Data Mining Report

CP3403 Data Mining

Tayla Black

Mitchell Fletcher

Table of Contents

[Introduction 3](#_Toc453080271)

[Related Work 3](#_Toc453080272)

[Description of Data 3](#_Toc453080273)

[Pre-processing 4](#_Toc453080274)

[Justifications 4](#_Toc453080275)

[Justification for chosen data 4](#_Toc453080276)

[Objective 4](#_Toc453080277)

[Association Rule Mining Algorithm 5](#_Toc453080278)

[Apriori and Predictive Apriori 5](#_Toc453080279)

[Hotspot 5](#_Toc453080280)

[Results 5](#_Toc453080281)

[Apriori 5](#_Toc453080282)

[Dalc (Workday Alcohol Consumption) 5](#_Toc453080283)

[Walc (Weekend Alcohol Consumption) 5](#_Toc453080284)

[Health (Current Health Status) 5](#_Toc453080285)

[Predictive Apriori 5](#_Toc453080286)

[Dalc (Workday Alcohol Consumption) 6](#_Toc453080287)

[Walc (Weekend Alcohol Consumption) 6](#_Toc453080288)

[Health (Current Health Status) 6](#_Toc453080289)

[Hotspot 6](#_Toc453080290)

[Health (Current Health Status) 6](#_Toc453080291)

[Dalc (Workday Alcohol Consumption) 6](#_Toc453080292)

[Walc (Weekend Alcohol Consumption) 6](#_Toc453080293)

[Absences 6](#_Toc453080294)

[Failures 6](#_Toc453080295)

[Comparison and Discussion 7](#_Toc453080296)

[Issues 7](#_Toc453080297)

[Conclusion 8](#_Toc453080298)

[Future Work 8](#_Toc453080299)

[References 8](#_Toc453080300)

[Appendix 9](#_Toc453080301)

[Figure 1 9](#_Toc453080265)

[2 Apriori 10](#_Toc453080266)

[3 Predictive Apriori 11](#_Toc453080267)

[4 Hotspot 14](#_Toc453080268)

# Introduction

The purpose of this report is to demonstrate the pre-processing method and data mining techniques applied to a data set. The data set consists of information regarding secondary school student performance and their alcohol consumption. From the reviewing of this data set, it was determined that analysis could be performed to determine a correlation between the students’ performance in their studies, their attendance and their alcohol consumption. An analysis regarding the student’s school performance and alcohol consumption compared to their current health was also performed to determine if their current health was a contributing factor to the other results. The many methods and tests that were performed for this report were pre-processing, association rule mining, multiple methods were performed to ensure the accuracy of the results. The methods and tests performed gave a greater understanding on how alcohol and health can affect the performance of secondary school students.

This research can be used to determine if the school needs to be implementing measures to ensure students are receiving help in regards to alcohol consumption and medical treatment.

# Related Work

Association Rule Mining (ARM) will be used for the analysis and report of this data set. ARM can be useful for evaluating data as a prediction and an association method. ARM is the discovery of associations between data items, these associations can develop information to help with data mining. ARM comprises of two parts the support and the confidence, the support being the probably of a data item and confidence was the conditional probability, how many times the data item was true.

An example of ARM is a study done on the India Cricket team (K. Antony and Panchapakesan Padma, 2013). This analysis found factors associated with the match outcomes including the Outcome of Toss, Batting First/second and Play at Home or Away. This analysis was to help the team to formulate a way to win in the One Day International cricket. The out of the study displayed information that showed the team had performed their best in the last ten years, when compared to their entire period since the team started playing.

# Description of Data

The data set “Student Alcohol Consumption” was obtained from the UCI Machine Learning Repository. This data was selected because it didn’t contain any missing values and would not require much data cleaning to ensure that all data could be reliable. The dataset consists of 33 attributes including school, sex, age, study time, failures, workday alcohol consumption, weekend alcohol consumption, health, absences, etc. Some of these attributes were not be required for this data mining report, as they aren’t relevant for reliable results. The data set contained 395 instances, of those instances majority of them were nominal data.

