

Web Mining (CSE3024)

Lab Assignment 9

Name: Kritika Mishra

Registration Number: 16BCI0041

Slot: L15+L16

Date: 23rd October 2018

Question:

Apriori

Find the results of a store transaction data using the Apriori algorithm with support threshold $S=35\%$ and confidence threshold $C=60\%$. Show the candidate and frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Code:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from apyori import apriori

dataset =
[['HotDogs', 'Buns', 'Ketchup'], ['HotDogs', 'Buns'], ['HotDogs', 'Coke', 'Chips'], ['Chips', 'Coke'],
['Chips', 'Ketchup'], ['HotDogs', 'Coke', 'Chips']]
print("Given Dataset:")
print('-----')
for i in dataset:
    for j in i:
        print(j + ' | ', end="")
```

```

print('\n-----')
ar = apriori(dataset, min_support=0.35, min_confidence=0.6, min_lift=1, max_length=3)
result = list(ar)
lofr = len(result)
finalres=[]
for i in result:
    for j in i[2]:
        finalres.append([j,i[1]])

finalres.reverse()
print("\n\n")

for i in finalres:
    p = i[0]
    item=[ x for x in p]
    if(len(item[0])>0):
        x1, *_ = item[0]
    else:
        x1="Null"
    if(len(item[1])>0):
        x2, *_ = item[1]
    else:
        x2="Null"
    print('\n\n-----')
    print("Association rule: " + x1 + " -> " + x2 )
    print("Confidence: " + str(i[0][2]))
    print("Support: " + str(i[1]))
    print("Lift: " + str(i[0][3]))
    print('-----')
    print()

```

Output:

The screenshot shows two windows. The left window is a Python 3.7.0 Shell with the following output:

```

Given Dataset:
-----
HotDogs | Buns | Ketchup |
HotDogs | Buns |
HotDogs | Coke | Chips |
Chips | Coke |
Chips | Ketchup |
HotDogs | Coke | Chips |

-----

Association rule: Coke -> Chips
Confidence: 1.0
Support: 0.5
Lift: 1.5

-----

Association rule: Chips -> Coke
Confidence: 0.75
Support: 0.5
Lift: 1.5

-----

Association rule: Null -> HotDogs
Confidence: 0.6666666666666666
Support: 0.6666666666666666
Lift: 1.0

```

The right window is a script editor showing the code used to generate this output. The code is as follows:

```

import pandas as pd
from apyori import apriori

dataset = [['HotDogs', 'Buns', 'Ketchup'], ['HotDogs', 'Buns'], ['HotDogs', 'Coke', 'Chips']]
print("Given Dataset:")
print('-----')
for i in dataset:
    for j in i:
        print(j + ' | ', end='')
    print('-----')
ar = apriori(dataset, min_support=0.35, min_confidence=0.6, min_lift=1, max_length=3)
result = list(ar)
lofr = len(result)
finalres=[]
for i in result:
    for j in i[2]:
        finalres.append([j,i[1]])

finalres.reverse()
print("\n\n")

for i in finalres:
    p = i[0]
    item=[ x for x in p]
    if(len(item[0])>0):
        x1, *_ = item[0]
    else:
        x1="Null"
    if(len(item[1])>0):
        x2, *_ = item[1]
    else:
        x2="Null"
    print('\n\n-----')
    print("Association rule: " + x1 + " -> " + x2 )
    print("Confidence: " + str(i[0][2]))
    print("Support: " + str(i[1]))
    print("Lift: " + str(i[0][3]))
    print('-----')
    print()

```