

Task 2

SUMMARISE

Section 1. Introduction

In this paper secondary student dataset evaluated in which the researcher wants to tell how the grades of the students get varied due to the other factors like Mjob, Pjob, Absence, internet, romance, finance, free time etc. In this the highest accuracy for models changed as the different situations were created so basically comparison of different model results mathematics and the Portuguese score. The factors on which different results were there are A- the final grade(G3) of all the student s were removed B – The second period(G2) score are also removed C – the first grade(G1) score is also removed so the evaluation is only done on the basis of other factors .Further more data was seen in 3 types Discrete data, 5 level data and Linearly.

- i) binary classification (pass/fail);
- ii) classification with five levels (from I very good or excellent to V - insufficient); and
- iii) regression, with a numeric output that ranges between zero (0%) and twenty (100%).

In the preprocessing the data was normalized and then standardized so the mean and the SD value will become 1 and 0 the models tested were Naïve – Bayes, Neural Network, SVM, Decision Tree and Random Forest for Binary and 5 level classification PCC was seen and for linear regression RMSE was seen. The changes in the model result are on the basis of other factors involved and the background and interests.

for the binary classification

Section 2: Naïve Bayesian Classifier

Following all the Factors as A, B & C the NVB showed the best result for Mathematics. For Portuguese data the model could not show the highest accuracy as it had the least accuracy.

Section 3: Neural Network

After the neurons also this model was not a great model for maths data as it obtained an accuracy of only 88.3. In the Portuguese results for G2 & G1 the model had accuracy of 87 % approx

Section 4: Support Vector Machine

As when different factors were involved the value of SVM also decreased as it was not coming as the best model for the dataset. For Portuguese data the SVM was not considered the best model

Section 5: Decision Tree

Decision Tree is considered as a better model than SVM but then also its value was varying around 90.1(not the best model). It has the highest accuracy of 93.3 % for Portuguese dataset

Section 6: Random Forest

In this model randomly data is picked and the evaluation is done on that to remove the biasness as this was not considered as the best as with the accuracy of 91.2 so there was no such biasness in the data. In the Portuguese data it could not show any Difference from maths data.

for the 5 – level classification

Naïve – Bayes shows the highest result with 78.5% for mathematics data and Decision tree with 76.1% as the highest where the data was divided into 5 parts.

for the linear regression While comparing all model and all the factors the best results were obtained for **Random Forest** and Root mean square error was seen for regression with the accuracy of around 1.75-1.32 value. As other were showing higher values.

Section 7: Experimental Evaluation

The accuracy of the model was decreasing as other factors were added as B and C. The best results can be seen when G3 was only removed and as we can say when there were no grades in the C Factor the result were not good at all because no grade were included in that. For the binary classification the data was sectioned into two parts as >10 pass otherwise fail so it showed different results and for the 5 level the data was sampled into parts from which NVB and DT were showing a good result and at last for the regression the data we could see in the error was showing the best result with Random Forest.

Sections 8 & 9: Related Work and Conclusion

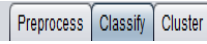
Related work applied different types of model which also have different results for different inputs which can conclude that for this dataset there can't be any particular model which we can consider as the best fit but we can say that there are so many other factors that can be considered as on the basis of factors which may lead to changes in the student results. We can also consider other subjects and the portioning we are doing can be considered with more variables and more diversity

EVALUATE

Experiment Plan

To replicate the experiments 2 datasets from the UCI repository were used, one with Mathematics data and grades and the other with Portuguese data and grade. While the paper also tested the algorithm on a large synthetic dataset, the focus in this experiment is on the results from the publicly available datasets. As the three types are not seen the experiment only binary classification and the linear regression is done.

Experiments

Following the methodology as delineated in the paper, the experiments were conducted in two phases: Binary Classification and Linear regression in which it was sectioned into 3 parts A, B & C as G3 removed, G2 also removed and finally G1 was also removed. As the experiment was conducted in Weka the preprocessing was done in was as the Mjob was classified further into its instances then it was filtered with normalization and then Standardization was done to make both the dataset of mat and portu skewed as the mean and the SD was changed to 0 and 1 there was no class for this processes. G3 was made as the Class and then data was filtered in unsupervised filtering from numerical to nominal then supervised discrete filter was applied which made the data into parts to verify it classifier was added as Logistic Regression which partitioned it into the best way. Now finally G3 is removed and then we go to classify section  where the models were applied one by one and there result was calculated .then again the process was repeated when G2 was removed then same steps for G1.

After this we undo all the steps and normalize and Standardize again and the in the Filter Linear Regression was added so the data was now classified as Linear regression and then in the classify section models were applied and evaluated to ROC value

Similar steps were done for the Portuguese data and then they are compared.

Analysis

Comparing the results of the experiments in WEKA with those of the original research paper, the following items are of note:

- In weka there is no coding and it is a tool which applies the model and the process which we tell it to do.
- In the sampling process it was not letting us give the sample it was taking a random sample which gave us a different result from the paper which we are replicating.
- When numeric or nominal data was involved some models we can't apply as we can see in the linear regression we can see NVB data and similarly in the binary Regression when no grades were involved.

- It gave very different results from the R-minor As the research paper has done the analysis in R, because of sampling and because of the preprocessing as I added some steps and excluded some.
- As seen in the table below are the results of the experiment conducted which are little bit of similar to those in the research paper but not exact.

Input Setup	Mathematics					Portuguese				
	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
A	75.18	74.68	<u>78.73</u>	74.17	76.15	83.3	75.5	<u>83.4</u>	81.0	81.1
B	74.17	76.96	<u>77.72</u>	<u>78.22</u>	76.70	<u>61.5</u>	56.9	<u>61.0</u>	58.1	58.8
C	na	<u>76.55</u>	75.96	65.02	64.72	<u>66.8</u>	60.9	50.0	65.7	64.8

Table1: BINARY CLASSIFICATION

Input Setup	Mathematics					Portuguese				
	NV	NN	SVM	DT	RF	NV	NN	SVM	DT	RF
A	na	0.98	<u>0.52</u>	0.59	0.64	na	0.98	<u>0.52</u>	0.56	0.58
B	na	1.45	<u>0.96</u>	1.06	<u>0.90</u>	na	1.69	<u>0.85</u>	0.90	<u>0.82</u>
C	na	1.67	0.98	1.01	<u>0.94</u>	na	1.84	0.98	1.02	<u>0.97</u>

Table2: LINEAR REGRESSION

Discussion

There are several reasons why the results diverge from those in the original research paper, including that WEKA uses a different version of the C4.5mdecision tree algorithm and the fact that appear to be each and every step was not mentioned in the paper which can be replicated for preprocessing which is a important part to receive some results whilst small, they could certainly affect the experiment results. Further some models can't be tested for some data as Weka does not allows doing that which could also have altered some results.

On the whole the Results in the paper are different from the experiment conducted but the process was same and Max Replication is done as the Results comparing the model is may not be the same but the best model for different situations is same and can be said in the paper including other factors may alter the results but the grade will eventually matter for the result and predicting final grades.

Appendix

Mathematics Results

Discrete DATA (binary classification)

A- Grade was removed

Naive-bayes

The screenshot shows the Orange3 Classifier window with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      297      75.1899 %
Incorrectly Classified Instances    98      24.8101 %
Kappa statistic                    0.439
Mean absolute error                 0.2541
Root mean squared error             0.4234
Relative absolute error             54.494 %
Root relative squared error         87.7109 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
               0.541    0.124    0.718     0.541    0.617     0.449    0.857     0.749    '(-inf--0.3227229953]'
               0.876    0.459    0.765     0.876    0.816     0.449    0.857     0.923    '(-0.3227229953-inf)'
Weighted Avg.   0.752    0.335    0.748     0.752    0.743     0.449    0.857     0.859

=== Confusion Matrix ===
      a  b  <-- classified as
79  67  |  a = '(-inf--0.3227229953]'
31  218 |  b = '(-0.3227229953-inf)'
```

Neural Network

The screenshot shows the Orange3 Classifier window with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
time taken to build model: 10.26 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      295      74.6835 %
Incorrectly Classified Instances    100      25.3165 %
Kappa statistic                    0.4593
Mean absolute error                 0.2518
Root mean squared error             0.4622
Relative absolute error             54.0117 %
Root relative squared error         95.7423 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
               0.664    0.205    0.655     0.664    0.660     0.458    0.839     0.703    '(-inf--0.3227229953]'
               0.795    0.336    0.802     0.795    0.798     0.458    0.839     0.916    '(-0.3227229953-inf)'
Weighted Avg.   0.747    0.287    0.748     0.747    0.747     0.458    0.839     0.837

=== Confusion Matrix ===
      a  b  <-- classified as
97  49  |  a = '(-inf--0.3227229953]'
51  198 |  b = '(-0.3227229953-inf)'
```

