An Exploratory Concentrate on FP-Growth Calculation for finding frequent Patterns

Purum Sasidhar

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract**—** *Market Basket examination figures out clients' buying designs by finding significant relationship among the items which they place in their shopping bins. It aids dynamic cycle as well as increments deals in numerous business associations. FP Growth is the most well-known calculations for mining successive itemsets. For this calculation predefined least help is expected to fulfill for distinguishing the successive itemsets. Be that as it may, when the base help is low, countless up-and-comer sets will be created which requires enormous calculation. In this paper, Market Bushel Bin Examination with FP-Growth calculation is proposed to decide the design and arranging of products accessibility. The use of FP-Growth calculation ended up being valuable in producing numerous and enlightening affiliation rules to figure out the shopper spending designs. In addition, from the principles of the client affiliation can be portioned independently to meet the particular requirements of clients with savvy by involving a few unique advancements for the general gathering. The exploratory outcomes demonstrate the way that FP-Growth calculation can break down rapidly and productively illuminating shopper shopping designs. That's what the outcomes show in the event that top selling things are utilized, it is feasible to get practically same successive itemsets and affiliation rules inside a brief time frame contrasting and that yields which are determined by processing every one of the things.*

I. INTRODUCTION

As of late, Data Mining has been carried out in different fields, including business and broadcast communications. Information digging is a strategy for removing and distinguishing designs in enormous informational collections that joins AI, measurements, and data set frameworks. One of the main use-cases in information mining is finding the high-recurrence designs between the arrangement of itemset called affiliation rules. Affiliation rule mining is a well-informed strategy for discovering a few relations between factors in huge data sets.

Data mining is a process to obtain potentially useful, previously unknown, and ultimately understandable knowledge from the data. Information disclosure in data sets (KDD) is characterized as the non-paltry extraction of substantial, verifiable, possibly helpful and eventually reasonable data in enormous data sets [1]. For quite a long time, a large number of uses in different spaces have profited from KDD strategies and many works has been directed on this theme. The issue of mining incessant itemsets emerged first as a sub-issue of mining affiliation rules [3]. Association rules mining is one of the important portions of data mining and is used to find the interesting associations or correlation relationships between item sets in mass data [2]. Discovering frequent item sets is a key technology and step in the applications of association rules mining [4].

II. ASSOCIATION RULES MINING

Association rules mining is a function of data mining research domain and arise many researchers interest to design a high efficient algorithm to mine association rules from transaction database [6]. Generally all the frequent item sets discovery from the database in the process of association rule mining shares of larger, the price is also spending more.

2.1 Frequent Item Sets.

Set $I = \{i1, i2,...,in\}$ as an assortment of all kinds of things in the data set, every exchange T is a subset of I, or at least, $T \subseteq I$, and data set D is an assortment of exchanges. For a given exchange data set D, the complete number of exchanges it contains is N. Characterize the help count(X) of thing set $X(X \subseteq I)$ as the quantity of exchanges T in D making $X \subseteq T$ and the help support(X) of thing set X as count $(X)/N$ [9]. The quantity of things in a thing set is called aspect or length of this thing set, on the off chance that the length of the thing set is k, called k-thing set [4][7].

Definition 1: For a given least help, minsup, if the thing set meets support(X) ≥ minsup, thing set X is known as a regular thing set and on the other hand thing set X is called a rare thing set. A set shows relationship between an incessant thing with different things, calling this set a successive thing affiliation set. The base help count, minCount, meets minCount = minsup$*N$. When count(X) ≥ minCount, one says support(X) ≥ minsup [4][5].

Definition 2: At the point when the length of the thing set X is k and support(X) ≥ minsup, one calls thing set Xk-thing regular set. In the event that k≥3, one can call thing set X multi-thing regular set[4][5].

III. FP-GROWTH ALGORITHM

The FP-Growth Algorithm is an elective method to discover continuous itemsets without utilizing applicant ages, accordingly further developing execution. The FP-Growth Algorithm, proposed by Han in [4], is a productive and adaptable technique for mining the total arrangement of successive examples by design piece development, utilizing an all-encompassing prefix-tree structure for putting away compacted and pivotal data about incessant examples named continuous example tree (FP-tree). FP-development calculation is an effective strategy for mining all continuous itemsets without competitor age. FP-development uses a blend of the vertical and even information base design to store the data set in primary memory.

The calculation mines the continuous itemsets by utilizing a gap and-vanquish methodology as follows: FP-development first packs the data set addressing successive itemset into a regular example tree, or FP-tree, which holds the itemset affiliation data also. The subsequent stage is to separate a compacted data set into set of restrictive All hubs relate to things have a counter.

