International Journal of Research in Science and Technology http://www.ijrst.com

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar e-ISSN: 2249-0604, p-ISSN: 2454-180X

ASSOCIATION RULE MINING WITH ECLAT ON A MALAYSIAN RETAIL STORE[1]

Ankita Nandy
Department of Computing
Asia Pacific University
Kuala Lumpur, Malaysia

Received: 02 December 2017; Accepted: 17 January 2018; Published: 29 January 2018

ABSTRACT

Association rule mining is a popular knowledge discovery algorithm in the retail industry. The knowledge is obtained in the form of rules. As what people buy is driven by reason and not entirely random, analysing their purchases gives an insight into their social, cultural and economic preferences. This understanding of the customer behaviour gives the store management a competitive edge, higher revenue and increased customer loyalty. This paper presents an implementation on the transactional dataset from a retail store in Malaysia. The weighted Eclat algorithm in R 'arules' package has been used to obtain the association rules for this dataset. A set of 110 rules with high lifts have been obtained which have been analysed to formulate solutions to optimize store layout, suggest cross selling opportunities, boost sales and customer satisfaction.

***Keywords**: ECLAT; retail; Malaysia; association rule mining*

INTRODUCTION

Association rule mining, also referred to as market basket analysis or affinity analysis, is a widely used and influential knowledge discovery approach in retail, and has found applications in diverse fields beyond retail. Conceived in 1993 and facing development till date, several algorithms and their optimized versions form a huge body of literature. Apriori, Eclat and FP Growth are three popular algorithms in this family. Each algorithm has its advantages and shortcomings, which have been enhanced and mitigated, respectively, with a range of improvements. Thus several variants are available on commonly used data analytics softwares. Additional criteria to mine significant rules, like weighted transactions, fuzzy logic et cetera have been incorporated. This family of algorithms identify the feature(s) that gets implied with a high probability by the presence (or absence) of other feature(s). This information is obtained in the form of rules, and can be used in diagnostic, predictive and prescriptive analytics. As an example, an association rule like the purchase of bread implies the purchase of butter, signals the management to maintain sufficient inventory of both the items at all times, to avoid customer displeasure. It can predict a rise in the sales of butter if a discount on the prices of breads is announced. It can guide the store staff into placing bread and butter within visible

[1] *How to cite the article:* Nandy A., Association Rule Mining with Eclat on a Malaysian retail Store; *International Journal of Research in Science and Technology*, Jan-Mar 2018, Vol 8, Issue 1, 31-49

range to each other around the store. In addition to the commercial benefit of association rules, socioeconomic and cultural trends that drive these rules can also be uncovered with these rules. As an example, luxury items are associated with the higher income groups. As an example, if it is known that purchasers of tea will purchase teacups also, announcing a promotion that bundles premium tea leaves with glassware will prove to be a better move than bundling tea with plastic cups.

In a multi-ethnicity country like Malaysia, with vast demographic and cultural diversity, catering to the customers' tastes can be a challenge. Thus, analyzing both the commercial and social aspects can benefit retail stores, increase revenues and promote customer loyalty. This paper presents a case study on a Malaysian retail store using weighted Eclat and discusses the results from various perspectives.

LITERATURE REVIEW

Association rule mining (ARM) is a knowledge discovery technique which to find factors/items such that the presence of one implies with high probability the occurrence of other. It is also known as market basket analysis and affinity analysis. Its inherent simplicity makes it a versatile tool for knowledge discovery. Reference [1] reported ARM as one of the most influential methods of knowledge discovery in the field of data science, alongside decision trees, neural networks and support vector machines. The following section reviews a subset of the vast literature available in this field.

ARM originated as the Apriori algorithm in [2]. Subsequent landmark developments in this field were the inventions of the Eclat algorithm [3] and the FP Growth [4]. Reference [5] presents a detailed survey on these algorithms. Reference [6] presents a classification system to select the best ARM algorithm for a dataset based on certain features. In [7], we find a comparative analysis of these three algorithms. Based on the maximum size of the basket and the dataset density, besides scalability and computational speed, FP Growth and Eclat are demonstrated to have better performance over Apriori.

The behavior and demands of the customer varies with the product or service offered, local demographics, culture and economic conditions among several other factors. Information on this variability can be give retail chains a competitive edge, especially which operate through online shopping portals and physical stores and thus cater to a diverse customer base. Within the retail sector, market basket analysis has been applied to several fields. Reference [8] performs a market basket analysis followed by multinomial regression on the dataset of transactions involving beauty products. Reference [9] applies this analysis to data from a sports equipment store. Association rules provide significant insight into such variations and are particularly useful for applications like online shopping which have a multicultural clientele. The advent of the Internet and the popularity of social media and e-commerce websites make the collection, storage and manipulation of huge volumes of transactional data easier, opening up possibilities for the development of further applications. Reference [10] proposes an ARM algorithm which can track the variations in the associations in the data over time. ARM systems are referred to as the poor man's recommender system and much work has been done in this direction. Beyond retail, [11] use a rank based ARM to find associations between genes. Reference [12] presents a book recommender system based on association rule mining.

Over the years, newer algorithms and optimizations over the aforementioned algorithms have been introduced. Reference [13] proposes a novel graph algorithm for mining association rules. Reference [14] presents fuzzy-genetic algorithm based approach to association rule mining. Optimization and enhancement in ARM is an on-going process.

