Department of Computer Engineering Exp. No.8

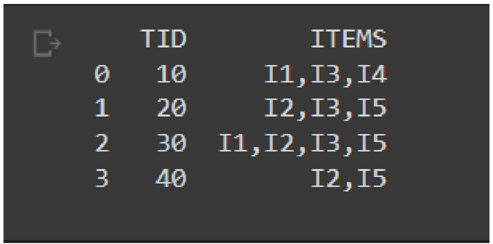
Semester	T.E. Semester VI – Computer Engineering
Subject	Data Warehousing and Mining
Subject Professor In-charge	Prof. Kavita Shirsat
Assisting Teachers	Prof. Kavita Shirsat
Laboratory	M-312B

Student Name	Vibodh Bhosure	
Roll Number	20102A0032	
Grade and Subject Teacher's Signature		

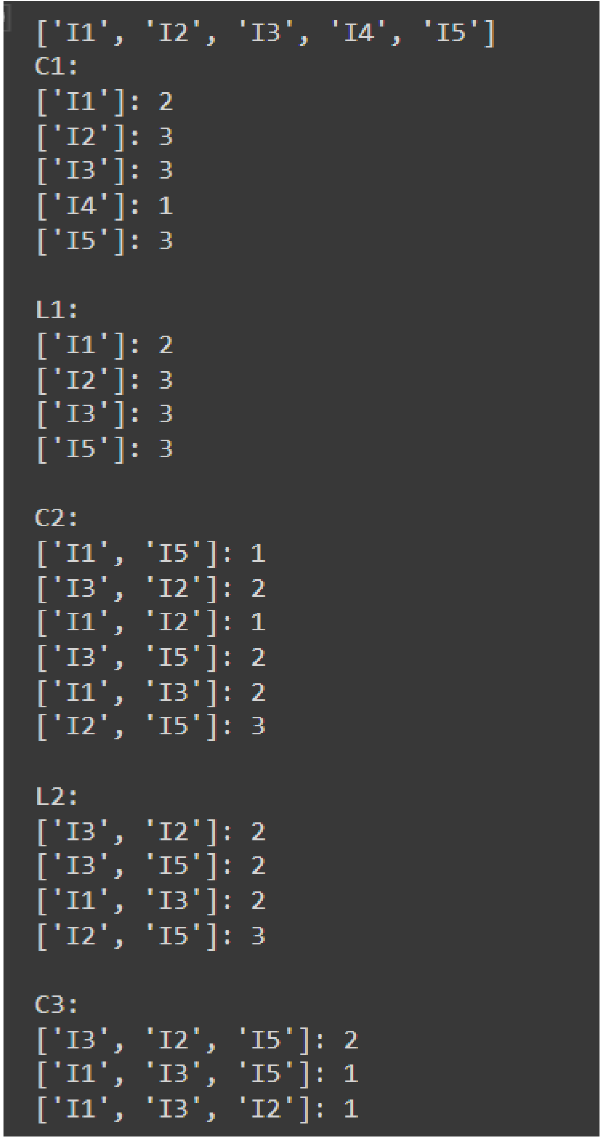
Experiment Number	08	
Experiment Title	Perform Apriori algorithm on a data set to find the relative data set	
Resources / Apparatus Required	Hardware: Computer system	Software: Python – Google Colab
Description :-	Apriori algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule leaning that analyzes that people who bought product A also bought product B. The primary objective of the apriori algorithm is to create the association rule between different objects. The association rule describes how two or more objects are related to one another. Apriori algorithm is also called frequent pattern mining. Generally, you operate the Apriori algorithm on a database that consists of a huge number of transactions. Let's understand the apriori algorithm with the help of an example; suppose you go to Big Bazar and buy different products. It helps the customers buy their products with ease and increases the sales performance of the Big Bazar.	
Implementation :-	# -*- coding: utf-8 -*- """AprioriAlgo.ipynb Automatically generated by Colaboratory. Original file is located at	

```
https://colab.research.google.com/drive/14GjwoxZW2bbNimZsDagVTAC
FI6YVRgXK

Code By - Vibodh Bhosure
"""

from google.colab import files
uploaded = files.upload()

import pandas as pd
import io

df = pd.read_csv(io.BytesIO(uploaded['Apriori.csv']))
print(df)

X = [[i] for i in df['TID'].tolist()]
Y = [i for i in df['ITEMS'].tolist()]
for key, value in enumerate(X):
  X[key].append(Y[key])

for i in X:
  i[1] = i[1].split(",")
data = X

init = []
for i in data:
    for q in i[1]:
        if(q not in init):
            init.append(q)
init = sorted(init)
print(init)

sp = 0.4
s = int(sp*len(init))
s

from collections import Counter

c = Counter()
for i in init:
    for d in data:
        if(i in d[1]):
            c[i]+=1
print("C1:")
for i in c:
    print(str([i])+": "+str(c[i]))
print()
l = Counter()
for i in c:
    if(c[i] >= s):
        l[frozenset([i])]+=c[i]
print("L1:")
for i in l:
    print(str(list(i))+": "+str(l[i]))
print()
pl = l
pos = 1
for count in range (2,1000):
    nc = set()
    temp = list(l)
    for i in range(0,len(temp)):
        for j in range(i+1,len(temp)):
            t = temp[i].union(temp[j])
            if(len(t) == count):
                nc.add(temp[i].union(temp[j]))
    nc = list(nc)
    c = Counter()
    for i in nc:
        c[i] = 0
        for q in data:
            temp = set(q[1])
            if(i.issubset(temp)):
                c[i]+=1
    print("C"+str(count)+":")
    for i in c:
        print(str(list(i))+": "+str(c[i]))
    print()
    l = Counter()
    for i in c:
        if(c[i] >= s):
            l[i]+=c[i]
    print("L"+str(count)+":")
    for i in l:
        print(str(list(i))+": "+str(l[i]))
    print()
    if(len(l) == 0):
        break
    pl = l
    pos = count
print("Result: ")
print("L"+str(pos)+":")
```

```
for i in pl:
    print(str(list(i))+": "+str(pl[i]))
print()
```

