



Market Basket Analysis for Determination of Consumer Behavior at XYZ Stores Using R Programming

Kelik Sussolaikah

Faculty of Information Technology, Universitas PGRI Madiun, Jl. Auri Madiun, East Jawa, Indonesia7

kelik@unipma.ac.id

Abstract. Data mining is one of the fields of science in the world of informatics which has an important role, especially with regard to data. There are many algorithms and methods that can be used to process data. The paper this time the author tries to conduct research on consumer behavior by using one of the data mining techniques, namely market basket analysis. This research uses the R Programming tool, where it is hoped that the research can be carried out effectively and efficiently. Based on the research conducted, it is known that there has been a significant purchase of several items that have been described as a plot. The tendency of consumers to buy several items followed by other items can be a consideration for arranging the layout of goods on the sales shelf or arranging product stock in a supermarket.

Keywords: Data Mining, Market Basket Analysis, Plot, R Programming

(Received 2021-05-25, Accepted 2021-05-31, Available Online by 2021-05-31)

1. Introduction

The world of marketing is growing and business competition is inevitable. In a supermarket, there is often a problem with the absence of stock in certain products. This results in consumer disappointment if it lasts for a long time, moreover, these products are in great demand by consumers. Therefore it is necessary to have good marketing management so that the fulfillment of consumer needs can be fulfilled. Research on product sales analysis has been carried out under the title Data Mining Estimated Banner Production Using the Apriori Algorithm (Case Study: CV. Mentari Persada Medan) in this research it is known that the most production of banners is Satin and Ritrama stickers, with a priori algorithm so you can see the types The most widely produced banners that meet support and minimum confidence.[1]

Another study conducted data mining implementation on medical equipment supplies entitled Implementation of Apriori Data Mining Algorithms in the Medical Equipment Supply System. In his research, it is known that the application of a priori algorithm in data mining techniques is very efficient and can accelerate the process of forming a trend of itemset combination patterns from the sale of medical devices at the Kelambir-2 Pharmacy in Medan, with the highest support and confidence is Uric

Acid Stick - Sugar Stick and Colestrol-Stick Sugar.[2] Pasaribu conducted research with the title Decision Support System for Analysis of Sales Patterns of Goods Using the Apriori Algorithm (Case study: Lucky Swalayan). In his research, Pasaribu created a program that can be used to process a large number of sales data into an alternative decision that can help supermarket owners see which products are often purchased by consumers and can also be used to assist in arranging the layout of goods at the supermarket so that the goods are frequently purchased.[3] Consumers buy simultaneously can be placed close together, while items that are rarely purchased can be given discounts to attract consumers. In research conducted by Ari and Laili, the system development model is carried out using a prototype where customers and users will be directly involved in the process where the research results are in the form of transaction data analysis using market basket analysis with 4 product combinations based on the largest support x confidence value. with the results in the form of numbers of possible transactions related to the products sold.[4]

Rina et al conducted a study entitled Perancangan Market Basket Analysis Menggunakan Association Rule untuk Pendukung Keputusan Promosi pada Sistem Penjualan Sun Young Cell. This study aims to design Market Basket Analysis for sales system at Sun Young Cell as a sales transaction information processing application which can provide analytical results representing sales rates according to trends in certain product sales. The technique used is association rule that describes the patterns of transaction by using rules to support decision making process. [5]

Based on this background and referring to previous research, the author is interested in conducting research with the title Market Basket Analysis to Determine Consumer Behavior Using R Programming.

2. Basic of Theory

2.1. Data Mining

Data mining is a process in making decisions by looking for patterns of information contained in data. Pattern search can be done using a query or with the help of certain applications so that it can automatically search for patterns in a database. Data mining is also a step in the Knowledge Discovery in Databases process.[6]

There are several data mining techniques as follows :[7]

1. Classification

Defines a new data record into a previously defined class.

Some of the applications of calcification are:

- Direct sales
- Fraud Detection
- Customer Attrition (Churn)

2. Clustering

It is unsupervised learning in which the data set is partitioned into groups so that the elements have a shared set of properties with a high level of similarity within one group, while between groups has a low level of similarity.

The applications of clustering are:

- Market segmentation
- Document clustering

3. Association Rule

Detect a set of attributes that occur together in frequent frequencies and form a number of rules from these sets.

The applications of the association rule are:

- Marketing and sales promotion
- Supermarket shelf management
- Inventory management

4. Sequential Patern

Look for a number of events that generally occur together.

5. Regression

Predict the value of a given continuous variable based on the value of the other variable based on the assumption of linear or nonlinear dependence.

