

MET CS699 Course Project

Sushant Khot

Introduction:

The Goal of the Project is to build Classifier Models and Test them using the 2018 BRFSS Survey Data prepared by CDC “**project-2018-BRFSS-arthritis.arff**”. It has 11933 tuples and 108 attributes. Each tuple is a person who participated in the survey and each attribute is an answer to a survey question. The class attribute is **havarth3** and its value is either 1 or 2. The value of 1 means that the person was ever told to have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia. Otherwise, the value is 2.

The requirement is to build and test classifier models multiple times using different attribute selection methods and different classification algorithms. We have divided the Original dataset into Training (66%) “**CS699_Sushant_Khot_Project-Training_Base.arff**” and Test (34%) “**CS699_Sushant_Khot_Project-Test_Base.arff**” datasets using stratified splitting to maintain the Class distribution in Training and Test datasets.

Preprocessing:

I converted the **project-2018-BRFSS-arthritis.arff** file into .csv file and imported it into R software for further analysis and processing. The data seemed to be fairly preprocessed and did not give me much of an option other than maybe I could have removed certain columns in the dataset that are redundant or of similar in nature. There was scope of also removing attributes which may be related to the class attribute. However, there was going to be an attribute selection process which would give me the opportunity to select specific attributes based on their ranks, bestfirst etc. Hence, I skipped the step to remove any attributes from the dataset.

Below is the code and some information that I could pull using R on our dataset.

1. Import the .csv data into R.
2. I saw that missing values were listed as “?” and hence I replaced them as blank while importing.
3. I wanted to know which columns had the maximum number of blank values.

```
library(dplyr)

# setting my working directory
setwd("C:\\Sushant\\CS699\\Project_Assignment")

# importing the .csv data into a dataframe in R called "project.dataset"
project.dataset <- read.csv("project-2018-BRFSS-arthritis.arff.csv", header = TRUE, na.strings = "?")
```

```

# Tuples with atleast 1 missing value column:
sum(rowSums(is.na(project.dataset)) > 0)

## [1] 4287

# Remaining tuples with no missing values:
nrow(project.dataset) - sum(rowSums(is.na(project.dataset)) > 0)

## [1] 7646

# Number of Columns that had missing values greater than 500
sum(colSums(is.na(project.dataset)) > 500)

## [1] 10

```

So based on my analysis, out of 11933 tuples, 4287 tuples had at least one column with missing values and we were left with 7646 tuples with no missing values.

Also, the number of columns that had missing values greater than 500 were only 10 out of the 108 total attributes / columns.

We decided to not go for any pre-processing due to the ability for us to select attributes using attribute selection methods. Also, One R classifier replaces all missing values with a global constant and J48 decision tree classifier replaces all missing values with either the attribute mean or attribute mode. If we handle missing values using a certain method, it may be appropriate for some classifiers but may not be appropriate for other classifiers. So, we left the handling of missing values to individual classifier on Weka instead of handling missing values ourselves.

Attribute Selection:

We decided to use 4 Attribute selection methods namely as follows:

1. InfoGain
2. OneR
3. GainRatio
4. Correlation

Steps used for Attribute Selection are listed below:

1. We open Training Base set “CS699_Sushant_Khot_Project-Training_Base” in Weka.
2. In Weka, click “Select attributes” and choose an Attribute Selection Method. E.g., Info Gain. Here default Search Method is Ranker.
3. We made sure that we have “Use Full training sets” selected.
4. Choose the class attribute **havarth3** and click Start.

This gave us a ranked list of attributes. We selected the top 6 attributes from the Ranked list. I selected attributes which seemed different than each other. E.g., We did not select **x.ageg5yr** and **x.age.g** as we already had **x.age80** in our ranked list. Instead, we chose to select the next ranked attribute that seemed different than the age attribute to avoid selection of similar attributes that are represented differently.

We will list the Attributes selected by each of the selection methods. All Attribute Selection methods gave us Ranked attributes. Of course, with each Attribute selection, we have maintained the presence of Class attribute **havarth3**.

1. **InfoGain:**

- **x.age80** - Imputed Age value collapsed above 80
- **employ1** - Employment Status
- **diffwalk** - Difficulty Walking or Climbing Stairs
- **genhlth** - General Health
- **x.hcvu651** - Respondents aged 18-64 with health care coverage
- **physhlth** - Number of Days Physical Health Not Good In last 30 days

2. **OneR:**

- **diffwalk** - Difficulty Walking or Climbing Stairs
- **employ1** - Employment Status
- **physhlth** - Number of Days Physical Health Not Good In last 30 days
- **x.phys14d** - 3 level not good physical health status: 0 days, 1-13 days, 14-30 days
- **x.rfhlth** - Adults with good or better health
- **genhlth** - General Health overall

3. **GainRatio:**

- **diffwalk** - Difficulty Walking or Climbing Stairs

- chccopd1 - (Ever told) you have chronic obstructive pulmonary disease, emphysema or chronic bronchitis?
- diffdres - Difficulty Dressing or Bathing
- x.rfhlth - Adults with good or better health
- x.age65yr - Two-level age category
- employ1 - Employment Status

4. Correlation:

- x.age80 - Imputed Age value collapsed above 80
- diffwalk - Difficulty Walking or Climbing Stairs
- employ1 - Employment Status
- x.rfhlth - Adults with good or better health
- x.hcvu651 - Respondents aged 18-64 with health care coverage
- x.exteth3 - Adults aged 18+ that have had permanent teeth extracted

We have saved the reduced Training datasets e.g., “CS699_Sushant_Khot_Project-**Training_AS01_InfoGain**” for attributes ranked by InfoGain method.

We have also done this for the reduced Test datasets keeping the 6 attributes and class attribute and remove rest. We have saved them as e.g., “CS699_Sushant_Khot_Project-**Test_AS01_InfoGain**”. The same has been done for all Attribute Selection methods.

We will add the screenshots of the Attribute Selection method outputs:

1. InfoGain Attribute Selection Method

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluato

Choose CorrelationAttributeEval

Search Method

Choose Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode

Use full training set
 Cross-validation Folds 10
Seed 1

(Nom) havarth3

Start Stop

Result list (right-click for options)

14:33:47 - Ranker + InfoGainAttributeEval
14:48:05 - Ranker + OneRAttributeEval
14:59:59 - Ranker + GainRatioAttributeEval
15:09:16 - Ranker + CorrelationAttributeEval

Attribute selection output

```
==== Run information ====
Evaluator: weka.attributeSelection.InfoGainAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: test-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C24-L3-4-weka
Instances: 7876
Attributes: 108
[list of attributes omitted]
Evaluation mode: evaluate on all training data
```



```
==== Attribute Selection on all input data ====
Search Method:
    Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 108 havarth3):
    Information Gain Ranking Filter

Ranked attributes:
0.11056128 64 x.age80
0.11001453 66 x.ageg5yr
0.10680984 62 x.age.g
0.09485689 2 employ1
0.07832746 22 diffwalk
0.06586308 34 genhlth
0.05819653 67 x.age65yr
0.0544134 97 x.hcvu651
0.05245567 31 physhlth
0.04877981 95 x.phys14d
0.04567538 27 rmvteth4
0.04277001 20 pneuvac4
0.04152785 87 x.rfhlth
0.03578182 102 x.exteth3
0.03578182 102 x.exteth3
```

2. OneR Attribute Selection Method

The screenshot shows the Weka interface with the following configuration:

- Preprocess**, **Classify**, **Cluster**, **Associate**, **Select attributes**, **Visualize** tabs are visible at the top.
- Attribute Evaluate** panel:
 - Choose: **CorrelationAttributeEval**
- Search Method** panel:
 - Choose: **Ranker -T -1.7976931348623157E308 -N -1**
- Attribute Selection Mode** panel:
 - Use full training set
 - Cross-validation Folds 10 Seed 1
- Result list (right-click for options)**:
 - 14:33:47 - Ranker + InfoGainAttributeEval
 - 14:48:05 - Ranker + OneRAttributeEval**
 - 14:59:59 - Ranker + GainRatioAttributeEval
 - 15:09:16 - Ranker + CorrelationAttributeEval
- Attribute selection output**:

```
==== Run information ===

Evaluator: weka.attributeSelection.OneRAttributeEval -S 1 -F 10 -B 6
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: test-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C24-L3-4-weka.filters.
Instances: 7876
Attributes: 108
[list of attributes omitted]
Evaluation mode: evaluate on all training data
```



```
==== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 108 havarth3):
OneR feature evaluator.

Using 10 fold cross validation for evaluating attributes.
Minimum bucket size for OneR: 6

Ranked attributes:
72.9812 22 diffwalk
70.3276 2 employl
69.68 31 physhlth
69.5531 95 x.phys1d
69.1595 87 x.rfhlth
69.1341 34 genhlth
68.9436 46 chccopdl
68.2834 66 x.ageg5yr
68.0675 25 diffalon
67.9787 67 x.age65yr
67.8898 64 x.age80
67.8898 24 diffdres
67.8898 67 ...
```

3. GainRatio Attribute Selection Method

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluate

Choose CorrelationAttributeEval

Search Method

Choose Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode

Use full training set
 Cross-validation Folds 10 Seed 1

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 14:33:47 - Ranker + InfoGainAttributeEval
- 14:48:05 - Ranker + OneRAttributeEval
- 14:59:59 - Ranker + GainRatioAttributeEval**
- 15:09:16 - Ranker + CorrelationAttributeEval

Attribute selection output

```
==== Run information ====
Evaluator: weka.attributeSelection.GainRatioAttributeEval
Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation: test-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C24-L3-4-weka.filter
Instances: 7876
Attributes: 108
[list of attributes omitted]
Evaluation mode: evaluate on all training data
```

==== Attribute Selection on all input data ====
Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 108 havarth3):
Gain Ratio feature evaluator

Ranked attributes:

0.1148546	22	diffwalk
0.0660318	46	chccopd1
0.0639112	24	diffdres
0.056881	87	x.rfhith
0.0550061	67	x.age65yr
0.0525771	2	employ1
0.045524	25	diffalon
0.044972	62	x.age.g
0.0439987	64	x.age80
0.0424946	97	x.hcvu651
0.0423869	53	cvdcrhd4
0.0405793	36	chckdny1
0.0397399	104	x.michd
0.0391022	66	x.ageg5yr
0.0374468	13	deaf

4. Correlation Attribute Selection Method

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluate

Choose CorrelationAttributeEval

Search Method

Choose Ranker-T-1.7976931348623157E308-N-1

Attribute Selection Mode	Attribute selection output																																
<input checked="" type="radio"/> Use full training set <input type="radio"/> Cross-validation Folds 10 Seed 1 (Nom) havarth3 Start Stop	<pre>== Run information == Evaluator: weka.attributeSelection.CorrelationAttributeEval Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1 Relation: test-weka.filters.unsupervised.instance.RemoveWithValues-S0.0-C24-L3-4-weka.filters.supplier Instances: 7876 Attributes: 108 [list of attributes omitted] Evaluation mode: evaluate on all training data</pre>																																
Result list (right-click for options)	<pre>== Attribute Selection on all input data == Search Method: Attribute ranking. Attribute Evaluator (supervised, Class (nominal): 108 havarth3): Correlation Ranking Filter Ranked attributes:</pre> <table> <tbody> <tr><td>0.36757</td><td>64 x.age80</td></tr> <tr><td>0.34918</td><td>66 x.ageg5yr</td></tr> <tr><td>0.34204</td><td>22 diffwalk</td></tr> <tr><td>0.32213</td><td>2 employl</td></tr> <tr><td>0.27488</td><td>67 x.age65yr</td></tr> <tr><td>0.24454</td><td>87 x.rfhlth</td></tr> <tr><td>0.22967</td><td>97 x.hcvu651</td></tr> <tr><td>0.2123</td><td>102 x.exteth3</td></tr> <tr><td>0.20866</td><td>20 pneuvac4</td></tr> <tr><td>0.20554</td><td>46 chccopd1</td></tr> <tr><td>0.17762</td><td>95 x.phys14d</td></tr> <tr><td>0.17273</td><td>31 phylth</td></tr> <tr><td>0.17164</td><td>6 children</td></tr> <tr><td>0.16995</td><td>25 diffalon</td></tr> <tr><td>0.1671</td><td>62 x.age.g</td></tr> <tr><td>0.1653</td><td>13 deaf</td></tr> </tbody> </table>	0.36757	64 x.age80	0.34918	66 x.ageg5yr	0.34204	22 diffwalk	0.32213	2 employl	0.27488	67 x.age65yr	0.24454	87 x.rfhlth	0.22967	97 x.hcvu651	0.2123	102 x.exteth3	0.20866	20 pneuvac4	0.20554	46 chccopd1	0.17762	95 x.phys14d	0.17273	31 phylth	0.17164	6 children	0.16995	25 diffalon	0.1671	62 x.age.g	0.1653	13 deaf
0.36757	64 x.age80																																
0.34918	66 x.ageg5yr																																
0.34204	22 diffwalk																																
0.32213	2 employl																																
0.27488	67 x.age65yr																																
0.24454	87 x.rfhlth																																
0.22967	97 x.hcvu651																																
0.2123	102 x.exteth3																																
0.20866	20 pneuvac4																																
0.20554	46 chccopd1																																
0.17762	95 x.phys14d																																
0.17273	31 phylth																																
0.17164	6 children																																
0.16995	25 diffalon																																
0.1671	62 x.age.g																																
0.1653	13 deaf																																

