



Design of Machine Learning Framework for Products Placement Strategy in Grocery Store

Olasehinde Olayemi O.
Department of Computer Science,
Federal polytechnic, Ile Oluji
Ondo State, Nigeria

Abiona Akeem A.
Department of Computer Science,
Federal polytechnic, Ile Oluji
Ondo State, Nigeria

Ibiyomi Michael A.
Department of Computer Science,
Federal polytechnic, Ile Oluji
Ondo State, Nigeria

ABSTRACT

The well-known and most used support-confidence framework for Association rule mining has some drawbacks when employ to generate strong rules; this weakness has led to its poor predictive performances. This framework predicts customers buying behavior based on the assumption of the confidence value, which limits its competent at making a good business decision. This work presents a better Association Rule Mining conceptualized framework for mining previous customers transactions dataset of the grocery store for the optimal prediction of products placed on the shelves, physical shelf arrangement and identification of products that needs promotion. Sampled transaction records were used to demonstrate the proposed framework. The proposed framework leverage the ability of lift metric at improving the predictive performance of Association Rule Mining. The Lift discloses how much better an association rule predicts products to be placed together on the shelf rather than assuming. The proposed conceptualized framework will assist retailers and grocery store owners to easily unlock the latent knowledge or patterns in collected grocery's store transaction dataset to make important business decisions that will make them competitive and maximize their profit margin.

Keywords

Association Rule Mining, Market Basket Analysis, Frequent Itemset, Support, Confidence, Lift

1. INTRODUCTION

Grocery stores are stores that involve in the primary sales of a general range of food products and daily needs. [1] Identified Cereals, Toothpaste, Beer, Butter, Cake Mix, Chips, Cookies, Facial Tissues, Laundry Detergent, Loaf Bread, Toilet paper, and Coffee, to be the twelve categories of products in a grocery store. These categories are selected for purchases based on certain parameter, which include price, always buy, satisfaction, recommendation, brand name, shelf space. Retailers regularly are faced with the challenges of allocating products to shelves due to shelf space being a scarce resource in retail stores and needs to increase the no of products to be included in the assortment [2]. Product shelving allocates products on the shelves in an optimized way that will maximize sales and profit. According to [3]. Products shelving tremendously affect consumer buying behaviors. Efficient allocation of product on shelves curtail the economic threats of unfilled product shelves, improves consumer satisfaction, healthier consumer relationship [4], and improve product sales [5].

Product shelving is a modern-day marketing strategy for products to get to end-users without using overt traditional

advertising. Product placement is becoming an increasingly important way for brands to reach their target audience in subtle ways. Businesses are exploiting product shelving to enhance brand awareness, increase sales and draw in customers without traditional marketing, shelving strategies are the various methods of arrangement of products on the shelves to induce impulse purchases and thereby increase sales and profit margin of the retailers. An ingenious display of products on shelf will increase customer's purchase decisions, which habitually influenced in-store factors [6]. The way customers pick items to purchase on shelves is based on certain behavioral patterns and factors. Analytic of the past consumer purchasing behavior's record using Machine Learning (ML) algorithms will enhance the store's overall profitability [7].

ML is an aspect of artificial intelligence that learns with the aids of an algorithm from data to obtain knowledge or pattern from it to make a decision without human intervention. ML automates the process of data analysis for model building. ML's goal is to make an excellent guess useful to the predictive (classification) problem [8]. Supervised ML algorithms extract valuable knowledge from the mapping of supplied inputs and its desired output (class label) of the training dataset, then validates the testing dataset's obtained knowledge. Regression and classification are examples of supervised ML techniques. Unsupervised learning draws knowledge from a dataset consisting of input data without label responses. It partitions the dataset into clusters based on similarities that exist among the dataset. It validates by assigning a new test instance into the appropriate cluster; clustering analysis and association rule mining are examples of unsupervised learning methods.

