**Objective**

We learn how we can mining frequent datasets and preprocess them.

Then used those data and applied algorithms of Apriori and FP-growth.

Based on threshold support and confidence values we tried to find the most associated rules data values.

Then we tried to compare our results with the apriori algorithm and FP-growth algorithm.Therefore getting a better understanding the differences between two algorithms. Moreover, different datasets show different graph values.

Problem Discussion

Find some frequent data problem and Using any programming language you are familiar with,

1.Implement two frequent itemset mining algorithms: (i) Apriori (ii) FP-growth. You have to do the following after implementation:

2.Compare the performance of these two algorithms with at least two datasets.

Data

We collected 2 dataset from [xx] and [xx] data.Each dataset contains different level of transactions and values.

|  |  |  |
| --- | --- | --- |
| Grocery dataset |  | Store Dataset |
| 9836 | Total transaction | 7502 |
| 30 | Total products | 20 |
| citrus fruit,grapes,whole milk,pastry,canned fruit,canned fish,chocolate,specialty bar,hygiene articles,napkins,shopping bags, berries,root vegetables,hard cheese,spread cheese,bottled water,soda,bottled beer and etc | Total products list | shrimp,almonds,avocado,vegetables mix,green grapes,whole weat flour,yams,cottage cheese,energy drink,tomato juice,low fat yogurt,green tea,honey,salad,mineral water,salmon,antioxydant juice,frozen smoothie,spinach,olive oil |

File formats

* Dataminingproject
  + GroceriesResults
    - CONF(0.1)SUP(0.005)
    - CONF(0.05)SUP(0.05)
    - CONF(0.25)SUP(0.05)
  + StoreResults
    - CONF(0.1)SUP(0.0045)
    - CONF(0.05)SUP(0.045)
    - CONF(0.25)SUP(0.05)
  + venv
  + Datasets
    - Groceries.csv
    - store.csv
  + AprioriAlgo.py
  + Fpgrowth.py
  + main.py

Algorithms

Apriori Algorithm

Input: D, a database of transactions; min sup, the minimum support count threshold.

Output: L, frequent itemsets in D.

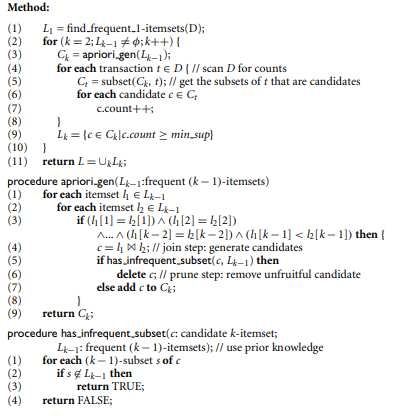


Fig:1 Apriori algorithm

Tasks

|  |  |
| --- | --- |
| Grocery dataset | Store Dataset |
|  |  |
|  |  |
|  |  |