Trevor Hodde

Matt Morrissey

James Forkey

Final Report

The dataset we have chosen for this project is a set of bank records that keep track of a person’s age, sex, marital status, children, income, and car. We decided that it would be interesting to determine the number of people married with and without children and see which is the larger group. Also, once we figure out these numbers, we would like to find out the number of people (married and single) with and without cars.

As for our algorithms, we have used the Naïve-Bayes, KStar, SimpleLogistic, and JRip algorithms. This way we can represent the Bayes, function, LAZY, and rules categories of algorithms.

Over the course of this project we found that not all algorithms are suited to run on every data set. For our project we found that the algorithm that was best for us was the Naïve-Bayes algorithm. After running a few tests we discovered that we could improve the accuracy of our results by modifying our attribute list. Therefore we trimmed our attribute list down to married, children, car, and PEP(Personal Equity Plan).

Using this algorithm and our reduced attribute list we discovered that, as we had suspected, married people generally have more children, cars, and are in general more financially stable. In fact, married couples have roughly twice as many children and cars as single people with children. Married couples are also roughly one and a half times more likely to have a PEP in place.

Attached are our results found by running the selected algorithms on the data set.

**Naïve-Bayes Classifier**

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: bank-weka.filters.unsupervised.attribute.Remove-R9

Instances: 300

Attributes: 8

age

sex

region

income

married

children

car

mortgage

Test mode: 10-fold cross-validation

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

Naive Bayes Classifier

Class

Attribute YES NO

(0.35) (0.65)

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

age

mean 41.6762 43.0513

std. dev. 13.6511 14.4583

weight sum 105 195

precision 1 1

sex

MALE 61.0 95.0

FEMALE 46.0 102.0

[total] 107.0 197.0

region

INNER\_CITY 53.0 86.0

RURAL 14.0 39.0

TOWN 37.0 52.0

SUBURBAN 5.0 22.0

[total] 109.0 199.0

income

mean 26611.6909 28216.1841

std. dev. 12208.1526 13280.0285

weight sum 105 195

precision 194.3675 194.3675

married

YES 69.0 135.0

NO 38.0 62.0

[total] 107.0 197.0

children

YES 54.0 119.0

NO 53.0 78.0

[total] 107.0 197.0

car

YES 54.0 95.0

NO 53.0 102.0

[total] 107.0 197.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 189 63 %

Incorrectly Classified Instances 111 37 %

Kappa statistic 0.0107

Mean absolute error 0.447

Root mean squared error 0.4831

Relative absolute error 98.1597 %

Root relative squared error 101.2642 %

Total Number of Instances 300

=== Detailed Accuracy By Class ===

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

0.086 0.077 0.375 0.086 0.14 0.549 YES

0.923 0.914 0.652 0.923 0.764 0.549 NO

Weighted Avg. 0.63 0.621 0.555 0.63 0.546 0.549

=== Confusion Matrix ===

a b <-- classified as

9 96 | a = YES

15 180 | b = NO

**Naïve-Bayes Classifier – Modified Attribute List**

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: bank-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelection.CfsSubsetEval-Sweka.attributeSelection.BestFirst -D 1 -N 5

Instances: 300

Attributes: 4

married

children

car

pep

Test mode: split 66.0% train, remainder test

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

Naive Bayes Classifier

Class

Attribute YES NO

(0.67) (0.33)

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

children

YES 114.0 59.0

NO 90.0 41.0

[total] 204.0 100.0

car

YES 101.0 48.0

NO 103.0 52.0

[total] 204.0 100.0

pep

YES 85.0 55.0

NO 119.0 45.0

[total] 204.0 100.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances 64 62.7451 %

Incorrectly Classified Instances 38 37.2549 %

Kappa statistic 0

Mean absolute error 0.4442

Root mean squared error 0.4871

Relative absolute error 98.6503 %

Root relative squared error 99.7815 %

Coverage of cases (0.95 level) 100 %

Mean rel. region size (0.95 level) 100 %

Total Number of Instances 102

=== Detailed Accuracy By Class ===

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

1 1 0.627 1 0.771 0.543 YES

0 0 0 0 0 0.543 NO

Weighted Avg. 0.627 0.627 0.394 0.627 0.484 0.543

=== Confusion Matrix ===

a b <-- classified as

64 0 | a = YES

38 0 | b = NO

***KStar Classifier***

=== Run information ===

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

Relation: bank-weka.filters.unsupervised.attribute.Remove-R9

Instances: 300

Attributes: 8

age

sex

region

income

married

children

car

mortgage

Test mode: evaluate on training data

=== 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

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances 300 100 %

Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0062

Root mean squared error 0.0421

Relative absolute error 1.4078 %

Root relative squared error 8.9716 %

Total Number of Instances 300

=== Detailed Accuracy By Class ===

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

1 0 1 1 1 1 YES

1 0 1 1 1 1 NO

Weighted Avg. 1 0 1 1 1 1

=== Confusion Matrix ===

a b <-- classified as

202 0 | a = YES

0 98 | b = NO

**SimpleLogistic Classifier**

=== Run information ===

Scheme: weka.classifiers.functions.SimpleLogistic -I 0 -M 500 -H 50 -W 0.0

Relation: bank-weka.filters.unsupervised.attribute.Remove-R9

Instances: 300

Attributes: 8

age

sex

region

income

married

children

car

mortgage

Test mode: evaluate on training data

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

SimpleLogistic:

Class 0 :

-0.03 +

[age] \* 0.01 +

[sex] \* 0.06 +

[region=INNER\_CITY] \* 0.15 +

[income] \* 0 +

[mortgage] \* 0.09

Class 1 :

0.03 +

[age] \* -0.01 +

[sex] \* -0.06 +

[region=INNER\_CITY] \* -0.15 +

[income] \* 0 +

[mortgage] \* -0.09

Time taken to build model: 0.12 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances 202 67.3333 %

Incorrectly Classified Instances 98 32.6667 %

Kappa statistic 0

Mean absolute error 0.4335

Root mean squared error 0.4651

Relative absolute error 98.4601 %

Root relative squared error 99.1606 %

Total Number of Instances 300

=== Detailed Accuracy By Class ===

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

1 1 0.673 1 0.805 0.586 YES

0 0 0 0 0 0.586 NO

Weighted Avg. 0.673 0.673 0.453 0.673 0.542 0.586

=== Confusion Matrix ===

a b <-- classified as

202 0 | a = YES

98 0 | b = NO

**JRip Classifier**

=== Run information ===

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

Relation: bank-weka.filters.unsupervised.attribute.Remove-R9

Instances: 300

Attributes: 8

age

sex

region

income

married

children

car

mortgage

Test mode: 10-fold cross-validation

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

JRIP rules:

===========

(age <= 42) and (age >= 32) => mortgage=YES (67.0/33.0)

=> mortgage=NO (233.0/71.0)

Number of Rules : 2

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 190 63.3333 %

Incorrectly Classified Instances 110 36.6667 %

Kappa statistic 0.0063

Mean absolute error 0.4557

Root mean squared error 0.4812

Relative absolute error 100.0704 %

Root relative squared error 100.8784 %

Total Number of Instances 300

=== Detailed Accuracy By Class ===

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

0.067 0.062 0.368 0.067 0.113 0.488 YES

0.938 0.933 0.651 0.938 0.769 0.488 NO

Weighted Avg. 0.633 0.628 0.552 0.633 0.539 0.488

=== Confusion Matrix ===

a b <-- classified as

7 98 | a = YES

12 183 | b = NO