

Analysis and Prediction of Customer's Purchasing Pattern Using Market Basket Analysis with Association Rule Mining

Santosh P. Shrikhande¹, Prashant P. Agnihotri²

^{1,2}*School of Technology, S.R.T.M. University, Sub-Campus, Latur (MS), India*

¹santoshshrikhande@gmail.com

²agnihotri@gmail.com

Abstract—Analysis of the customer's purchasing patterns or behaviour is very much important in supermarket management to design their marketing policies by gaining insight into which items are often bought together by the customers. Recently, the market basket analysis (MBA) technique of data mining is widely used for interpreting the relationships among the various items often purchased together. It plays a very important role in supermarket management that helps for arrangement of the goods and selves, designing the sales promotions and giving offers and discounts on specific items to the customer so that sales and revenue will be increased. This research paper presents an analysis of relationship among the various products purchased often together using market basket analysis with association rule mining technique. Experiment conducted on supermarket transactional data set using an Apriori algorithm in Weka software to generate the best association rules. The association rules generated from this experiment, its analysis and interpretations can be useful for the supermarket managers to strategically promote the products and improve sales revenue. These results are also useful to provide valuable insights for new products integration and selves arrangements task, cross-selling and up-selling in the supermarket.

Keywords—Data Mining, Association Rule Mining, Market Basket Analysis, Apriori Algorithm.

I. INTRODUCTION

Data mining is the process of extracting the novel, meaningful, potentially useful and non-trivial pattern from the huge amount of data from different sources. It is also known as knowledge discovery in databases (KDD) [1]. Recently, data mining techniques are widely used in various fields, such as marketing, banking, insurance sector, bioinformatics, education field and medical field for examining the interesting pattern and their analysis [2]. Association rule mining is one of the widely used data mining method in supermarkets to examine the relationship among the various items purchased often together [3]. Market basket analysis (MBA) is a typical example of association rule mining that helps the marketing analyst to understand the purchasing behaviour of customer by analysing the market basket such as which products are being bought often together [4]. Therefore, market basket analysis is most prevalently used in supermarket to analyse customer's shopping pattern [1], [2], [3].

A. Market Basket Analysis (MBA)

Market basket analysis is a typical example of association rule mining that performs analysis of customer's purchasing

pattern through identifying association and/or correlation among various items from the shopping market basket [1]. In supermarket, market basket is the combination of various items purchased together by the customer in a single transaction. In precise, market basket analysis identifies the most frequently purchased items, interesting relationships in between the purchased items and provides understanding that helps the supermarket managers to make the right decision in determining the sales policies and knowing the consumer's buying habits regarding what products are often purchased [2]. Moreover, it helps in placing the most frequently purchased products simultaneously into one specific area or rack.

B. Association Rule Mining (ARM)

Association rule mining is one of the data mining tasks that examine the relationship in between two items from the market basket of supermarket shopping [1]. Generally, association rules are expressed in the if-then form such as $A \rightarrow B$ where A and B are two items or item sets [2], [3], [4]. Let's consider I is a set of n items $I = I_1, I_2, I_3, \dots, I_n$ and T is a set of transactions $T = T_1, T_2, T_3, \dots, T_m$ in the transactions database D . The transaction T_i has some I , i.e. $T_i \subseteq I$. Then association rule is represented in the form such as $A \Rightarrow B$, where $A, B \subseteq I$ items or item sets and $A \cap B = \emptyset$, here A is an antecedent and B is a consequent. The rule $A \Rightarrow B$ indicates that if item A is purchased then B is also purchased along with item A [5]. Following example shows the list of transactions with purchased items from the supermarket in Table 1.

TABLE 1: AN EXAMPLE OF MARKET BASKET TRANSACTIONS

Transaction Id (TID)	Items Purchased
1	Butter, Cheese, Burger
2	Milk, Cheese, Butter
3	Butter, Milk, Bread
4	Milk, Bread, Cheese

The interesting relationships from the above transactions can be represented in the form of association rules such as: Milk \rightarrow Bread. This rule shows that there is a strong relationship between Milk and Bread that means most of the customer buys Milk and Bread together. If Milk is purchased

by the customer then Bread is also purchased along with Milk. The strength and probability of this association rule is measured using three factors known as support, confidence and lift [6], [7].

