CS F415: DATA MINING
SECOND SEMESTER 2018-19
Assignment-2

INSTRUCTOR: DR. ARUNA MALAPATI

Submission date and time: 17th Feb, 2019 23:59 hrs

Maximum marks: 30

The goal of this assignment is to generate frequent itemsets and interesting association rules using **Apriori and FP-growth algorithm**. Experiment with different values of support and confidence to generate frequent itemsets and association rules. You must also identify the **maximal frequent itemsets and closed frequent itemsets**. The output must strictly be stored in a file in the following format:

- File1: **Freq_Items_sup:s** (experiment with different s)
  Format: Frequent Itemset (count)
  Example: safety=med, persons=4, class=acc, (90)

- File2: **Assn_Rules_sup:s,conf:c**(experiment with different values of s and c)
  Format: LHS (item set (count)) ---> RHS (item set (count)) - confidence value
  Example: maint=vhigh, persons=4 (144) ----> CLASS=unacc (1210) - conf(0.75)

Datasets:
- **Groceries dataset**
- **Car evaluation dataset**

You are free to you any other dataset provided that it is of comparable size as the datasets provided.

Programming Languages: C, C++, Java, Python

Team Size: 3

Report:

- Name and ID of team members.
- Dataset used.
- Pre-processing done on the data(if any).
- Number of frequent item sets & association rules for different values of support & confidence (at least 3 values).
- The set of maximal frequent itemsets and closed frequent itemsets and the association rules (that have not been generated) which become redundant due to the use of these. (refer to page 354 in textbook)

Submission Files:

- Source code files
- Frequent Item sets & Association Rules for different support & confidence values
- Report in PDF format
- README

Remarks:

- Strictly follow the format provided for the output files. Any other format will result in deduction of marks.
- All submission documents should be zipped together and submitted to CMS through one of the group member's account before deadline. Name of the file should be DM_ASSN1_201x0xxx_201x0xxx_201x0xxx.zip
- All source codes will be checked for plagiarism on Moss (for a Measure of Software Similarity). Any kind of plagiarism will lead to severe penalization.
- You are expected to demo your code and present your results as per the schedule that will be made available on CMS later.

Evaluation:

- Code & comments (15 marks)
- Frequent Itemsets & Association rules (5 Marks)
- Report (5 marks)
- Viva (5 marks)

Please contact the following teaching assistants for any queries:

1. Harshith Thonupunoori(f20150071@hyderabad.bits-pilani.ac.in)