

Question 7.

- Done in code.
- It took about 3.2-3.4 seconds to convert the file from .dat to .arff.
- It took around 3-4 seconds to load the file on WEKA.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **None** Apply Stop

Current relation
Relation: kosarak
Instances: 990002
Attributes: 41270
Sum of weights: 990002

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> i1
2	<input type="checkbox"/> i2
3	<input type="checkbox"/> i3
4	<input type="checkbox"/> i4
5	<input type="checkbox"/> i5
6	<input type="checkbox"/> i6
7	<input type="checkbox"/> i7
8	<input type="checkbox"/> i8
9	<input type="checkbox"/> i9
10	<input type="checkbox"/> i10
11	<input type="checkbox"/> i11
12	<input type="checkbox"/> i12
13	<input type="checkbox"/> i13
14	<input type="checkbox"/> i14
15	<input type="checkbox"/> i15
16	<input type="checkbox"/> i16
17	<input type="checkbox"/> i17
18	<input type="checkbox"/> i18
19	<input type="checkbox"/> i19

Remove

Selected attribute
Name: i1
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	0	792480	792480
2	1	197522	197522

Class: i41270 (Nom) Visualize All

792480

197522

Status
OK Log x 0

d.

The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. A dialog box titled 'weka.gui.GenericObjectEditor' is open, displaying the configuration for the 'weka.associations.FPGrowth' class. The configuration includes various parameters such as 'delta', 'doNotCheckCapabilities', 'findAllRulesForSupportLevel', 'lowerBoundMinSupport', 'maxNumberOfItems', 'metricType', 'minMetric', 'numRulesToFind', 'positiveIndex', 'rulesMustContain', 'transactionsMustContain', 'upperBoundMinSupport', and 'useORForMustContainList'. The 'Start' button is visible in the background.

Weka Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **FPGrowth -P 2 -I -1 -N 10 -T 0**

Start Stop

Result list (right-click for options)

weka.associations.FPGrowth

About

Class implementing the FP-growth algorithm for finding large item sets without candidate generation.

More Capabilities

delta 0.05

doNotCheckCapabilities False

findAllRulesForSupportLevel True

lowerBoundMinSupport 0.05

maxNumberOfItems -1

metricType Confidence

minMetric 0.99

numRulesToFind 10

positiveIndex 2

rulesMustContain

transactionsMustContain

upperBoundMinSupport 1.0

useORForMustContainList False

Weka Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose **FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.99 -D 0.05 -U 1.0 -M 0.05 -S**

Start Stop

Result list (right-click f...)

18:28:40 - FPGrowth

Associator output

=== Run information ===

Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.99 -D 0.05 -U 1.0 -M 0.05 -S

Relation: kosarak

Instances: 990002

Attributes: 41270

[list of attributes omitted]

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

FPGrowth found 2 rules

1. [i11=1, i218=1, i148=1]: 50098 ==> [i6=1]: 49866 <conf:(1)> lift:(1.64) lev:(0.02) conv:(84.4)

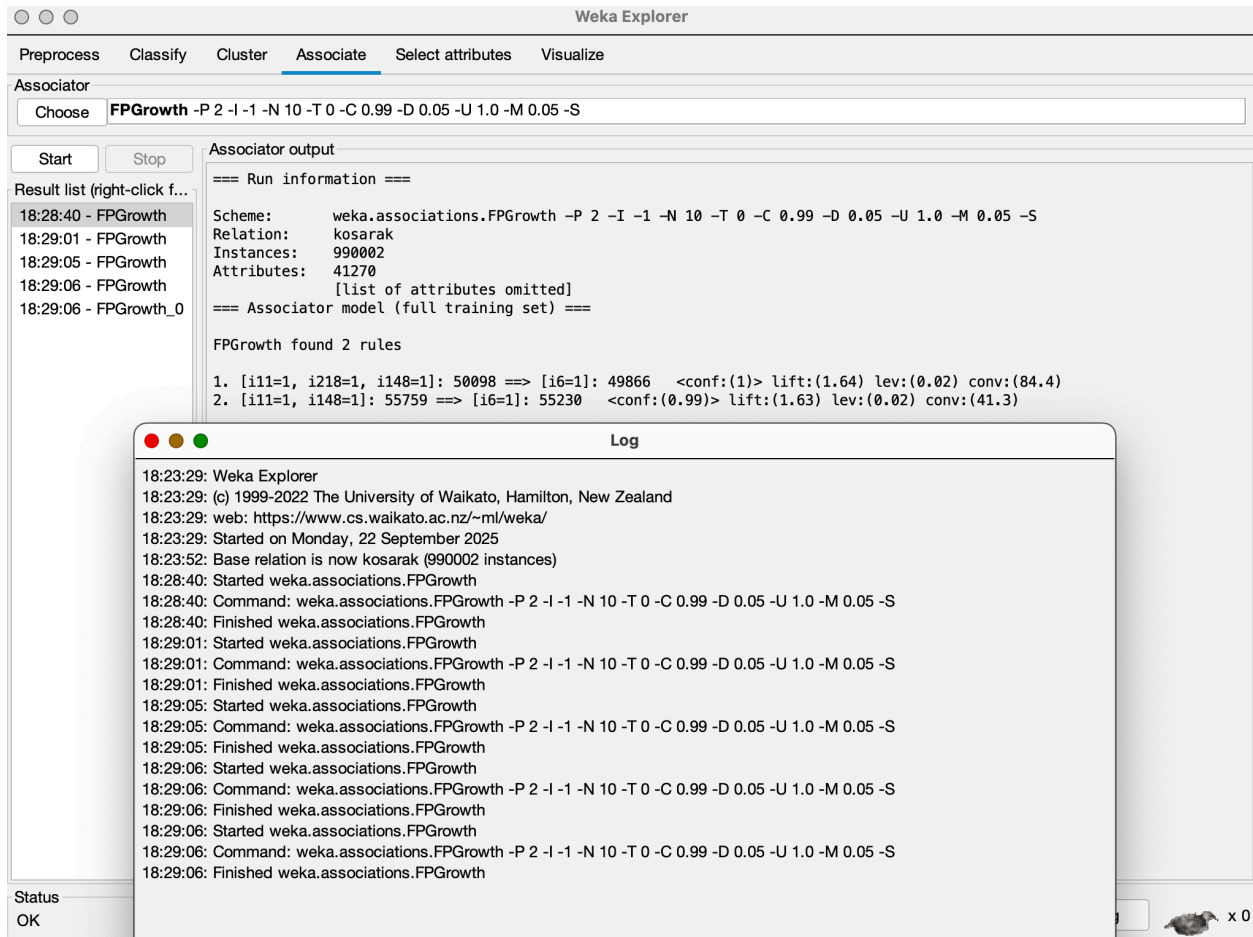
2. [i11=1, i148=1]: 55759 ==> [i6=1]: 55230 <conf:(0.99)> lift:(1.63) lev:(0.02) conv:(41.3)

Status

OK

Log x 0

e.



It each of the 5 runs it took < 1 second to run the algorithm. It takes around a total of ~7.5 seconds to convert the file to arff and then for it to load on weka, hence time taken to run FP growth is significantly lower.