Support is the probability that both A and B item sets occur together in a given transaction. The higher value of support indicates the more frequency of item set occurrence [7], [8], [10]. Support is computed using following formula:

$$S(AUB) = \frac{\text{Count}(AUB)}{N} \dots\dots\dots(1)$$

Where *S*-is the support and *N*-is the total number of transactions.

Confidence is also the probability that, if a transaction has item set A then it will also contains B item set. It means probability of purchasing the item set B along with the item set A. Confidence is calculated using following formula [7], [8],[10]

$$\text{Confedence of } (A \rightarrow B) = \frac{\text{Number of transactin that contain A and B}}{\text{Number of transaction that contain A}}$$

$$C(A \rightarrow B) = \frac{S(AUB)}{S(A)} \dots\dots\dots(2)$$

Where *C*-is the confidence and *S(AUB)*-is support of A with B item set.

The lift measure is a probability of co-occurring items in rules divided to the multiplication of supports of left and right sides. It means lift sum up the strength of association rule between items in the left and right sides [7], [8], [10], [11]. Lift is computed using following formula:

$$C(A \rightarrow B) = \frac{S(AUB)}{S(A)S(B)} = \frac{S(AUB)}{S(B)} \dots\dots\dots(3)$$

The higher value of the lift indicates the stronger connection or relationship between the two items or item sets.

The rules generated from association rule mining task are said to be interesting, if these rules meets minimum support and confidence threshold set by the user. An example of association rule mining is given below:

Milk→Bread [support = 50%, confidence = 100%].

This association rule indicates that 50% of the total customers purchased Milk, Bread and there is a 100% confidence that if a customer buys Milk, then he buys Bread too [12].

This research paper presents a methodology for extracting and analysing the purchasing behaviour of supermarket customer based on market basket analysis using an Apriori algorithm. This will help super market managers in marketing, targeted advertising, floor planning, inventory control, and churning management. The organization of this research paper is as follows: Section I, provides the introduction of association rule mining and market basket analysis with their basics. The literature review of existing methodologies is illustrated in the section II. Section III, presents the methodology for extracting association rules for analysing the

customer’s purchasing behaviour. Section IV has provided the results and discussion. The conclusions and suggestions based on the results and discussion are given in Section V.

II. RELATED WORK

Association rule mining technique of data mining has been widely used in super market especially for the purpose of market basket analysis to find best association rules between frequently purchased items by the customer. Therefore, many researchers have used market basket analysis technique to understand the purchasing behaviour of customer by analysing the market basket such as which products are being bought often together [12],[13]. The detail literature review of existing association rule mining techniques is given below.

In [6], Yuksel Akay et al. performed a market basket analysis technique using Apriori and FPgrowth algorithm on 255 different items of super market data from Vancouver Island University website using Weka software. Author generated the top ten strong association rules using support, confidence, lift and conviction to suggest super market manager for arrangement of products into the selves so that customer would purchase the product and increase the revenue of supermarket.

In [7], Alfiqra and Khasanah proposed Overall Variability of Association Rule (OCVR) that focuses on the market basket analysis by considering the high variability in costumers buying behaviour. Author used transactional data of Yogyakarta Supermarket, Indonesia for generating association rules and then these rules were used for further analysis using OCVR. After conducting experiment author concluded that the rules with OCVR value smaller than 30% can be used for making marketing strategies such as product bundling and shelves product arrangement.

In [8], A. R. Efrat et al. used an Apriori algorithm for analysing customers purchasing patterns while shopping into the supermarket. Author used supermarket transactional data and with minimum support, confidence generated top association rules using Apriori algorithm in Weka software. Afterwards author concluded that these rules determine the consumer purchasing behaviour and overall sale of supermarket.

In [9], Maliha Hossain et al. proposed an approach to avoid the large computation of large item set by considering only the top selling products such as 30%, 40%, 50%, 55% selling products for both Apriori and FP growth algorithm and generated topmost association rules. Experimental results have shown that if top selling products are used, then almost same frequent item sets and association rules are shown within a short computation time. Author has also concluded that FPgrowth has taken less computation time than Apriori algorithm.

