2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)

Shopping Cart Analysis Using Apriori Algorithm and Association Rules

1st Prajna K B
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bangalore, India
prajna.kb@nmit.ac.in

2nd Abhishek K
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bangalore, India
abhishekkreddy@hotmail.com

3rd Ashwin A Shanbhag
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bangalore, India
ashshanbhag9945@gmail.com

4th B V Shashank
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bangalore, India
shashankbola10@gmail.com

5th Ekshith P
Department of Electronics and Communication Engineering
Nitte Meenakshi Institute of Technology
Bangalore, India
ekpujar@gmail.com

2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC) | 979-8-3503-2379-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAISC58445.2023.10200434

***Abstract*—Machine learning and artificial intelligence have many applications in today's world. Computers can process large amounts of data very quickly. One of the applications of big data analysis is data mining. Shopping cart analysis is a data-only technique that retailers use to better understand shoppers and increase sales. Shopping cart analysis is a powerful tool for retailers because it allows them to understand the purchasing behavior of customers and create strategies to increase sales and customer loyalty. Association rules are used in shopping cart analysis to predict the probability of two products being purchased together. The point is, they create or organize the right mining, it is the basis of the work basket analysis. We found that the Apriori algorithm is 100 times faster than the current Apriori implementation.**

Keywords—Data Mining, Apriori algorithm, Association rule mining.

I. INTRODUCTION

Machine Learning and Artificial Intelligence have many applications in today's world. Computers can process huge sets of data very quickly. This can be accelerated by using machine learning algorithms. One of the applications of big data analysis is data mining. Finding anomalies, trends, and correlations within huge data sets to forecast outcomes is known as data mining. E-commerce websites use data mining techniques to analyse shopping baskets of the customers to find associations between different items.[17][18].

Retailers use Shopping cart analysis, a data mining technique, to better understand customer buying habits and boost sales. Utilising Shopping cart analysis became simpler with the introduction of electronic point-of-sale (POS) systems. The digital records created by POS systems made it simpler for apps to handle and analyse large amounts of purchase data compared to the handwritten records kept by store owners.[11,12].

Shopping cart Analysis forms association rules among the different items in the data sets. The associations formed are used to group frequently bought items or products together and even provide discounts on them. They can also be used to bundle these products with less frequently bought items to boost sales. Association rule mining uses support, confidence and lift to filter the items that have been bought less frequently.[13].

II. RELATED WORKS

A. Association Rules Mining

Association rule mining discovers interesting relationships and relationships between the vast amount of information about products. This rule indicates how many products appear on the market. A classic example is market analysis. Given a business process, we can find rules that predict the outcome of an item based on the outcome of other items in the swap.

According to Patwary et al. [1]. the relationship between association rules and IF-THEN can be thought of as an IF-THEN relationship. If a customer buys item A, it is estimated how likely it is that they will choose item B with the same Transaction ID. These guidelines consist of two parts:

- Antecedent (IF): This is a category of item or collection of items that may be found in datasets or item sets.
- Consequent (THEN): This item is accompanied by an antecedent or a group of antecedents

In the process of forming association rules, useful association rules are formed by support and confidence values. Useful rules are formed if the association rules exceed minimum support and confidence value. [2].

1) Support: The percentage of transactions that contain each item in a list of items (for example, rubber, paper, and pencil). The level of support raises the frequency of the itemset. High support rules are suggested because they are more likely to be applicable to a significant portion of potential transactions. [3].

The support calculation formula is given below in Eq. (1):

$$Support(A) = \frac{(transaction\ amount\ contains\ A)}{(total\ transaction)} \quad (1)$$

For the support value of more than one item, the formula is given below in Eq. (2):

$$Support(A, B) = \frac{(transaction\ amount\ contains\ A\ and\ B)}{(total\ transaction)} \quad (2)$$

2) Confidence: A variable with an object (pen and paper in our example) on the left side of the rule has a probability of having an object (eraser) on the right side of the rule. The more reliable it is, the more you buy the product on the right,

979-8-3503-2379-5/23/$31.00 ©2023 IEEE

meaning the higher returns you can expect for a particular policy.

The value of confidence in forming Association rules is given below in Eq. (3):

$$Confidence\ P(A|B) = \frac{(transaction\ amount\ contains\ A\ and\ B)}{(total\ transaction\ contains\ A)} \quad (3)$$

3) Lift: The support of a rule is calculated by dividing the chance of all the items occurring simultaneously by the product of the probabilities of the items occurring separately (as if there were no link between them). The lift parameter shows the validity of the association rule created.

$$Lift = \frac{Support(A+B)}{Support(A)\ X\ Support(B)} \quad (4)$$

One way to find association rules is A priori algorithm proposed by Agrawal and Srikan in 1994 [4]. The Apriori algorithm is commonly used in market basket.