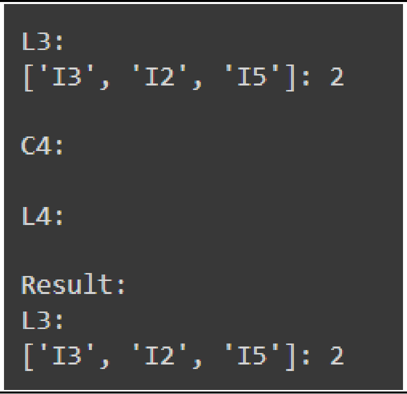


```
     TID        ITEMS
0     10     I1,I3,I4
1     20     I2,I3,I5
2     30  I1,I2,I3,I5
3     40        I2,I5
```



```
['I1', 'I2', 'I3', 'I4', 'I5']
C1:
['I1']: 2
['I2']: 3
['I3']: 3
['I4']: 1
['I5']: 3

L1:
['I1']: 2
['I2']: 3
['I3']: 3
['I5']: 3

C2:
['I1', 'I5']: 1
['I3', 'I2']: 2
['I1', 'I2']: 1
['I3', 'I5']: 2
['I1', 'I3']: 2
['I2', 'I5']: 3

L2:
['I3', 'I2']: 2
['I3', 'I5']: 2
['I1', 'I3']: 2
['I2', 'I5']: 3

C3:
['I3', 'I2', 'I5']: 2
['I1', 'I3', 'I5']: 1
['I1', 'I3', 'I2']: 1
```

<table>
<tr>
<td></td>
<td>



```
L3:
['I3', 'I2', 'I5']: 2

C4:

L4:

Result:
L3:
['I3', 'I2', 'I5']: 2
```

</td>
</tr>
<tr>
<td>Conclusion :-</td>
<td>Here, we have performed apriori algorithm on a dataset and have found a associative data set with a certain frequency.</td>
</tr>
</table>