# Pre-processing

Upon download of the dataset it was found the data file was in CSV file format, under the impression that Weka did not support this format an online converter was used to convert the CSV file to a ARFF format. This online converter required manual input of the data types, for if they were nominal or numeric. After the completion of the conversion, the file was loaded into Weka, Weka rejected the file complaining that it contained errors. Instead the CSV file was loaded straight into the Weka, Weka prompted for the file to be converted, so the file was converted and processed. Once this was loaded all numeric data was converted to nominal as Apriori does not support numeric data.

For the tests to reveal more accurate results over half of the attributes were removed, these attributes focussed more on the home and social life of the student and not on the grades or alcohol consumption. In other tests these attributes could be used to study a correlation between the students’ home and social life, their drinking patterns and their performance in school. This information was not relevant for this study and were removed before any further processes were implemented. **Refer appendix Fig. 1**

# Justifications

## Justification for chosen data

The data set was chosen for two reasons, the first being that there was a lot of attributes available. These attributes would be helpful for mining the data, as more data could be used to formulate a more accurate result. The second reason was that the data did not contain any mission or invalid data. This helped to reduce the amount of time required for pre-processing and data cleaning.

## Objective

The objective of the data mining was to find all information that could be relevant to determining if drinking alcohol had an effect on a student’s performance. Also to determine if a student’s performance would be lower if their overall health was lower whilst they were consuming alcohol.

## Association Rule Mining Algorithm

### Apriori and Predictive Apriori

Apriori is the main association rule mining algorithm and is commonly used to find associations in data sets and prepare a comparison for the data. It uses a find and test method, which tests for frequent data sets and then tests the data sets if they are true. With use of Predictive Apriori certain conditions can implemented to determine a more precise result. The results of using Apriori would be determined by finding out which students would succeed best at school based on their alcohol consumption.

### Hotspot

Hotspot algorithm is used to find associated rules to one specific data item. The algorithm discovers the specified data item and assigns tests to run while the probability of the target is still increasing (Mark Hall, 2010). This algorithm is helpful to discover association between one data item and others in the data set. The target can be changed to any attribute with any index of that attribute. The attributes chosen for this test are: G1, G2, G3, Health, Dalc (Workday alcohol consumption), Walc (weekend alcohol consumption), absences and Failures. All of these attributes will be tested using their first index, being the lowest capable result of that attribute. The result of using Hotspot should demonstrate an association between all the chosen attributes.

# Results

## Apriori

Apriori algorithm was run using the default configuration and produced 10 results. Of the 10 results it was clear that students who consumed the lowest amount of alcohol were more likely to produce a lower failure score. These students also represented that were more likely to take part in higher education. **Refer to appendix Fig. 2 Apriori**

### Dalc (Workday Alcohol Consumption)

When assessing the association of Dalc with other attributes you can see a relation between Dalc and higher. It was clear that students who drink the lowest amount on were more likely to want to continue with higher education. Once case also displayed that students with zero failures were also producing a low score on Dalc, which suggests that drinking less means less failures.

### Walc (Weekend Alcohol Consumption)

For this test there were no results that had a confidence above 90% that involved Walc.

### Health (Current Health Status)

There were no results that received a 90% confidence or above that included Health.

## Predictive Apriori

Predictive Apriori was run using the default configuration, which produced 100 results. These 100 results displayed several associations of interest including ones suggesting that drinking less meant failing less and having a higher current health status. **Refer to appendix Fig. 3 Predictive Apriori**

### Dalc (Workday Alcohol Consumption)

Within these results there was a supporting pattern to the Apriori results, suggesting that with a low score of Dalc also meant that higher education was considered. There were also several instances that that displayed a correlation between drinking less and receiving zero fails.