Budhi et al. [5] explains that based on the current circumstances, namely the fulfilment of the required sales information systems-based data mining for the business to be computerized and enhance control over business processes in the firm. Making this application has led to the creation of methodologies for Shopping cart analysis using data mining, market purchasing, and sales processes. et analysis was used to create association rules for the dataset's items.

Cong et al. [6]. points out the problems with the Apriori algorithm in practice:

The Apriori algorithm needs to scan the database many times this will increase the I/O load when the database is too large.

It mines out association rules that have no practical significance.

This paper improves the algorithm by pruning the unwanted items after each iteration by adding a judgement dataset and reduce the cost of time consumption.

Mlambo et al. [7]. introduced a system to reduce the resource requirements of the algorithm by using the MapReduce. The MapReduce software framework is utilized for processing and generating large datasets in a parallel manner and hence reducing the overhead on the processor.

Hasan et al. [8]. proposes a methodology to mine frequent item sets using FP growth algorithm. It creates association rules without creating sub-sets of the dataset. It applies a divide and conquer method to find the support value. The paper also proposes a method to find the support value through Binomial Distribution adaptively.

Gayathri et al. [9]. shows the implementation of a variation of the FP Growth algorithm called the FP-Bonsai. As it will never support any itemset matching the constraint, a transaction that does not satisfy a monotone constraint can be erased from the input database.

Apriori Algorithm finds its application in E-Commerce websites. X. Ren [10]. shows the applications as

1) Find potential customers by comparing the late purchase direction of a new client to the old customer.
2) Realizes customer residency by using product page placement and customized advertising display in accordance with consumer choices, it aids retailers in staying on top of customers' psychological activity and genuine demands.
3) Big data computations strongly influence how the site's region, which includes all the products and related products, is designed. Utilizing this strategy enhances both the stickiness of customers and dealers while also giving customers a positive perception of dealers.
4) Data mining on the web makes it simple to make market forecasts by studying what customers might buy in the future and calculating the projected return on investment. In essence, this encourages the correctness of business decisions and significantly lowers operating expenses for the corporation.

This paper consists of research methodology where the flow of the algorithm is discussed, results and discussions where the results of the implementation are displayed and discussed and finally the conclusion section where the final findings of the paper are presented.

III. METHODS

The flowchart in Fig 1. shows the steps we used in the project.

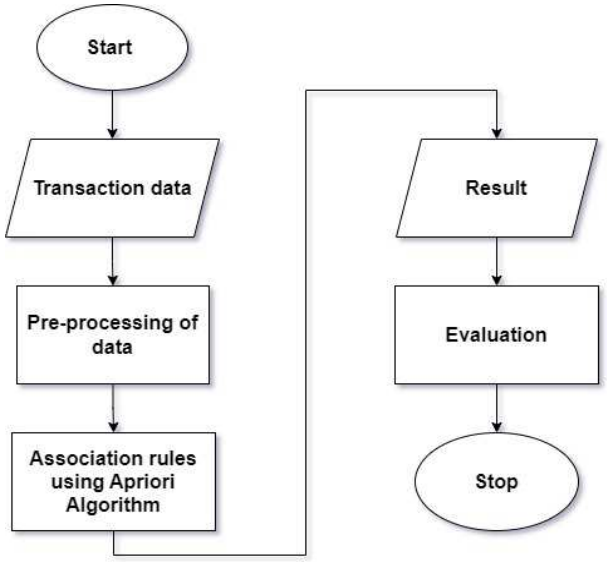

Fig. 1. Research Flow

- Transactional data. In order to conduct our research, we used a publicly available data collection of 25,900 online retail transactions for a UK-based online retailer that occurred between December 2010 and December 2011 [6]. There were 4,372 customers in total, hailing from 38 different countries, and 4,223 unique products.
- The spaces and data without proper transaction IDs are filtered out so that the data to be processed is cleaner to process the results.
- Apriori algorithm association rules. In this study, Shopping cart analysis was used to assess sales transaction data using association criteria implementing the Apriori algorithm. The frequent itemset search method benefited from the a priori algorithm.[15].

- Evaluation. The next stage after performing data mining is to assess the outcomes of the analysis procedure. The results analysis was used as a guide for making inferences and recommendations for this study.