Association Rule Mining (ARM) is rule-based ML algorithm for the discovering of interesting relationship among entities of a transactional dataset, ARM aim to identify patterns (combinations of events that occurred together) of entities in a transaction that frequently appear together among the whole transaction dataset. It generates rules that summarize these patterns and uses the generated rules to predict the presence of one or more products based on the occurrences of some products in a new transaction. Products that are capable of influencing the presence of other products in the transaction are predicted to be placed together on the shelf to create impulse purchases. Grocery store generates a lot of data on daily basis from customer's transactions; this dataset contains hidden knowledge and patterns that can be used to make important business intelligent decision, unlocking this knowledge and patterns remains a mirage to several grocery stores, provision of a framework for discovering latent pattern or knowledge from the transactional dataset will help grocery

store's owners to make an important business decision that will make them competitive and maximize their profit margin. This work presents an Association Rule Mining framework for mining previous transactions of consumers' buying patterns for the optimal prediction of products placed on the shelves, physical shelf arrangement, and identification of products that needs promotion.

2. REVIEW OF RELATED LITERATURE

Several authors have applied Association Rule Mining algorithms to provide solutions to different problems; Olasehinde et al. (2018) applied ARM to mine customers buying behavior to improve customers relationship management, results from the research suggest products that should be shelved close to each other, products that need promotion and products that promotion will not improve it sales [9].

[10] Applied ARM to extract knowledge from the Market Basket Analysis (MBA) to predict products that will be bought together and hence be placed close to each other on the shelf to induce and increase impulse buying. Serban et al. proposed the application of relational ARM to predict the probability of certain diseases and predicts likely therapy [11]. Gupta et al. adopted ARM to determine the relationship among sequence of protein [12]. The research in [13] applied ARM to the analysis of huge supermarket data exploiting the customer behavior to make market competitive decisions. Liu et al. (2007) applied ARM to generate important rules to extract strategic Business Intelligence (BI) from the mining of organization transactions. The experimenter results from the application of ARM to records of business transactions and customer data analysis show interesting patterns for customer satisfaction and improvement of quality of service and profit [14]. [15] Applied ARM to determine the probability of purchases in online stores; result from this work shows that customers that have spent 10 to 25 minutes in an online book store and have opened thirty to seventy pages has a probability of 92% to confirm a purchase. The work in [16] applied ARM to the historic customer's transaction data from a grocery store to segment customers for targeted marketing. [17] conducted a research on Market Basket Analysis, Apriori Algorithm was used to discover frequent item sets among products stored in a large database, rules generated from this work were used to cluster customers based on their buying patterns and further subjected to selective marketing

3. ASSOCIATION RULE MINING

Association mining concerns the discernment of rules that cut across a good percentage of the dataset [18]. Given a set of transactions, T, the goal of ARM is to find all rules that predict products to be placed closer to each other on the shelf and products that need promotion. ARM involves two stages; in the first stage, frequent itemsets from the transaction dataset are generated that satisfied the predefined minimum support level. The second stage involves the generation of association rules that satisfied the minimum user's defined confidence rate among the frequent item-sets. Item-sets are one or more products in each record of the transaction dataset. A frequent itemset is a pattern that occurred frequently than a predefined threshold [19], frequent itemsets are products combinations that satisfied the user predefined minimum

support. All subsets of a frequent itemsets are also frequent itemsets, while subsets of infrequent itemsets are infrequent item-sets. ARM is defined as follow:

$$\text{let } P = P_1, P_2, \dots, P_n \quad (1)$$

Be a set of n binary attributes called products.

$$\text{let } D = T_1, T_2, \dots, T_n \quad (2)$$

be set of all possible transactions D.

where each transaction T_i is a set of products such that $T_i \subseteq P$

Each transaction in D has a unique transaction ID and contains a subset of the products in P. A rule is defined as an implication of the form $X \Rightarrow Y$ interpreted as X implies Y, where

$$X, Y \subseteq I \text{ and } X \cap Y = \emptyset \quad (3)$$

To select interesting rules for optimal product placement strategy, Support and confidence constraints are applied to all the generated rules from the transaction dataset.

