

COMP7103 Assignment 2

Suggested solutions

Question 1 Classification

a), b) Using only weather of one previous day, it is possible to achieve 47% accuracy by using CVPParameterSelection with J48 classifier, and a 57% accuracy by using Naïve Bayesian Classifier. Further discussion is expected.

For your information, the sunshine and radiation data, as well as the best option on 2018-01-01 is shown below.

Day	Sunshine	Radiation	Energy production	Best
2017-02-10	0.4	9.42	$E_{high} = 9.42 \times 60 \times 20\% - 24 \times 4 = 17.04$ $E_{low} = 9.42 \times 60 \times 18\% - 24 \times 3 - 0.4 \times 1 = 29.34$ $E_{off} = 0$	Low
2017-05-30	7.4	21.51	$E_{high} = 21.51 \times 60 \times 20\% - 24 \times 4 = 186.12$ $E_{low} = 21.51 \times 60 \times 18\% - 24 \times 3 - 7.4 \times 1 = 152.91$ $E_{off} = 0$	High
2017-08-28	0	2.97	$E_{high} = 2.97 \times 60 \times 20\% - 24 \times 4 = -36.36$ $E_{low} = 2.97 \times 60 \times 18\% - 24 \times 3 - 0 \times 1 = -39.92$ $E_{off} = 0$	Off
2017-11-18	3.2	7.21	$E_{high} = 7.21 \times 60 \times 20\% - 24 \times 4 = -9.48$ $E_{low} = 7.21 \times 60 \times 18\% - 24 \times 3 - 3.2 \times 1 = 2.67$ $E_{off} = 0$	Low
2018-01-01	3.0	9.99	$E_{high} = 9.99 \times 60 \times 20\% - 24 \times 4 = 23.88$ $E_{low} = 9.99 \times 60 \times 18\% - 24 \times 3 - 3.0 \times 1 = 32.90$ $E_{off} = 0$	Low

Question 2 Association analysis

a) $\{\{B, D\}, \{C\}, \{A, B\}, \{A\}\}, \{\{B, D\}, \{C\}, \{A\}, \{A\}\}, \{\{B, D\}, \{C\}, \{B\}, \{A\}\},$
 $\{\{B\}, \{C\}, \{A, B\}, \{A\}\}, \{\{B\}, \{C\}, \{A\}, \{A\}\}, \{\{B\}, \{C\}, \{B\}, \{A\}\},$
 $\{\{D\}, \{C\}, \{A, B\}, \{A\}\}, \{\{D\}, \{C\}, \{A\}, \{A\}\}, \{\{D\}, \{C\}, \{B\}, \{A\}\}$

Alternatively if item is considered as element:

$\{\{B, D\}, \{C\}, \{A\}\}, \{\{B, D\}, \{C\}, \{B\}\}, \{\{B, D\}, \{A, B\}\}, \{\{B, D\}, \{A\}, \{A\}\}, \{\{B, D\}, \{B\}, \{A\}\},$
 $\{\{B\}, \{C\}, \{A, B\}\}, \{\{B\}, \{C\}, \{A\}, \{A\}\}, \{\{B\}, \{C\}, \{B\}, \{A\}\}, \{\{B\}, \{A, B\}, \{A\}\},$
 $\{\{D\}, \{C\}, \{A, B\}\}, \{\{D\}, \{C\}, \{A\}, \{A\}\}, \{\{D\}, \{C\}, \{B\}, \{A\}\}, \{\{D\}, \{A, B\}, \{A\}\}, \{\{C\}, \{A, B\}, \{A\}\}$

b) Total number of objects is 5, therefore a support count of 3 is needed for the requirement of $\text{Support} \geq 60\% . (5 \times 60\% = 3)$

Support count of 1-item sequence (all of them are frequent):

Sequence	$\{\{A\}\}$	$\{\{B\}\}$	$\{\{C\}\}$	$\{\{D\}\}$
Count	5	5	5	3

Candidate Generation (2nd pass), candidate 2-sequence generated:

$\{\{A, B\}\}, \{\{A, C\}\}, \{\{A, D\}\}, \{\{B, C\}\}, \{\{B, D\}\}, \{\{C, D\}\},$
 $\{\{A\}, \{A\}\}, \{\{A\}, \{B\}\}, \{\{A\}, \{C\}\}, \{\{A\}, \{D\}\}, \{\{B\}, \{A\}\}, \{\{B\}, \{B\}\}, \{\{B\}, \{C\}\}, \{\{B\}, \{D\}\},$
 $\{\{C\}, \{A\}\}, \{\{C\}, \{B\}\}, \{\{C\}, \{C\}\}, \{\{C\}, \{D\}\}, \{\{D\}, \{A\}\}, \{\{D\}, \{B\}\}, \{\{D\}, \{C\}\}, \{\{D\}, \{D\}\}$