In [10], M. Qisman et al. designed an approach for customer purchasing behaviour analysis using market basket analysis. Author used transactional data of computer retail shop and applied Apriori algorithm using PHP programming language for generating the best association rules. Experimental results

have shown that if a consumer buys a Laptop and Joystick then he would also buy a mouse.

In [11], Hidayat et al. used market basket analysis technique on Breilant store cosmetics data with Apriori and FPgrowth algorithm for generating the best association rules. It is concluded that the combination of products which have strong support and confidence were Original Liquid Bleaching Seeds, Harva Peeling Gel and Castor oil.

In [12], Manpreet Kaura et al. proposed an algorithm that work on static and changing data that helpful in analysing the customer behaviour and assists in increasing the sales. Author concluded that this approach useful in finding the patterns due to changes in the data and tracking the changes in facts from previous data. Moreover, it is useful in predicting future association rules to find out outliers and fraud.

In [13], Yusuf Kurnia et al. used market basket analysis technique with an Apriori algorithm on fish restaurant to find the customers purchase pattern for knowing what products are often purchased simultaneously. The results generated from this study were in the form of website-based application for analysing the customers purchasing behaviour that can be used as recommendations in determining the promotion activities.

In [14], Mustakim et al. proposed a method using market basket analysis with FPgrowth algorithm to generate informative association rules useful for layout and planning of goods availability in Berkah Mart in Pekanbaru, Indonesia. Experiments conducted using FPgrowth algorithm on real data of Berkah Mart has shown that FPgrowth algorithm gives customer purchase behaviour quickly and efficiently.

In [15], Kutuzova Tatiana et al. integrated heterogeneous data sources from a grocery supermarket based on market basket analysis methods for customer purchase behaviour analysis. This analysis has been used for the improvement of the recommendation system for the customer suggestions.

In [16], Fachrul Kurniawan et al. carried the market basket analysis using an Apriori algorithm on transaction data of UIN Malang supermarket, Indonesia and generated 30 best association rules with average confidence value 46.69% and support value 1.78%.

III. METHODOLOGY

This research study have used the transactional data set of Pramod Super Bazar Shop at Latur City in Maharashtra and generated the top best association rules using Market Basket Analysis (MBA) technique with Apriori algorithm in the Weka 3.9.4 software. An Apriori algorithm performs iterative level-wise search where k item sets are used to extract the $(k+1)$ frequent item sets and generates best association rules. This dataset contains operational transactions of different customer shopping at supermarket over four week period of time. Every single bill of a customer has been considered as one transaction that contains different items purchased by the customer. The data pre-processing operations such as data cleaning is done to remove unwanted attributes and records with inappropriate values such as price and quantity from the source data.

The duplication of the items record due its quantity are also removed and kept them as one item with its number of quantity. Then, data transformation operation is done to bring the transactional data into the .ARFF (Attribute-Relation File Format) format and used for conducting an experiment because Weka supports for this file format. Following figure shows the working of proposed methodology.

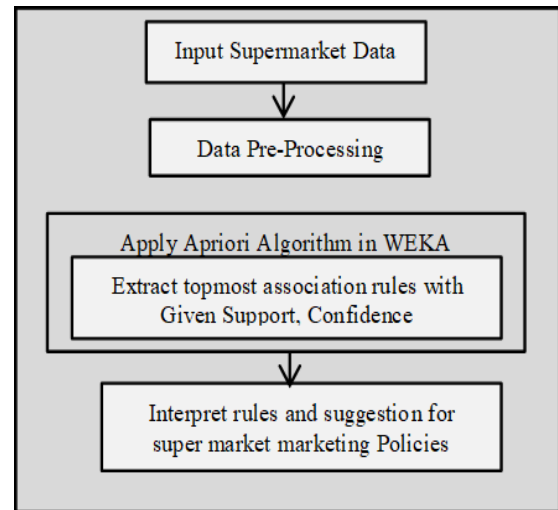


Fig.1: Methodology for generating best association rules

In the proposed methodology, transactional dataset of supermarket has undergone with the pre-processing operations for bringing the dataset into an appropriate .ARFF file format and removing the unwanted attributes and records. Then, Apriori algorithm has been applied on the pre-processed transactional dataset in the Weka software for generating the best association rules using support and confidence measures. Finally, in the last step analysis and interpretations of the best association rules are presented and suggestions are provided to the super market management based on interpretations.

