

# Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data

Fachrul Kurniawan <sup>a, 1, \*</sup>, Binti Umayah <sup>a</sup>, Jihad Hammad <sup>c</sup>,  
Supeno Mardi Susiki Nugroho <sup>b</sup>, Mochammad Hariadi <sup>b</sup>

<sup>a</sup>Dept. of Informatics Eng., Maulana Malik Ibrahim State Islamic Univ., Jl. Gajayana No.50, Malang 65144, Indonesia

<sup>b</sup>Dept. of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Surabaya 60111, Indonesia

<sup>c</sup>ICT Faculty, AlQuds Open University, Beit Jalla-The Main road-Khallat Al Badd, Bethlehem, Palestine

<sup>1</sup>fachrulk@ti.uin-malang.ac.id\*

\* corresponding author

## ARTICLE INFO

### Article history:

Received 30 August 2017

Revised 13 September 2017

Accepted 1 October 2017

Published online 8 January 2018

### Keywords:

Data mining

Market basket analysis

Association rule

Apriori

## ABSTRACT

Transaction data is a set of recording data result in connections with sales-purchase activities at a particular company. In these recent years, transaction data have been prevalently used as research objects in means of discovering new information. One of the possible attempts is to design an application that can be used to analyze the existing transaction data. That application has the quality of market basket analysis. In addition, the application is designed to be desktop-based whose components are able to process as well as re-log the existing transaction data. The used method in designing this application is by way of following the existing steps on data mining technique. The trial result showed that the development and the implementation of market basket analysis application through association rule method using apriori algorithm could work well. With the means of confidence value of 46.69% and support value of 1.78%, and the amount of the generated rule was 30 rules.

This is an open access article under the CC BY-SA license  
(<https://creativecommons.org/licenses/by-sa/4.0/>).

## I. Introduction

Consumer behavior is a consumer activity in deciding to purchase, use, as well as consume the purchased goods and services including in the customer factors which can give a rise to their decisions whether to purchase and use products. Every customers have different needs and inclinations as well as have different behaviors in fulfilling those things. However, in the event of different behaviors to fulfill their needs, they still share some similarities, one of them is desiring to maximize their satisfactions in consuming a necessary product or service.

Of that consumption activity, that can be inferred as to the behavior, pattern, or habit that the customers do in fulfilling their needs and desires. That behavior can be identified by way of the logging carried out by the intermediary provider of consumer needs (supermarket). The logging is brought off in view of the requisite of documentation and history data identification over the carrying out of transaction activity.

In these recent years, transaction data have been commonly used as research and analysis objects for researchers. In this study, also, transaction data are to be re-processed/re-explored to generate more valuable information. For instance, information of an item whose sales is the highest. Besides, information can be utilized in regard with the stock addition of that item. Moreover, from transaction data there can be utilized as to the relation of each purchased item inside the customer' basket. By that information, we can make use of it for effective product display/assortment to attract customers' interest. The commonly-used application to analyze transaction data customers' shopping basket is market basket analysis.

Market basket analysis is one of the modes from data mining technique prevalently employed to analyze items/goods in one or more shopping baskets that a customer has in one particular moment

[1]. Market basket analysis application ought to be designed and implemented at a supermarket not only owing to being able to help the sales promotion design but also able to be made as a reference to re-manage item stock' incoming and outcoming in the warehouse. In this study, market analysis application will be implemented at BC UIN Malang supermarket, in regard with its inability to use transaction data. This application is expected to work well and is able to generate the desired result.

#### A. Market Basket Analysis

Market Basket Analysis is an analysis mode performed on customer behavior whilst shopping at a supermarket through the means identifying association and connections among various items placed by the customers in their shopping baskets [2]. In specific, Market Basket Analysis aims at simultaneously identifying the most frequently-purchased items by customers. Here, item is depicted as several kinds of products in supermarket.

Using market basket analysis mode, a knowledge of what are the items oftenly purchased by the customers simultaneously and having an opportunity to be promoted can be obtained. With regards to the objective of market basket analysis mode to decide which products that customers purchase at the same time, whereby the name of this mode is taken from the behavior of the customers in placing shopping products into their shopping baskets or shopping list. Over identifying shopping basket pattern of a customer will significantly be able to help a company in using that information in respect of business strategy needs, one of them is placing the most frequently-purchased products simultaneously into one specific area.

