

DESIGN OF PRODUCT PLACEMENT LAYOUT IN RETAIL SHOP USING MARKET BASKET ANALYSIS

Isti Surjandari dan Annury Citra Seruni

Industrial Engineering Department, Faculty of Engineering, University of Indonesia, Depok 16424, Indonesia

E-mail: surjandari.2@osu.edu, a_citra_s@yahoo.com

Abstract

Retailing is an industry with high level of competition. It is a customer-based industry which depends on how it could be aware of what the customers' needs and requirements are. One technique most used in supermarkets is the mix merchandise. The purpose of this paper is to identify associated products, which then grouped in mix merchandise with the use of market basket analysis. This association between products then will be applied in the design layout of the product in the supermarket. The process of identifying the related products bought together in one transaction is done by using data mining technique. Apriori algorithm is chosen as a method in the data mining process. Using WEKA (Waikato Environment for Knowledge Analysis) software, the association rule between products is calculated. The results found five category association rules and fourteen sub-category association rules. These associations then will be interpreted as confidence and support to become consideration for the product layout.

Keywords: layout, market basket analysis, retail

1. Introduction

Retail industry is a kind of business with high level of competition. The success of retail business is influenced by its fast response and its ability in understanding consumers' behaviors. Retail business must focus to its consumer since retail business plays its role at the end of distribution channel.

Consumer buying behaviors can be comprehended by observing how someone interacts and reacts to the marketing mix. According to Cohen (1991), company determines the decisions related to the 4P (Product, Place, Promotion, and Price) by focusing to its consumers; while each individual considers the option to buy which products under the psychological influences of culture, attitude, experience, 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, where to buy, which brand to buy, and other choices.

Aside of observing customers reactions to marketing mix, another way for understanding customers behavior is by using historical data, which is transactional data. From customer transactional

database, it can be observed customers' shopping patterns which show associated categories or even associated products.

The following study is guided by the hypothesis, that products chosen on a shopping trip in a supermarket are an indicator of the preference interdependencies between different products or brands. The bundle chosen on the trip can be regarded as an indicator of a global utility function. In other words, the existence of such a function implies a cross-category dependence of brand choice behavior. It is hypothesized, that the global utility function related to a product bundle is the result of the marketing-mix of the underlying brands.

The objectives of this study were to discover the associated products, and to determine the proposed product placement layout based on the associated products. This study was carried out with several boundaries.

2. Method

This study used market basket analysis to find association rules between sets of items in transactional database and the Apriori algorithm is used in the data mining process. The goal of the association discovery

was to find items that imply the presence of other items. For example, 90% of customers that purchase frozen pizza also buy soda.

The problem of finding association rules was first introduced by Agrawal, et. al. [1], which is described as follows: Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of literals called items. Let D be a set of transactions where each transaction T is a set of items such that $T \subset I$. An association rule is an implication of the form $X \Rightarrow Y$ where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c if $c\%$ of transactions in D that contain X also contain Y . The rule $X \Rightarrow Y$ holds in the transaction set D with support s if $s\%$ of transactions in D contain $X \cup Y$. Given a set of transactions D , the problem of mining association rules is to generate all the association rules that have support and confidence greater than the user-specified minimum support (called *minsup*) and minimum confidence (called *minconf*), respectively.

Apriori Algorithm was also discussed by Agrawal et. al. [2], which is considered as one of the most contributions to this subject. Its main algorithm, Apriori, has affected not only the association rule mining community, but other data mining fields as well. The Apriori algorithm for finding all large item

sets makes multiple passes over the database. In the first pass, the algorithm counts item occurrences to determine large 1-item sets. The subsequent pass, say pass k , consist of two steps. First, the large item sets L_{k-1} found in the $(k-1)$ -th pass are used to generate the candidate item sets C_k . Then, all those item sets which have some $(k-1)$ subset that is not in L_{k-1} are deleted, yielding C_k . Figure 1 gives the Apriori algorithm as described by Agrawal et. al. [2].

```

1)  $L_1 = \{\text{large 1-item sets}\};$ 
2) for (  $k = 2; L_{k-1} \neq \emptyset; k++$  ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1});$  // New candidates
4)   forall transactions  $t \in D$  do begin
5)      $C_t = \text{subset}(C_k, t);$  // Candidates contained in  $t$ 
6)     forall candidates  $c \in C_t$  do
7)        $c.\text{count}++;$ 
8)   end
9)    $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{minsup} \}$ 
10) end
11) Answer =  $\bigcup_k L_k;$ 

```

Figure 1. Apriori Algorithm

Table 1. Product Categories

rice-flour-sugar	eggs	fat	toiletries	apparel
freshener-insecticides	seasoning	milk	sanitary	accessory
instant food	snack	baby products	cleaner	stationery
ready-to-eat	tea-coffee	cigarettes	electronics	houseware
ready-to-drink	bread filling	healthcare	amusements	etc.