A. Association Rule Mining Terminology

A basket is defined as all the products a customer purchases under on transaction. These multiple products do not form a random collection; rather they hint the purpose of the purchase. As an example, butter or jam or some form of spread invariably accompanies the purchase of bread in most countries and hints at the general dietary preference. It can also uncover some unexpected interesting associations like the famous beer-diaper rule.

Let **I** be a set of items. A set **X** = $\{i_1,\ldots, i_k\}\subseteq \mathbf{I}$ is called an itemset. It is called a k-itemset if it contains k items. A transaction over **I** is a couple **T** = (tid, **I**) where tid is the transaction identifier. A transaction **T** = (tid, **I**) is said to support an itemset $\mathbf{X} \subseteq \mathbf{I}$, if $\mathbf{X} \subseteq \mathbf{I}$. A transaction database **D** over **I** is a set of transactions over **I**.

The cover of an itemset X in D consists of the set of transaction identifiers of transactions in **D** that support **X**:

$$\text{cover}(\mathbf{X},\mathbf{D}):=\{\text{tid}|(\text{tid},\mathbf{I}) \in \mathbf{D}, \mathbf{X}\subseteq \mathbf{I}\} \quad (1)$$

The support of an itemset **X** in **D** is the number of transactions in the cover of **X** in **D**:

$$\text{support}(\mathbf{X},\mathbf{D}):=|\text{cover}(\mathbf{X},\mathbf{D})| \quad (2)$$

The frequency of an itemset X in D is the probability of X occurring in a transaction $\mathbf{T} \in \mathbf{D}$:

$$\text{frequency}(\mathbf{X},\mathbf{D}):=\text{support}(\mathbf{X},\mathbf{D})/|\mathbf{D}| \quad (3)$$

If this itemset appears more often than a predetermined threshold, it is termed frequent.

The conditional probability of the occurrence of one member in the itemset, provided that the other member(s) has occurred, is termed confidence.

$$\text{confidence}(x_i\rightarrow x_j)=\text{support}(x_i \wedge x_j)/\text{support}(x_i), \text{ where } i\neq j \quad (4)$$

A rule (A => B) with confidence 60% would imply that if the baskets contains product A, 6 out of 10 such baskets will also contain product B. Combinations that have a confidence above a predetermined threshold are termed significant.

The LHS is the antecedent or body of the rule and RHS is consequent or head of the rule. If products A and B are bound by a rule A=>B then the management can apply promotions like discount, clearance sales et cetera on A and find that sales of B also show an increase without any promotions on B. Placing A and B in close vicinity in the store add to convenience. However, placing A and B far apart in the store in some cases makes the customer spend more time in the shop browsing the aisles along the way and thus, end up buying more products.

B. Redundancy Reduction in Association Rules

A handful of products can generate several rules. The concept of redundancy is introduced at this point. A rule is said to be redundant if the association it depicts is already covered by a more generic

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

rule of higher or equal confidence. Filtering out the redundant rules leaves the user with fewer rules and makes it easier to interpret.

C. *Significance of Lift*

It is often encountered that two much purchased products appear as an association rule. As an example, milk=>water may appear simply because people buy these products often. However, fluctuations in prices of milk may in no way affect the sales of water. Such combinations occur by chance and do not have high usability. An additional measure, lift, is used to check the quality of rules. A high value of lift > 1 makes the rule interesting.

$$\text{lift}(x_i \rightarrow x_j) = \text{support}(x_i \wedge x_j)/(\text{support}(x_i)\text{support}(x_j)) \quad (5)$$

D. *Weighted association rule mining*

With the broadening of the domain of application of ARM, the presence and absence of items/ attributes/ factors were not considered informative enough and additional parameters/ features about the data were included, one of them being weight. Every item does not hold the same significance in a basket. A rarely bought but costly item, like premium tea leaves, may be considered infrequent and get left out in the analysis. In such a scenario, associating a weight with each basket/ item highlights those rules which get filtered out due to low support however are important in the field of study. References [15][16] discuss the drawbacks of presence-absence based association rule mining and introduce novel algorithms to include weights into the mining of rules.

The weight assigned to each basket, can be selected from a range of parameters like the profit per product, revenue et cetera. In such an implementation, support is redefined t include the weight component. Reference [11] uses rank as weight to uncover biologically significant association in genes.

E. *Negative association rules*

For some products, as an example milk and milk powder, it is observed that the increased sales of one might lead to drop in the sales of the other. Such rules are referred to as negative association rules. References [17][18] work on algorithms which mine negative association rules. Together with positive association rules, these add to the available information which can be exploited to optimize decision making. As an example, a positive rule might indicate placing sanitary napkins in the vicinity of grooming products as a welcome move. However, the same products placed near the unisex grooming products or grooming products targeted at men can make the customers uncomfortable and discourage visits. Mining negative association rules can avoid such layouts. Reference [19] presents an algorithm to mine both positive and negative association rules.

F. *ECLAT algorithm*

Eclat works on a recursive depth first search, thus reducing the number of passes on the database increasing the processing speed and reducing the memory consumption. The following algorithm explains the working of Eclat as mentioned in [5].

D is defined as a transaction database over a set of items **I**, and **σ** a minimal support threshold. **F[I](D, σ)** is defined as the frequent itemsets of I present in D which satisfy the minimal support criterion.

