

Exploring Association Rule Mining with the Apriori Algorithm: A Case Study on Online Store Data

H S Bharath

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

Abstract— Association rule mining is a powerful technique widely used in data mining and machine learning for discovering interesting patterns and relationships in large datasets. Established in market crate examination, there are an extraordinary number of methods created for affiliation rule mining. They incorporate incessant example revelation, intriguing quality, complex affiliations, and various information source mining. In this research paper, we apply the Apriori algorithm to analyze transaction data from an online store. Our study focuses on uncovering associations between products purchased by customers, shedding light on potential buying patterns and preferences.

I. INTRODUCTION

With the quick improvement of the Web, the present social data cooperation is all the more intently, client experience is all the more top notch, individuals' way of behaving and exercises can be digitized and mimicked, so a wide range of information are growing quickly. Huge information assumes a vital part in big business direction, business association and business process. The advancement of Data Innovation has created a lot of data sets and enormous information in different regions [1]. The exploration in data sets and data innovation has led to a way to deal with store and control this valuable information for additional direction. Information mining coming from the quick development of data, is the cycle to extricate, recognize and track down the possibly valuable and at last reasonable information from the information. Information mining innovation is to recognize examples of information in the information previously existed, to assist clients with grasping the current data and anticipate for what's to come conditions based on existing data. Information mining is a course of extraction of helpful data and examples from immense information. It is additionally called as information revelation process, information mining from information, information extraction or information/design examination [2]. Different calculations and procedures like Order, Bunching, Relapse, Computerized reasoning, Brain Organizations, Affiliation Rules, Choice Trees, Hereditary Calculation, Closest Neighbor strategy and so on, are utilized for information disclosure from data sets. In this examination paper, we will talk about Affiliation rules mining.

II. ASSOCIATION RULE MINING

Affiliation rules mining is one of the significant parts of information mining and is utilized to find the fascinating affiliations or connection connections between thing sets in mass information. Finding successive thing sets is a critical innovation and move toward the uses of affiliation rules mining [3]. Affiliation rule mining is a significant exploration part of information mining, predominantly used to find the important connections between things in the informational collection. Affiliation Rule is a strategy for information mining that can track down the interconnections of the relationship between the information things in an information exchange. Mining Successive Itemsets from exchange information bases is an essential errand for a few types of information disclosure, for example, affiliation rules, consecutive examples, and grouping [4]. An itemset is continuous on the off chance that the subsets in an assortment of sets of things happen regularly. Successive itemsets is by and large embraced to produce affiliation rules. The goal of Successive Thing set Mining is the recognizable proof of things that co-happen over a client given worth of recurrence, in the exchange data set [5]. Affiliation rule mining is one of the chief issues treated in KDD and can be characterized as separating the fascinating connection and connection among colossal measure of exchanges.

2.1 Affiliation Rule Plan

The motivation behind affiliation rule mining is to figure out the secret connection between various information thing sets in the data set. By and large, given an exchange information base, the issue of affiliation rule mining is the cycle to find the affiliation rules through a client determined least help and least certainty. Let $I = \{I_1, I_2, \dots, I_n\}$ be the arrangement of all information things. D is the arrangement of all exchanges in the exchange data set, in which every exchange T is an assortment

of things, and every exchange T has a special TID identifier [3][4]. For the thing set $X \subseteq I$, which is called that T contains X provided that $X \subseteq T$. On the off chance that $|X| = k$, the assortment is called k request set. In the event that the things organized by the word reference request, k request thing set X can be communicated as $[X_1] [X_2] \dots [X_k]$. For the thing set $C = XUY$, in the event that Y is m request thing set, Y is the m request development of X . In the event that X, Y are project sets, and $X \cap Y = \Phi$, a containing equation $X \Rightarrow Y$ is known as affiliation rules, and X, Y is independently called the reason and finish of the affiliation rule $X \Rightarrow Y$.