IV. RESULTS AND DISCUSSION

The proposed methodology using Apriori algorithm has been applied on the pre-processed transactional dataset of supermarket into the Weka software and topmost best association rules are extracted with a specific threshold value of support and confidence measures. The strength and accuracy of association rules is be determined by changing the threshold value of support and confidence measures. The higher the value of support and confidence indicates the higher the strength of association rule. If an association rule with a high degree of accuracy or strength is expected then the value of support and confidence measures should be increased. If the value of support and confidence measures is decreased then more number of rules with average degree of strengths is obtained. The best top ten association rules generated after conducting an experiment using Apriori algorithm with the threshold of 50% support and 100% confidence is shown in the following Fig. 2:

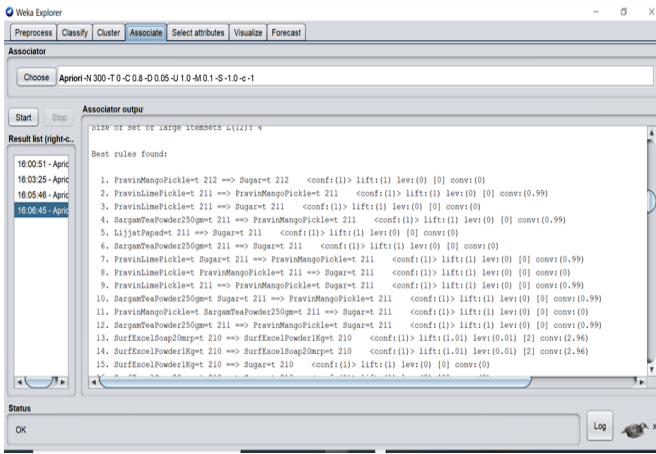


Fig.2: Top best association rules with 100% confidence using Apriori algorithm in Weka software

The association rules shown in the above figure reflects the strong relationship between purchased products by the customer. The rule with highest degree of confidence, lift, and conviction ratio represents the best rule. In above figure, the rule {SurfExcelSoap20mrp→SurfExcelPowder1kg} is with 100% confidence and 1.01 lift indicates that the probability of purchasing SurfExcelSoap20mrp with SurfExcelPowder1kg increases 1.01 times more. The conviction ratio 2.96 indicates the highest rank of the rule among all. Then, the rule {PravinLimePickle→PravinMangoPickle} with {Confidence:(1), Lift:(1) and Conv:(0.99)} indicates that the probability of purchasing PravinLimePickle with PravinMangoPickle increases 1 times more and 0.99 indicates the second highest rank of the rule among all. Then, the rule {PravinLimePickle, PravinMangoPickle→Sugar} with {Confidence:(1), Lift:(1) and Conv:(0.99)} indicates that the probability of purchasing PravinLimePickle, PravinMangoPickle with Sugar increases 1 time more and 0.99 indicates the third highest rank of the rule among all.

If the threshold value of minimum confidence is reduced to 80% and top fifty best association rules using an Apriori algorithm are shown in the following figure 3(a) and 3(b).

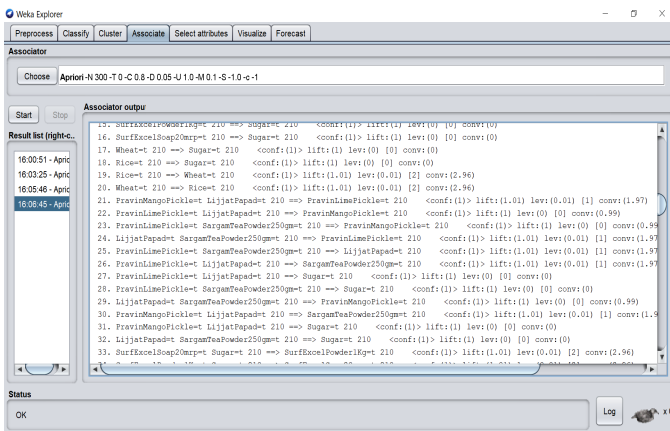


Fig. 3(a):