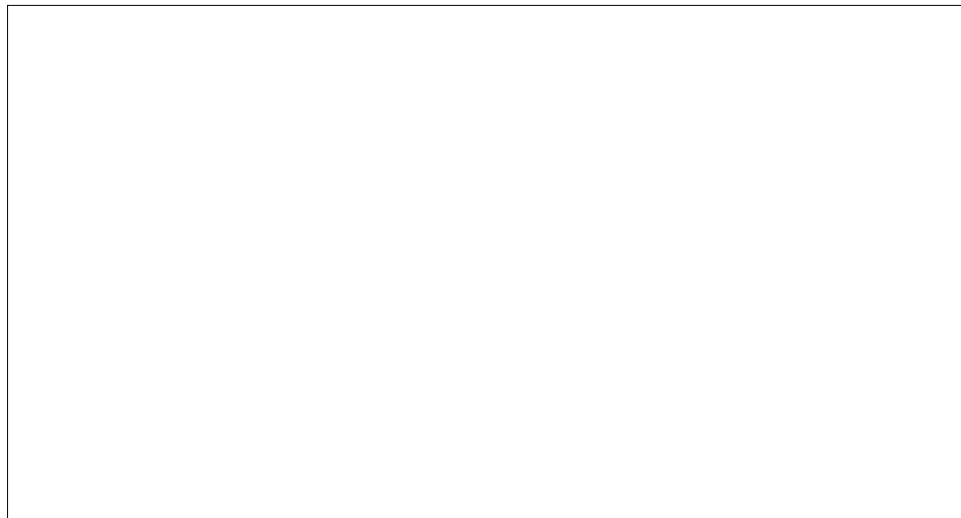


Figure 2. ARFF Sample of Transactional Data

For the purpose of this study, empirical data is obtained from a retail store at Kukusan, Depok. A three-months transactional data of the retail's costumers was obtained and processed using WEKA (Waikato Environment for Knowledge Analysis) [3]. The products which were studied in this research consist of 704 products. These products then classified as more specific categories (i.e., 25 categories) and listed in Table 1. The main objective of the classification was to discover a more detail of customers shopping patterns. Another objective was to avoid bias in the data processing.

The first step was to choose which transactions to be included in this study. We sorted and selected transactions that had two or more product categories, which consist of 1049 transactions. These transactions were then converted to ARFF format [4]. WEKA expects the data to be in ARFF format, therefore it is necessary to have information about the type of each attribute, which cannot be automatically deduced from the attribute values [3]. Figure 2 gives the ARFF sample of the data.

3. Result and Discussion

Initial processing of the 1049 transactions using WEKA has given the results of 5 category associations from 18 iterations with minimum support of 0.1 and minimum confidence of 0.9, as can be seen in Table 2. The minimum support with a value of 0.1 indicate that a category must be present in at least 0.1 of the whole three-months data. While a value of 0.9 of the minimum confidence indicate that we are sure by at least 100% that the associations are true based on the historical data. For example, the first association in table 2, specify that we are sure by 100% that if a

customer buy a product in stationery category, he/she will also buy an amusement category product. The value of 1 (100%) comes from the amusements' consequent support divided by stationery's antecedent support.

In order to obtain more specific results, a succeeding processing of the category association was conducted. The following study has given the results of 14 subcategory associations from 18 iterations with minimum support of 0.1 and minimum confidence of 0.9. These subcategory associations are listed in Table 3.

The current product layout applied in the store is the grid layout, which is the best layout to be applied in a convenience store [5]. Figure 3 illustrates current product layout in the retail store. Figure 4 illustrates proposed product layout for the retail store. The proposed layout shows customer shopping flows based on category associations which are symbolized by the colored arrows. The layout also shows various colors of display which denote various kinds and sizes of display. The proposed layout will expectantly increase customers' purchase because it is designed to accommodate their shopping patterns, based on their behaviors in the past.