The support often expressed as a percentage of the total number of transactions in the dataset is basically the number of transactions that include all items in the antecedent and consequent parts of the rule [20]. The support of item-set containing products X and Y [21], written as $\text{supp}(X \Rightarrow Y)$ is the ratio of number of the transactions that contains item-set X and Y to the total number of the transaction in the dataset as shown in equation 4. Support of 0.75 for itemset X implies that 75% of the whole transactions in the dataset contain itemset X. Itemsets that satisfied the minimum support threshold value are considered to be frequent.

$$\text{supp}(X \Rightarrow Y) = \frac{(X \cup Y)}{\text{Total No of the trasactions}} \quad (4)$$

The confidence of a frequent itemset (rule) is the percentage of all transactions that contain all products in both the consequent and the antecedent of the rule to the number of transactions that contain products in the antecedent [20]. The confidence of a frequent itemset (rule) $X \Rightarrow Y$ is a conditional probability that Y will occur whenever X occurred [22], its the ratio of the support of $X \cup Y$ to support of X given in equation 5. The implication of the confidence of a rule $X \Rightarrow Y$ to be 0.90 implies that, 90% of customers that buys X also buys Y.

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)} \quad (5)$$

3.1 Proposed Framework for Association Rules Products Shelving Strategy

The proposed framework for the products placement (shelving) strategy is based on the horizontal dataset layout. Basically the framework consists of the following major components, as shown in Figure 1

