

CSE 601: DATA MINING AND BIOINFORMATICS

PROJECT 1: PCA AND APRIORI ALGORITHM

PART 2: ASSOCIATION ANALYSIS

REPORT

Keerthana Baskaran UB NO: 50288944 (UBIT: kbaskara)

Sriram Venkataramanan UB NO : 50289666 (UB IT: sv84)

Vivek Nagaraju UB NO : 50288711 (UB IT: viveknag)

Apriori Algorithm:

It is a classic algorithm which is used in mining frequent itemsets and finding the relevant association rules. It is generally applied on database containing a large number of transactions. It identifies the frequent individual items in database and then it extends to larger item sets as long as those sets appear frequently in the transactions.

The Apriori algorithm works on two basic key ideas,

- Ø All subsets of a frequent itemset must be frequent.
- Ø If an itemset is infrequent, all its supersets will be infrequent.

Implementation Logic:

Frequent Itemset Generation

- a) Read the input file “associationrulesdata.txt” and get the support value from the user.
- b) Preprocess the data and convert them into required format by adding the required prefix and suffix (ie adding “G” , “_” etc)
- c) Use a dictionary to store the frequency of itemsets . The key for this dictionary will be each individual item converted to a tuple . The value will be the frequency of this itemset.
- d) We start with length = 1 to generate the frequent itemsets and increment the length by 1 in each subsequent step
- e) In each step , we prune the itemsets based on the support value.
- f) We also update the dictionary to update the frequency of the itemsets.
- g) We continue doing this until there are no more frequent itemsets.

Rule Generation

- a) Get Confidence value from the user.
- b) For each itemset in the frequent itemset list , we generate rules based on the confidence value provided by the user.

Template Formatting

Template 1

- a) We extract the head and body parts from the rules and store them in sets
- b) We then take the intersection of two sets based on the parameters provided by the user and return the final set of rules

Template 2

- a) Extract the head and body parts from the rules and store them in sets.
- b) Return the set if the size of the required set is greater than the threshold value obtained from the user.

Template 3

- a) Split the first parameter based on either “or” or “and” keyword.
- b) Return the intersection or union of two template rules depending on the input provided by the user.

RESULTS OBTAINED

FREQUENT ITEMSET LENGTH OUTPUTa

1) Min_support = 50. (Default given in question)

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_SOLN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 50
Support is set to be 50%
number of length-1 frequent itemsets: 109
number of length-2 frequent itemsets: 63
number of length-3 frequent itemsets: 2
number of length-4 frequent itemsets: 0
number of all lengths frequent itemsets 174
Enter confidence percentage:
```

2) Min_support =30

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_SOLN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 30
Support is set to be 30%
number of length-1 frequent itemsets: 196
number of length-2 frequent itemsets: 5340
number of length-3 frequent itemsets: 5287
number of length-4 frequent itemsets: 1518
number of length-5 frequent itemsets: 438
number of length-6 frequent itemsets: 88
number of length-7 frequent itemsets: 11
number of length-8 frequent itemsets: 1
number of length-9 frequent itemsets: 0
number of all lengths frequent itemsets 12879
Enter confidence percentage: _
```

3) Min_support=40

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_SOLN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 40
Support is set to be 40%
number of length-1 frequent itemsets: 167
number of length-2 frequent itemsets: 753
number of length-3 frequent itemsets: 149
number of length-4 frequent itemsets: 7
number of length-5 frequent itemsets: 1
number of length-6 frequent itemsets: 0
number of all lengths frequent itemsets 1077
Enter confidence percentage: _
```

4) Min_support=60

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_SOLN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 60
Support is set to be 60%
number of length-1 frequent itemsets: 34
number of length-2 frequent itemsets: 2
number of length-3 frequent itemsets: 0
number of all lengths frequent itemsets 36
Enter confidence percentage: _
```

5) Min_support=70

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_S0LN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 70
Support is set to be 70%
number of length-1 frequent itemsets: 7
number of length-2 frequent itemsets: 0
number of all lengths frequent itemsets 7
Enter confidence percentage:
```

6) Min_support=80

```
E:\3rdSEM\DATA MINING J GAO\PROJECTS\PROJECT1_S0LN\ASS>python apriori.py
Enter the name of the input file: associationruletestdata.txt
Enter the Minimum Support Count: 80
Support is set to be 80%
number of length-1 frequent itemsets: 0
number of all lengths frequent itemsets 0
Enter confidence percentage: _
```

Once Support reaches 80, there are no frequent item set.

RULE GENERATION OUTPUTS FOR THE TEMPLATES PROVIDED

min_support=50

Conf_value =70

```
Enter the Minimum Support Count: 50
Support is set to be 50%
number of length-1 frequent itemsets: 109
number of length-2 frequent itemsets: 63
number of length-3 frequent itemsets: 2
number of length-4 frequent itemsets: 0
number of all lengths frequent itemsets 174
Enter confidence percentage: 70
rule count ---- 117
Enter the template number:
```

TOTAL RULE COUNT =117

TEMPLATE 1

Template 1: {RULE|HEAD|BODY} HAS ({ANY|NUMBER|NONE}) OF (ITEM1, ITEM2, ..., ITEMn)

For template 1, we have 9 possible keywords combinations:

```
(result11, cnt) = asso_rule.template1("RULE", "ANY", ['G59_UP'])
(result12, cnt) = asso_rule.template1("RULE", "NONE", ['G59_UP'])
(result13, cnt) = asso_rule.template1("RULE", 1, ['G59_UP', 'G10_Down'])
(result14, cnt) = asso_rule.template1("HEAD", "ANY", ['G59_UP'])
(result15, cnt) = asso_rule.template1("HEAD", "NONE", ['G59_UP'])
(result16, cnt) = asso_rule.template1("HEAD", 1, ['G59_UP', 'G10_Down'])
(result17, cnt) = asso_rule.template1("BODY", "ANY", ['G59_UP'])
(result18, cnt) = asso_rule.template1("BODY", "NONE", ['G59_UP'])
(result19, cnt) = asso_rule.template1("BODY", 1, ['G59_UP', 'G10_Down'])
```

RESULTS OBTAINED FOR VARIOUS TEST CASES: Template 1:

1) (result11, cnt) = asso_rule.template1("rule", "any", ['g59_up'])