SVM

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
☒ Cross-validation Folds 10
☐ Percentage split % 66

More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

```
time taken to build model: 0.11 seconds

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

Correctly Classified Instances      311          78.7342 %
Incorrectly Classified Instances    84           21.2658 %
Kappa statistic                    0.5538
Mean absolute error                 0.2127
Root mean squared error             0.4611
Relative absolute error             45.6144 %
Root relative squared error         95.5309 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.767   0.201   0.691    0.767   0.727     0.556   0.783    0.616   '(-inf--0.3227229953)'
          0.799   0.233   0.854    0.799   0.826     0.556   0.783    0.809   '(-0.3227229953-inf)'
Weighted Avg.  0.787   0.221   0.794    0.787   0.789     0.556   0.783    0.738

=== Confusion Matrix ===

  a  b  <-- classified as
112 34 | a = '(-inf--0.3227229953)'
 50 199 | b = '(-0.3227229953-inf)'
```

Status

OK Log

Decision Tree

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
☒ Cross-validation Folds 10
☐ Percentage split % 66

More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

```
time taken to build model: 0.02 seconds

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

Correctly Classified Instances      293          74.1772 %
Incorrectly Classified Instances    102          25.8228 %
Kappa statistic                    0.4784
Mean absolute error                 0.2751
Root mean squared error             0.402
Relative absolute error             59.0055 %
Root relative squared error         83.2842 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===

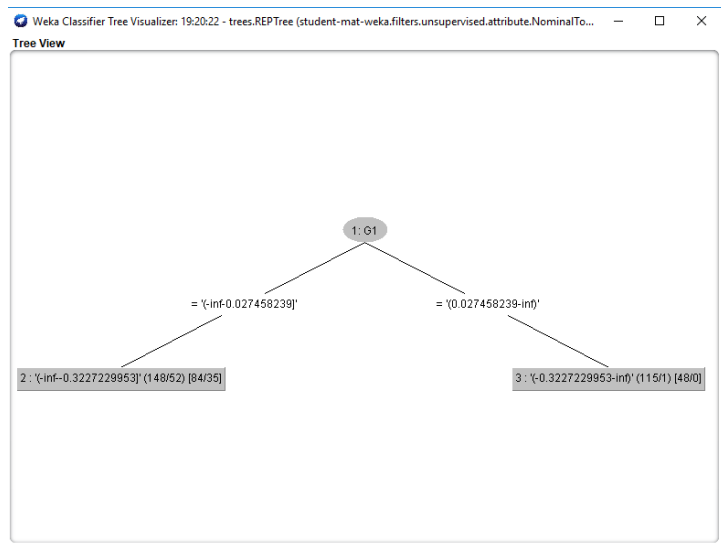
          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0.801   0.293   0.616    0.801   0.696     0.491   0.812    0.645   '(-inf--0.3227229953)'
          0.707   0.199   0.859    0.707   0.775     0.491   0.812    0.880   '(-0.3227229953-inf)'
Weighted Avg.  0.742   0.234   0.769    0.742   0.746     0.491   0.812    0.793

=== Confusion Matrix ===

  a  b  <-- classified as
117 29 | a = '(-inf--0.3227229953)'
 73 176 | b = '(-0.3227229953-inf)'
```

Status

OK Log



Random Forest

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

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

```

time taken to build model: 0.3 seconds

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

Correctly Classified Instances      302      76.4557 %
Incorrectly Classified Instances    93      23.5443 %
Kappa statistic                    0.4882
Mean absolute error                 0.3115
Root mean squared error             0.3755
Relative absolute error             66.8182 %
Root relative squared error         77.7936 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC   ROC Area  PRC Area  Class
      -----  -
Weighted Avg.  0.765    0.283    0.762    0.765    0.763    0.489    0.874    0.876    '(-inf--0.3227229953]'
      -----  -
      a  b  <-- classified as
      95  51 |  a = '(-inf--0.3227229953]'
      42  207 |  b = '(-0.3227229953-inf)'

=== Confusion Matrix ===
  
```

Status

OK Log

B - G2 also removed

Naive Bayes

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes**
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

```
time taken to build model: 0 seconds

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

Correctly Classified Instances      293      74.1772 %
Incorrectly Classified Instances    102      25.8228 %
Kappa statistic                    0.4851
Mean absolute error                 0.2735
Root mean squared error             0.4416
Relative absolute error             56.4177 %
Root relative squared error         89.6882 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0.698    0.196    0.835    0.698    0.761    0.494    0.816    0.867    '(-inf-0.027458239)'
      0.804    0.302    0.652    0.804    0.720    0.494    0.816    0.710    '(0.027458239-inf)'
Weighted Avg.    0.742    0.240    0.759    0.742    0.744    0.494    0.816    0.802

=== Confusion Matrix ===
      a  b  <-- classified as
162  70 |  a = '(-inf-0.027458239)'
 32 131 |  b = '(0.027458239-inf)'
```

Status

OK x0

Neural network

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron**
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

```
time taken to build model: 19.62 seconds

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

Correctly Classified Instances      304      76.962 %
Incorrectly Classified Instances     91      23.038 %
Kappa statistic                    0.5269
Mean absolute error                 0.2356
Root mean squared error             0.4456
Relative absolute error             48.5834 %
Root relative squared error         90.513 %
Total Number of Instances          395

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0.793    0.264    0.811    0.793    0.802    0.527    0.844    0.911    '(-inf-0.027458239)'
      0.736    0.207    0.714    0.736    0.725    0.527    0.844    0.701    '(0.027458239-inf)'
Weighted Avg.    0.770    0.240    0.771    0.770    0.770    0.527    0.844    0.824

=== Confusion Matrix ===
      a  b  <-- classified as
184  48 |  a = '(-inf-0.027458239)'
 43 120 |  b = '(0.027458239-inf)'
```

Status

OK x0

SVM

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

☒ Cross-validation Folds 10

☐ Percentage split % 66

(Num) absences

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

Time taken to build model: 0.22 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	307	77.7215 %
Incorrectly Classified Instances	88	22.2785 %
Kappa statistic	0.5597	
Mean absolute error	0.2228	
Root mean squared error	0.472	
Relative absolute error	45.9486 %	
Root relative squared error	55.8704 %	
Total Number of Instances	395	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.707	0.123	0.891	0.707	0.788	0.577	0.792	0.802	'(-inf-0.027458239]'
	0.877	0.293	0.678	0.877	0.765	0.577	0.792	0.645	'(0.027458239-inf)'
Weighted Avg.	0.777	0.193	0.803	0.777	0.779	0.577	0.792	0.737	

=== Confusion Matrix ===

a	b	<-- classified as
164	68	a = '(-inf-0.027458239]'
20	143	b = '(0.027458239-inf)'

Status

OK

x0

Decision tree

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

☒ Cross-validation Folds 10

☐ Percentage split % 66

(Num) absences

Result list (right-click for options)

- 19:16:03 - bayes.NaiveBayes
- 19:17:05 - functions.MultilayerPerceptron
- 19:20:09 - functions.SMO
- 19:20:22 - trees.REPTree
- 19:20:41 - trees.RandomForest
- 19:21:52 - bayes.NaiveBayes
- 19:22:11 - functions.MultilayerPerceptron
- 20:37:13 - functions.SMO
- 20:37:31 - trees.REPTree
- 20:37:40 - trees.RandomForest
- 20:39:37 - trees.RandomForest

Classifier output

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	309	78.2278 %
Incorrectly Classified Instances	86	21.7722 %
Kappa statistic	0.5704	
Mean absolute error	0.2753	
Root mean squared error	0.3859	
Relative absolute error	56.7838 %	
Root relative squared error	78.3779 %	
Total Number of Instances	395	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.707	0.110	0.901	0.707	0.792	0.589	0.835	0.879	'(-inf-0.027458239]'
	0.890	0.293	0.681	0.890	0.771	0.589	0.835	0.699	'(0.027458239-inf)'
Weighted Avg.	0.782	0.186	0.810	0.782	0.784	0.589	0.835	0.805	

=== Confusion Matrix ===

a	b	<-- classified as
164	68	a = '(-inf-0.027458239]'
18	145	b = '(0.027458239-inf)'

Status

OK

x0

C – G1 also removed

Neural network

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following data:

```
Attrib nursery=no 0.143232800548
Attrib higher=no 0.2808098574880528
Attrib internet=yes -0.16937228699372708
Attrib romantic=yes 0.01944085475692102
Attrib famrel -0.09183393108266274
Attrib freetime -0.06604415091304126
Attrib goout 0.07313960289388502
Attrib dalc 0.11536883471222975
Attrib walc -0.010224619346560904
Attrib health -0.07655077853329864

Class
Input
Node 0
```

Time taken to build model: 9.66 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.1191
Mean absolute error	1.1477
Root mean squared error	1.6788
Relative absolute error	176.0589 %
Root relative squared error	167.4587 %
Total Number of Instances	395

The 'Result list' on the left shows various models, with '20.44.43 - functions.MultiplePerceptron' selected.