IV. Results And Discussions

Implementation of Shopping cart analysis can be seen as follows. Creating a website enables a retailer to upload his data and view the results of most products that are purchased by customers. The retailer can understand customer purchase patterns from the result obtained through Apriori Algorithm and increase sales accordingly. The retailer must upload the file to be processed in the upload section of the website. [16].

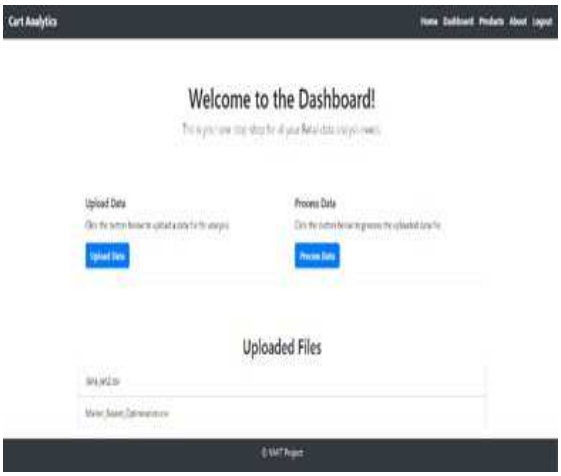

Fig. 2. Dashboard of the website

Fig. 2 is the dashboard page where the files to be processed can be uploaded in the .csv, comma-separated values, format. After the file is uploaded, the data can be processed and the results can be viewed in the form of a table as shown below.

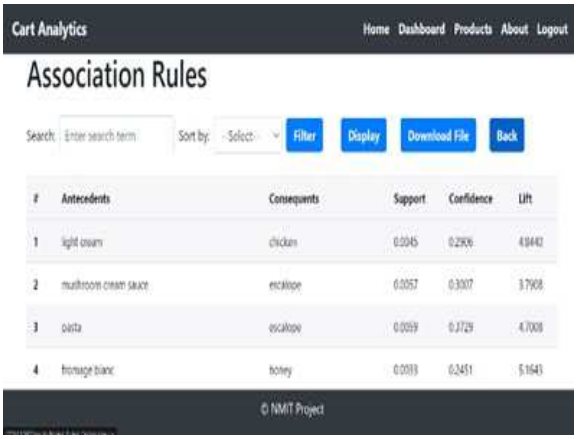

Fig. 3. Apriori Results page

Fig. 3 shows the association rules for the items in the uploaded dataset after processing namely the support, confidence, and lift values can be seen for different items of the dataset and for the combination of the items. The results can be filtered to only show the rules of a single item by searching for the item. The code for the results page that we have written is shown in the snippet below in Fig 4.

```
def results(filename):
    user_email = current_user.email
    file = os.path.join(app.config['UPLOAD_FOLDER'],
user_email,'Datasets',filename)

    df = pd.read_csv(file, header=None)
    df.fillna(0, inplace=True)

    # Transform DataFrame into list of transactions
    transactions = []
    for i in range(0, len(df)):
        transactions.append([str(df.values[i, j]) for j
 in range(0, len(df.columns)) if str(df.values[i, j
]) != '0'])

    # Perform Apriori algorithm
    rules = apriori(transactions, min_support=0.003,
min_confidence=0.2, min_lift=3, min_length=1)
```

Fig. 4. Code snippet for the results page

Fig. 5 shows the association rules of "chicken" with other items in the data set. This page can be downloaded for to study the buying patterns for the item "chicken". The results of the association rules page is displayed after carrying out the apriori algorithm process. The function that we have written to implement the search function is shown in Fig. 6.

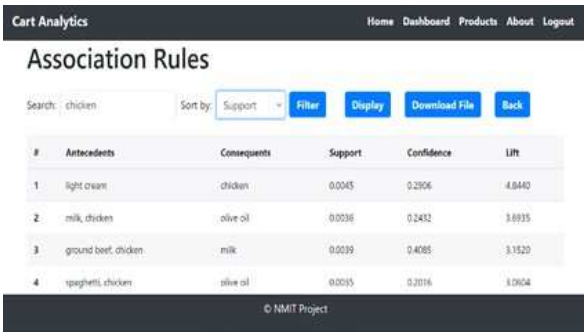

Fig. 5. Association Rules page for "chicken"

```
    # Search the results based on the 'search'
parameter
    if search:
        results = [result for result in results if
search.lower() in result['items'].lower()]
```

Fig. 6. Code snippet of the search function

A bar graph can be plotter for the association rules that indicate the support, life, and confidence values of each itemset as shown in Fig.7.