The FP-development calculation comprises of the accompanying advances:

1. Scan DB once, discover incessant 1-itemset (single thing design)
2. Sort continuous things in recurrence diving request, f-list
3. Scan DB once more, develop FP-tree
4. Construct the restrictive FP tree in the succession of opposite request of F - List - produce incessant thing set

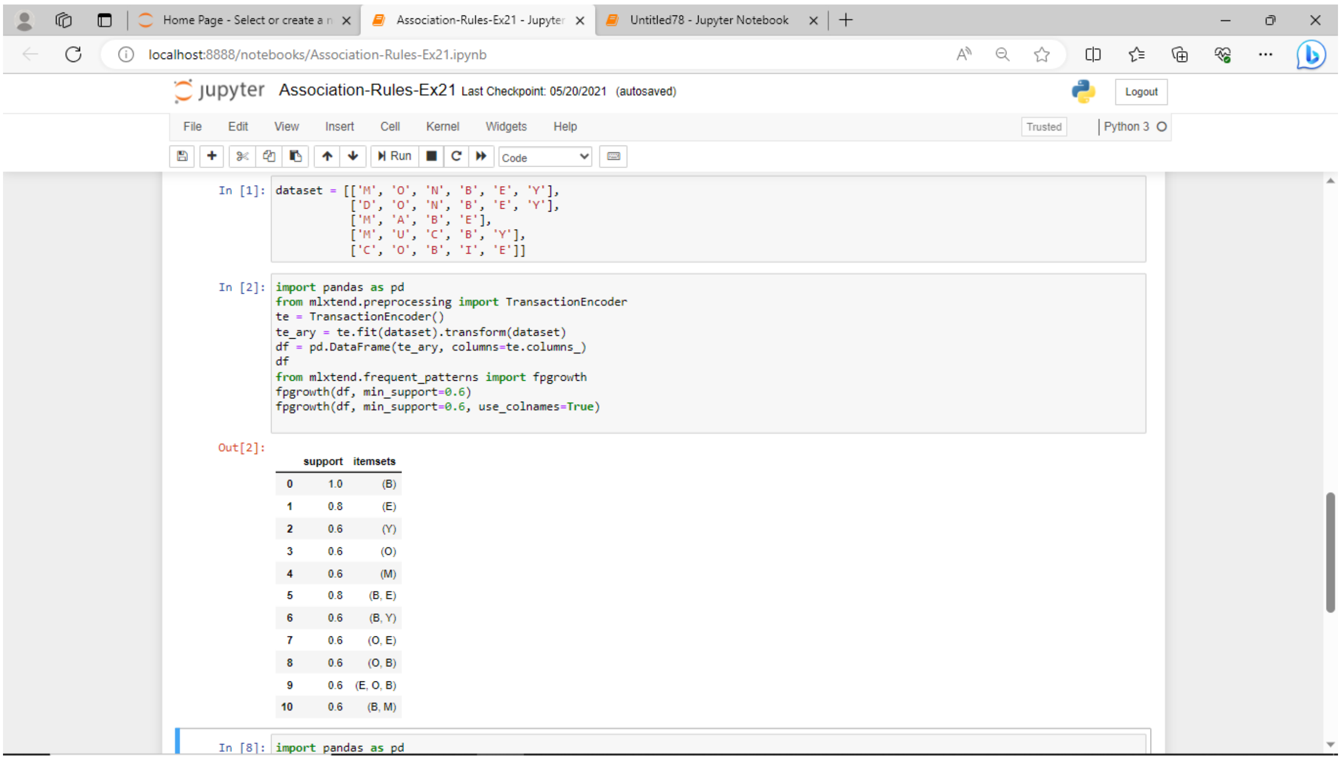
IV. EXPERIMENTAL RESULTS

The experiment was conducted using Python Programming Language. This section comprises the experimental analysis of FP-Growth algorithm for supermarket transaction dataset was used using python programming language. This dataset contains 5 transactions and 10 items. The Supermarket transaction dataset are shown in the table-1.

TABLE 1
TRANSACTION DATA

Transaction-ID	List of items in the transaction
T1	M, O, N, B, E, Y
T2	D, O, N, B, E, Y
T3	M, A, B, E
T4	M, U, C, B, Y
T5	C, O, B, I, E

The experimental results are shown in the figure-1 and figure-2.