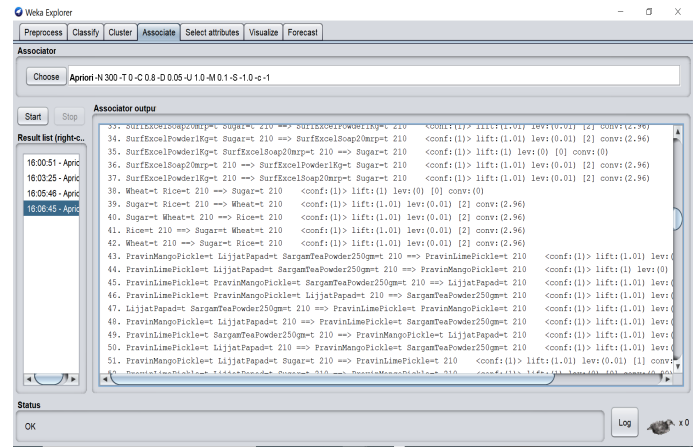


Fig. 3(b):

Fig. 3(a) and 3(b) : Top 50 association rules using Apriori algorithm in Weka software

From the above figure 3(a) and 3(b) it has been observed that the rules with 100% confidence shows the best strength of rule that means probability of purchasing item along with other item. The rules with two item set combinations such as {PravinLimePickle→PravinMangoPickle} and {Wheat→Rice} have shown 100% confidence indicates that if customer purchase PravinLimePickle, Wheat then there is a 100% probability that s/he also purchase PravinMangoPickle, Rice along with the PravinLimePickle and Wheat respectively. The rules with three item set combinations such as {Wheat, Rice→Sugar}, {PravinMangoPickle, PravinLimePickle→LijjatPapad} has shown 100% confidence with 1.01 lift ratio and 2.96 conviction ratio. It indicates that if customer purchases Wheat and Rice then s/he purchases Sugar with 100% confidence and its probability increases 1.01 times more. Likewise if customer purchases PravinMangoPickle and PravinLimePickle then there is a 100% probability that s/he also purchases LijjatPapad. This rule has the 1.01 lift ratio and 1.97 conviction ratio indicates that probability of purchasing PravinMangoPickle, PravinLimePickle with LijjatPapad increases 1.01 time more. Some of the rules with three and four item set combinations with 100% confidence are {PravinMangoPickle,PravinLimePickle→Sugar} ,{PravinMangoPickle,PravinLimePickle→SargamTeaPowder}, {PravinLimePickle,PravinMangoPickle,LijjatPapad→Sugar}, {PravinLimePickle,PravinMangoPickle,LijjatPapad→SargamTeaPowder}. This rule indicates that when customer buys the PravinLimePickle, PravinMangoPickle, and LijjatPapad then there is a 100% guarantee that he also buys the Sugar, SargamTeaPowder along with it. After interpreting best association rules from experimental results, it is been observed that the minimum support and confidence threshold values plays a very important role in filtering the best association rules. If a threshold value is too restrictive then some interesting association rules are not obtained and if threshold value is too relaxed then some non-interesting rules are obtained. Therefore, it is very important to set an appropriate threshold value by experimenting on several combinations of threshold values that provides best association rules from the market basket analysis.

