# Investigation of Beauty Accessories Sales and Marketing by using Association Rule Mining

Than Yaw Zin, Daw Mar Mar Lwin
thanyawzin247@gmail.com,dawmarmarlwin@gmail.com
Computer University (Maubin)

## Abstract

*Many business enterprises accumulate large quantities of data from their day to day operations. Therefore, in this system, huge amounts of customer purchase beauty accessories data are collected daily by supplier representatives in transaction database. Supplier representatives are interesting in analyzing fivecategories beauty accessories to investigate about increase amount of purchasing of their customers. Such valuable information can be used to support a variety of business-related applications such as sales and marketing promotions. In this paper, we present an investigation framework by executing a methodology of association rule mining and correlation analysis. This system is mainly composed of three parts which are data collection, choice of frequent itemsets by generating association rules with Apriori algorithm and computation with correlation analysis to relate data items from each rule. According to interestingness measures, such as generating association rules and correlation, this system can give investigation of beauty accessories sales and marketing for supplier representatives.*

**Keywords:** data collection, transaction database, association rule mining, correlation analysis, investigation

## 1. Introduction

Nowadays, most of the organizations have implemented analysis on selling products that allow supplier representatives to perform investigation at the point of marketing. The marketing process involves ways that value can be created for the customers. Consequently, a supplier representative to plan the procedure of selling and promoting for customers is extensive on the product, at times taking up to several months. A supplier representative's main responsibility is to promote products or services to wholesale or retail buyers or purchasing agents. Supplier representatives demonstrate or explain the product or service to the customer and answer questions. Supplier representatives have several duties beyond selling products. They analyze sales statistics, prepare reports, and handle administrative duties. They also read about new and existing products and monitor the sales, prices and products of their competitors. If supplier representatives can not investigate on the sales, they make no sense to consider the role looks in the sale marketing.

On the other hand, supplier representatives must focus to its customers since they play their role at the end of distribution channel. Customers which buy sale products can be comprehended by observing how someone interacts and reacts to the marketing mix. According to this idea, supplier representatives determine the decisions related to the 2P (Product, Price) by focusing to its customers; while each individual considers the option to buy which products under the psychological influences of culture, attitude, previous usage of the products and personal perception. Effects of both inputs (marketing and psychological) somewhat influences the customers to decide whether they will buy or not, which brand to buy, which product associates with other products and other choices. Accordingly, there has to a line of action which may serve as a marketing for organizations and allow many supplier representatives sufficient scope to handle their performance situations. [5] Thus, in this paper, keeping customers' data of beauty accessories and using them as a part of decision making process has been investigated by working out the association rule with apriori algorithm in mining process.

Mining process is a rapidly advancing field of strategies for finding connections between elements in large database. The association rule is an essential study aspect in mining field. Association rules reflect the inner relationship of data. Discovering these associations is beneficial to the correct and appropriate decision made by decision-makers. The association rules provide an effective means to find the potential link between the data, reflecting a built-in association between the data. As well, the Apriori algorithm is a popular and foundational member of correlation-based data mining kernels used today. It is also an efficient algorithm for knowledge mining in form of association rules. The original Apriori algorithm is applied to a transaction database of market baskets. In our system, instead of a market basket, we work with basket of beauty accessories occurring in customers for supplier representatives and the transactional database is in fact a set of descriptions of beauty accessories and their prices.

In our system, the following study is guided by the hypothesis, that products chosen on a selling trip of supplier representative are an indicator of the

preference interdependencies between different products or related items. The objective of this paper is to discover associated beauty accessories, and to determine the proposed product type based on the associated products. As a consequence, we carry out with several explanations in this paper for supplier representatives.

## 2. Methods

This system use market basket analysis to find association rules between sets of items in transactional database. The goal of association rules mining is to find interesting associations or correlation relationships among a large dataset, i.e., to identify the sets of attributes-items, which frequently occur together and then to formulate the rules characterizing these relationships.

