



ADVANCED INTERNATIONAL JOURNAL OF BUSINESS, ENTREPRENEURSHIP AND SMES (AIJBES) www.aijbes.com



MARKET BASKET ANALYSIS USING APRIORI ALGORITHM: GROCERY ITEMS RECOMMENDATION

Nor Aidatul Hazirah Mohd Rosli¹, Noor Hasimah Ibrahim Teo^{2*}

- ¹ Web Integration Department, Blue Ocean Data Solutions, Malaysia Email: aidatul@blueoceands.com
- ² College of Computing, Informatics and Media, Universiti Teknologi MARA (Melaka Branch), Malaysia Email: shimateo@uitm.edu.my
- * Corresponding Author

Article Info:

Article history: Received date: 30.10.2022

Revised date: 07.11.2022 Accepted date: 15.11.2022 Published date: 31.12.2022

To cite this document:

Mohd Rosli, N. A. H., & Teo, N. H. I. (2022). Market Basket Analysis using Apriori Algorithm: Grocery Items Recommendation. Advanced International Journal of Business, Entrepreneurship and SMEs, 4 (14), 01-09.

DOI: 10.35631/AIJBES.414001.

This work is licensed under <u>CC BY 4.0</u>

Abstract:

A grocery list is frequently used to make sure that only essential items are bought and nothing else is forgotten. Due to the extensive list of items, they must buy each month, people frequently forget what they need to buy when they get to the supermarket. Even with a list, there are still things that get left out. Therefore, an application called MyGrocery is proposed that offers suggestions for items users frequently buy together and aids users in creating grocery lists. This might help them make a personalised grocery list with items that are suggested based on their past purchases. The aim is to help them keep track of what they have bought as well as to help them create personalised grocery lists. Market basket analysis is used to find associations between items in MyGrocery using past purchased history. The result shows the significant recommendation by MyGrocery where the Market Basket Analysis, which is carried out using the Apriori Algorithm, displayed the recommended item following the calculation of lift value.

Keywords:

Market Basket Analysis, Personalization, Association rules, Recommendation, Apriori Algorithm

Introduction

People always make a grocery list before heading to the supermarket to ensure that they don't forget anything important. The reason for this is that when individuals go grocery shopping, they normally buy a big number of goods. As a result, some people may find it challenging to recall things without writing a list, which can increase the likelihood that they will forget anything on the list.

Copyright © GLOBAL ACADEMIC EXCELLENCE (M) SDN BHD - All rights reserved



Forgetful are a common occurrence in people's lives (Fernandes et al., 2016), and this includes forgetting items during grocery shopping. This circumstance is more unpleasant when people forget a moment before entering the grocery store. This issue may waste people's time and effort due to forgetting concerns. Even when a shopping list is created, certain products are often forgotten. This is especially true for items that are purchased infrequently but have a long lifespan. Creating a shopping list in a rush can also add to this problem (Nath, 2021).

The problem of forgetting and overlooking can be reduced by creating an application that can recommend items based on a user's past purchases. This means that the next items they intend to purchase will be suggested based on their previous purchases and the frequency with which they make those purchases. The algorithm will suggest and recommend the next items to be purchased based on the purchasing trends that have been logged. In addition, relevant products with the same purchase date will be suggested to the consumer, allowing them to swiftly add them to their shopping cart. Additionally, they can compare the prices to the current prices of the same items in the store.

Related Work

Many grocery list applications have been developed to make it easier for users to make their lists before grocery shopping. It has the ability to create, store and manage the list for future reference (Firoz & Ratnayaka, 2020). However, each of the applications has different purposes and functions. The method used to design the application may also vary from one another. The development of the shopping list application, which focuses on creating a grocery list for the user, is the only aspect of the comparable applications.

The Smart Shopping List application (Jayawilal & Premeratne, 2017) is one of the many current or comparable programmes that offer the ability to create grocery lists. The application's feature allows users to send their shopping lists to others via SMS (Short Message Service) (SMS). To generate item recommendations, this application can also analyse historical user data and find association patterns. Additionally, the app can show local supermarkets with typical iconography. The text-to-app function known as "Bring Me!" is the method that the programme uses. Apriori algorithm and geolocation services are also used to power this Android mobile app solution. The user-created list content, the user's prior sales data, and overall sales data are where the data comes from. Other than that, it made use of the Mapper "Find a store" component, which is in charge of handling the geolocation and navigational needs.

