

**East West University**

*Department of Computer Science & Engineering*

|  |  |
| --- | --- |
| Course No. | CSE522 |
| Course Name | Data Mining |
| Assignment No. | 01 |

**Apply Apriori and FP-growth algorithm**

**Submitted To:**

Dr. Hasan Mahmud

Department of CSE,IUT

**Submitted By:**

|  |  |
| --- | --- |
| **Name** | Md Salah Uddin |
| **ID No.** | 160204046 |
| DOS | 11/12/2023 |
|  |  |

**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 Use any programming language you are familiar with,

1.Implement two frequent itemset mining algorithms: (i) Apriori (ii) FP-growth. I 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 kaggle 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 |

Table 01: Dataset values and types

**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 and FP-growths Algorithm

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

Output: L, frequent itemsets in D.

|  |  |  |
| --- | --- | --- |
|  | **Apriori Algorithm(SUP=0.05,CON=0.25)** | **FP-Growth(SUP=0.05,CON=0.25)** |
| Groceries | ASSOCIATION RULES yogurt ( 1372 ) ---> whole milk ( 2513 ) confidence = 0.40160349854227406 other vegetables ( 1903 ) ---> whole milk ( 2513 ) confidence = 0.38675775091960063 whole milk ( 2513 ) ---> other vegetables ( 1903 ) confidence = 0.29287703939514526 rolls/buns ( 1809 ) ---> whole milk ( 2513 ) confidence = 0.30790491984521834 | ASSOCIATION RULES  yogurt ( 1372 ) ---> whole milk ( 2513 ) - conf( 0.40160349854227406 ) other vegetables ( 1903 ) ---> whole milk ( 2513 ) - conf( 0.38675775091960063 ) whole milk ( 2513 ) ---> other vegetables ( 1903 ) - conf( 0.29287703939514526 ) rolls/buns ( 1809 ) ---> whole milk ( 2513 ) - conf( 0.30790491984521834 ) |
| T.runtime | 2 min | 45s |
| StoreItems | ASSOCIATION RULES eggs ( 1348 ) ---> mineral water ( 1788 ) confidence = 0.28338278931750743 mineral water ( 1788 ) ---> spaghetti ( 1306 ) confidence = 0.2505592841163311 spaghetti ( 1306 ) ---> mineral water ( 1788 ) confidence = 0.3430321592649311 chocolate ( 1230 ) ---> mineral water ( 1788 ) confidence = 0.32113821138211385 | ASSOCIATION RULES eggs ( 1348 ) ---> mineral water ( 1788 ) - conf( 0.28338278931750743 ) mineral water ( 1788 ) ---> spaghetti ( 1306 ) - conf( 0.2505592841163311 ) spaghetti ( 1306 ) ---> mineral water ( 1788 ) - conf( 0.3430321592649311 ) chocolate ( 1230 ) ---> mineral water ( 1788 ) - conf( 0.32195121951219513 ) |
| T.runtime | 1.6 min | 40s |

Table 01: Sample output of 2 datasets based on Confidence=0.25 and Support=0.05

From our experiments, we can see that apriori and FP-growth algorithm shows similar results for both datasets but their main difference is in total time transaction completion FP-growth handles data faster than Apriori algorithm.Fig1 shows the time variance of two algorithms.

Fig 1:Total transactions (transactions vs time)

We also conduct more experiments based on different confidence values as 0.1,0.05 and support values 0.050,0.045 .Then compare transaction with time in graphs.All graphs are given below.

**Experimental Tasks**

|  |  |  |
| --- | --- | --- |
| Grocery dataset |  | Store Dataset |
|  | C=0.1 S=0.005 |  |
|  | C=0.05 S=0.05 |  |
|  | C=0.25 S=0.05 |  |

Table 02: Experiments results (Transactions,Runtime)