### 2.1 Mining Association Rules

Association rule mining often results in a huge amount of rules. The task of association rules mining is to generate all association rules which have values of the parameters support and confidence, exceeding the previously given respectively minimal support and minimal confidence. Attempts to reduce the size of the result for easier inspection can be roughly divided to two categories. (i) In the subjective approaches, the user is offered some tools to specify which rules are potentially interesting and which are not, such as templates and constraints. (ii) In the objective approaches, user-independent quality measures are applied on association rules. While interestingness is user-dependent to a large extent, objective measures are needed to reduce the redundancy inherent in a collection of rules. [1, 4]

The objective approaches can be further categorized by whether they measure each rule independently of other rules (e.g., using support, confidence, or lift) or address rule redundancy in the presence of other rules (e.g., being a rule with the most general condition and the most specific consequent among those having certain support and confidence values). Obviously only approaches of the latter type can potentially address redundancy between rules. Our work will be in this category.

### 2.2 Basic Concept of Apriori Algorithm

Apriori algorithm uses a search order of the cycle-level approach ( also known as the layers of iterative search method) to complete the set of frequent excavation work, use the following Apriori nature to compress the search space, increasing the frequency of collection efficiency.

The basic concept of Apriori algorithm is to generate a specific size of the candidate projects set, and then scan the database times line counts, to determine whether the candidate frequent item sets. The process is the first concrete realization of scanning the database of all matters, in terms of the number of each item, having a candidate $C_i$, again based on pre-set to determine the minimum support of a frequent $L_1$. $L_1 * L_1$ and then by the operator to connect two candidates generating sets $C_2$, once again scanning the database of all services, calculated each element of $C_2$ in the number of occurrences, and in accordance with pre-set to determine the minimum support of frequent $L_2$. This process repeats until the formation k frequent $L_k$ and can no longer meet the minimum generate support for the project k+1. [2]

### 2.3 Correlation Analysis

Before making any investment decisions based upon correlation analysis, one must recognize that correlation does not imply one security is responsible for the movement of another. Correlation simply measures the relationship of movement between two sides (association between selling products at the left side and other products at right side). In market basket analysis, the correlation between two sides is a statistical measure of the relationship between the movements of two sides. This relationship, which is expressed by what is known as the correlation coefficient, is represented by a value within the range of -1.00 to +1.00.

A correlation coefficient of +1.00 indicates that two sides move in the same direction at all times. If the products of side X gains in value, we would expect the products of side Y to gain as well. A correlation coefficient of 0 indicates that the side X and Y movements are totally random. A gain by side X provides no insight into expected movement of side Y. A correlation coefficient of -1.00 indicates that two sides move in the opposite direction at all times. If side X gains in value, we would expect side Y to decline in value. Otherwise, correlation quantifies the extent to which two quantitative variables, X and Y, "go together". When high values of X are associated with high values of Y, a positive correlation exists. When high values of X are associated with low values of Y, a negative correlation exists. [3]

## 3. Evolution of Investigation Framework

Data mining, which is the exploration of knowledge from the large set of data, generated as a result of the various data processing activities. In this paper, we express the evolution of investigation framework according to support three technologies as shown in Figure 1. The three technologies are (i) massive data collection, (ii) data mining algorithm and (iii) high performance computing.

**(i)Massive Data Collection:** For the purpose of this technology, empirical data are collected from daily-Boucher of customers by supplier representative. A

one year transaction data of the retail's customers is obtained and processed using association rules. The beauty accessories which are executed in this system consist of five categories such as hair color, shampoo, conditioner, vitamin-E and hair coat for hair treatment. In this system, 66 items are chosen which transactions to be included acquirement from customers' purchase. Moreover, this system executes transactions that had two or more product categories which consist of 1200 transaction data set for beauty accessories.