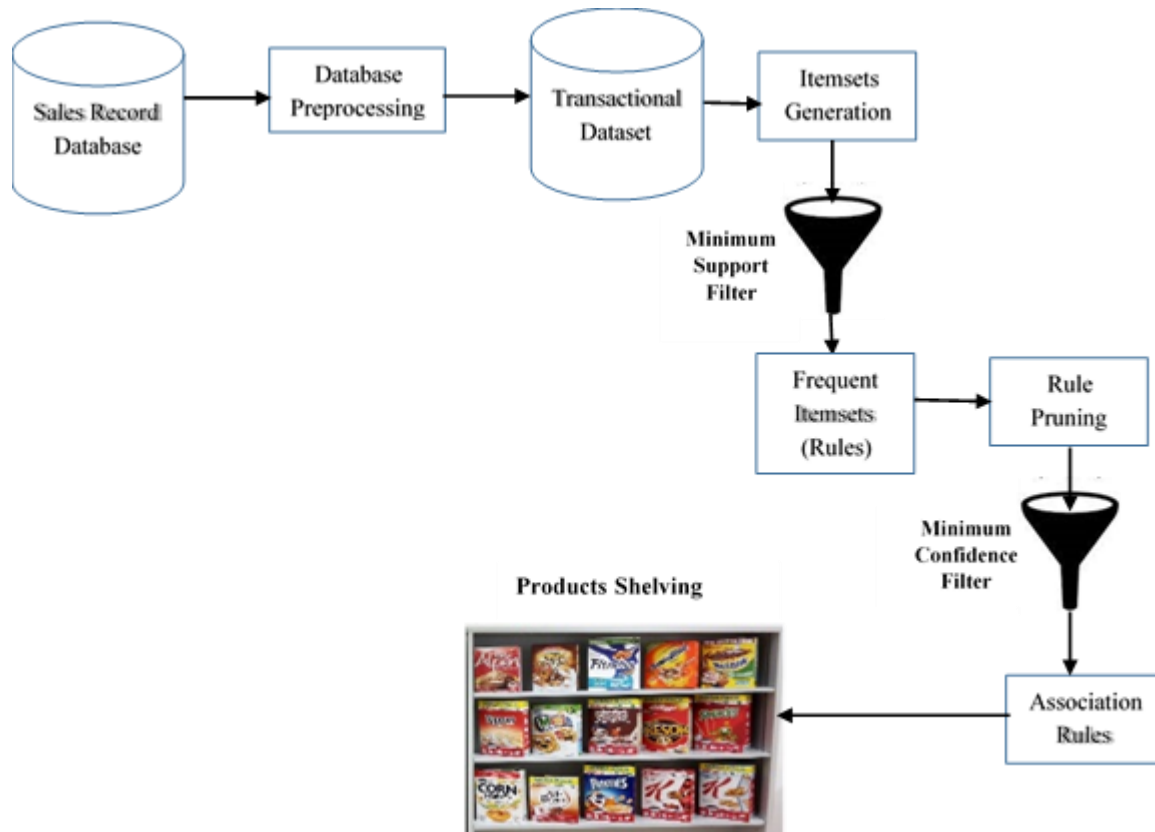


Fig 1: Proposed Framework for Association Rule Product Shelving Strategy

3.1 Sales Record Dataset

Availability and quality of data is the bedrock of any successful database modeling projects; availability involves the existence of a relevant and suitable quantity of data; while the quality of data involves the fitness of data for the purpose for which it is intended for. High quality data are free of redundancy, defects and possess desired features fit for the modeling purpose. Grocery stores generate and store a large amounts of data on a daily basis; extraction of sensitive information implicitly contained in data will provide a lot of direct benefits to the store’s owners. To obtain desirable results of great benefits to the store, the data must be large enough to represent all the possible patterns of events in the store; a transaction data of 6 to 24 months is recommended in order to provide good and effective decision that will benefit the grocery store [23]. Sales record contains many items such as transaction ID, customer’s name, customer’s ID, Product(s) bought, quantity bought, data and time of sales, unit price, product(s) code, Receipt number, product description.

3.2 Sales Record Database Preprocessing

All the constitutes of the sales record are not relevant for the modeling of the products placement strategy; there is a need to select the relevant constitute and represent them appropriately for ARM algorithms to be able to model them. Data preprocessing is critical to a successful data modeling process, the presence of missing data, noise and irrelevant attributes will degrade the quality of the modeling results. For products placement strategy, transaction ID and the list of products in the transaction are the two relevant attributes; Table 2 shows a sampled preprocessed of a sales record containing five transactions depicted in Table 1, each row of

the sales record represent a transaction, and each column represent a product (an item). The present of an item in every transaction is represented with 1’s while 0’s implies an absence of a product in the trasaction

Table 1: Horizontal Representation of Transaction Record

Transaction ID	Transaction Details
T1	{Bread, Egg}
T2	{Milk, Bread, Egg}
T3	{Bread, Butter, Egg}
T4	{Bread, Butter}
T5	{Milk, Bread, Butter, Egg}

Table 2; Sampled Preprocessed Five Transactions

Transaction ID	Milk	Bread	Butter	Egg
1	0	1	0	1
2	1	1	0	1
3	0	1	1	1
4	0	1	1	0
5	1	1	1	1

3.3 Itemset Mining

Itemsets are one or more than one products bought together by customers and recorded in the transaction dataset, Itemsets mining, the process of determining itemsets in the transaction dataset was first introduced by [24] in 1993, and it is nowadays called Frequent Itemset Mining (FIM). Frequent itemsets are the pattern that transpired repeatedly than the predefined verge denoted as L_K , where K is the no of elements in the itemset. FIM mines a group of items bought regularly together. Any itemset X with its frequency of occurrences in the transaction dataset is more than the predefined user verge known as minimum support threshold (i.e; $\text{sup}(X) \geq \text{minsup}$)

is called frequent itemset. In a transaction dataset with n distinct items (products), there will be $2^n - 1$ possible itemsets. The five transactions represented in Table 1 have four distinct items; {Milk, Bread, Butter, and Egg} with possible itemsets = $2^4 - 1 = 15$ itemsets. The fifteen itemsets from Table 1 with their support are ;{ Bread}: 1.0, it appeared in all the five transactions in the dataset, its support is $5/5 = 1.0$, {Egg}: 0.8, {Milk}: 0.4, {Butter}: 0.6. {Milk and Bread}:0.4, {Milk and Butter}:0.2, {Milk and Egg}:0.4, {Butter and Egg}:0.4, {Bread and Egg}:0.8, {Bread and Butter}:0.6, {Milk, Bread and Butter}: 0.2, {Milk, Bread and Egg}:0.4, {Bread, Butter and Egg}:0.4, {Milk, Butter and Egg}:0.2, {Milk, Bread, Butter and Egg}: 0.2.

Given a user-defined minimum support of 0.5, itemsets that have it support equals or greater than 0.5 will be filtered as the frequent itemsets; from Table 1, the itemsets that meet the minimum support threshold (0.5) set by the user are: {Bread}:1.0, {Egg}: 0.8, {Butter}: 0.6. {Bread and Egg}:0.8, {Bread and Butter}:0.6. Considering all conceivable itemsets, the mining of frequent itemsets is huge, naive, time-consuming, expensive in terms of computer resources employed and not efficient particularly when the number of items under consideration are many. Efficient way to mine frequent itemsets is via the design of algorithms that circumvent exploring the search space of all conceivable itemsets and analyses each itemset in the search space as efficient as possible.