The Smart List application proposed by Katuk et al., 2019 can set reminders whenever items are running out of stock. Additionally, it will be organising and producing grocery lists using modern smartphones. This application was created using the Unified Modelling Language software (UML). The system requirements are modelled and shown using UML. Additionally, it is done using the Rapid Application Development (RAD) technique. RAD is an adaptive software development methodology that uses prototyping to obtain the requirements for the application's systems. It is a mobile application that extracted data using information technology. This application incorporates augmented reality (AR) into the "AR-Assisted Mobile Grocery Shopping" system.



The Digital Grocery List (DGL) application (Heinrichs et al., 2011) provides an auto-complete list of matching items when the users start to enter an item into the list. The items in the lists will be displayed by their name and an icon. The items can optionally and instantly be tagged as purchased, which is likewise indicated by a PPUI. In DGL applications, a client/server architecture is the method employed. A graphical user interface will be provided by client apps (GUI). While the PPUI component is based on Letras, the mobile client application was created on the Android platform utilising the Application Programming Interface (API) version 2.1.

The List It application's function suggested by Adaji (2018) allows users to pick whatever products they want from a list of options the system provides. The list then includes information on each product's calorie count. Following that, this application can let users become friends. This application uses the ionic method as its integrated technology. Ionic is an open-source framework for creating online applications or hybrid mobile applications. Given that it is a hybrid approach, any platform can be used.

A feature that enables the shopping list to be rapidly updated by numerous users is offered by Sunil et al. in their 2018 proposal for an improved smart shopping mobile application. This feature can stop the user from purchasing any identical goods on the list. People who use this service can view the most recent grocery list. Once the grocery list has been modified by any user, a notification will ring. The data was extracted by the Improved Smart Shopping programme using an information system. Then, for mobile applications, it also applied the Agile approach. This programme was created specifically with Android users in mind.

Apriori Algorithm

Market Basket Analysis is a strategy that employs the Apriori algorithm to produce suggestions. The theory behind the technique is that if a user buys one thing, he or she will predictably buy other items too (Ting et al., 2010). Support, lift, and confidence are the three main components of the Apriori algorithm (Deora et al., 2013). Support essentially indicates how popular an item is. The number of orders or transactions with product B out of the total number of orders is used to determine support. If item A is purchased, confidence refers to the possibility that item B will be purchased as well. It is calculated by dividing the total number of transactions in which A and B are purchased jointly by the total number of transactions in which A is purchased. For lift, when A is sold, the lift refers to the increase in the sale ratio of B. A lift value of 1 indicates that the products are unrelated. If the lift value is larger than one, the products are more likely to be purchased together (Singh et al., 2021). When the lift value is less than one, it indicates that the products are unlikely to be purchased together.

Market Basket Analysis has the advantages of using a large item set, being easily parallelized, and being simple to implement (Kumar & Rukmani, 2010). Although the technique has many benefits, it also has drawbacks, such as the necessity for numerous data repetitions. The same minimal support threshold is used. Difficulties in obtaining rare occurrences, the technique of mining association rules takes a long time with a huge database, therefore competency becomes a critical aspect (Agarwal, 2013). However, this algorithm is still a popular choice for market basket analysis and have been used for different purposes such as predicting sales pattern (Kurnia, Isharianto, Giap, & Hermawan, 2019) and consumer purchasing pattern (Aldino, Pratiwi, Sintaro, & Putra, 2021).



Methodology

System Design

Flow Diagram for MyGrocery Recommendation

Figure 1 shows the flow diagram on the MuGrocery using APriori Algorithm for suggesting grocery items. The Apriori algorithm will be used once the user has chosen or added any products to the shopping list, and it will identify the item that is typically chosen alongside the selected item based on past purchases. The list will not have any recommendations if no items have been added.



Figure 1: Flowchart of MyGrocery

Figure 2: Apriori Algorithm

Apriori Algorithm

Figure 2 depicts the Apriori algorithm steps. When a user adds an item to their shopping list, the system reads the information for each item and calculates its support before recommending it to the user. The estimated support will be entered as a frequent item set if it is higher than the minimal support or equal to it. The confidence will then be determined, and if it is larger than the minimal confidence or equal to it, it will be placed into strong rules.

