IJDSEA
International Journal of Data Science Engineering, and Analytics

Implementation of Data Mining in Shopping Cart Analysis using the Apriori Algorithm

Susy Rahmawati [1], Miftahul Nuril Silviyah [2], and Nur Syifa'ul Husna [3]

[1] UPN "Veteran" Jawa Timur; susyrahmawati7771@gmail.com
[2] UPN "Veteran" Jawa Timur; miftahulnurilsilviyah@gmail.com
[3] UPN "Veteran" Jawa Timur, nursiva3112@gmail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Citation: Rahmawati, S.; Silviyah, M.N.; Husna, N.S. Title. *C* **2021**, Vol *1*, Page 30-36.
https://doi.org/10.3390/xxxx
Academic Editor: Susy Rahmawati
Received: 10 July, 2021
Accepted: 15 July, 2021
Published: 19 July, 2021

Abstract: Market basket analysis is one of the techniques of knowledge mining used in a broad dataset or database to find a collection of items that are interwoven. Generally used in a sale, the most relevant shopping cart data is used. This methodology has been widely applied in different multinational or foreign industries and is very useful in consumer buying preferences. Technology advances change business trends dramatically, shifting customer demands require increased surgical accuracy of business. In this research, the writer wants to analyze the shopping cart using a priori algorithm, with a dataset from the Kaggle web. Using anaconda software features with the Python programming language which is expected to create knowledge overwriting consumer buying patterns. In conclusion, this pattern can be used to support an industry in managing its company activities.

Keywords: Data Mining, Market Cart Analysis, Apriori Algorithm, Anaconda

1. Introduction

Internet-based Advances in technology support the growth of technology. In various fields, one that is facing the consequences of technological advances is that this business trend has an impact. Businesses in this information age must also compete in a rapidly changing, environmental, highly competitive and customer-focused, regional market [6]. These aspects can be changed over time, so developers and business actors must think of strategies that better ensure the continuity of their business. That is by observing sales transaction data. In the retail business, one method that can be tried is to identify the state of the market (customers) [6]. From the transaction data it can be processed to find information, this method is usually called market basket research. Data processing focuses on finding purchasing patterns by extracting associations and events from store data for a transaction, namely one a process system [10]. In this case, it aims to carry out identification or groups that are linked together in buying transactions. Because it can help promotions, can help business actors implement policies to improve sales strategies. Market baskets or market basket analysis this market is commonly referred to as analysis.

Referring to [1,6] If then' which serves to form a mixed pattern of items and rules such as knowledge and meaningful data from sales transaction information' The process of association patterns that use encouragement a priori to create laws of association with patterns.

The market basket analysis uses an a priori algorithm. The a priori algorithm is a business basket research law of association algorithm used to create [11]. The a priori algorithm lists the type of association terms on information mining, thereby creating a collection of frequent items from yahoo a priori that aim to run on information. (Market cart analysis, the commonly used approach is terms [11]. This study condemns the a priori algorithm because it is very efficient, by practicing

associations. The interrelation between products due to the effects between transactions to provide product suggestions.

In this study, the author wants to analyze the market basket using the a priori algorithm and also the dataset from the Kaggle web. This research is assisted by using the anaconda application which is expected to create knowledge about consumer buying patterns. And in conclusion, the existing pattern can be used as a support for an industry in managing its company activities.

2. Related Works

In the previous research, Apriori technique was used to mining frequent item sets and interesting association in transaction database. Apriori algorithm one of the most popular algorithm in implementation of data mining. After we was literature review, traditional Apriori algorithms have two major bottlenecks, for scanning the database frequently, and generating a large number of candidates sets [14]. In their research, the algorithm has improvements to mapping using new database and scanning the database repeatedly, to improve efficiency for pruning frequent and candidate item sets, and further to achieve high efficiency with count support using overlapping strategy.

In the other research, association rule mining to be an important part of data mining process. The rules was use for finding some hidden information in the database [15]. Many various algorithms already provided to solve problems in the data mining.

Hence, in this paper will be used Apriori Algorithm for Shopping Cart Analysis.