### Walc (Weekend Alcohol Consumption)

Walc and Dalc appeared several times together suggesting that students who drink less during the week also drink less during the weekend. Walc also very closely followed the pattern of Dalc, with a Failure rate of 0, and Walc equal to 1. It was also clear that less drinks in Walc also meant that there were less absences.

### Health (Current Health Status)

The test results revealed that students with a higher health rating were drinking less during the workday and weekends. They were also more likely to be less absent, have less failures and want to continue with higher education.

## Hotspot

A variety of attributes were tested for the Hotspot tests; these attributes were key factors for this data mining report. From the tests performed it was found that students who rated low in Dalc and Walc were less likely to receive fails or have any absences. It was unclear of the effects of high ratings on Dalc and Walc contributed to low health ratings but low health ratings did not contribute to any fails. **Refer to appendix Fig. 4 Hotspot**

### Health (Current Health Status)

It revealed that students with low current health who rated low in Dalc, still managed to not receive any fails. They also wanted to continue to higher education.

### Dalc (Workday Alcohol Consumption)

From the test, it shows that students who rated 1 on Dalc were more likely to rate 1 on Walc and to have 0 in their failures.

### Walc (Weekend Alcohol Consumption)

This test also showed the relation between low rating for Dalc and Walc. It was shown that students who rated low on Walc received no fails, no absences and were aiming for higher education.

### Absences

For the students who rated 0 on absences also had associations with low Walc and zero failures.

### Failures

Students who received zero fails displayed low rating for both Walc and Dalc. This implying that student who drink less overall are less likely to fail.

# Comparison and Discussion

The tests for the Apriori and the Hotspot algorithms produced very similar results. The Apriori and Predictive Apriori produced almost identical results. These results gave an overall representation of all the associations with the data, these associations ranged from helpful to useless results. The Hotspot algorithm data allowed for more accurate results for individual attributes, this helped to focus on the particular attributes that were in question.

The Apriori results only looked at 10 different rules and helped to discovered the most prominent associations within the data set. These associations were helpful to understand what information could be immediately found within this dataset. Although these results were helpful they did not include all the attributes which were important for this study.

The Predictive Apriori algorithm produced 100 rules which managed to include all of the desired attributes for this study. The more rules helped with developing a better understanding of this data and it produced more associations that are considered helpful to this study.

The Hotspot algorithm was capable of looking at the individual attributes that were the focus of this report. Although it was tedious task to test each attribute, it provided a better look at the associations that effected singular attributes. This algorithm may produce a bias view of the results as it can be configured to ensure that associations will be produced by certain attributes.

From each of the algorithm tests helpful and insightful results were produced. All of these results suggested similar associations and produced a trend that could be interpreted. The interpretation of the results was that student who drank less overall were less likely to badly in school (failures and absences). These students also were interested in undertaking higher education.

From these findings, schools could implement programs to ensure that students monitor their alcohol consumption to ensure they do well in school. Schools could also ensure that student who do drink during the workday and weekend are given access to information to help them with their studies and reduce their drinking.

# Issues

The first issue that was found, was that the file was in CSV format, this was an incompatible file format for Weka. The CSV file was converted to ARFF file using an online convert but due to the Field Separator being a semi-colon, the file was incompatible once again. This issue was overcome by using the Weka converter, ensuring that the Field Separator was set to a semi-colon as the default is a comma.

The second issue was that there were too many attributes to provide any helpful data, so to solve this issues all unnecessary attributes were removed. Also the Association algorithms did not support any Numeric data, so all numeric data was converted to nominal in pre-processing.

# Conclusion

From comparing the data produced by all algorithms, two conclusions for student alcohol consumption and students school performance were found.

The first conclusion is, student who drink less during the week and on weekends are less likely to receive any failures or be absent from any classes. They also strive to continue on to higher educations.

The second conclusion is, there is no association with students who do well in school and students who are unhealthy and drink rarely.

# Future Work

For the future work on this project, it would require looking closely at each students’ academic records and closely compare this to how much they consume alcohol. This would be better than just trying to compare the students’ drinking habits to the combined failures.