V. CONCLUSION AND FUTURE SCOPE

This research paper has presented the methodology for analysing and predicting the customer's purchasing behaviour or pattern using market basket analysis technique of the data mining. The proposed methodology using Apriori algorithm has been applied on the pre-processed transactional dataset of supermarket in the Weka software for generating the best association rules using a threshold value of minimum support and confidence measures. There are some top and best association rules of two item set combinations with the minimum support and 100% confidence threshold value generated from the conducted experiment such as {SurfExcelSoap20mrp→SurfExcelPowder1kg},{PravinLimePickle→PravinMangoPickle},{Wheat→Rice},{Wheat→Sugar},{Sugar→SurgamTeaPowder},{Rice→Sugar}. Moreover, there are some association rules of three item set combinations obtained with 100% confidence from experiment are {PravinLimePickle,PravinMangoPickle→LijjatPapad},{PravinMangoPickle,PravinLimePickle→Sugar},{Wheat,Rice→Sugar},{PravinMangoPickle,SurgamTeaPowder→Sugar},{SurfExcelPowder,SurfExcelSoap→Sugar},{PravinMangoPickle,LijjatPapad→SurgamTeaPowder}. Based on these results it is concluded that if a customer buys PravinLimePickle, PravinMangoPickle, then s/he will also buy LijjatPapad, Sugar together respectively. Based on this interpretation and conclusion it is recommended that these items can be kept together in racks so that its sales will be increased. It is also concluded from the results that an appropriate threshold value of support and confidence is very much important for filtering the best association rules. Based on the best association rules generated from experiment, its analysis and interpretations, it is suggested to the supermarket management for arrangement of the products shelves, designing the sales promotions and giving offers and discounts on specific products to the targeted customers so that the sales and revenue of supermarket will be increased. The future work of the research is to apply SVM and neural network based methods for extracting more accurate and best association rules from transactional data of supermarket.

REFERENCES

- [1] Anisha Maske, Bela Joglekar, "Survey on Frequent Item-Set Mining Approaches in Market Basket Analysis", Fourth International (IEEE) Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-5, 2018.
- [2] Afrin F Pathan, Snehal Palande, Tejal Shende, Ruchita Patil, Prof. Vitthal S Gutte, "A Study on Market Basket Analysis and Association Mining", Proceedings of National Conference on Machine Learning, pp., 2019
- [3] Jagmeet Kaur, Neena Madan, "Association Rule Mining: A Survey", International Journal of Hybrid Information Technology, Vol.8, No.7, pp.239-242, 2015.
- [4] Savi Gupta, Roopal Mamtorra, "A Survey on Association Rule Mining in Market Basket Analysis", International Journal of Information and Computation Technology, ISSN 0974-2239 Vol.4, No.4, pp. 409-414, 2014.
- [5] Anjana Gosain, Maneela Bhugra, "A Comprehensive Survey Of Association Rules On Quantitative Data In Data Mining", Proceedings of IEEE Conference on Information and Communication Technologies, pp. 1003-1008, 2013.
- [6] Yüksel Akay Ünvan, "Market basket analysis with association rules", Communications in Statistics-Theory and Methods Journal, ISSN: 0361-0926, pp.1-14, 2020.
- [7] Alfiqra, A U Khasanah, "Implementation of Market Basket Analysis based on Overall Variability of Association Rule (OCVR) on Product Marketing Strategy", IOP Conference Series: Materials Science and Engineering, pp.1-8, 2020.
- [8] A. R. Efrat, R. Gernowo, Farikhin, "Consumer purchase patterns based on market basket analysis using apriori algorithms", IOP, Journal of Physics: Conference Series, pp. 1-7, 2020.
- [9] Maliha Hossain, A H M Sarowar Sattar, Mahit Kumar Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm", 22nd IEEE International Conference on Computer and Information Technology (ICCIIT), pp. 1-6, 2019.
- [10] M Qisman, R Rosadi, 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)", IOP Publishing, Journal of Physics: Conference Series, pp. 1-14, 2020.
- [11] Manpreet Kaura, Shivani Kanga, "Market Basket Analysis: Identify the changing trends of market data using association rule mining", International Conference on Computational Modeling and Security, Published by Elsevier Procedia Computer Science, pp. 78 – 85, 2016.
- [12] A. Hidayat, A. Rahman, R. M. Wangi, R. J. Abidin, R. S. Fuadi, W. Budiawan, "Implementation and comparison analysis of apriori and fpgrowth algorithm performance to determine market basket analysis in Breiliant shop", 4th Annual Applied Science and Engineering Conference, Published by IOP Publishing Journal of Physics: Conference Series 1402, pp. 1-6, 2019.
- [13] Yusuf Kurnia, Yohanes Isharianto, Yo Ceng Giap, Aditiya Hermawan, Riki, "Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm", 1st International Conference on Advance and Scientific Innovation (ICASI), IOP Conf. Series: Journal of Physics, pp. 1-6, 2019.
- [14] Mustakim, Della Maulina Herianda et al., "Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau", WMA-Mathcomtech, IOP Journal of Physics: Conf. Series 1114, pp. 1-10, 2018.
- [15] Kutuzova Tatiana, Melnik Mikhaila, "Market basket analysis of heterogeneous data sources for recommendation system improvement", 7th International Young Scientist Conference on Computational Science, Published by Elsevier Procedia Computer Science 136, pp. 246–254, 2018.
- [16] Fachrul Kurniawan, Binti Umayah, "Market Basket Analysis to Identify Customer Behaviour by Way of Transaction Data", Knowledge Engineering and Data Science, Vol. 1, No 1, pp. 20–25, 2018.
- [17] Nurfitriana Tri Utami, Isti Surjandari, "Identifying Consumer Buying Behavior Differences through Market Basket Analysis in Multiple Outlet Types", Association for Computing Machinery, ACM ISBN 978-1-4503-5276-5/17/07, pp. 82-86, 2017.
- [18] Alexander Setiawan et al. "Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company", International Conference on Soft Computing, Intelligent System and Information Technology national Conference on Soft Computing, Intelligent System and Information Technology (ICSII), pp.337-340, 2017.
- [19] Mohamed M. Mostafa, "Knowledge discovery of hidden consumer purchase behaviour: a market basket analysis", International Journal of Data Analysis Techniques and Strategies, Vol. 7, No. 4, pp. 384-405, 2015.
- [20] Wan Faezah Abbas, Nor Diana Ahmad, Nurlina Binti Zaini, "Discovering Purchasing Pattern of Sport Items Using Market Basket Analysis", IEEE International Conference on Advanced Computer Science Applications and Technologies, pp.120-125, 2013.
- [21] M. Hemalatha, "Market basket analysis – a data mining application in Indian retailing" International Journal of Business Information Systems, Vol. 10, No. 1, pp.109-129, 2012.
- [22] Santosh P. Shrikhande, Prashant P. Agnihotri, "Comparative Study of Various Data Mining Techniques for Early Prediction of Diabetes Disease", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol. 8, Issue 1, pp. 287-295, 2022.