Interface Design

The application's interface was created using the Android Studio software and comprises six modules: Login, Register, Add Items, List, Shopping List, and Recommendation List.

- Login Module: Users need to enter their email and password to login into the system.
- Registration Module: Users need to enter their full name, email, password and phone number to register into the system. Once registration is complete, the information will be added to the database, and the user can then log in.
- Add Item Module: To add new data to the database, the user must input the date, the item's name, and the phone number.

Copyright © GLOBAL ACADEMIC EXCELLENCE (M) SDN BHD - All rights reserved



- List Module: The user can search for any item using the item's name in the search field, and related items with the same keyword will display. This will list all the data in the database that has been entered by the user. After that, if the user wants to add the item to their shopping list, they may do so by simply selecting it.
- Shopping List Module: The user will refer to their grocery list from the Shopping List interface when doing their grocery shopping. Any chosen items will be retrieved and placed on this shopping list. The objects that have been listed in this shopping list are viewable, editable, and deletable by the user.
- Recommendation Module: The recommendation modules recommended things displayed based on the most popular user's purchase of the item with the item that has been chosen for the shopping list. The recommended items can be added to the user's shopping list. Apriori Algorithm is implemented in this module.

Implementation of Market Basket Analysis using Apriori Algorithm

Market Basket Analysis is written in Python and implemented in Jupyter Notebook (anaconda 3). The Apriori algorithm and the association rule were used to determine the frequent item set to implement the Market Basket Analysis. The first step is to import the related function into the system to ensure that the calculation works properly.

Dataset

The CSV file containing the 6-month grocery list dataset is utilised for market basket analysis. The items are organised by month. As seen in Figure 3, each column reflects products purchased for a month.

