

Report

Data mining – Assignment 1

Group members

Pranav Kumar Asthana : 2015A7PS0961H
Uttara Ravi : 2015A7PS0032H
Sahil Sangwan : 2015A7PS0965H
Utkarsh Grover : 2017A7PS1428H

Dataset used

Groceries dataset:

http://www.sci.csueastbay.edu/~esuess/classes/Statistics_6620/Presentations/ml13/groceries.csv

Results (Actual itemsets and rules are stored in output files according to the specified format)

1. sup:100, conf:50

Number of maximal frequent itemsets = 239
Number of closed frequent itemsets = 326
Total number of frequent itemsets = 326
Total number of association rules = 14
Total number of redundant rules = 0

2. sup:100, conf:30

Number of maximal frequent itemsets = 239
Number of closed frequent itemsets = 326
Total number of frequent itemsets = 326
Total number of association rules = 122
Total number of redundant rules = 0

3. sup:250, conf:20

Number of maximal frequent itemsets = 68
Number of closed frequent itemsets = 86
Total number of frequent itemsets = 86
Total number of association rules = 39
Total number of redundant rules = 0

4. sup:500, conf:30

Number of maximal frequent itemsets = 27
Number of closed frequent itemsets = 31
Total number of frequent itemsets = 31
Total number of association rules = 3
Total number of redundant rules = 0

* Maximal and closed itemsets are also stored in the output files for apriori at /output/

* Output files for apriori are at /output/ and for fp growth are at /output_fp/