```
Enter the Minimum Support Count: 50
Support is set to be 50%
number of length-1 frequent itemsets: 109
number of length-2 frequent itemsets: 63
number of length-3 frequent itemsets: 2
number of length-4 frequent itemsets: 0
number of all lengths frequent itemsets 174
Enter confidence percentage: 70
rule count ---- 117
Enter the template number: 1
Enter First parameter for template 1: rule
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G88_DOWN->G59_UP', 'G82_DOWN->G72_UP, G59_UP', 'G96_DOWN->G59_UP', 'G87_UP->G59_UP', 'G6_UP->G59_UP', 'G1_UP->G59_UP', 'G72_UP, G82_DOWN->G59_UP', 'G72_UP->G59_UP', 'G72_UP, G96_DOWN->G59_UP', 'G59_UP, G96_DOWN->G72_UP', 'G59_UP->G96_DOWN', 'G82_DOWN->G59_UP', 'G32_DOWN->G59_UP', 'G38_DOWN->G59_UP', 'G96_DOWN->G72_UP, G59_UP', 'G59_UP, G82_DOWN->G72_UP', 'G59_UP, G72_UP->G96_DOWN', 'G28_DOWN->G59_UP', 'G59_UP->G72_UP', 'G72_UP->G59_UP, G82_DOWN', 'G59_UP->G88_DOWN', 'G59_UP->G13_DOWN', 'G13_DOWN->G59_UP', 'G10_DOWN->G59_UP', 'G59_UP, G72_UP->G82_DOWN', 'G59_UP->G82_DOWN'}
the count of the template rules are ... 26
```

Rules Obtained: 26

2) (result12, cnt) = asso_rule.template1("rule", "none", ['g59_up'])

```
Enter First parameter for template 1: rule
Enter Second parameter for template 1: none
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G1_UP->G70_DOWN', 'G65_DOWN->G38_DOWN', 'G38_DOWN->G28_DOWN', 'G91_UP->G38_DOW
N', 'G32_DOWN->G72_UP', 'G2_DOWN->G38_DOWN', 'G6_UP->G28_DOWN', 'G88_DOWN->G38_D
OWN', 'G10_DOWN->G70_DOWN', 'G94_UP->G38_DOWN', 'G54_UP->G1_UP', 'G8_UP->G88_DOW
N', 'G24_DOWN->G88_DOWN', 'G41_DOWN->G38_DOWN', 'G72_UP->G96_DOWN', 'G38_DOWN->G
1_UP', 'G10_DOWN->G38_DOWN', 'G38_DOWN->G70_DOWN', 'G6_UP->G13_DOWN', 'G28_DOWN-
->G88_DOWN', 'G13_DOWN->G72_UP', 'G28_DOWN->G2_DOWN', 'G88_DOWN->G24_DOWN', 'G10_
DOWN->G47_UP', 'G1_UP->G67_UP', 'G10_DOWN->G28_DOWN', 'G13_DOWN->G28_DOWN', 'G2_
DOWN->G28_DOWN', 'G70_DOWN->G10_DOWN', 'G38_DOWN->G52_DOWN', 'G82_DOWN->G97_DOW
N', 'G28_DOWN->G13_DOWN', 'G1_UP->G38_DOWN', 'G10_DOWN->G94_UP', 'G72_UP->G82_DOW
N', 'G38_DOWN->G32_DOWN', 'G96_DOWN->G82_DOWN', 'G28_DOWN->G10_DOWN', 'G38_DOWN-
->G91_UP', 'G1_UP->G54_UP', 'G82_DOWN->G96_DOWN', 'G47_UP->G38_DOWN', 'G28_DOWN->
G38_DOWN', 'G96_DOWN->G72_UP', 'G10_DOWN->G88_DOWN', 'G88_DOWN->G10_DOWN', 'G28_
DOWN->G52_DOWN', 'G88_DOWN->G41_DOWN', 'G67_UP->G38_DOWN', 'G47_UP->G28_DOWN', '
G70_DOWN->G38_DOWN', 'G6_UP->G38_DOWN', 'G1_UP->G72_UP', 'G13_DOWN->G6_UP', 'G70_
DOWN->G1_UP', 'G24_DOWN->G54_UP', 'G10_DOWN->G1_UP', 'G67_UP->G1_UP', 'G13_DOWN
->G82_DOWN', 'G32_DOWN->G28_DOWN', 'G41_DOWN->G28_DOWN', 'G72_UP->G13_DOWN', 'G8
8_DOWN->G54_UP', 'G88_DOWN->G87_UP', 'G38_DOWN->G47_UP', 'G88_DOWN->G8_UP', 'G54_
UP->G24_DOWN', 'G28_DOWN->G41_DOWN', 'G88_DOWN->G28_DOWN', 'G87_UP->G88_DOWN',
'G32_DOWN->G6_UP', 'G32_DOWN->G38_DOWN', 'G52_DOWN->G28_DOWN', 'G72_UP->G1_UP',
'G6_UP->G32_DOWN', 'G87_UP->G28_DOWN', 'G52_DOWN->G38_DOWN', 'G1_UP->G10_DOWN',
'G94_UP->G10_DOWN', 'G82_DOWN->G13_DOWN', 'G54_UP->G88_DOWN', 'G97_DOWN->G72_UP',
'G28_DOWN->G6_UP', 'G38_DOWN->G10_DOWN', 'G28_DOWN->G87_UP', 'G28_DOWN->G47_UP',
'G41_DOWN->G88_DOWN', 'G82_DOWN->G72_UP', 'G47_UP->G10_DOWN', 'G28_DOWN->G32_
DOWN', 'G97_DOWN->G82_DOWN'}
the count of the template rules are ... 91
```

Rules obtained : 91

3) (result13, cnt) = asso_rule.template1("rule", 1, ['g59_up', 'g10_down'])

