Market-Basket Optimisation using Sales Pattern of Supermarket

Oyebode E. O. & Agbalaya M. O.

Ajayi Crowther University

ABSTRACT

Finding association among numerous items that are related together can be challenging. At times such association may vary from individuals. Large supermarkets are often faced with such a puzzle which if well addressed can boost or adversely affect the business space, time and profit. In this study, the use of a priority algorithm has been applied using the market pattern of items that are sold as related item in big supermarket like Shoprite etc to optimize item arrangement. We implemented the approach of the algorithm in Python language with hypothetical sales pattern of items from such supermarket, we obtained support, confidence, and lift as criteria from the sales pattern that gave the association rules from the customers.

Market-Basket Optimisation using Sales Pattern of Supermarket

Oyebode E. O.[α] & Agbalaya M. O.[σ]

ABSTRACT

Finding association among numerous items that are related together can be challenging. At times such association may vary from individuals. Large supermarkets are often faced with such a puzzle which if well addressed can boost or adversely affect the business space, time and profit. In this study, the use of a priority algorithm has been applied using the market pattern of items that are sold as related item in big supermarket like Shoprite etc to optimize item arrangement. We implemented the approach of the algorithm in Python language with hypothetical sales pattern of items from such supermarket, we obtained support, confidence, and lift as criteria from the sales pattern that gave the association rules from the customers.

The results of the sales pattern can be applied to rearrange items in the big supermarket for improved packaging, faster sales and resource utilization of such market.

Keywords: association, optimize, support, confidence, lift.

Author α: Computer Science Department, Faculty of Natural Sciences, Ajayi Crowther University, Oyo.
σ: Computer Science Department, Faculty of Natural Sciences, Ajayi Crowther University, Oyo.

I.   INTRODUCTION

Finding association among numerous items that are related together can be challenging. At times, such association may vary from individuals depending on choice. Large supermarkets having numerous goods to display on shelve may not know the best of arrangement on shelve that can showcase and meet with the interest of the buyers. Which product or combination of products to invest on so as to make maximum profit may not be clear. Large supermarkets are often faced with such puzzles which if well addressed can boost the business via space, time and profit management amidst other benefits. Market basket analysis is one of the techniques that can be used to discover association among items in   large retailing (Babu *et al.,* 2021). It can reveal the relationships among entities in list of transactions. Market basket analysis can be used to compute product association to other products, arrange store and product layout for display array of items, decide market trends/inventory, predict future demands of product, formulate new market strategies, promote sales growth, develop merchants' marketing strategies, facilitate product promotion, product region division, slow down the loss and   improve decision making etc (Kurniawan *et al.,* 2021; Alghanam *et al.,* 2022).

Market basket analysis is a form of data mining techniques. Data mining offers methods that can be used to explore meaniful information via its techniques that can be supervised and non-supervised.

Supervised machine learning method requires that data should be preprocessed and made to have instance labels (Osisanwo *et al.,* 2019). For a given sample output, some variables must have contributed to it and their respective contributions ultimately resulted to an output.

Learning with supervised machine algorithm might give weight according to size of the entire dataset and use that for estimation of co-efficients such that when a situation that has resemblance or variables identified as determinant, the result will be determined according to the co-efficient(s) of such determining variables during training.

Classification and predictive models are good examples of supervised learning. Examples of supervised learning models are Naïve Bayes Classifier, Perceptron, Support Vector Machine, Quadratic Classifiers, K-Means Clustering, Boosting, Decision Tree, Random Forest (RF); Neural Networks, Bayesian Networks, Linear Classifiers and Logistic Regression.

Using unsupervised learning method, the sample datasets usually cannot be arranged to determine any sample that will lead to output. Yet some methods are meant to mine association rules that exist in patterns of the dataset. Such rules can then be applied to subsequent future cases to determine the degree of association. Such learning methods are called unsupervised learning. Majority of text mining algorithms are examples of unsupervised learning method. Unsupervised learning algorithms can be used to recognize previously unknown patterns from dataset (Doersch *et al.,* 2015). It is similar to learning from the data itself and finding relation among different inputs and it is also possible not to have useful relationship among the inputs. If there is relationship, how strong are the relationships empirically? These represent the characteristics of unsupervised learning.