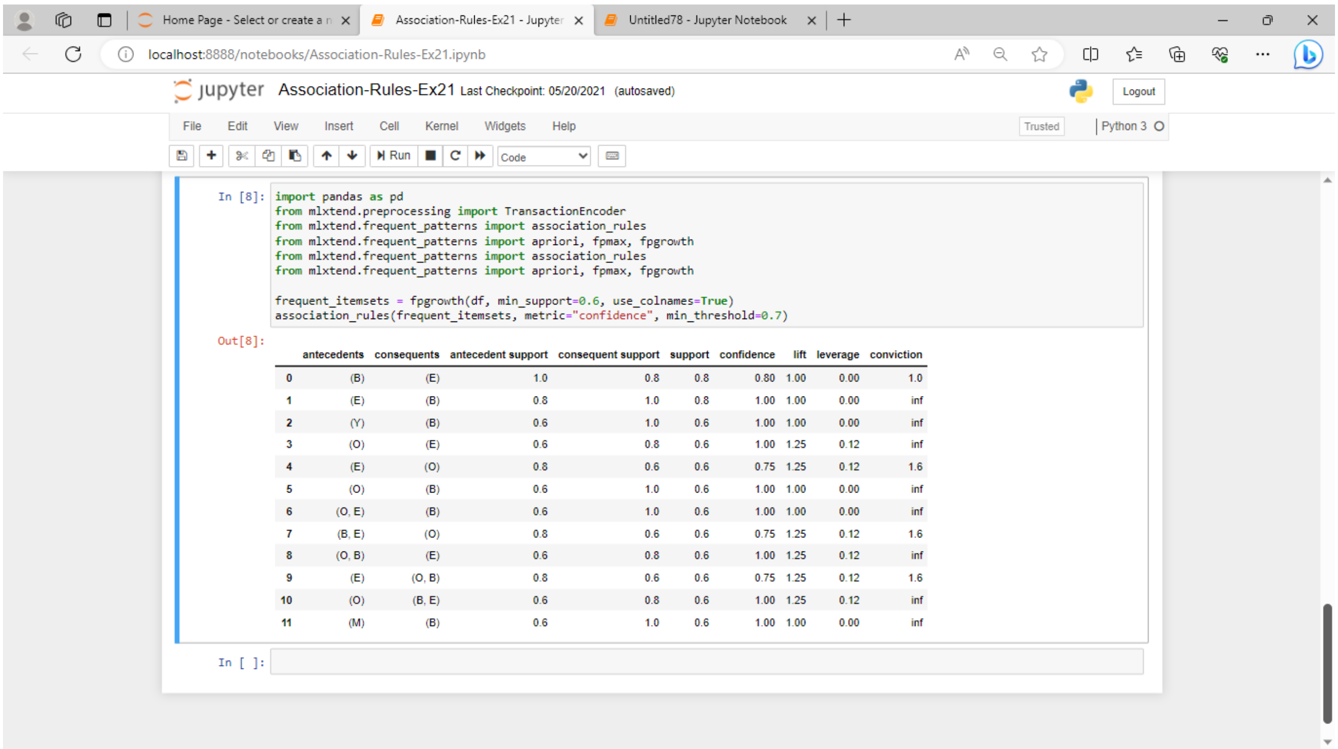
```
In [1]: dataset = [['M', 'O', 'N', 'B', 'E', 'Y'],
                   ['D', 'O', 'N', 'B', 'E', 'Y'],
                   ['M', 'A', 'B', 'E'],
                   ['M', 'U', 'C', 'B', 'Y'],
                   ['C', 'O', 'B', 'I', 'E']]

In [2]: import pandas as pd
        from mlxtend.preprocessing import TransactionEncoder
        te = TransactionEncoder()
        te_ary = te.fit(dataset).transform(dataset)
        df = pd.DataFrame(te_ary, columns=te.columns_)
        df
        from mlxtend.frequent_patterns import fpgrowth
        fpgrowth(df, min_support=0.6)
        fpgrowth(df, min_support=0.6, use_colnames=True)

Out[2]:
```

	support	itemsets
0	1.0	(B)
1	0.8	(E)
2	0.6	(Y)
3	0.6	(O)
4	0.6	(M)
5	0.8	(B, E)
6	0.6	(B, Y)
7	0.6	(O, E)
8	0.6	(O, B)
9	0.6	(E, O, B)
10	0.6	(B, M)