Candidate Pruning (2nd pass), remove 2-sequence with infrequent 1-sequence (none removed):

$\{\{A, B\}\}, \{\{A, C\}\}, \{\{A, D\}\}, \{\{B, C\}\}, \{\{B, D\}\}, \{\{C, D\}\},$
 $\{\{A\}, \{A\}\}, \{\{A\}, \{B\}\}, \{\{A\}, \{C\}\}, \{\{A\}, \{D\}\}, \{\{B\}, \{A\}\}, \{\{B\}, \{B\}\}, \{\{B\}, \{C\}\}, \{\{B\}, \{D\}\},$
 $\{\{C\}, \{A\}\}, \{\{C\}, \{B\}\}, \{\{C\}, \{C\}\}, \{\{C\}, \{D\}\}, \{\{D\}, \{A\}\}, \{\{D\}, \{B\}\}, \{\{D\}, \{C\}\}, \{\{D\}, \{D\}\}$

Support counting (2nd pass):

$\{\{A, B\}\}$	$\{\{A, C\}\}$	$\{\{A, D\}\}$	$\{\{B, C\}\}$	$\{\{B, D\}\}$	$\{\{C, D\}\}$	$\{\{A\}, \{A\}\}$	$\{\{A\}, \{B\}\}$
5	3	1	1	1	2	3	1
$\{\{A\}, \{C\}\}$	$\{\{A\}, \{D\}\}$	$\{\{B\}, \{A\}\}$	$\{\{B\}, \{B\}\}$	$\{\{B\}, \{C\}\}$	$\{\{B\}, \{D\}\}$	$\{\{C\}, \{A\}\}$	$\{\{C\}, \{B\}\}$
3	1	4	3	5	2	5	5
$\{\{C\}, \{C\}\}$	$\{\{C\}, \{D\}\}$	$\{\{D\}, \{A\}\}$	$\{\{D\}, \{B\}\}$	$\{\{D\}, \{C\}\}$	$\{\{D\}, \{D\}\}$		
3	1	3	2	2	1		

Frequent 2-sequences:

$\{\{A, B\}\}, \{\{A, C\}\}, \{\{A\}, \{A\}\}, \{\{A\}, \{C\}\}, \{\{B\}, \{A\}\}, \{\{B\}, \{B\}\}, \{\{B\}, \{C\}\}, \{\{C\}, \{A\}\},$
 $\{\{C\}, \{B\}\}, \{\{C\}, \{C\}\}, \{\{D\}, \{A\}\}$

Candidate Generation (3rd pass), candidate 3-sequence generated:

$\{\{A, B\}\} \times \{\{B\} \dots\} \Rightarrow \{\{A, B\}, \{A\}\}, \{\{A, B\}, \{B\}\}, \{\{A, B\}, \{C\}\}$
 $\{\{A, C\}\} \times \{\{C\} \dots\} \Rightarrow \{\{A, C\}, \{A\}\}, \{\{A, C\}, \{B\}\}, \{\{A, C\}, \{C\}\}$
 $\{\{A\}, \{A\}\} \times \{\{A\} \dots\} \Rightarrow \{\{A\}, \{A, B\}\}, \{\{A\}, \{A, C\}\}, \{\{A\}, \{A\}, \{A\}\}, \{\{A\}, \{A\}, \{C\}\}$
 $\{\{A\}, \{C\}\} \times \{\{C\} \dots\} \Rightarrow \{\{A\}, \{C\}, \{A\}\}, \{\{A\}, \{C\}, \{B\}\}, \{\{A\}, \{C\}, \{C\}\}$
 $\{\{B\}, \{A\}\} \times \{\{A\} \dots\} \Rightarrow \{\{B\}, \{A, B\}\}, \{\{B\}, \{A, C\}\}, \{\{B\}, \{A\}, \{A\}\}, \{\{B\}, \{A\}, \{C\}\}$
 $\{\{B\}, \{B\}\} \times \{\{B\} \dots\} \Rightarrow \{\{B\}, \{B\}, \{A\}\}, \{\{B\}, \{B\}, \{B\}\}, \{\{B\}, \{B\}, \{C\}\}$
 $\{\{B\}, \{C\}\} \times \{\{C\} \dots\} \Rightarrow \{\{B\}, \{C\}, \{A\}\}, \{\{B\}, \{C\}, \{B\}\}, \{\{B\}, \{C\}, \{C\}\}$
 $\{\{C\}, \{A\}\} \times \{\{A\} \dots\} \Rightarrow \{\{C\}, \{A, B\}\}, \{\{C\}, \{A, C\}\}, \{\{C\}, \{A\}, \{A\}\}, \{\{C\}, \{A\}, \{C\}\}$
 $\{\{C\}, \{B\}\} \times \{\{B\} \dots\} \Rightarrow \{\{C\}, \{B\}, \{A\}\}, \{\{C\}, \{B\}, \{B\}\}, \{\{C\}, \{B\}, \{C\}\}$
 $\{\{C\}, \{C\}\} \times \{\{C\} \dots\} \Rightarrow \{\{C\}, \{C\}, \{A\}\}, \{\{C\}, \{C\}, \{B\}\}, \{\{C\}, \{C\}, \{C\}\}$
 $\{\{D\}, \{A\}\} \times \{\{A\} \dots\} \Rightarrow \{\{D\}, \{A, B\}\}, \{\{D\}, \{A, C\}\}, \{\{D\}, \{A\}, \{A\}\}, \{\{D\}, \{A\}, \{C\}\}$

Candidate Pruning (3rd pass), remove 3-sequence with infrequent 2-sequence:

$\{\{A, B\}, \{A\}\}, \{\{A, B\}, \{B\}\}, \{\{A, B\}, \{C\}\}, \{\{A, C\}, \{A\}\}, \{\{A, C\}, \{B\}\}, \{\{A, C\}, \{C\}\}, \{\{A\}, \{A, B\}\},$
 $\{\{A\}, \{A, C\}\}, \{\{A\}, \{A\}, \{A\}\}, \{\{A\}, \{A\}, \{C\}\}, \{\{A\}, \{C\}, \{A\}\}, \{\{A\}, \{C\}, \{B\}\}, \{\{A\}, \{C\}, \{C\}\}$
 $\{\{B\}, \{A, B\}\}, \{\{B\}, \{A, C\}\}, \{\{B\}, \{A\}, \{A\}\}, \{\{B\}, \{A\}, \{C\}\}, \{\{B\}, \{B\}, \{A\}\}, \{\{B\}, \{B\}, \{B\}\},$
 $\{\{B\}, \{B\}, \{C\}\}, \{\{B\}, \{C\}, \{A\}\}, \{\{B\}, \{C\}, \{B\}\}, \{\{B\}, \{C\}, \{C\}\}, \{\{C\}, \{A, B\}\}, \{\{C\}, \{A, C\}\},$
 $\{\{C\}, \{A\}, \{A\}\}, \{\{C\}, \{A\}, \{C\}\}, \{\{C\}, \{B\}, \{A\}\}, \{\{C\}, \{B\}, \{B\}\}, \{\{C\}, \{B\}, \{C\}\}, \{\{C\}, \{C\}, \{A\}\},$
 $\{\{C\}, \{C\}, \{B\}\}, \{\{C\}, \{C\}, \{C\}\}, \{\{D\}, \{A, B\}\}, \{\{D\}, \{A, C\}\}, \{\{D\}, \{A\}, \{A\}\}, \{\{D\}, \{A\}, \{C\}\}$

Support counting (3rd pass):