SVM

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following data:

```
- 0.045 * (normalized) nigner=no
+ 0.0191 * (normalized) internet=yes
+ 0.0032 * (normalized) romantic=yes
- 0.0223 * (normalized) famrel
- 0.0056 * (normalized) freetime
+ 0.0206 * (normalized) goout
+ 0.0225 * (normalized) dalc
+ 0.0333 * (normalized) walc
+ 0.005 * (normalized) health
+ 0.0556
```

Number of kernel evaluations: 78210 (99.578% cached)

Time taken to build model: 2.43 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.2639
Mean absolute error	0.5747
Root mean squared error	0.9877
Relative absolute error	88.1563 %
Root relative squared error	98.5198 %
Total Number of Instances	395

The 'Result list' on the left shows various models, with '20.46.54 - functions.SMOreg' selected.

Decision tree

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
☒ Cross-validation Folds 10
☐ Percentage split % 66

More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 20:40:48 - functions.MultilayerPerceptron
- 20:43:06 - functions.SMO
- 20:43:19 - trees.REPTree
- 20:43:28 - trees.RandomForest
- 20:44:43 - functions.MultilayerPerceptron
- 20:46:54 - functions.SMOreg
- 20:47:33 - trees.REPTree
- 20:47:45 - trees.RandomForest
- 20:54:30 - functions.LinearRegression
- 20:54:42 - functions.MultilayerPerceptron
- 20:56:45 - functions.SMOreg

Classifier output

```
Ualc
Walc
health
absences

Test mode: 10-fold cross-validation

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

REPTree
=====
: -0 (263/1.16) [132/0.67]

Size of the tree : 1

Time taken to build model: 0.03 seconds

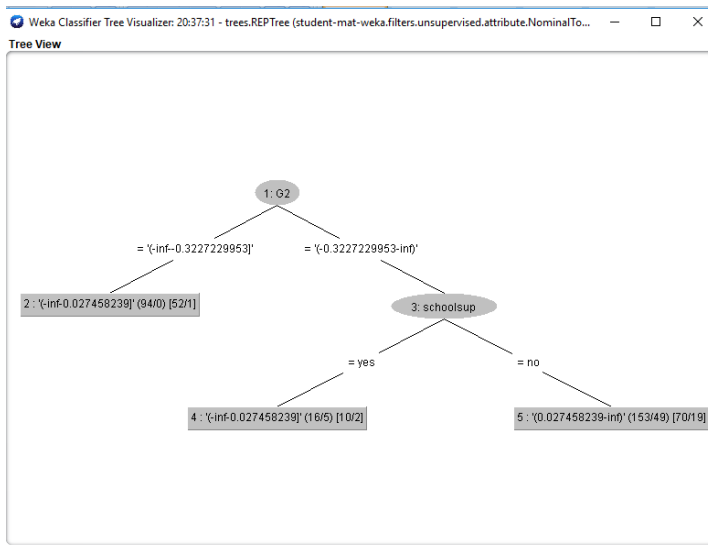
=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.0308
Mean absolute error         0.6582
Root mean squared error     1.0155
Relative absolute error     100.9627 %
Root relative squared error 101.2988 %
Total Number of Instances   395
```

Status

OK

Log x0



Random Forest

The screenshot shows the Weka GUI with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the following text:

```
goout
Dalc
Walc
health
absences

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: 0.37 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.1894
Mean absolute error         0.6471
Root mean squared error     0.9896
Relative absolute error     99.2618 %
Root relative squared error  98.7091 %
Total Number of Instances   395
```

The 'Result list' on the left shows a list of models, with '20.47.45 - trees RandomForest' selected. The 'Status' bar at the bottom shows 'OK'.

Linear regression

A- Final Grade was removed

Neural network

This screenshot is identical to the one above, showing the Weka GUI with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the same text as above:

```
goout
Dalc
Walc
health
absences

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: 0.37 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.1894
Mean absolute error         0.6471
Root mean squared error     0.9896
Relative absolute error     99.2618 %
Root relative squared error  98.7091 %
Total Number of Instances   395
```

The 'Result list' on the left shows a list of models, with '20.47.45 - trees RandomForest' selected. The 'Status' bar at the bottom shows 'OK'.

SVM

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
- 0.0243 * (normalized) romantic=yes
- 0.0192 * (normalized) famrel
- 0.0133 * (normalized) freetime
- 0.018 * (normalized) goout
+ 0.0228 * (normalized) dalc
- 0.0009 * (normalized) walc
- 0.0045 * (normalized) health
- 0.0579 * (normalized) absences
+ 0.7721 * (normalized) G1
+ 0.2422
```

Number of kernel evaluations: 78210 (99.549% cached)
Time taken to build model: 1.96 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.8487
Mean absolute error	0.3488
Root mean squared error	0.5291
Relative absolute error	44.4091 %
Root relative squared error	52.7163 %
Total Number of Instances	395

The 'Result list' on the left shows various models, with '20.56.45 - functions.SMOreg' selected. The 'Status' bar at the bottom shows 'OK'.

Decision Tree

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
| | | | | failures >= 0.22 : 0.21 (2/0.02) [0/0]
| | | | | address = R : 0.3 (6/0.03) [0/0]
| | | | | absences >= 0.16 : 0.36 (10/0.06) [4/0.08]
| | | | | G1 >= 1.08
| | | | | G1 < 1.38
| | | | | age < -0.15 : 1.21 (5/0.01) [6/0.04]
| | | | | age >= -0.15 : 0.79 (6/0.09) [7/0.04]
| | | | | G1 >= 1.38
| | | | | G1 < 1.99
| | | | | Pstatus = A : 1.87 (4/0.01) [0/0]
| | | | | Pstatus = T : 1.3 (18/0.11) [8/0.1]
| | | | | G1 >= 1.99 : 1.96 (6/0.07) [5/0.01]
```

Size of the tree : 59
Time taken to build model: 0.33 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.809
Mean absolute error	0.4098
Root mean squared error	0.5937
Relative absolute error	52.1739 %
Root relative squared error	59.1428 %
Total Number of Instances	395

The 'Result list' on the left shows various models, with '20.57.15 - trees.REPTree' selected. The 'Status' bar at the bottom shows 'OK'.

Random forest

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the following text:

```
waic
health
absences
G1
G2
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: 0.41 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.7979
Mean absolute error         0.4518
Root mean squared error     0.6412
Relative absolute error     57.5266 %
Root relative squared error 63.8932 %
Total Number of Instances   395
```

The 'Result list' on the left shows a list of models with their scores. The 'Status' bar at the bottom indicates 'OK'.

B – G2 also removed

neural network

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the following text:

```
Attrib nigner=no 0.06452308971306797
Attrib internet=yes -0.28935794787452107
Attrib romantic=yes 1.0586397413276494
Attrib famrel 0.24931712719647142
Attrib freetime -0.5381631523609565
Attrib goout 1.0496356575107244
Attrib dalc -0.6549481418934696
Attrib walc -2.3937213707067015
Attrib health -2.2007095510651884
Attrib absences -0.6657499072224784

Class
Input
Node 0

Time taken to build model: 14.26 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.2206
Mean absolute error         1.1469
Root mean squared error     1.4544
Relative absolute error     137.8781 %
Root relative squared error 145.2513 %
Total Number of Instances   395
```

The 'Result list' on the left shows a list of models with their scores. The 'Status' bar at the bottom indicates 'OK'.