6. Deviation Detection

Perform anomaly detection automatically to identify the habits of an entity and establish a number of norms through pattern discovery.

2.2. Market Basket Analysis

Shopping basket analysis or market basket analysis is a discussion in data mining which aims to determine the relationship between products purchased by consumers in one transaction. In the market basket analysis, we will find out what products consumers often buy at the same time.

With the existence of a market basket analysis, it will be possible to know the pattern of purchases that occur so that it is expected to be able to help supermarkets and marketing businesses to maximize the arrangement of goods and inventory

2.3. Association Rule Analysis

Is a procedure in Market Basket Analysis that is used to find relationships between items in a data set and display them in the form of Association Rules. The function of Association Rules is often called "market basket analysis", which is used to find the relation or correlation between a set of items². The type of association rule can be stated for example: "70% of the people who buy noodles, juice and sauce will also buy plain bread".

Association rules capture items or events in large data sets containing transaction data.

With advances in technology, sales data can be stored in large amounts called "data basket."

The association rules defined in the data basket are used for promotional purposes, catalog design, customer segmentation and marketing targets.

2.4. R Programming

R programming is a programming language that can be used to perform data manipulation, simulation, calculation and graphic display. R has the ability to analyze data very effectively and is equipped with array and matrix processing operators. R also has the ability to display graphics that are very sophisticated as well as modeling for his data. The R language is similar to the S language developed by Rick Becker, John Chambers and Allan Wilks at AT&T Bell Laboratories.[8]

Software R is an open-source software from the commercial version of the S programming language based on the S language is S plus. Software R has the capability that is not inferior to commercial data processing packages, even in some ways its capabilities are much better. The R language has not been widely known by the Indonesian people but the R language has received a good response from statisticians around the world. Apart from being open source, the R language also has other advantages, namely that it is multi-platform so that it can run on various operating systems. The R language also has thousands of packages that can be used as needed and can be developed via github or the developer version

3. Methods

3.1. Research Methods and Data Collection

In this study the authors used a descriptive research method because in this study a number of facts will be presented. The research that will be carried out begins with formulating the problem, then conducts a literature review by referring to references to previous studies. Then, data will be collected through secondary data which is taken for analysis using R Programming.

The results of the shopping cart analysis that has been carried out with R Programming will be displayed in the form of a plot, so that the patterns of consumer purchases that occur will be seen. The next step is to arrange the research flow so that the research runs according to the steps determined by the author with the hope that the research will run smoothly and the research results can be maximized.

3.2. Data analysis

The research carried out took secondary data so that the existing data would be further analyzed as a shopping cart analysis so that the support and confidence were determined first. From determination

3.3. Research Flow

The research flow carried out by the author is as follows:

- a. Retrieving secondary data
- b. Save data into csv
- c. Open R Programming and load the required libraries
- d. Perform shopping cart analysis process
- e. Visualization of analysis results'

The research flow is described as in Figure 1.

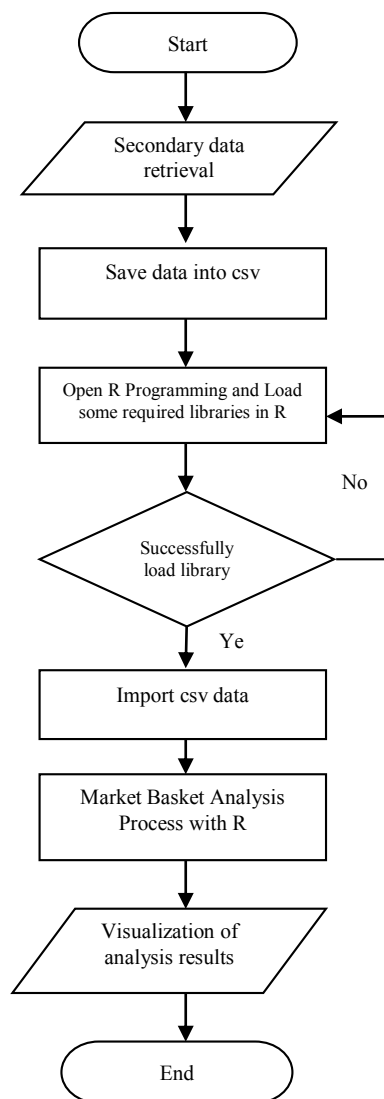


Figure 1. Research Flow

4. Results and Discussion

4.1. Problem analysis

Some of the problems that arise in the marketing world, especially retail, are a shortage of stock of goods and improper placement of goods that affect purchasing patterns. Starting from a problem like this, an analysis of the buying pattern can be carried out with a shopping cart analysis.