Classifier Algorithms:

The 6 Models that have used in our project are as follows:

1. Naïve Bayes
2. J48 – Decision tree
3. K-Nearest Neighbors (KNN with K = 7)
4. MultilayerPerceptron
5. Logistic
6. Random Forest

We ran the 6 Classifier Models each on the 4 reduced Test datasets. The steps are listed below:

1. Open the reduced Training set e.g., CS699_Sushant_Khot_Project-
Training_AS01_InfoGain.
2. Click Classify and Choose a Classifier e.g., Naïve Bayes.
3. Default values set for Cross validation select class attribute.
4. Click Start.
5. Training Model is created.
6. To TEST the Model, choose “supplied test set” and select the reduced test dataset e.g.,
“CS699_Sushant_Khot_Project-**Test_AS01_InfoGain**”
7. Re-run the Model. This is the Model running on the reduced Test dataset.

Repeat Steps 1 to 7 for remaining 5 other Classifier Models like J48, KNN etc. on the reduced Attribute selected reduced Test datasets.

Screenshots of Test Results for the 24 Models:

1. Naïve Bayes:

a. Training the Model: Naïve Bayes (NB) - AS01 InfoGain

The screenshot shows the Weka interface for training a Naïve Bayes classifier. The top menu bar includes Preprocess, Classify, Cluster, Associate, Select attributes, and Visualize. The 'Classifier' tab is selected. Under 'Choose', 'NaiveBayes' is selected. In the 'Test options' section, 'Cross-validation' is chosen with 'Folds' set to 10. The 'Classifier output' pane displays statistical information for two datasets: x.age60v and x.hcvu651. The 'Result list' pane shows the command used (20:14:34 - bayes.NaiveBayes), the start time, and the time taken to build the model (0.04 seconds). The summary statistics include correctly classified instances (5778, 73.3621 %), incorrectly classified instances (2098, 26.6379 %), Kappa statistic (0.4157), Mean absolute error (0.294), Root mean squared error (0.4405), Relative absolute error (65.7461 %), Root relative squared error (93.1609 %), and Total Number of Instances (7876). Detailed accuracy by class is provided for two classes (2 and 1), along with a confusion matrix.

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

20:14:34 - bayes.NaiveBayes

Time taken to build model: 0.04 seconds

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

	Correctly Classified Instances	73.3621 %
Incorrectly Classified Instances	2098	26.6379 %
Kappa statistic	0.4157	
Mean absolute error	0.294	
Root mean squared error	0.4405	
Relative absolute error	65.7461 %	
Root relative squared error	93.1609 %	
Total Number of Instances	7876	

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.779	0.355	0.812	0.779	0.795	0.416	0.793	0.879	2
1	0.645	0.221	0.598	0.645	0.620	0.416	0.793	0.633	1
Weighted Avg.	0.734	0.310	0.739	0.734	0.736	0.416	0.793	0.796	

==== Confusion Matrix ====

	a	b	-- classified as
4063	1155		a = 2
943	1715		b = 1

On Test: Naïve Bayes - AS01 InfoGain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

Use training set
 Supplied test set [Set...](#)
 Cross-validation Folds 10
 Percentage split % 66
[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes**

Classifier output

```
weight sum      5210    2050
precision          1        1

x.hcvu651
1            3323.0  1081.0
9            1418.0   1466.0
2            480.0    114.0
[total]       5221.0  2661.0
```

Time taken to build model: 0.01 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.09 seconds

==== Summary ===

	Correctly Classified Instances	2949	72.6892 %
Incorrectly Classified Instances	1108	27.3108 %	
Kappa statistic	0.4006		
Mean absolute error	0.2938		
Root mean squared error	0.4396		
Relative absolute error	65.7054 %		
Root relative squared error	92.9735 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.774	0.366	0.806	0.774	0.790	0.401	0.797	0.882	2
	0.634	0.226	0.588	0.634	0.610	0.401	0.797	0.648	1
Weighted Avg.	0.727	0.319	0.733	0.727	0.729	0.401	0.797	0.803	

==== Confusion Matrix ===

	a	b	<- classified as
2081	607		a = 2
501	868		b = 1

b. Training Model: NB - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options	Classifier output																																																																
<input type="radio"/> Use training set <input type="radio"/> Supplied test set Set... <input checked="" type="radio"/> Cross-validation Folds 10 <input type="radio"/> Percentage split % 66 More options...	<pre> 1 4504.0 1700.0 2 644.0 873.0 9 13.0 8.0 [total] 5221.0 2661.0 x.phys14d 1 3633.0 1255.0 2 1104.0 662.0 9 81.0 92.0 3 404.0 653.0 [total] 5222.0 2662.0 </pre>																																																																
(Nom) havarth3																																																																	
Start																																																																	
Stop																																																																	
Result list (right-click for options)																																																																	
<pre> 20:14:34 - bayes.NaiveBayes 20:16:27 - bayes.NaiveBayes 20:22:35 - trees.J48 20:27:39 - trees.J48 20:30:49 - lazy.IBk 20:34:52 - lazy.IBk 20:37:48 - functions.MultilayerPerceptron 20:41:54 - functions.MultilayerPerceptron 20:54:41 - bayes.NaiveBayes </pre>																																																																	
<p>Time taken to build model: 0.01 seconds</p> <p>==== Stratified cross-validation ===</p> <p>==== Summary ===</p> <table border="1"> <thead> <tr> <th></th> <th>Correctly Classified Instances</th> <th>72.2956 %</th> </tr> </thead> <tbody> <tr> <td>Incorrectly Classified Instances</td> <td>2182</td> <td>27.7044 %</td> </tr> <tr> <td>Kappa statistic</td> <td>0.3514</td> <td></td> </tr> <tr> <td>Mean absolute error</td> <td>0.3167</td> <td></td> </tr> <tr> <td>Root mean squared error</td> <td>0.4587</td> <td></td> </tr> <tr> <td>Relative absolute error</td> <td>70.8117 %</td> <td></td> </tr> <tr> <td>Root relative squared error</td> <td>97.0079 %</td> <td></td> </tr> <tr> <td>Total Number of Instances</td> <td>7876</td> <td></td> </tr> </tbody> </table> <p>==== Detailed Accuracy By Class ===</p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>MCC</th> <th>ROC Area</th> <th>PRC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.837</td> <td>0.502</td> <td>0.766</td> <td>0.837</td> <td>0.800</td> <td>0.355</td> <td>0.741</td> <td>0.828</td> <td>2</td> </tr> <tr> <td>1</td> <td>0.498</td> <td>0.163</td> <td>0.610</td> <td>0.498</td> <td>0.548</td> <td>0.355</td> <td>0.741</td> <td>0.592</td> <td>1</td> </tr> <tr> <td>Weighted Avg.</td> <td>0.723</td> <td>0.387</td> <td>0.713</td> <td>0.723</td> <td>0.715</td> <td>0.355</td> <td>0.741</td> <td>0.749</td> <td></td> </tr> </tbody> </table> <p>==== Confusion Matrix ===</p> <pre> a b <- classified as 4370 848 a = 2 1334 1324 b = 1 </pre>			Correctly Classified Instances	72.2956 %	Incorrectly Classified Instances	2182	27.7044 %	Kappa statistic	0.3514		Mean absolute error	0.3167		Root mean squared error	0.4587		Relative absolute error	70.8117 %		Root relative squared error	97.0079 %		Total Number of Instances	7876			TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	0	0.837	0.502	0.766	0.837	0.800	0.355	0.741	0.828	2	1	0.498	0.163	0.610	0.498	0.548	0.355	0.741	0.592	1	Weighted Avg.	0.723	0.387	0.713	0.723	0.715	0.355	0.741	0.749	
	Correctly Classified Instances	72.2956 %																																																															
Incorrectly Classified Instances	2182	27.7044 %																																																															
Kappa statistic	0.3514																																																																
Mean absolute error	0.3167																																																																
Root mean squared error	0.4587																																																																
Relative absolute error	70.8117 %																																																																
Root relative squared error	97.0079 %																																																																
Total Number of Instances	7876																																																																
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class																																																								
0	0.837	0.502	0.766	0.837	0.800	0.355	0.741	0.828	2																																																								
1	0.498	0.163	0.610	0.498	0.548	0.355	0.741	0.592	1																																																								
Weighted Avg.	0.723	0.387	0.713	0.723	0.715	0.355	0.741	0.749																																																									

On Test: NB - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

Classifier output

[total] 5221.0 2662.0

x.phys14d

1	3633.0	1255.0	
2	1104.0	662.0	
9	81.0	92.0	
3	404.0	653.0	
[total]	5222.0	2662.0	

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes

Time taken to build model: 0.01 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.02 seconds

==== Summary ===

Correctly Classified Instances	2968	73.1575 %	
Incorrectly Classified Instances	1089	26.8425 %	
Kappa statistic	0.3647		
Mean absolute error	0.3124		
Root mean squared error	0.4541		
Relative absolute error	69.8604 %		
Root relative squared error	96.0272 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.855	0.510	0.767	0.855	0.808	0.808	0.371	0.746	0.827	2
0.490	0.145	0.632	0.490	0.552	0.552	0.371	0.746	0.601	1
Weighted Avg.	0.732	0.387	0.721	0.732	0.722	0.371	0.746	0.751	

==== Confusion Matrix ===

	a	b	-- classified as
2297	391		a = 2
698	671		b = 1

c. Training Model: NB - AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

(Nom) havarth3

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes**

Classifier output

```
x.age0y1
1      3821.0 1199.0
2      1282.0 1418.0
3      118.0   44.0
[total] 5221.0 2661.0

x.rfhlth
1      4564.0 1780.0
2      644.0   873.0
9      13.0    8.0
[total] 5221.0 2661.0
```

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances 5746 72.9558 %
Incorrectly Classified Instances 2130 27.0442 %
Kappa statistic 0.3705
Mean absolute error 0.3141
Root mean squared error 0.4356
Relative absolute error 70.2422 %
Root relative squared error 92.1131 %
Total Number of Instances 7876

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.837	0.481	0.774	0.837	0.804	0.373	0.756	0.824	2	
0.519	0.163	0.618	0.519	0.564	0.373	0.756	0.617	1	
Weighted Avg.	0.730	0.374	0.721	0.730	0.723	0.373	0.756	0.754	

==== Confusion Matrix ====

		a	b	<- classified as
		4366	852	a = 2
		1278	1380	b = 1

On Test: NB - AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

Classifier output

```
x.rfhlt
 1           110.0   44.0
 [total]      5221.0 2661.0

x.rfhlt
 1           4564.0 1780.0
 2           644.0   873.0
 9            13.0    8.0
 [total]      5221.0 2661.0
```