The first algorithm used to mine frequent itemsets and association rules was Artificial Immune System (AIS) algorithm proposed by [25], improvement on AIS renamed as Apriori [24], other algorithms proposed for FIM include , Frequent pattern (FP) Growth algorithm [26], Equivalence Class Transformation (ECLATt) [27], Hyper-links Mine [28], Linear time Closed Mining (LCM) [29] and SET-oriented Mining (SETM) [30]. Apriori algorithm has been a predominantly implemented algorithm for mining frequent itemsets, but it is not efficient in its high overhead and consumption of the computer resources, an improvement to overcome it inefficiency was proposed in vertical representation of its dataset, AprioriTID [31] improve the efficiency of Apriori by avoiding multiple scan of the dataset during it valuation process. All these algorithms employ different strategies and data structures to discover frequent itemsets efficiently. According to [32], FIM algorithms differs in the following areas;

- 1) Mode of dataset representation, and how to compute minimum support
- 2) Search Space techniques, such as Depth-first or Breadth-first search and how it determine the next itemsets to explore in the search universe

The two dataset representation formats used in FIM algorithms are Horizontal and vertical data format, horizontal format is presented in Table 1, it represents each transactions by its transaction ID, the vertical format is depicted in Table

3, it represents transactions with same items together, horizontal format can be easily converted to vertical format, the vertical format is more effective than horizontal format, it scan the dataset once to compute the support for each itemsets, it is faster than horizontal format in computing the support, but it also required more computer memory space to store the transactions ID. FIM algorithms employs Breadth-first and Depth-First search to mine frequent itemsets, Breadth-First search (BFS) explore all available nodes and select the shortest path between the starting node and other nodes, its memory consumption is higher than the Depth-First Search (DFS). in Breadth-first Search, the algorithm first evaluate single itemsets {Bread}, {Milk},{Butter}, {Egg}, then itemsets with two itemsets such as{{Milk and Bread}, {Milk and Butter}, {Milk and Egg}, {Butter and Egg}, {Bread and Egg}, {Bread and Butter}, follows by three elements, {Milk, Bread and Butter}, {Milk, Bread and Egg}, {Bread, Butter and Egg}, {Milk, Butter and Egg}and so on until all the number of items has been generated. On the other hand, depth-first search explore itemsets starting with single itemset and then recursively append items to the existing itemset to create another itemset, in the following order: {Milk}, {Milk, Bread}, {Milk, Bread, Egg}, {Milk, Bread, Butter}, {Milk, Bread, Butter, Egg}, {Milk, Butter}, {Milk, Butter, Egg}, {Milk, Egg}, {Bread}, {Bread, Egg}, {Bread, Butter}, {Bread, Butter, Egg}, {Butter},{Butter, Milk},{Butter, Egg}, {Egg}. Table 4 depletes the features of some FIM algorithms.

Table 3: Vertical Representation of Transactions in Table 1

Itemsets	Transaction ID
Milk	T2, T5.
Bread	T1, T2, T3, T4, T5.
Butter	T3, T4, T5.
Egg	T1, T2, T3, T5.
Milk and Bread	T2, T5.
Milk and Butter	T5.
Milk and Egg	T2, T5.
Bread and Butter	T3, T4, T5.
Bread and Egg	T1, T2, T3, T5.
Butter and Egg	T3, T5.
Milk, Bread and Butter	T5.
Milk, Bread and Egg	T2, T5.
Bread, Butter and Egg	T3, T5.
Milk, Butter and Egg	T5.
Milk, Bread, Butter and Egg	T5.

