

MARKET BASKET ANALYTICS DEFINE CUSTOMER BEHAVIOUR THROUGH TRANSACTION DATA

Kalpana S Kumaran¹

¹- Associate Professor, ITM B-School, Navi Mumbai, India

¹- kalpanas@itm.edu

Gauri Surve²

²Student, ITM B-School, Navi Mumbai, India

²- 2022.gauris@itm.edu

Abstract-The objective of this research is to design an application which can be used to analyse the existing transaction data. The feature of the application is market basket analysis. It is significant to understand the ever-changing customer behaviour and to identify product relationships that are frequently purchased together by customers to enable cross-selling & upselling opportunities, rise product placement, increase recommendations, better inventory management, and optimize pricing tactics which will contribute to profit maximization. Transaction data has frequently been utilized as a research object to find new knowledge. Utilizing transaction data is a better strategy than using the standard, time- and money-consuming customer surveys. Transaction data from Kaggle, a free data portal was used to find the best method for the implementation of market basket analysis. The trail result showed that association rule method using Apriori algorithm works very well. This study also demonstrates that while completing any market basket analysis, time is a crucial consideration that should be considered. This application can be used not only in retail, also in telecommunications, healthcare and anywhere transactional data can expose item associations.

Keywords-Market basket analytics, Association rule, Apriori algorithm, Transaction data.

I. INTRODUCTION

Consumer behavior is an activity that involves making decisions about whether to use and consume the goods and services that have been acquired, as well as customer aspects that may influence these choices. Every consumer has unique wants, preferences, and habits when it comes to addressing those demands. The logging done by the intermediary provider of customer demands can be used to identify that conduct.[1]

In recent years, researchers have increasingly employed transaction data as study and analytical items. Market basket analysis is a frequently used application to examine transaction data from the customer's purchasing cart. The term "market basket analysis" refers to a broad range of analytical methods used to identify relationships and connections between particular commodities, identify consumer behavior, and determine how things are related to one another. If a customer purchases one set of goods, they are more (or less) likely to purchase a different group of goods, according to the following theory. For instance,

it is commonly known that customers typically purchase chips and beer together. The businesses selling these products are interested in the actions that result in purchases [2], [3]

The market basket analysis help businesses selling these products to identify the effective technique for putting up-selling, cross-selling, and inventory management methods into practice [4]. Market basket analysis, sometimes referred to as association rule mining or affinity analysis, has been used to analyze consumer purchasing patterns. It is a data mining technique that was first employed in the marketing industry to comprehend client purchasing patterns by pulling relationships and co-occurrences from transactional databases (i.e., market basket data) [5]

II. LITERATURE REVIEW

X (formerly known as Twitter) is a data mine where practically every user's tweets are exposed to the public. The social media platform enables regular people and well-known public personalities to widely disseminate their uncensored ideas. Users tweet at different rates; some may tweet multiple times each day, while others may only tweet infrequently [6]. These days, other social media platforms have also evolved into data hubs, which aid organizations in creating winning competitive strategies, whereas market basket analysis improve the efficiency of these marketing strategies and tactics. The study reveals whether goods and services are bought concurrently or in a specific order [7].

A. Data Mining

Data mining, often referred to as "facts acquisition information" on occasion, is the computational process of obtaining reliable, obscure, intelligible, and usable information from sizable databases that are used for crucial decision-making. In the process of retrieving knowledge from a database, data mining is regarded as a crucial phase. Users can categorize, analyze, and summarize data as well as the connections between different types of data [1], [8], [9].

Data mining techniques have been used quite successfully up until this point in many different fields. Customer segmentation, in the context of marketing, is the division of a large customer base into smaller customer groups, each of which includes customers who are similar across all of the segments. Using this segmentation technique, it is possible to recognize the customers and group them according to their traits and qualities. The examination of

transactional data is a key area of application for data mining. Each transaction is a collection of objects in a recorded transactional database. Market basket analysis is the most effective method for examining and identifying connections and patterns among different commodities [4], [10].

B. Association Rules

The assertion of "what's with what" is related to the association rule. The rules of association, also known as association rules, are one of the key data mining techniques and the format that is most frequently used to extract a pattern or patterns from a set of data. Association rule mining is typically used to uncover intriguing connections, patterns, and associations between groups of items in a dataset [1], [11], [12]. An example is data about transactions in supermarkets. Transaction data lists all items purchased by a customer in a single purchase transaction. Data on purchases made in supermarkets is one example. Data on transactions includes a list of all the things a consumer has bought in a single transaction. Managers want to determine if a particular set of goods is consistently purchased as a bundle. These managers can use this data to design supermarket layouts where products are placed in the best possible relationship with one another, for marketing initiatives, customer segmentation, product cataloging, or to observe consumer behavior [13], [14].