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes

Time taken to build model: 0 seconds

== Evaluation on test set ==

Time taken to test model on supplied test set: 0.02 seconds

== Summary ==

Correctly Classified Instances	3001	73.9709 %
Incorrectly Classified Instances	1056	26.0291 %
Kappa statistic	0.3626	
Mean absolute error	0.3147	
Root mean squared error	0.4374	
Relative absolute error	70.3834 %	
Root relative squared error	92.4982 %	
Total Number of Instances	4057	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.894	0.562	0.757	0.894	0.820	0.379	0.750	0.828	2
1	0.438	0.106	0.677	0.438	0.531	0.379	0.750	0.614	1
Weighted Avg.	0.740	0.409	0.730	0.740	0.723	0.379	0.758	0.756	

== Confusion Matrix ==

a	b	<- classified as
2402	286	a = 2
770	599	b = 1

d. Training Model: NB - AS04 Correlation

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) havarth3

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes**
- 21:04:32 - bayes.NaiveBayes

Classifier output

```
x.hcvu051
1      3323.0 1081.0
9      1418.0 1466.0
2      480.0   114.0
[total] 5221.0 2661.0

x.exteth3
1      3068.0 941.0
2      2063.0 1641.0
9      90.0    79.0
[total] 5221.0 2661.0
```

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances	5671	72.0036 %
Incorrectly Classified Instances	2205	27.9964 %
Kappa statistic	0.3972	
Mean absolute error	0.3027	
Root mean squared error	0.4518	
Relative absolute error	67.6811 %	
Root relative squared error	95.5572 %	
Total Number of Instances	7876	

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
a	0.749	0.336	0.814	0.749	0.780	0.400	0.780	0.870	2
b	0.664	0.251	0.574	0.664	0.616	0.400	0.780	0.618	1
Weighted Avg.	0.720	0.307	0.733	0.720	0.724	0.400	0.780	0.785	

==== Confusion Matrix ====

a	b	<- classified as
3906	1312	a = 2
893	1765	b = 1

On Test: NB - AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes**

Classifier output

	2	400.0	114.0
[total]	5221.0	2661.0	
x.exteth3			
1	3068.0	941.0	
2	2063.0	1641.0	
9	90.0	79.0	
[total]	5221.0	2661.0	

Time taken to build model: 0.01 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.03 seconds

==== Summary ===

	Correctly Classified Instances	2885	71.1117 %
Incorrectly Classified Instances	1172	28.8883 %	
Kappa statistic	0.3737		
Mean absolute error	0.3031		
Root mean squared error	0.4522		
Relative absolute error	67.7828 %		
Root relative squared error	95.6377 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.749	0.364	0.802	0.749	0.775	0.375	0.782	0.871	2
	0.636	0.251	0.564	0.636	0.598	0.375	0.782	0.627	1
Weighted Avg.	0.711	0.326	0.721	0.711	0.715	0.375	0.782	0.789	

==== Confusion Matrix ===

	a	b	<-- classified as
2014	674	1	a = 2
498	871	1	b = 1

2. J48

a. Training Model: J48 - AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48-C 0.25-M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48

Classifier output

```

physhlth > 77.2 (1525.29/502.11)
| diffwalk = 1
| | physhlth <= 77: 1 (814.21/192.56)
| | physhlth > 77
| | | x.age80 <= 51: 2 (20.39/4.63)
| | | x.age80 > 51: 1 (328.81/108.73)
| diffwalk = 7: 1 (19.49/7.32)
| diffwalk = 9: 2 (6.16/2.05)

```

Number of Leaves : 76

size of the tree : 124

Time taken to build model: 0.11 seconds

==== Stratified cross-validation ===

==== Summary ===

	Correctly Classified Instances	73.9843 %
Incorrectly Classified Instances	2049	26.0157 %
Kappa statistic	0.3732	
Mean absolute error	0.3479	
Root mean squared error	0.4236	
Relative absolute error	77.7932 %	
Root relative squared error	89.5886 %	
Total Number of Instances	7876	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.878	0.532	0.764	0.878	0.817	0.384	0.764	0.837	2	
0.468	0.122	0.662	0.468	0.548	0.384	0.764	0.600	1	
Weighted Avg.	0.740	0.394	0.730	0.740	0.727	0.384	0.764	0.757	

==== Confusion Matrix ===

		a b <- classified as
4583	635	a = 2
1414	1244	b = 1

On Test: J48 - AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48

Classifier output

```
|   |   physician > /|\n|   |   |   x.age80 <= 51: 2 (20.39/4.63)\n|   |   |   x.age80 > 51: 1 (328.81/108.73)\n|   |   diffwalk = 7: 1 (19.49/7.32)\n|   |   diffwalk = 9: 2 (6.16/2.05)\n\nNumber of Leaves : 76\n\nSize of the tree : 124\n\nTime taken to build model: 0.07 seconds\n\n==== Evaluation on test set ====\n\nTime taken to test model on supplied test set: 0.01 seconds\n\n==== Summary ====\n\nCorrectly Classified Instances      3054          75.2773 %\nIncorrectly Classified Instances    1003          24.7227 %\nKappa statistic                      0.4064\nMean absolute error                  0.3395\nRoot mean squared error              0.4175\nRelative absolute error              75.9264 %\nRoot relative squared error         88.3007 %\nTotal Number of Instances            4057\n\n==== Detailed Accuracy By Class ====\n\n          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class\n          0.885    0.506    0.774    0.885    0.826    0.417    0.777    0.841    2\n          0.494    0.115    0.606    0.494    0.574    0.417    0.777    0.611    1\nWeighted Avg.    0.753    0.374    0.744    0.753    0.741    0.417    0.777    0.764\n\n==== Confusion Matrix ====\n\n      a      b  <- classified as\n2378  310 |  a = 2\n693   676 |  b = 1
```

b. Training Model: J48 - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48**

Classifier output

```

| | | x.physica = 3: 1 (34.54/10.54)
| | | genhlth = 1: 2 (17.05/5.33)
| | | genhlth = 5: 1 (46.3/8.3)
| | | genhlth = 9: 2 (1.0)
| | | genhlth = 7: 1 (1.0)
| | employl > 6: 1 (935.46/238.31)
| | diffwalk = 7: 1 (23.63/7.47)
| | diffwalk = 9: 2 (6.16/2.04)

Number of Leaves : 27

Size of the tree : 40

Time taken to build model: 0.18 seconds

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

Correctly Classified Instances      5792          73.5399 %
Incorrectly Classified Instances   2084          26.4601 %
Kappa statistic                   0.3557
Mean absolute error               0.3634
Root mean squared error           0.428
Relative absolute error            81.2724 %
Root relative squared error      90.519 %
Total Number of Instances         7876

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

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.885    0.558    0.757     0.885    0.816     0.370   0.723    0.791     2
          0.442    0.115    0.662     0.442    0.530     0.370   0.723    0.587     1
Weighted Avg.    0.735    0.409    0.725     0.735    0.719     0.370   0.723    0.722

==== Confusion Matrix ====

      a     b  <- classified as
4618  600 |  a = 2
1484 1174 |  b = 1

```

On Test: J48 - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48-C 0.25-M 2

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48

Classifier output

```
|   |   genmrth = 9: 2 (1.0)
|   |   genhlth = 7: 1 (1.0)
|   employl > 6: 1 (935.46/238.31)
diffwalk = 7: 1 (23.63/7.47)
diffwalk = 9: 2 (6.16/2.04)

Number of Leaves :      27

Size of the tree :      40

Time taken to build model: 0.05 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances      3013          74.2667 %
Incorrectly Classified Instances   1044          25.7333 %
Kappa statistic                   0.3674
Mean absolute error               0.3614
Root mean squared error           0.4262
Relative absolute error            80.8208 %
Root relative squared error       90.1454 %
Total Number of Instances         4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.900    0.565    0.758     0.900    0.822     0.386   0.730    0.795     2
          0.435    0.100    0.688     0.435    0.533     0.386   0.730    0.568     1
Weighted Avg.    0.743    0.408    0.734     0.743    0.725     0.386   0.730    0.718

==== Confusion Matrix ====

      a     b  <- classified as
2418  270 |  a = 2
    774  595 |  b = 1
```

c. Training Model: J48 - AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25-M 2**

Test options

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

(Nom) havarth3

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48**

Classifier output

```

chccopar = 2
|   x.rfhlth = 1: 1 (10.49/2.4)
|   x.rfhlth = 2: 2 (6.11/2.05)
|   x.rfhlth = 9: 2 (1.0)
chccopd1 = 1: 1 (6.03/0.01)
chccopd1 = 7: 1 (0.0)
chccopd1 = 9: 2 (0.0)
diffwalk = 9: 2 (6.16/2.04)

Number of Leaves : 26
size of the tree : 38

Time taken to build model: 0.05 seconds

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

Correctly Classified Instances      5786          73.4637 %
Incorrectly Classified Instances  2090          26.5363 %
Kappa statistic                   0.3436
Mean absolute error               0.3626
Root mean squared error           0.4275
Relative absolute error            81.0916 %
Root relative squared error       90.4039 %
Total Number of Instances         7876

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

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.899    0.589    0.750     0.899    0.818     0.364   0.728    0.798     2
          0.411    0.101    0.676     0.411    0.511     0.364   0.728    0.590     1
Weighted Avg.    0.735    0.424    0.725     0.735    0.714     0.364   0.728    0.728

==== Confusion Matrix ====

      a     b  <- classified as
4693  525 |   a = 2
1565 1093 |   b = 1

```

On Test: J48 – AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48

Classifier output

```
| X.TINITH = 9. 2 (1.0)
| chccopd1 = 1: 1 (6.03/0.01)
| chccopd1 = 7: 1 (0.0)
| chccopd1 = 9: 2 (0.0)
diffwalk = 9: 2 (6.16/2.04)

Number of Leaves : 26

Size of the tree : 38

Time taken to build model: 0.04 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances 2985 73.5765 %
Incorrectly Classified Instances 1072 26.4235 %
Kappa statistic 0.3431
Mean absolute error 0.3635
Root mean squared error 0.4275
Relative absolute error 81.2841 %
Root relative squared error 90.4019 %
Total Number of Instances 4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.905   0.596   0.749   0.905   0.819   0.365   0.727   0.793   2
          0.404   0.095   0.684   0.404   0.508   0.365   0.727   0.567   1
Weighted Avg.  0.736   0.427   0.727   0.736   0.714   0.365   0.727   0.717

==== Confusion Matrix ===

      a     b  <- classified as
2432  256 |    a = 2
  816  553 |    b = 1
```

d. Training Model: J48 - AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48-C 0.25-M 2

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48

Classifier output

```

diffwalk =
|   |   x.rfhlth = 1: 1 (10.37/2.26)
|   |   x.rfhlth = 2
|   |   |   x.hcvu651 = 1: 1 (2.04/0.03)
|   |   |   x.hcvu651 = 9: 2 (6.07/2.05)
|   |   |   x.hcvu651 = 2: 1 (0.01/0.0)
|   |   x.rfhlth = 9: 2 (1.0)
|   diffwalk = 9: 2 (6.16/2.05)

```

Number of Leaves : 35

Size of the tree : 58

Time taken to build model: 0.07 seconds

==== Stratified cross-validation ===

==== Summary ===

	Correctly Classified Instances	73.6668 %
Incorrectly Classified Instances	2074	26.3332 %
Kappa statistic	0.3459	
Mean absolute error	0.3547	
Root mean squared error	0.4231	
Relative absolute error	79.3283 %	
Root relative squared error	89.4752 %	
Total Number of Instances	7876	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.905	0.593	0.750	0.905	0.820	0.368	0.763	0.842	2
	0.407	0.095	0.685	0.407	0.510	0.368	0.763	0.601	1
Weighted Avg.	0.737	0.425	0.728	0.737	0.715	0.368	0.763	0.761	

==== Confusion Matrix ===

a	b		-- classified as
4721	497		a = 2
1577	1081		b = 1

On Test: J48 – AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48

Classifier output

```
| | | x.hcvu651 = 1: 1 (2.04/0.05)
| | | x.hcvu651 = 9: 2 (6.07/2.05)
| | | x.hcvu651 = 2: 1 (0.01/0.0)
| | | x.rfhlth = 9: 2 (1.0)
| | diffwalk = 9: 2 (6.16/2.05)

Number of Leaves : 35

Size of the tree : 58

Time taken to build model: 0.06 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances 3002 73.9956 %
Incorrectly Classified Instances 1055 26.0044 %
Kappa statistic 0.3524
Mean absolute error 0.3489
Root mean squared error 0.4202
Relative absolute error 78.0283 %
Root relative squared error 88.8734 %
Total Number of Instances 4057

==== Detailed Accuracy By Class ===



|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| 0.910         | 0.593   | 0.751   | 0.910     | 0.823  | 0.376     | 0.767 | 0.834    | 0.2      |       |
| 0.407         | 0.090   | 0.696   | 0.407     | 0.514  | 0.376     | 0.767 | 0.591    | 0.1      |       |
| Weighted Avg. | 0.740   | 0.423   | 0.732     | 0.740  | 0.718     | 0.376 | 0.767    | 0.752    |       |



==== Confusion Matrix ===



```
a b <- classified as
2445 243 | a = 2
812 557 | b = 1
```


```

3. KNN (K = 7)

a. Training Model: KNN - AS01 Info Gain

The screenshot shows the Weka interface with the following configuration and output:

Classifier Selection: IBk-K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \\"weka.core.EuclideanDistance -R first-last\\"", chosen from the Choose dropdown.