Table 4: Features of Frequent Itemsets Mining Algorithms

Algorithms	Search Methods	Dataset Representation
AIS [25]	BFS (Candidate generation)	Horizontal
Apriori [24]	BFS (Candidate generation)	Horizontal
Apriori TID [31]	BFS (Candidate generation)	Vertical (TID)
SETM [30]	BFS (Candidate generation)	Horizontal (Sql)
ECLAT [27]	DFS (Candidate generation)	Vertical (TID-List)
FP-GROWTH [26]	DFS (Pattern Growth)	Horizontal (Prefix-tree)
H-MINE [28]	DFS (Pattern Growth)	Horizontal (Hyperlink Structure)
LCM [29]	DFS (Pattern Growth)	Horizontal (transaction merging)

3.4 Association Rules Generations

Association Rules (AR) generation in ARM involves two stages, in the first stage, frequent itemsets were generated, while the second stage has to do with creation of all possible rules from each of identified frequent itemsets that satisfied the minimum confidence threshold. AR are conditional probability that indicate the likelihood of a customers to buy a certain product provided if he or she had bought another product in the same purchase. AR is of the form $\{X \Rightarrow Y\}$ has two part; the antecedent and the consequent, X is the antecedent (if) and Y (then is the consequent. Antecedent are items found within the data while consequent are items found in combination with the antecedent. AR are created from binary partitioning of each itemsets, the following binary rules will be generated from {Bread, Egg, Milk} frequent itemset; {Bread \Rightarrow Egg}, {Bread \Rightarrow Milk}, {Bread \Rightarrow Egg, Milk}, {Egg \Rightarrow Bread}, {Egg \Rightarrow Milk}, {Egg \Rightarrow Bread, Milk}, {Milk \Rightarrow Egg}, {Milk \Rightarrow Bread}, {Milk \Rightarrow Bread, Egg}, {Bread, Egg \Rightarrow Milk}, {Bread, Milk \Rightarrow Egg}, {Egg, Milk \Rightarrow Bread}, etc. The total number of possible binary rules R, generated from an itemset with d no of items is given in equation 6

$$R = 3^d - 2^{d+1} + 1 \quad (6)$$

AR generate a lot of rules, most these rules are not relevant and important, to prune the rules and obtain important rules, confidence of the each rule are computed using equation 5 based on the user defined minimum confidence threshold filter. AR that does not meet the minimum confidence threshold will be discarded. Note that the confidence of the rule {Bread \Rightarrow Egg} may not be same with the confidence of rule {Egg \Rightarrow Bread}.

From Table 1, the itemsets that meet the minimum support threshold (0.5) set by the user are: {Bread}:1.0, {Egg}: 0.8, {Butter}: 0.6. {Bread and Egg}:0.8, {Bread and Butter}:0.6. Given a user defined confidence of 60% (0.6). The number of AR with their support and confidence values are listed below;

- Rule 1: {Bread \Rightarrow Egg}, support: 0.8, confidence: 0.8
- Rule 2: {Egg \Rightarrow Bread}, support: 0.8, confidence: 1.0
- Rule 3: {Bread \Rightarrow Butter}, support: 0.6, confidence: 0.6
- Rule 4: {Butter \Rightarrow Bread}, support: 0.6, confidence: 1.0

The rules are interpreted as follows, in rule 1, 80% of

customers that bought Bread also bought Eggs. In rule 2, all the customers that bought Egg also bought Bread. 60% of customers that bought bread in rule 3 also bought Butter, while all the customers that bought butter in rule 4, also bought Bread. Rules that satisfied the minimum support and confidence threshold are strong rules. Often, an AR with high confidence implies a strong rule, this can be misleading and deceptive when the antecedent and/or the consequent have a high support. Whenever the consequent of any AR is very frequent, its confidence will high. High confidence may be misleading at times, and does not always implies strong rules.

Lift ratio is a better metric to measure the strength of AR, it is the ratio of confidence of a rule to the expected confidence a rule. The expected confidence of a rule is probability of buying the consequent of the AR without any knowledge about antecedent. The lift ratio of AR $(X \Rightarrow Y)$ is given in equation 7.

$$Lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{conf(Y)} \quad (7)$$

A Lift value greater than one (1) implies positive association (correlation) between the antecedent and consequent of the AR, it implies that if a customer buy products in the antecedent there is great chances that products in the consequent will also be bought also. A lift value less than one (1) implies negative association between the antecedent and consequent of the AR, lift value of one (1) indicates no association between the antecedent and consequent of the AR. Applying Equation 7 to Table 1 gives the following lift values for the Rules 1, 2, 3 and 4.

