0)

Number of Leaves : 5

Size of the tree : 8

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

Kappa statistic 0.186

Mean absolute error 0.2857

Root mean squared error 0.4818

Relative absolute error 60 %

Root relative squared error 97.6586 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.778 0.6 0.7 0.778 0.737 0.789 yes

0.4 0.222 0.5 0.4 0.444 0.789 no

Weighted Avg. 0.643 0.465 0.629 0.643 0.632 0.789

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

3 2 | b = no

Rule Based Classifiers

JRip

The JRip class uses a propositional rule learner to build a pruned or unpruned tree. Repeated incremental pruning us used in a pruned tree to produce error reduction. The algorithm can be used on many different data types and may or may not use binary splits.

JRip Results

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

JRIP rules:

===========

=> play=yes (14.0/5.0)

Number of Rules : 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.1351

Mean absolute error 0.5516

Root mean squared error 0.6149

Relative absolute error 115.8462 %

Root relative squared error 124.6318 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.889 1 0.615 0.889 0.727 0.122 yes

0 0.111 0 0 0 0.122 no

Weighted Avg. 0.571 0.683 0.396 0.571 0.468 0.122

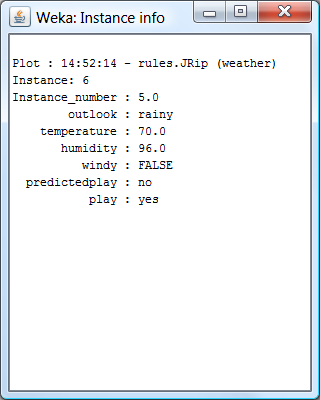
=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

5 0 | b = no

JRip results visualization given below. One of the misclassified instances is pointed out. The predicted play value was no with the actual value of yes. The particular combination of instance values shown in the call out box cannot be classified correctly with the current JRip classification.



Decision Table Classifier

The decision table class uses a simple decision table majority classifier.

=== Run information ===

Scheme: weka.classifiers.rules.DecisionTable -X 1 -R -S "weka.attributeSelection.BestFirst -D 1 -N 5"

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

Decision Table:

Number of training instances: 14

Number of Rules : 1

Non matches covered by Majority class.

Best first.

Start set: no attributes

Search direction: forward

Stale search after 5 node expansions

Total number of subsets evaluated: 13

Merit of best subset found: 64.286

Evaluation (for feature selection): CV (leave one out)

Feature set: 5

Rules:

================

play

================

yes

================

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.1351

Mean absolute error 0.4859

Root mean squared error 0.5123

Relative absolute error 102.0357 %

Root relative squared error 103.8359 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.889 1 0.615 0.889 0.727 0.367 yes

0 0.111 0 0 0 0.367 no

Weighted Avg. 0.571 0.683 0.396 0.571 0.468 0.367

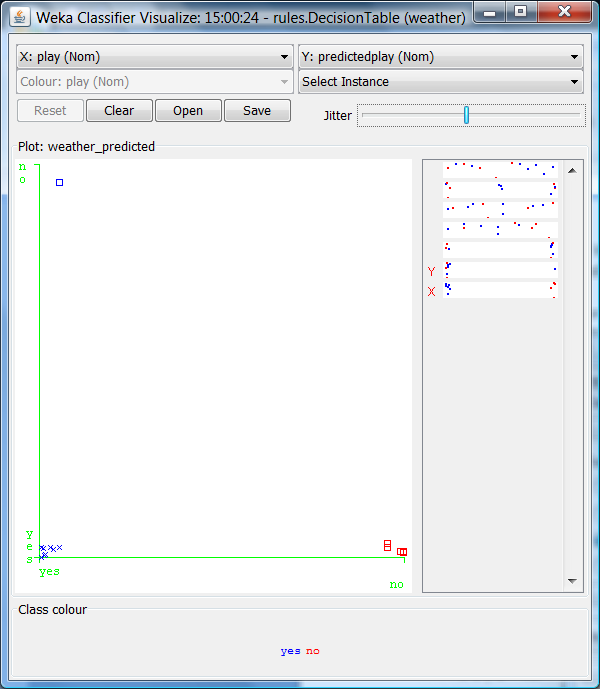
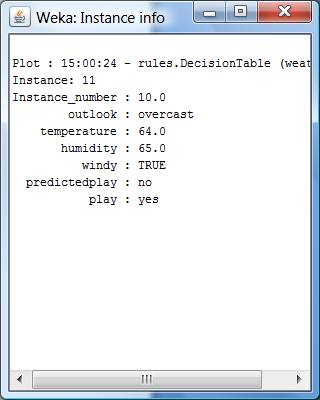
=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

