Project 3.1

Apriori:

As is common in association rule mining, given a set of itemsets, the algorithm attempts to find subsets which are common to at least a minimum number C of the itemsets.

Output:

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
                outlook
                temperature
                humidity
                windy
                play
=== Associator model (full training set) ===

Apriori
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4    conf:(1)
 2. temperature=cool 4 ==> humidity=normal 4    conf:(1)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4    conf:(1)
 4. outlook=sunny play=no 3 ==> humidity=high 3    conf:(1)
 5. outlook=sunny humidity=high 3 ==> play=no 3    conf:(1)
```

6. outlook=rainy play=yes 3 ==> windy=FALSE 3 conf:(1)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 conf:(1)
8. temperature=cool play=yes 3 ==> humidity=normal 3 conf:(1)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2 conf:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2 conf:(1)

Analysis: Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

PredictiveApriori:

Class implementing the predictive apriori algorithm to mine association rules. It searches with an increasing support threshold for the best n rules concerning a support-based corrected confidence value.

Output:

```
=== Run information ===

Scheme:       weka.associations.PredictiveApriori -N 100 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===

PredictiveApriori
==================

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4 acc:(0.95323)
 2. temperature=cool 4 ==> humidity=normal 4 acc:(0.95323)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4 acc:(0.95323)
 4. outlook=sunny humidity=high 3 ==> play=no 3 acc:(0.92093)
 5. outlook=sunny play=no 3 ==> humidity=high 3 acc:(0.92093)
 6. outlook=rainy windy=FALSE 3 ==> play=yes 3 acc:(0.92093)
 7. outlook=rainy play=yes 3 ==> windy=FALSE 3 acc:(0.92093)
 8. outlook=sunny temperature=hot 2 ==> humidity=high play=no 2 acc:(0.86233)
 9. outlook=sunny humidity=normal 2 ==> play=yes 2 acc:(0.86233)
10. outlook=sunny play=yes 2 ==> humidity=normal 2 acc:(0.86233)
11. outlook=overcast temperature=hot 2 ==> windy=FALSE play=yes 2 acc:(0.86233)
```

12. outlook=overcast windy=FALSE 2 ==> temperature=hot play=yes 2 acc:(0.86233)
13. outlook=rainy humidity=high 2 ==> temperature=mild 2 acc:(0.86233)
14. outlook=rainy windy=TRUE 2 ==> play=no 2 acc:(0.86233)
15. outlook=rainy play=no 2 ==> windy=TRUE 2 acc:(0.86233)
16. temperature=hot play=yes 2 ==> outlook=overcast windy=FALSE 2 acc:(0.86233)
17. temperature=hot play=no 2 ==> outlook=sunny humidity=high 2 acc:(0.86233)
18. temperature=mild humidity=normal 2 ==> play=yes 2 acc:(0.86233)
19. temperature=mild play=no 2 ==> humidity=high 2 acc:(0.86233)
20. temperature=cool windy=FALSE 2 ==> humidity=normal play=yes 2 acc:(0.86233)
21. windy=FALSE play=no 2 ==> outlook=sunny humidity=high 2 acc:(0.86233)
22. temperature=mild windy=FALSE play=yes 2 ==> outlook=rainy 2 acc:(0.86233)
23. humidity=normal 7 ==> play=yes 6 acc:(0.69497)
24. play=no 5 ==> humidity=high 4 acc:(0.59096)
25. windy=FALSE 8 ==> play=yes 6 acc:(0.56435)
26. temperature=hot 4 ==> humidity=high 3 acc:(0.54473)
27. temperature=hot 4 ==> windy=FALSE 3 acc:(0.54473)
28. temperature=cool 4 ==> humidity=normal play=yes 3 acc:(0.54473)
29. humidity=high play=no 4 ==> outlook=sunny 3 acc:(0.54473)
30. play=yes 9 ==> humidity=normal 6 acc:(0.53808)
31. play=yes 9 ==> windy=FALSE 6 acc:(0.53808)
32. temperature=mild 6 ==> humidity=high 4 acc:(0.51949)
33. temperature=mild 6 ==> play=yes 4 acc:(0.51949)
34. outlook=sunny humidity=high 3 ==> windy=FALSE play=no 2 acc:(0.49529)
35. outlook=sunny humidity=high 3 ==> temperature=hot play=no 2 acc:(0.49529)