- Rule 1: {Bread \Rightarrow Egg}, support: 0.8, confidence: 0.8, lift: 1.0
- Rule 2: {Egg \Rightarrow Bread}, support: 0.8, confidence: 1.0, lift: 1.25
- Rule 3: {Bread \Rightarrow Butter}, support: 0.6, confidence: 0.6, lift: 1.0
- Rule 4: {Butter \Rightarrow Bread}, support: 0.6, confidence: 1.0, lift: 1.25

The values of the lift of the rules above shows that there is no association between the rules {Bread \Rightarrow Egg} and {Bread \Rightarrow Butter}, while there is a positive correlation between the antecedent and the consequent of rules {Egg \Rightarrow Bread} and {Butter \Rightarrow Bread}, with 25% more chances of buying the antecedent and the consequent products together. Considering



the confidence of an AR alone will limit the competency of making good business decision, The Lift discloses how much better an AR is at predicting products to be placed together on the shelf rather than assuming, confidence assumes, *Lift* is a measure that assist store managers to determine the products to be placed together on shelf.

4. CONCLUSION

The vast amount of transaction dataset being generated by grocery store remain useless unless the latent knowledge and patterns hidden in it is unlock and discovered. Discovered latent pattern or knowledge from transactional dataset will help grocery store's owners to make important business decision that will make them competitive and maximized their profit margin. The well-known and most used support-confidence framework for Association Rule Mining has some drawbacks when employ to generate strong rules, this weakness has led to it poor predictive performances. This framework predict customers buying behavior based on the assumption of the confidence value, which limits it competent at making good business decision. This work presents a better Association Rule Mining framework for mining data of previous transactions of consumers' buying patterns for the optimal prediction of products placement on the shelves, physical shelf arrangement and identification of products that needs promotion. The proposed framework leverage on the ability of lift metric at improving the predictive performance of association rule mining. The Lift discloses how much better an AR is at predicting products to be placed together on the shelf rather than assuming, confidence assumes. *Lift* is a measure that assist store managers to determine the products to be placed together on shelf. The proposed framework will assist retailers and grocery store's owners on products placement on the shelves, physical shelf arrangement and identification of products that needs promotion

5. REFERENCES

- [1] Hoyer, W.D & Walgren, C.J.C. (1988). Consumer Decision Making Across Product Categories: The Influence of Task Environment. John Wiley & Sons Inc., 5(1) 45-69
- [2] Hübner, A. H. and Kuhn, H. (2012). Retail category management: State-of-the-art review of quantitative research and software applications in assortment and shelf space management. *Omega*, 40(2):199–209
- [3] Dreze, X., Hoch, S. J., & Purk, M. E. (1994). Shelf management and space elasticity. *Journal of Retailing*, 70, 301–326.
- [4] Fancher, L. A. (1991). Computerized space management: A strategic weapon. *Discount Merchandiser*, 31(3), 64-65.
- [5] Hwang, H., Choi, B., & Lee, M. J. (2005). A model for shelf space allocation and inventory control considering location and inventory level effects on demand. *International Journal of Production Economics*, 97(2), 185-195.
- [6] Hübner, A. H. and Schaal, K. (2017). An integrated assortment and shelf-space optimization model with demand substitution and space-elasticity effects. *European Journal of Operational Research*, 261(1):302–316.
- [7] Lufás F. M. S. (2018). Optimizing Shelf Space Allocation under Merchandising Rules. A Master's Dissertation submitted University of Porto.
- [8] Olasehinde O. O. (2020). A Stacked Ensemble Intrusion Detection Approach for Security of Information System, *International Journal for Information Security Research (IJISR)*, 10(1).
- [9] Olasehinde, O.O., Williams, K.O. and Ajinaja, M.O. (2018): Application of Association Rule Learning in Customer Relationship Management. Proceedings of the 14th iSTEAMS International Multidisciplinary Conference, AlHikmah University, Ilorin, Nigeria, 14: 29-36
- [10] Sherdiwala B., Khanna O. (2015). Association Rule Mining: An Overview", *International Multidisciplinary Research Journal (RHIMRJ)*.
- [11] Serban G., Czibula I. G. and Campan A. (2016). A Programming Interface For Medical diagnosis Prediction", *Studia Universitatis, "Babes-Bolyai", Informatica, LI(1)*, pages 21-30, 2006.
- [12] Gupta N., Mangal N, Tiwari K. and Mitra P. (200). Mining Quantitative Association Rules in Protein Sequences", In Proceedings of Australasian Conference on Knowledge Discovery and Data Mining –AUSDM, 273-281, 2000
- [13] Raorane A.A., Kulkarni R.V. and Jitkar B.D. (2012). Association Rule – Extracting Knowledge Using Market Basket Analysis, *Research Journal of Recent Sciences*, 1(2): 19-27.
- [14] Liu, H., Su, B., & Zhang, B. (2007). The Application of Association Rules in Retail Marketing Mix. *2007 IEEE International Conference on Automation and Logistics*, 2514-2517.
- [15] Suchacka, G., Chodak, G. (2017) Using association rules to assess purchase probability in online stores. *Inf Syst E-Bus Manage* **15**, 751–780. <https://doi.org/10.1007/s10257-016-0329-4>
- [16] March N., Reutterer T. (2008) Building an Association Rules Framework for Target Marketing. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_52
- [17] Phani, P., and Murlidher, M. (2013). A Study on Market Basket Analysis Using a Data Mining Algorithm. *International Journal of Emerging Technology and Advanced Engineering*, 3(6): 361- 363. Retrieved from www.ijetea.com
- [18] Zaki M.J. (2000). Scalable Algorithms for Association Mining. *IEEE Transaction on Knowledge and Data Engineering*, 12(3): 372-390.
- [19] P. Yazgan, A.O. Kusakci. (2016). A Literature Survey on Association Rule Mining Algorithms, *Southeast Europe Journal of Soft Computing* 5(1), Doi: 10.21533/scjour(nal.v5i1.102
- [20] Vidhate, D. (2014). To improve Association Rule Mining using New Technique: Multilevel Relationship