5 0 | b = no

Decision Table Classifier results visualization shown below. Again in this case a play of yes was incorrectly predicted as no, but a different set of values results in this compared to the JRip class.



Ridor Class

The Ridor class is a ripple down rule learner. This algorithm generates a default rule then looks at exceptions to rule generating the "best" exceptions for each exception be minimizing the error. This performs a tree like expansion till each leaf has one default rule with no exceptions.

=== Run information ===

Scheme: weka.classifiers.rules.Ridor -F 3 -S 1 -N 2.0

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

RIpple DOwn Rule Learner(Ridor) rules

--------------------------------------

play = no (14.0/9.0)

Except (humidity <= 82.5) => play = yes (5.0/1.0) [2.0/0.0]

Total number of rules (incl. the default rule): 2

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 5 35.7143 %

Incorrectly Classified Instances 9 64.2857 %

Kappa statistic -0.4651

Mean absolute error 0.6429

Root mean squared error 0.8018

Relative absolute error 135 %

Root relative squared error 162.5137 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.556 1 0.5 0.556 0.526 0.278 yes

0 0.444 0 0 0 0.278 no

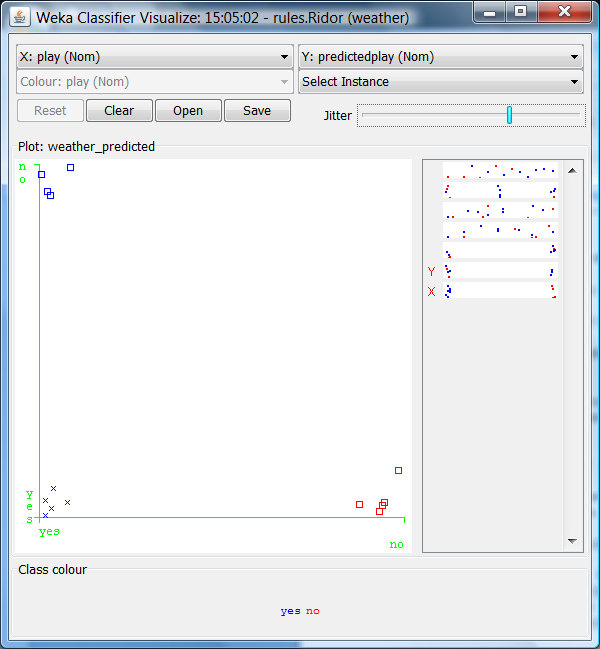
Weighted Avg. 0.357 0.802 0.321 0.357 0.338 0.278

=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

5 0 | b = no

Ridor visualization of classification matrix. In this case there were 0 correctly classified instance for the class play with a value of no.

Nearest Neighbor Classifiers

1Bk

1Bk is a simple nearest neighbor classifier that depends on user defined K nearest neighbors and window size. Multiple runs were done with the 1Bk class and the results from two runs are given below. The results from 1Bk are highly sensitive to the specified K and W (window) parameters. The 1st run with a K of 2 and a W of 2 resulted in classifying everything as play-yes. The second run with a K of 2 and W of 1 resulted in a lower % correctly classified but it did not just give one value for predicted play.