Table 2. Category Associations

Antecedent	Antecedent Support (*1/1049)	Consequent	Consequent Support (*1/1049)	Confidence
stationery	181	amusements	181	1
tea-coffee	158	rice-flour-sugar	158	1
cigarettes	143	snack	132	0.92
rice-flour-sugar & fat	120	eggs	109	0.91
fats & cleaner	118	eggs	107	0.91

Table 3. Subcategory Associations

Antecedent	Antecedent Support (*1/1049)	Consequent	Consequent Support (*1/1049)	Confidence
drawing tools	34	toys	34	1
coffee	97	sugar	97	1
tea	60	sugar	60	1
cigarette	83	candy	83	1
cigarette	36	chocolate	36	1
butter	38	eggs	36	0.95
tepung & mentega	19	eggs	18	0.95
flour	48	eggs	45	0.94
rice & butter	14	eggs	13	0.93
flour & cooking oil	24	eggs	22	0.92
cooking oil	72	eggs	65	0.90
cooking oil & detergent	33	eggs	30	0.91
cooking oil	101	eggs	91	0.90
detergent	40	eggs	36	0.90

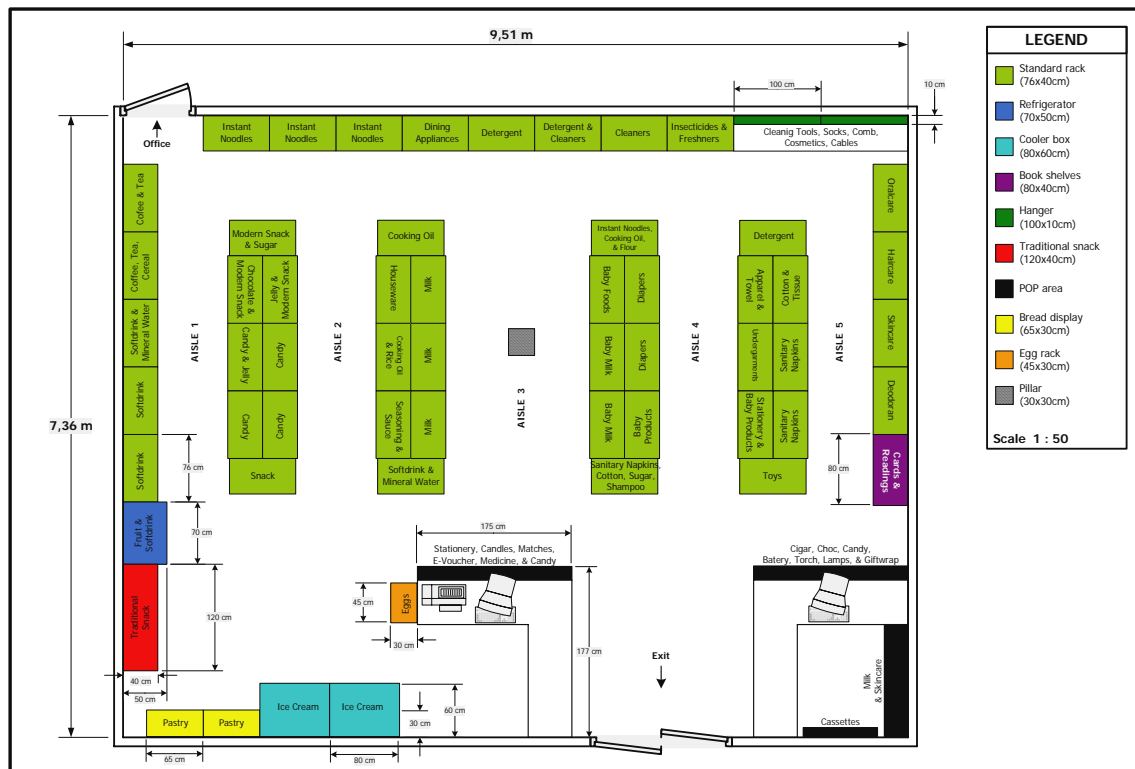


Figure 3. Current Layout



Figure 4. Proposed Layout

4. Conclusions

Using historical data of customers' shopping behaviors, this study has discovered several shopping patterns in five category associations and fourteen subcategory associations. The absolute relations for category associations are shown with the confidence value of 1. These associations are the relation of stationery to amusements, and tea-coffee to rice-flour-sugar. The absolute relations for subcategory associations are also shown with the confidence value of 1. These associations are the relation of drawing tools to toys, coffee to sugar, tea to sugar, cigarette to candy, and cigarette to chocolate. Based on those category associations, a design of product placement layout was proposed. This proposed layout is expected to give more shopping convenience to the customers, because it is based on the bundle most frequent chosen by the customers on a shopping trip in the retail store

that might regarded as an indicator of the customers' utility.

References

- [1] R. Agrawal, T. Imielinski, A. Swami, Proceeding of the ACM SIGMOD Conference on Management of Data, 1993.
- [2] R. Agrawal, R. Srikant. The International Conference on Very Large Databases, 1994.
- [3] I. H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, chapter 8, <http://www.cs.waikato.ac.nz>, 2000.
- [4] A. Roberts, Guide to WEKA, <http://www.comp.leeds.ac.uk/andyr>, 2005.
- [5] M. Levy, B. Weitz, Retailing Management, McGraw-Hill, New York, 2001.