3. Experiment and Analysis

3.1. Methodology Experiment

In carrying out this research there are several stages that must be tried by the author. The sessions are as follows:

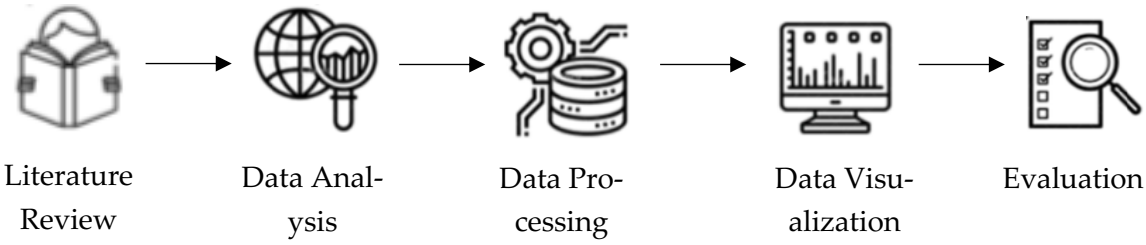

Figure 1. Flowchart Methodology for Research

1. Literature Review
   Before carrying out the research, the author first searches for and reads scientific literature studies that are related to research such as from the internet, as well as previously available journals. To understand the market basket analyzer using a priori algorithm.
2. Data Analysis
   In this study, the object used by the author is a collection of data from the Kaggle web. Where this information contains records of customer shopping carts in a mini market, which will later be used to analyze using information mining.
3. Data Processing
   The data obtained after going through the stages of data collection, then the data is processed with a computer program and converted into data in the form of numbers.

Information processing includes several sessions such as preprocessing, implementation of information mining as material for this research.

4. Data Visualization
   Data visualization is the stage where this session is in the form of program results that have been made with a display that is easier to understand.
5. Evaluation
   After the research process is complete, the researcher conducts an evaluation session. In this session, we try to check whether the patterns or data found match those compared by the researchers.

3.1.1. Data Mining

Useful information and linked knowledge from large databases, extraction and identification. That is a process that uses statistical methods, mathematics, artificial intelligence, and machine learning. information mining [7]. Information mining is a process that aims at one or more machine education processes to analyze [11]. Information mining, which is to make connections and patterns that can provide useful indications [6]. In this study we use information from the Kaggle website is a website that provides many datasets that are used for scientific information purposes [13]. One of them is the food ingredients market basket dataset, a dataset containing 9835 transactions by grocery shopping customers, but in this paper we only used 20 pieces of information.

3.1.2. Apriori Algorithm

Agrawal and Srikant in 1994 for frequent object set for Boolean association conditions, the a priori algorithm found the basic algorithm proposed by. The a priori algorithm finds the basic algorithm proposed by [7]. It is very popular for creating large frequency pattern algorithms [13]. This algorithm uses an iterative approach and there are 2 steps in each iteration. The first step is to create a set of candidates [9]. The main idea of the algorithm is a priori:

1. Searches for a regular item set from the transaction information base (the set of items that meet minimal support). [11]
2. With an already low frequency sourced at the specified minimum help level previously set Eliminates the item set. [11]
3. Set that meets the minimum value of trust in the information base by constructing the association conditions of the object. [11]

In this session look for a mix of items that meet the minimum requirements of support in the database. Items are obtained by applying the following formula Support value:

$$Support\ (A) = \frac{\sum Transaksi\ Mengandung\ A}{\sum Transaksi} \quad (1)$$

Meanwhile, the support value of 2 items using the formula is obtained by:

$$Support\ (A \cup B) = \frac{\sum Transaksi\ Mengandung\ A\ dan\ B}{\sum Transaksi} \quad (2)$$

3.1.3. Shopping Cart Analysis

Basket analysis is a method of information mining that is used to search for items found in large collections of information or databases [2]. The market basket procedure can help a business to gain insight into the buying behavior of their customers [4]. The market basket initiates the analysis of transactions containing one or more items, vice versa and some data [1]. To carry out the market basket, the following steps:

1. Specifies the desired minimum Help value. This is the minimum support threshold for the number of item sets that are allowed, if the number of items is at the bottom of the threshold until the item is eliminated, the minimum support should be eliminated.
2. Set a frequent itemset (a collection of items that stick out simultaneously), which has an item set frequency by taking a minimum itemset of Minimum previously supported.
3. From any Frequent Itemset, Limited Assistance, generate the association terms that meet the value.

3.1.4. Recommendation System

Which suggests useful information or what presumptions the customer will try to achieve his goals, such as sorting, the suggestion device is a system [5]. Hence customers can choose products that are more effective to determine the desired product. The suggestion system will guide users to create products that are relevant and useful from the many existing products. [14]. Product Promotion Advice uses customer comments on an object to help customers choose products [15].