SVM

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Result list' on the left shows various models, with '21:03:37 - functions.SMOreg' selected. The 'Classifier output' pane displays the following results:

```
- 0.0027 * (normalized) internet=yes
- 0.0086 * (normalized) romantic=yes
- 0.0244 * (normalized) famrel
+ 0.1264 * (normalized) freetime
- 0.1577 * (normalized) goout
+ 0.0318 * (normalized) dalc
- 0.0192 * (normalized) walc
+ 0.0156 * (normalized) health
- 0.0015 * (normalized) absences
+ 0.3694
```

Number of kernel evaluations: 78210 (98.978% cached)
Time taken to build model: 0.97 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.3983
Mean absolute error	0.7884
Root mean squared error	0.9665
Relative absolute error	94.7853 %
Root relative squared error	96.5178 %
Total Number of Instances	395

Status: OK

Decision tree

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Result list' on the left shows various models, with '21:03:37 - functions.SMOreg' selected. The 'Classifier output' pane displays the following results:

```
absences
G1
Test mode: 10-fold cross-validation

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

REPTree
=====
failures < 0.22 : 0.17 (213/0.87) [59/1]
failures >= 0.22 : -0.63 (50/0.8) [33/0.92]

Size of the tree : 3

Time taken to build model: 0.02 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.2349
Mean absolute error         0.8245
Root mean squared error     1.0043
Relative absolute error     99.1168 %
Root relative squared error 100.3 %
Total Number of Instances   395
```

Status: OK

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 20:57:15 - trees REPTree
- 20:58:41 - trees RandomForest
- 21:00:20 - functions.MultilayerPerceptron
- 21:03:37 - functions.SMOreg
- 21:03:57 - trees REPTree
- 21:04:10 - trees RandomForest
- 21:04:51 - functions.MultilayerPerceptron
- 21:06:36 - functions.SMOreg
- 21:07:07 - trees REPTree
- 21:07:16 - trees RandomForest
- 21:08:47 - trees RandomForest

Classifier output

```

Attrib nursery=no      0.143132800348
Attrib higher=no       0.2808098574880528
Attrib internet=yes    -0.16937228699372708
Attrib romantic=yes    0.01944085475692102
Attrib famrel          -0.09163393108256274
Attrib freetime        -0.06604415091304126
Attrib goout           0.07313960289388502
Attrib dalc            0.11536883471222975
Attrib walc            -0.010224619346560984
Attrib health          -0.07655077853329864

Class
Input
Node 0


Time taken to build model: 9.61 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.1191
Mean absolute error         1.1477
Root mean squared error     1.6788
Relative absolute error     176.0589 %
Root relative squared error 167.4587 %
Total Number of Instances   395

```

Status

OK Log  x 0

SVM

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 20:57:15 - trees REPTree
- 20:58:41 - trees RandomForest
- 21:00:20 - functions.MultilayerPerceptron
- 21:03:37 - functions.SMOreg
- 21:03:57 - trees REPTree
- 21:04:10 - trees RandomForest
- 21:04:51 - functions.MultilayerPerceptron
- 21:06:36 - functions.SMOreg
- 21:07:07 - trees REPTree
- 21:07:16 - trees RandomForest
- 21:08:47 - trees RandomForest

Classifier output

```

- 0.045 * (normalized) higher=no
+ 0.0191 * (normalized) internet=yes
+ 0.0032 * (normalized) romantic=yes
- 0.0223 * (normalized) famrel
- 0.0056 * (normalized) freetime
+ 0.0206 * (normalized) goout
+ 0.0225 * (normalized) dalc
+ 0.0333 * (normalized) walc
+ 0.005 * (normalized) health
+ 0.0556

Number of kernel evaluations: 78210 (99.578% cached)


Time taken to build model: 2.73 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.2639
Mean absolute error         0.5747
Root mean squared error     0.9877
Relative absolute error     88.1563 %
Root relative squared error 98.5198 %
Total Number of Instances   395

```

Status

OK Log  x 0

Decision Tree

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 20:57:15 - trees.REPTree
- 20:58:41 - trees.RandomForest
- 21:00:20 - functions.MultilayerPerceptron
- 21:03:37 - functions.SMOreg
- 21:03:57 - trees.REPTree
- 21:04:10 - trees.RandomForest
- 21:04:51 - functions.MultilayerPerceptron
- 21:06:36 - functions.SMOreg
- 21:07:07 - trees.REPTree
- 21:07:16 - trees.RandomForest
- 21:08:47 - trees.RandomForest

Classifier output

```

Ualc
Walc
health
absences
Test mode: 10-fold cross-validation

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

REPTree
=====
: -0 (263/1.16) [132/0.67]

Size of the tree : 1

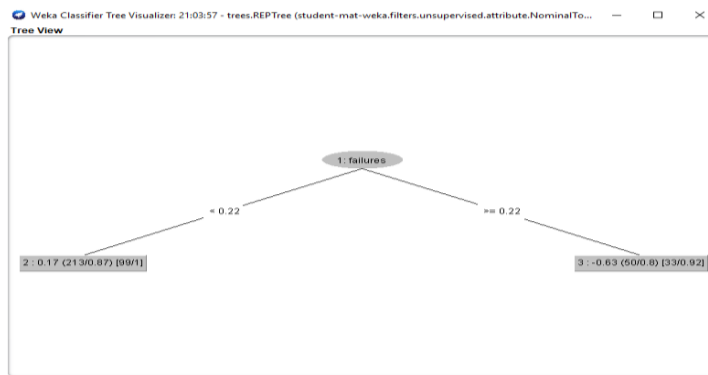
Time taken to build model: 0.02 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.0308
Mean absolute error         0.6582
Root mean squared error     1.0155
Relative absolute error     100.9627 %
Root relative squared error 101.2988 %
Total Number of Instances   395
  
```

Status

OK Log



Random Forest

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 20:57:15 - trees.REPTree
- 20:58:41 - trees.RandomForest
- 21:00:20 - functions.MultilayerPerceptron
- 21:03:37 - functions.SMOreg
- 21:03:57 - trees.REPTree
- 21:04:10 - trees.RandomForest
- 21:04:51 - functions.MultilayerPerceptron
- 21:06:36 - functions.SMOreg
- 21:07:07 - trees.REPTree
- 21:07:16 - trees.RandomForest
- 21:08:47 - trees.RandomForest

Classifier output

```

gout
Dalc
Walc
health
absences
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: 0.4 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.1894
Mean absolute error         0.6471
Root mean squared error     0.9896
Relative absolute error     99.2618 %
Root relative squared error 98.7091 %
Total Number of Instances   395
  
```

Status

OK Log

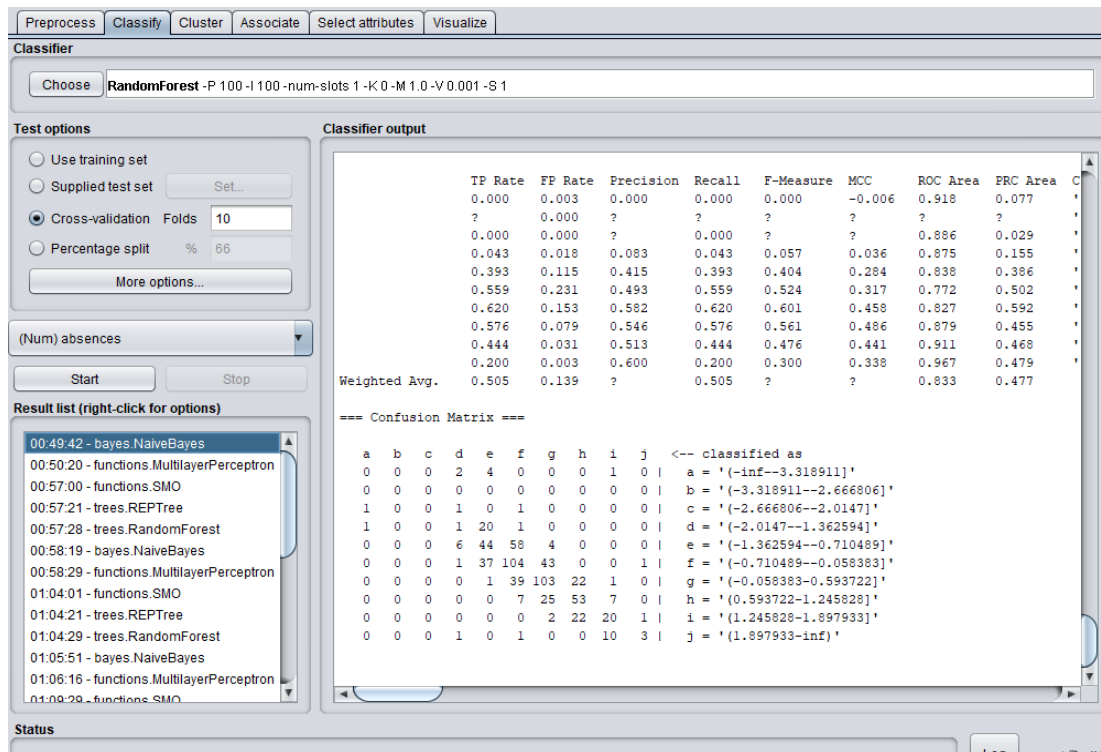


PROTUGESE RESULTS

BINARY CLASSIFICATION

A – Final grade removed

Naive bayes



Neural network

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.000	0.008	0.000	0.000	0.000	-0.009	0.735	0.036
	?	0.000	?	?	?	?	?	?
	0.000	0.000	?	0.000	?	?	0.417	0.005
	0.043	0.030	0.050	0.043	0.047	0.014	0.736	0.097
	0.384	0.130	0.381	0.384	0.382	0.253	0.747	0.354
	0.505	0.246	0.452	0.505	0.477	0.251	0.713	0.466
	0.440	0.172	0.468	0.440	0.453	0.274	0.734	0.457
	0.478	0.092	0.463	0.478	0.471	0.382	0.818	0.419
	0.333	0.046	0.349	0.333	0.341	0.293	0.855	0.315
	0.200	0.009	0.333	0.200	0.250	0.245	0.858	0.337
Weighted Avg.	0.421	0.155	?	0.421	?	?	0.752	0.404

