# Association Rules Analysis on Frequent Pattern-Growth Method in Predicting Sales of Equipments

[1]Vishal Dholariya　　　　　　　　　　　　[2]Asha Patel
[1]UG Student　　　　　　　　　　　　　　[2]Assistant Professor

[1,2]Department of Computer Engineering
[1,2]Sal Institute of Technology and Engineering Research, Ahmedabad, India

*Abstract:* The word Data mining is a broad branch of science that integrate techniques from all other branches like pattern recognize processing, statistics, databases, machine learning, deep learning and visualization to handle the problem of fetching the information from large amount of databases. FP Growth algorithm is a scalable technique for hiring frequent pattern in database. FP Growth improves Apriori Algorithm to a big extend. Frequent Itemset Mining is possible without candidate generation only "two scan" to the database is needed in FP-growth Algorithm. In this paper we present the frequent-pattern growth Algorithm peculiar method for multitude-targeted mining: finding the count of a given large list of itemsets in large data.

**Keywords**: Data Mining, FP-Growth, FP-Growth Algorithm, Equipment set.

## I. INTRODUCTION:

From past 5-10 years, the economy development is not stable and the level of information enhances unceasingly, so the organizations and agencies have collected the massive business data generally. However, we have these massive data does not mean that we had the rich commercial information. The business organization urgent needs to discover the valuable information and knowledge from the magnanimous data. The most common example of mining a frequent item (data) set is "market basket analysis", through discovering the relation between the different enterprises that the customer puts in "the basket", we can analyze a customer's buying choice and habit [1] [2]. This kind of discovery of connection relation may help the retail enterprises to understand that which products are also purchased by the customer frequently, thus this can help them to develop the better marketing strategy and policy. [5]

We can take one example if one customer purchase mobile as a product then how much is the possibly to purchase charger and hands free with mobile? Or we can say how much is the possibly to purchase earphone or tuffen glass with mobile? So, this is the case when market is based on how much of selling of product? So, every entrepreneur has to analyze this case. Here I am showing one example of analysis of frequently buying products.

Apriori is an association rule mining algorithm to calculate mining of data but in this algorithm, we have to calculate candidate sets of patterns and it scan the data repeatedly so it takes more time to execute and more ever it is expensive [11]. So, to overcome this problem Frequent pattern growth mining algorithm is used. It is an association rule mining algorithm which reduces the time by calculating support and drawing of FP-Growth tree. It avoids time complexity and avoids costly and repeated database scans [12].

**Support:**

This data mining objective measure specifies the probability of occurrence that a particular transaction contains all the considered data items. For instance, there are two items A and B, then their support is count as the probability that both A and B occur in the same transaction of database.  This can be obtained by finding the ratio of having number of transactions containing A and B both and total number of transactions occur.  Thus, for association rule Support (S) is represented as [6]:

$$A \Rightarrow B$$

$$S\{A \Rightarrow B\} = \frac{Number\ of\ transactions\ having\ both\ A\ and\ B}{Total\ number\ of\ transactions} \qquad (1)$$

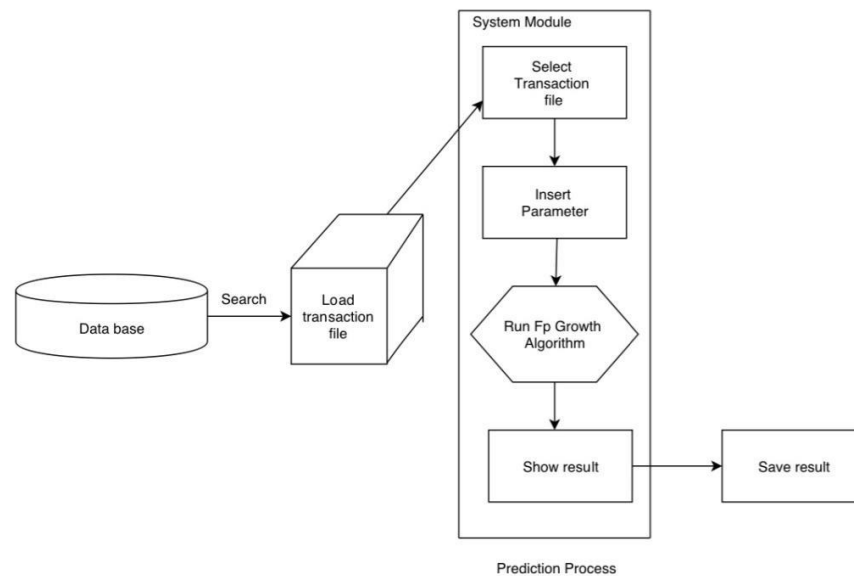**Block Diagram of prediction process of Fp growth algorithm**