II. RELATED LITERATURE REVIEW

Agrawal et al (1993) invented the apriori algorithm for exploring and finding relations from the items in the database. The items in the database are usually represented in a relational form as tables. The items are used to form list of lists. Each combination of item alongside with another form can be used to form new list in which subsets can be generated and intersection is established. Danabhakyam and Punithavalli (2011) conducted study using Classification Dependent Predictive Association Rules (CPAR), Associative Classication, Classication Association Rule Mining (CARM), Distributed Apriori Association Rule, Six Sigma Technique and the Apriori algorithm and made comparism on each of the technique to detect their strength and weakness. The Apriori algorithm was found to be the best even though, it has some drawbacks.

E-commerce requires logistics to appropriately manage enormous client bases and develop long-term and profitable connections (Zineb *et al.,* 2021). This requires logistics to appropriately manage enormous client bases and develop long-term and profitable connections. This requires intelligent customer identification.

E-commerce represents "Big Data", extracting knowledge from it can be challenging due to size. Approaches such as the use of analytical solutions to extract patterns knowledge and establish relationships can be worse. Therefore, association rule mining has been used (Qiu *et al.,* 2015).

Market basket analysis has been applied to diagnosis of patients to find relationship among symptoms of diseases to help improve diagnosis (Awodun and Adedara, 2017). The study used medical data inform of symptoms and diseases to detect the association between group of symptoms and disease. The rules generated in the study provided a method that measured joint frequencies for different symptoms and corresponding disease.

Association rule in data mining operation seeks to find connections among the features of a dataset. From the association rule, association analysis that seeks to study the features that are related to each other can be carried out. Association rule mining forms the basis of market basket analysis (Hamid and Khafaji, 2021). It can be carried out in two phases. In the first phase, all combination of items having support above the minimum support count are first treated. Those identified in the first phase are usually classified as frequent item sets that can be used to generate the desired rules. The association rule has both antecedent (if) and consequent (then). The antecedent represents an item found in the data while the consequent represents an item found in combination using the antecedent.

Formal definition of association rules

$$\text{Let } A = \{a_1, a_2, a_3 \dots a_m\} \text{ be set of items} \qquad (1)$$

$$\text{Let } T = \{t_1, t_2, t_3 \dots t_n\} \text{ be set of transactions} \qquad (2)$$

Transaction t contains X, a set of items in A, if

$$X \sqsubseteq T \qquad (3)$$

An association rule represents an implication of the form

$$X \Rightarrow Y, \qquad (4)$$

where

$$X \subset T, \qquad (5)$$

$$Y \subset T \quad \text{and} \qquad (6)$$

$$X \cap Y = \emptyset \qquad (7)$$

2.1 Properties of Apriori Algorithm

There are two important properties of Apriori algorithm which can be used to reduce the search complexity of the algorithm and the amount of computation involve in the mining process so as to concentrate on the needed frequent item sets. Firstly, suppose X is a frequent item-sets, it means all subsets of X are frequent itemset. Secondly, if Y is a non-frequent item-set, then all supersets of Y are non-frequent itemset. The steps involve in the Apriori algorithm are shown below

2.2 Apriori Algorithm

Input: Transaction Database, min support, min Confidence
Output: Support, Confidence, Lift

Step 1: Tabulate items in database to form list
Step 2: Start with itemsets containing just a single item
Step 3: Formulate list of each item and obtain frequency

Step 4: Compute the support for itemsets
Step 5: Using the itemsets from Step 1, generate all possible itemset combinations.
Step 6: Repeat steps 1 - 4 until there are no more new itemsets.
Step 7: Keep the itemsets that meet the minimum support threshold and remove itemsets that do not support minimum support
Step 8: Extract all the subsets having higher value of support than minimum threshold.
Step 9: Select all the rules from the subsets with confidence value higher than minimum threshold.
Step 10: Order the rules by descending order of Lift.

III. PERFORMANCE EVALUATION

The three Major Components of association rule mining are support, confidence and lift.

