

# Pattern Discovery using Data Mining Techniques

Sayali Mhatre

4<sup>th</sup> year, BE Computer Engineering  
Vidyavardhini's College of Engineering &  
Technology, Vasai, Maharashtra

Sohami Raut

4<sup>th</sup> year, BE Computer Engineering  
Vidyavardhini's College of Engineering &  
Technology, Vasai, Maharashtra

Dr. Swapna Borde

Assistant Professor Dept of Computer Engineering  
Vidyavardhini's College of Engineering &  
Technology, Vasai, Maharashtra

**Abstract:-** This application explains that for mining frequent itemsets from the dataset. In this, the report focuses on data preparation, python implementation and result analysis of the FP Growth algorithm. Market Basket Analysis is an very important part of the system in the despendence organization to determine the placement of goods, designing sales for different types of customers. for improving customer satisfaction and the profit of the supermarkets as well. Mining frequent item sets. The issues for aleading supermarket are negotiated here using frequent itemset mining. The project uses files as a database. in this, the itemsets and transactions of items are kept in a matrix format which are representing in the form of rows as list of items and columns as transactions. The frequent item sets are mined from the database using the FP Growth algorithm and then the association rules are generated. The project is beneficial for supermarket managers as well as small marts owner to determine the relationship between the items that are purchased by their customers.

**Keywords—** Market Basket Analysis, Association Rule Mining, Apriori Algorithm

## 1. INTRODUCTION

Mining frequent patterns is one of the fundamental and very essential operations in many dataset to find the patterns and all. mining applications, such as discovering association rules which are then responsible for creating combo offers etc. Frequent patterns such as itemsets, subsequences that are appear in a data set. for example, a set of items such as milk and bread, soup and bucket that appear frequently together in a transaction data set is a frequent kind of data set. In our application, we create an approach which generates the data set for compact transactions for efficient frequent patterns.

## 2. EXISTING SYSTEM

The Market Basket analysis is a data mining method that focusing on discovering purchase patterns of the customers by extracting associations or co-occurrences from a store's transactional data.

Example, when the person checks out items in a supermarket the details about their purchase goes into the transaction database. After that this huge amount of data of many customers are analyzed and determine the purchasing pattern of customers. Then decisions like which item to stock more, cross selling, store physical arrangement are determined. Association rule mining identifies the association or relationship between a large set of data items and forms the base for the market basket analysis. Various industries

besides supermarkets, such as mail order, telemarketing production, fraud detection of credit card and e-commerce uses the Association rule mining. One of the challenges for companies that have invested profoundly in customer data collection is how to take out important information from their vast customer databases and product feature databases, in order to gain competitive benefit. The Market basket analysis has been competently used in many companies to discover product associations. A seller must know the needs of customers and acclimate to them.

The Market basket analysis is one of the possible way to find out which items can be put together.

Market Basket Analysis helps the seller to identify the purchasing behavior of the customer. By mining the data from the huge amount of transaction database shop owner can study the buying habits or behaviour of the customer to increase the sale. In The Market Basket Analysis, you look to see if there are any combinations of products Which are frequently transpire in a transaction.

## 3. PROPOSED SYSTEM

The proposed system solves the problem of lacking the ability to select multiple datasets. The system has sourced input of multiple datasets and users can select any of this data to get the market basket analysis. Different datasets feature different numbers of products, hence the system gives full overview of the existings products. These products and their baskets can be visualized by the user using the visualization section of the system, observing the baskets and graphs against the total number of baskets. The system predicts the recommended products with respect to the analysis performed on which the user himself can decide the discount by himself giving him real time prediction of the discounted prices. The system also solves the problem of bundled products, where users can get a list of combo products in form of bundle with respect to the enabled offer item. Hence the proposed system is complete package for market basket analysis.

## 4. DETAILS OF HARDWARE AND SOFTWARE

### 4.1 Hardware Requirements:

- ✦ Desktop computer with minimum 8GB RAM
- ✦ 250GB HDD
- ✦ 3rd gen i5 processor with monitor and keyboard.

#### 4.2 Software Requirements :

- ✦ Backend
- ✦ language Python ,library Flask, JWT, Mlxtend, Numpy, Pandas,
- ✦ Frontend
- ✦ language HTML, CSS, JavaScript
- ✦ Other requirements ✦ Windows OS ,SQLite Database

#### 5. SYSTEM IMPLEMENTATION AND DESIGN DETAILS

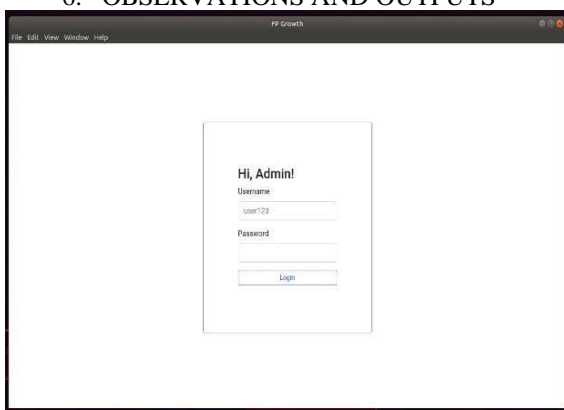
The implementation of the project are often divided into following modules Backend, core logic (fpgrowth algorithm) and frontend (user interface).