INPUT: **D**, **σ**, i ⊆**I**

OUTPUT: **F**[**I**](**D**,**σ**)

1. F[I] := {}
2. for all i I occurring in **D** do
3. **F**[**I**] := **F**[**I**] ∪ {**I** ∪ {i}}
4. // Create $\mathbf{D}_i$
5. $\mathbf{D}_i$:= {}
6. for all j ∈**I** occurring in **D** such that j > i do
7. C := cover({i}) ∩ cover({j})
8. if |C| ⩾ **σ** then
9. $\mathbf{D}_i$:= $\mathbf{D}_i$ ∪ {(j, C)}
10. end if
11. end for
12. // Depth-first recursion
13. Compute **F**[**I** ∪ {i}]($\mathbf{D}_i$,σ)
14. **F**[**I**] := **F**[I] ∪ **F**[**I** ∪ {i}]
15. end for

The current analysis uses the weighted version of Eclat, as included in the R package.

G. *Applications of these rules*

Association rules provide a guideline to understand the dynamics of the sales in the store. This information provides clues to the management to place butter and in the vicinity of bread and indicates a rise in the sales of these spreads whenever some promotion on bread is announced. Such trends in customer behavior contribute in enhancing customer satisfaction and winning their loyalty.

Reference [20] carries out an interview based analysis of the ways retail sector utilizes these rules and comes up with a list of best practices. Retail sector can use these rules for coming up with more targeted advertisements and promotions. It can improvise the shop layout, inventory management, loyalty benefits and thus influence the customers into buying more. An understanding of the customer's dietary and cultural preferences can be known from their shopping baskets, which opens up opportunities to establish a more personalized relationship with them. These contribute to increase in sales, customer satisfaction and loyalty, and brand building.

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

Rules of the form A=> B are useful in planning promotions. A discount on A can boost the sales of B. A=> not C will prepare the management for drop in sales of C due to the same promotions and they can announce clearance sales on C or introduce additional offers on C to mitigate the risks.

Rules of the type A=>B can be used to plan the layout of the store. Although it seems logical that A and B should be placed together to increase customer convenience, it has been learnt through corporate experience that placing the products apart can encourage the customer to browse the store and thus get incited into purchasing more.

Certain combination of products can be traced to specific traditions. An in-depth study of association rules can highlight cultural preferences of the customer. As every culture places different importance to the components in the overall shopping experience, it prepares the management to understand local tastes and to customize the packaging, advertising and interactions with the customers.

MATERIALS AND METHODS

The data for this analysis was obtained as a MySQL dump of 13 tables, of which two tables, henceforth referred to as A and B, were used in the analysis. Table A contained records for each product sold at the store, with its description, department ID, units sold and total revenue generated. Table B listed the details of the transactions carried out at the Point of Sale (PoS), with product description, discounts (if any) and the net cost. Each transaction in the table is mapped to an alphanumeric reference ID.

For this analysis, several transformations and filters were applied to convert the raw data into usable format.

(1) The reference IDs were mapped to unique numeric values to aid in database indexing for fast queries.

(2) The reference IDs recording the purchase of only one product were filtered out.

(3) Reference IDs with a total cost of less than RM 50 were filtered out.

(4) Some departments recorded very low revenue in comparison to the overall revenue generated in the period of study. They were filtered out.

The data in Table A was aggregated and used for visualizing the contribution made by each department to the total revenue and volume sales of the store. Visualization of this information on PowerBI indicates that some products are purchased often, however do not generate much revenue in comparison to the products which are purchased less often but make a significant contribution. As visualization is a window to the data, it justifies the implementation of weighted association rules mining in this work.

R is a language and environment for statistical computing and graphics. It is open source software covered by the GNU license and its functionalities can be easily extended using packages. The package 'arules 1.5-5' provides implementation of several ARM algorithms. The current analysis used the weighted Eclat on a real dataset of retail store in Malaysia. Negative association rules have not mined, however wherever speculated they have been hinted at.

RESULTS AND DISCUSSION

PowerBI provides an interactive interface to visualize the data at various levels of detail. The report (Figure 1) has been prepared with three visuals and one selector. It gives the viewer an option to compare and contrast the contributions each department makes to the overall revenue generated in a class. The word cloud displays the variety of products sold under the selected department. It can be observed that some products end up short in the sales by quantity bar chart but form a greater share in the revenue earned treemap. This justifies the selection of weighted Eclat for this analysis.