==== Confusion Matrix ====

```
a b c d e f g h i j <-- classified as
0 0 0 1 3 1 1 0 1 0 | a = '(-inf--3.318911)'\n
0 0 0 0 0 0 0 0 0 0 | b = '(-3.318911--2.666806)'\n
0 0 0 0 0 0 0 1 0 2 | c = '(-2.666806--2.0147)'\n
0 0 0 1 16 5 0 0 1 0 | d = '(-2.0147--1.362594)'\n
2 0 0 7 43 50 9 1 0 0 | e = '(-1.362594--0.710489)'\n
1 0 0 7 39 94 37 7 0 1 | f = '(-0.710489--0.058383)'\n
2 0 0 1 10 50 73 25 5 0 | g = '(-0.058383-0.593722)'\n
0 0 0 1 1 7 26 44 13 0 | h = '(0.593722-1.245828)'\n
0 0 0 1 1 0 9 16 15 3 | i = '(1.245828-1.897933)'\n
0 0 0 1 0 0 1 1 1 8 | j = '(1.897933-inf)'
```

SVM

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.000	0.005	0.000	0.000	0.000	-0.007	0.915	0.061
	?	0.000	?	?	?	?	?	?
	0.000	0.000	?	0.000	?	?	0.309	0.005
	0.043	0.000	1.000	0.043	0.083	0.205	0.930	0.253
	0.643	0.153	0.468	0.643	0.541	0.435	0.845	0.427
	0.570	0.168	0.576	0.570	0.573	0.403	0.780	0.502
	0.578	0.124	0.615	0.578	0.596	0.464	0.819	0.525
	0.652	0.083	0.566	0.652	0.606	0.537	0.886	0.481
	0.422	0.028	0.528	0.422	0.469	0.437	0.898	0.391
	0.267	0.008	0.444	0.267	0.333	0.333	0.980	0.380
Weighted Avg.	0.552	0.120	?	0.552	?	?	0.834	0.465

==== Confusion Matrix ====

```
a b c d e f g h i j <-- classified as
0 0 0 0 7 0 0 0 0 0 | a = '(-inf--3.318911)'\n
0 0 0 0 0 0 0 0 0 0 | b = '(-3.318911--2.666806)'\n
1 0 0 0 2 0 0 0 0 0 | c = '(-2.666806--2.0147)'\n
0 0 0 1 22 0 0 0 0 0 | d = '(-2.0147--1.362594)'\n
2 0 0 0 72 37 1 0 0 0 | e = '(-1.362594--0.710489)'\n
0 0 0 0 47 106 32 1 0 0 | f = '(-0.710489--0.058383)'\n
0 0 0 0 4 40 96 26 0 0 | g = '(-0.058383-0.593722)'\n
0 0 0 0 0 1 25 60 6 0 | h = '(0.593722-1.245828)'\n
0 0 0 0 0 0 2 19 19 5 | i = '(1.245828-1.897933)'\n
0 0 0 0 0 0 0 0 0 11 | j = '(1.897933-inf)'
```

Decision Tree

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

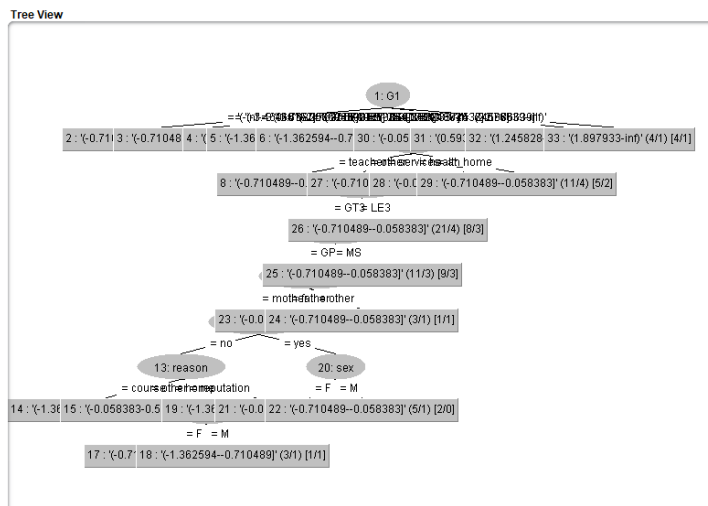
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.429	0.008	0.375	0.429	0.400	0.394	0.885	0.178	
?	0.000	?	?	?	?	?	?	
0.000	0.000	?	0.000	?	?	0.457	0.005	
0.000	0.006	0.000	0.000	0.000	-0.015	0.753	0.120	
0.464	0.115	0.456	0.464	0.460	0.346	0.798	0.420	
0.629	0.216	0.539	0.629	0.581	0.396	0.781	0.515	
0.578	0.122	0.619	0.578	0.598	0.467	0.806	0.528	
0.630	0.084	0.552	0.630	0.589	0.517	0.852	0.476	
0.489	0.025	0.595	0.489	0.537	0.508	0.874	0.433	
0.400	0.005	0.667	0.400	0.500	0.508	0.979	0.443	
Weighted Avg.	0.545	0.127	?	0.545	?	?	0.810	0.469

==== Confusion Matrix ====

```

a b c d e f g h i j <-- classified as
3 0 0 0 3 1 0 0 0 0 | a = '(-inf--3.318911)'
0 0 0 0 0 0 0 0 0 0 | b = '(-3.318911--2.666806)'
1 0 0 1 1 0 0 0 0 0 | c = '(-2.666806--2.0147)'
1 0 0 0 19 3 0 0 0 0 | d = '(-2.0147--1.362594)'
2 0 0 3 52 54 1 0 0 0 | e = '(-1.362594--0.710489)'
1 0 0 0 37 117 29 2 0 0 | f = '(-0.710489--0.058383)'
0 0 0 0 2 41 96 27 0 0 | g = '(-0.058383-0.593722)'
0 0 0 0 0 1 27 58 6 0 | h = '(0.593722-1.245828)'
0 0 0 0 0 0 2 18 22 3 | i = '(1.245828-1.897933)'
0 0 0 0 0 0 0 0 0 9 6 | j = '(1.897933-inf)'
  
```

Status



Random forest

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest**
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:06:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.003	0.000	0.000	0.000	0.000	-0.006	0.903	0.088
?	0.000	?	?	?	?	?	?	?
0.000	0.000	?	0.000	?	?	?	0.523	0.006
0.043	0.006	0.200	0.043	0.071	0.078	0.078	0.846	0.158
0.429	0.119	0.429	0.429	0.429	0.309	0.309	0.809	0.389
0.554	0.233	0.488	0.554	0.519	0.309	0.309	0.755	0.526
0.578	0.176	0.530	0.578	0.553	0.391	0.391	0.813	0.537
0.489	0.072	0.529	0.489	0.508	0.431	0.431	0.868	0.479
0.467	0.035	0.500	0.467	0.483	0.446	0.446	0.856	0.406
0.333	0.009	0.455	0.333	0.385	0.377	0.377	0.977	0.372
Weighted Avg.	0.492	0.146	?	0.492	?	?	0.811	0.466

==== Confusion Matrix ====

```

a b c d e f g h i j <-- classified as
0 0 0 0 3 4 0 0 0 0 | a = '(-inf--3.318911)'
```

B – G2 also removed

Naïve- Bayes

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes**
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:06:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.002	0.000	0.000	0.000	0.000	-0.002	0.934	0.023
?	0.000	?	?	?	?	?	?	?
0.000	0.003	0.000	0.000	0.000	0.000	-0.006	0.821	0.034
0.071	0.061	0.075	0.071	0.073	0.011	0.011	0.750	0.127
0.243	0.100	0.325	0.243	0.278	0.162	0.162	0.710	0.321
0.366	0.359	0.291	0.366	0.324	0.007	0.007	0.526	0.291
0.429	0.305	0.304	0.429	0.356	0.111	0.111	0.605	0.289
0.189	0.094	0.282	0.189	0.226	0.112	0.112	0.639	0.263
0.026	0.005	0.250	0.026	0.048	0.064	0.064	0.601	0.120
0.000	0.000	?	0.000	?	?	?	0.460	0.012
Weighted Avg.	0.284	0.211	?	0.284	?	?	0.615	0.263

==== Confusion Matrix ====

```

a b c d e f g h i j <-- classified as
0 0 0 0 0 1 0 0 0 | a = '(-inf--3.460167)'
```