Fig. Block Diagram of Prediction process of Fp Growth Algorithm

## II. LITERATURE SURVEY

I study patent on distributed fp-growth with node table for large-scale rule mining which relates to technology for mining data in a database by recursively mining  a conditional frequent pattern for frequent items for each conditional base for each node to obtain frequent patterns. For each branch  in Fp tree ,a single item node table is generated for a selected one of the frequent items. However, frequent items appears in the fp tree and a corresponding frequent items are counted. All this single item NT of each branch generated for the frequent items , the frequent item count of each frequent items is summed to form a combined single item NT , and association rules based on the frequent patterns are generated for each of the frequent items and the combined single-item NT[8]

Songyao Lian Tian an author had published a patent on kind of song recommendations method based on FP-growth. The method conveys the song recommendations method based on Fp Growth that the present invention belongs to data mining. It first establishes a user and listening  the list of songs database on song , user account is combined to transaction set  D as an affairs T with an identifier Tid, corresponding to songs; First time scanning  is carried out to D , calculates and retains frequent episode song ; then second is carried out to D to scan , construct frequent patterns and extract frequent items according to it. However, generating corresponding Strong association rule by each frequent item set, and calculate the confidence level size of each Strong association rule; finally these calculation of the confidence level size of each strong association rule, compared to prior interest of users it shows the songs level wise and arrange songs according to it[9].

 The patent on  Data Mining Algorithms in Frequent Pattern Mining is the technique to abstract the useful data from the large dataset for different perspectives. Frequent pattern mining has become an important data mining technique to find the frequent patterns from the data set that appears frequently. However, frequent Pattern Technique is widely used in financial, retail, telecommunication and many more. The major concern of these industries is that it has a faster processing of a very large amount of data. Various techniques and algorithms have been proposed for this purpose. Apriori, FP-tree are the pioneer techniques among them. It have been analysed algorithms for finding frequent patterns with the purpose of discovering how these algorithms can be used to obtain frequent patterns over large transactional databases with most of the efficient way in various aspects[4].

 The research paper FP-Growth Method in Predicting Sales is based on association Rules Analysis.Nowadays, sales transaction data has been increasing on  a company day by day. Large amounts of data can be problematic for a company if it is not managed properly. Data mining is a field of science that unifies techniques from machine learning, pattern processing, statistics, databases, and visualization to handle the problem of retrieving information from large databases. The algorithm included in association rules in data mining is the Frequent Pattern Growth (FP-Growth) algorithm is one of the alternatives that can be used to determine the most frequent itemset in a data set. Fp Growth method conveys the information implemented by using a sales database to find trends in pattern itemsets combination so it can be used as valuable information in making decisions to prepare the stock of goods and marketing strategies required later. The application of the FP-Growth algorithm for marketing technique has a very efficient ease and it can accelerate the process of sales item[3].

The paper on Fake Product review monitoring and removal for genuine online review Journal Network Communications and Emerging Technologies explains information .In today's world reviews on online websites play a vital role in sales of the product so people try to get all the pros and cons of any product before buying the product as there are many options for the same product in the market and  there might be some difference in the procedure that is taken while buying the product so the reviews are directly related to the sales of the product and thus it necessary for the online websites to spot fake reviews as it's their own reputation that comes into consideration as well, so a Fake Review Detection is used to spot any fraudulent which is going on because it's not possible to verify every product and sales manually in the market, so a program comes into the picture that tries to detect any pattern in the reviews given by the customers[10].