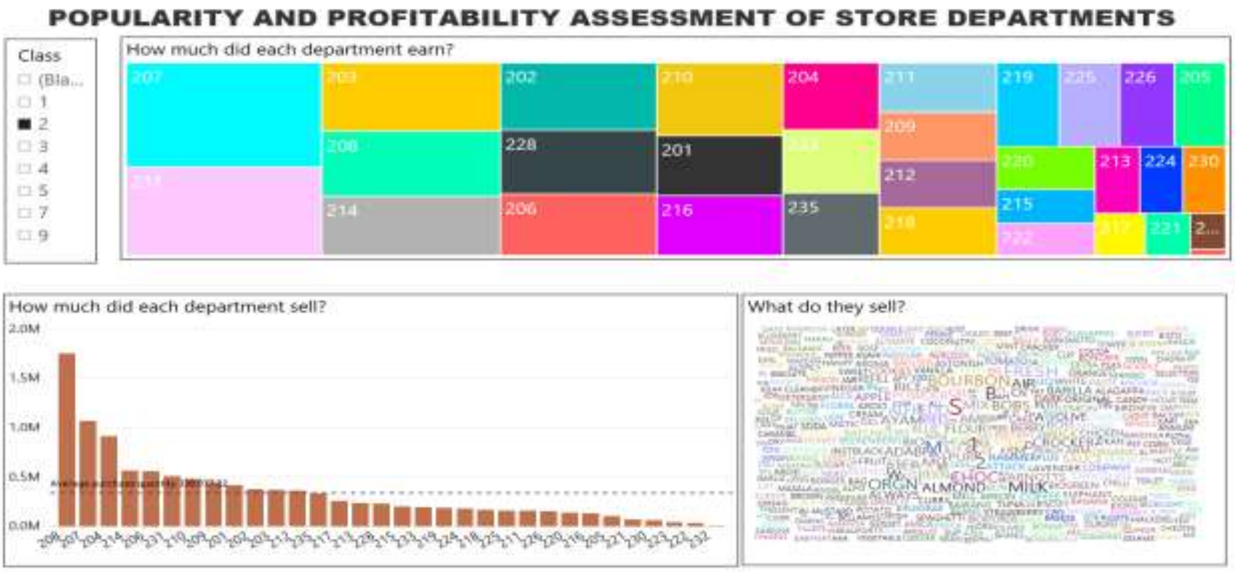

Figure 1: PowerBI data visualization

An execution of the weighted ECLAT algorithm generates 121 rules which have been presented as a graph in Figure 2. It has the support count along the x axis, the confidence levels along the y-axis. The color saturation indicates the lift. The values of lift corresponding to a particular color can be read from the colour map to the right. A high value of lift for a particular rule indicates that the rule is not the outcome of chance purchases and highlights its interestingness.

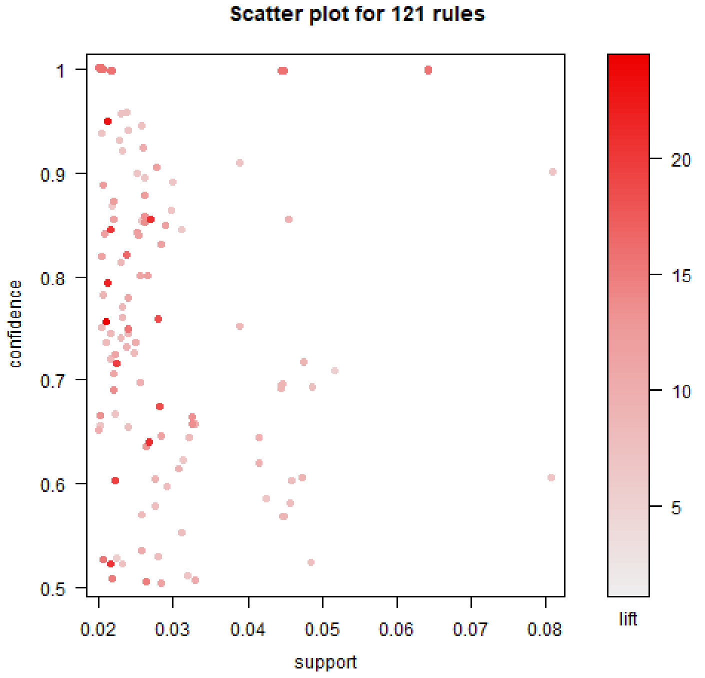

Figure 2: Association rules in graph

On removing redundant rules, 110 rules are obtained which are clubbed to enhance interpretation. The rules have been arranged in descending order of their lifts. The following section presents an overview of the related departments and what the association discovered between them may possibly suggest.

A. *Products in departments 302 and 304 are bought together*

Department 304 sells products essential to personal hygiene like deodorants, shower gels and soaps. The department (302) houses toiletries for both sexes, like shaving creams, cleansers and range of skincare essentials. The set of rules (Table I) indicates that people purchase bath essentials and grooming products together. Rule 302=>304 carries greater confidence. Placing clearance sales or promotional discount on products of 302 can boost sales for both the departments.

Table I. Rules 302-304

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept304	dept302	0.02166320	0.5216968	20.332293
dept302	dept304	0.02166320	0.8442884	20.332293

B. *Products in departments 303, 304 and 305 are bought together*

The department (303) comprises of products related to hair care and styling for both sexes. The department (304) sells products essential to personal hygiene like deodorants, shower gels and soaps. The department (305) is related to oral health care products like toothpastes, mouthwash, chewable breath neutralizers et cetera. This set of association rules (Table II) indicates that people purchase bath essentials together with aids to keep their oral and aural hygiene. Of all the rules, the one with highest confidence indicates that the baskets with 305 and 303 also contain 304. Promotions which include products from both 303 and 305, as an example, shampoo satchels with toothpaste tubes, can boost sales for 304.

Table II. Rules 303-304-305

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept304, dept305	dept303	0.02114481	0.7562591	24.326771
dept303, dept305	dept304	0.02114481	0.9522776	22.932908
dept303, dept304	dept305	0.02114481	0.7939931	21.571230
dept304	dept303	0.02663097	0.6413315	20.629869
dept303	dept304	0.02663097	0.8566451	20.629869
dept305	dept303	0.02220446	0.6032515	19.404938
dept303	dept305	0.02220446	0.7142563	19.404938
dept305	dept304	0.02795975	0.7596112	18.293083
dept304	dept305	0.02795975	0.6733312	18.293083

C. *Products of 223 are bought with 222*

The department 223 sells insecticides and fume sprays. The department 222 sells diapers for infants and adults. This association (Table III) indicates that families who pay special attention towards maintaining pest free homes might be having kids at their place. These departments should be in visible range to department selling infant and toddler related products (504).