4.2. Sales Itemset Data

The following is sales data which is used as a shopping cart analysis.

citrus fruit semi-finished bread
tropical fruit yogurt
whole milk
pip fruit yogurt
other vegetables whole milk
whole milk butter
rolls / buns
other vegetables UHT-milk
pot plants
whole milk cereals
tropical fruit other vegetables
citrus fruit tropical fruit
beef
frankfurter rolls / buns
chicken tropical fruit
butter sugar
fruit / vegetable juice
packaged fruit / vegetables
chocolate
specialty bar
other vegetables

The data item above is a sample item from a number of 10,000 existing transactions. By observing the existing transaction data, it can be seen that several items are the same, which means that when someone buys item A, he also buys items B, C and so on.

When described in mathematics are:

- Item set: $I = \{i_1, i_2, \dots, i_n\}$
- Transaction: $tn = \{ij, ik, \dots, in\}$
- Rules: $\{i_1, i_2\} \Rightarrow \{ik\}$

If we describe it further, it is as follows:

"If someone buys an item in the item set on the left, that person will also buy the item on the right."

$\{\text{coffee, sugar}\} \Rightarrow \{\text{milk}\}$

When a consumer buys coffee and sugar, he also buys milk.

From this it can be seen that there are three important things in association rules, namely:

- Support: shows the percentage of the number of transactions that contain the item.
- Confidence: shows the percentage contained in transactions containing items
- Lift ratio: is an important parameter besides support and confidence. The lift ratio measures how important the rule is based on the value of support and confidence.

4.3. Market Basket Analysis with R

After successfully importing the data set, we plot the frequency items for the top 20 items. This is depicted in Figure 4.

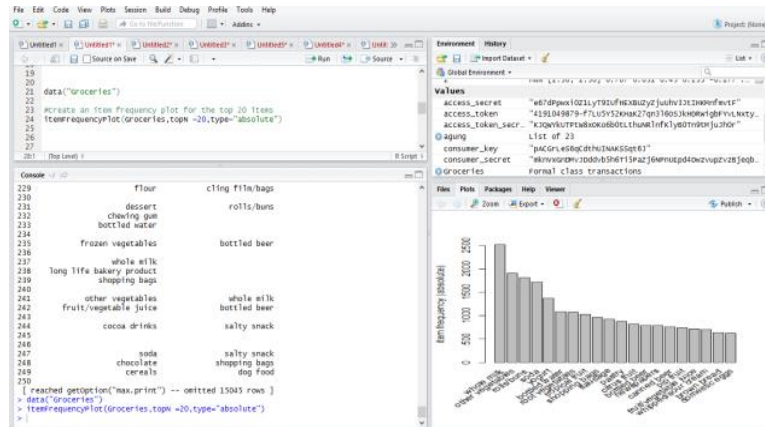


Figure 4. Process of Create Item Frequency

The results of the plot of the frequency items are as shown in Figure 5

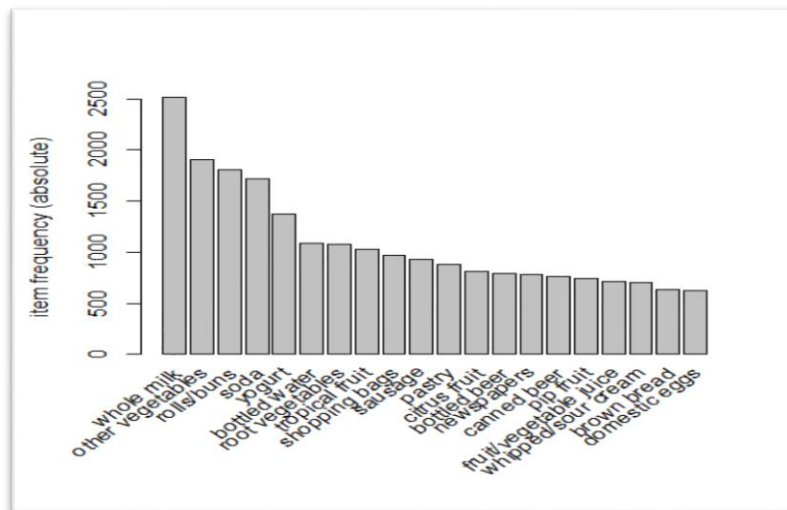


Figure 5. Item Frequency Plot

Successfully plotting frequency items for the top 20 items, the next step is to make rules, namely:

- determine the minimum support 0.001
- determine the minimum confidence 0.8
- determine the top 5 rules

this step is depicted as in Figure 6.