#### B. Association Rule

Association rule is related to the statement of "what' with what". This matter can be in a form of statement on transaction activity carried out by the customers at a supermarket. From that statement, there has a strong relation to the study of customer transaction data database to determine the habit of a purchased product with what product, thus, association rule is frequently referred as market basket analysis [1], [2].

The significance of an associative rule can be figured in the presence of two parameters, namely support and confidence. Support (supporting value) is the percentage of combinations of product items in the database. While confidence (certainty value) is a value to determine the strength of inter-item relationships in association rules.

$$S = \frac{\sum(Ta+Tc)}{\sum(T)} \quad (1)$$

$$\text{Support (A)} = \frac{\text{The Number of Transaction that Contains A}}{\text{Total Transaction}} \quad (2)$$

where:

$$S = \text{Support}$$

$$\sum(Ta + Tc) = \text{the number of transaction that contains } \textit{antecedent} \text{ and } \textit{consequent}.$$

$$\sum(T) = \text{the number of transaction.}$$

$$C = \frac{\sum(Ta+Tc)}{\sum(Ta)} \quad (3)$$

$$\text{Confidence} = P(A | B) = \frac{\text{The Number of Transaction that Contains A and B}}{\text{Total Transaction that Contains A}} \quad (4)$$

where:

$$C = \text{Confidence}$$

$$\sum(Ta + Tc) = \text{the number of transaction that contains } \textit{antecedent} \text{ and } \textit{consequent}.$$

$$\sum(Ta) = \text{the number of transaction that contains } \textit{antecedent}.$$

#### C. Customer Behaviour

Definition of customer behavior is a dynamic interaction between cognition, affection, behavior, and its environment whereby someone performs exchange activities in their regular basis [3]. In view of this statement, there are three significant matters to grasp, namely:

1. Customer behavior bears dynamic characteristic, thus, hard to predict.
2. Involving interaction, like cognition, affection, behavior, and the occurrences around customers,
3. Involving exchange, like the exchange of item and money from merchant to customer

Four factors that could give a rise to customer purchase in shopping [4], some of them were:

1. Cultural Factor
2. Social Factor
3. Personal Factor
4. Psychological Factor

There were three variables that must be regarded in understanding customer behavior, namely stimulus variable, response variable, and intervening variable[5].

#### D. Data Mining

The researcher argued that Data Mining had garnered the interest of information industry and public during in that recent years, this was brought about by massive data availability and the necessities to render information data to usable knowledge [6], [7].

Data Mining is prevalently depicted as retrieval pattern on a large set of crude data so that some unveiled knowledge in the data can be found. Data Mining is considered a major step in the process of knowledge retrieval in the database. Data Mining inherits many aspects and techniques from established fields of science first. Data Mining also has long roots from fields such as artificial intelligence (Artificial Intelligent), machine learning, statistics, database and also information retrieval.

## II. Methods

In association rule mode, the implemented algorithm in the making of market basket analysis application is Apriori algorithm. Algorithm is used to develop frequent itemset using 1-item first, then value support of every item would be later counted. Item whose support value was above the minimum support value was selected as 1-itemset high frequency pattern and as 2-itemset candidates. By that of 1-itemset, the development of frequent itemset into 2-itemset which would then the value of confidence be calculated next was recursively brought off.

## III. Results and Discussion

This application was implemented in private supermarket of UIN Malang Business Center, over inputting the collected transaction data of which the data used were transaction data in BC UIN Malang on October 1st, 2014 with 1553 collected data. In additions of the stored transaction receipts of 890753 up to 891319 transaction receipts.

The further step was inputting minimum support and minimum value of confidence. In illustration, user wanted to insert transaction data with receipts from 890753 up to 890853. From that receipt there were 56 transaction data with 20 items could be retrieved. Those 56 transaction data were obtained from the transaction receipts whose item amounts were minimumly 2 in 1 single transaction

The trial system was carried out in means of system trial of market basket analysis application by way of association rule counting mode using apriori algorithm. In this study, a discussion of the tests that had been carried out on the system and evaluation of the results responded by the system. From this test, how patterns differences of the generated data would be known. The test was performed on 56 data by entering the initial value