Table III. Rules 223-222

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept223	dept222	0.02630217	0.8543292	13.674374

D. *Products from departments 212, 215 and 217 sell together*

The department (212) sells a variety of products from sauces, seasonings to essences and canned products. The department (215) sells a variety of baked products and essentials for baking enthusiasts. This department (217) houses staple grains like flour and corn, and other essentials in the kitchen like sugar and salt.

This set (Table IV) indicates that the customers prefer home-cooked meals and semi-processed foods. Placing these products in visible range to electrical appliances used in the kitchen can be beneficial.

Table IV. Rules 212-215-217

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept212, dept217	dept215	0.02388975	0.8191749	16.537429
dept212, dept215	dept217	0.02388975	0.7484453	15.313343
dept217	dept215	0.03253347	0.6656413	13.437907
dept215	dept217	0.03253347	0.6567827	13.437907

dept215, dept217	dept212	0.02388975	0.7343130	9.349547
dept215	dept212	0.03191916	0.6443811	8.204501
dept217	dept212	0.02916318	0.5966846	7.597212

E. *Products from departments 202, 203, 205 and 208 sell together*

The department (202) houses spices, condiments and breakfast cereals. The department (203) sells a variety of essentials in preparing hot beverages like tea leaves, coffee powders and dairy creamers. The department (205) houses pastry and bread spreads. The department (208) is dedicated to sweetened and/or carbonated beverages.

This set of rules indicates a preference for meals which can be prepared quickly with little or no efforts at cooking. A hectic pace of life might be leaving people with little time to prepare traditional breakfast dishes. Reference [21] in their study on school children in Kuala Lumpur report a growing trend of having ready-to-eat cereals and other food items as breakfast or substitutes for skipped breakfasts. They can be placed near the department selling breads (204, 702, 703, 704 and 706) and snacks and cookies (206). Further, considering the time constraints these people face, traditional dishes in heat-and-eat mixes like in (226) might attract these customers. Free samples can be distributed with cereals to increase awareness of such products.

Table V. Rules 202-203-205-208

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept203, dept208	dept205	0.02198673	0.6900570	13.796064
dept202, dept208	dept205	0.02026345	0.6671467	13.338025
dept202, dept205	dept203	0.02639320	0.8580085	13.277651
dept202, dept203	dept205	0.02639320	0.6345681	12.686693
dept203, dept205	dept202	0.02639320	0.8031446	12.006369
dept202, dept208	dept203	0.02205230	0.7260422	11.235476
dept205, dept208	dept203	0.02198673	0.7061397	10.927487
dept203, dept208	dept202	0.02205230	0.6921148	10.346562
dept205	dept203	0.03286233	0.6570048	10.167125
dept203	dept205	0.03286233	0.5085433	10.167125
dept205, dept208	dept202	0.02026345	0.6507936	9.728843
dept203	dept202	0.04159239	0.6436407	9.621913
dept202	dept203	0.04159239	0.6217729	9.621913
dept205	dept202	0.03076100	0.6149937	9.193664
dept203, dept205	dept208	0.02198673	0.6690558	6.884210
dept202, dept205	dept208	0.02026345	0.6587382	6.778047
dept205	dept208	0.03113652	0.6225013	6.405190
dept202, dept203	dept208	0.02205230	0.5302003	5.455464

F. *Products from departments 209, 210 and 213 are bought together*

The department (209) presents a range of noodles and pasta products. The department (210) sells sauces and salad dressings. The department (213) sells spices and seasoning items. Noodles, ketchups and seasonings make an obvious set (Table VI). Placing these products far apart but in visible range will add to customer convenience.

Table VI. Rules 209-210-213

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept209, dept210	dept213	0.02592698	0.5349766	10.114419
dept209, dept213	dept210	0.02592698	0.9261012	10.020517
dept213	dept210	0.04529768	0.8564107	9.266459
dept210, dept213	dept209	0.02592698	0.5723689	8.178235
dept213	dept209	0.02799584	0.5292973	7.562811
dept210	dept209	0.04846377	0.5243833	7.492598
dept209	dept210	0.04846377	0.6924696	7.492598

G. *Products from 211, 212 and 221 are bought together*

The department (211) sells liquids like edible oils and fats of plant and animal origin. The department (212) is dedicated to canned foods. Department (221) is related to essentials in female menstrual hygiene like sanitary napkins, disposable briefs and tampons. The presence of sanitary napkins in this set of rules (Table VII) indicates that these set of customers are women, and they cook. The store can create a positive impact by offering free goodies and innovative recipes with their purchases.

Table VII. Rules 211-212-221

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept221	dept211	0.06440371	1.0000000	15.527056
dept211	dept221	0.06440371	1.0000000	15.527056
dept212	dept211	0.04472150	0.5694109	8.841275
dept211	dept212	0.04472150	0.6943933	8.841275
dept221	dept212	0.04472150	0.6943933	8.841275
dept212	dept221	0.04472150	0.5694109	8.841275

H. *Products from departments 218, 219 and 220 are bought together*

The department (218) mainly deals with detergents and conditioners related to laundry. The department (219) sells cleaners for home care, including air fresheners and floor cleaners. The department (220) deals with paper products related to personal hygiene like napkins and wet wipes. These departments are related to cleanliness and are thus purchased together (Table VIII). For the purpose of promotion, these products can be housed in proximity to other aids in housekeeping like mothballs, insecticides, organizers, cloth hangers et cetera to present a more comprehensive package.