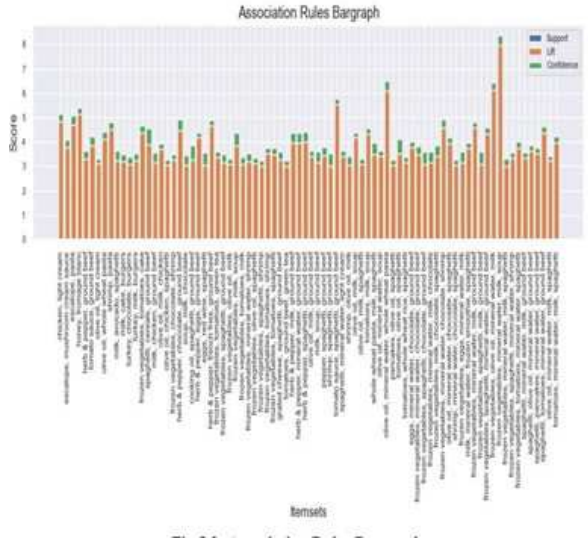

Fig. 7. Association rule bar graph

The association rules of the transactions are plotted in the form of a heat map as shown in Fig .8. Lighter colour indicates the chances of those items bought together is higher.

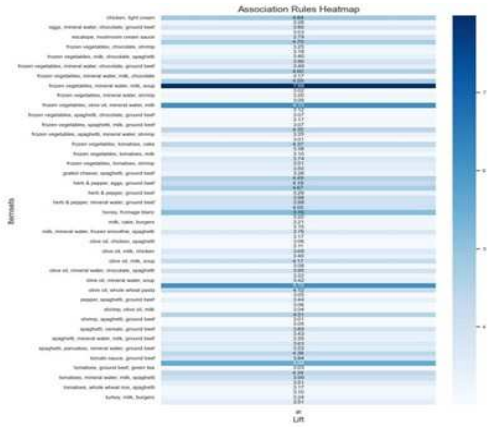

Fig. 8. Heat Map for Association Rule

Some of the methods are implemented to improve the run time and memory efficiency of the Apriori algorithm. By running the optimised Apriori algorithm for the dataset, we obtain the results as tabulated below: (All numbers are in seconds)

TABLE I. EXECUTION TIME RESULTS

Min support	Max length	Optimised Apriori	mlxtend	Apyori
0.01	1	0.1	0.1	0.2
0.01	2	1.8	38.8	1.9
0.01	3	1.9	42.8	2.4
0.01	4	2	42.5	2.4
0.01	5	2	43.8	2.4
0.005	1	0.2	0.2	0.2
0.005	2	5.4	190	5.6
0.005	3	5.6	301.7	18.5
0.005	4	5.4	315.1	37.2
0.005	5	5.7	307.4	37.6

We can see a major speedup when the minimum support value is set to 0.005, when the pruned dataset contains a lot more items. The Apriori algorithm from mlxtend package completes execution in 301.7 seconds for a minimum support value of 0.005 and a maximum length of data of 3. For the same set of parameters, the Apriori algorithm from Apyori package completes execution in 18.5 seconds. The mlxtend was allocating about 150GB of memory and crashing. The optimised Apriori is also memory efficient. [14].

Fig 9. shows the execution time comparison between Apriori algorithm of Apyori package with the orange plot indicating the execution time of the optimised Apriori and the blue line indicating the execution time of Apyori package.

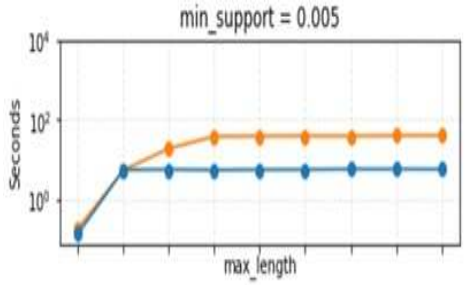

Fig. 9. Graph of execution time for optimised apriori

V. CONCLUSION

By identifying the association rule, the Apriori algorithm analyses sales transaction data for the online retail data set items over a year, from December 1, 2010, to November 30, 2011. The programming languages HTML, PHP, and CSS were used to create the website. Among the five pages are the login page, the upload page, the dashboard page, the process page, and the display page.

The results of the rule-making procedure can be utilized to recommend new strategies for marketing retail goods. One of these is holding promotions where customers who buy product bundles receive bonuses for unsold products.