Support is a proportion of measurable meaning of affiliation rules in the whole exchange data set, which exhibits the level of portrayal of the standard in all exchanges. The more noteworthy the help, the more significant the standard. If Backing (X) isn't not exactly the client indicated least help MinSup , then, at that point, X is called incessant thing set, alluded to as the successive or huge thing set, also called non-regular thing set, alluded to as non-continuous or little thing set. The certainty is a proportion of affiliation rules' precision. In the event that $\text{Support}(X \Rightarrow Y) \geq \text{MinSup}$ and $\text{conf}(X \Rightarrow Y) \geq \text{MinConf}$, the affiliation rule is successful or solid affiliation rule, in any case, said the affiliation rule to be invalid or powerless affiliation rule. The extent of the standard $X \Rightarrow Y$ in the exchange data set portrays the appearance likelihood of the thing set Y in all exchanges without a trace of condition impact.

III. METHODOLOGY

This study means to find instances of client purchasing affiliations. The exhibiting procedures used in this assessment is the methodology for Alliance Rules while the computation used in this study is the estimation Apriori calculation. The apriori calculation has become one of the most generally involved calculations for continuous itemset mining and affiliation rule learning. It has been applied to various applications, including market bushel examination, suggestion frameworks, and extortion identification, and has roused the improvement of numerous different calculations for comparable errands.

3.1 Apriori calculation

The Apriori works by first distinguishing the continuous itemsets in the dataset (itemsets that show up in a specific number of exchanges). It then utilizes these regular itemsets to produce affiliation rules, which are proclamations of the structure "in the event that thing A_n is bought, thing B is likewise prone to be bought"[3][6]. The Apriori calculation utilizes a granular perspective, beginning with individual things and step by step moving toward more complicated itemsets.

The apriori calculation begins by setting the base help limit. This is the base number of times a thing should happen in the data set for it to be considered a successive itemset [7]. The calculation then, at that point, sift through any up-and-comer itemsets that don't meet the base help edge [8].

The calculation then creates a rundown of all potential mixes of continuous itemsets and counts the times every mix shows up in the data set. The calculation then, at that point, creates a rundown of affiliation rules in light of the regular itemset mixes.

The calculation then sift through any affiliation decides that don't meet a base certainty limit. These guidelines are alluded to areas of strength for as rules. At last, the calculation then returns the rundown of solid affiliation rules as result.

The calculation comprises of the accompanying advances:

- Begin with itemsets containing only a solitary thing (Individual things)
- Decide the help for itemsets
- Keep the itemsets that meet the base help edge and eliminate itemsets that don't uphold least help
- Utilizing the itemsets kept from Stage 1, create all the conceivable itemset blends.
- Rehash stages 1 and 2 until there are not any more new thing sets.

IV. EXPERIMENTAL RESULTS

We utilized python language for Apriori algorithm implementation. We employed a online supermarket dataset. The experimental results are shown in the figure-1.