Table VIII. Rules 218-219-220

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept218, dept220	dept219	0.02755456	0.9042196	11.543002
dept219	dept218	0.04752149	0.6066454	9.148311
dept218	dept219	0.04752149	0.7166317	9.148311
dept219	dept220	0.04555540	0.5815470	7.676857
dept220	dept219	0.04555540	0.6013655	7.676857

I. *Products of 222 are bought with 220*

The department (222) sells infant and adult diapers and related products. The department (220) houses paper based products like tissues, napkins, toilet paper rolls et cetera.

The association (Table IX) indicates that young parents accompany their diaper purchases with paper based wipes to keep their child dry. Offers like add two RM and get a handy cloth carry bag and similar schemes can be an optimum strategy to boost sales for department 413(Bags).

Table IX. Rules 222-220

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept222	dept220	0.03199371	0.5120906	6.759981

J. *Products of 226 are bought with 208*

The department (226) sells instant meals. Department (208) sells sweetened and/or carbonated beverages. Reference [22] study the trends and factors in fast food consumption in the Malaysian market and highlight that busy lifestyles make instant meals a convenient and tasty option. The motivation of buying ready to eat foods is foremost convenience followed by taste. This association (Table X) indicates a preference towards grabbing a quick meal and accompanying it with ready to consume beverages. Placing these products nearby in the store and preferable close to the entrance and cash counters will make it easier to grab-pay-run and can boost sales.

Table X. Rules 226-208

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept226	dept208	0.02297304	0.8143128	8.378823

K. *Products in 101, 106 and 204 are bought together*

The department (101) sells fresh green groceries. The department (106) sells an assortment of beef and pork products, together with handmade servings. The department (204) deals with pastries and baked goods like breads and buns.

This set of rules (Table XI) indicates the popular choice of having fresh greens and meat products with a variety of breads. As the economic conditions across Asian cultures get better, a preference to obtain more calories and nutrients from greens and animal products and substitute rice with wheat and alternative grains is observed.

Table XI. Rules 101-106-204

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept101, dept204	dept106	0.02187994	0.8558757	11.783680
dept204	dept106	0.02516064	0.8408954	11.577432
dept204	dept102	0.02156496	0.7207238	8.024678
dept106, dept204	dept101	0.02187994	0.8696098	6.525269
dept204	dept101	0.02556439	0.8543891	6.411058

L. *Rules between 101, 102, 106 and other departments*

The department (101) sells fresh green groceries. The department (102) sells a variety of fruits and edible items derived from them like juices and nectars. The department (106) sells an assortment of butchery products, together with handmade servings. These occur together with one of the following departments (Table XII). There are four sets of rules (Table XIV-XVI) concerning departments 101, 102, 106, 201, 111, 104 and 214. 106, 102 and 106 are commonly purchased together.

Table XII. Department with Descriptions

Department No.	*Description*
201	Poultry lay products, fresh and preserved eggs
111	Dairy products, cheese
104	Poultry meat products
214	Fisheries and related products

Table XIII. Rules connecting 101, 102, 106 and 111

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept101,dept102, dept106	dept111	0.02041896	0.5266299	15.330248
dept102, dept106	dept111	0.02173346	0.5098168	14.840819
dept101, dept106	dept111	0.02610552	0.5070157	14.759278
dept101,dept102, dept111	dept106	0.02041896	0.8886292	12.234629
dept101, dept111	dept106	0.02610552	0.8781091	12.089789
dept102, dept111	dept106	0.02173346	0.8725774	12.013630
dept111	dept106	0.02913488	0.8481194	11.676891
dept101,dept106, dept111	dept102	0.02041896	0.7821701	8.708833
dept101, dept111	dept102	0.02297804	0.7729104	8.605733
dept111	dept102	0.02490720	0.7250511	8.072859
dept106, dept111	dept102	0.02173346	0.7459601	8.305664
dept102,dept106, dept111	dept101	0.02041896	0.9395170	7.049831

dept102, dept111	dept101	0.02297804	0.9225461	6.922486
dept106, dept111	dept101	0.02610552	0.8960228	6.723464
dept111	dept101	0.02972925	0.8654213	6.493840

Table XIV. Rules connecting 101, 102, 106 and 104

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept101, dept104	dept106	0.02419764	0.7787645	10.722015
dept104	dept106	0.02570408	0.6992137	9.626761
dept101, dept104	dept102	0.02300845	0.7404921	8.244782
dept104	dept102	0.02403388	0.6537804	7.279317
dept102, dept104	dept101	0.02300845	0.9573337	7.183521
dept104,dept106	dept101	0.02419764	0.9413931	7.063908
dept102	dept101	0.08089515	0.9007022	6.758577
dept101	dept102	0.08089515	0.6070109	6.758577
dept104	dept101	0.03107183	0.8452298	6.342330

Table XV. Rules connecting 101, 102, 106 and 201

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept101, dept201	dept106	0.02531056	0.8432269	11.609532
dept102, dept201	dept106	0.02069605	0.8425261	11.599883
dept201	dept106	0.02810923	0.8327743	11.465621
dept101, dept201	dept102	0.02288463	0.7624065	8.488782
dept106, dept201	dept102	0.02069605	0.7362725	8.197800
dept201	dept102	0.02456428	0.7277505	8.102915
dept102, dept201	dept101	0.02288463	0.9316221	6.990590
dept102	dept101	0.08089515	0.9007022	6.758577
dept101	dept102	0.08089515	0.6070109	6.758577
dept106, dept201	dept101	0.02531056	0.9004360	6.756579
dept201	dept101	0.03001631	0.8892742	6.672825