Also in future work, using data from a wider range of student and using different algorithms to determine the result, may result in more accurate results.

# References

Fabio Pagnotta, Hossain Mohammad Amran (2016) Student Alcohol Consumption Data set, Retrieved June 20, 2016, From  
<https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION>

K. Antony and Panchapakesan Padma (2013)Application of Association Rule Mining: A case study on team India Retrieved June 25, 2016, From  
<http://Ieeexploree.ieee.org>

Mark Hall (2010) About Hotspot Algorithm in Weka, Retrieved June 27, 2016, From  
<https://list.waikato.ac.nz/pipermail/wekalist/2010-April/048196.html>

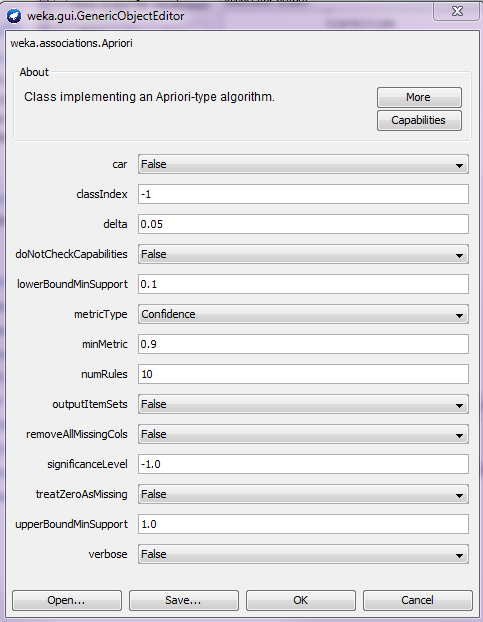
Data Mining and Knowledge Discovery: ARM (PowerPoint) (n.d.). Retrieved from <https://learnjcu.jcu.edu.au>

# Appendix

|  |  |  |
| --- | --- | --- |
|  | Default Attributes | Required Attributes |
| 1 | School | Sex |
| 2 | Sex | Age |
| 3 | Age | Studytime |
| 4 | Address | Failure |
| 5 | famsize | Nursey |
| 6 | Pstatus | Higher |
| 7 | Medu | Internet |
| 8 | Fedu | Dalc |
| 9 | Mjob | Walc |
| 10 | Fjob | Health |
| 11 | Reason | Absences |
| 12 | Guardian | G1 |
| 13 | Traveltime | G2 |
| 14 | Studytime | G3 |
| 15 | Failures |  |
| 16 | Schoolsup |  |
| 17 | Famsup |  |
| 18 | Paid |  |
| 19 | Activities |  |
| 20 | Nursey |  |
| 21 | Higher |  |
| 22 | Internet |  |
| 23 | Romantic |  |
| 24 | Famrel |  |
| 25 | Freetime |  |
| 26 | Goout |  |
| 27 | Dalc |  |
| 28 | Walc |  |
| 29 | Health |  |
| 30 | Absences |  |
| 31 | G1 |  |
| 32 | G2 |  |
| 33 | G3 |  |

Figure 1

**Apriori**



2 Apriori

*Apriori Test*

1. failures=0 Dalc=1 232 ==> higher=yes 228 <conf:(0.98)> lift:(1.04) lev:(0.02) [7] conv:(2.35)

2. sex=F 208 ==> higher=yes 204 <conf:(0.98)> lift:(1.03) lev:(0.02) [6] conv:(2.11)

3. failures=0 312 ==> higher=yes 305 <conf:(0.98)> lift:(1.03) lev:(0.02) [8] conv:(1.97)

4. failures=0 nursery=yes 254 ==> higher=yes 248 <conf:(0.98)> lift:(1.03) lev:(0.02) [6] conv:(1.84)

5. internet=yes Dalc=1 228 ==> higher=yes 222 <conf:(0.97)> lift:(1.03) lev:(0.01) [5] conv:(1.65)