Test options:

- Use training set (radio button)
- Supplied test set (radio button)
- Cross-validation** (radio button) Folds: 10
- Percentage split (radio button) %: 66

Result list (right-click for options):

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk

Classifier output:

```

uirrwalk
physlth
genlth
x.age80
x.hcvu651
havarth3
Test mode: 10-fold cross-validation

== Classifier model (full training set) ==

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      5735           72.8162 %
Incorrectly Classified Instances   2141            27.1838 %
Kappa statistic                      0.3543
Mean absolute error                  0.3377
Root mean squared error              0.4328
Relative absolute error              75.5119 %
Root relative squared error        91.5333 %
Total Number of Instances          7876

== Detailed Accuracy By Class ==

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.856    0.522    0.763     0.856    0.807     0.361   0.759    0.851     2
          0.478    0.144    0.628     0.478    0.543     0.361   0.759    0.599     1
Weighted Avg.    0.728    0.395    0.717     0.728    0.718     0.361   0.759    0.766

== Confusion Matrix ==

  a   b   <- classified as
4465 753 |   a = 2
1388 1270 |   b = 1

```

On Test: KNN - AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \weka.core.EuclideanDistance -R first-last"

Test options

Use training set
 Supplied test set [Set...](#)
 Cross-validation Folds 10
 Percentage split % 66
[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk

Classifier output

```
x.ageov
x.hcvu651
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

==== Classifier model (full training set) ====
IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

==== Evaluation on test set ====
Time taken to test model on supplied test set: 1.59 seconds

==== Summary ====
Correctly Classified Instances 2973 73.2807 %
Incorrectly Classified Instances 1084 26.7193 %
Kappa statistic 0.3608
Mean absolute error 0.3319
Root mean squared error 0.4292
Relative absolute error 74.2293 %
Root relative squared error 90.7631 %
Total Number of Instances 4057

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


|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| 0.866         | 0.529   | 0.763   | 0.866     | 0.811  | 0.369     | 0.766 | 0.846    | 0.2      |       |
| 0.471         | 0.134   | 0.642   | 0.471     | 0.543  | 0.369     | 0.766 | 0.606    | 0.1      |       |
| Weighted Avg. | 0.733   | 0.396   | 0.722     | 0.733  | 0.721     | 0.369 | 0.766    | 0.765    |       |



==== Confusion Matrix ====
a b <- classified as
2328 360 | a = 2
724 645 | b = 1


```

b. Training Model: KNN - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **IBk-K 7-W 0-A "weka.core.neighboursearch.LinearNNSearch-A \weka.core.EuclideanDistance -R first-last"**

Test options

- Use training set
- Supplied test set **Set...**
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start **Stop**

Result list (right-click for options)

- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk

Classifier output

```

girth
phylth
genhlth
x.rfhlth
x.phys14d
havarth3
Test mode: 10-fold cross-validation

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

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      5723          72.6638 %
Incorrectly Classified Instances   2153          27.3362 %
Kappa statistic                   0.3411
Mean absolute error               0.3543
Root mean squared error           0.4298
Relative absolute error            79.2176 %
Root relative squared error       90.8854 %
Total Number of Instances         7876

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.869    0.552    0.755     0.869    0.808     0.351   0.750    0.826     2
          0.448    0.131    0.635     0.448    0.525     0.351   0.750    0.600     1
Weighted Avg.    0.727    0.410    0.715     0.727    0.713     0.351   0.750    0.750     0.750

==== Confusion Matrix ===

  a    b  <- classified as
4533  685 |  a = 2
1468 1190 |  b = 1

```

On Test: KNN – AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk

Classifier output

```
x.titanic
x.phys14d
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

== Classifier model (full training set) ==

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0.01 seconds

== Evaluation on test set ==

Time taken to test model on supplied test set: 1.69 seconds

== Summary ==

Correctly Classified Instances      2987           73.6258 %
Incorrectly Classified Instances   1070            26.3742 %
Kappa statistic                      0.3472
Mean absolute error                  0.3522
Root mean squared error              0.4256
Relative absolute error              78.7673 %
Root relative squared error        90.008 %
Total Number of Instances          4057

== Detailed Accuracy By Class ==

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.901    0.587    0.751    0.901    0.819     0.368   0.757    0.837     2
          0.413    0.099    0.680    0.413    0.514     0.368   0.757    0.608     1
Weighted Avg.    0.736    0.423    0.727    0.736    0.716     0.368   0.757    0.760

== Confusion Matrix ==

a     b  <- classified as
2422  266 |  a = 2
     804  565 |  b = 1
```

c. Training Model: KNN - AS03 Gain Ratio

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: IBk-K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options

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

(Nom) havarth3

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk

Classifier output

```

havarth3
diffdres
chccopd1
x.age65yr
x.rfhlth
havarth3
Test mode: 10-fold cross-validation

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

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances      5768          73.2351 %
Incorrectly Classified Instances   2108          26.7649 %
Kappa statistic                   0.3426
Mean absolute error               0.3526
Root mean squared error           0.4232
Relative absolute error            78.849 %
Root relative squared error       89.495 %
Total Number of Instances         7876

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

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.891    0.579    0.751    0.891    0.815    0.359    0.755    0.829    2
          0.421    0.109    0.663    0.421    0.515    0.359    0.755    0.603    1
Weighted Avg.    0.732    0.420    0.721    0.732    0.714    0.359    0.755    0.753

==== Confusion Matrix ===

     a     b  <- classified as
4649  569 |  a = 2
1539 1119 |  b = 1

```

On Test: KNN - AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose IBk -K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options Classifier output

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk

x.age0jy1
x.rfhlth
havarth3

Test mode: user supplied test set: size unknown (reading incrementally)

==== Classifier model (full training set) ====
IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

==== Evaluation on test set ====
Time taken to test model on supplied test set: 1.69 seconds

==== Summary ====
Correctly Classified Instances 2987 73.6258 %
Incorrectly Classified Instances 1070 26.3742 %
Kappa statistic 0.3496
Mean absolute error 0.3484
Root mean squared error 0.4185
Relative absolute error 77.9041 %
Root relative squared error 88.5026 %
Total Number of Instances 4057

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.898	0.581	0.752	0.898	0.819	0.368	0.771	0.844	2
1	0.419	0.102	0.676	0.419	0.518	0.368	0.771	0.617	1
Weighted Avg.	0.736	0.419	0.727	0.736	0.717	0.368	0.771	0.767	

==== Confusion Matrix ====
a b <- classified as
2413 275 | a = 2
795 574 | b = 1

d. Training Model: KNN - AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **IBk-K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last!"**

Test options

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

(Nom) havarth3

Result list (right-click for options)

- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk

Classifier output

```

diffwark
x.age80
x.rfhlth
x.hcvu651
x.exteth3
havarth3
Test mode: 10-fold cross-validation

== Classifier model (full training set) ==

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

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

Correctly Classified Instances      5656           71.8131 %
Incorrectly Classified Instances   2220           28.1869 %
Kappa statistic                      0.3196
Mean absolute error                  0.3471
Root mean squared error              0.4317
Relative absolute error              77.6223 %
Root relative squared error        91.2984 %
Total Number of Instances          7876

== Detailed Accuracy By Class ==

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
                           0.864    0.568    0.749    0.864    0.802    0.330    0.755    0.848    2
                           0.432    0.136    0.618    0.432    0.509    0.330    0.755    0.594    1
Weighted Avg.       0.718    0.422    0.705    0.718    0.703    0.330    0.755    0.762

== Confusion Matrix ==

     a     b  <- classified as
4507  711 |   a = 2
1509 1149 |   b = 1

```

On Test: KNN - AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **IBk -K 7 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \\"weka.core.EuclideanDistance -R first-last\\""**

Test options

- Use training set
- Supplied test set [Set...](#)
- Cross-validation Folds 10
- Percentage split % 66

[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk
- 21:36:54 - lazy.IBk

Classifier output

```

x.nvduo51
x.exteth3
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

== Classifier model (full training set) ==

IB1 instance-based classifier
using 7 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

== Evaluation on test set ==

Time taken to test model on supplied test set: 1.48 seconds

== Summary ==
Correctly Classified Instances      2953          72.7878 %
Incorrectly Classified Instances    1104          27.2122 %
Kappa statistic                      0.3424
Mean absolute error                  0.3421
Root mean squared error              0.4266
Relative absolute error              76.4997 %
Root relative squared error        90.2222 %
Total Number of Instances           4057

== Detailed Accuracy By Class ==

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC   ROC Area  PRC Area  Class
          0.872    0.555    0.755     0.872    0.809     0.353   0.770    0.860     2
          0.445    0.128    0.639     0.445    0.525     0.353   0.770    0.603     1
Weighted Avg.   0.728    0.411    0.716     0.728    0.713     0.353   0.770    0.773

== Confusion Matrix ==

      a      b  <- classified as
2344  344 |  a = 2
      760  609 |  b = 1

```

4. MultilayerPerceptron

a. Training Model: MP - AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options Classifier output

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

20:34:52 - lazy.IBk
20:37:48 - functions.MultilayerPerceptron
20:41:54 - functions.MultilayerPerceptron
20:54:41 - bayes.NaiveBayes
20:57:55 - bayes.NaiveBayes
21:00:48 - bayes.NaiveBayes
21:01:59 - bayes.NaiveBayes
21:03:56 - bayes.NaiveBayes
21:04:32 - bayes.NaiveBayes
21:09:10 - trees.J48
21:11:10 - trees.J48
21:15:06 - trees.J48
21:16:32 - trees.J48
21:19:10 - trees.J48
21:20:28 - trees.J48
21:32:45 - lazy.IBk
21:34:24 - lazy.IBk
21:35:55 - lazy.IBk
21:36:54 - lazy.IBk
21:39:12 - lazy.IBk
21:39:50 - lazy.IBk

Attrib genmirth=5 2.200990222000000
Attrib genhlth=7 4.777126332273161
Attrib x.age80 -8.911467223548932
Attrib x.hcvu651=1 -5.191225390708772
Attrib x.hcvu651=9 3.686132386653871
Attrib x.hcvu651=2 -0.3929049661266807

Class 2
Input
Node 0
Class 1
Input
Node 1

Time taken to build model: 11.12 seconds

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

	Correctly Classified Instances	73.5272 %
Incorrectly Classified Instances	2085	26.4728 %
Kappa statistic	0.3716	
Mean absolute error	0.3338	
Root mean squared error	0.4179	
Relative absolute error	74.6521 %	
Root relative squared error	88.3775 %	
Total Number of Instances	7876	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.735	0.385	0.725	0.735	0.725	0.378	0.784	0.871	2

== Confusion Matrix ==

	a	b	<- classified as
4499	729	1	a = 2
1356	1302	1	b = 1

On Test: MP - AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk
- 21:36:54 - lazy.IBk
- 21:39:12 - lazy.IBk
- 21:39:50 - lazy.IBk
- 21:42:40 - functions.MultilayerPerceptron
- 21:45:22 - functions.MultilayerPerceptron