```
Enter the template number: 1
Enter First parameter for template 1: rule
Enter Second parameter for template 1: 1
Enter Third parameter for template 1: g59_up,g10_down
the final set of template rules are:
{'G70_DOWN->G10_DOWN', 'G38_DOWN->G59_UP', 'G6_UP->G59_UP', 'G59_UP,G72_UP->G96_
DOWN', 'G59_UP,G72_UP->G82_DOWN', 'G10_DOWN->G38_DOWN', 'G82_DOWN->G72_UP,G59_UP
', 'G59_UP,G82_DOWN->G72_UP', 'G47_UP->G10_DOWN', 'G10_DOWN->G94_UP', 'G72_UP->G
59_UP', 'G59_UP,G96_DOWN->G72_UP', 'G10_DOWN->G1_UP', 'G10_DOWN->G70_DOWN', 'G59_
UP->G13_DOWN', 'G96_DOWN->G72_UP,G59_UP', 'G59_UP->G72_UP', 'G59_UP->G88_DOWN',
'G28_DOWN->G59_UP', 'G82_DOWN->G59_UP', 'G72_UP,G82_DOWN->G59_UP', 'G59_UP->G96_
DOWN', 'G72_UP->G82_DOWN,G59_UP', 'G59_UP->G82_DOWN', 'G10_DOWN->G47_UP', 'G10_
DOWN->G28_DOWN', 'G10_DOWN->G88_DOWN', 'G72_UP,G96_DOWN->G59_UP', 'G1_UP->G10_DOW
N', 'G32_DOWN->G59_UP', 'G88_DOWN->G59_UP', 'G38_DOWN->G10_DOWN', 'G96_DOWN->G5
9_UP', 'G28_DOWN->G10_DOWN', 'G88_DOWN->G10_DOWN', 'G94_UP->G10_DOWN', 'G87_UP->
G59_UP', 'G13_DOWN->G59_UP', 'G1_UP->G59_UP'}
the count of the template rules are ... 39
```

Rules obtained : 39

4) (result14, cnt) = asso_rule.template1("head", "any", ['g59_up'])

```
Enter the template number: 1
Enter First parameter for template 1: head
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G59_UP->G96_DOWN', 'G59_UP->G72_UP', 'G59_UP,G72_UP->G96_DOWN', 'G59_UP,G72_UP
->G82_DOWN', 'G59_UP->G13_DOWN', 'G59_UP,G96_DOWN->G72_UP', 'G59_UP->G82_DOWN',
'G59_UP->G88_DOWN', 'G59_UP,G82_DOWN->G72_UP'}
the count of the template rules are ... 9
```

Rules obtained: 9

5) (result15, cnt) = asso_rule.template1("head", "none", ['g59_up'])

```

Enter First parameter for template 1: head
Enter Second parameter for template 1: none
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G38_DOWN->G52_DOWN', 'G72_UP->G13_DOWN', 'G1_UP->G67_UP', 'G94_UP->G10_DOWN',
'G38_DOWN->G10_DOWN', 'G88_DOWN->G28_DOWN', 'G88_DOWN->G38_DOWN', 'G72_UP->G96_D
OWN', 'G72_UP->G82_DOWN->G59_UP', 'G28_DOWN->G41_DOWN', 'G82_DOWN->G96_DOWN', 'G1
3_DOWN->G82_DOWN', 'G28_DOWN->G88_DOWN', 'G41_DOWN->G38_DOWN', 'G72_UP->G82_DOWN',
'G67_UP->G1_UP', 'G28_DOWN->G10_DOWN', 'G97_DOWN->G72_UP', 'G32_DOWN->G28_DOWN',
'G1_UP->G10_DOWN', 'G41_DOWN->G88_DOWN', 'G82_DOWN->G59_UP', 'G24_DOWN->G54_
UP', 'G88_DOWN->G24_DOWN', 'G96_DOWN->G59_UP', 'G94_UP->G38_DOWN', 'G67_UP->G38_
DOWN', 'G82_DOWN->G72_UP', 'G88_DOWN->G8_UP', 'G28_DOWN->G2_DOWN', 'G72_UP->G59_
UP', 'G28_DOWN->G47_UP', 'G28_DOWN->G87_UP', 'G88_DOWN->G87_UP', 'G6_UP->G28_DOWN',
'G10_DOWN->G28_DOWN', 'G38_DOWN->G47_UP', 'G82_DOWN->G97_DOWN', 'G91_UP->G38_
DOWN', 'G1_UP->G72_UP', 'G10_DOWN->G88_DOWN', 'G13_DOWN->G72_UP', 'G10_DOWN->G3
8_DOWN', 'G10_DOWN->G47_UP', 'G1_UP->G54_UP', 'G38_DOWN->G59_UP', 'G82_DOWN->G72_
UP', 'G59_UP', 'G6_UP->G59_UP', 'G2_DOWN->G38_DOWN', 'G28_DOWN->G32_DOWN', 'G87_UP
->G59_UP', 'G54_UP->G1_UP', 'G24_DOWN->G88_DOWN', 'G96_DOWN->G82_DOWN', 'G32_DOWN
->G59_UP', 'G70_DOWN->G10_DOWN', 'G10_DOWN->G70_DOWN', 'G13_DOWN->G28_DOWN', 'G
1_UP->G59_UP', 'G1_UP->G70_DOWN', 'G38_DOWN->G91_UP', 'G72_UP->G59_UP', 'G82_DOWN',
'G6_UP->G32_DOWN', 'G2_DOWN->G28_DOWN', 'G87_UP->G28_DOWN', 'G97_DOWN->G82_DOWN',
'G28_DOWN->G13_DOWN', 'G65_DOWN->G38_DOWN', 'G88_DOWN->G59_UP', 'G70_DOWN->G3
8_DOWN', 'G54_UP->G88_DOWN', 'G10_DOWN->G59_UP', 'G6_UP->G13_DOWN', 'G32_DOWN->G
72_UP', 'G96_DOWN->G72_UP', 'G13_DOWN->G6_UP', 'G88_DOWN->G41_DOWN', 'G82_DOWN->
G13_DOWN', 'G38_DOWN->G28_DOWN', 'G54_UP->G24_DOWN', 'G38_DOWN->G70_DOWN', 'G87_
UP->G88_DOWN', 'G47_UP->G28_DOWN', 'G8_UP->G88_DOWN', 'G10_DOWN->G1_UP', 'G28_DOWN->
G38_DOWN', 'G72_UP->G96_DOWN->G59_UP', 'G28_DOWN->G52_DOWN', 'G28_DOWN->G59_U
P', 'G52_DOWN->G38_DOWN', 'G88_DOWN->G54_UP', 'G88_DOWN->G10_DOWN', 'G10_DOWN->G
94_UP', 'G72_UP->G1_UP', 'G70_DOWN->G1_UP', 'G47_UP->G10_DOWN', 'G38_DOWN->G1_UP',
'G96_DOWN->G72_UP', 'G59_UP', 'G6_UP->G38_DOWN', 'G28_DOWN->G6_UP', 'G52_DOWN->G
28_DOWN', 'G38_DOWN->G32_DOWN', 'G32_DOWN->G38_DOWN', 'G1_UP->G38_DOWN', 'G32_DOWN
->G6_UP', 'G13_DOWN->G59_UP', 'G41_DOWN->G28_DOWN', 'G47_UP->G38_DOWN'}
the count of the template rules are ... 108

```

Rules obtained: 108

6) (result16, cnt) = asso_rule.template1("head", 1, ['g59_up', 'g10_down'])