Algorithm towards Cooperative Learning 241–246.

- [21] Cuzzocrea A., Leung C.K., MacKinnon R.K. (2015). Approximation to expected support of frequent itemsets in mining probabilistic sets of uncertain data, *Procedia Comput. Sci.* 613–622.
- [22] Naresh P. and Suguna R. (2019). "Association Rule Mining Algorithms on Large and Small Datasets: A Comparative Study," *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, pp. 587-592, doi: 10.1109/ICCS45141.2019.9065836.
- [23] Al-hagery, M.A. (2015). Knowledge Discovery in the Data Sets of Hepatitis Disease for Diagnosis and Prediction to Support and Serve Community. *Int. J. Comput. Electron. Res.* 4, 118–125.
- [24] Agrawal, R., Srikant, R. (1994) Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB 1994, Santiago de Chile, Chile, (487–499).
- [25] Aggarwal, C. C. *Data mining: the textbook*. Heidelberg: Springer; 2015.
- [26] Han, J, Pei, J, Ying, Y, Mao, R. Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Min. Knowl. Discov.*, 2004, 8(1):53–87.
- [27] Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering.* 12 (3), 371- 390. doi:10.1109/69.846291
- [28] Pei J., Han, J., Lu H., Nishio S, Tang S, Yang D. (2001) H-mine: Hyper-structure mining of frequent patterns in large databases. In: Proc. 2001 IEEE Intern. Conf. Data Mining, 441–448.
- [29] T. Uno, M. Kiyomi, and H. Arimura, (2004) "Lcm ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets," in Fimi, vol. 126, Proc. ICDM'04 Workshop on Frequent Itemset Mining Implementations, CEUR, 2004.
- [30] Houtsma M. and Swami A. (1995) Set-oriented mining of association rules". Elsevier journal *Data & Knowledge Engineering* 245-262 Technical Report RJ 9567
- [31] Gosain, A., & Bhugra, M. (2013). A comprehensive survey of association rules on quantitative data in data mining. *2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES*, 1003-1008. DOI: <https://doi.org/10.1109/cict.2013.6558244>.
- [32] Fournier-Viger, P., Lin, C., Vo, B., Truong, T.C., Zhang, J., & Le, H. (2017). A survey of itemset mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(4) <https://doi.org/10.1002/widm.1207>