Table XVI. Rules connecting 101, 102, 106 and 214

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept102, dept214	dept106	0.02037155	0.8202071	11.292596
dept101, dept214	dept106	0.02569822	0.8030808	11.056802
dept101, dept106	dept102	0.03877288	0.7530384	8.384475
dept106, dept214	dept102	0.02037155	0.7503734	8.354802
dept101, dept214	dept102	0.02382836	0.7446468	8.291041
dept102, dept214	dept101	0.02382836	0.9593862	7.198923
dept106, dept214	dept101	0.02569822	0.9465778	7.102812

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar e-ISSN: 2249-0604, p-ISSN: 2454-180X

dept102, dept106	dept101	0.03877288	0.9095223	6.824760
dept102	dept101	0.08089515	0.9007022	6.758577
dept101	dept102	0.08089515	0.6070109	6.758577
dept106	dept102	0.04262994	0.5869282	6.534972
dept106	dept101	0.05148858	0.7088938	5.319309

They correspond to green groceries, fresh fruits and fresh butchery products which indicate a preference for cooking at home. But, departments 104, 111, 201 and 214, are observed to occur together less often. This indicates a possibility for the offerings in these four departments to be complementary. Due to allergies, health conscious choices, cultural preferences, economic preferences, storage convenience and similar reasons, egg lovers avoid the dairy, chicken and seafood. Similar trends exist for customers who prefer meat, fish or dairy. To investigate the matter further, the following tables (XVII-XXII) have been presented.

Each table takes a combination of two of four departments, say A and B, and presents four counts,

(1) Transactions with products from both departments

(2) Transactions containing A but not B

(3) Transactions containing B but not A

(4) Transactions containing neither

In all tables it is invariably observed that counts (2) and (3) exceed counts (1), thus indicating a possibility of negative association of the form A=>~B and ~A=>B. Such uniformity in results cannot be attributed to mere chance.

Table XVII. Combination matrix for 104 and 201

104	***201***	***TRUE***	***FALSE***
TRUE		445	750
FALSE		608	27713

Table XVIII. Combination matrix for 104 and 111

104	***111***	***TRUE***	***FALSE***
TRUE		361	834
FALSE		627	27694

Table XIX: Combination matrix for 104 and 214

104	***214***	***TRUE***	***FALSE***
TRUE		478	717
FALSE		3695	24626

Table XX. Combination matrix for 201 and 111

201	***111***	***TRUE***	***FALSE***
TRUE		451	602
FALSE		537	27926

Table XXI. Combination matrix for 201 and 214

201	*214*	*TRUE*	*FALSE*
TRUE		465	588
FALSE		3708	24755

Table XXII. Combination matrix for 111 and 214

111	*214*	*TRUE*	*FALSE*
TRUE		383	605
FALSE		3790	24738

In such scenarios, activities to improve sales in one might have negative impact on others. The management will need to generate cross selling opportunities. The customers can be made aware of the health benefits of taking a combination of these complementary products through advertising. As the customers are mostly cooking enthusiasts, offering innovative recipes through hand-outs might encourage them to purchase a combination.

M. *Products from departments 106, 206 and 207 are bought together*

The department (106) sells an assortment of beef and pork products, together with handmade servings. The department (206) sells cookies and snacks. The department (207) comprises of sweets and confectionary items.

The rule 206=>207 suggests that these departments should be placed far apart, with one of them close to baked goods (703, 704). Those with a sweet tooth might get enticed by the cakes and muffins. The products from 206 and 207 can also be bundled up as a promotion, boosting sales of products in 106.

Table XXIII. Rules 106-206-207

LHS	*RHS*	*Support*	*Confidence*	*Lift*
dept206	dept207	0.02833731	0.6454097	11.475808
dept207	dept206	0.02833731	0.5038561	11.475808
dept206	dept106	0.02303525	0.5246500	7.223372
dept207	dept106	0.03118347	0.5544626	7.633830

N. *Products of 228 are bought with 106*

Department (228) sells a diverse range of organic products. Department (106) sells fresh butchery products and handmade servings.

Increased accessibility to information on environmental impacts of the commercial chemical methods has shifted the consumer focus towards ethical practices of production of grains, vegetables, fruits and animal products, leading to a growing demand for organic products. For the purchasers of organic goods ethical beliefs are the primary motivators of consumption of organic goods. The results are reinforced by a study on Kluang, Johor, Malaysia in [23] that the consumers of organic foods are driven by intention over product quality.

Such customers can be enticed into purchasing products that carry a charity component, as an example, for each purchase of product A, the store donates RM 1 towards a social or environmental cause. However organic products contribute to a very small share of the total revenue of class 2 products and sales need a boost. Buyer behavior is intrinsically linked to the emotions a product evokes. Educating the shoppers on the benefits of the organic products through visual aids might encourage them to get overcome the price barrier and try organic products.

Table XXIV. Rules 228-106

LHS	***RHS***	***Support***	***Confidence***	***Lift***
dept228	dept106	0.02497393	0.7379300	10.159806

CONCLUSIONS

The store organization, promotion planning and cross selling measure at the ubiquitous retail stores are based on the intuition and subjective understanding of the local tastes and demographics. Association rules mining uncovers some not-so-intuitive purchasing trends and enables the store management to take a data driven approach in monetizing cross-selling opportunities, planning promotions and maximizing the revenue. Instead of limiting the study to baskets in one shopping trip, a study that covers baskets over successive trips can uncover further aspects of the customer behavior. As an example, those who purchase a microwave may purchase microwave safe utensils in the consecutive weeks. Store management can offer attractive deals on potential future purchases and ensure customer loyalty towards the store. Mining such sequences comes under sequential pattern mining and is proposed as the future work.