The association rule will generate a pattern with two (two) parts: the antecedent, which is declared as if, and the consequent, which is declared as then. If represents the item found in the data, then represents the item discovered in conjunction with the if item. Association rules typically take the following form:

$$A \rightarrow B$$

where: A = antecedent if, and B = consequent then so that this concludes that if x buys A, then x will also buy B [11].

Association rules are essentially made up of three metrics: support, confidence, and lift, which characterize the relationship between a collection of antecedents and consequents.

- *Support*: How often a rule applies to a certain data collection is determined by its support. It is calculated by dividing the number of transactions involving both A and B by the overall number of transactions [6], [15], [16].

$$A \rightarrow B = \frac{\sigma(A \cup B)}{N}$$

where A and B are the antecedent and consequent, respectively; N is the total number of transactions in the dataset. Usually, a minimum support threshold is established to only include regulations that have support over the threshold. From study to study, the proper support threshold could be different. Setting the threshold too high could cause intriguing patterns to be missed, while

setting it too low could cause a large number of uninteresting patterns to be studied [6].

- *Confidence*: The reliability of the inference drawn by the rule is gauged by confidence. It details how often the items A and B are purchased together based on the number of times A is purchased.[6], [15], [16].

$$A \rightarrow B = \frac{\sigma(A \cup B)}{\sigma A}$$

where A and B are the antecedent and consequent, respectively; A is the total number of transactions of A in the dataset.

- *Lift*: It specifies the strength of a rule over the randomness of A and B being bought together. It basically measures the strength of any association rule. Depending on the analysis, lift may be useful as it solves the problem of misleading high-confidence rules, which are enabled by confidence not taking support into account.[6], [15], [16].

$$A \rightarrow B = \frac{\text{Support}}{\text{Support } A \times \text{Support } B}$$

If lift is larger than 1, it means that there is a greater chance that the items on the right-hand side (RHS) will occur in this transaction as a result of the left-hand side's (LHS) presence. If the lift is less than 1, it means that the likelihood that the items on the right-hand side (RHS) will be purchased is less likely as a result of the left-hand side's (LHS) presence. If the lift is 1, it means that the left-hand side (LHS) and right-hand side (RHS) item's presence are independent; the likelihood that things will appear on the right-hand side (RHS) is unaffected by the left-hand side's (LHS) item's presence [4], [7].

C. Apriori Algorithm

The Apriori algorithm initially splits the customers into many segments, groups, and clusters. The frequent item sets and association rules for each of those categories are then discovered. The Apriori algorithm attempts to find client behaviors as groups in order to efficiently satisfy those specific groups of persons [11], [17].

There are numerous narration or pass stages in the apriori algorithm.

- Step 1: When creating an itemset candidate, the k-itemset antedate is created by combining the itemset from the previous iteration with (k-1). The trimming of k-itemset candidates whose subset contains k-1 items is not included in the high frequency pattern with length k-1 is one method of the apriori algorithm.
- Step 2: Support computation for each candidate in the k-itemset. The number of transactions that

contain every item in each k-itemset candidate is determined by scanning the database for each candidate. This is a characteristic of the apriori process as well, where it is necessary to find the lengthiest k-itemset for the entire database.

- Step 3: Establish a high-frequency pattern. a high frequency pattern with k items or k item sets whose support is higher than the minimal support.

The process is terminated if no high frequency pattern is discovered. If it is obtained once more, k is increased by one and the position is returned to Step 1.

III. METHODOLOGY AND ANALYSIS

One of the earliest and fastest tools for association mining is the apriori algorithm. The Apriori algorithm is the implemented algorithm in the market basket analysis application in association rule mode. We have sourced the data from Kaggle portal and have used Excel for the calculation, Tableau for data interpretation & Google Looker Studio for visualizations to perform a simple market basket analysis based upon dataset.

In dataset, a chain of grocery stores from Tamil Nadu was considered for the analysis. The transaction details are from 2015 – 2019 with 7 categories and 24 sub-categories of grocery stores in 24 cities from Tamil Nadu. Total number of transactions are 9995 from 5 different regions of Tamil Nadu State. Using association rule (Apriori algorithm), 2-way lift and 3-way lift was considered for the analysis.