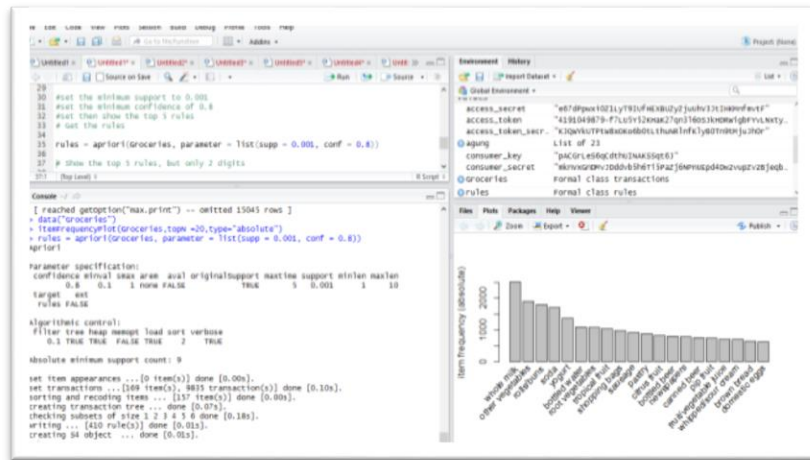


Figure 6. Top 5 Rules

Top 5 Frequency Items as shown in Figure 7.

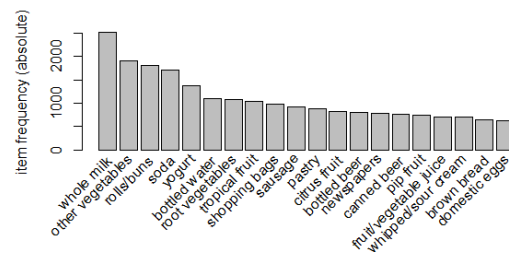


Figure 7. Top 5 Frequency Items

The output is as follows (Figure 8):

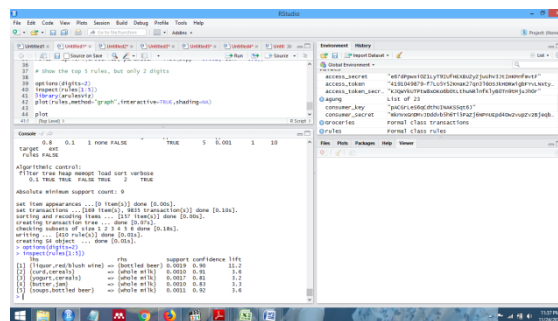


Figure 8. Output on R

Illustrated by the table can be seen in Table 1.

Table 1. Example Output

No	Lhs	Rhs	Support	Confidence	Lift
1	{liquor,red/blush wine}	=>{bottled beer}	0.0019	0.90	11.2
2	{curd,cereals}	=>{whole milk}	0.0010	0.91	3.6
3	{yogurt,cereals}	=>{whole milk}	0.0017	0.81	3.2
4	{butter,jam}	=>{whole milk}	0.0010	0.83	3.3
5	{soups,bottled beer}	=>{whole milk}	0.0011	0.92	3.6

Visualization of the results in the form of a plot (random) as shown in Figure 9.

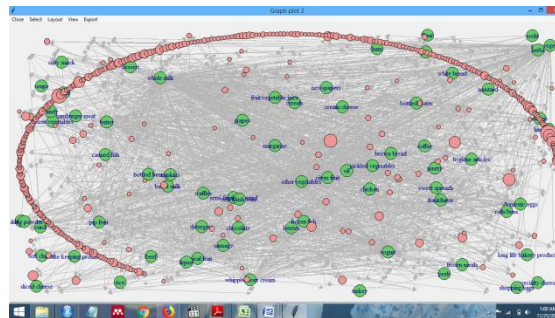


Figure 9. Visualization of Random's Layout Plot

Visualization of the results in the form of a plot (circle) as shown in Figure 10

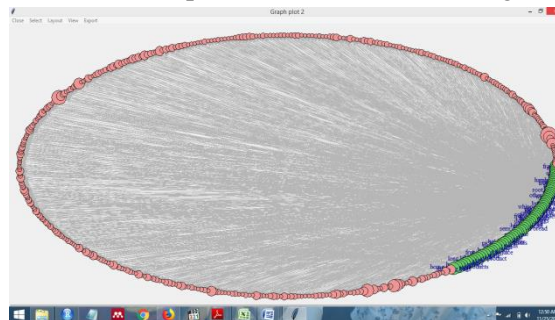


Figure 10. Visualization of Circle's Layout plot

Visualization of the Fruchterman-Reingol plot as shown in Figure 11

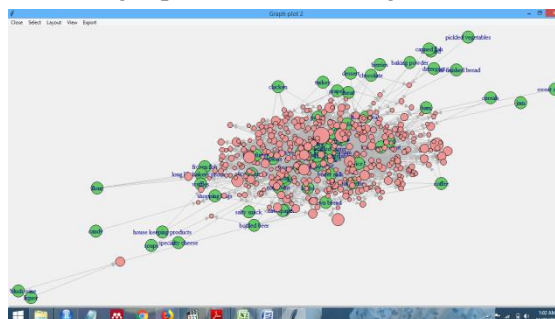


Figure 11. Visualization of the Fruchterman-Reingol plot

Visualization of Kamada Kawai Plots as shown in Figure 12.