Neural Network

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron**
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.000	?	0.000	?	?	?	0.060	0.002
?	0.000	?	?	?	?	?	?	?
0.000	0.008	0.000	0.000	0.000	0.000	-0.009	0.690	0.039
0.095	0.059	0.100	0.095	0.098	0.037	0.646	0.110	?
0.271	0.131	0.290	0.271	0.280	0.144	0.620	0.250	?
0.296	0.311	0.276	0.296	0.286	-0.015	0.504	0.304	?
0.305	0.240	0.283	0.305	0.294	0.063	0.581	0.287	?
0.198	0.153	0.202	0.198	0.200	0.046	0.578	0.203	?
0.079	0.039	0.111	0.079	0.092	0.047	0.592	0.135	?
0.000	0.012	0.000	0.000	0.000	-0.012	0.478	0.013	?
Weighted Avg.	0.245	0.199	?	0.245	?	?	0.569	0.245

=== Confusion Matrix ===

```

a b c d e f g h i j <-- classified as
0 0 0 0 1 0 0 0 0 0 | a = '(-inf--3.460167)'
```

SVM

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO**
- 01:04:21 - trees.REPTree
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.002	0.000	0.000	0.000	0.000	-0.002	0.049	0.002
?	0.000	?	?	?	?	?	?	?
0.000	0.002	0.000	0.000	0.000	0.000	-0.004	0.767	0.031
0.000	0.002	0.000	0.000	0.000	0.000	-0.010	0.775	0.154
0.327	0.059	0.522	0.327	0.402	0.327	0.707	0.332	?
0.591	0.523	0.313	0.591	0.409	0.062	0.540	0.307	?
0.338	0.283	0.271	0.338	0.301	0.051	0.588	0.282	?
0.104	0.044	0.314	0.104	0.156	0.097	0.615	0.221	?
0.000	0.000	?	0.000	?	?	0.582	0.113	?
0.000	0.000	?	0.000	?	?	0.476	0.012	?
Weighted Avg.	0.320	0.234	?	0.320	?	?	0.610	0.263

=== Confusion Matrix ===

```

a b c d e f g h i j <-- classified as
0 0 0 0 0 0 1 0 0 0 | a = '(-inf--3.460167)'
```

Decision Tree

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
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

(Num) absences **▼**

Result list (right-click for options)

- 00:49:42 - bayes.NaiveBayes
- 00:50:20 - functions.MultilayerPerceptron
- 00:57:00 - functions.SMO
- 00:57:21 - trees.REPTree
- 00:57:28 - trees.RandomForest
- 00:58:19 - bayes.NaiveBayes
- 00:58:29 - functions.MultilayerPerceptron
- 01:04:01 - functions.SMO
- 01:04:21 - trees.REPTree**
- 01:04:29 - trees.RandomForest
- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.000	?	0.000	?	?	?	0.477	0.002
?	0.000	?	?	?	?	?	?	?
0.000	0.003	0.000	0.000	0.000	0.000	-0.006	0.575	0.015
0.000	0.020	0.000	0.000	0.000	0.000	-0.036	0.642	0.091
0.290	0.089	0.392	0.290	0.333	0.333	0.228	0.657	0.347
0.505	0.486	0.295	0.505	0.372	0.372	0.018	0.520	0.294
0.435	0.289	0.319	0.435	0.368	0.368	0.133	0.611	0.293
0.038	0.039	0.160	0.038	0.061	0.061	-0.002	0.544	0.185
0.000	0.002	0.000	0.000	0.000	0.000	-0.010	0.609	0.091
0.000	0.002	0.000	0.000	0.000	0.000	-0.004	0.442	0.014
Weighted Avg.	0.302	0.230	?	0.302	?	?	0.581	0.253

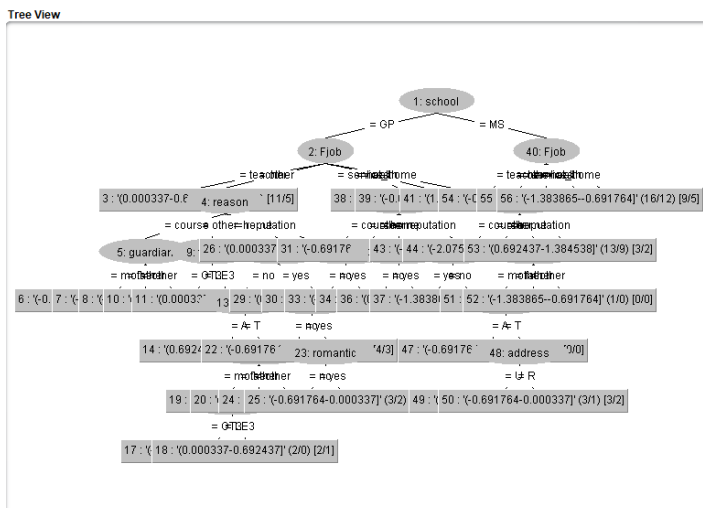
==== Confusion Matrix ====

```

a b c d e f g h i j <-- classified as
0 0 0 0 0 0 1 0 0 0 | a = '(-inf--3.460167)!'
0 0 0 0 0 0 0 0 0 0 | b = '(-3.460167--2.768066)!'
0 0 0 0 2 5 0 0 0 0 | c = '(-2.768066--2.075965)!'
0 0 1 0 12 23 5 1 0 0 | d = '(-2.075965--1.383865)!'
0 0 1 5 31 52 16 2 0 0 | e = '(-1.383865--0.691764)!'
0 0 0 4 21 94 53 14 0 0 | f = '(-0.691764--0.000337)!'
0 0 0 2 6 76 67 2 1 0 | g = '(0.000337--0.692437)!'
0 0 0 1 6 52 43 4 0 0 | h = '(0.692437--1.384538)!'
0 0 0 0 1 13 22 1 0 1 | i = '(1.384538--2.076639)!'
0 0 0 0 0 4 3 1 0 0 | j = '(2.076639--inf)!'

```

Status



Random forest

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...

(Num) absences

Start Stop

Result list (right-click for options)

00:49:42 - bayes.NaiveBayes

00:50:20 - functions.MultilayerPerceptron

00:57:00 - functions.SMO

00:57:21 - trees.REPTree

00:57:28 - trees.RandomForest

00:58:19 - bayes.NaiveBayes

00:58:29 - functions.MultilayerPerceptron

01:04:01 - functions.SMO

01:04:21 - trees.REPTree

01:04:29 - trees.RandomForest

01:05:51 - bayes.NaiveBayes

01:06:16 - functions.MultilayerPerceptron

01:06:29 - functions.SMO

Classifier output

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.000	0.000	?	0.000	?	?	?	0.471	0.002
?	0.000	?	?	?	?	?	?	?
0.000	0.002	0.000	0.000	0.000	0.000	-0.004	0.707	0.026
0.143	0.038	0.207	0.143	0.169	0.125	0.125	0.718	0.150
0.262	0.109	0.322	0.262	0.289	0.166	0.166	0.685	0.315
0.419	0.365	0.316	0.419	0.360	0.051	0.051	0.549	0.347
0.357	0.265	0.296	0.357	0.324	0.087	0.087	0.568	0.278
0.132	0.105	0.197	0.132	0.158	0.032	0.032	0.558	0.190
0.026	0.039	0.040	0.026	0.032	-0.016	-0.016	0.516	0.067
0.000	0.005	0.000	0.000	0.000	-0.008	-0.008	0.510	0.012
Weighted Avg.	0.280	0.207	?	0.280	?	?	0.588	0.263

Weighted Avg.

0.280 0.207 ? 0.280 ? ? 0.588 0.263

Confusion Matrix

```

a b c d e f g h i j <-- classified as
0 0 0 0 0 0 1 0 0 0 | a = '(-inf--3.460167)''
0 0 0 0 0 0 0 0 0 0 | b = '(-3.460167--2.768066)''
0 0 0 1 3 3 0 0 0 0 | c = '(-2.768066--2.075965)''
0 0 0 6 14 12 5 2 2 1 | d = '(-2.075965--1.383865)''
0 0 0 8 28 43 17 7 4 0 | e = '(-1.383865--0.691764)''
0 0 1 8 22 78 49 22 5 1 | f = '(-0.691764--0.000337)''
0 0 0 3 14 58 55 16 7 1 | g = '(0.000337--0.692437)''
0 0 0 1 4 41 40 14 6 0 | h = '(0.692437--1.384538)''
0 0 0 2 2 8 16 9 1 0 | i = '(1.384538--2.076639)''
0 0 0 0 0 4 3 1 0 0 | j = '(2.076639--inf)''

```

Status

C – G1 is also removed

Naïve - Bayes

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...