```

Enter the template number: 1
Enter First parameter for template 1: head
Enter Second parameter for template 1: 1
Enter Third parameter for template 1: g59_up,g10_down
the final set of template rules are:
{'G10_DOWN->G59_UP', 'G59_UP->G82_DOWN', 'G59_UP->G72_UP->G82_DOWN', 'G59_UP->G82_
DOWN->G38_DOWN', 'G10_DOWN->G38_DOWN', 'G59_UP->G88_DOWN', 'G59_UP->G72_UP->G96_DOWN',
'G10_DOWN->G88_DOWN', 'G59_UP->G72_UP', 'G59_UP->G13_DOWN', 'G59_UP->G96_DOWN',
'G10_DOWN->G94_UP', 'G10_DOWN->G28_DOWN', 'G10_DOWN->G47_UP', 'G10_DOWN->G70_
DOWN', 'G10_DOWN->G1_UP', 'G59_UP->G96_DOWN->G72_UP'}
The count of the template rules are ... 17

```

Rules obtained: 17

7) (result17, cnt) = asso_rule.template1("body", "any", ['g59_up'])

```

Enter the template number: 1
Enter First parameter for template 1: body
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G96_DOWN->G59_UP', 'G72_UP', 'G10_DOWN->G59_UP', 'G1_UP->G59_UP', 'G87_UP->G59_UP',
'G72_UP->G59_UP', 'G13_DOWN->G59_UP', 'G82_DOWN->G59_UP', 'G82_DOWN->G59_UP', 'G
72_UP', 'G88_DOWN->G59_UP', 'G96_DOWN->G59_UP', 'G72_UP->G82_DOWN->G59_UP', 'G32_
DOWN->G59_UP', 'G72_UP->G82_DOWN', 'G59_UP', 'G28_DOWN->G59_UP', 'G6_UP->G59_UP',
'G38_DOWN->G59_UP', 'G72_UP', 'G96_DOWN->G59_UP'}
the count of the template rules are ... 17

```

Rules obtained: 17

8) (result18, cnt) = asso_rule.template1("BODY", "NONE", ['G59_UP'])

