PESIT Department of Computer Science and Engineering

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

Assignment: 05
Topic: Constrained Association Mining – FP Growth
Due by: Midnight on **Tuesday, February 9, 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. First, delete the "total" attribute (as part of the program).

Write a program that will use the Weka FP Growth API to extract and print the rules that satisfy the following requirements:

- Minimum Lift should be 1.2 (Use `Apriori.TAGS_SELECTION` to set this parameter)
- Minimum Support must be 0.3
- Rule must have "vegetables" in the Premise (LHS), but not in the Consequence (RHS)

The rules must be printed in order of decreasing confidence. For each rule, the program must output the following:

- Premise (i.e. LHS)
- Consequence (i.e. RHS)
- Total Support
- Lift
- Confidence

Answer the following questions:

1. What do the measures "Leverage" and "Conviction" mean?
2. How are they calculated in Weka? (Browse the source code for FPGrowth.java. Note that the Weka install includes a jar file which contain the sources)
3. Notice that Weka can print out a string representation of a rule (try it out). Suppose you wanted to change default way in which a rule is printed, which method in which class needs to be modified?

See also: http://facweb.cs.depaul.edu/mobasher/classes/ect584/WEKA/associate.html

Submit a zip archive using the standard naming conventions that contains the following files:

- Program (.java)
- Output (.txt)
- Answers to the above three questions (.pdf)