- Minimum support= 3
- Minimum confidence= 10

Fig. 2 shows the data that would be used for the test, which were the receipts between 890753 up to 890853 with the number of transaction data was 56 data assorted in respect of “Receipts”

The calculation of the support value of 1 item can be obtained through the following formula:

$$\text{Support (A)} = \frac{\text{The Number of Transaction which contains A}}{\text{Total Transaction}} \quad (5)$$

While the support value for 2 items can be obtained with the following formula:

$$Support(A, B) = \frac{Transaction\ Number\ Which\ Contains\ A\ and\ B}{Total\ Transaction} \tag{6}$$

Confidence value calculation was generated of this following formula:

$$Confidence = P(A|B) = \frac{The\ Number\ of\ Transaction\ Which\ Contains\ A\ and\ B}{Total\ Transaction\ which\ contains\ A} \tag{7}$$

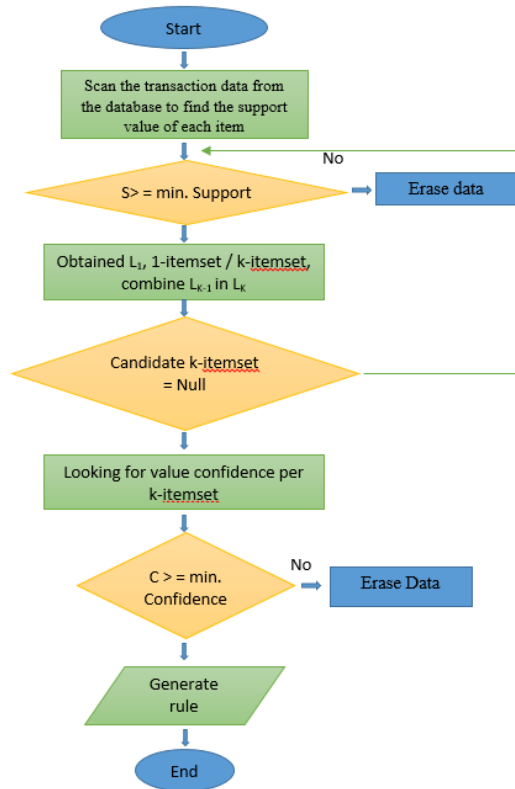


Fig. 1. Apriori algorithm design

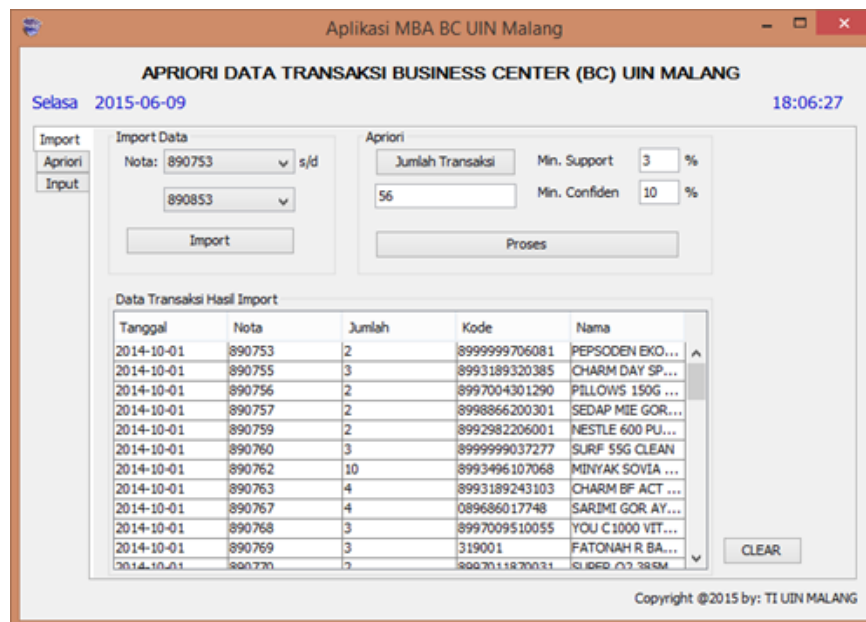


Fig. 2. Application implementation

In calculating support 1-itemset with 2-itemset, the inputted minimum support was differentiated. Minimum support for 2-itemset calculation was deducted as if it was evened out, there would be no candidate for 2-itemset. Hence, the generated output was only constrained o 1-itemset. Thereby, on the test of this study, for the calculation of 2-itemset, the researcher deliberate entered different minimum support value with minimum support value on 1-itemset calculation by means of generating the desired recommendation outcome.