Support: Refers to the default popularity of an item and can be calculated by finding number of transactions containing a particular item divided by total number of transactions.

$$\text{Support}(A) = \frac{\textit{Transactions containing } (A)}{\textit{Total Transactions}} \qquad (8)$$

Confidence is an index of reliability of the association rule. Confidence refers to the likelihood that an item B is also bought if item A is bought. It can be calculated by finding the number

of transactions where A and B are bought together, divided by total number of transactions where A is bought.

$$\text{Confidence}(A \rightarrow B) = \frac{(Transactions\ containing\ both\ (A\ and\ B))}{(Transactions\ containing\ A)} \qquad (9)$$

Lift(A -> B) refers to the increase in the ratio of sale of B when A is sold. Lift(A –> B) can be calculated by dividing Confidence(A -> B) divided by Support(B).

$$\text{Lift}(A \rightarrow B) = \frac{(Confidence\ (A \rightarrow B))}{(Support\ (B))} \qquad (10)$$

The dataset used in this study contains the record of transaction of customers in a supermarket including the names of various items purchased by the customers. There are twenty different products and seven thousand five hundred transactions [https://raw]. The apriori algorithm was used to convert the items into list, extract frequent item set and find subsets that meet up with minimum support count, we were able to obtain various lists such as the number count for all items. The screen shot is shown in Figure 1.0. The sample count is shown in Figure 1.1. We were also able to use the model to obtain the association of items containing 2 items, three, four etc. Figure 1.2 shows the sample results obtained using association rule containing 2 items. While Figure 1.3 shows the confidence, support and lift obtained from the model. We vary the support probability from 0.01 to 0.15 and obtain the plot for execution times. The plot is shown in Figure 1.4. The figure shows the reduction in execution time as the support value increases, there is reduction in the number of association rule because of increasing value of support.

	items	incident_count	all
0	mineral water	1787	all
1	eggs	1348	all
2	spaghetti	1306	all
3	french fries	1282	all
4	chocolate	1230	all
...	...	...	...
115	bramble	14	all
116	cream	7	all
117	napkins	5	all
118	water spray	3	all
119	asparagus	1	all

Figure 1.0: Association Rule for Two Item Sets

Market-Basket Optimisation using Sales Pattern of Supermarket

	support	itemsets	length
23	0.059573	(, a)	2
27	0.080373	(, e)	2
34	0.056340	(, n)	2
36	0.072653	(, r)	2
38	0.056087	(t,)	2
45	0.093927	(e, a)	2
51	0.052973	(l, a)	2
53	0.849033	(n, a)	2
56	0.052827	(r, a)	2
57	0.055747	(s, a)	2
58	0.075867	(t, a)	2

Figure 1.1: Association rule for two item sets

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	levera
64885	(o, r, b, , n)	(g, e, f)	0.011547	0.010707	0.010360	0.897229	83.800931	0.0102
64978	(g, e, f)	(o, r, b, , n)	0.010707	0.011547	0.010360	0.967621	83.800931	0.0102
64912	(g, e, r, f)	(, n, o, b)	0.010707	0.011547	0.010360	0.967621	83.800931	0.0102
47910	(, n, o, b)	(g, e, f)	0.011547	0.010707	0.010360	0.897229	83.800931	0.0102
64951	(, n, o, b)	(g, e, r, f)	0.011547	0.010707	0.010360	0.897229	83.800931	0.0102
...	...	...	...	...	...	...	...	
1237	(g)	(c, e)	0.066073	0.061507	0.010053	0.152154	2.473783	0.0059
2315	(t, l)	(s)	0.042880	0.098067	0.010147	0.236629	2.412944	0.0059
6901	(t, l, a)	(s)	0.042880	0.098067	0.010147	0.236629	2.412944	0.0059
2318	(s)	(t, l)	0.098067	0.042880	0.010147	0.103467	2.412944	0.0059
6910	(s)	(t, l, a)	0.098067	0.042880	0.010147	0.103467	2.412944	0.0059