Classifier output

```
Attrib x.hcvu651=1      3.1912233390700772
Attrib x.hcvu651=9      3.686132386653871
Attrib x.hcvu651=2      -0.3929049661266807

Class 2
Input
Node 0

Class 1
Input
Node 1

Time taken to build model: 11.41 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances      3019          74.4146 %
Incorrectly Classified Instances   1038          25.5854 %
Kappa statistic                   0.3953
Mean absolute error               0.3243
Root mean squared error           0.4141
Relative absolute error           72.5259 %
Root relative squared error      87.5842 %
Total Number of Instances         4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.863    0.489    0.776     0.863    0.817     0.401   0.795    0.881     2
          0.511    0.137    0.655     0.511    0.574     0.401   0.795    0.644     1
Weighted Avg.  0.744    0.370    0.735     0.744    0.735     0.401   0.795    0.801

==== Confusion Matrix ===

      a     b  <- classified as
2319  369 |  a = 2
       669  700 |  b = 1
```

b. Training Model: MP - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options		Classifier output	
<input type="radio"/> Use training set			
<input type="radio"/> Supplied test set	Set...		
<input checked="" type="radio"/> Cross-validation	Folds 10		
<input type="radio"/> Percentage split	% 66		
More options...			
(Nom) havarth3			
Start		Stop	
Result list (right-click for options) <pre> 20:37:48 - functions.MultilayerPerceptron 20:41:54 - functions.MultilayerPerceptron 20:54:41 - bayes.NaiveBayes 20:57:55 - bayes.NaiveBayes 21:00:48 - bayes.NaiveBayes 21:01:59 - bayes.NaiveBayes 21:03:56 - bayes.NaiveBayes 21:04:32 - bayes.NaiveBayes 21:09:10 - trees.J48 21:11:10 - trees.J48 21:15:06 - trees.J48 21:16:32 - trees.J48 21:19:10 - trees.J48 21:20:28 - trees.J48 21:32:45 - lazy.IBk 21:34:24 - lazy.IBk 21:35:55 - lazy.IBk 21:36:54 - lazy.IBk 21:39:12 - lazy.IBk 21:39:50 - lazy.IBk 21:42:40 - functions.MultilayerPerceptron 21:45:22 - functions.MultilayerPerceptron 21:47:55 - functions.MultilayerPerceptron </pre>			

Attrib x.rflnth=2 1.1975107522550723
Attrib x.rfhlth=9 2.327069498085988
Attrib x.phys14d=1 2.1016811339875057
Attrib x.phys14d=2 -0.30250235983567786
Attrib x.phys14d=9 2.1167770871650706
Attrib x.phys14d=3 0.28930452891652716

Class 2
Input
Node 0

Class 1
Input
Node 1

Time taken to build model: 15.21 seconds

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

```

Correctly Classified Instances      5761          73.1463 %
Incorrectly Classified Instances   2115          26.8537 %
Kappa statistic                   0.3504
Mean absolute error               0.3516
Root mean squared error           0.4265
Relative absolute error            78.6286 %
Root relative squared error       90.1995 %
Total Number of Instances         7876

```

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.876	0.552	0.757	0.876	0.812	0.362	0.754	0.832	2	
0.448	0.124	0.648	0.448	0.530	0.362	0.754	0.608	1	
Weighted Avg.	0.731	0.408	0.720	0.731	0.717	0.362	0.754	0.757	

==== Confusion Matrix ====

```

a    b    <- classified as
4570 648 |    a = 2
1467 1191 |    b = 1

```

On Test: MP - AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk
- 21:36:54 - lazy.IBk
- 21:39:12 - lazy.IBk
- 21:39:50 - lazy.IBk
- 21:42:40 - functions.MultilayerPerceptron
- 21:45:22 - functions.MultilayerPerceptron
- 21:47:55 - functions.MultilayerPerceptron

Classifier output

```
Attrib x.phys14d=2      0.302302333903307700
Attrib x.phys14d=9      2.1167770871650706
Attrib x.phys14d=3      0.28938452891652716
Class 2
Input
Node 0
Class 1
Input
Node 1

Time taken to build model: 14.77 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ====

Correctly Classified Instances      2945          72.5906 %
Incorrectly Classified Instances   1112          27.4094 %
Kappa statistic                   0.35
Mean absolute error               0.3519
Root mean squared error           0.426
Relative absolute error           78.6947 %
Root relative squared error      90.0848 %
Total Number of Instances         4057

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

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.852    0.522    0.762     0.852    0.805     0.356   0.754    0.826     2
          0.478    0.148    0.622     0.478    0.540     0.356   0.754    0.615     1
Weighted Avg.  0.726    0.396    0.715     0.726    0.716     0.356   0.754    0.755

==== Confusion Matrix ====

      a     b  <- classified as
2291  397 |     a = 2
715   654 |     b = 1
```

c. Training Model: AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk
- 21:36:54 - lazy.IBk
- 21:39:12 - lazy.IBk
- 21:39:50 - lazy.IBk
- 21:42:40 - functions.MultilayerPerceptron
- 21:45:22 - functions.MultilayerPerceptron
- 21:47:55 - functions.MultilayerPerceptron
- 21:51:37 - functions.MultilayerPerceptron
- 21:52:41 - functions.MultilayerPerceptron

Classifier output

```

Attrib x.age65yr=1 2.12552375075090
Attrib x.age65yr=2 -0.05042600640237855
Attrib x.age65yr=3 1.5192087248022716
Attrib x.rfhlth=1 1.2902263898571449
Attrib x.rfhlth=2 -1.9505775010633017
Attrib x.rfhlth=9 4.3471341187764345

Class 2
Input
Node 0

Class 1
Input
Node 1

Time taken to build model: 13.2 seconds

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

Correctly Classified Instances      5774          73.3113 %
Incorrectly Classified Instances   2102          26.6887 %
Kappa statistic                   0.3353
Mean absolute error               0.3478
Root mean squared error           0.4255
Relative absolute error            77.7752 %
Root relative squared error      89.9788 %
Total Number of Instances         7876

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

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.905   0.603   0.746   0.905   0.818   0.358   0.755   0.825   2
          0.397   0.095   0.679   0.397   0.501   0.358   0.755   0.601   1
Weighted Avg.   0.733   0.432   0.724   0.733   0.711   0.358   0.755   0.750

==== Confusion Matrix ====

     a     b  <- classified as
4720  498 |  a = 2
1604 1054 |  b = 1

```

On Test: MP - AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

- Use training set
- Supplied test set [Set...](#)
- Cross-validation Folds 10
- Percentage split % 66

[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:10 - trees.J48
- 21:15:06 - trees.J48
- 21:16:32 - trees.J48
- 21:19:10 - trees.J48
- 21:20:28 - trees.J48
- 21:32:45 - lazy.IBk
- 21:34:24 - lazy.IBk
- 21:35:55 - lazy.IBk
- 21:36:54 - lazy.IBk
- 21:39:12 - lazy.IBk
- 21:39:50 - lazy.IBk
- 21:42:40 - functions.MultilayerPerceptron
- 21:45:22 - functions.MultilayerPerceptron
- 21:47:55 - functions.MultilayerPerceptron
- 21:51:37 - functions.MultilayerPerceptron
- 21:52:41 - functions.MultilayerPerceptron

Classifier output

```

Attrib x.LININCT=1  1.2502203090371445
Attrib x.rfhlth=2  -1.9505775010633017
Attrib x.rfhlth=9  4.3471341187764345

Class 2
Input
Node 0
Class 1
Input
Node 1

Time taken to build model: 13.35 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.02 seconds

==== Summary ===

Correctly Classified Instances      3003          74.0202 %
Incorrectly Classified Instances   1054          25.9798 %
Kappa statistic                   0.3479
Mean absolute error               0.3339
Root mean squared error           0.4233
Relative absolute error            75.8049 %
Root relative squared error       89.5297 %
Total Number of Instances         4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.917    0.607    0.748     0.917    0.824     0.376   0.753    0.819     2
          0.393    0.083    0.707     0.393    0.505     0.376   0.753    0.612     1
Weighted Avg.  0.740    0.430    0.734     0.740    0.716     0.376   0.753    0.749

==== Confusion Matrix ===

      a     b  <- classified as
2465  223 |  a = 2
 831  538 |  b = 1

```

d. Training Model: MP - AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron

Classifier output

```

Attrib x.hcvu651=1      1.040504055571051
Attrib x.hcvu651=9      5.067121494683076
Attrib x.hcvu651=2      -1.7949127042524404
Attrib x.exteth3=1      1.2032976850593238
Attrib x.exteth3=2      -1.8727009111088817
Attrib x.exteth3=9      2.156372883713926

Class 2
Input
Node 0
Class 1
Input
Node 1

Time taken to build model: 9.7 seconds

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

Correctly Classified Instances      5729          72.74   %
Incorrectly Classified Instances   2147          27.26   %
Kappa statistic                   0.3316
Mean absolute error               0.3431
Root mean squared error           0.4224
Relative absolute error            76.7331 %
Root relative squared error       89.3361 %
Total Number of Instances         7876

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

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.886    0.583    0.749     0.886    0.811     0.347    0.772    0.860     2
          0.417    0.114    0.650     0.417    0.508     0.347    0.772    0.611     1
Weighted Avg.    0.727    0.425    0.715     0.727    0.709     0.347    0.772    0.776

==== Confusion Matrix ====

      a      b  <- classified as
4621  597 |  a = 2
1550 1108 |  b = 1
  
```

On Test: MultilayerPerceptron – AS04 Correlation

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron

Classifier output

```
Attrib x.exteth3=1    1.2032970050595250
Attrib x.exteth3=2    -1.8727009111088817
Attrib x.exteth3=9    2.156372883713926
Class 2
Input
Node 0
Class 1
Input
Node 1

Time taken to build model: 9.43 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.02 seconds

==== Summary ===

Correctly Classified Instances      2991           73.7244 %
Incorrectly Classified Instances   1066           26.2756 %
Kappa statistic                   0.325
Mean absolute error               0.337
Root mean squared error          0.4177
Relative absolute error          75.3608 %
Root relative squared error     88.344 %
Total Number of Instances        4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.936    0.653    0.738     0.936    0.825     0.365    0.783    0.864     2
          0.347    0.064    0.734     0.347    0.471     0.365    0.783    0.641     1
Weighted Avg.  0.737    0.454    0.737     0.737    0.706     0.365    0.783    0.789

==== Confusion Matrix ===

      a     b  <-- classified as
2516  172 |   a = 2
      894  475 |   b = 1
```

5. Logistic

a. Training Model: Logistic – AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose Logistic -R 1.0E-8-M-1-num-decimal-places 4

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

01:32:13 - functions.Logistic

Time taken to build model: 0.35 seconds

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

	5873	74.5683 %
Correctly Classified Instances	5873	74.5683 %
Incorrectly Classified Instances	2003	25.4317 %
Kappa statistic	0.3898	
Mean absolute error	0.3386	
Root mean squared error	0.4129	
Relative absolute error	75.7095 %	
Root relative squared error	87.3123 %	
Total Number of Instances	7876	

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.879	0.515	0.770	0.879	0.821	0.400	0.796	0.878	0.878	2
0.485	0.121	0.670	0.485	0.563	0.400	0.796	0.636	0.636	1
Weighted Avg.	0.746	0.382	0.736	0.746	0.734	0.400	0.796	0.796	

==== Confusion Matrix ====

	a	b	<-- classified as
4585	633		a = 2
1370	1288		b = 1

On Test: Logistic – AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

01:32:13 - functions.Logistic
01:35:08 - functions.Logistic

Classifier output

genhlth=4	0.0547
genhlth=1	2.0154
genhlth=5	0.6024
genhlth=9	1.4322
genhlth=7	1.814
x.age80	0.9514
x.hcvu651=1	0.8472
x.hcvu651=9	1.1637
x.hcvu651=2	1.0856

Time taken to build model: 0.14 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances	3026	74.5871 %
Incorrectly Classified Instances	1031	25.4129 %
Kappa statistic	0.3881	
Mean absolute error	0.3354	
Root mean squared error	0.4098	
Relative absolute error	75.0133 %	
Root relative squared error	86.675 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.882	0.522	0.769	0.882	0.821	0.399	0.801	0.885	2
	0.478	0.118	0.674	0.478	0.560	0.399	0.801	0.659	1
Weighted Avg.	0.746	0.385	0.737	0.746	0.733	0.399	0.801	0.809	