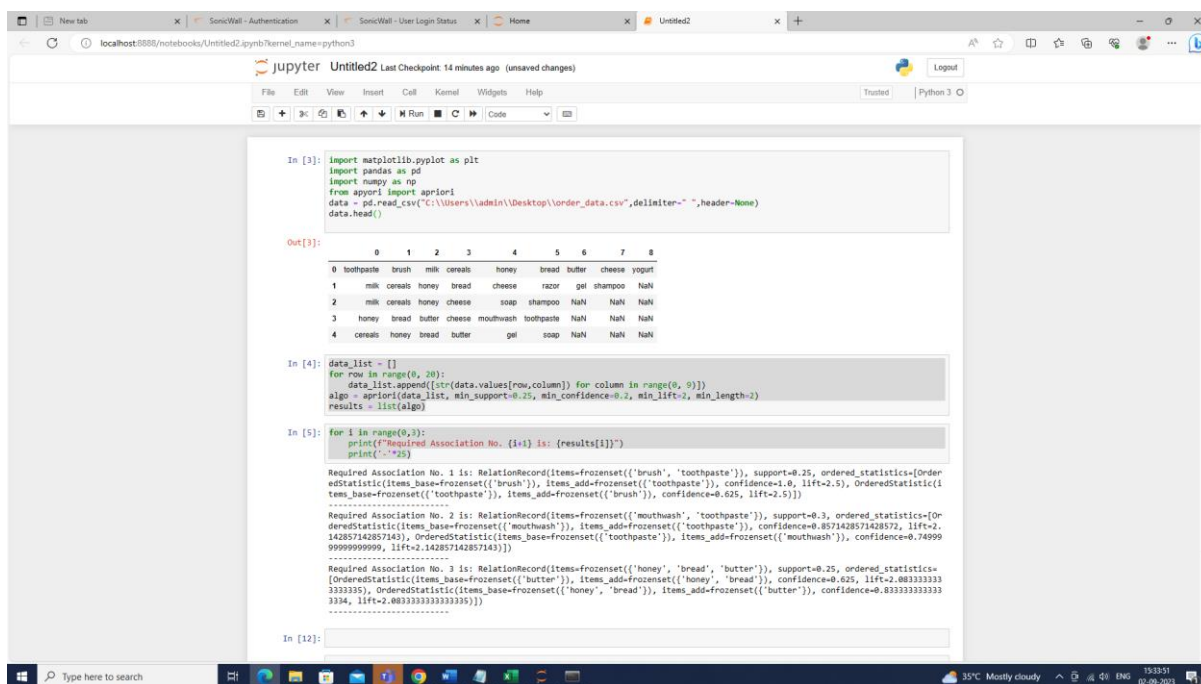


Figure-1: Experimental results of Association rules

4.1 Results

Using the Apriori algorithm, we identified three significant association rules within the online store dataset:

- The association between "brush" and "toothpaste" with a support of 0.25, a confidence of 1.0, and a lift of 2.5. This suggests that customers who purchase a toothbrush are highly likely to purchase toothpaste as well, indicating a strong complementary relationship between these products.
- The association between "mouthwash" and "toothpaste" with a support of 0.3, a confidence of 0.857, and a lift of 2.143. This rule suggests that customers often buy mouthwash together with toothpaste, indicating that these products are frequently purchased as a pair, possibly for oral hygiene reasons.
- The association among "honey," "bread," and "butter" with a support of 0.25, a confidence of 0.625, and a lift of 2.083. This rule indicates that customers who purchase butter are likely to purchase honey and bread together, possibly for breakfast or cooking purposes.

Our analysis reveals strong associations between specific product pairs, indicating potential cross-selling opportunities for the online store. These associations can be leveraged for marketing and inventory management strategies to enhance customer experience and increase revenue.

4.2 Discussion

The Apriori algorithm is a valuable tool for discovering association rules in transaction data, enabling businesses to better understand customer behavior and make informed decisions. In our study, we found several interesting associations that provide insights into customer preferences and buying patterns.

The first association rule between "brush" and "toothpaste" suggests that these products are frequently bought together, and the online store can optimize its product recommendations and promotions to capitalize on this association. For example, offering discounts on toothbrushes when purchasing toothpaste can encourage customers to buy both products.

The second association between "mouthwash" and "toothpaste" highlights the importance of bundling or promoting these products together. By offering a discount or special deal for customers who purchase both mouthwash and toothpaste, the store can increase sales of both items and potentially improve customer satisfaction by catering to their oral hygiene needs.

The third association involving "honey," "bread," and "butter" suggests that these items are commonly purchased together, potentially indicating a preference for breakfast combinations or baking ingredients. The online store could use this insight to

create special breakfast bundles or recipe suggestions that include these products, enhancing the shopping experience and increasing sales.

V. CONCLUSION

In conclusion, association rule mining with the Apriori algorithm is a valuable technique for uncovering hidden patterns in online store transaction data. By leveraging these insights, businesses can make data-driven decisions to improve customer satisfaction and maximize revenue. Further research could involve analyzing larger datasets and exploring more complex association rules to gain deeper insights into customer behavior and preferences.

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] Zhi-Chao Li; Pi-Lian He; Ming Lei, "A high efficient AprioriTid algorithm for mining association rule", 2005 International Conference on Machine Learning and Cybernetics, Year: 2005, Volume: 3,Pages: 1812 – 1815.