Figure 1.0: Support, Confidence and Lift

Market-Basket Optimisation using Sales Pattern of Supermarket
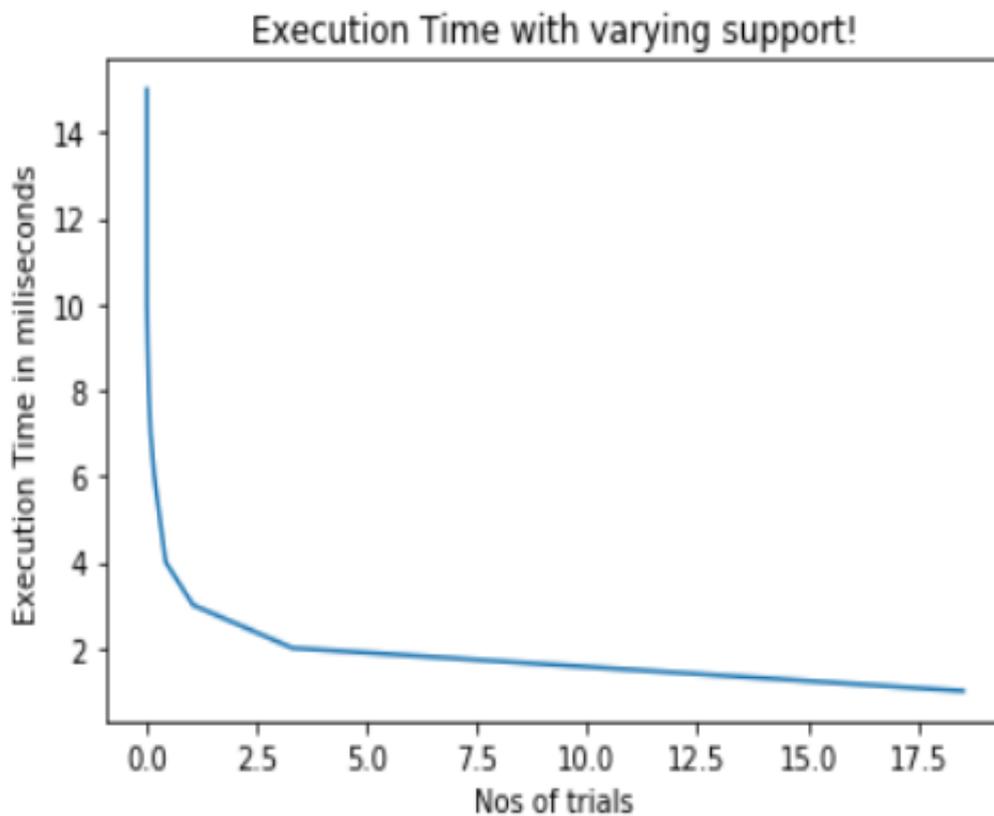
IV. CONCLUSION

Databases are used in storing transaction records in form of relational databases in many supermarkets. Physical attempt to mine useful patterns out of such transactions can be difficult and time consuming, but useful patterns of information can be obtained through the use of association rules for illustrating the relationships in the database. This study has applied the association rules and provided various answers to queries that can be used to boost decision making in supermarkets with large number of items and customers alongside with execution time of such results of the decision. The study showed that large supermarkets have a lot to benefit using the association rule.

REFERENCES

1. Babu S. K., Kavya, M., Kumari, K. R. and Swarupa, J. 2021. Market Basket Analysis using Datamining. International Journal for Modern Trends in Science and Technology 7, pp. 291-296. https://doi.org/10.46501/IJMTST0707051.

2. Hamid, Z. and Khafaji H. K. (2021). A General Algorithm of Association Rule-Based Machine Learning Dedicated for Text Classification. Journal of Physics: Conference Series 1-11. https://raw.githubusercontent.com/gitganeshnethi/Datasets/master/store_data.csv.

3. Qiu, J. Lin, Z. and Li, Y. (2015). Predicting customer purchase behavior in the e-commerce context, Electronic commerce research, vol. 15, no. 4,pp. 427–452.

4. Dhanabhakyam, M. and Punithavalli, M. (2011). A survey on data mining algorithm for market basket analysis,ʻ Global Journal of Computer Science and Technology 11(11): 23-28.

5. Kurniawan, F., Umayah, B., Hammad, J., Nugroho, S. M. S. and Hariadi, M. (2021). Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data. Knowledge Engineering and Data Science (KEDS). 1(1):20–25.

6. Zineb, E. F., Rafalia, N. and Abouchabaka, J. 2021. An Intelligent Approach for Data Analysis and Decision Making in Big Data: A Case Study on E-commerce Industry. International Journal of Advanced Computer Science and Applications (IJACSA). 12(7):723 - 736. NU

7. Alghanam, O. A., Al-Khatib, S. N. and Hiari, M. O. (2022). Data Mining Model for Predicting Customer Purchase Behavior in e-Commerce Context. International Journal of Advanced Computer Science and Applications. 13(2):421-428.

8. Awodun, M. A. and Adedara, R. O. 2017. An Efficient Rule-Mining for Medical Diagnosis: A Market-Basket Approach. International Journal of Research and Analytical Reviews 4(1):96-107.

9. Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O. and Akinjobi J. Supervised Machine Learning Algorithms: Classification and Comparison. International Journal of Computer Trends and Technology (IJCTT) – Volume 48(3):128-138.

10. Doersch, C., Gupta, A., and Efros, A. A. 2015. Unsupervised visual representation learning by context prediction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1422–1430.