PESIT Department of Computer Science and Engineering

Course: Data Mining
Semester: 2016 Spring (January – May)
Instructor: BNR (Dr. B. Narsing Rao)

Assignment: 04A
Topic: Association Mining - Apriori
Due by: Midnight on **Tuesday, February 2, 2016**
Method: See below for details; email to bnrao@pes.edu

For this assignment, use the file **supermarket.arff** which you will in the Weka install directory. Also use the Weka GUI.

Task 0 (Preprocessing)

Delete all attributes whose names start with "department". For the tasks below, use the following attributes:

Task 1: All attributes except those with start with "department" and "total (i.e. delete the attribute "total" also.)
Task 2: All attributes except those with start with "department" (i.e. retain "total")

Task 1

Use the Apriori method to generate association rules as follows:

Metric Type = Confidence, Minimum MetricValue = 0.9, Number of Rules = 10

Question 1: Explain the top 5 rules in simple English.

Task 2

Generate, using the Apriori method, rules that will predict the attribute **total** based on other attributes (Hint: Use the **car** option with a minimum confidence of 0.8. Note that this attribute has two values, namely **low** and **high**).

Question 2: Explain the top 5 rules in simple English.
Question 3: In what way are the rules generated by this task useful?

Please submit **three** files in text format:

- Answers to Questions 1, 2, and 3 (Name of file: DM-A04A-USN-Name)
- Result Buffers for tasks 1 and 2 (Weka allows you to save this) (Names of files:DM-04A-RB1-USN-Name, DM-04A-RB2-USN-Name)

The above files must be in a zip archive using the standard naming convention.

Assignment: 04B
Topic: Association Mining – FP Growth
Due by: 1:15 P.M. on **Wednesday, February 3, 2016**
Method: Submit handwritten solution (on stapled A4 sheets) at the beginning of the class
Each sheet should have your name, USN, and sheet number at the top

A database has the following transactions:

TID	Items
1	E, A, D, B
2	D, A, C, E, B
3	C, A, B. E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C

Determine the frequent item sets using the FP Growth algorithm using a minimum support count of 3.

For each iteration, show the following (as applicable):

- Header table
- Conditional FP Base
- Conditional FP Tree
- Generated frequent itemsets