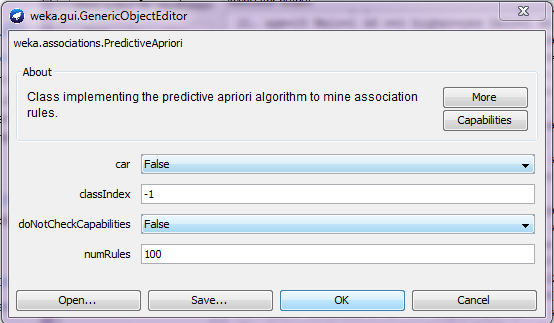
6. failures=0 internet=yes 265 ==> higher=yes 258 <conf:(0.97)> lift:(1.03) lev:(0.02) [6] conv:(1.68)

7. nursery=yes Dalc=1 222 ==> higher=yes 216 <conf:(0.97)> lift:(1.02) lev:(0.01) [5] conv:(1.61)

8. failures=0 nursery=yes internet=yes 215 ==> higher=yes 209 <conf:(0.97)> lift:(1.02) lev:(0.01) [4] conv:(1.56)

9. Dalc=1 276 ==> higher=yes 267 <conf:(0.97)> lift:(1.02) lev:(0.01) [4] conv:(1.4)

10. nursery=yes 314 ==> higher=yes 300 <conf:(0.96)> lift:(1.01) lev:(0) [1] conv:(1.06)



3 Predictive Apriori

*Predictive Apriori Test*

1. higher=yes Walc=1 144 ==> Dalc=1 144 acc:(0.99401)

2. sex=F Walc=1 94 ==> Dalc=1 94 acc:(0.99266)

3. age=15 82 ==> higher=yes 82 acc:(0.99199)

4. Walc=1 151 ==> Dalc=1 150 acc:(0.99188)

5. age=16 failures=0 Dalc=1 67 ==> higher=yes 67 acc:(0.99068)

6. age=16 internet=yes Dalc=1 66 ==> higher=yes 66 acc:(0.99057)

7. studytime=3 65 ==> higher=yes 65 acc:(0.99045)

8. failures=0 nursery=yes Dalc=1 health=5 64 ==> higher=yes 64 acc:(0.99033)

9. nursery=yes internet=yes Dalc=1 health=5 60 ==> higher=yes 60 acc:(0.98978)

10. age=16 nursery=yes Dalc=1 59 ==> higher=yes 59 acc:(0.98963)

11. sex=F age=17 58 ==> higher=yes 58 acc:(0.98947)

12. sex=F age=16 54 ==> higher=yes 54 acc:(0.98877)

13. failures=0 nursery=yes Dalc=1 Walc=1 109 ==> higher=yes 108 acc:(0.98849)

14. sex=M failures=0 Dalc=1 Walc=1 50 ==> higher=yes 50 acc:(0.98794)

15. failures=0 Dalc=1 Walc=1 absences=0 50 ==> higher=yes 50 acc:(0.98794)

16. sex=F nursery=yes 166 ==> higher=yes 164 acc:(0.98752)

17. G3=11 47 ==> higher=yes 47 acc:(0.9872)

18. failures=0 internet=no 47 ==> higher=yes 47 acc:(0.9872)

19. age=15 Walc=1 46 ==> higher=yes Dalc=1 46 acc:(0.98693)

20. age=17 studytime=2 failures=0 46 ==> higher=yes 46 acc:(0.98693)

21. studytime=2 failures=0 nursery=yes health=5 46 ==> higher=yes 46 acc:(0.98693)

22. sex=F Walc=3 45 ==> higher=yes 45 acc:(0.98664)

23. studytime=2 nursery=yes internet=yes health=5 44 ==> higher=yes 44 acc:(0.98634)

24. sex=M failures=0 Dalc=1 93 ==> higher=yes 92 acc:(0.986)