3.1.5. Anaconda

Anaconda is one of many open source platforms that facilitate the use of open source programming languages (R, Python) for large-scale data processing, analytics, predictive and scientific computing [3]. The area study community can choose to adapt the use of one of the programming languages R or Python to analyze information science problems on the Anaconda platform [12].

3.2. Results and Discusses

In this section the author should carry out the results of research on market basket analysis using the a priori algorithm. The following dataset is used by the author:

Table 1. Dataset for Apriori Algorithm

Number	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
1	Citrus fruit	Semi-finished bread	Margarine	Ready soups			
2	Tropical fruit	Yogurt	Coffee				
3	Whole milk						
4	Pip fruit	Yogurt	Cream cheese	Meat spreads			
5	Other vegetables	Whole milk	Condensed milk	Long life bakery product			
6	Whole milk	Butter	Yogurt	Rice	Abrasive cleaner		
7	Rolls/buns						
8	Other vegetables	UHT-milk	Rolls/buns	Bottled Beer	Liquor (appertizer)		
9	Potted plants						
10	Whole milk	Cereals					
11	Tropical fruit	Other vegetables	White bread	Bottled water	Chocolate		
12	Citrus fruit	Tropical fruit	Whole milk	Butter	Curd	Yogurt	Flour

Number	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7
13	Beef						
14	Frankfurter	Rolls/buns	Soda				
15	Chicken	Tropical fruit					
16	Butter	Sugar	Fruit/vegeta-ble juice	Newspaper			
17	Fruit/vegetable juice						
18	Packaged fruit/vegetables						
19	Chocolate						
20	Specialty bar						

In Table 1. Dataset, where the dataset contains baskets of groceries by 20 customers.

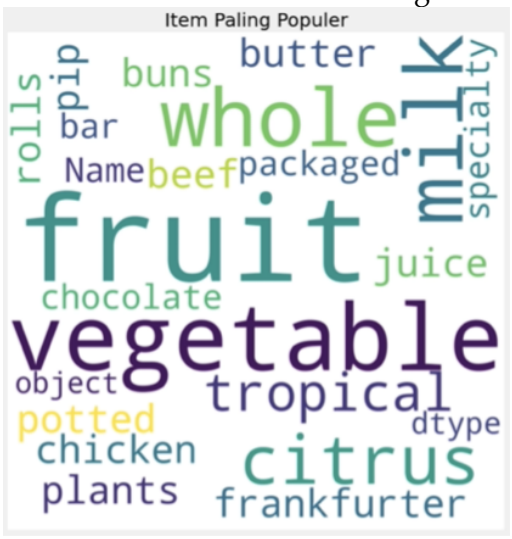

Figure 2. Word-cloud

In Figure 2 there is a word-cloud image, where the word-cloud itself functions as text data visually to make it easier to understand. The more often the word is used, the larger the size pattern on the word-cloud display. In the picture 2 words that have large sizes are vegetable, whole milk, and fruit.

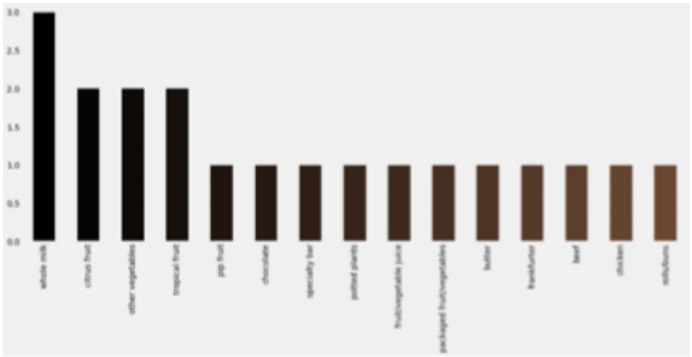

Figure 3. Box Diagram for Attribute

In Figure 3, the attribute that has the highest value is whole milk, followed by citrus fruits, other vegetables, and tropical fruits that have almost the same value.

	support	itemsets
0	1.00	(Food)
1	1.00	(Second Choice)
2	1.00	(Third Choice)
3	0.05	(UHT-milk)
4	0.05	(abrasive cleaner)
...	...	...
3618	0.05	(citrus fruit, curd, Food, yogurt, flour, whol...
3619	0.05	(butter, citrus fruit, curd, Food, yogurt, flo...
3620	0.05	(butter, citrus fruit, curd, Food, yogurt, flo...
3621	0.05	(butter, citrus fruit, curd, flour, yogurt, wh...
3622	0.05	(butter, citrus fruit, curd, Food, yogurt, flo...