□ 5 • ♂• •		aprioridata - Excel (Produ	ct Activation Failed)	🛕 Aidatul Hazirah 🎑	• - • ×
File Home Insert Page Layou	ut Formulas Data Review	View Help Q Tell me what			오. Share
Paste Clipboard 5 Font	A* A* = =	arap Text General arge & Center * Single & Mumber	Conditional Format as Cell Formatting * Table * Styles *	Insert Delete Format Cells E	rt & Find & ter * Select * diting Sensitivity
AR3 • : × ~ fx	NaN				
A	в	C	D	E	F G H
2 ADABI 250G KARI AYAM & DAGING	ADABI 250G PERENCAH SUP	ADABI KETUPAT MINI 20G *30	ASAM KEPING	AYAM BRAND COCONUT MILK 1L	BABAS SEF BUI KETUN BUMI HUA
3 SAILCOOKING OIL 5KG	OO EGGS GRADE A 30'S	TEMBIKAI MERAH	PRALGULA PASIR KASAR 1KG	CILI MERAH	DAUN PUE KACANG T AYAM SEC
4 TELUR MASIN (NP) 4'S	NaN	NaN	NaN	NaN	NaN NaN NaN
5 ADAB BAWANG GORENG KECIL 500G	ADAB BELACAN 130G	BAWANG BESAR CHINA 1.5KG	BAWANG PUTIH GUNI 500G	BS WH PLASTIC TALL 8*9 500G	BUMI HUA BUMI HUA CAP CIKGI
6 ADABI KETUPAT MINI LAZAT 20G*30	AJINOMOTO 1KG	APPLE HIJAU (S) 10PCS	BAWANG BESAR MERAH HOLLAND 8KG	BS WH DISHWASH 10L STRAWBERRY	BUMI HUA CARNATIC CARROT C
7 AYAM SEGAR	DAUN KUNYIT	DAUN LIMAU	DAUN SUP	IKAN MASIN GELAMA BELAH	JAMBU BA KACANG P ORANGE 1
8 QL TELUR GRED B 30PCS	DAUN PUDINA	HALIA TUA	ADABI 500G SERBUK KUNYIT	AYAM BRAND COCONUT MILK 1L	BAWANG I BS WH DIS BS WH PL
9 KAMBING-LAMB CHOP	AYAM SEGAR	PAPADAM CHIPS	BUMI HUAU SOS LADA HITAM 2KG	BUMI HIJAU SOS TIRAM 2KG	CILI HUAU CILI MERA DAUN SUF
10 BRAND COCONUT MILK 1L	BAWANG MERAH BESAR PAKISTAN 5KG	BAWANG PUTIH GUNI 500G	BESTARI FRIED CHICKEN COATING 1KG OR	BUMI HIJAU SOS CILI 2KG	BUMI HUA CARNATIC CARROT C
11 SUPERIOR FACIAL TISSUE 4*70'S	TEMBIKAI MERAH	AYAM SEGAR	GULA MELAKA A	DAUN SUP	NaN NaN NaN
12 AYAM BRAND COCONUT MILK 1L	BAWANG PUTIH KG	BOH TEA LEAVES 250G	BROCOLI KG	CARROT CINA KG	CILI MERA CILI PADI DAUN SUF
13 BS WH PLASTIC BOX 500	BS WH PLASTIC BOX 650	BS WH POLY FOAM CUP B"	BS WH POLY FOAM LUNCH BOX SHALLOW	BS WH POLY FOAM PLATE 9"	ASAM JIU JASMINE F CARNATIC
14 SAWI	SERAI (BESAR)	томато	BABAS SERBUK 70G BRIYANI SPICE	QL TELUR GRED B 30PCS	NaN NaN NaN
15 AYAM SEGAR	NENAS MADU	NENAS 1 BUI	JASMINE PUSA CREAM BASMATHI SELLA 5	A1 GARAM HALUS 800G	ADABI 50C BOH TEH F BROCOLI I
16 SAJI COOKING OIL 5KG	SAJI SUSU CAIR 390G	AJINOMOTO 1KG	SALAD CAMERON	LOBAK MERAK	TIMUN A MAKANaN NaN
17 SUNQUICK BES MINUMAN BLACKCURRAN	PRAI GULA PASIR KASAR 1KG	NaN	NaN	NaN	NaN NaN NaN
18 BAWANG BESAR INDIA GUNI 9KG	BAWANG PUTIH KG	BROCOLI KG	BUMI HIJAU SOS LADA HITAM 2KG	BUMI HIJAU SOS TIRAM 2KG	CAP SAUH CENDAWA DAUN KES
19 APPLE HIJAU (S) 10PCS	BESTARI FRIED CHICKEN COATING 1KG OF	I BILAL BRAND 656G LYCHEES IN SYRU	CARNATION BERAS WANGI IMPORT 5KG A	CARROT CINA KG	CENDAWA CILI BOLA CILI MERA
20 BAWANG GUNI	CARNATION BERAS SUPER IMPORT APPLE	HOSEN FIESTA FRUITCOCKTAIL 8360	DAUN KARI (AMSOON)	SUNQUICK BES MINUMAN OREN 840M	SAJI SUSU SAJI SUSU AYAM BR/
21 KIJANG TEPUNG RANGUP 400G	BLACK GOLD 11 SLICE BONELESS 900G	BUMI HIJAU SOS LADA HITAM 2KG	CAP HELANG GARAM PUTIH KASAR 500G*	IKAN MASIN GELAMA BATANG	KAILAN KC KANGKUN TAUGE KG
22 ADABI 250G PERENCAH SUP	ADAD BAWANG GORENG 160G	ADAD BELACAN 130G	APPLE HIJAU (S) 10PCS	AYAM BRAND COCONUT MILK 1L	BAWANG I BROCOLI I BUNGA RA
23 PRAI GULA PASIR KASAR 1KG	AYAM SEGAR	LADA KERING GUNI 1KG	MUSHROOM FISH BALL GORENG 160G	MUSHROOM FISH CAKE (SQUARE)	NaN NaN NaN
24 AYAM BRAND COCONUT MILK 1L	BAWANG PUTIH KG	CARROT CINA KG	CILI BOLA HIJAU KG	DAUN SUP	HALIA TU/ KACANG P KENTANG
25 APPLE HIJAU (S) 10PCS	BAWANG BESAR INDIA GUNI 1.5KG	BAWANG PUTIH KG	BUMI HIJAU SOS LADA HITAM 2KG	CAP BUNGA MERAH ASAM JAWA 1KG	CARNATIC DAUN LIM DAUN SUF
26 SAJI SUSU CAIR 390G	UBI KENTANG CINA GUNI	AYAM SEGAR	ADAB BAWANG GORENG KECIL 500G	AYAM BRAND COCONUT MILK 1L	BUMI HIJA DAUN SUP HOSEN FIL
aprioridata 🕀			Ξ		Þ
Peadu				COL (C) (T)	

Figure 3: Grocery Dataset for 6 Months

EDA

To carry out the Market Basket Analysis, the frequency of the most popular items must be acquired to determine the most frequently listed items by the user.