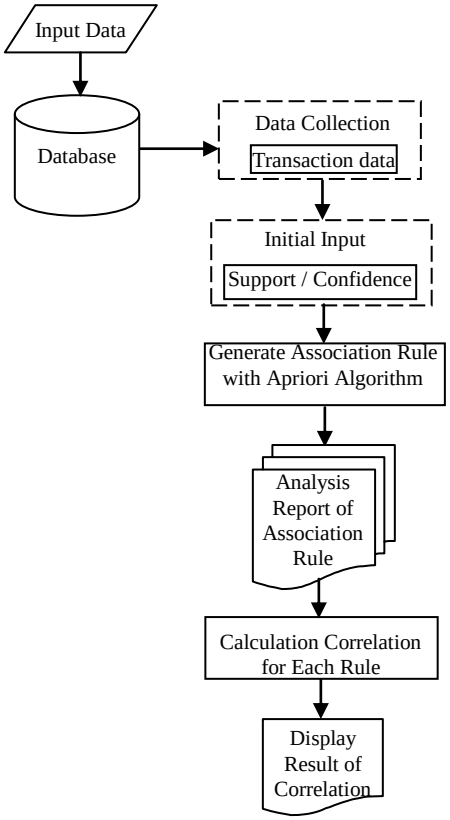


**Figure 1: Overview of Investigation Framework**

**(ii)Data Mining Algorithm:** After massive amounts of beauty accessories data items in marketing are continuously being collected and stored, mining frequent patterns or itemsets are a fundamental and essential problem in this system. This system includes the association rules with apriori algorithm. Therefore, in this system, association rule mining finds interesting association or correlation relationships among a large set of transaction beauty accessories. The association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Association rules are used to identify the relationships with frequent items amount a set of items in transaction database. These relationships are not based on inherent properties of the data themselves, but rather based on occurrence of the frequent itemsets. [7, 8]

As a result, the Apriori algorithm is not only influenced the association rule mining community, but it affected the frequent itemset mining as well. The Apriori algorithm employs the downward closure property if an item set is not frequent, any susperset of it cannot be frequent either. The Apriori algorithm performs a breath-first search in the search space by generating candidate k+1 itemsets from frequent k itemsets as illustrate as Figure 2.

**Join Step:** $C_k$ is generated by joining $L_{k-1}$with itself

**Prune Step**: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

**Pseudo-code**:

$C_k$: Candidate itemset of size k
$L_k$ : frequent itemset of size k
$L_1$ = {frequent items};
**for** ($k$ = 2; $L_k$ !=∅; $k$++) **do begin**
$C_{k+1}$ = candidates generated from $L_k$;
**for each** transaction $t$ in database do increment the count of all candidates in $C_{k+1}$ that are contained in $t$
$L_{k+1}$ = candidates in $C_{k+1}$ with min_support
**end**
**return** $\cup_k L_k$;

**Figure 2: Apriori Algorithm**

In this system, the frequency of an itemset is computed by counting its occurrence in each transaction of beauty accessories. The frequency of an item set is computed by counting its occurrence in each transaction.

Apriori is an influential algorithm for mining frequent itemsets of Boolean association rules. Apriori is an iterative level wise search Algorithm, where k itemsets are used to expore (k+1) itemsets. First, the set of frequent 1-itemsets is found. This set is denoted by $L_1$. $L_1$ is used to find $L_2$, the set of frequent 2 itemsets, which is used to find $L_3$ and so on, until no more frequent k-itemset can be found. The finding of each $L_k$ requires one full scan of beauty accessories in transaction database.

In this system we use two steps for understanding that how frequent itemset of beauty accessories $L_{k-1}$ is used to find $L_k$:-

1. The join step:-

To find $L_k$, a set of candidate k itemsets is generated by joining $L_{k-1}$ with itself. This set of candidates is denoted by $C_k$ (candidate itemset of beauty accessories).