The optimized Apriori is approximately 54 times faster than the traditional Apriori available in the mlxtend package. As the minimum support values increase, the optimised apriori performs better than the traditional apriori algorithm.

The author is aware of the numerous flaws and limitations in the current state of the research. In order to increase efficiency and shorten the data processing procedure, the author will encourage further research to create an automatic data processing system (data pre-processing). The author is aware that the research conducted still has many flaws and limits. Additionally, a Python-based a priori process can be created to speed up execution.

REFERENCES

[1] 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), 2021.

[2] S. Hadi, Implementation of Data Mining with Association Rule in Decision Making for Correlation of Product Purchases Using Apriori Algorithm (Case Study: ASR JEANS Jember). Jember: Muhammadiyah University of Jember, 2016I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] D. Widiastuti, N. Sofi, “Comparative Analysis of Apriori Algorithm and FP-Growth in Cooperative Transactions.” UG Journal, 21, 2014.

[4] M. Badrul, “Association Algorithm With Apriori Algorithm For Sales Data Analysis.” Journal of Pilar Nusa Mandiri Vol. XII, No. 2, 125, 2016.M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[5] A. Setiawan, G. S. Budhi, D. H. Setiabudi and R. Djunaidy, "Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company," 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), 2017, pp. 337-340, doi: 10.1109/ICSIIT.2017.39.

[6] Y. Cong, "Research on Data Association Rules Mining Method Based on Improved Apriori Algorithm," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2020, pp. 373-376, 2020.

[7] M. N. Mlambo, N. Gasela and M. B. Esiefarienrhe, "Implementation and Analysis of Enhanced Apriori Using MapReduce," 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD), 2018, pp. 1-6, doi: 10.1109/ICABCD.2018.8465426.

[8] M. M. Hasan and S. Zaman Mishu, "An Adaptive Method for Mining Frequent Itemsets Based on Apriori And FP Growth Algorithm," 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018.

[9] B. Gayathri, "Efficient market basket analysis based on FP-Bonsai," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 788-792, doi: 10.1109/I-SMAC.2017.8058287.

[10] X. Ren, "Application of Apriori Association Rules Algorithm to Data Mining Technology to Mining E-commerce Potential Customers,"

2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin City, China, 2021, pp. 1193-1196, doi: 10.1109/IWCMC51323.2021.9498773.

[11] T. Lim, "K-Means Clustering-Based Market Basket Analysis: U.K. Online E-Commerce Retailer," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 126-131, doi: 10.1109/ICIT52682.2021.9491703.

[12] Berlilana, T. Hariguna and A. R. Hananto, "Analysis of Customer Product Interests using the Market Basket Analysis Model with Hash-Based Algorithm and Association Rules," 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Prapat, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICORIS56080.2022.10031409.

[13] A. Gehlot and R. Singh, "Execution of Market Basket Analysis and Recommendation Systems in Physical Retail Stores to Advance Sales Revenues," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-5, doi: 10.1109/INCOFT55651.2022.10094471.

[14] A. A. Aldino, E. D. Pratiwi, Setiawansyah, S. Sintaro and A. Dwi Putra, "Comparison Of Market Basket Analysis To Determine Consumer Purchasing Patterns Using Fp-Growth And Apriori Algorithm," 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Banyuwangi, Indonesia, 2021, pp. 29-34, doi: 10.1109/ICOMITEE53461.2021.9650317.

[15] S. S. Khedkar and S. Kumari, "Market Basket Analysis using A-Priori Algorithm and FP-Tree Algorithm," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670981.

[16] Prajna K B, Akshat Kothiyal,Dhanush N, Ankit Kumar "Implementation Of A Hybrid Recommendation System For Movies",2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon),Date of Conference: 16-17 October 2022, DOI: 10.1109/MysuruCon55714.2022.9972580

[17] C. Zhang and Kushalatha. M R, "Deep Learning Technology in Film and Television Post-Production," 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2023, pp. 1-5, doi: 10.1109/ICICACS57338.2023.10100148.

[18] R. Dilip, N. Samanvita, R. Pramodhini, S. G. Vidhya & Bhagirathi S. Telkar ,"Performance Analysis of Machine Learning Algorithms in Intrusion Detection and Classification", Part of the Communications in Computer and Information Science book series (CCIS,volume 1591) ICETCE 2022: Emerging Technologies in Computer Engineering: Cognitive Computing and Intelligent IoT