==== Confusion Matrix ===

	a	b	-- classified as
2371	317		a = 2
714	655		b = 1

b. Training Model: Logistic – AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8-M-1-num-decimal-places 4**

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic**

Classifier output

```

gennlth=4      0.0003
genhlth=1     1.8884
genhlth=5     0.9167
genhlth=9     0.8566
genhlth=7     1.6068
x.rfhlth=1    1.142
x.rfhlth=2    0.8708
x.rfhlth=9    1.3176
x.phys14d=1   1.2467
x.phys14d=2   0.9108
x.phys14d=9   0.9348
x.phys14d=3   0.7443

```

Time taken to build model: 0.15 seconds

== Stratified cross-validation ==

== Summary ==

	Correctly Classified Instances	73.2986 %
Incorrectly Classified Instances	2103	26.7014 %
Kappa statistic	0.3364	
Mean absolute error	0.3617	
Root mean squared error	0.4262	
Relative absolute error	80.8901 %	
Root relative squared error	90.1365 %	
Total Number of Instances	7876	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.902	0.600	0.747	0.902	0.817	0.358	0.754	0.834	2
	0.400	0.098	0.676	0.400	0.503	0.358	0.754	0.608	1
Weighted Avg.	0.733	0.430	0.723	0.733	0.711	0.358	0.754	0.758	

== Confusion Matrix ==

```

a      b      <-- classified as
4709  509 |  a = 2
1594 1064 |  b = 1

```

On Test: Logistic – AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8 -M-1 -num-decimal-places 4**

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

Classifier output

gemmirth=9	0.0500
genhlth=7	1.6068
x.rfhlth=1	1.142
x.rfhlth=2	0.8708
x.rfhlth=9	1.3176
x.phys14d=1	1.2467
x.phys14d=2	0.9108
x.phys14d=9	0.9348
x.phys14d=3	0.7443

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic**

Time taken to build model: 0.09 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

==== Summary ===

Correctly Classified Instances	3000	73.9463 %
Incorrectly Classified Instances	1057	26.0537 %
Kappa statistic	0.3457	
Mean absolute error	0.3606	
Root mean squared error	0.4243	
Relative absolute error	80.6304 %	
Root relative squared error	89.7374 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.917	0.609	0.747	0.917	0.823	0.373	0.757	0.833	2
	0.391	0.083	0.706	0.391	0.503	0.373	0.757	0.615	1
Weighted Avg.	0.739	0.432	0.733	0.739	0.715	0.373	0.757	0.760	

==== Confusion Matrix ===

a	b	<-- classified as
2465	223	a = 2
834	535	b = 1

c. Training Model: Logistic – AS03 Gain Ratio

The screenshot shows the Weka interface for training a classifier. The top navigation bar includes Preprocess, Classify (selected), Cluster, Associate, Select attributes, and Visualize. Below the navigation bar, the title "Classifier" is displayed. A "Choose" button is followed by the selected classifier: "Logistic -R 1.0E-8 -M -1 -num-decimal-places 4".

Test options:

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

Classifier output:

attribute	value
airfare=7	1.7774
diffdres=9	0.2996
chccopd1=2	1.4236
chccopd1=1	0.7078
chccopd1=7	0.2559
chccopd1=9	41492914.8473
x.age65yr=1	1.499
x.age65yr=2	0.6427
x.age65yr=3	1.3515
x.rfhlth=1	1.2901
x.rfhlth=2	0.7664
x.rfhlth=9	1.7781

Result list (right-click for options):

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic

Summary Statistics:

```

Time taken to build model: 0.13 seconds

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

Correctly Classified Instances      5818           73.87 %
Incorrectly Classified Instances   2058           26.13 %
Kappa statistic                      0.3492
Mean absolute error                  0.357
Root mean squared error              0.4234
Relative absolute error              79.8327 %
Root relative squared error         89.5376 %
Total Number of Instances            7876

```

Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.909	0.595	0.750	0.909	0.822	0.373	0.755	0.822	0.752	2
0.405	0.091	0.693	0.405	0.511	0.373	0.755	0.613	0.752	1
Weighted Avg.	0.739	0.425	0.731	0.739	0.717	0.373	0.755	0.752	

Confusion Matrix:

	a	b	-- classified as
a	4742	476	a = 2
b	1582	1076	b = 1

On Test: Logistic – AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose Logistic -R 1.0E-8 -M 1 -num-decimal-places 4

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic

Classifier output

chccopar=1	0.7070
chccopd1=7	0.2559
chccopd1=9	41492914.8473
x.age65yr=1	1.499
x.age65yr=2	0.6427
x.age65yr=3	1.3515
x.rfhlth=1	1.2901
x.rfhlth=2	0.7664
x.rfhlth=9	1.7781

Time taken to build model: 0.12 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

==== Summary ===

Correctly Classified Instances	2990	73.6998 %
Incorrectly Classified Instances	1067	26.3002 %
Kappa statistic	0.344	
Mean absolute error	0.3559	
Root mean squared error	0.4225	
Relative absolute error	79.588 %	
Root relative squared error	89.3547 %	
Total Number of Instances	4057	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.909	0.600	0.748	0.909	0.821	0.368	0.761	0.829	2	
0.400	0.091	0.691	0.400	0.506	0.368	0.761	0.614	1	
Weighted Avg.	0.737	0.429	0.729	0.737	0.715	0.368	0.761	0.757	

==== Confusion Matrix ===

	a	b	-- classified as
2443	245		a = 2
822	547		b = 1

d. Training Model: Logistic – AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Logistic -R 1.0E-8-M -1-num-decimal-places 4**

Test options

- Use training set
- Supplied test set Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) havarth3

Start **Stop**

Classifier output

```

diffwalk=7 0.5052
diffwalk=9 2.3899
x.age80 0.955
x.rfhlth=1 1.3293
x.rfhlth=2 0.7413
x.rfhlth=9 2.1789
x.hcvu65l=1 0.8437
x.hcvu65l=9 1.1675
x.hcvu65l=2 1.0898
x.exteth3=1 1.1222
x.exteth3=2 0.8916
x.exteth3=9 0.9891

```

Time taken to build model: 0.07 seconds

==== Stratified cross-validation ===

==== Summary ===

	Correctly Classified Instances	73.5653 %
Incorrectly Classified Instances	2082	26.4347 %
Kappa statistic	0.3565	
Mean absolute error	0.3467	
Root mean squared error	0.4175	
Relative absolute error	77.5203 %	
Root relative squared error	88.2892 %	
Total Number of Instances	7876	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.885	0.558	0.757	0.885	0.816	0.371	0.784	0.869	2
	0.442	0.115	0.662	0.442	0.530	0.371	0.784	0.626	1
Weighted Avg.	0.736	0.408	0.725	0.736	0.720	0.371	0.784	0.787	

==== Confusion Matrix ===

```

a   b   <- classified as
4618 600 |   a = 2
1482 1176 |   b = 1

```

On Test: Logistic – AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options		Classifier output	
<input type="radio"/> Use training set		x.rfmln=1	1.3293
<input checked="" type="radio"/> Supplied test set	Set...	x.rfhlh=2	0.7413
<input type="radio"/> Cross-validation	Folds 10	x.rfhlh=9	2.1789
<input type="radio"/> Percentage split	% 66	x.hcvu651=1	0.8437
More options...		x.hcvu651=9	1.1675
		x.hcvu651=2	1.0898
		x.exteth3=1	1.1222
		x.exteth3=2	0.8916
		x.exteth3=9	0.9891

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic**

Time taken to build model: 0.07 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

==== Summary ===

	Correctly Classified Instances	2963	73.0343 %
Incorrectly Classified Instances	1094	26.9657 %	
Kappa statistic	0.3438		
Mean absolute error	0.3436		
Root mean squared error	0.4154		
Relative absolute error	76.8436 %		
Root relative squared error	87.8533 %		
Total Number of Instances	4057		

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.881	0.565	0.754	0.881	0.812	0.357	0.788	0.874	2	
0.435	0.119	0.650	0.435	0.521	0.357	0.788	0.641	1	
Weighted Avg.	0.730	0.414	0.719	0.730	0.714	0.357	0.788	0.795	

==== Confusion Matrix ===

a	b	<-- classified as
2367	321	a = 2
773	596	b = 1

6. Random Forest

a. Training Model: RandomForest – AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest**

Classifier output

```
gennrich
x.age80
x.hcvu651
havarth3
Test mode: 10-fold cross-validation
==== Classifier model (full training set) ====
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 1.32 seconds
==== Stratified cross-validation ====
==== Summary ===
Correctly Classified Instances      5586          70.9243 %
Incorrectly Classified Instances   2290          29.0757 %
Kappa statistic                   0.3156
Mean absolute error               0.3366
Root mean squared error           0.4448
Relative absolute error            75.2743 %
Root relative squared error       94.7426 %
Total Number of Instances         7876
==== Detailed Accuracy By Class ====

TP Rate   FP Rate   Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0.833    0.533    0.754     0.833    0.791     0.320   0.741    0.843    2
0.467    0.167    0.587     0.467    0.520     0.320   0.742    0.574    1
Weighted Avg. 0.709    0.409    0.698     0.709    0.700     0.320   0.741    0.753
==== Confusion Matrix ====

a     b  <-- classified as
4344  874 |   a = 2
1416 1242 |   b = 1
```

On Test: RandomForest – AS01 Info Gain

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest-P 100-l 100-num-slots 1-K 0-M 1.0-V 0.001-S 1**

Test options

Use training set
 Supplied test set [Set...](#)
 Cross-validation Folds 10
 Percentage split % 66
[More options...](#)

(Nom) havarth3

[Start](#) [Stop](#)

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest
- 01:49:41 - trees.RandomForest**

Classifier output

```
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 1.04 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.22 seconds

==== Summary ===

Correctly Classified Instances 2903 71.5553 %
Incorrectly Classified Instances 1154 28.4447 %
Kappa statistic 0.3174
Mean absolute error 0.3299
Root mean squared error 0.4428
Relative absolute error 73.7741 %
Root relative squared error 93.6418 %
Total Number of Instances 4057

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

      TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
      0.856 0.560 0.750 0.856 0.800 0.326 0.748 0.841 2
      0.440 0.144 0.609 0.440 0.511 0.326 0.748 0.592 1
Weighted Avg. 0.716 0.420 0.702 0.716 0.702 0.326 0.748 0.757

==== Confusion Matrix ====

      a     b  <- classified as
2301 387 |  a = 2
    767 602 |  b = 1
```

b. Training model: RandomForest – AS02 OneR

The screenshot shows the Weka interface with the following configuration and output:

Classifier Selection: Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options:

- Use training set
- Supplied test set [Set...](#)
- Cross-validation Folds 10
- Percentage split % 66 [More options...](#)

Classifier output:

```

gennirth
x.rfhlth
x.phys14d
havarth3
Test mode: 10-fold cross-validation

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.6 seconds

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

Correctly Classified Instances      5708      72.4733 %
Incorrectly Classified Instances   2168      27.5267 %
Kappa statistic                   0.3332
Mean absolute error               0.3533
Root mean squared error           0.4359
Relative absolute error            79.0065 %
Root relative squared error       92.1874 %
Total Number of Instances         7876

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.872    0.564    0.752     0.872    0.808    0.345    0.739    0.815     2
          0.436    0.128    0.634     0.436    0.517    0.345    0.739    0.585     1
Weighted Avg.  0.725    0.417    0.712     0.725    0.709    0.345    0.739    0.737

==== Confusion Matrix ===

      a      b  <- classified as
4550  668 |  a = 2
1500 1158 |  b = 1
  
```

Result list (right-click for options):

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest
- 01:49:41 - trees.RandomForest**
- 01:51:10 - trees.RandomForest

On Test: RandomForest – AS02 OneR

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options Classifier output

Use training set
 Supplied test set Set...
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest
- 01:49:41 - trees.RandomForest
- 01:51:10 - trees.RandomForest
- 01:52:49 - trees.RandomForest

havarth3

Test mode: user supplied test set: size unknown (reading incrementally)

