International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Study on Market Basket Analysis with Apriori Algorithm Approach

Himanshu Singh[1], Nikhil Shelke[2], Aniket Bavaskar[3], Shradha Nikam[4], Prof. Pradip Shewale[5], Prof. Deepa Mahajan[6]

[1-6]*Dr.D.Y.Patil Institute of Technology, Pimpri, Pune*

Abstract - *Since the introduction of electronic sale , retailers have had at into their inventory a vast amount of data. The challenge has been how to utilize that data to produce business inference. Most retailers have already figured out a way to understand the basics of the business: what are they selling, how many units are moving and the sales amount. However, few have deployed enough model to analyze the information at lowest level of granularity: the market transaction. The main reason for this is, perhaps, the pre notion that looking at data at this level of granularity is very much expensive and has limited business authority. This article will explore value of market basket analysis through real scenarios, outlining along the way why the users don't need a strong statistics background to understand it and benefit from it.*

Market basket analysis , is the process of analyzing transaction-level data to drive business value. At this level, the information is very useful as it provides the business users with direct visibility into the market of each of the customers who shopped at their store. The data becomes a gateway into the events as they happened, understanding not only the quantity of the items that were purchased in that particular basket.

***Key Words*: Graphology, Market Basket Analysis, Machine Learning, Apriori Algorithm, Jupiter Notebook**

1. INTRODUCTION

Apriori is an concept for frequent item set association rule learning over relational databases. It proceeds by processing the frequent individual items in the database and extending them to larger item sets as long as those item sets appear frequently in database. The frequent item determined by Apriori can be used to determine association rules which determine general phenomenon in the database given or loaded: this has applications in fields such as market basket analysis.

The Apriori algorithm was introduced by Agrawal and Srikant in 1994. Apriori is designed to operate on databases which contains transactions (for example, lists of items bought by customers, or details of a website frequently visited or IP addresses searches regularly). Other algorithms are dedicated for designing and determining association rules in data having no transactions (or having no timestamps) .Each transaction is seen as set of items (an itemset).

Apriori uses an approach called bottom up approach, where frequent subsets are extended one item at one time known as candidate generation, and other groups of candidates are tested against the data. The algorithm terminates the code when no further successful extensions found in the dataset.

2. LITERATURE SURVEY

[1]. Paper Name: Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data

Authors: Fachrul Kurniawan, Binti Umayah , Jihad Hammad , Supeno Mardi Susiki Nugroho , Mochammad Hariadi

Description: In this survey the author analyze Consumer behavior in deciding to purchase, use, as well as consume the purchased materials and services including in the consumer factors which can give a increase to the decisions of whether to buy and use products. Every customer defined by their different needs as well as has different behaviors in fulfilling those things. However, in the process of different behaviors to fulfill their needs, they commonly share some similarities, one of them in many is to desiring to maximize their satisfaction level in consuming a necessary item or services. Of that purchasing behavior, that can be inferred as to the pattern, or habit that the customers do to fulfill their needs and desires. In these recent years and also in coming many years, transaction data have been generally used as research and analysis objects for research and students. In the study created by author, also, transaction data are to be re-processed/re-explored to generate more valuable and calculable information. For example, information of an item whose sales is the lowest or highest and also combination of product. Besides, information can be make use of in regard with the stock summation of that product. Moreover, from settlement data there can be make use of as to the relation of each acquire item inside the customer' basket. By that details, we can make use of it for effective product

display/assortment to attract interest. The commonly-used application to examine transaction data customers' shopping basket is market basket analysis.

[2]. Paper Name: Market Basket Analysis with Data Mining methods

Authors: Andrej Trnka

Description: In this paper the author describes the way of Market Basket Analysis implementation by using the Six Sigma methodology in introduction with data mining for large group of data. Data mining methods come up with a lot of chances in the market neighborhood. Basket Market Analysis is one of them. Six Sigma methodologies uses several statistical methods. With execution of Market Basket Analysis to Six Sigma we can increase the credibility's of the results and change the Sigma performance level of the process. In our research we used GRI (General Rule Induction) algorithm to create the association rules in between the products in the market purchased item. These associations between different products displays a variety between the described products. To implement and getting a view of the dependence between the products and buying capacities we used Web plotting. The last used in the algorithm for analysis was C5.0. This algorithm here was used to implement and build rule-based profiles.