(Num) absences

Start Stop

Result list (right-click for options)

00:49:42 - bayes.NaiveBayes

00:50:20 - functions.MultilayerPerceptron

00:57:00 - functions.SMO

00:57:21 - trees.REPTree

00:57:28 - trees.RandomForest

00:58:19 - bayes.NaiveBayes

00:58:29 - functions.MultilayerPerceptron

01:04:01 - functions.SMO

01:04:21 - trees.REPTree

01:04:29 - trees.RandomForest

01:05:51 - bayes.NaiveBayes

01:06:16 - functions.MultilayerPerceptron

01:06:29 - functions.SMO

Classifier output

Time taken to build model: 0 seconds

Stratified cross-validation

Summary

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error	Total Number of Instances
	490	159	0.1084	0.3228	0.4174	90.2895 %	98.786 %	649

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
0.252	0.092	0.452	0.252	0.323	0.201	0.668	0.364	
0.908	0.748	0.800	0.908	0.850	0.201	0.668	0.852	
Weighted Avg.	0.755	0.596	0.719	0.755	0.728	0.201	0.668	0.739

Confusion Matrix

```

a b <-- classified as
38 113 | a = no
46 452 | b = yes

```

Status

Neural Network

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

Time taken to build model: 16.99 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	458	70.5701 %
Incorrectly Classified Instances	191	29.4299 %
Kappa statistic	0.1583	
Mean absolute error	0.3003	
Root mean squared error	0.5089	
Relative absolute error	83.9861 %	
Root relative squared error	120.4332 %	
Total Number of Instances	649	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	C
Weighted Avg.	0.338	0.183	0.359	0.338	0.348	0.158	0.609	0.331	n
	0.817	0.662	0.803	0.817	0.810	0.158	0.609	0.817	y
	0.706	0.551	0.700	0.706	0.702	0.158	0.609	0.704	

=== Confusion Matrix ===

a	b	<-- classified as
51	100	a = no
91	407	b = yes

Status

SVM

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

Time taken to build model: 0.15 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	498	76.7334 %
Incorrectly Classified Instances	151	23.2666 %
Kappa statistic	0	
Mean absolute error	0.2327	
Root mean squared error	0.4824	
Relative absolute error	65.0709 %	
Root relative squared error	114.1565 %	
Total Number of Instances	649	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	C
Weighted Avg.	0.000	0.000	?	0.000	?	?	0.500	0.233	n
	1.000	1.000	0.767	1.000	0.868	?	0.500	0.767	y
	0.767	0.767	?	0.767	?	?	0.500	0.643	

=== Confusion Matrix ===

a	b	<-- classified as
0	151	a = no
0	498	b = yes

Status

Decision Tree

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
☒ Cross-validation Folds
☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	498	76.7334 %
Incorrectly Classified Instances	151	23.2666 %
Kappa statistic	0.0843	
Mean absolute error	0.3287	
Root mean squared error	0.4137	
Relative absolute error	91.9191 %	
Root relative squared error	97.5044 %	
Total Number of Instances	649	

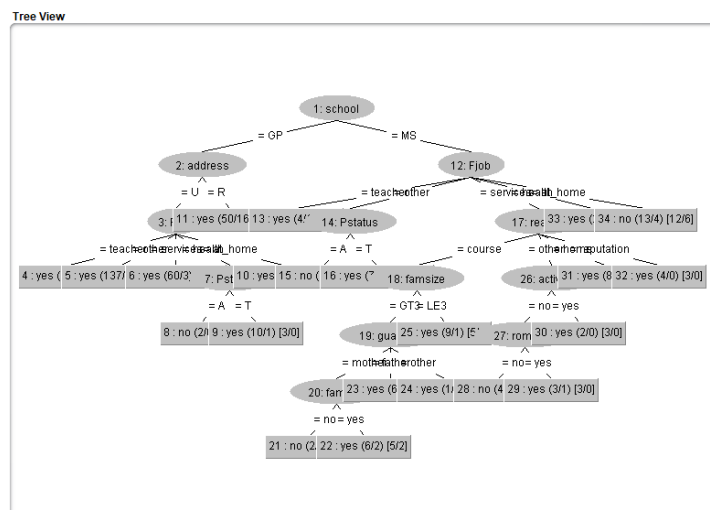
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	C
no	0.086	0.026	0.500	0.086	0.147	0.129	0.657	0.353	n
yes	0.974	0.914	0.778	0.974	0.865	0.129	0.657	0.846	y
Weighted Avg.	0.767	0.707	0.714	0.767	0.698	0.129	0.657	0.731	

=== Confusion Matrix ===

a	b	<-- classified as
13	138	a = no
13	485	b = yes

Status



Random Forest

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
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest**
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

Time taken to build model: 0.37 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      483           74.4222 %
Incorrectly Classified Instances    166           25.5778 %
Kappa statistic                    0.1115
Mean absolute error                 0.324
Root mean squared error             0.4252
Relative absolute error             90.6044 %
Root relative squared error         100.6197 %
Total Number of Instances          649

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  C
Weighted Avg.    0.744    0.654    0.693    0.744    0.705    0.125    0.648    0.342    n
               0.918    0.828    0.785    0.918    0.846    0.125    0.648    0.848    y

=== Confusion Matrix ===
      a  b  <-- classified as
26 125 | a = no
41 457 | b = yes
  
```

Status

Linear regression

A - Final grade removed

Neural Network

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
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron**
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

Attrib higher=no    0.5320237721567196
Attrib innernet=yes 0.28750619291392254
Attrib romantic=yes -0.5600499094943088
Attrib famrel      -0.7245216610996504
Attrib freetime     0.6136458339503047
Attrib goout       -0.5197541022733801
Attrib Dalc        -0.25023244554508356
Attrib Walc        -0.18278744252331475
Attrib health      -0.3251826057594989
Attrib absences     0.5164250657809732
Attrib G1          -2.2115704978314485

Class
Input
Node 0

Time taken to build model: 14.62 seconds

=== Cross-validation ===
=== Summary ===
Correlation coefficient            0.6215
Mean absolute error                0.7253
Root mean squared error           0.9816
Relative absolute error            93.7281 %
Root relative squared error        97.9672 %
Total Number of Instances          649
  
```

Status

SVM

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

+ 0.0118 * (normalized) internet=yes
- 0.0043 * (normalized) romantic=yes
+ 0.0237 * (normalized) famrel
+ 0.0066 * (normalized) freetime
- 0.0277 * (normalized) goout
+ 0.0202 * (normalized) dalc
- 0.0152 * (normalized) walc
- 0.0209 * (normalized) health
- 0.0292 * (normalized) absences
+ 0.8405 * (normalized) G1
+ 0.0791

Number of kernel evaluations: 210925 (99.478% cached)

Time taken to build model: 6.16 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.8537
Mean absolute error         0.3625
Root mean squared error     0.5213
Relative absolute error     46.8418 %
Root relative squared error 52.026 %
Total Number of Instances   649

```

Status

Decision Tree

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...

(Num) absences

Start Stop

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

| | G1 < 0.4 : 0.16 (60/0.17) [22/0.12]
| | G1 >= 0.4
| | | age < -0.2
| | | | G1 < 0.77 : 0.38 (33/0.12) [15/0.13]
| | | | G1 >= 0.77 : 0.73 (24/0.09) [14/0.19]
| | | | age >= -0.2 : 0.87 (33/0.21) [24/0.24]
| | G1 >= 1.13
| | | G1 < 1.86
| | | | G1 < 1.49 : 1.13 (23/0.16) [12/0.11]
| | | | G1 >= 1.49 : 1.5 (15/0.15) [7/0.03]
| | | | G1 >= 1.86
| | | | age < -0.2 : 1.75 (3/0.03) [3/0.03]
| | | | age >= -0.2 : 2.15 (14/0.03) [4/0.02]

Size of the tree : 25

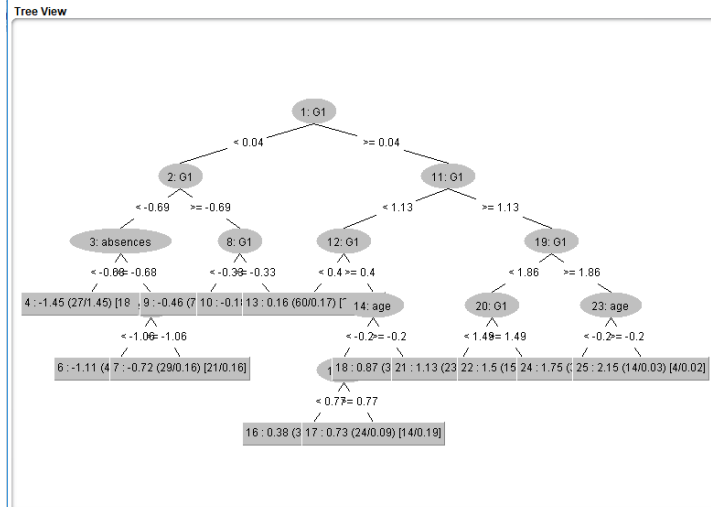
Time taken to build model: 0.03 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.8279
Mean absolute error         0.3916
Root mean squared error     0.5619
Relative absolute error     50.6053 %
Root relative squared error 56.0795 %
Total Number of Instances   649