$\{\{A, B\}, \{A\}\}$	$\{\{A, B\}, \{C\}\}$	$\{\{A, C\}, \{A\}\}$	$\{\{A, C\}, \{C\}\}$	$\{\{A\}, \{A, C\}\}$	$\{\{A\}, \{A\}, \{A\}\}$
2	3	1	2	1	0
$\{\{A\}, \{A\}, \{C\}\}$	$\{\{A\}, \{C\}, \{A\}\}$	$\{\{A\}, \{C\}, \{C\}\}$	$\{\{B\}, \{A, B\}\}$	$\{\{B\}, \{A, C\}\}$	$\{\{B\}, \{A\}, \{A\}\}$
1	0	0	3	2	1
$\{\{B\}, \{A\}, \{C\}\}$	$\{\{B\}, \{B\}, \{A\}\}$	$\{\{B\}, \{B\}, \{B\}\}$	$\{\{B\}, \{B\}, \{C\}\}$	$\{\{B\}, \{C\}, \{A\}\}$	$\{\{B\}, \{C\}, \{B\}\}$
1	1	0	1	3	3
$\{\{B\}, \{C\}, \{C\}\}$	$\{\{C\}, \{A, B\}\}$	$\{\{C\}, \{A, C\}\}$	$\{\{C\}, \{A\}, \{A\}\}$	$\{\{C\}, \{A\}, \{C\}\}$	$\{\{C\}, \{B\}, \{A\}\}$
1	5	2	2	3	2
$\{\{C\}, \{B\}, \{B\}\}$	$\{\{C\}, \{B\}, \{C\}\}$	$\{\{C\}, \{C\}, \{A\}\}$	$\{\{C\}, \{C\}, \{B\}\}$	$\{\{C\}, \{C\}, \{C\}\}$	$\{\{D\}, \{A\}, \{A\}\}$
0	3	0	0	1	1

Frequent 3-sequences:

$\{\{A, B\}, \{C\}\}, \{\{B\}, \{A, B\}\}, \{\{B\}, \{C\}, \{A\}\}, \{\{B\}, \{C\}, \{B\}\}, \{\{C\}, \{A, B\}\}, \{\{C\}, \{A\}, \{C\}\},$
 $\{\{C\}, \{B\}, \{C\}\}$

Candidate Generation (4th pass), candidate 4-sequence generated:

$\{\{A, B\}, \{C\}\} \times \{\{B\}, \{C\} \dots\} \Rightarrow \{\{A, B\}, \{C\}, \{A\}\}, \{\{A, B\}, \{C\}, \{B\}\}$
 $\{\{B\}, \{A, B\}\} \times \{\{A, B\} \dots\} \Rightarrow \{\{B\}, \{A, B\}, \{C\}\}$
 $\{\{B\}, \{C\}, \{A\}\} \times \{\{C\}, \{A\} \dots\} \Rightarrow \{\{B\}, \{C\}, \{A, B\}\}, \{\{B\}, \{C\}, \{A\}, \{C\}\}$
 $\{\{B\}, \{C\}, \{B\}\} \times \{\{C\}, \{B\} \dots\} \Rightarrow \{\{B\}, \{C\}, \{B\}, \{C\}\}$
 $\{\{C\}, \{A, B\}\} \times \{\{A, B\} \dots\} \Rightarrow \{\{C\}, \{A, B\}, \{C\}\}$
 $\{\{C\}, \{A\}, \{C\}\} \times \{\{A\}, \{C\} \dots\} \Rightarrow (\text{none})$
 $\{\{C\}, \{B\}, \{C\}\} \times \{\{B\}, \{C\} \dots\} \Rightarrow \{\{C\}, \{B\}, \{C\}, \{A\}\}, \{\{C\}, \{B\}, \{C\}, \{B\}\}$

Candidate Pruning (4th pass), remove 4-sequence with infrequent 3-sequence:

$\{\{A, B\}, \{C\}, \{A\}\}, \{\{A, B\}, \{C\}, \{B\}\}, \{\{B\}, \{A, B\}, \{C\}\}, \{\{B\}, \{C\}, \{A, B\}\}, \{\{B\}, \{C\}, \{A\}, \{C\}\},$
 $\{\{B\}, \{C\}, \{B\}, \{C\}\}, \{\{C\}, \{A, B\}, \{C\}\}, \{\{C\}, \{B\}, \{C\}, \{A\}\}, \{\{C\}, \{B\}, \{C\}, \{B\}\}$

Support counting (4th pass):

$\{\{B\}, \{C\}, \{A, B\}\}$	$\{\{C\}, \{A, B\}, \{C\}\}$
3	3

All of them are frequent.

Candidate Generation (5th pass), candidate 5-sequence generated:

$\{\{B\}, \{C\}, \{A, B\}\} \times \{\{C\}, \{A, B\} \dots\} \Rightarrow \{\{B\}, \{C\}, \{A, B\}, \{C\}\}$
 $\{\{C\}, \{A, B\}, \{C\}\} \times \{\{A, B\}, \{C\} \dots\} \Rightarrow (\text{none})$

Candidate Pruning (5th pass), remove 5-sequence with infrequent 4-sequence:

$\{\{B\}, \{C\}, \{A, B\}, \{C\}\}$

No more frequent sequence found in this pass, algorithm stops here.