## III. MINING FREQUENT PATTERN USING FP-GROWTH ALGORITHM [4]

*Input*: A database DB, represented by FP-tree constructed according to input of Equipment set.

*Output*: The complete set of frequent patterns.

*Method*: call FP-growth (FP-tree, null).

Procedure: FP-growth (Tree, a) {
1. If Tree contains a single prefix path then { // Mining single prefix-path FP-tree
2. Let A be the single prefix-path part of Tree;
3. Let B be the multipath part with the top branching node replaced by a null root;
4. For each combination (denoted as ß) of the nodes in the path A do
5. Generate pattern ß ∪ a with support = minimum support of nodes in ß;
6. Let freq pattern set(A) be the set of patterns as generated;
}7. Else let B be Tree;
8. For each item ai in B do {// Mining multipath part in FP-tree
9. Generate pattern ß = ai ∪ a with support = ai. support;
10. Construct ß's conditional pattern-base and then ß's conditional FP-Tree ß;
11. If Tree ß ≠ Ø then
12. Call FP-growth (Tree ß, ß);
13. Let freq pattern set(B) be the set of patterns as generated;
}
14. return (freq pattern set(A) ∪ freq pattern set(B) ∪ (freq pattern set(A) × freq pattern set(B)))
}

## IV. RESULT & ANALYSIS:

 Here the FP-Growth algorithm method  will help us to solve the problem of equipment sales analysis which will be used to know marketing and promotion strategy equipment set in improving the quality of sales. In this part will discuss the analysis of sales date of equipment that affect the sales ratio of equipments by using FP growth algorithm.

| Sr No | Equipment Name | Initial |
|-------|----------------|---------|
| 1 | Mobile | Mb |
| 2 | Charger | CH |
| 3 | OTG | OTG |
| 4 | Pen Drive | PD |
| 5 | Earphone | Ep |
| 6 | Cover | CV |
| 7 | Screen Guard | SG |
| 8 | TuffenGlass | TG |

Table 1 Equipment Name

| T id | Equipment's set | Ordered Equipment sets |
|------|-----------------|------------------------|
| 1 | Mb,Ch,Cv,SG | Mb,Ch,SG |
| 2 | Mb, OTG | Mb,OTG |
| 3 | Mb,PD | Mb |
| 4 | PD,Mb | Mb |
| 5 | TG,Ep, OTG | OTG,TG |
| 6 | Ch, OTG | Ch,OTG |
| 7 | Mb,SG | Mb, SG |
| 8 | Mb,SG,Ch,TG | Mb,Ch,SG,TG |
| 9 | SG, TG | SG,TG |
| 10 | CV,Ep | √ |

Table 2 Equipment Sets

If we are consider support as a 3 then equipment set are as shown below

| Sr No | Equipment Name | No of item occurred |
|-------|----------------|---------------------|
| 1 | Mb | 6 |
| 2 | CH | 3 |
| 3 | CV | 2 |
| 4 | SG | 4 |
| 5 | OTG | 3 |
| 6 | PD | 2 |
| 7 | TG | 3 |
| 8 | Ep | 2 |

Table 3 Sales rate

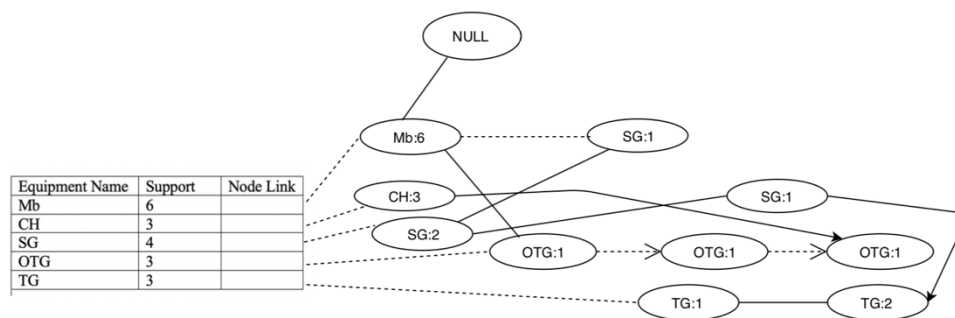| r No | Equipment | Support | Conditional Pattern Base | Conditional Fp Tree |
|------|-----------|---------|--------------------------|---------------------|
| 1 | TG | 3 | {SG:1},{Mb,CH,SG:1},{OTG:1} | [Mb:1] |
| 2 | OTG | 3 | { Mb :1},{TG:1},{CH:1} | [Mb:1] |
| 3 | CH | 3 | {OTG:1},{Mb,CH,SG:2} | [Mb:2] |
| 4 | SG | 3 | {Mb,CH:2},{Mb:1},{TG:1} | [Mb:3] |
| 5 | Mb | 6 | {CH,SG:2},{ OTG:1},{SG:1},{CH,SG,TG:1} | √ |