```
Enter the template number: 1
Enter First parameter for template 1: body
Enter Second parameter for template 1: none
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G10_DOWN->G38_DOWN', 'G88_DOWN->G38_DOWN', 'G28_DOWN->G47_UP', 'G10_DOWN->G88_DOWN', 'G24_DOWN->G54_UP', 'G91_UP->G38_DOWN', 'G82_DOWN->G96_DOWN', 'G47_UP->G28_DOWN', 'G96_DOWN->G82_DOWN', 'G10_DOWN->G28_DOWN', 'G59_UP->G13_DOWN', 'G32_DOWN->G6_UP', 'G10_DOWN->G1_UP', 'G59_UP->G72_UP', 'G28_DOWN->G38_DOWN', 'G6_UP->G13_DOWN', 'G2_DOWN->G38_DOWN', 'G32_DOWN->G72_UP', 'G28_DOWN->G10_DOWN', 'G88_DOWN->G28_DOWN', 'G54_UP->G1_UP', 'G65_DOWN->G38_DOWN', 'G38_DOWN->G28_DOWN', 'G1_UP->G54_UP', 'G32_DOWN->G38_DOWN', 'G82_DOWN->G13_DOWN', 'G13_DOWN->G72_UP', 'G88_DOWN->G10_DOWN', 'G10_DOWN->G70_DOWN', 'G1_UP->G70_DOWN', 'G1_UP->G72_UP', 'G38_DOWN->G70_DOWN', 'G41_DOWN->G28_DOWN', 'G38_DOWN->G47_UP', 'G10_DOWN->G47_UP', 'G88_DOWN->G8_UP', 'G13_DOWN->G82_DOWN', 'G47_UP->G38_DOWN', 'G54_UP->G88_DOWN', 'G47_UP->G10_DOWN', 'G88_DOWN->G41_DOWN', 'G52_DOWN->G28_DOWN', 'G41_DOWN->G38_DOWN', 'G72_UP->G96_DOWN', 'G88_DOWN->G54_UP', 'G28_DOWN->G41_DOWN', 'G59_UP->G72_UP', 'G96_DOWN->G28_DOWN', 'G28_DOWN->G2_DOWN', 'G59_UP->G82_DOWN', 'G38_DOWN->G52_DOWN', 'G88_DOWN->G24_DOWN', 'G28_DOWN->G13_DOWN', 'G70_DOWN->G38_DOWN', 'G28_DOWN->G6_UP', 'G28_DOWN->G87_UP', 'G13_DOWN->G6_UP', 'G54_UP->G24_DOWN', 'G28_DOWN->G88_DOWN', 'G24_DOWN->G88_DOWN', 'G67_UP->G38_DOWN', 'G88_DOWN->G87_UP', 'G87_UP->G28_DOWN', 'G2_DOWN->G28_DOWN', 'G70_DOWN->G10_DOWN', 'G6_UP->G38_DOWN', 'G13_DOWN->G28_DOWN', 'G38_DOWN->G10_DOWN', 'G38_DOWN->G1_UP', 'G52_DOWN->G38_DOWN', 'G38_DOWN->G91_UP', 'G1_UP->G38_DOWN', 'G87_UP->G88_DOWN', 'G70_DOWN->G1_UP', 'G59_UP->G72_UP->G82_DOWN', 'G72_UP->G82_DOWN', 'G1_UP->G10_DOWN', 'G72_UP->G13_DOWN', 'G59_UP->G96_DOWN', 'G10_DOWN->G94_UP', 'G6_UP->G28_DOWN', 'G97_DOWN->G82_DOWN', 'G28_DOWN->G52_DOWN', 'G59_UP->G96_DOWN->G72_UP', 'G59_UP->G82_DOWN->G72_UP', 'G38_DOWN->G32_DOWN', 'G1_UP->G67_UP', 'G8_UP->G88_DOWN', 'G32_DOWN->G28_DOWN', 'G6_UP->G32_DOWN', 'G82_DOWN->G72_UP', 'G94_UP->G38_DOWN', 'G96_DOWN->G72_UP', 'G41_DOWN->G88_DOWN', 'G82_DOWN->G97_DOWN', 'G67_UP->G1_UP', 'G97_DOWN->G72_UP', 'G28_DOWN->G32_DOWN', 'G72_UP->G1_UP', 'G59_UP->G88_DOWN', 'G94_UP->G10_DOWN'}
the count of the template rules are ... 100
E:\3rdSEM\DATA MINING J GA0\PROJECTS\PROJECT1 SOLN\ASS>
```

Rules obtained: 100

9) (result19, cnt) = asso_rule.template1("BODY", 1, ['G59_UP', 'G10_Down'])

```
Enter the template number: 1
Enter First parameter for template 1: body
Enter Second parameter for template 1: 1
Enter Third parameter for template 1: g59_up,g10_down
the final set of template rules are:
{'G13_DOWN->G59_UP', 'G28_DOWN->G59_UP', 'G72_UP->G96_DOWN->G59_UP', 'G82_DOWN->G59_UP', 'G10_DOWN->G59_UP', 'G72_UP->G59_UP', 'G38_DOWN->G10_DOWN', 'G72_UP->G82_DOWN', 'G1_UP->G59_UP', 'G1_UP->G10_DOWN', 'G70_DOWN->G10_DOWN', 'G72_UP->G59_UP', 'G82_DOWN->G10_DOWN', 'G96_DOWN->G59_UP', 'G38_DOWN->G59_UP', 'G87_UP->G59_UP', 'G32_DOWN->G59_UP', 'G47_UP->G10_DOWN', 'G6_UP->G59_UP', 'G88_DOWN->G59_UP', 'G96_DOWN->G72_UP', 'G59_UP', 'G88_DOWN->G10_DOWN', 'G94_UP->G10_DOWN', 'G82_DOWN->G72_UP', 'G59_UP'}
the count of the template rules are ... 24
```

Rules obtained: 24

TEMPLATE 2:

Template 2: $\text{SizeOf}(\{\text{HEAD}|\text{BODY}|\text{RULE}\}) \geq \text{NUMBER}$.

For template 2, we have 3 keywords choices:

```
(result21, cnt) = asso_rule.template2("RULE", 3)
(result22, cnt) = asso_rule.template2("HEAD", 2)
(result23, cnt) = asso_rule.template2("BODY", 1)
```

RESULTS OBTAINED FOR VARIOUS TEST CASES: TEMPLATE 2

1) $(\text{result21}, \text{cnt}) = \text{asso_rule.template2}(\text{"RULE"}, 3)$

```
Enter the Minimum Support Count: 50
Support is set to be 50%
number of length-1 frequent itemsets: 109
number of length-2 frequent itemsets: 63
number of length-3 frequent itemsets: 2
number of length-4 frequent itemsets: 0
number of all lengths frequent itemsets 174
Enter confidence percentage: 70
rule count ---- 117
Enter the template number: 2
Enter First parameter for template 2: rule
Enter Second parameter for template 2: 3
the final set of template rules are:
{'G72_UP,G96_DOWN->G59_UP', 'G59_UP,G96_DOWN->G72_UP', 'G72_UP,G82_DOWN->G59_UP',
 'G72_UP->G82_DOWN,G59_UP', 'G59_UP,G72_UP->G96_DOWN', 'G59_UP,G82_DOWN->G72_UP',
 'G96_DOWN->G72_UP,G59_UP', 'G82_DOWN->G72_UP,G59_UP', 'G59_UP,G72_UP->G82_DOWN'}
the count of the template rules are ... 9
```

Rules obtained: 9

2) $(\text{result22}, \text{cnt}) = \text{asso_rule.template2}(\text{"HEAD"}, 2)$