36. outlook=sunny windy=FALSE 3 ==> humidity=high play=no 2 acc:(0.49529)
37. outlook=sunny play=no 3 ==> humidity=high windy=FALSE 2 acc:(0.49529)
38. outlook=sunny play=no 3 ==> temperature=hot humidity=high 2 acc:(0.49529)
39. outlook=rainy temperature=mild 3 ==> windy=FALSE play=yes 2 acc:(0.49529)
40. outlook=rainy humidity=normal 3 ==> temperature=cool 2 acc:(0.49529)
41. outlook=rainy humidity=normal 3 ==> windy=FALSE play=yes 2 acc:(0.49529)
42. outlook=rainy windy=FALSE 3 ==> humidity=normal play=yes 2 acc:(0.49529)
43. outlook=rainy windy=FALSE 3 ==> temperature=mild play=yes 2 acc:(0.49529)
44. outlook=rainy play=yes 3 ==> temperature=mild windy=FALSE 2 acc:(0.49529)
45. outlook=rainy play=yes 3 ==> humidity=normal windy=FALSE 2 acc:(0.49529)
46. temperature=hot humidity=high 3 ==> outlook=sunny play=no 2 acc:(0.49529)
47. temperature=hot windy=FALSE 3 ==> outlook=overcast play=yes 2 acc:(0.49529)
48. temperature=mild windy=FALSE 3 ==> outlook=rainy play=yes 2 acc:(0.49529)
49. temperature=cool play=yes 3 ==> humidity=normal windy=FALSE 2 acc:(0.49529)
50. humidity=high windy=TRUE 3 ==> temperature=mild 2 acc:(0.49529)
51. humidity=high windy=TRUE 3 ==> play=no 2 acc:(0.49529)
52. humidity=high play=yes 3 ==> outlook=overcast 2 acc:(0.49529)
53. humidity=high play=yes 3 ==> temperature=mild 2 acc:(0.49529)
54. humidity=normal windy=TRUE 3 ==> temperature=cool 2 acc:(0.49529)
55. windy=TRUE play=yes 3 ==> outlook=overcast 2 acc:(0.49529)

56. windy=TRUE play=yes 3 ==> temperature=mild 2 acc:(0.49529)
57. windy=TRUE play=no 3 ==> outlook=rainy 2 acc:(0.49529)
58. humidity=high 7 ==> temperature=mild 4 acc:(0.49376)
59. humidity=high 7 ==> windy=FALSE 4 acc:(0.49376)
60. humidity=normal 7 ==> temperature=cool 4 acc:(0.49376)
61. humidity=normal 7 ==> windy=FALSE play=yes 4 acc:(0.49376)
62. outlook=sunny 5 ==> windy=FALSE 3 acc:(0.48941)
63. outlook=sunny 5 ==> humidity=high play=no 3 acc:(0.48941)
64. outlook=rainy 5 ==> temperature=mild 3 acc:(0.48941)
65. outlook=rainy 5 ==> humidity=normal 3 acc:(0.48941)
66. outlook=rainy 5 ==> windy=FALSE play=yes 3 acc:(0.48941)
67. play=no 5 ==> windy=TRUE 3 acc:(0.48941)
68. play=no 5 ==> outlook=sunny humidity=high 3 acc:(0.48941)
69. windy=FALSE 8 ==> humidity=high 4 acc:(0.46991)
70. windy=FALSE 8 ==> humidity=normal play=yes 4 acc:(0.46991)
71. temperature=mild 6 ==> outlook=rainy 3 acc:(0.46058)
72. temperature=mild 6 ==> windy=TRUE 3 acc:(0.46058)
73. windy=TRUE 6 ==> temperature=mild 3 acc:(0.46058)
74. windy=TRUE 6 ==> humidity=high 3 acc:(0.46058)
75. windy=FALSE play=yes 6 ==> outlook=rainy 3 acc:(0.46058)
76. outlook=overcast 4 ==> humidity=high play=yes 2 acc:(0.44678)
77. outlook=overcast 4 ==> temperature=hot windy=FALSE 2 acc:(0.44678)
78. temperature=hot 4 ==> outlook=overcast windy=FALSE 2 acc:(0.44678)
79. temperature=hot 4 ==> outlook=sunny humidity=high 2 acc:(0.44678)
80. temperature=cool 4 ==> humidity=normal windy=TRUE 2 acc:(0.44678)
81. temperature=cool 4 ==> outlook=rainy humidity=normal 2 acc:(0.44678)
82. outlook=overcast play=yes 4 ==> windy=TRUE 2 acc:(0.44678)
83. temperature=mild humidity=high 4 ==> play=no 2 acc:(0.44678)
84. temperature=mild play=yes 4 ==> outlook=rainy windy=FALSE 2 acc:(0.44678)
85. humidity=high play=no 4 ==> temperature=hot 2 acc:(0.44678)
86. humidity=normal windy=FALSE 4 ==> outlook=rainy 2 acc:(0.44678)
87. play=yes 9 ==> outlook=overcast 4 acc:(0.44601)
88. play=yes 9 ==> temperature=mild 4 acc:(0.44601)
89. play=yes 9 ==> humidity=normal windy=FALSE 4 acc:(0.44601)
90. humidity=high 7 ==> temperature=hot 3 acc:(0.43225)
91. humidity=high 7 ==> outlook=sunny play=no 3 acc:(0.43225)
92. humidity=normal 7 ==> outlook=rainy 3 acc:(0.43225)
93. humidity=normal 7 ==> windy=TRUE 3 acc:(0.43225)
94. humidity=normal 7 ==> temperature=cool play=yes 3 acc:(0.43225)
