

Day-3

Question-1

Weka Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **FPGrowth** -P 2 -I -1 -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.3

Start Stop

Result list (right-click for ...)

- 18:44:02 - Apriori
- 18:48:49 - FPGrowth

Associator output

```
=== Run information ===

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.3
Relation:    tester
Instances:   10
Attributes:  5
              a
              b
              c
              d
              e

=== Associator model (full training set) ===

FPGrowth found 12 rules (displaying top 10)

1. [d=0]: 4 ==> [b=1]: 4   <conf:(1)> lift:(1.67) lev:(0.16) conv:(1.6)
2. [b=1, c=1]: 3 ==> [d=0]: 3   <conf:(1)> lift:(2.5) lev:(0.18) conv:(1.8)
3. [c=1, d=0]: 3 ==> [b=1]: 3   <conf:(1)> lift:(1.67) lev:(0.12) conv:(1.2)
4. [d=0]: 4 ==> [c=1]: 3   <conf:(0.75)> lift:(1.5) lev:(0.1) conv:(1)
5. [d=0]: 4 ==> [b=1, c=1]: 3   <conf:(0.75)> lift:(2.5) lev:(0.18) conv:(1.4)
6. [b=1, d=0]: 4 ==> [c=1]: 3   <conf:(0.75)> lift:(1.5) lev:(0.1) conv:(1)
7. [b=1]: 6 ==> [d=0]: 4   <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(1.2)
8. [c=1]: 5 ==> [b=1]: 3   <conf:(0.6)> lift:(1) lev:(0) conv:(0.67)
9. [c=1]: 5 ==> [d=0]: 3   <conf:(0.6)> lift:(1.5) lev:(0.1) conv:(1)
10. [c=1]: 5 ==> [b=1, d=0]: 3   <conf:(0.6)> lift:(1.5) lev:(0.1) conv:(1)
```

Question-2

Apriori

=====

Minimum support: 0.55 (5 instances)

Minimum metric <confidence>: 0.5

Number of cycles performed: 9

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 5

Size of set of large itemsets L(3): 1

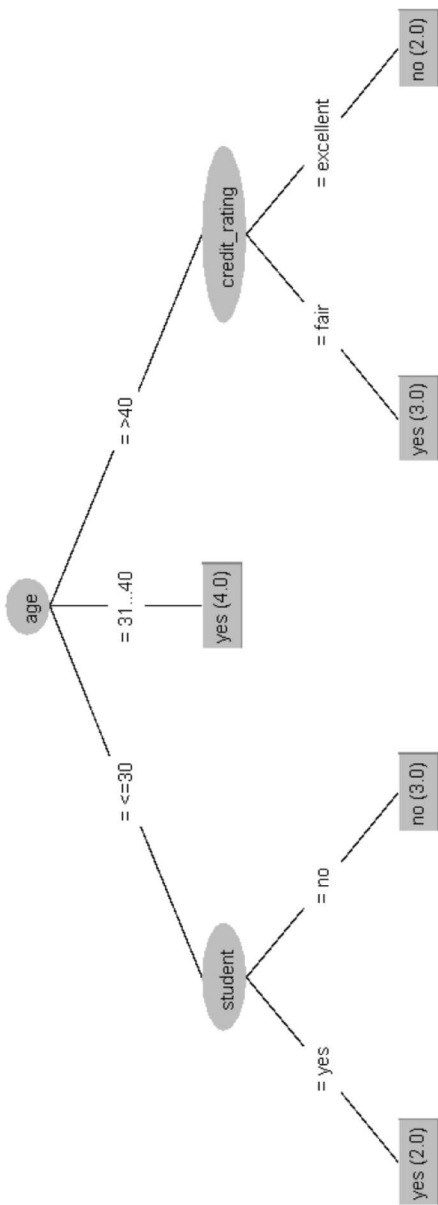
Best rules found:

1. Bread=true 5 ==> Beer=false 5 <conf:(1)> lift:(1.67) lev:(0.2) [2] conv:(2)
2. Butter=true 5 ==> Beer=false 5 <conf:(1)> lift:(1.67) lev:(0.2) [2] conv:(2)
3. Butter=true 5 ==> Bread=true 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
4. Bread=true 5 ==> Butter=true 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
5. Butter=false 5 ==> Bread=false 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
6. Bread=false 5 ==> Butter=false 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
7. Bread=true Butter=true 5 ==> Beer=false 5 <conf:(1)> lift:(1.67) lev:(0.2) [2] conv:(2)
8. Beer=false Butter=true 5 ==> Bread=true 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
9. Beer=false Bread=true 5 ==> Butter=true 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)
10. Butter=true 5 ==> Beer=false Bread=true 5 <conf:(1)> lift:(2) lev:(0.25) [2] conv:(2.5)

FPGrowth found 14 rules (displaying top 10)

1. [Butter=false]: 5 ==> [Bread=false]: 5 <conf:(1)> lift:(2) lev:(0.25) conv:(2.5)
2. [Bread=false]: 5 ==> [Butter=false]: 5 <conf:(1)> lift:(2) lev:(0.25) conv:(2.5)
3. [Milk=false, Butter=false]: 3 ==> [Bread=false]: 3 <conf:(1)> lift:(2) lev:(0.15) conv:(1.5)
4. [Milk=false, Bread=false]: 3 ==> [Butter=false]: 3 <conf:(1)> lift:(2) lev:(0.15) conv:(1.5)
5. [Cookies=false]: 6 ==> [Beer=false]: 4 <conf:(0.67)> lift:(1.11) lev:(0.04) conv:(0.8)
6. [Beer=false]: 6 ==> [Cookies=false]: 4 <conf:(0.67)> lift:(1.11) lev:(0.04) conv:(0.8)
7. [Milk=false]: 5 ==> [Butter=false]: 3 <conf:(0.6)> lift:(1.2) lev:(0.05) conv:(0.83)
8. [Butter=false]: 5 ==> [Milk=false]: 3 <conf:(0.6)> lift:(1.2) lev:(0.05) conv:(0.83)
9. [Milk=false]: 5 ==> [Bread=false]: 3 <conf:(0.6)> lift:(1.2) lev:(0.05) conv:(0.83)
10. [Bread=false]: 5 ==> [Milk=false]: 3 <conf:(0.6)> lift:(1.2) lev:(0.05) conv:(0.83)

Question-3



Naive Bayes Classifier

Attribute	Class	
	yes (0.63)	no (0.38)
=====		
age		
<=30	3.0	4.0
31...40	5.0	1.0
>40	4.0	3.0
[total]	12.0	8.0
income		
high	3.0	3.0
medium	5.0	3.0
low	4.0	2.0
[total]	12.0	8.0
student		
yes	7.0	2.0
no	4.0	5.0
[total]	11.0	7.0
credit_rating		
fair	7.0	3.0
excellent	4.0	4.0
[total]	11.0	7.0

Time taken to build model: 0 seconds

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

Correctly Classified Instances	8	57.1429 %
Incorrectly Classified Instances	6	42.8571 %
Kappa statistic	-0.0244	
Mean absolute error	0.4374	
Root mean squared error	0.4916	
Relative absolute error	91.8631 %	
Root relative squared error	99.6492 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.778	0.800	0.636	0.778	0.700	-0.026	0.578	0.697	yes
	0.200	0.222	0.333	0.200	0.250	-0.026	0.578	0.557	no
Weighted Avg.	0.571	0.594	0.528	0.571	0.539	-0.026	0.578	0.647	

=== Confusion Matrix ===

```
a b  <-- classified as
7 2 | a = yes
4 1 | b = no
```

Question-4

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

Variable	Class
preg	-0.1232
plas	-0.0352
pres	0.0133
skin	-0.0006
insu	0.0012
mass	-0.0897
pedi	-0.9452
age	-0.0149
Intercept	8.4047

Odds Ratios...