2. The prune step:-

$C_k$ is a superset of $L_k$, that is, its members may or may not be frequent, but all of the frequent k itemsets of beauty accessories are included in $C_k$. A scan of the beauty accessories database t to determine

the count of each candidate in $C_k$ would result in the determination of $L_k$. $C_k$, however, can be huge and so this could involve heavy computation. To reduce the size of $C_k$, the Apriori property is used as follow.

Thus, in this system, any (k-1) item set of beauty accessories that is not frequent cannot be a subset of frequent k-item set of beauty accessories.

Hence, if (k-1) subset of a candidate k item set is not in $L_{k-1}$ then the candidate can not be frequent either and so can be removed from $C_k$.

As a result, in this system, these rules must satisfy minimum support and minimum confidence. Rules support and confidence are two measures of rule interestingness. Association Rules are considered interestingness if they satisfy both a minimum support threshold and minimum confidence threshold.

Single-Dimensional Boolean Association Rules

buys (X, Shampoo) => buys (X, Conditioner)
[support=2%, Confidence=60%]

A support 2% of all the transactions under analysis shows that Shampoo and Conditioner are purchased together.

A confidence 60% means that 60% of customers who purchased a Shampoo also bought the Conditioner.

For each frequent itemset *l* in beauty accessories database, this system generates all non-empty subsets of *l* and all empty subsets of *s*.

For every nonempty subsets of *l*, this system outputs the rules

$$\text{“}s \Rightarrow (l - s)\text{” if} \frac{\text{Support\_count (l)}}{\text{Support\_count (s)}} \geq \text{min\_conf,}$$

where min_conf, is the minimum confidence threshold.

We describe the frequent item set of beauty accessories in transaction database as an example as illustrate as Table 1.

**Table 1: Frequent Data Set in Transaction Database**

| TID | Items |
|---|---|
| 1 | $I_1, I_3, I_4$ |
| 2 | $I_2, I_3, I_5$ |
| 3 | $I_1, I_2, I_3, I_5$ |
| 4 | $I_2, I_5$ |

In the above beauty accessories transaction database, $I_1$ means Pantene Shampoo, $I_2$ and $I_3$ express Rejoice Shampoo and Clear Shampoo, $I_4$ and $I_5$are Pantene Conditioner and Head&Shoulder Conditioner.

Subsequently, this system operates the frequent item set $l = \{I_2,I_3,I_5\}$

$l = \{I_2,I_3,I_5\}$
Non-empty sets of *l* are $\{I_2,I_3\}$, $\{I_2,I_5\}$, $\{I_3,I_5\}$, $\{I_2\}$, $\{I_3\}$, $\{I_5\}$.
The resulting association rule,

$s \Rightarrow (l - s)$

E.g., $s = \{I_2,I_3\}$

$l - s = \{I_5\}$

Confidence, $I_2 \wedge I_3 \Rightarrow I_5$
( Shampoo , Conditioner ) => Hair Coat

The confidence can be calculated as below

$$P(I_5 \setminus I_2 \wedge I_3) = \frac{P((I_2,I_3) \cup I_5)}{P(I_2,I_3)} = \frac{2}{2} * 100$$

$$= 100\%$$

**(iii) High Performance Computing:** In this system, we consider the suggestion of each association rule for finding interesting relationships between beauty accessories data items based on correlation. Therefore, in this system, for investigation of beauty accessories, the occurrence of itemset X at left side is independent of the itemset Y at right side if $P(X \cup Y)=P(X)\ P(Y)$, otherwise itemsets X and Y are dependent and correlated as events. The correlation (dependence) between the occurrence of X and Y can be measured by correlation (X, Y), which we measure below. [6]

$$correlation\ (X, Y) = \frac{P(X \cup Y)}{P(X)P(Y)} = \frac{P(Y \mid X)}{P(Y)} = \frac{P(X \mid Y)}{P(X)} \quad (1)$$

Consider the relationship between P(Y|X) and P(Y) [or P(X|Y) and P(X)], correlation (X, Y) has the following three possible cases:

(1) If *correlation*(X, Y)=1 or P(Y|X)=P(Y) [or P(X|Y)=P(X)] then Y and X are independent and there is no correlation between them.