95. outlook=sunny 5 ==> temperature=mild 2 acc:(0.41235)
96. outlook=sunny 5 ==> temperature=hot humidity=high 2 acc:(0.41235)
97. outlook=rainy 5 ==> temperature=mild humidity=high 2 acc:(0.41235)
98. outlook=rainy 5 ==> temperature=mild windy=FALSE 2 acc:(0.41235)
99. play=no 5 ==> outlook=sunny temperature=hot 2 acc:(0.41235)

Analysis: This algorithm describe a detailed rule generated. From the accuracy, we can judge which rules are more interesting and useful.

Tertius:

It finds rules according to confirmation measure.

Output:

=== Run information ===

Scheme: weka.associations.Tertius -K 10 -F 0.0 -N 1.0 -L 4 -G 0 -c 0 -I 0 -P 0
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===

Tertius
=======

1. /* 0.633754 0.071429 */ play = yes ==> humidity = normal or outlook = overcast
2. /* 0.607625 0.000000 */ humidity = normal ==> temperature = cool or play = yes
3. /* 0.607625 0.000000 */ temperature = cool ==> humidity = normal
4. /* 0.594071 0.214286 */ humidity = normal ==> temperature = cool
5. /* 0.590214 0.000000 */ humidity = high and outlook = sunny ==> play = no
6. /* 0.555556 0.000000 */ play = no ==> windy = TRUE or outlook = sunny
7. /* 0.486606 0.000000 */ play = no and outlook = sunny ==> humidity = high
8. /* 0.486606 0.000000 */ humidity = normal ==> play = yes or outlook = rainy
9. /* 0.469374 0.000000 */ outlook = overcast ==> play = yes
10. /* 0.469374 0.000000 */ windy = FALSE and outlook = overcast ==> temperature = hot
11. /* 0.469374 0.000000 */ outlook = overcast ==> temperature = hot or windy = TRUE
12. /* 0.469374 0.000000 */ temperature = hot and play = yes ==> outlook = overcast
13. /* 0.469374 0.000000 */ play = no ==> humidity = high or windy = TRUE
14. /* 0.469374 0.000000 */ temperature = hot ==> play = no or outlook = overcast
15. /* 0.469374 0.000000 */ temperature = hot ==> humidity = high or outlook = overcast
16. /* 0.469374 0.000000 */ humidity = high and play = no ==> temperature = mild or outlook = sunny
17. /* 0.469374 0.000000 */ temperature = mild and play = yes ==> windy = TRUE or outlook = rainy
18. /* 0.469374 0.000000 */ outlook = sunny ==> temperature = cool or windy = TRUE or play = no
19. /* 0.467119 0.357143 */ play = yes ==> outlook = overcast

20. /* 0.458333 0.071429 */ play = yes ==> windy = FALSE or outlook = overcast
21. /* 0.458333 0.071429 */ humidity = high and play = no ==> outlook = sunny
22. /* 0.439100 0.071429 */ play = no ==> humidity = high
23. /* 0.439100 0.071429 */ humidity = high ==> temperature = mild or play = no
24. /* 0.439100 0.071429 */ humidity = high ==> temperature = mild or outlook = sunny

Number of hypotheses considered: 1724
Number of hypotheses explored: 689

HotSpot:

It is just a greedy search subject to support and improvement constraints.

I used non-nominal version of dataset in this part.

Output:

```
=== Run information ===

Scheme:       weka.associations.HotSpot -c last -V first -S 0.33 -M 2 -I 0.01
Relation:     weather
Instances:    14
Attributes:   5
                outlook
                temperature
                humidity
                windy
                play
=== Associator model (full training set) ===

Hot Spot
========
Total population: 14 instances
Target attribute: play
Target value: yes [value count in total population: 9 instances (64.29%)]
Minimum value count for segments: 5 instances (33% of total population)
Maximum branching factor: 2
Minimum improvement in target: 1%

play=yes (64.29% [9/14])
   humidity <= 80 (85.71% [6/7])
   |    temperature > 65 (100% [5/5])
   windy = FALSE (75% [6/8])
   |    temperature <= 83 (85.71% [6/7])
   |    |    humidity <= 86 (100% [5/5])
   |    humidity <= 86 (83.33% [5/6])
```

Analysis: This algorithm has very good performance for targeting temperature.