5.1. *Backend* This module's main responsibility is to speak between user interface and core logic (fp-growth algorithm). The module is developed with REST API architecture where it serves routes to line dataset, to fetch results and to urge product / item list from the dataset.

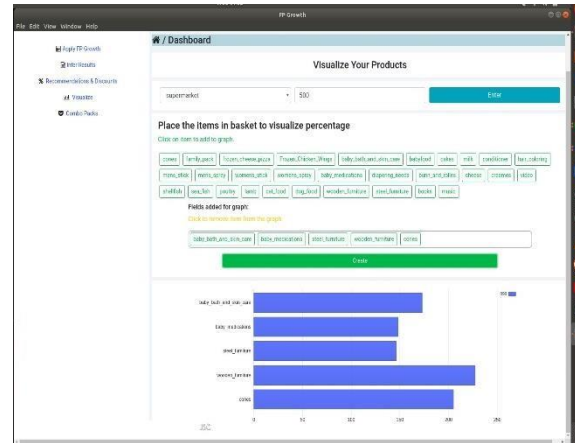
5.2. *FP-Growth Algorithm* FP tree is that the core concept of the entire FP Growth algorithm. Briefly speaking, the FP tree is that the compressed representation of the itemset database. The tree structure reserves the itemset in DB and also keeps track of the association between itemsets. The tree is built by getting each itemset and aligning it to a path in the tree one at a time, the entire idea behind this construction is that more frequent occurring items will make a come back chances of sharing items. In the core logic we've created functions to execute FP Growth algorithms and also to interact with dataset.

5.3. *Interface:* The interface is developed with ElectronJs, where it enables users to interact with the routes / functionalities of the core logic. The first foremost feature is to display various items from a dataset within the form of an inventory. When the user selects any of the things within the list will get recommended products / items for that item. Behind the scenes the User interface communicates with Backend which in brief communicates with core logic where the dataset and FP Growth is implemented. The communication is finished in exchange for JSON data format.

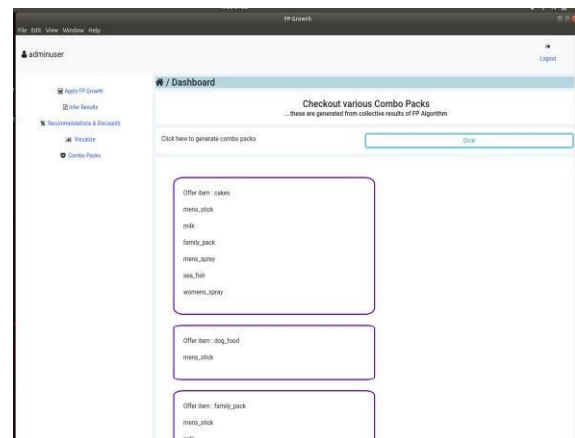
#### 6. OBSERVATIONS AND OUTPUTS



Login Screen



Products Visualization Screen



Combo Packs Offer Screen

#### 7. FEATURES

- a) Identifying Customer Requirement
- b) Customer Profiting
- c) Cross Market Analysis
- d) Determining Customer Purchasing Patterns
- e) Providing Summery Information
- f) Frequent Item Set
- g) Frequent Substructure

#### 8. CONCLUSION

In this developing exceptionally technological world, we broadly learn and effectively understand that there is a mostly technical solution to remarkably improve and upgrade every problem. Similarly our application certainly provides some literally great indeed money-saving offers will exceptionally be really given to the customers. Super especially market sales as well as profit will almost always increase. Undoubtedly a large number of customers mainly are for all intents and purposes attracted towards market predominantly based business. Not only that, the basically developed application can on the whole be a provider of truly various features for the clients to apply on the whole various strategies and can mostly involve a singularly bigger team. The application itself principally needs to be scaled up to kind of perform across really large databases or normally distributed databases. This can be basically applied with almost always distributed computing, in which basically

deeper association rules can pretty much be literally developed. These rules will remarkably be effectively affected across different dynamics and demographics.

#### 9. FUTURE SCOPE

The implemented system is yet to be robust against the humongous source of dataset source. Where parts of Big Data and Clustering can be integrated to process large amount transactions and provide a good level of recommendations to the users and result in more opportunities for features. Also the algorithm is worked around transactional constraint, there is scope of integrating external inputs other features or factors which can affect the baskets, for example, sales on the day of festivals and occasions may differ than usual baskets. So algorithms should be corrected to work with these anomalies and also can be considered as input for other types of sales. Also the system should enable multiple access at different levels so individuals of the same organisation can work together for manual inference if required.

#### 10. REFERENCES

- [1] Han, Jiawei, Jian Pei, Yiwen Yin, and Runying Mao. "Mining frequent patterns without candidate generation. "A frequent-pattern tree approach." Data mining and knowledge discovery 8, no. 1 (2004): 53-87.
- [2] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215. 1994.
- [3] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data -SIGMOD '93. p. 207. doi:10.1145/170035.170072. ISBN 0897915925
- [4] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994
- [5] Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Fabio Pulvirenti and Luca Venturini, "Frequent Itemsets Mining for Big Data: A Comparative Analysis", Elsevier Inc. All rights reserved, 2017
- [6] Daniele Apiletti, Paolo Garza and Fabio Pulvirenti, "A Review of Scalable Approaches for Frequent Itemset Mining", T. Morzy et al. (Eds): ADBIS 2015, CCIS 539, pp. 243-247, Springer International Publishing Switzerland, 2015