(2) If *correlation*(X, Y)>1 or P (Y|X)>P(Y) [or P(X|Y)>P(X)], then X and Y are positively correlated, meaning the occurrence of one implies the occurrence of the other. In this case, one has $0<P(X|Y)-P(Y)\leq 1-P(Y)$, $0<[P(Y|X)-P(Y)]/(1-P(Y))\leq 1$. The bigger the ratio $(P(Y|X)-P(Y))/(1-P(Y))$, the stronger Y is positively dependent on X.

(3) If *correlation*(X, Y)<1 or P(Y|X)<P(Y)[ or P(X|Y)<P(X)], then X and Y are negatively correlated, meaning the occurrence of one discourage the occurrence of the other. In this case, one has $-P(Y)<P(Y|X)-P(Y)<0$, $-1<P(Y|X)-P(Y)<0$. The smaller ratio $(P(Y|X)-P(Y)/P(Y))$, the stronger the Y is negatively dependent on X.

In order to make results easier to understand, potential frequent items of beauty accessories are displayed in the order of the top-down support. By using positive, zero and negative integers to indicate

the former and latter of all association rules generate by a certain frequent itemsets are namely convened the association rules (positive, negative, no related) of given frequent itemsets of beauty accessories.

In this system, the correlation can be calculated in Equation (1) as follows:

$$\text{lift}\ (I_5\ ,\ (I_2,I_3)) = \frac{P(I_5\ U\ (I_2,I_3))}{P(I_5).P((I_2,I_3))}$$

$$= .5/(.5 * .75)$$

$$= 1.33$$

The correlation of lift $(I_5,\ (I_2,I_3))$ is greater than 1. According to the result of correlation is positively correlated, this system shows Positive result for this example.

## 4. Implementation of Framework

Sales and marketing of beauty accessories are a kind of business with high level of competition. The success of marketing's beauty accessories is influenced by its fast response and its ability in investigating of supplier representatives. Using historical data of customers' buying products, this system has investigated several buying patterns in five category associations. Thus, in this system, the sales patterns in daily applied from supplier representatives are collected and stored in transaction database as shown in Figure 3. During the process of investigation framework, we implement by using Microsoft C# language with 2005 SQL database server.

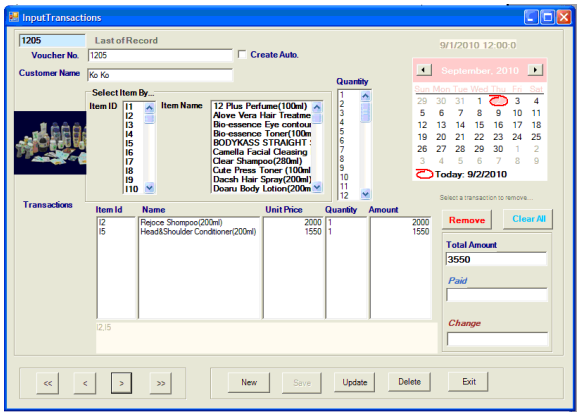


**Figure 3: Data Collection by Supplier Representative**

Figure 4 illustrates the planned beauty accessories layout of this system by generating association rules with apriori algorithm. The planned layout shows customer buying flows based on category associations which are symbolized by the daily selling. The layout also shows total number of rules which denote associate kinds of items and confidence amount of each itemsets. This planned layout will expectantly increase customer's purchase because it is designed to accommodate their buying patterns based on buying products in the past.

In this system, based on these category associations, a design of relationship (correlation) product layout is also implemented in Figure 5. This product layout is expected to give more selling convenience to the customers and investigate the beauty accessories sales and marketing for supplier representatives because it is based on the bundle most frequent chosen by the customers on a selling trip of supplier representative that might regarded as an indicator of the supplier representatives' utility.