Getting the List of Transactions

After reading the dataset, the system must obtain the list of items in each transaction. As a result, there will be two loops, one for the total number of transactions and the other for the total number of columns in each transaction. As indicated in Figure 4, this list will be utilised



as a training set to generate a list of association rules. The next stage is to extract the unique items, after which the items with a 'nan' value will be removed.



Figure 4: List of Transaction Code

One-Hot Encoding Transaction Data

The next step is to use a lambda function to break each transaction string into a list, transforming the column into a list of lists. The transactions are then turned into a one-hot encoded DataFrame, with 'TRUE' and 'FALSE' values in each column indicating whether or not an item was included in a transaction.

Support Metric

Following that, it will go through the metrics and pruning processes, as shown in Figure 5, where metrics are used to monitor the performance of the rules and pruning is used to eliminate the rules. The support statistic calculates the percentage of transactions that contain an itemset by dividing the number of transactions with items by the total number of transactions.

In [16]:	<pre># Compute the support support = onehot.mean() support = pd.DataFrame(support, columns=['support']).sort_values('support', ascending=False # Print the support support.head()</pre>					
Out[16]:		support				
		Support				
	Item	1.000000				
	AYAM SEGAR	0.581818				
	SAWI PUTIH KG	0.381818				
	QL TELUR GRED B 30PCS	0.381818				
	TIMUN A KG	0.345455				
1						

Figure 5: Support Metric Implementation

Confidence and Lift Metric

It then moves on to the confidence and lifts metric after computing the support. The confidence measure will compute the likelihood that the user will purchase the goods along with the second product, whilst the lift metric will evaluate the relationship between the items. To determine



the likelihood of a user purchasing the second product, compute the support, confidence, and lift of the two most frequent items, as illustrated in Figure 6.

Figure 6: Confidence Metric Code

Results

The association analysis is conducted on the grocery dataset. It is done by calculating support, confidence, and lift of the two most frequent items as discussed in the previous section. The 6 months' past grocery bills are collected and extracted into a CSV file. Table 1 shows the association result produce from this analysis.

No	Items 1	Items 2	Lift Value
1.	Bawang Goreng Kecil 500G	Ayam Segar	1.145833
2.	Bawang Goreng Kecil 500G	Cili Merah KG	2.820513
3.	Bawang Goreng Kecil 500G	Mee Hoon 3KG	2.619048
4.	Bawang Goreng Kecil 500G	Telur Gred B 30PCS	1.746032
5.	Ayam Brand Coconut Milk 1L	Telur Gred B 30PCS	2.070000
6.	Suria Fish Stick 250G	Cili Merah KG	2.644231
7.	Suria Fish Stick 250G	Mega Mix Vegetables 1KG	6.875000
8.	Suria Fish Stick 250G	Tepung Gandum Cap Sauh 1KG	3.437500
9.	Suria Fish Stick 250G	Kapal Api Kopi Campuran 500G	1.250000
10.	Suria Fish Stick 250G	Mega Mix Vegetables 1KG	6.875000
11.	Asam Jiu Zhi Chen Pi 100G	Apple Royal Gala (M) 8PCS	55.000000
12.	Ubi Kentang Cina	Meatball	11.000000
13.	Mushroom Fish Ball 160G	Tepung Gandum Cap Sauh 1KG	18.333333
14.	Kismis	Ayam Hp	11.000000
15.	Brocolli	Daun Kunyit	55.000000

Table 1: Lift Value for Calculation Product Association

The products are unrelated when the lift value is 1. If the lift value is larger than one, the products are more likely to be purchased together (Singh et al., 2021). When the lift value is less than one, the products are unlikely to be purchased together. The validation test results demonstrate that the lift value for the tested item set is greater than one, which is the expected result.