We search for rules that have a lift of more than one when performing market basket analysis. Furthermore, it is better to have rules with high support and confidence, as these will be relevant to a large number of transactions and will

City	Sales
Vellore	174073
Bodi	173655.1
Kanyakumari	172217.7
Perambalur	171132.2
Karur	169305.9
Tirunelveli	165169
Pudukottai	164072.6
Chennai	160921.3
Salem	160899.3
Krishnagiri	160477.5
Ramanadhapuram	158951
Coimbatore	157399.4
Cumbum	156355.1
Tenkasi	156230.7
Madurai	152548.6
Virudhunagar	150816.7
Ooty	150078.9
Namakkal	145502.1
Dindigul	144873
Viluppuram	144200.6
Theni	142739.8
Dharmapuri	141593.1
Nagercoil	137848.5
Trichy	136059.9

Table 1: 24cities and their sales in the state of Tamil Nadu

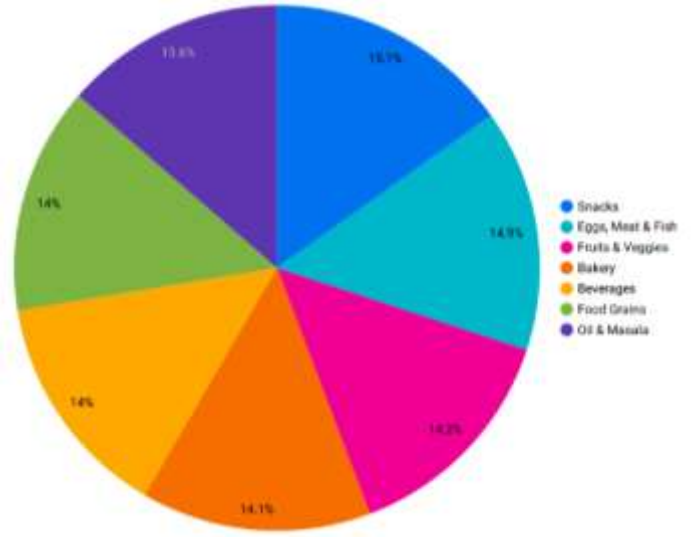


Figure 1: 7 categories sale in the 24 cities of Tamil Nadu

increase the likelihood of an item being on the right-hand side (RHS) given the presence of items on the left-hand side (LHS) [4], [7].

For 2-way lift, city as left-hand side (LHS) and sub-category as right-hand side (RHS) was chosen to see the lift value. The highest lift value was 1.9 for the city “Namakkal” and the sub-category “mutton” shows that chances are more for the people from “Namakkal” prefer “mutton”. The second highest lift value is 1.6 for the city “Coimbatore” and the sub-category “organic vegetables” show that the customers from “Coimbatore” are more health conscious and prefer Organic products. The third highest lift value is 1.5 for the city “Trichy” and the sub-category “Breads & Buns”.

For 3-way lift, region as left-hand side (LHS) and city & sub-category as right-hand side (RHS) were chosen to see the lift value. The highest lift value was 2.7 for west region, the city “Dharmapuri”, and the sub-category was “Soft Drinks”. The second highest lift value was 1.8 for central region, the city “Krishnagiri”, and the sub-category was “Eggs”. The third highest lift value was 1.7 for south region, the city “Ooty”, and the sub-category was “Cookies”.

	City	Sub-category	Lift Value
2-way lift	Namakkal	Mutton	1.9
	Coimbatore	Organic vegetables	1.6
	Trichy	Breads & Buns	1.5
3-way lift	Generate	Soft drinks	2.7
	Krishnagiri	Eggs	1.8
	Ooty	Cookies	1.7

Table 2: 2-way lift & 3-way lift was performed using Tableau to generate the lift value

IV. CONCLUSION

In conclusion, this research focused on the development of an application for market basket analysis using transaction data sourced from Kaggle. Market basket analysis is a powerful tool for understanding customer behavior, product relationships, and opportunities for cross-selling and upselling. The Apriori algorithm was employed to uncover association rules, considering both 2-way and 3-way lift values. The analysis revealed interesting insights into consumer preferences across different cities and sub-categories of grocery stores in Tamil Nadu. For instance, the high lift values indicated that people from Namakkal prefer mutton, while Coimbatore residents lean towards organic vegetables. Additionally, the study highlighted that the west region, specifically Dharmapuri, has a strong affinity for soft drinks. Such findings have significant implications for businesses in terms of product placement, marketing strategies, and inventory management. Overall, this application not only has relevance in the retail sector but also holds potential applications in other industries where transactional data can reveal valuable item associations, such as telecommunications and healthcare.