Nota	Kode	1-Itemset	Support(%)
890753	8999999706081	PEPSODEN EKONOMI 75	5.357142857142857
890753	8886008101053	AQUA 600 BTL	5.357142857142857
890757	8998866200301	SEDAP MIE GORENG	3.571428571428571
890759	8992982206001	NESTLE 600 PURE LIFE BTL	3.571428571428571
890760	89926581	POLO MINT	3.571428571428571
890762	8999999706081	PEPSODEN EKONOMI 75	5.357142857142857
890762	910002	GULA BIASA 1 KG	3.571428571428571
890763	8993379260835	LERVIA 250 MILK ROSE BTL	3.571428571428571
890767	8995899250143	KOBE BON CABE LVL 15 7,5G	3.571428571428571
890767	8992388111145	ABC MI SUP TOMAT	3.571428571428571
890768	8996001304990	ROMA SARI GANDUM 155...	3.571428571428571
890771	8992931025202	MULTI MPH 200-048 POP	5.357142857142857
890772	8998009050053	ULTRA 250 KACANG IJO	5.357142857142857
890791	8992388111145	ABC MI SUP TOMAT	3.571428571428571
890791	8998866200578	SEDAP MIE KARI SPL	3.571428571428571
890791	227001	YAKULT (5)	3.571428571428571

Fig. 3. The calculation result of 1 item support value

**REKOMENDASI**

- Jika membeli " SANGHAI IKAN MUTIARA " maka akan membeli " ROMA SARI GANDUM 155G S " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " ROMA SARI GANDUM 155G SUS " maka akan membeli " SANGHAI IKAN MUTIA " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " INDOMI GOR PDS " maka akan membeli " PEPSODEN HERBAL 75 " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " PEPSODEN HERBAL 75 " maka akan membeli " INDOMI GOR PDS " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " INDOMI GOR PDS " maka akan membeli " PEPSODEN CENTER F 160 " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " PEPSODEN CENTER F 160 " maka akan membeli " INDOMI GOR PDS " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " YAKULT (5) " maka akan membeli " ABC MI SUP TOMAT " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " ABC MI SUP TOMAT " maka akan membeli " YAKULT (5) " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " YAKULT (5) " maka akan membeli " SEDAP MIE KARI SPL " dengan Supp 1.7857142857142856% dan Conf 50.0%
- Jika membeli " SEDAP MIE KARI SPL " maka akan membeli " YAKULT (5) "

Fig. 4. Confidence value calculation result

#### IV. Conclusions

In the test and analysis that have been carried through, it is concluded that of development and implementation of market basket analysis over association rule mode using apriori algorithm on the transaction data of Business Centre (BC) UIN Malang supermarket, can perform well. With the confidence value average of 46.69% from support value 1.78% and the generated rules are 30 rules. The data transaction rule/ pattern which was retrieved had a low association tendency. That was due to the owned data and the analyzed data did not really support association analyses/ the relation between items.

#### References

- [1] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016.
- [2] A. Mansur and T. Kuncoro, "Product Inventory Predictions at Small Medium Enterprise Using Market Basket Analysis Approach-Neural Networks," *Procedia Econ. Financ.*, vol. 4, no. Icsmed, pp. 312–320, 2012.
- [3] X. Su, "Intertemporal Pricing with Strategic Customer Behavior," *Manage. Sci.*, vol. 53, no. 5, pp. 726–741, 2007.
- [4] G. Armstrong, S. Adam, S. Denize, and P. Kotler, *Armstrong, G., Adam, S., Denize, S., & Kotler, P.* Pearson Australia., 2014.
- [5] E. Sherman, A. Mathur, and R. B. Smith, "Store Environment and Consumer Purchase Behavior: Mediating Role of Consumer Emotions," *Psychol. Mark.*, vol. 14, no. 4, pp. 361–378, 1997.
- [6] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015.
- [7] A. Bertoni and T. Larsson, "ScienceDirect Data Mining in Product Service Systems Design: Literature Review and Research Questions," *Procedia CIRP*, vol. 64, pp. 306–311, 2017.