Figure-1: Experimental results of Frequent patterns



```
In [8]: import pandas as pd
        from mlxtend.preprocessing import TransactionEncoder
        from mlxtend.frequent_patterns import association_rules
        from mlxtend.frequent_patterns import apriori, fpmax, fpgrowth
        from mlxtend.frequent_patterns import association_rules
        from mlxtend.frequent_patterns import apriori, fpmax, fpgrowth

        frequent_itemsets = fpgrowth(df, min_support=0.6, use_colnames=True)
        association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)

Out[8]:
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(B)	(E)	1.0	0.8	0.8	0.80	1.00	0.00	1.0
1	(E)	(B)	0.8	1.0	0.8	1.00	1.00	0.00	inf
2	(Y)	(B)	0.6	1.0	0.6	1.00	1.00	0.00	inf
3	(O)	(E)	0.6	0.8	0.6	1.00	1.25	0.12	inf
4	(E)	(O)	0.8	0.6	0.6	0.75	1.25	0.12	1.6
5	(O)	(B)	0.6	1.0	0.6	1.00	1.00	0.00	inf
6	(O, E)	(B)	0.6	1.0	0.6	1.00	1.00	0.00	inf
7	(B, E)	(O)	0.8	0.6	0.6	0.75	1.25	0.12	1.6
8	(O, B)	(E)	0.6	0.8	0.6	1.00	1.25	0.12	inf
9	(E)	(O, B)	0.8	0.6	0.6	0.75	1.25	0.12	1.6
10	(O)	(B, E)	0.6	0.8	0.6	1.00	1.25	0.12	inf
11	(M)	(B)	0.6	1.0	0.6	1.00	1.00	0.00	inf

Figure-2: Experimental results of Association rules

From the experimental results, the frequent itemsets are satisfied minimum support 60% are shown in the table-2.

TABLE 2
FREQUENT ITEMSET

Item-ID	Support
B	100
E	80
Y	60
O	60
M	60
B, E	80
B, Y	60
O, E	60
O, B	60
E, O, B	60
B, M	60

4.1 Results and Discussion

The FP-Growth algorithm was applied to a supermarket dataset with a minimum support of 60% and a minimum confidence of 80%.

The top 10 association rules with the highest confidence values are presented in the table-3.

TABLE 3
ASSOCIATION RULES

Antecedents	Consequents	Support	Confidence
B	E	80	80
E	B	80	100
Y	B	60	100
O	E	60	100
E	O	60	75
O	B	60	100
O, E	B	60	100
B, E	O	60	75
O, B	E	60	100
E	O, B	60	75
O	B, E	60	100
M	B	60	100

1. If a customer purchases O and E together, there is an 100% chance they will also buy B in the same transaction.
2. Customers who buy O and B have an 100% likelihood of purchasing E in the same transaction.
3. If a customer buys O, there is an 100% chance they will also purchase B and E.

By analyzing the generated sets of large itemsets, we discovered several high-confidence association rules that highlight strong relationships between different product categories. For instance, customers who purchased biscuits and vegetables were highly likely to buy bread and cake in the same transaction. Similarly, when the total purchase value was high, customers tended to purchase bread and cake as well. These associations can be leveraged by supermarkets to optimize product placement, design targeted promotions, and enhance cross-selling opportunities.

V. CONCLUSION

In this review, we directed affiliation rule mining utilizing the FP-Growth calculation on a true grocery store dataset with a base help of 60 and a base certainty of 80. The outcomes uncovered fascinating and important examples in client buying conduct, which can have huge ramifications for grocery store the board and promoting systems.

All in all, the affiliation rule mining results acquired in this study have useful ramifications for grocery store the executives, empowering them to go with information driven choices to support deals, further develop consumer loyalty, and advance store tasks. As the retail business keeps on developing, affiliation rule mining methods like FP-Growth will stay important apparatuses for separating significant experiences from huge conditional datasets, prompting more proficient and compelling retail systems.

REFERENCES

[1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)

[2] F. H. AL-Zawaidah, Y. H. Jbara, and A. L. Marwan, “An Improved Algorithm for Mining Association Rules in Large Databases,” Vol. 1, No. 7, 311-316, 2011

[3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.

[4] J.Han and M.Kamber,”Data Mining concepts and Techniques”, the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[5] N.Michael, “Artificial Intelligence – A Guide to Intelligent Systems”, 2nd Edition, Addison Wesley 2005

[6] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” in ACM SIGMOD Record, vol. 22, pp. 207–216, 1993

[7] R. Srikant, “Fast algorithms for mining association rules and sequential patterns,” UNIVERSITY OF WISCONSIN, 1996.

[8] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/ datasets.