```
Enter the Minimum Support Count: 50
Support is set to be 50%
number of length-1 frequent itemsets: 109
number of length-2 frequent itemsets: 63
number of length-3 frequent itemsets: 2
number of length-4 frequent itemsets: 0
number of all lengths frequent itemsets 174
Enter confidence percentage: 70
rule count ---- 117
Enter the template number: 2
Enter First parameter for template 2: head
Enter Second parameter for template 2: 2
the final set of template rules are:
{'G59_UP,G82_DOWN->G72_UP', 'G59_UP,G72_UP->G96_DOWN', 'G59_UP,G72_UP->G82_DOWN',
 'G72_UP,G96_DOWN->G59_UP', 'G72_UP,G82_DOWN->G59_UP', 'G59_UP,G96_DOWN->G72_UP'}
the count of the template rules are ... 6
```

Rules obtained: 6

3) $(\text{result23}, \text{cnt}) = \text{asso_rule.template2}(\text{"BODY"}, 1)$

```

Enter the template number: 2
Enter First parameter for template 2: body
Enter Second parameter for template 2: 1
the final set of template rules are:

```

```

{ 'G87_UP->G59_UP', 'G38_DOWN->G91_UP', 'G96_DOWN->G82_DOWN', 'G67_UP->G1_UP', 'G
65_DOWN->G38_DOWN', 'G38_DOWN->G47_UP', 'G32_DOWN->G28_DOWN', 'G88_DOWN->G24_DOW
N', 'G6_UP->G32_DOWN', 'G6_UP->G38_DOWN', 'G6_UP->G59_UP', 'G70_DOWN->G10_DOWN',
'G59_UP->G72_UP->G82_DOWN', 'G72_UP->G1_UP', 'G59_UP->G82_DOWN', 'G88_DOWN->G59_
UP', 'G1_UP->G38_DOWN', 'G87_UP->G28_DOWN', 'G72_UP->G59_UP', 'G59_UP->G96_DOWN->
G72_UP', 'G87_UP->G88_DOWN', 'G13_DOWN->G59_UP', 'G72_UP->G82_DOWN', 'G59_UP', 'G54
_UP->G1_UP', 'G32_DOWN->G38_DOWN', 'G10_DOWN->G38_DOWN', 'G59_UP->G96_DOWN', 'G1
0_DOWN->G28_DOWN', 'G28_DOWN->G88_DOWN', 'G59_UP->G72_UP', 'G96_DOWN->G72_UP',
'G88_DOWN->G28_DOWN', 'G38_DOWN->G59_UP', 'G38_DOWN->G70_DOWN', 'G38_DOWN->G32_DOW
N', 'G94_UP->G10_DOWN', 'G91_UP->G38_DOWN', 'G28_DOWN->G2_DOWN', 'G41_DOWN->G38
_DOWN', 'G82_DOWN->G59_UP', 'G41_DOWN->G28_DOWN', 'G13_DOWN->G72_UP', 'G88_DOWN-
->G10_DOWN', 'G59_UP->G72_UP->G96_DOWN', 'G28_DOWN->G59_UP', 'G10_DOWN->G94_UP',
'G41_DOWN->G88_DOWN', 'G47_UP->G10_DOWN', 'G54_UP->G88_DOWN', 'G24_DOWN->G54_UP',
'G1_UP->G54_UP', 'G2_DOWN->G38_DOWN', 'G1_UP->G72_UP', 'G88_DOWN->G8_UP', 'G82_
_DOWN->G59_UP->G72_UP', 'G72_UP->G13_DOWN', 'G8_UP->G88_DOWN', 'G47_UP->G28_DOWN',
'G72_UP->G96_DOWN', 'G70_DOWN->G38_DOWN', 'G28_DOWN->G52_DOWN', 'G10_DOWN->G59_
UP', 'G28_DOWN->G10_DOWN', 'G28_DOWN->G47_UP', 'G38_DOWN->G52_DOWN', 'G96_DOWN->
G59_UP', 'G28_DOWN->G13_DOWN', 'G47_UP->G38_DOWN', 'G10_DOWN->G70_DOWN', 'G6_UP-
->G13_DOWN', 'G59_UP->G82_DOWN->G72_UP', 'G32_DOWN->G6_UP', 'G59_UP->G13_DOWN', 'G
82_DOWN->G96_DOWN', 'G59_UP->G88_DOWN', 'G96_DOWN->G59_UP', 'G72_UP', 'G67_UP->G38_
_DOWN', 'G28_DOWN->G32_DOWN', 'G1_UP->G70_DOWN', 'G72_UP->G82_DOWN', 'G10_DOWN->G
47_UP', 'G32_DOWN->G72_UP', 'G1_UP->G67_UP', 'G38_DOWN->G28_DOWN', 'G38_DOWN->G1
_UP', 'G97_DOWN->G72_UP', 'G1_UP->G10_DOWN', 'G13_DOWN->G6_UP', 'G72_UP->G82_DOWN
->G59_UP', 'G82_DOWN->G72_UP', 'G88_DOWN->G54_UP', 'G70_DOWN->G1_UP', 'G6_UP->G2
8_DOWN', 'G28_DOWN->G6_UP', 'G13_DOWN->G82_DOWN', 'G82_DOWN->G13_DOWN', 'G28_DOW
N->G38_DOWN', 'G52_DOWN->G28_DOWN', 'G52_DOWN->G38_DOWN', 'G97_DOWN->G82_DOWN',
'G88_DOWN->G87_UP', 'G2_DOWN->G28_DOWN', 'G10_DOWN->G1_UP', 'G88_DOWN->G38_DOWN',
'G72_UP->G96_DOWN->G59_UP', 'G24_DOWN->G88_DOWN', 'G28_DOWN->G87_UP', 'G88_DOWN
->G41_DOWN', 'G13_DOWN->G28_DOWN', 'G28_DOWN->G41_DOWN', 'G1_UP->G59_UP', 'G54_U
P->G24_DOWN', 'G38_DOWN->G10_DOWN', 'G32_DOWN->G59_UP', 'G82_DOWN->G97_DOWN', 'G
94_UP->G38_DOWN', 'G10_DOWN->G88_DOWN' }
the count of the template rules are ... 117

```

Rules obtained: 117

TEMPLATE 3:

Template 3: Any combined templates using AND or OR. For example: BODY HAS (1) OF (Disease) AND HEAD HAS (NONE) OF (Disease)

For template 3, you need to implement AND/OR logical operator to connect two parts which can be from either template 1 or template 2. For example, "1or1" means "a query of Template1 OR another query of Template1".