Table 4 Frequent Equipment set



Fig: Fp Growth tree for Equipment

## V. CONCLUSION & FUTURE WORK

As a result of this process we can conclude following parameters

1. It consumes less time

2. It contains less complexity

Due to these parameters a user can search about the product which is high in demand. Based on the testing done in the previous stage, it can be concluded that the data mining can be implemented by a sales database to find trends in pattern itemsets combination. However, it can be used as a valuable information in making decisions to prepare the stock of equipment's and marketing strategies is required later. The comparison shows that all the approaches are efficient in reducing mining time to some extent.This efficient application of fp growth algorithm enhance the process  and also further it accelerate the process of formation of the trends of equipment's.

More focused research can have a possibility  of getting accurate results about equipment's which can be utilised for the planning and forecasting. From the growth of the algorithm results it can predict the mobile models and it might arrange the sales according to it.

**REFERENCES**:

1. Lior Shabtay , Rami Yaari  and Itai Dattner| A Guided FP-growth algorithm for multitude-targeted mining of big data|arxiv:1803.06632v2

2. Huang-Cheng Kuo and Pei-Cheng Tsai| Mining Time-delayed Gene Regulation Patterns from Gene Expression Data| GSTF Journal on Computing (JoC) Vol.2 No.1

3.Supiyandi, Mohammad Iswan Perangin-angin, Andre Hasudungan Lubis, Ali Ikhwan, Mesran5Andysah Putera Utama Siahaan "Association Rules Analysis on FP-Growth Method in Predicting Sales" DOI: 10.23883/IJRTER.2017.3453.DHCOA

4.https://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R/Frequent_Pattern_Mining/The_FP-Growth_Algorithm   [last  visited 15 Feb. 2019 12: oo pm ]

5. Pamba Pravallika, K. Narendra "Analysis on Medical Data sets using Apriori Algorithm Based 2018 IJSRSET | Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN: 2394-4099 Themed Section: Engineering and Technology.

6. Asha Patel "A Survey paper on Efficient FP-growth Tree for large Data Set"
IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 05, 2016 | ISSN (online): 2321-0613.

7.Draw.io [last visited 15 feb-2019 2:00pm]

8.KaiYang,TaoQuan,YanjiaSun|https://patentimages.storage.googleapis.com/1b/57/0d/d9f6f93bc72f1e/US20180107695A1 |distributed fp-growth with node table for large-scale association rule mining

9. Songyao Lian Tian Yu Jie Wu Xu Wenlin Wang Huidong Longhua|A kind of song recommendations method based on FP-growth|http://patents.google.com/patent/CN109299313A/en?q=Fp&q+Growth&q=tree&oq=FpGrowth

10. Carloine El Fiorenza, Aditya Singh Kashyap, Kartikey Chauhan, Kishan Mokaria, Aashutosh Chandra | Fake Product review monitoring and removal for genuine online review Journal Network Communications and Emerging Technologies (JNCET) Volume Issue 4, April (2018).

11. Pamba Pravallika, K. Narendra| Analysis on Medical Data sets using Apriori Algorithm Based on Association Rules| Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN: 2394-4099

12. Chunhua Fu, Xiaojing Wang, Lijun Zhang and Liying Qiao| Mining Algorithm for Association Rules in Big Data Based on Hadoop| AIP Conf. Proc. 1955, 040035-1–040035-6; https://doi.org/10.1063/1.5033699|Published: 978-0-7354-1654-3