Variable	Class
preg	0.8841
plas	0.9654
pres	1.0134
skin	0.9994
insu	1.0012
mass	0.9142
pedi	0.3886
age	0.9852

Time taken to build model: 0.03 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances	601	78.2552 %
Incorrectly Classified Instances	167	21.7448 %
Kappa statistic	0.4966	
Mean absolute error	0.3063	
Root mean squared error	0.3908	
Relative absolute error	67.3928 %	
Root relative squared error	81.9907 %	
Total Number of Instances	768	

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: 0.16 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	582	75.7813 %
Incorrectly Classified Instances	186	24.2188 %
Kappa statistic	0.4566	
Mean absolute error	0.3106	
Root mean squared error	0.4031	
Relative absolute error	60.3405 %	
Root relative squared error	84.5604 %	
Total Number of Instances	768	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.836	0.388	0.801	0.836	0.818	0.458	0.820	0.886	tested_negative
	0.612	0.164	0.667	0.612	0.638	0.458	0.820	0.679	tested_positive
Weighted Avg.	0.758	0.310	0.754	0.758	0.755	0.458	0.820	0.814	

=== Confusion Matrix ===

a	b	<-- classified as
418	82	a = tested_negative
104	164	b = tested_positive

KStar Beta Verion (0.1b).

Copyright (c) 1995-97 by Len Trigg (trigg@cs.waikato.ac.nz).

Java port to Weka by Abdelaziz Mahoui (am14@cs.waikato.ac.nz).

KStar options : -B 20 -M a

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	531	69.1406 %
Incorrectly Classified Instances	237	30.8594 %
Kappa statistic	0.2895	
Mean absolute error	0.3275	
Root mean squared error	0.4969	
Relative absolute error	72.055 %	
Root relative squared error	104.2509 %	
Total Number of Instances	768	

Question-5

FPGrowth found 9 rules (displaying top 9)

1. [E=true]: 4 ==> [K=true]: 4 <conf:(1)> lift:(1) lev:(0) conv:(0)
2. [Y=true]: 3 ==> [K=true]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
3. [O=true]: 3 ==> [K=true]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
4. [M=true]: 3 ==> [K=true]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
5. [O=true]: 3 ==> [E=true]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
6. [O=true]: 3 ==> [K=true, E=true]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
7. [K=true, O=true]: 3 ==> [E=true]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
8. [E=true, O=true]: 3 ==> [K=true]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
9. [K=true]: 5 ==> [E=true]: 4 <conf:(0.8)> lift:(1) lev:(0) conv:(0.5)

Question-6

```
J48 pruned tree
-----

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances      646          84.1146 %
Incorrectly Classified Instances    122          15.8854 %
Kappa statistic                    0.6319
Mean absolute error                0.2383
Root mean squared error            0.3452
Relative absolute error            52.4339 %
Root relative squared error        72.4207 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0.936   0.336   0.839    0.936   0.885     0.642   0.888    0.915   tested_negative
      0.664   0.064   0.848    0.664   0.745     0.642   0.888    0.808   tested_positive
Weighted Avg.   0.841   0.241   0.842    0.841   0.836     0.642   0.888    0.878

=== Confusion Matrix ===

  a    b  <-- classified as
468  32 |  a = tested_negative
 90 178 |  b = tested_positive
```

```
Logistic Regression with ridge parameter of 1.0E-8

=== Summary ===

Correctly Classified Instances      601          78.2552 %
Incorrectly Classified Instances    167          21.7448 %
Kappa statistic                    0.4966
Mean absolute error                0.3063
Root mean squared error            0.3908
Relative absolute error            67.3928 %
Root relative squared error        81.9907 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0.890   0.418   0.799    0.890   0.842     0.504   0.839    0.897   tested_negative
      0.582   0.110   0.739    0.582   0.651     0.504   0.839    0.730   tested_positive
Weighted Avg.   0.783   0.310   0.778    0.783   0.775     0.504   0.839    0.839

=== Confusion Matrix ===

  a    b  <-- classified as
445  55 |  a = tested_negative
112 156 |  b = tested_positive
```


Question-7

Apriori

=====

Minimum support: 0.33 (2 instances)

Minimum metric <confidence>: 0.6

Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 4

Size of set of large itemsets L(3): 1

Best rules found:

1. Coke=true 3 ==> Chips=true 3 <conf:(1)> lift:(1.5) lev:(0.17) [1] conv:(1)
2. Buns=true 2 ==> Hot_Dogs=true 2 <conf:(1)> lift:(1.5) lev:(0.11) [0] conv:(0.67)
3. Hot_Dogs=true Chips=true 2 ==> Coke=true 2 <conf:(1)> lift:(2) lev:(0.17) [1] conv:(1)
4. Hot_Dogs=true Coke=true 2 ==> Chips=true 2 <conf:(1)> lift:(1.5) lev:(0.11) [0] conv:(0.67)
5. Chips=true 4 ==> Coke=true 3 <conf:(0.75)> lift:(1.5) lev:(0.17) [1] conv:(1)
6. Coke=true 3 ==> Hot_Dogs=true 2 <conf:(0.67)> lift:(1) lev:(0) [0] conv:(0.5)
7. Coke=true Chips=true 3 ==> Hot_Dogs=true 2 <conf:(0.67)> lift:(1) lev:(0) [0] conv:(0.5)
8. Coke=true 3 ==> Hot_Dogs=true Chips=true 2 <conf:(0.67)> lift:(2) lev:(0.17) [1] conv:(1)

FPGrowth found 8 rules (displaying top 8)

1. [Buns=true]: 2 ==> [Hot_Dogs=true]: 2 <conf:(1)> lift:(1.5) lev:(0.11) conv:(0.67)
2. [Coke=true]: 3 ==> [Chips=true]: 3 <conf:(1)> lift:(1.5) lev:(0.17) conv:(1)
3. [Hot_Dogs=true, Chips=true]: 2 ==> [Coke=true]: 2 <conf:(1)> lift:(2) lev:(0.17) conv:(1)
4. [Hot_Dogs=true, Coke=true]: 2 ==> [Chips=true]: 2 <conf:(1)> lift:(1.5) lev:(0.11) conv:(0.67)
5. [Chips=true]: 4 ==> [Coke=true]: 3 <conf:(0.75)> lift:(1.5) lev:(0.17) conv:(1)
6. [Coke=true]: 3 ==> [Hot_Dogs=true]: 2 <conf:(0.67)> lift:(1) lev:(0) conv:(0.5)
7. [Coke=true]: 3 ==> [Hot_Dogs=true, Chips=true]: 2 <conf:(0.67)> lift:(2) lev:(0.17) conv:(1)
8. [Chips=true, Coke=true]: 3 ==> [Hot_Dogs=true]: 2 <conf:(0.67)> lift:(1) lev:(0) conv:(0.5)

Question-8

```
348 pruned tree
-----
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)
|   petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves :    5

Size of the tree :    9

Time taken to build model: 0 seconds

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

Correctly Classified Instances      144      96      %
Incorrectly Classified Instances     6       4      %
Kappa statistic                     0.94
Mean absolute error                  0.035
Root mean squared error              0.1586
Relative absolute error              7.8705 %
Root relative squared error          33.6353 %
Total Number of Instances           150

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.980  0.000  1.000    0.980  0.990  0.985  0.990  0.987  Iris-setosa
0.940  0.030  0.940    0.940  0.940  0.910  0.952  0.880  Iris-versicolor
0.960  0.030  0.941    0.960  0.950  0.925  0.961  0.905  Iris-virginica
Weighted Avg.   0.960  0.020  0.960    0.960  0.960  0.940  0.968  0.924

=== Confusion Matrix ===
 a b c  <-- classified as
49 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
 0 2 48 | c = Iris-virginica
```

```
PART decision list
-----
petalwidth <= 0.6: Iris-setosa (50.0)

petalwidth <= 1.7 AND
petallength <= 4.9: Iris-versicolor (48.0/1.0)

: Iris-virginica (52.0/3.0)

Number of Rules :      3

Time taken to build model: 0 seconds

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

Correctly Classified Instances      141      94      %
Incorrectly Classified Instances     9       6      %
Kappa statistic                     0.91
Mean absolute error                  0.0482
Root mean squared error              0.1794
Relative absolute error              10.8379 %
Root relative squared error          38.0567 %
Total Number of Instances           150

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.980  0.000  1.000    0.980  0.990  0.985  0.990  0.987  Iris-setosa
0.940  0.060  0.887    0.940  0.913  0.868  0.954  0.878  Iris-versicolor
0.900  0.030  0.938    0.900  0.918  0.879  0.959  0.914  Iris-virginica
Weighted Avg.   0.940  0.030  0.941    0.940  0.940  0.911  0.968  0.926

=== Confusion Matrix ===
 a b c  <-- classified as
49 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
 0 5 45 | c = Iris-virginica
```