- [23] C. Chapman and E. M. Feit, "Association Rules for Market Basket Analysis", R For Marketing Research and Analytics, Published by Springer Nature, pp. 341-362, 2019.
- [24] Yasemin Boztug, Thomas Reutterer, "A combined approach for segment-specific market basket analysis", Elsevier's European Journal of Operational Research, pp. 294-312, 2008.
- [25] Santosh P. Shrikhande, H. S. Fadewar, "Personal Identification using Local and Global Feature of Finger Vein Patterns using SVM Based Classification", International Journal of Computer Sciences and Engineering, Vol.6, Issue-12, pp. 138-145, 2018.
- [26] Ling Zhou, Stephen Yau, "Efficient association rule mining among both frequent and infrequent items", Elsevier's International Journal of Computers and Mathematics with Applications, 54, pp.737-749, 2007.
- [27] Andrej Trnka, "Market Basket Analysis with Data Mining Methods", IEEE, International Conference on Networking and Information Technology, pp.446-450, 2010.
- [28] Loraine Charlet Annie M.C., Ashok Kumar D, "Market Basket Analysis for a Supermarket based on Frequent Itemset Mining", International Journal of Computer Science Issues (IJCSI), Vol. 9, Issue 5, No 3, pp.257-262, 2012.
- [29] Andres Musalem, Luis Aburto, Maximo Bosch, "Market basket analysis insights to support category management", European Journal of Marketing, pp. 1-25, 2018.
- [30] Mu-Chen Chen, l Chuang-Min Chao, Kuan-Ting Wu, "Pattern filtering and classification for market basket analysis with profit-based measures", Expert Systems Journal of Knowledge Engineering, 2011, pp.1-13, 2011.