

FP Growth Algorithm

Team Members:

1. Sandesh Thakar - 2017A7PS0181H
2. Kunal Verma - 2017A7PS0120H
3. Bhargav N - 2017A7pS1574H

Dataset Used:

groceries.csv

Pre-Processing and Processing:

1. Importing the dataset using CSV reader from pandas library.
2. Assigning threshold support and calculating minimum support count from it.
3. Finding frequency of each item in the transactions.
4. Removing all items which have frequency less than the min. support count.
5. Ordering remaining items in descending order of frequencies.
6. Removing all non-supported items from each transaction and ordering remaining items in the above defined order.
7. Inserting all transactions into the FP tree.
8. Mining for frequent itemsets from the FP tree.
9. Finding maximal frequent itemsets.
10. Finding closed frequent itemsets.

Statistics for different support values:

- 1) Support Threshold = 3% :
 - 1.1. Total frequent itemsets: 63
 - 1.2. Maximal frequent itemsets: 50
 - 1.3. Closed frequent itemsets: 63
- 2) Support Threshold = 5%
 - 1.1. Total frequent itemsets: 31
 - 1.2. Maximal frequent itemsets: 28
 - 1.3. Closed frequent itemsets: 31
- 3) Support Threshold = 10%
 - 1.1. Total frequent itemsets: 8
 - 1.2. Maximal frequent itemsets: 8
 - 1.3. Closed frequent itemsets: 8