V. REFERENCES

- [1] F. Kurniawan, B. Umayah, J. Hammad, S. Mardi, S. Nugroho, and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data," vol. 1, no. 1, pp. 20–25, 2018, doi: 10.17977/um017v1i12018p20-25.
- [2] G. Kapadia and K. Kalyandurgmath, "Market Basket Analysis of Consumer Buying Behaviour of a Lifestyle Store."
- [3] N. Mak, "International Research Market Basket Analysis." [Online]. Available: www.ijtsrd.com
- [4] S. M. Zafar Ansari, "Market Basket Analysis."
- [5] Dr. Y. Arora and Dr. Y. Arora, "Market Basket Analysis using Apriori Algorithm," *International Journal of Innovative Research in Computer Science & Technology*, pp. 62–66, May 2022, doi: 10.55524/ijirest.2022.10.3.12.
- [6] R. Kruse, T. Lokukatagoda, and S. Alkhushayni, "A framework for association rule learning with social media networks," *IOP SciNotes*, vol. 3, no. 1, p. 015001, Mar. 2022, doi: 10.1088/2633-1357/abe9be.
- [7] "Market Basket Analysis."
- [8] W. P. Nurmawati *et al.*, "Market Basket Analysis with Apriori Algorithm and Frequent Pattern Growth (Fp-Growth) on Outdoor Product Sales Data", [Online]. Available: <https://ijersc.org/>
- [9] PakSree, "."
- [10] W. : Ww, P. Prasad, and M. Mourya, "International Journal of Emerging Technology and Advanced Engineering A Study on Market Basket Analysis Using a Data Mining Algorithm," 2008. [Online]. Available: www.ijetae.com
- [11] M. Qisman, R. Rosadi, and A. S. Abdullah, "Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1722/1/012020.
- [12] Vj. Sai Sree, Aa. Lakshmi, and J. Jayanthi, "Market Basket Analysis using Apriori Algorithm," 2022.
- [13] I. Yustiana, I. Sutarha, N. Maulawati, and I. M. Yusuf, "BASKET MARKET ANALYSIS USING R-BASED APRIORI ALGORITHM TO FIND INFORMATION FROM SALES DATA," *International Journal of Engineering and Applied Technology (IJEAT)*, vol. 4, no. 2, pp. 79–92, 2021, doi: 10.52005/ijeat.v3i1.xxx.
- [14] "INDIAN JOURNAL OF APPLIED RESEARCH X 633 Market Basket Analysis by Using Apriori Algorithm in Terms of Their Effectiveness Against Various Food Product." [Online]. Available: <http://en>.
- [15] T. Raeder and N. V. Chawla, "Market basket analysis with networks," *Soc Netw Anal Min*, vol. 1, no. 2, pp. 97–113, Jan. 2011, doi: 10.1007/s13278-010-0003-7.
- [16] A. Ramesh Jadhav, A. Ramesh Jadhav, and R. D. Jadhav, "Association Rule Mining in retail: Exploring Market Basket Analysis with Apriori Algorithm." [Online]. Available: <https://ssrn.com/abstract=4461121>
- [17] K. Sussolaikah, "Market Basket Analysis for Determination of Consumer Behavior at XYZ Stores Using R Programming," *Advance Sustainable Science, Engineering and Technology*, vol. 3, no. 2, p. 0210206, Nov. 2021, doi: 10.26877/asset.v3i2.8547.
- [18] A. Trmka, "Market Basket Analysis with Data Mining methods," 2010 International Conference on Networking and Information Technology, Manila, Philippines, 2010, pp. 446–450, doi: 10.1109/ICNIT.2010.5508476.
- [19] Sanket Sandip Khedkar, Sangeeta Kumari, "Market Basket Analysis using A-Priori Algorithm and FP-Tree Algorithm", 2021 *International Conference on Artificial Intelligence and Machine Vision (AIMV)*, pp.1-6, 2021.
- [20] Abishek B. Rao, Jammula Surya Kiran, Poornalatha G, "Application of market-basket analysis on healthcare", *International Journal of System Assurance Engineering and Management*, 2021.
- [21] Ruben Nicolas-Sans, Daniel González Ibáñez, "Customer basket heterogeneity: how to measure it and some possible business applications", *Economic Research-Ekonomska Istraživanja*, pp.1, 2020.
- [22] Ruben Nicolas-Sans, Daniel González Ibáñez, "Customer basket heterogeneity: how to measure it and some possible business applications", *Economic Research-Ekonomska Istraživanja*, pp.1, 2020.
- [23] A. H. Patwary, M. T. Eshan, P. Debnath and A. Sattar, "Market Basket Analysis Approach to Machine Learning," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-9, doi: 10.1109/ICCCNT51525.2021.9580175.