25. failures=0 health=1 42 ==> higher=yes 42 acc:(0.98568)

26. nursery=yes G2=9 42 ==> higher=yes 42 acc:(0.98568)

27. Dalc=1 Walc=3 42 ==> higher=yes 42 acc:(0.98568)

28. Dalc=1 absences=4 41 ==> higher=yes 41 acc:(0.98532)

29. age=16 studytime=2 Dalc=1 41 ==> higher=yes 41 acc:(0.98532)

30. failures=0 nursery=yes health=5 88 ==> higher=yes 87 acc:(0.98498)

31. nursery=yes health=1 40 ==> higher=yes 40 acc:(0.98495)

32. failures=0 Dalc=1 health=4 40 ==> higher=yes 40 acc:(0.98495)

33. G1=11 39 ==> higher=yes 39 acc:(0.98455)

34. age=17 studytime=2 internet=yes 39 ==> higher=yes 39 acc:(0.98455)

35. G3=0 38 ==> absences=0 38 acc:(0.98413)

36. sex=M internet=yes Dalc=1 health=5 38 ==> higher=yes 38 acc:(0.98413)

37. nursery=yes Dalc=1 Walc=1 health=5 38 ==> higher=yes 38 acc:(0.98413)

38. age=17 failures=0 84 ==> higher=yes 83 acc:(0.98405)

39. G2=13 37 ==> higher=yes 37 acc:(0.98368)

40. sex=F health=4 37 ==> higher=yes 37 acc:(0.98368)

41. sex=M studytime=2 nursery=yes internet=yes Dalc=1 36 ==> higher=yes 36 acc:(0.9832)

42. failures=0 nursery=yes Dalc=1 192 ==> higher=yes 189 acc:(0.9829)

43. studytime=2 internet=yes Walc=3 35 ==> higher=yes 35 acc:(0.98269)

44. sex=F Dalc=1 Walc=1 absences=0 35 ==> higher=yes 35 acc:(0.98269)

45. sex=F nursery=yes Walc=1 78 ==> higher=yes Dalc=1 77 acc:(0.98244)

46. G2=15 34 ==> failures=0 higher=yes 34 acc:(0.98215)

47. Dalc=1 health=1 34 ==> higher=yes 34 acc:(0.98215)

48. studytime=2 failures=0 Walc=3 34 ==> higher=yes 34 acc:(0.98215)

49. age=16 Dalc=1 76 ==> higher=yes 75 acc:(0.98183)

50. failures=0 Dalc=1 Walc=1 126 ==> higher=yes 124 acc:(0.98158)

51. G3=15 33 ==> higher=yes 33 acc:(0.98157)

52. sex=F Dalc=2 33 ==> higher=yes 33 acc:(0.98157)

53. studytime=3 Walc=1 33 ==> higher=yes Dalc=1 33 acc:(0.98157)

54. failures=0 G1=8 33 ==> higher=yes 33 acc:(0.98157)

55. internet=yes Dalc=1 health=5 74 ==> higher=yes 73 acc:(0.98119)

56. studytime=2 nursery=yes Walc=3 32 ==> higher=yes 32 acc:(0.98094)

57. failures=0 Dalc=1 232 ==> higher=yes 228 acc:(0.98059)

58. absences=6 31 ==> higher=yes 31 acc:(0.98028)

59. G1=9 31 ==> higher=yes 31 acc:(0.98028)

60. sex=F health=1 31 ==> higher=yes 31 acc:(0.98028)

61. studytime=2 failures=0 internet=yes absences=2 31 ==> higher=yes 31 acc:(0.98028)

62. sex=F internet=yes 170 ==> higher=yes 167 acc:(0.97975)

63. sex=F G3=10 30 ==> higher=yes 30 acc:(0.97956)

64. failures=0 G1=12 30 ==> higher=yes 30 acc:(0.97956)

65. studytime=2 nursery=yes higher=yes absences=2 30 ==> failures=0 30 acc:(0.97956)