Conclusion

The primary issue with this work is that most people forget what they want to get when they go grocery shopping. As a result, introducing a 'Grocery List Application' will assist consumers in managing their grocery lists, ensuring that forgetting difficulties are reduced. The results of utilising the Apriori algorithm for Market Basket Analysis reveal that it is successfully recommending the item sets based on the past grocery purchased dataset. To recommend the items, the data's support, lift, and confidence have been calculated to indicate the relationship between the items that are suitable for showing as recommended items. This work has contributed towards the market basket analysis approach for grocery items recommendation based on past grocery purchases using Apriori Algorithm. Although MyGrocery fulfils its primary objective, further features can be added in the future to make it more user-friendly and to improve its performance.

Acknowledgement

The registration fee is funded by *Pembiayaan Yuran Prosiding Berindeks (PYPB), Tabung Dana Kecemerlangan Pendidikan (DKP),* Universiti Teknologi MARA (UiTM), Malaysia.

References

- Adaji, I., Oyibo, K., & Vassileva, J. (2018, July). List it: a shopping list app that influences healthy shopping habits. In Proceedings of the 32nd International BCS Human Computer Interaction Conference 32 (pp. 1-4).
- Agarwal, P. (2013). Study on Apriori Algorithm and its Application in Grocery Store. 74(14), 1–8.
- Aldino, A. A., Pratiwi, E. D., Sintaro, S., & Putra, A. D. (2021, October). Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm. In 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE) (pp. 29-34). IEEE.
- Deora, C., Arora, S., & Makani, Z. (2013). Comparison of Interestingness Measures: Support-Confidence Framework versus Lift-Irule Framework. International Journal of Engineering Research and Applications (IJERA), 3(2), 208–215.
- Fernandes, D., Puntoni, S., van Osselaer, S. M. J., & Cowley, E. (2016). When and why we forget to buy. Journal of Consumer Psychology, 26(3), 363–380.
- Firoz, A., & Ratnayaka, G. (2020, March). Shoplister–a grocery list management application. In 2020 International Conference on Image Processing and Robotics (ICIP) (pp. 1-6).
- Heinrichs, F., Schreiber, D., & Schöning, J. (2011, August). The hybrid shopping list: Bridging the gap between physical and digital shopping lists. In Proceedings of the 13th international conference on Human computer interaction with mobile devices and services (pp. 251-254).
- Jayawilal, W. H., & Premeratne, S. (2017, November). The smart shopping list: An effective mobile solution for grocery list-creation process. In 2017 IEEE 13th Malaysia International Conference on Communications (MICC) (pp. 124-129). IEEE.
- Katuk, N., Jayasangar, T., & Yusof, Y. (2019). Design and development of smart list: a mobile app for creating and managing grocery lists. Baghdad Science Journal, 16(2), 462-476.
- Kumar, B., & Rukmani, K. (2010). Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms. Int. J. of Advanced Networking and Applications, 404(06), 400–404.
- Kurnia, Y., Isharianto, Y., Giap, Y. C., & Hermawan, A. (2019, March). Study of application of data mining market basket analysis for knowing sales pattern (association of items)



at the O! Fish restaurant using apriori algorithm. In *Journal of Physics: Conference Series* (Vol. 1175, No. 1, p. 012047). IOP Publishing.

- Nath, G., & Sen, J. (2021, June). An Algorithm for Recommending Groceries Based on an Item Ranking Method. In 2021 International Conference on Intelligent Technologies (CONIT) (pp. 1-7). IEEE.
- Singh, P. K., Othman, E., Ahmed, R., Mahmood, A., Dhahri, H., & Choudhury, P. (2021). Optimized recommendations by user profiling using apriori algorithm. Applied Soft Computing, 106, 107272.
- Sunil, S., Fahad, K., Mohid, K. (2018, April). An improved smart shopping mobile application. International Journal of Advance Research in Computer Science; Udaipur 9(2), 168-173.
- Ting, P. H., Pan, S., & Chou, S. S. (2010). Finding ideal menu items assortments: An empirical application of market basket analysis. Cornell Hospitality Quarterly, 51(4), 492–501. https://doi.org/10.1177/1938965510378254