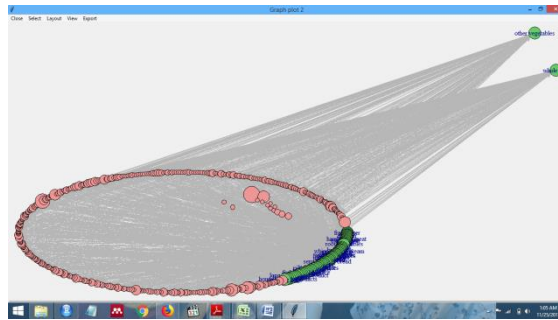


Figure 12. Visualization of Kamada Kawai Plots

Visualization of the Reingold_Tilford Plot as shown in Figure 13.

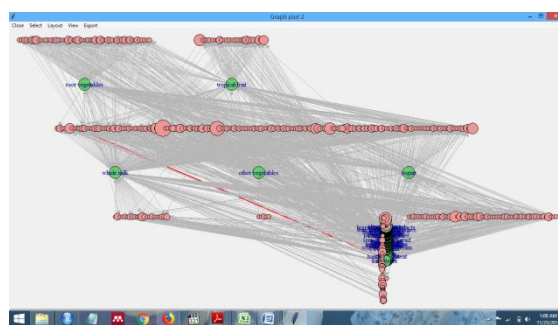


Figure 13. Visualization of the Reingold_Tilford Plot

5. Conclusions

Based on the shopping cart analysis, conclusions can be drawn:

1. The implementation and utilization of R Programming is able to make the market basket analysis process run faster and more effectively, even with very large data. This is done by utilizing the available libraries in R Programming.
2. Looking at the output results and visualization that shows the patterns that occur in transactions can be known quickly, for example, as in no.3 table 1, that consumers who buy yogurt are certain 81% also buy whole milk, and so on.

During the analysis process, there were several obstacles faced, one of which was when running R Programming experienced an error in program execution. This research is built on the author's assumptions and conclusions and visualization are still not optimal.

The author provides suggestions to the next researcher to make improvements in related research, a more interactive visualization can be added and the output may be added to how the layout of goods at supermarkets or retail is able to attract consumer interest and increase sales productivity.

References

- [1] Dewi Puspitasari, "Data Mining Perkiraan Produksi Spanduk Dengan Algoritma Apriori (Studi Kasus: CV. Mentari Persada Medan)," vol. IX, Mar. 2015.
- [2] K. Tampubolon, H. Saragih, B. Reza, K. Epicentrum, A. Asosiasi, and A. Apriori, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan," *Inf. dan Teknol. Ilm.*, pp. 93–106, 2013, doi: 2339-210X.
- [3] Safar Riduan Pasaribu, "Sistem Pendukung Keputusan Analisa Pola Penjualan Barang Dengan Algoritma Apriori (Studi Kasus: Lucky Swalayan)," *Pelita Infomatika Budi Darma*, vol. VI, no. Nomor 2, Apr. 2014.
- [4] A. Muzakir and L. Adha, "Market Basket Analysis (Mba) Pada Situs Web E-Commerce Zakiyah

- Collection,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 2, p. 459, 2016, doi: 10.24176/simet.v7i2.755.
- [5] R. Halim and P. A. Jusia, “Perancangan Market Basket Analysis Menggunakan Association Rule untuk Pendukung Keputusan Promosi pada Sistem Penjualan Sun Young Cell,” vol. 12, no. 1, pp. 974–983, 2017.
- [6] Sri Rahayu Siregar, “Implementasi Data Mining Pada Penjualan Tiket Pesawat Menggunakan Algoritma Apriori (Studi Kasus: Jumbo Travel Medan),” *Pelita Inform. Budi Darma*, vol. VII, no. Nomor 1, Jul. 2014.
- [7] Fajar Astuti Hermawati, *Data Mining*. Yogyakarta: Andi Offset, 2013.
- [8] “www.r-project.org,” Feb. 19, 2021. .