66. sex=F failures=0 168 ==> higher=yes 165 acc:(0.97943)

67. sex=M studytime=2 failures=0 68 ==> higher=yes 67 acc:(0.97899)

68. sex=F G2=10 29 ==> higher=yes 29 acc:(0.97879)

69. studytime=2 health=1 29 ==> higher=yes 29 acc:(0.97879)

70. failures=0 Walc=2 67 ==> higher=yes 66 acc:(0.97858)

71. failures=0 health=5 110 ==> higher=yes 108 acc:(0.97781)

72. failures=0 nursery=yes Walc=1 110 ==> higher=yes 108 acc:(0.97781)

73. sex=F 208 ==> higher=yes 204 acc:(0.97772)

74. Dalc=1 Walc=2 65 ==> higher=yes 64 acc:(0.97772)

75. studytime=4 27 ==> higher=yes 27 acc:(0.97707)

76. G3=14 27 ==> higher=yes 27 acc:(0.97707)

77. age=17 Walc=3 27 ==> higher=yes 27 acc:(0.97707)

78. failures=0 Walc=3 63 ==> higher=yes 62 acc:(0.97679)

79. nursery=yes G2=11 26 ==> higher=yes 26 acc:(0.97609)

80. age=15 Dalc=1 health=5 26 ==> failures=0 26 acc:(0.97609)

81. sex=M age=15 Walc=1 25 ==> failures=0 25 acc:(0.97504)

82. failures=0 internet=yes G1=13 25 ==> higher=yes 25 acc:(0.97504)

83. failures=0 nursery=no 58 ==> higher=yes 57 acc:(0.97418)

84. failures=0 absences=2 58 ==> higher=yes 57 acc:(0.97418)

85. G1=15 24 ==> higher=yes 24 acc:(0.97389)

86. nursery=yes higher=yes internet=yes G3=15 24 ==> failures=0 24 acc:(0.97389)

87. failures=0 nursery=yes 254 ==> higher=yes 248 acc:(0.9729)

88. G2=14 23 ==> higher=yes 23 acc:(0.97263)

89. age=16 Walc=2 23 ==> higher=yes 23 acc:(0.97263)

90. age=16 studytime=2 54 ==> higher=yes 53 acc:(0.97171)

91. failures=0 health=4 54 ==> higher=yes 53 acc:(0.97171)

92. absences=8 22 ==> higher=yes 22 acc:(0.97126)

93. G1=16 22 ==> failures=0 higher=yes 22 acc:(0.97126)

94. age=16 health=3 22 ==> higher=yes 22 acc:(0.97126)

95. age=17 nursery=no 22 ==> failures=0 22 acc:(0.97126)

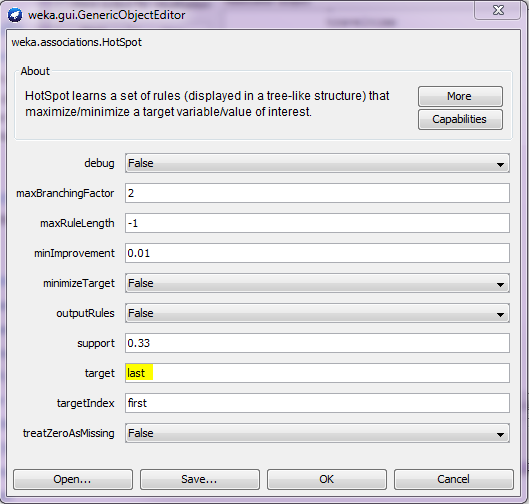
96. Dalc=1 G3=14 22 ==> failures=0 22 acc:(0.97126)

97. sex=M failures=0 health=4 22 ==> internet=yes 22 acc:(0.97126)

98. failures=0 internet=yes Walc=1 health=3 22 ==> higher=yes 22 acc:(0.97126)

99. studytime=2 failures=0 internet=yes Walc=1 absences=0 22 ==> higher=yes 22 acc:(0.97126)