ACKNOWLEDGMENT

The author would like to acknowledge Crest Business Solutions Sdn. Bhd., Selangor, Malaysia for sharing the anonymized dataset and providing the knowledge support required in completing this work.

REFERENCES

[1] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, S. Y. Philip and Z. H. Zhou, (2008), Top 10 algorithms in data mining, *Knowledge and information systems*, 14(1), 1-37.

[2] R. Agrawal, T. Imieliński and A. Swami, (1993), Mining association rules between sets of items in large databases, *ACM sigmod record*, Vol. 22, No. 2, pp. 207-216, ACM.

[3] M. J. Zaki, (2000), Scalable algorithms for association mining, *IEEE Transactions on Knowledge and Data Engineering*, 12(3), 372-390.

[4] C. Borgelt, (2005), An Implementation of the FP-growth Algorithm, *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, pp. 1-5, ACM .

[5] B. Goethals, (2003), *Survey on frequent pattern mining*, University of Helsinki, 19, 840-852.

[6] M. HooshSadat, H. W. Samuel, S. Patel and O. R. Zaïane, (2011), Fastest association rule mining algorithm predictor (FARM-AP), *In Proceedings of The Fourth International C* Conference on Computer Science and Software Engineering* (pp. 43-50). ACM.

[7] J. Heaton, (2016), Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms, *SoutheastCon*, pp. 1-7. IEEE.

[8] V. Gancheva, (2013), *Market basket analysis of beauty products*, Master of Science in Economics and Business, Erasmus University Rotterdam, Erasmus School of Economics.

[9] H. Kaur and K. Singh, (2013), Market basket analysis of sports store using Association Rules. *International Journal of Recent Trends in Electrical & Electronics Engg*, 3(1), 81-85.

[10] M. Kaur and S. Kang, (2016), Market basket analysis: Identify the changing trends of market data using association rule mining, *Procedia Computer Science*, 85, 78-85.

[11] S. Mallik, A. Mukhopadhyay and U. Maulik, (2015), RANWAR: rank-based weighted association rule mining from gene expression and methylation data, *IEEE transactions on nanobioscience*, 14(1), 59-66.

[12] J. V. Joshua, O. D. Alao, A. O. Adebayo, G. A. Onanuga, E. O. Ehinlafa and O. E. Ajayi, (2016), Data Mining: A book recommender system using frequent pattern algorithm, *Quest Journals Journal of Software Engineering and Simulation*, Volume 3 ~ Issue 3, pp: 01-13 ISSN(Online) : 2321-3795 ISSN (Print): 2321-3809.

[13] I. F. Videla-Cavieres and S. A. Ríos, (2014), Extending market basket analysis with graph mining techniques: A real case, *Expert Systems with Applications*, 41(4), 1928-1936.

[14] H. Kalia, S. Dehuri, A. Ghosh and S. B. Cho, (2016), On the mining of fuzzy association rule using multi-objective genetic algorithms, *International Journal of Data Mining, Modelling and Management*, 8(1), 1-31.

[15] G. D. Ramkumar, S. Ranka and S. Tsur, (1998), Weighted association rules: Model and algorithm, *Proc. ACM SIGKDD.*

[16] C. H. Cai, A. W. C. Fu, C. H. Cheng and W. W. Kwong, (1998), Mining association rules with weighted items, *Database Engineering and Applications Symposium*, Proceedings. IDEAS'98. International (pp. 68-77). IEEE.

[17] X. Yuan, B. P. Buckles, Z. Yuan and J. Zhang, (2002), Mining negative association rules, *Computers and Communications*, 2002. Proceedings. ISCC 2002. Seventh International Symposium on (pp. 623-628). IEEE.

[18] A. A. W. Al-Opahi and B. M. Al-Maqaleh, (2015), An Efficient Algorithm to Automated Discovery of Interesting Positive and Negative Association Rules, *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 6, No. 6.

[19] L. Zhou and S. Yau, (2007), Efficient association rule mining among both frequent and infrequent items, *Computers & mathematics with applications*, 54(6), 737-749.

[20] L. Gordon, (2008), Leading practices in market basket analysis: How top retailers are using market basket analysis to win margin and market share, *Factpoint Group.*

(IJRST) 2018, Vol. No. 8, Issue No. I, Jan-Mar **e-ISSN: 2249-0604, p-ISSN: 2454-180X**

[21] Hui-Chin Koo, Siti Nurain Abdul Jalil and A. T. Ruzita, (2015), Breakfast eating pattern and ready-to-eat cereals consumption among schoolchildren in Kuala Lumpur, *The Malaysian journal of medical sciences: MJMS,* 22(1), 32.

[22] F. Q. Habib, R. Abu Dardak and S. Zakaria, (2011), Consumers' preference and consumption towards fast food: Evidences from Malaysia, *Business & Management Quarterly Review,* 2(1), 14-27.

[23] C. S. Wee, M. S. B. M. Ariff, N. Zakuan, M. N. M. Tajudin, K. Ismail and N. Ishak, (2014). Consumers' perception, purchase intention and actual purchase behavior of organic food products, *Review of Integrative Business and Economics Research*, 3(2), 378.