[3]. Paper Name: Market Basket Analysis: Identify the changing trends of market data using association rule mining

Authors: Manpreet Kaura, Shivani Kang

Description: Today, the large amount of data that are being processed is being maintained in the databases which can be includes by various fields like retail markets, banking sector, medical field etc. But it is not necessary that the whole information is useful for the user. That is why, it is very important to get and implement further from the useful information that are being provided by large amount of data. This process of implementing and extracting useful data is known as data mining or A Knowledge Discovery and Data (KDD) process. The overall process of getting and interpreting patterns from data generally involves many steps such as selection, preprocessing, transformation, data mining and interpretation. Data mining helps in the business of many sectors for marketing. The work of getting a hold of market basket analysis in management research has been implemented by Aguinis et al. Market basket analysis generally also known as association rule mining. It helps the marketing enthusiast to understand the current trends of customers e.g. which products are being selected together by the customer. There are various algorithms and techniques that are available to perform data mining

3. EXISTING SYSTEM

The existing system involving Association Rules are implemented by getting hold of some calculation involving calculus or different programming languages like python, R etc. Today you have to calculate different parameters involving hard and error prone calculations over a large amount of data basically for a big super market which can consume more computational and management time. Cases involving major changes and more number of items to select with the above calculations involves more computation.

3.1 Disadvantages

- More computation power
- Less Accurate
- More investment on resources

4. PROPOSED SYSTEM

In order to shorten the work load of the analysis environment, we proposed the apriori algorithm that can be widely used as a supplement of the hectic and time consuming mathematics involved in the analysis. It can be done in both python and R but we prefer doing it by using Python language.

In this we first take the billing process of the counter where all the billings of the super market happening, then we convert it into software readable file like csv, tsv etc. then we upload it into real time cloud environment like IBM Databases or Google Cloud. After uploading it into cloud we can access it from anywhere in the world (needs internet connections though). We can use Jupiter notebook or any other script writer like PyCharm to connect our analysis code from cloud and we can use SQL to query into our databases created inside cloud. After getting data from the cloud we starts our analysis involving many steps like pre-processing, applying algorithm, Validating Data, testing according to the percentage of customers buying from it with one or more combination of items.

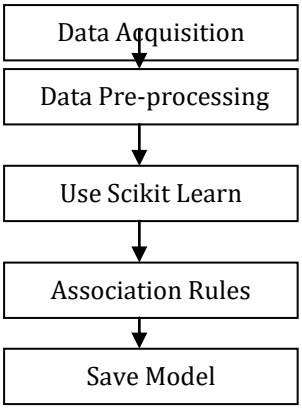

Fig. 1 Proposed Framework

4.1 Advantages

- Less Resources utilization due to more computations is being done by libraries.
- More flexibility
- Higher accuracy since done by pre-compiled library
- Cost effective architecture involving less time to compute.

5. CONCLUSIONS AND FUTURE SCOPE

We are implementing the application in which, the input will be the Market receipt to the application, and the that will be forwarded to the system for pre-processing. The dataset variables from the document are broken up into the predefined features. This process is continuous which is determining the relative positions of these features and comparing them with the database of feature-graphs goes on until a match is obtained. The output will be the predicted behaviour at market of that person.

MATKET BASKET ANALYSIS using APRIORI is a project for Grocers and retail market owners where people has to choose what they want and put that into the cart they are carrying just like shopping mart where customers has to move into stall per stall to pick up the things.This project help them to place analysed item side by side so that if customer picked itemA then chances of picking itemB will increase by 30% that means increasing sales by 30 – 40%, which is good for their growth as well as profit. For future using this project a shop owner can placed some items close together in future for their customers to pick more than one items whether they were previous only going to buy a single item.

REFERENCES

[1] Tan, Pang-Ning; Kumar, Vipin; Srivastava, Jaideep (2004). "Selecting the right objective measure for association analysis". Information Systems. 29 : 293–313

[2] Tan, Pang-Ning; Michael, Steinbach; Kumar, Vipin (2005). "Chapter 6. Association Analysis: Basic Concepts and Algorithms" (PDF).

[3] Zaki, Mohammed J. (2001); SPADE: An Efficient Algorithm for Mining Frequent Sequences, Machine Learning Journal, 42, pp. 31–60.

[4] 2-Aurélien-Géron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O'Reilly-Media-2019.

[5] Gionis, Aristides; Mannila, Heikki; Mielikäinen, Taneli; Tsaparas, Panayiotis (2007). "Assessing data mining results via swap randomization". ACM Transactions on Knowledge Discovery from Data.

[6] Menzies, T.; Ying Hu (2003). "Computing practices - Data mining for very busy people". Computer

[7] Liu, Jinze; Paulsen, Susan; Sun, Xing; Wang, Wei; Nobel, Andrew; Prins, Jan (2006). "Mining Approximate Frequent Itemsets in the Presence of Noise: Algorithm and Analysis". Proceedings of the 2006 SIAM International Conference on Data Mining. pp. 407–418.

[8] Zaki, Mohammed J. (2001); SPADE: An Efficient Algorithm for Mining Frequent Sequences, Machine Learning Journal, 42, pp. 31–60

[9] Zimek, Arthur; Assent, Ira; Vreeken, Jilles (2014). Frequent Pattern Mining. pp. 403–423.

[10] King, R. D.; Srinivasan, A.; Dehaspe, L. (Feb 2001). "Warmr: a data mining tool for chemical data". J Comput Aided Mol Des. 15 (2): 173–81.