3623 rows × 2 columns

Figure 4. Minimum Summarize for Support 5%

In the creating process C1 or one item set with minimum summarize for support 5 % present at figure 4.

	support	itemsets	length
41	1.00	(Food, Second Choice)	2
42	1.00	(Third Choice, Food)	2
43	0.05	(UHT-milk, Food)	2
44	0.05	(abrasive cleaner, Food)	2
45	0.05	(Food, beef)	2
...	...	...	...
267	0.05	(rolls/buns, soda)	2
268	0.05	(white bread, tropical fruit)	2
269	0.05	(tropical fruit, whole milk)	2
270	0.10	(tropical fruit, yogurt)	2
271	0.10	(yogurt, whole milk)	2

231 rows × 3 columns

Figure 5. Combination for two item sets

This stage provides a description of overriding the results of implementing the a priori algorithm using the Python Anaconda programming language. The a priori algorithm that is tried is shown in Figures 4 & 5 showing the results of the implementation of each process of the a priori algorithm, some of the results showing each process.

4. Conclusions

Notification is one method that not only works as a marker that there is information waiting to be read. But along with the times, notifications are increasingly filled with information that is considered less important for device users. So there needs to be a breakthrough to overcome this. This study aims to design a system that can help users to sort out notifications that are considered important and not.

References

1. Ghozali, M. I., & Pati, S. M. A. N. J. (2017). ANALISA POLA BELANJA MENGGUNAKAN ALGORITMA FP GROWTH , SELF ORGANIZING MAP (SOM) DAN K MEDOIDS. 8(1), 317–326.
2. Isa, N., Ramlan, M. A., & Puteh, M. (2018). Market Basket Analysis of Customer Buying Patterns at Corm Café Market Basket Analysis of Customer Buying Patterns at Corm Café. October 2019. https://doi.org/10.14419/ijet.v7i4.42.25692
3. Kadiyala, A., & Kumar, A. (2017). Applications of Python to Evaluate Environmental Data Science Problems. 00(00), 1–7. https://doi.org/10.1002/ep.12786
4. Kaur, M., & Kang, S. (2016). Market Basket Analysis : Identify the changing trends of market data using association rule mining. Procedia - Procedia Computer Science, 85(Cms), 78– 85. https://doi.org/10.1016/j.procs.2016.05.180
5. Kurniawan, A. (2016). SISTEM REKOMENDASI PRODUK SEPATU DENGAN MENGGUNAKAN. 2016(Sentika), 18–19.
6. Listriani, D., Setyaningrum, A. H., & A, F. E. M. (2016). PENERAPAN METODE ASOSIASI MENGGUNAKAN ALGORITMA APRIORI PADA APLIKASI ANALISA POLA BELANJA KONSUMEN (Studi Kasus Toko Buku Gramedia Bintaro). 9(2), 120–127.
7. Marisa, F., & Purnomo, D. (n.d.). Penerapan Algoritma Apriori Terhadap Data Penjualan di Toko Gudang BM. 35, 1–5.
8. Ponsel, P., Metode, M., & Apriori, A. (2020). Data mining. 5(2), 81–95.
9. Putra, T. D. (2020). Analisis Keranjang Belanja dengan Algoritma Apriori Klasik pada Data Mining. Jurnal Kajian Ilmiah, 20(1), 59–66. https://doi.org/10.31599/jki.v20i1.70
10. Santosa, S. (2016). Analisis Keranjang Pasar Untuk Rekomendasi Produk (Consumer Good) Menggunakan Fp-Growth. Teknik Informatika, 12, 103–115.
11. Sianturi, F. A., Informatika, T., & Utara, S. (2018). PENERAPAN ALGORITMA APRIORI UNTUK PENENTUAN TINGKAT. 2(1), 50–57. Sistem, R., Buah, D., & Jenis, B. (2021). Jurnal resti. 1(10), 476–481.
12. Sodik, F., Dwi, B., & Kharisudin, I. (2020). Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python. 3, 689–694.
13. Wulandari, R., & Mursidah, I. (2019). Pola Pembelian Produk Berdasarkan Association Rule Data Mining. 3, 1–6.
14. Yuan, X. "An improved Apriori algorithm for mining association rules". AIP Conference Proceedings 1820, 080005 (2017). https://doi.org/10.1063/1.4977361
15. Sharma, N., and Verma., D. "An Improved Apriori-Based Algorithm For Association Rule Mining". 2014.