```

Status



Random Forest

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

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

Dalc
Walc
health
absences
G1
G2

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: 0.58 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.8353
Mean absolute error         0.4105
Root mean squared error     0.5908
Relative absolute error     53.0559 %
Root relative squared error  57.9676 %
Total Number of Instances   649
  
```

Status

B – G2 also removed

Neural Network

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

☒ Cross-validation Folds

☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

Attrib nursery=no      -1.1286032374651727
Attrib higher=no       -0.33024761694780846
Attrib internet=yes    -1.1331894783463348
Attrib romantic=yes    -0.5347485354894026
Attrib famrel          -0.18473393797524607
Attrib freetime        -2.5232051457638875
Attrib goout           2.2475397841677265
Attrib dalc             0.24718065664634073
Attrib walc             0.5159487881120541
Attrib health           0.3568483515684939
Attrib absences         0.8271468213314647

Class
Input
Node 0

Time taken to build model: 15.62 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.267
Mean absolute error         1.2661
Root mean squared error     1.6393
Relative absolute error     156.3145 %
Root relative squared error 163.5804 %
Total Number of Instances   649

```

Status

SVM

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

☒ Cross-validation Folds

☐ Percentage split %

(Num) absences

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

Classifier output

```

- 0.0641 * (normalized) higher=no
+ 0.0011 * (normalized) internet=yes
- 0.0163 * (normalized) romantic=yes
+ 0.0003 * (normalized) famrel
- 0.0025 * (normalized) freetime
+ 0.0102 * (normalized) goout
+ 0.0168 * (normalized) dalc
- 0.034 * (normalized) walc
- 0.0315 * (normalized) health
- 0.1433 * (normalized) absences
+ 0.6196

Number of kernel evaluations: 210925 (99.06% cached)

Time taken to build model: 3 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.5258
Mean absolute error         0.6592
Root mean squared error     0.8553
Relative absolute error     81.3891 %
Root relative squared error 85.3448 %
Total Number of Instances   649

```

Status

Decision Tree

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
☒ Cross-validation Folds: 10
☐ Percentage split % 66
More options...

(Num) absences

Start Stop

Result list (right-click for options)

- 01:05:51 - bayes.NaiveBayes
- 01:06:16 - functions.MultilayerPerceptron
- 01:09:29 - functions.SMO
- 01:09:40 - trees.REPTree
- 01:09:51 - trees.RandomForest
- 01:12:16 - functions.MultilayerPerceptron
- 01:16:09 - functions.SMOreg
- 01:21:33 - trees.REPTree
- 01:21:40 - trees.RandomForest
- 01:22:24 - functions.MultilayerPerceptron
- 01:26:07 - functions.SMOreg
- 01:26:48 - trees.REPTree
- 01:27:12 - trees.RandomForest

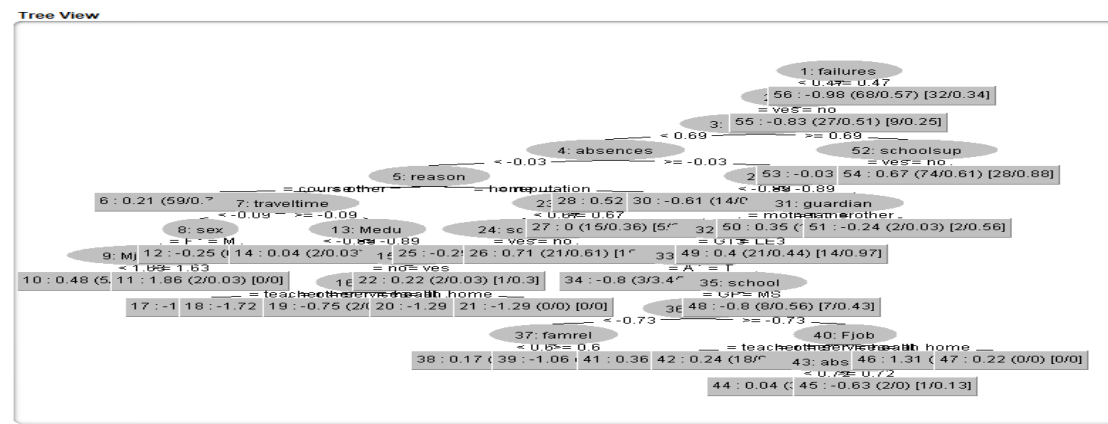
Classifier output

```
absences < 0.72 : 0.04 (3/0.03) [3/0.4]
absences >= 0.72 : -0.63 (2/0) [1/0.13]
Fjob = health : 1.31 (1/0) [1/0]
Fjob = at_home : 0.22 (0/0) [0/0]
school = MS : -0.8 (8/0.56) [7/0.43]
famsize = LE3 : 0.4 (21/0.44) [14/0.97]
guardian = father : 0.35 (19/0.62) [6/1.02]
guardian = other : -0.24 (2/0.03) [2/0.56]
studytime >= 0.69
schoolsup = yes : -0.03 (12/0.34) [4/0.49]
schoolsup = no : 0.67 (74/0.61) [28/0.88]
higher = no : -0.83 (27/0.51) [9/0.25]
failures >= 0.47 : -0.98 (68/0.57) [32/0.34]

Size of the tree : 56
Time taken to build model: 0.03 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.4371
Mean absolute error         0.7136
Root mean squared error     0.9092
Relative absolute error     88.097 %
Root relative squared error 90.725 %
Total Number of Instances   649
```



C – G1 also removed

Neural Network

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
Attrib activities=yes -2.4056883663405664
Attrib nursery=no -0.2442735802603659
Attrib higher=no 1.6803166292359046
Attrib internet=yes -1.334056070386168
Attrib romantic=yes 0.0240404834854716
Attrib famrel -1.1101051139764446
Attrib freetime -0.39683780144606273
Attrib goout 0.8262814248905724
Attrib Dalc 0.6600733381027147
Attrib Walc -0.3830101667151966
Attrib health 0.8080604337262255

Class
Input
Node 0

Time taken to build model: 13.54 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient 0.1142
Mean absolute error 1.327
Root mean squared error 1.8425
Relative absolute error 179.483 %
Root relative squared error 184.0155 %
Total Number of Instances 649
```

The 'Result list' on the left shows a list of models, with '01:27:42 - functions.MultilayerPerceptron' selected.

SVM

The screenshot shows the Orange3 software interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```
+ 0.0093 * (normalized) nursery=no
+ 0.0147 * (normalized) higher=no
- 0.019 * (normalized) internet=yes
+ 0.0094 * (normalized) romantic=yes
- 0.0699 * (normalized) famrel
- 0.0228 * (normalized) freetime
+ 0.0295 * (normalized) goout
+ 0.0328 * (normalized) Dalc
+ 0.0098 * (normalized) Walc
- 0.0163 * (normalized) health
+ 0.1374

Number of kernel evaluations: 210925 (99.191% cached)

Time taken to build model: 3.25 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient 0.2805
Mean absolute error 0.6832
Root mean squared error 0.9886
Relative absolute error 92.4099 %
Root relative squared error 96.7323 %
Total Number of Instances 649
```

The 'Result list' on the left shows a list of models, with '01:30:08 - functions.SMOreg' selected.

Decision Tree

The screenshot shows the Weka GUI with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the following text:

```
goout
Dalc
Walc
health
absences
Test mode: 10-fold cross-validation

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

REPTree
=====
: -0 (432/1.11) [217/0.78]

Size of the tree : 1

Time taken to build model: 0.05 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.0242
Mean absolute error         0.7578
Root mean squared error     1.0256
Relative absolute error     102.4926 %
Root relative squared error 102.4255 %
Total Number of Instances   649
```

The 'Result list' on the left shows a list of classifiers, with '01:30:45 - trees.REPTree' selected.

Random Forest

The screenshot shows the Weka GUI with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'RandomForest'. The 'Test options' section has 'Cross-validation' selected with 'Folds' set to 10. The 'Classifier output' pane displays the following text:

```
freetime
goout
Dalc
Walc
health
absences
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: 0.59 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient      0.2495
Mean absolute error         0.7213
Root mean squared error     0.9701
Relative absolute error     97.5621 %
Root relative squared error 96.8662 %
Total Number of Instances   649
```

The 'Result list' on the left shows a list of classifiers, with '01:31:05 - trees.RandomForest' selected.

Graph representing the attributes and their pre-processed form