Ist Run

Scheme: weka.classifiers.lazy.IBk -K 2 -W 2 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

IB1 instance-based classifier

using 2 nearest neighbour(s) for classification

using a maximum of 2 (windowed) training instances

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

Kappa statistic 0

Mean absolute error 0.4524

Root mean squared error 0.5

Relative absolute error 95 %

Root relative squared error 101.3451 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.643 1 0.783 0.522 yes

0 0 0 0 0 0.522 no

Weighted Avg. 0.643 0.643 0.413 0.643 0.503 0.522

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

5 0 | b = no

2nd Run

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 2 -W 1 -E -I -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

=== Confusion Matrix ===

a b <-- classified as

6 3 | a = yes

3 2 | b = no

KStar

KStar is nearest neighbor classifier that use a similarity function to test instances based on similarity to those of training instances. This method results in a different confusion matrix from 1Bk class with a 50 % split on correctly classified instances.

=== Run information ===

Scheme: weka.classifiers.lazy.KStar -B 20 -M a

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 2-fold cross-validation

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

KStar Beta Verion (0.1b).

Copyright (c) 1995-97 by Len Trigg (trigg@cs.waikato.ac.nz).

Java port to Weka by Abdelaziz Mahoui (am14@cs.waikato.ac.nz).

KStar options : -B 20 -M a

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 7 50 %

Incorrectly Classified Instances 7 50 %

Kappa statistic -0.2564

Mean absolute error 0.5188

Root mean squared error 0.6668

Relative absolute error 108.9426 %

Root relative squared error 135.6609 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.778 1 0.583 0.778 0.667 0.222 yes

0 0.222 0 0 0 0.222 no

Weighted Avg. 0.5 0.722 0.375 0.5 0.429 0.222

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

5 0 | b = no

---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

NNge

NNge is a nearest neighbor classifier that uses a form of generalization. Compared to the other classification methods NNge resulted in the highest percentage, 78.57%, of correctly classified instances. Examining the rules given below the windy class is the most uncertain variable resulting in the misclassified instances.

=== Run information ===

Scheme: weka.classifiers.rules.NNge -G 5 -I 5

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

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

NNGE classifier

Rules generated :

class no IF : outlook in {rainy} ^ temperature in {mild,cool} ^ humidity in {high,normal} ^ windy in {TRUE} (2)

class yes IF : outlook in {overcast,rainy} ^ temperature in {hot,mild,cool} ^ humidity in {high,normal} ^ windy in {FALSE} (5)

class yes IF : outlook in {overcast} ^ temperature in {mild,cool} ^ humidity in {high,normal} ^ windy in {TRUE} (2)

class yes IF : outlook in {sunny} ^ temperature in {mild,cool} ^ humidity in {normal} ^ windy in {TRUE,FALSE} (2)

class no IF : outlook in {sunny} ^ temperature in {hot,mild} ^ humidity in {high} ^ windy in {TRUE,FALSE} (3)

Stat :

class yes : 3 exemplar(s) including 3 Hyperrectangle(s) and 0 Single(s).

class no : 2 exemplar(s) including 2 Hyperrectangle(s) and 0 Single(s).

Total : 5 exemplars(s) including 5 Hyperrectangle(s) and 0 Single(s).

Feature weights : [0.24674981977443894 0.029222565658954563 0.15183550136234153 0.04812703040826924]

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 11 78.5714 %

Incorrectly Classified Instances 3 21.4286 %

Kappa statistic 0.5116

Mean absolute error 0.2143

Root mean squared error 0.4629

Relative absolute error 45 %

Root relative squared error 93.8273 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.889 0.4 0.8 0.889 0.842 0.744 yes

0.6 0.111 0.75 0.6 0.667 0.744 no

Weighted Avg. 0.786 0.297 0.782 0.786 0.779 0.744

=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

2 3 | b = no