```

(result31, cnt) = asso_rule.template3("1or1", "HEAD", "ANY",
['G10_Down'], "BODY", 1, ['G59_UP'])
(result32, cnt) = asso_rule.template3("1and1", "HEAD", "ANY",
['G10_Down'], "BODY", 1, ['G59_UP'])
(result33, cnt) = asso_rule.template3("1or2", "HEAD", "ANY",
['G10_Down'], "BODY", 2)
(result34, cnt) = asso_rule.template3("1and2", "HEAD", "ANY",
['G10_Down'], "BODY", 2)

```


RESULTS OBTAINED FOR VARIOUS TEST CASES: TEMPLATE 3

- 1) (result31, cnt) = asso_rule.template3("1or1", "HEAD", "ANY", ['G10_Down'], "BODY", 1, ['G59_UP'])

```
Enter the template number: 3
Enter the first parameter: 1or1
you have chosen template 1 for first part . please provide the relevant parameter details...
Enter First parameter for template 1: head
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g10_down
you have chosen template 1 for second part . please provide the relevant parameter details...
Enter First parameter for template 1: body
Enter Second parameter for template 1: 1
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G10_DOWN->G59_UP', 'G10_DOWN->G94_UP', 'G96_DOWN->G59_UP', 'G96_DOWN->G72_UP', 'G59_UP->G10_DOWN->G1_UP', 'G10_DOWN->G70_DOWN', 'G82_DOWN->G72_UP', 'G59_UP', 'G10_DOWN->G38_DOWN', 'G72_UP->G59_UP', 'G10_DOWN->G47_UP', 'G6_UP->G59_UP', 'G87_UP->G59_UP', 'G72_UP->G82_DOWN', 'G59_UP', 'G13_DOWN->G59_UP', 'G72_UP', 'G82_DOWN->G59_UP', 'G10_DOWN->G28_DOWN', 'G10_DOWN->G88_DOWN', 'G72_UP', 'G96_DOWN->G59_UP', 'G32_DOWN->G59_UP', 'G28_DOWN->G59_UP', 'G38_DOWN->G59_UP', 'G1_UP->G59_UP', 'G82_DOWN->G59_UP', 'G88_DOWN->G59_UP'}
the count of the template rules are ... 24
```

Rules obtained: 24

- 2) (result32, cnt) = asso_rule.template3("1and1", "HEAD", "ANY", ['G10_Down'], "BODY", 1, ['G59_UP'])

```
Enter the template number: 3
Enter the first parameter: 1and1
you have chosen template 1 for first part . please provide the relevant parameter details...
Enter First parameter for template 1: head
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g10_down
you have chosen template 1 for second part . please provide the relevant parameter details...
Enter First parameter for template 1: body
Enter Second parameter for template 1: 1
Enter Third parameter for template 1: g59_up
the final set of template rules are:
{'G10_DOWN->G59_UP'}
the count of the template rules are ... 1
```

Rules obtained : 1

- 3) (result33, cnt) = asso_rule.template3("1or2", "HEAD", "ANY", ['G10_Down'], "BODY", 2)

```
Enter the template number: 3
Enter the first parameter: 1or2
you have chosen template 1 for first part . please provide the relevant parameter details...
Enter First parameter for template 1: head
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g10_down
Enter First parameter for template 2: body
Enter Second parameter for template 2: 2
the final set of template rules are:
{'G10_DOWN->G1_UP', 'G10_DOWN->G47_UP', 'G72_UP->G59_UP', 'G82_DOWN', 'G10_DOWN->G70_DOWN', 'G10_DOWN->G38_DOWN', 'G10_DOWN->G28_DOWN', 'G10_DOWN->G59_UP', 'G10_DOWN->G88_DOWN', 'G82_DOWN->G59_UP', 'G72_UP', 'G10_DOWN->G94_UP', 'G96_DOWN->G59_UP', 'G72_UP'}
the count of the template rules are ... 11
```

Rules obtained: 11

- 4) (result34, cnt) = asso_rule.template3("1and2", "HEAD", "ANY",
['G10_Down'], "BODY", 2)

```
Enter the template number: 3
Enter the first parameter: 1and2
you have chosen template 1 for first part . please provide the relevant parameter details...
Enter First parameter for template 1: head
Enter Second parameter for template 1: any
Enter Third parameter for template 1: g10_down
Enter First parameter for template 2: body
Enter Second parameter for template 2: 2
the final set of template rules are:
set()
the count of the template rules are ... 0
```

Rules obtained: 0

- 5) (result35, cnt) = asso_rule.template3("2or2", "HEAD", 1, "BODY", 2)

```
Enter details...
Enter First parameter for template 2: head
Enter Second parameter for template 2: 1
Enter First parameter for template 2: body
Enter Second parameter for template 2: 2
```