100. sex=F studytime=2 failures=0 higher=yes internet=yes Dalc=1 Walc=1 30 ==> nursery=yes 26 acc:(0.84342)



4 Hotspot

*Health*

health=1 (11.9% [47/395])

sex = F (14.9% [31/208])

| studytime = 2 (17.7% [20/113])

| | failures = 0 (20% [18/90])

| | | internet = yes (22.22% [16/72])

| | | | higher = yes (23.19% [16/69])

| | | Dalc = 1 (21.92% [16/73])

| | | | higher = yes (22.86% [16/70])

| | internet = yes (19.78% [18/91])

| | | higher = yes (20.45% [18/88])

| internet = yes (17.06% [29/170])

| | nursery = yes (18.52% [25/135])

| | | failures = 0 (19.47% [22/113])

| | | | higher = yes (19.82% [22/111])

| | | Dalc = 1 (18.87% [20/106])

| | | | higher = yes (19.23% [20/104])

| | | | | failures = 0 (19.57% [18/92])

| | | | failures = 0 (19.15% [18/94])

| | failures = 0 (18.25% [25/137])

| | | higher = yes (18.66% [25/134])

studytime = 2 (14.65% [29/198])

| failures = 0 (17.09% [27/158])

| | Dalc = 1 (18.64% [22/118])

| | | nursery = yes (20.62% [20/97])

| | | | internet = yes (22.5% [18/80])

| | | | | higher = yes (23.08% [18/78])

| | | | higher = yes (21.05% [20/95])

| | | internet = yes (19.39% [19/98])

| | | | higher = yes (20% [19/95])

| | internet = yes (17.69% [23/130])

| | | nursery = yes (19.81% [21/106])

| | | | higher = yes (20.39% [21/103])

| | | higher = yes (18.25% [23/126])

| Dalc = 1 (16.31% [23/141])

| | sex = F (19.32% [17/88])

| | | higher = yes (20.24% [17/84])

| | nursery = yes (17.7% [20/113])

| | | internet = yes (19.78% [18/91])

| | | | higher = yes (20.22% [18/89])

| | | higher = yes (18.18% [20/110])

*Dalc (Workday Alcohol Consumption)*

Dalc=1 (69.87% [276/395])

Walc = 1 (99.34% [150/151])

sex = F (80.29% [167/208])

| Walc = 1 (100% [94/94])

| failures = 0 (82.74% [139/168])

| | nursery = yes (84.29% [118/140])

*Walc (Weekend Alcohol Consumption)*

Walc=1 (38.23% [151/395])

Dalc = 1 (54.35% [150/276])

| absences = 0 (66.28% [57/86])

| | failures = 0 (71.43% [50/70])

| | | higher = yes (72.46% [50/69])

| | higher = yes (67.47% [56/83])

| nursery = yes (56.31% [125/222])

| | sex = F (58.21% [78/134])

| | studytime = 2 (57.52% [65/113])

| | | higher = yes (58.18% [64/110])

absences = 0 (50.43% [58/115])

| failures = 0 (57.3% [51/89])

| higher = yes (52.34% [56/107])

| | failures = 0 (57.47% [50/87])

*Absences*

absences=0 (29.11% [115/395])

G3 = 0 (100% [38/38])

Walc = 1 (38.41% [58/151])

| internet = yes (40.65% [50/123])

| | failures = 0 (43.27% [45/104])

| failures = 0 (40.16% [51/127])

*Failures*

failures=0 (78.99% [312/395])

Walc = 1 (84.11% [127/151])

| nursery = yes (87.3% [110/126])

| | higher = yes (88.52% [108/122])

| higher = yes (86.11% [124/144])

Dalc = 1 (84.06% [232/276])

| nursery = yes (86.49% [192/222])

| | sex = F (88.06% [118/134])

| | internet = yes (88.04% [162/184])

| internet = yes (85.53% [195/228])