==== Classifier model (full training set) ====
RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.56 seconds

==== Evaluation on test set ====
Time taken to test model on supplied test set: 0.14 seconds

==== Summary ====
Correctly Classified Instances 2987 73.6250 %
Incorrectly Classified Instances 1070 26.3742 %
Kappa statistic 0.3501
Mean absolute error 0.3521
Root mean squared error 0.4308
Relative absolute error 78.748 %
Root relative squared error 91.1136 %
Total Number of Instances 4057

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
a	0.897	0.579	0.752	0.897	0.818	0.369	0.744	0.815	2
b	0.421	0.103	0.675	0.421	0.518	0.369	0.744	0.600	1
Weighted Avg.	0.736	0.419	0.726	0.736	0.717	0.369	0.744	0.743	

==== Confusion Matrix ====
a b <- classified as
2411 277 | a = 2
793 576 | b = 1

c. On Training Model: RandomForest – AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options	Classifier output																																																																																																																																												
<input type="radio"/> Use training set <input type="radio"/> Supplied test set Set... <input checked="" type="radio"/> Cross-validation Folds 10 <input type="radio"/> Percentage split % 66 More options...	chccopur x.age65yr x.rfhlth havarth3 Test mode: 10-fold cross-validation === Classifier model (full training set) === RandomForest Bagging with 100 iterations and base learner weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities Time taken to build model: 0.44 seconds === Stratified cross-validation === === Summary === Correctly Classified Instances 5772 73.2859 % Incorrectly Classified Instances 2104 26.7141 % Kappa statistic 0.3425 Mean absolute error 0.3521 Root mean squared error 0.4258 Relative absolute error 78.7322 % Root relative squared error 90.052 % Total Number of Instances 7876 === Detailed Accuracy By Class === <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>MCC</th> <th>ROC Area</th> <th>PRC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td>01:32:13 - functions.Logistic</td> <td>0.893</td> <td>0.582</td> <td>0.751</td> <td>0.893</td> <td>0.816</td> <td>0.360</td> <td>0.749</td> <td>0.821</td> <td>2</td> </tr> <tr> <td>01:35:08 - functions.Logistic</td> <td>0.418</td> <td>0.107</td> <td>0.666</td> <td>0.418</td> <td>0.514</td> <td>0.360</td> <td>0.749</td> <td>0.597</td> <td>1</td> </tr> <tr> <td>01:36:46 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:38:33 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:40:18 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:41:36 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:43:20 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:44:22 - functions.Logistic</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:46:54 - trees.RandomForest</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:49:41 - trees.RandomForest</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:51:10 - trees.RandomForest</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:52:49 - trees.RandomForest</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>01:54:28 - trees.RandomForest</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> === Confusion Matrix === a b <- classified as 4661 557 a = 2 1547 1111 b = 1		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	01:32:13 - functions.Logistic	0.893	0.582	0.751	0.893	0.816	0.360	0.749	0.821	2	01:35:08 - functions.Logistic	0.418	0.107	0.666	0.418	0.514	0.360	0.749	0.597	1	01:36:46 - functions.Logistic										01:38:33 - functions.Logistic										01:40:18 - functions.Logistic										01:41:36 - functions.Logistic										01:43:20 - functions.Logistic										01:44:22 - functions.Logistic										01:46:54 - trees.RandomForest										01:49:41 - trees.RandomForest										01:51:10 - trees.RandomForest										01:52:49 - trees.RandomForest										01:54:28 - trees.RandomForest									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class																																																																																																																																				
01:32:13 - functions.Logistic	0.893	0.582	0.751	0.893	0.816	0.360	0.749	0.821	2																																																																																																																																				
01:35:08 - functions.Logistic	0.418	0.107	0.666	0.418	0.514	0.360	0.749	0.597	1																																																																																																																																				
01:36:46 - functions.Logistic																																																																																																																																													
01:38:33 - functions.Logistic																																																																																																																																													
01:40:18 - functions.Logistic																																																																																																																																													
01:41:36 - functions.Logistic																																																																																																																																													
01:43:20 - functions.Logistic																																																																																																																																													
01:44:22 - functions.Logistic																																																																																																																																													
01:46:54 - trees.RandomForest																																																																																																																																													
01:49:41 - trees.RandomForest																																																																																																																																													
01:51:10 - trees.RandomForest																																																																																																																																													
01:52:49 - trees.RandomForest																																																																																																																																													
01:54:28 - trees.RandomForest																																																																																																																																													

On Test: RandomForest – AS03 Gain Ratio

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options

Use training set
 Supplied test set **Set...**
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest
- 01:49:41 - trees.RandomForest
- 01:51:10 - trees.RandomForest
- 01:52:49 - trees.RandomForest
- 01:54:28 - trees.RandomForest
- 01:55:38 - trees.RandomForest**

Classifier output

```
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.42 seconds

==== Evaluation on test set ====

Time taken to test model on supplied test set: 0.1 seconds

==== Summary ===

Correctly Classified Instances      2986          73.6012 %
Incorrectly Classified Instances   1071          26.3988 %
Kappa statistic                   0.3498
Mean absolute error               0.3483
Root mean squared error          0.4204
Relative absolute error           77.8903 %
Root relative squared error     88.9109 %
Total Number of Instances        4057

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

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.896    0.579    0.753     0.896    0.818     0.368   0.767    0.841     2
          0.421    0.104    0.674     0.421    0.519     0.368   0.767    0.607     1
Weighted Avg.    0.736    0.418    0.726     0.736    0.717     0.368   0.767    0.762

==== Confusion Matrix ====

      a     b  <- classified as
2409  279 |  a = 2
  792  577 |  b = 1
```

d. Training Model: RandomForest – AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest-P 100-l 100-num-slots 1-K 0-M 1.0-V 0.001-S 1**

Test options		Classifier output																																																																																																					
<input type="radio"/> Use training set		x.tinith x.hcvu651 x.exteth3 havarth3																																																																																																					
<input type="radio"/> Supplied test set	Set...	Test mode: 10-fold cross-validation																																																																																																					
<input checked="" type="radio"/> Cross-validation	Folds 10																																																																																																						
<input type="radio"/> Percentage split	% 66	==== Classifier model (full training set) ==== RandomForest Bagging with 100 iterations and base learner weka.classifiers.trees.RandomTree -K 0 -M 1.0 -v 0.001 -S 1 -do-not-check-capabilities																																																																																																					
More options...		Time taken to build model: 0.84 seconds																																																																																																					
<input type="button" value="Start"/> <input type="button" value="Stop"/>		=== Stratified cross-validation === === Summary === <table border="1"> <thead> <tr> <th></th> <th>Correctly Classified Instances</th> <th>69.9594 %</th> </tr> </thead> <tbody> <tr> <td>01:32:13 - functions.Logistic</td> <td>5510</td> <td></td> </tr> <tr> <td>01:35:08 - functions.Logistic</td> <td></td> <td>30.0406 %</td> </tr> <tr> <td>01:36:46 - functions.Logistic</td> <td>2366</td> <td></td> </tr> <tr> <td>01:38:33 - functions.Logistic</td> <td></td> <td></td> </tr> <tr> <td>01:40:18 - functions.Logistic</td> <td>0.2812</td> <td></td> </tr> <tr> <td>01:41:36 - functions.Logistic</td> <td>0.3473</td> <td></td> </tr> <tr> <td>01:43:20 - functions.Logistic</td> <td>0.4456</td> <td></td> </tr> <tr> <td>01:44:22 - functions.Logistic</td> <td>77.6713 %</td> <td></td> </tr> <tr> <td>01:46:54 - trees.RandomForest</td> <td>94.2303 %</td> <td></td> </tr> <tr> <td>01:49:41 - trees.RandomForest</td> <td>7876</td> <td></td> </tr> <tr> <td>01:51:10 - trees.RandomForest</td> <td></td> <td></td> </tr> <tr> <td>01:52:49 - trees.RandomForest</td> <td></td> <td></td> </tr> <tr> <td>01:54:28 - trees.RandomForest</td> <td></td> <td></td> </tr> <tr> <td>01:55:38 - trees.RandomForest</td> <td></td> <td></td> </tr> <tr> <td>01:57:05 - trees.RandomForest</td> <td></td> <td></td> </tr> </tbody> </table> === Detailed Accuracy By Class === <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>MCC</th> <th>ROC Area</th> <th>PRC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td>0.841</td> <td>0.578</td> <td>0.741</td> <td>0.841</td> <td>0.788</td> <td>0.288</td> <td>0.732</td> <td>0.832</td> <td>0.2</td> <td></td> </tr> <tr> <td>0.422</td> <td>0.159</td> <td>0.575</td> <td>0.422</td> <td>0.487</td> <td>0.288</td> <td>0.732</td> <td>0.570</td> <td>1</td> <td></td> </tr> <tr> <td>Weighted Avg.</td> <td>0.700</td> <td>0.437</td> <td>0.685</td> <td>0.700</td> <td>0.686</td> <td>0.288</td> <td>0.732</td> <td>0.744</td> <td></td> </tr> </tbody> </table> === Confusion Matrix === <table border="1"> <thead> <tr> <th></th> <th>a</th> <th>b</th> <th><-- classified as</th> </tr> </thead> <tbody> <tr> <td>a</td> <td>4389</td> <td>829</td> <td> a = 2</td> </tr> <tr> <td>b</td> <td>1537</td> <td>1121</td> <td> b = 1</td> </tr> </tbody> </table>			Correctly Classified Instances	69.9594 %	01:32:13 - functions.Logistic	5510		01:35:08 - functions.Logistic		30.0406 %	01:36:46 - functions.Logistic	2366		01:38:33 - functions.Logistic			01:40:18 - functions.Logistic	0.2812		01:41:36 - functions.Logistic	0.3473		01:43:20 - functions.Logistic	0.4456		01:44:22 - functions.Logistic	77.6713 %		01:46:54 - trees.RandomForest	94.2303 %		01:49:41 - trees.RandomForest	7876		01:51:10 - trees.RandomForest			01:52:49 - trees.RandomForest			01:54:28 - trees.RandomForest			01:55:38 - trees.RandomForest			01:57:05 - trees.RandomForest				TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	0.841	0.578	0.741	0.841	0.788	0.288	0.732	0.832	0.2		0.422	0.159	0.575	0.422	0.487	0.288	0.732	0.570	1		Weighted Avg.	0.700	0.437	0.685	0.700	0.686	0.288	0.732	0.744			a	b	<-- classified as	a	4389	829	a = 2	b	1537	1121	b = 1
	Correctly Classified Instances	69.9594 %																																																																																																					
01:32:13 - functions.Logistic	5510																																																																																																						
01:35:08 - functions.Logistic		30.0406 %																																																																																																					
01:36:46 - functions.Logistic	2366																																																																																																						
01:38:33 - functions.Logistic																																																																																																							
01:40:18 - functions.Logistic	0.2812																																																																																																						
01:41:36 - functions.Logistic	0.3473																																																																																																						
01:43:20 - functions.Logistic	0.4456																																																																																																						
01:44:22 - functions.Logistic	77.6713 %																																																																																																						
01:46:54 - trees.RandomForest	94.2303 %																																																																																																						
01:49:41 - trees.RandomForest	7876																																																																																																						
01:51:10 - trees.RandomForest																																																																																																							
01:52:49 - trees.RandomForest																																																																																																							
01:54:28 - trees.RandomForest																																																																																																							
01:55:38 - trees.RandomForest																																																																																																							
01:57:05 - trees.RandomForest																																																																																																							
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class																																																																																														
0.841	0.578	0.741	0.841	0.788	0.288	0.732	0.832	0.2																																																																																															
0.422	0.159	0.575	0.422	0.487	0.288	0.732	0.570	1																																																																																															
Weighted Avg.	0.700	0.437	0.685	0.700	0.686	0.288	0.732	0.744																																																																																															
	a	b	<-- classified as																																																																																																				
a	4389	829	a = 2																																																																																																				
b	1537	1121	b = 1																																																																																																				

On Test: RandomForest – AS04 Correlation

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **RandomForest-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options

Use training set
 Supplied test set **Set...**
 Cross-validation Folds 10
 Percentage split % 66
More options...