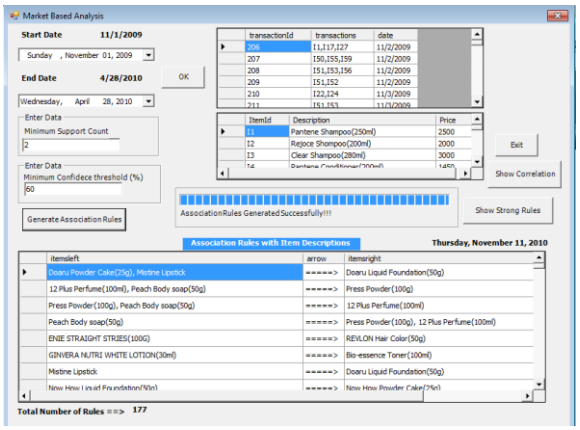


**Figure 4: Generating Association Rules**

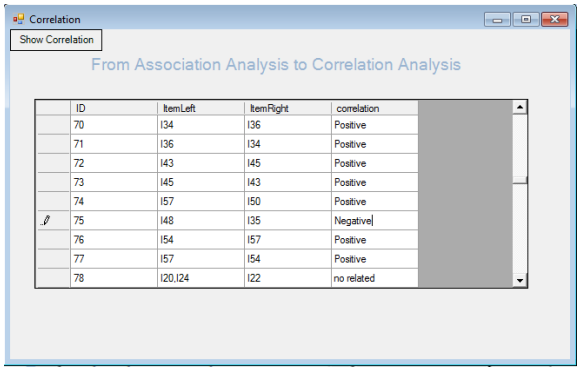


**Figure 5: Design of Correlation Product Layout**

## 5. Conclusions and Future Work

Apriori is the simplest algorithm which is used for mining of frequent patterns from the transaction database. This system's goal is to implement association rules mining for sale data using Apriori algorithm for a valuable service to the customers and profitable service to the retailer. This paper concludes that we implement a system as aside of observing customers reactions to marketing mix is by using association rule mining, another way for investigating of sales and marketing based on beauty accessories is by using correlation analysis to give correlation results.

T1, T2, T3, T5

In future work, this system enhances the understandability of association rule mining the large transaction dataset to adopt association rule mining with market basket analysis for real world transaction data. It can also be applied to the application with theoretical data mining of Market based analysis and any real world transaction database.

## Reference


[1] B. Goethals, J. Muhonen, and H. Toivonen, "Mining Non-Derivable Association Rules", Helsinki Institute for Information Technology, Department of Computer Science, University of Helsinki, Finland.

[2] D. Chandra, P. Sao and S. Gajbhiye, "Comparative Study of Apriori and DIC Algorithm", *Proceedings of the Int. Conf. on Information Science and Applications, ICISA 2010,* Chennai, India, 6th February, 2010.

[3] Direxionfunds, "Correlation Analysis: A Key Practice in Achieving Portfolio Diversification".

[4] H. Feng, Z. Shu-mao, D. Ying-shuang, "The Analysis and Improvement of Apriori Algorithm:, North China Electric Power University, Boading 071003, China.

[5] I. Surjandari and A. C. Seruni, "Design of Product Placement Layout in Retail Shop using Market Basket Analysis", Industrial Engineering Department, Faculty of Engineering, University of Indonesia, Makara Teknologi, Vol. 9, No. 2, pp.. 43-47, November 2005.

[6] L. Zhou and S. Yau, " Association Rule and Quantitative Association Rule Mining among Infrequent Items", Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago.

[7] R. Srikant* and R. Agrawal, "Mining Generalized Association Rules", IBM Almaden Research Center.

[8] V. K. Singh, V. Shah, and Y. K. Jain, "Proposing an Efficient Method for Frequent Pattern Matching", World Academy of Science, Engineering and Technology, 4th , August, 2008.