```
{ 'G28_DOWN->G38_DOWN', 'G67_UP->G38_DOWN', 'G54_UP->G24_DOWN', 'G91_UP->G38_DOWN',
'G97_DOWN->G72_UP', 'G13_DOWN->G82_DOWN', 'G38_DOWN->G47_UP', 'G72_UP->G59_UP',
'G82_DOWN', 'G82_DOWN->G59_UP', 'G1_UP->G70_DOWN', 'G28_DOWN->G32_DOWN', 'G70_DOWN->G10_DOWN',
'G13_DOWN->G72_UP', 'G28_DOWN->G52_DOWN', 'G8_UP->G88_DOWN', 'G59_UP->G72_UP->G82_DOWN',
'G10_DOWN->G94_UP', 'G2_DOWN->G38_DOWN', 'G67_UP->G1_UP',
'G6_UP->G28_DOWN', 'G72_UP->G96_DOWN->G59_UP', 'G6_UP->G13_DOWN', 'G2_DOWN->G28_DOWN',
DOWN', 'G1_UP->G59_UP', 'G82_DOWN->G96_DOWN', 'G88_DOWN->G41_DOWN', 'G72_UP->G96_DOWN',
DOWN', 'G6_UP->G32_DOWN', 'G13_DOWN->G6_UP', 'G28_DOWN->G41_DOWN', 'G70_DOWN->G38_DOWN',
'G10_DOWN->G38_DOWN', 'G47_UP->G38_DOWN', 'G32_DOWN->G28_DOWN', 'G38_DOWN->G1_UP',
'G32_DOWN->G72_UP', 'G96_DOWN->G59_UP', 'G28_DOWN->G13_DOWN', 'G1_UP->G67_UP',
'G88_DOWN->G28_DOWN', 'G59_UP->G72_UP->G96_DOWN', 'G72_UP->G82_DOWN',
'G72_UP->G1_UP', 'G1_UP->G38_DOWN', 'G96_DOWN->G72_UP->G59_UP', 'G24_DOWN->G54_UP',
'G59_UP->G82_DOWN->G72_UP', 'G59_UP->G96_DOWN', 'G87_UP->G59_UP', 'G54_UP->G1_UP',
UP', 'G72_UP->G59_UP', 'G59_UP->G72_UP', 'G10_DOWN->G59_UP', 'G32_DOWN->G38_DOWN',
'G10_DOWN->G70_DOWN', 'G65_DOWN->G38_DOWN', 'G38_DOWN->G91_UP', 'G88_DOWN->G8_UP',
'G59_UP->G82_DOWN', 'G59_UP->G13_DOWN', 'G47_UP->G10_DOWN', 'G87_UP->G88_DOWN',
DOWN', 'G88_DOWN->G38_DOWN', 'G38_DOWN->G32_DOWN', 'G82_DOWN->G97_DOWN', 'G82_DOWN->G13_DOWN',
'G32_DOWN->G59_UP', 'G87_UP->G28_DOWN', 'G38_DOWN->G70_DOWN', 'G32_DOWN->G6_UP',
'G47_UP->G28_DOWN', 'G1_UP->G10_DOWN', 'G10_DOWN->G88_DOWN', 'G28_DOWN->G10_DOWN',
'G88_DOWN->G10_DOWN', 'G88_DOWN->G87_UP', 'G38_DOWN->G59_UP',
'G96_DOWN->G72_UP', 'G41_DOWN->G28_DOWN', 'G1_UP->G72_UP', 'G28_DOWN->G87_UP',
'G28_DOWN->G88_DOWN', 'G94_UP->G10_DOWN', 'G10_DOWN->G28_DOWN', 'G97_DOWN->G82_DOWN',
DOWN', 'G59_UP->G96_DOWN->G72_UP', 'G94_UP->G38_DOWN', 'G38_DOWN->G28_DOWN', 'G28_DOWN->G6_UP',
'G70_DOWN->G1_UP', 'G88_DOWN->G24_DOWN', 'G38_DOWN->G52_DOWN', 'G41_DOWN->G88_DOWN',
'G82_DOWN->G72_UP', 'G96_DOWN->G82_DOWN', 'G52_DOWN->G28_DOWN',
'G6_UP->G59_UP', 'G10_DOWN->G47_UP', 'G72_UP->G82_DOWN->G59_UP', 'G88_DOWN->G54_UP',
'G28_DOWN->G59_UP', 'G82_DOWN->G72_UP->G59_UP', 'G13_DOWN->G28_DOWN', 'G10_DOWN->G1_UP',
'G1_UP->G54_UP', 'G59_UP->G88_DOWN', 'G41_DOWN->G38_DOWN', 'G28_DOWN->G47_UP',
'G54_UP->G88_DOWN', 'G6_UP->G38_DOWN', 'G24_DOWN->G88_DOWN', 'G13_DOWN->G59_UP',
'G52_DOWN->G38_DOWN', 'G88_DOWN->G59_UP', 'G38_DOWN->G10_DOWN',
'G72_UP->G13_DOWN', 'G28_DOWN->G2_DOWN'}
```

Rules obtained: 117

6) (result36, cnt) = asso_rule.template3("2and2", "HEAD", 1, "BODY", 2)

```
rule count --- 117
Enter the template number: 3
Enter the first parameter: 2and2
you have chosen get_template 2 for first part. please provide the relevant parameter details...
Enter First parameter for template 2: head
Enter Second parameter for template 2: 1
Enter First parameter for template 2: body
Enter Second parameter for template 2: 2
the final set of template rules are:
{'G82_DOWN->G59_UP,G72_UP', 'G96_DOWN->G59_UP,G72_UP', 'G72_UP->G82_DOWN,G59_UP'}
}
the count of the template rules are ... 3
```

Rules obtained: 3

CONCLUSION:

- 1) Apriori algorithm is implemented in python.
- 2) The frequent itemset is generated and the rules are generated.
- 3) The results are tested according to the 3 templates.

REFERENCES:

<https://docs.python.org/3/>