(Nom) havarth3

Start Stop

Result list (right-click for options)

- 01:32:13 - functions.Logistic
- 01:35:08 - functions.Logistic
- 01:36:46 - functions.Logistic
- 01:38:33 - functions.Logistic
- 01:40:18 - functions.Logistic
- 01:41:36 - functions.Logistic
- 01:43:20 - functions.Logistic
- 01:44:22 - functions.Logistic
- 01:46:54 - trees.RandomForest
- 01:49:41 - trees.RandomForest
- 01:51:10 - trees.RandomForest
- 01:52:49 - trees.RandomForest
- 01:54:28 - trees.RandomForest
- 01:55:38 - trees.RandomForest
- 01:57:05 - trees.RandomForest
- 01:58:33 - trees.RandomForest**

Classifier output

```
havarth3
Test mode: user supplied test set: size unknown (reading incrementally)

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.86 seconds

==== Evaluation on test set ===

Time taken to test model on supplied test set: 0.21 seconds

==== Summary ===

Correctly Classified Instances      2906          71.6293 %
Incorrectly Classified Instances   1151          28.3707 %
Kappa statistic                   0.3219
Mean absolute error               0.3399
Root mean squared error           0.4372
Relative absolute error            76.0158 %
Root relative squared error       92.4606 %
Total Number of Instances         4057

==== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
          0.853    0.551    0.752     0.853    0.799     0.329    0.748    0.839     2
          0.449    0.147    0.608     0.449    0.516     0.329    0.748    0.589     1
Weighted Avg.    0.716    0.415    0.704     0.716    0.704     0.329    0.748    0.755

==== Confusion Matrix ===

      a     b  <- classified as
2292  396 |  a = 2
  755  614 |  b = 1
```

Best Performance:

List of Classifiers and Attribute Selection Methods based on **Overall** Accuracy:

Overall Accuracy	InfoGain	OneR	GainRatio	Correlation
Naïve Bayes	0.727 (72.7%)	0.732 (73.2%)	0.740 (74%)	0.711 (71.1%)
J48	0.753 (75.3%)	0.743 (74.3%)	0.736 (73.6%)	0.740 (74%)
KNN (K = 7)	0.733 (73.3%)	0.736 (73.6%)	0.736 (73.6%)	0.728 (72.8%)
Multilayer Perceptron	0.744 (74.4%)	0.726 (72.6%)	0.740 (74%)	0.737 (73.7%)
Logistic	0.746 (74.6%)	0.739 (73.9%)	0.737 (73.7%)	0.730 (73%)
Random Forest	0.716 (71.6%)	0.736 (73.6%)	0.736 (73.6%)	0.716 (71.6%)

List of Classifiers and Attribute Selection Methods based on **Class 1** Accuracy: (People were ever told to have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia)

Class 1 Accuracy	InfoGain	OneR	GainRatio	Correlation
Naïve Bayes	0.634 (63.4%)	0.490 (49%)	0.438 (43.8%)	0.636 (63.6%)
J48	0.494 (49.4%)	0.435 (43.5%)	0.404 (40.4%)	0.407 (40.7%)
KNN (K = 7)	0.471 (47.1%)	0.413 (41.3%)	0.419 (41.9%)	0.445 (44.5%)
Multilayer Perceptron	0.511 (51.1%)	0.478 (47.8%)	0.393 (39.3%)	0.347 (34.7%)
Logistic	0.478 (47.8%)	0.391 (39.1%)	0.400 (40%)	0.435 (43.5%)
Random Forest	0.440 (44%)	0.421 (42.1%)	0.421 (42.1%)	0.449 (44.9%)

List of Classifiers and Attribute Selection Methods based on **Class 2** Accuracy: (People **never** told to have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia)

Class 2 Accuracy	InfoGain	OneR	GainRatio	Correlation
Naïve Bayes	0.744 (74.4%)	0.855 (85.5%)	0.894 (89.4%)	0.749 (74.9%)
J48	0.855 (85.5%)	0.900 (90%)	0.905 (90.5%)	0.910 (91%)
KNN (K = 7)	0.866 (86.6%)	0.901 (90.1%)	0.898 (89.8%)	0.872 (87.2%)
Multilayer Perceptron	0.863 (86.3%)	0.852 (85.2%)	0.917 (91.7%)	0.936 (93.6%)
Logistic	0.882 (88.2%)	0.917 (91.7%)	0.909 (90.9%)	0.881 (88.1%)
Random Forest	0.856 (85.6%)	0.897 (89.7%)	0.896 (89.6%)	0.853 (85.3%)

Summary:

Data Mining Goal	Best Model
A model that has the highest overall accuracy	InfoGain attribute selection method with J48 classification algorithm has the highest overall accuracy of 0.753 (75.3 %)
A model that predicts people were ever told to have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia (class = 1) with the highest accuracy	Correlation attribute selection method with Naïve Bayes classification algorithm (class = 1) has highest accuracy of 0.636 (63.6 %)
A model that predicts people never told to have any form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia (class = 2) with the highest accuracy	Correlation attribute selection method with Multilayer Perceptron classification algorithm (class = 2) has highest accuracy of 0.936 (93.6 %)

Overall, **InfoGain** attribute selection method with **J48** classification algorithm has the highest overall accuracy of **0.753 (75.3 %)**.

List of attributes that are in the best training or best test dataset):

InfoGain attribute selection method with **J48** classification algorithm was my best test with highest overall accuracy of **0.753 (75.3 %)**.

The attributes in the reduced test dataset were:

- employ1 - Employment Status
- diffwalk - Difficulty Walking or Climbing Stairs
- physhlth - Number of Days Physical Health Not Good In last 30 days
- genhlth - General Health
- x.age80 - Imputed Age value collapsed above 80
- x.hcvu651 - Respondents aged 18-64 with health care coverage

Class attribute: havarth3 - (Ever told) you have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia? (Arthritis diagnoses include: rheumatism, polymyalgia rheumatica; osteoarthritis (not osteoporosis); tendonitis, bursitis, bunion, tennis elbow; carpal tunnel syndrome, tarsal tunnel syndrome; joint infection, etc.)

Best Test result: (Overall Accuracy) - InfoGain attribute selection method with J48 classification algorithm

The screenshot shows the Weka interface with the following configuration and output:

Test options:

- Supplied test set (selected)
- Cross-validation Folds 10
- Percentage split % 66

Classifier output:

```

|   |   physntr > 11
|   |   |   x.age80 <= 51: 2 (20.39/4.63)
|   |   |   x.age80 > 51: 1 (328.81/108.73)
|   |   diffwalk = 7: 1 (19.49/7.32)
|   |   diffwalk = 9: 2 (6.16/2.05)

Number of Leaves :      76

Size of the tree :      124

Time taken to build model: 0.07 seconds

```

Result list (right-click for options):

- 20:14:34 - bayes.NaiveBayes
- 20:16:27 - bayes.NaiveBayes
- 20:22:35 - trees.J48
- 20:27:39 - trees.J48
- 20:30:49 - lazy.IBk
- 20:34:52 - lazy.IBk
- 20:37:48 - functions.MultilayerPerceptron
- 20:41:54 - functions.MultilayerPerceptron
- 20:54:41 - bayes.NaiveBayes
- 20:57:55 - bayes.NaiveBayes
- 21:00:48 - bayes.NaiveBayes
- 21:01:59 - bayes.NaiveBayes
- 21:03:56 - bayes.NaiveBayes
- 21:04:32 - bayes.NaiveBayes
- 21:09:10 - trees.J48
- 21:11:08 - trees.J48

Summary Statistics:

	Correctly Classified Instances	75.2773 %
Incorrectly Classified Instances	1003	24.7227 %
Kappa statistic	0.4064	
Mean absolute error	0.3395	
Root mean squared error	0.4175	
Relative absolute error	75.9264 %	
Root relative squared error	88.3007 %	
Total Number of Instances	4057	

Detailed Accuracy By Class:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.885	0.506	0.774	0.885	0.826	0.417	0.777	0.841	2	
0.494	0.115	0.686	0.494	0.574	0.417	0.777	0.611	1	
Weighted Avg.	0.753	0.374	0.744	0.753	0.741	0.417	0.777	0.764	

Confusion Matrix:

a	b	-- classified as
2378	310	a = 2
693	676	b = 1

Discussion:

(1). Describe what criteria you used when you were choosing your *best model*. In other words, you need to justify why you chose that particular model as the best model. If your justification is not based on sound technical criteria, you will lose points.

Solution 1:

I ran the 6 classifier models on the 4 attribute selection methods and selected 3 models for inspection.

1. Best Overall Accuracy (75.3%)
2. Best Class 1 Accuracy (63.6%)
3. Best Class 2 Accuracy (93.6%)

I selected the model that classifies based on Overall Accuracy as my Best Model.

The model that classifies Class 2 accurately (**Correlation** attribute selection method with **Multilayer Perceptron** classification algorithm) with **93.6%** has very high accuracy but the issue is that it can classify people correctly based on attributes that they DO NOT have or were never told that they have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia. It is always good to know that a model can classify and tell that someone does not have a disease. However, it is more important to classify whether a person has some form of arthritis (Class = 1). Hence, we did not choose the model which classifies Class 2 Accuracy (93.6%) as our best model.

I would have been very happy if we could have got a model that classifies based on the attributes selected to classify with higher accuracy that someone was told or has some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia (Class = 1). Our model that has the highest accuracy to classify for Class = 1 is **63.6%** (**Correlation** attribute selection method with **Naïve Bayes** classification algorithm). The accuracy percentage is very low, and it is closer to 50%. Hence it is not a model that is highly accurate in classifying someone with Class = 1. Hence, we did not choose the model which classifies Class 1 Accuracy (63.6%) as our best model.

Now we are having a model that classifies people whether they were ever told or have some form of Arthritis (Class = 1) OR do not have or were never told that they have some form of Arthritis (Class = 2) with Overall Accuracy of **75.3%**. This is not a very high accuracy, but it is still a better model.

Hence, we selected **InfoGain** attribute selection method with **J48** classification algorithm has the highest overall accuracy of **0.753 (75.3 %)** as our **Best Model**.

(2). List five attributes that you think are most relevant to the class attribute. You need to justify why you selected those five attributes.

Solution 2:

We ran 4 attribute selection methods on our Training dataset to find the best ranked attributes based on different criteria of the methods.

We see that the below 5 attributes are highly ranked on all or most of these attribute selection methods.

- employ1 - Employment Status of the person ()
- diffwalk - Difficulty Walking or Climbing Stairs
- physlth - Number of Days Physical Health Not Good In last 30 days
- genhlth - General Health
- x.age80 - Imputed Age value collapsed above 80

These attributes seem to be most relevant to the class attribute **havarth3**.

- **employ1** – Indicates the employment (self – employed, student, homemaker, out of work etc.)
- **diffwalk** – Indicates whether the person asked this question has serious difficulty walking or climbing stairs. (Yes, No, don't know, refused to answer etc.)
- **physlth** – Includes physical illness and injury, for how many days during the past 30 days was the person's physical health not good (1-30 number of days, None, refused to answer etc.)
- **genhlth** – Indicates the general overall health of a person (Excellent, good, fair, poor etc.)
- **x.age80** – Categorizes people into different age categories. (18-24, 25-29, 30 -34,...80 or older)

All these attributes do indicate a good measure of a person's health and specifically related to our class attribute **havarth3**, which is an indicator of whether someone was ever told if they had some form of Arthritis or not.

Employment Status of a person tells us how active the person is and may be also indicates the age of the person if he/she is a student, or the person stays at home most of the times or he/she is unable to work due to some reason.

Serious **Difficulty in walking up and down the stairs** is a very good indicator of some form of Arthritis as Arthritis is basically a disease that causes the swelling and tenderness of one or more of your joints.

The **Physical illness** or injury in the past 30 days also indicate whether a person could have had the injury related to some sort of pre-existing disease OR may be just an accident that could cause a disease like Arthritis.

General health of a person is a good indicator of how the person is feeling overall and tells us about the overall health condition of a person.

The **Age categorization** of people into different range can tell a lot about the sample population and would be a good indicator based on other attributes whether the person has chances of having Arthritis.