Assignment 3: Wine Data

CIS 435

Section 56

Summer Quarter

School of Continuing Studies

Northwestern University

Daniel Prusinski

Business Intelligence Data Analyst

Target Corporation

Minneapolis, MN

In Compliance with Master of Science Predictive Analytics

Wine data in Weka using J48 classifier:

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

Relation: Wine

Instances: 178

Attributes: 14

Test mode:10-fold cross-validation

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

J48 pruned tree

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

Flavanoids <= 1.57

| Color\_Intensity <= 3.8: B (13.0)

| Color\_Intensity > 3.8: C (49.0/1.0)

Flavanoids > 1.57

| Proline <= 720: B (54.0/1.0)

| Proline > 720

| | Color\_Intensity <= 3.4: B (4.0)

| | Color\_Intensity > 3.4: A (58.0)

Number of Leaves : 5

Size of the tree : 9

Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 167 93.8202 %

Incorrectly Classified Instances 11 6.1798 %

Kappa statistic 0.9058

Mean absolute error 0.0486

Root mean squared error 0.2019

Relative absolute error 11.0723 %

Root relative squared error 43.0865 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

0.983 0.034 0.935 0.983 0.959 0.977 A

0.944 0.056 0.918 0.944 0.931 0.937 B

0.875 0.008 0.977 0.875 0.923 0.946 C

Weighted Avg. 0.938 0.036 0.94 0.938 0.938 0.953

=== Confusion Matrix ===

a b c <-- classified as

58 1 0 | a = A

3 67 1 | b = B

1 5 42 | c = C

Wine data in Weka using NaiveBayesSimple:

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes

Relation: Wine

Instances: 178

Attributes: 14

Test mode:10-fold cross-validation

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

Naive Bayes Classifier

Class

Attribute A B C

(0.33) (0.4) (0.27)

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

Alcohol

mean 13.7434 12.2782 13.1537

std. dev. 0.4587 0.5351 0.5252

weight sum 59 71 48

precision 0.0304 0.0304 0.0304

Malic\_Acid

mean 2.0115 1.9329 3.3334

std. dev. 0.6824 1.0078 1.0749

weight sum 59 71 48

precision 0.0383 0.0383 0.0383

Ash

mean 2.4555 2.2451 2.4354

std. dev. 0.2253 0.3139 0.1817

weight sum 59 71 48

precision 0.024 0.024 0.024

Ash\_Alcalinity

mean 17.0506 20.2594 21.4208

std. dev. 2.5279 3.3209 2.2327

weight sum 59 71 48

precision 0.3129 0.3129 0.3129

Magnesium

mean 106.3338 94.5915 99.2981

std. dev. 10.4831 16.6495 10.8441

weight sum 59 71 48

precision 1.7692 1.7692 1.7692

Total\_Phenols

mean 2.8396 2.2618 1.681

std. dev. 0.3357 0.5412 0.3553

weight sum 59 71 48

precision 0.0302 0.0302 0.0302

Flavanoids

mean 2.983 2.0793 0.7802

std. dev. 0.3944 0.7013 0.2896

weight sum 59 71 48

precision 0.0362 0.0362 0.0362

Nonflavanoid\_Phenols

mean 0.2908 0.3646 0.4478

std. dev. 0.0701 0.1229 0.123

weight sum 59 71 48

precision 0.0139 0.0139 0.0139

Proanthocyanins

mean 1.8982 1.631 1.1518

std. dev. 0.4095 0.5992 0.4046

weight sum 59 71 48

precision 0.0317 0.0317 0.0317

Color\_Intensity

mean 5.5241 3.0796 7.3996

std. dev. 1.2265 0.9159 2.2849

weight sum 59 71 48

precision 0.0895 0.0895 0.0895

Hue

mean 1.0611 1.0559 0.6836

std. dev. 0.1151 0.2013 0.1129

weight sum 59 71 48

precision 0.016 0.016 0.016

OD280\_OD315

mean 3.1579 2.7843 1.6842

std. dev. 0.3543 0.4923 0.2688

weight sum 59 71 48

precision 0.0226 0.0226 0.0226

Proline

mean 1115.8573 519.8261 629.683

std. dev. 220.0034 154.7719 113.0791

weight sum 59 71 48

precision 11.6833 11.6833 11.6833

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 172 96.6292 %

Incorrectly Classified Instances 6 3.3708 %

Kappa statistic 0.9489

Mean absolute error 0.0217

Root mean squared error 0.1294

Relative absolute error 4.9371 %

Root relative squared error 27.6176 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

0.949 0 1 0.949 0.974 0.998 A

0.958 0.028 0.958 0.958 0.958 0.997 B

1 0.023 0.941 1 0.97 1 C

Weighted Avg. 0.966 0.017 0.967 0.966 0.966 0.998

=== Confusion Matrix ===

a b c <-- classified as

56 3 0 | a = A

0 68 3 | b = B

1. 0 48 | c = C

Wine data in Weka using IBK (kNN algorithm):

Scheme:weka.classifiers.lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: Wine

Instances: 178

Attributes: 14

Test mode:10-fold cross-validation

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

IB1 instance-based classifier

using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 169 94.9438 %

Incorrectly Classified Instances 9 5.0562 %

Kappa statistic 0.9238

Mean absolute error 0.0413

Root mean squared error 0.1821

Relative absolute error 9.3973 %

Root relative squared error 38.8682 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

1 0.042 0.922 1 0.959 0.983 A

0.873 0 1 0.873 0.932 0.941 B

1 0.031 0.923 1 0.96 0.983 C

Weighted Avg. 0.949 0.022 0.953 0.949 0.949 0.966

=== Confusion Matrix ===

a b c <-- classified as

59 0 0 | a = A

5 62 4 | b = B

0 0 48 | c = C

Wine data in Weka using jRP:

=== Run information ===

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

Relation: Wine

Instances: 178

Attributes: 14

Alcohol

Malic\_Acid

Ash

Ash\_Alcalinity

Magnesium

Total\_Phenols

Flavanoids

Nonflavanoid\_Phenols

Proanthocyanins

Color\_Intensity

Hue

OD280\_OD315

Proline

Type

Test mode:10-fold cross-validation

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

JRIP rules:

===========

(Flavanoids <= 1.39) and (Color\_Intensity >= 4) => Type=C (46.0/0.0)

(OD280\_OD315 <= 1.3) => Type=C (2.0/0.0)

(Proline >= 760) => Type=A (61.0/4.0)

=> Type=B (69.0/2.0)

Number of Rules : 4

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 167 93.8202 %

Incorrectly Classified Instances 11 6.1798 %

Kappa statistic 0.9059

Mean absolute error 0.048

Root mean squared error 0.1991

Relative absolute error 10.9333 %

Root relative squared error 42.4986 %

Total Number of Instances 178

=== Detailed Accuracy By Class ===

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

0.932 0.034 0.932 0.932 0.932 0.973 A

0.944 0.056 0.918 0.944 0.931 0.97 B

0.938 0.008 0.978 0.938 0.957 0.97 C

Weighted Avg. 0.938 0.036 0.939 0.938 0.938 0.971

=== Confusion Matrix ===

a b c <-- classified as

55 4 0 | a = A

3 67 1 | b = B

1 2 45 | c = C