Jurnal Mantik
Volume 4, Number 2, Aug 2020, pp. 1272-1277 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Menu Package Recommendation using Combination of K-Means and FP-Growth Algorithms at Bakery Stores

N P Dharshinni[1]*, Elvana Bangun[2], Sarah Karunia[3], Ruth Damayanti[4], Gabriel Rophe[5], Roy Pandapotan[6]

[1,2,3,4,5,6] Faculty of Technology and Computer Science, Universitas Prima Indonesia, Indonesia

*E-mail: priyadharshinninaidu@gmail.com

ARTICLE INFO

Article history:
Received: 12/01/2020
Revised: 22/07/2020
Accepted: 01/08/2020

Keywords:
Menu Package Recommendation, K-Means, FP-Growth

ABSTRACT

Bakery shop is a shop that sells variants menu like bread, cakes, and drinks. The main problem with this store's sales is still not knowing which product items are best sellers and the shop still markets a lot of non-selling menus, causing the shop to lose money. So it takes the right strategy to increase the sales of bakery shop menus by making a menu package recommendations from the menus most frequently purchased by customers. The k-means algorithm performs grouping on menus to get menu packages. Furthermore, the fp-growth algorithm looks for linkages between frequently purchased menus to get menu package recommendations. The results of the research that the dominant items often purchased in cluster$_0$ packages are hotdogs, pancakes, milk, garlic breadsticks with a confidence value of 92%, cluster$_1$ packages are garlic breadsticks, hotdogs, chicken sand, pancakes with a confidence value of 92% and the last cluster$_2$ packages are garlic breadstick, pastry, milk with a confidence value of 79%.

1. Introduction

Bakery Shop is a bakery that provides several variants of cake, bread, and beverage menus which are certainly in demand by the public. However, the management at the bakery shop is unable to regulate the availability of bakery products based on the ones most often ordered by customers and still markets products that are not in demand, so there are often products that are not sold and cause the shop to lose money. (Andi Pribadi, Max Nur Alam, 2017). The bakery shop has sales transaction data, but it is only stored as an archive and is not used properly, while from this data the shop can obtain information which is a solution to problems in the bakery shop. Therefore, the right strategy is needed to increase sales of the menu at the Bakery Shop to get the maximum profit (Desak Made Dwi Utama Putra, 2017). Sales transaction data can be processed to find customer purchasing patterns by utilizing data mining (Setyawan, 2018). Data mining is the process of analyzing data, where through this process, it can find hidden information in large amounts of data in running a business (Larose, 2005).

The K-Means algorithm is one of the algorithms in the clustering technique, which is the division of data into groups so that data that has similarities are grouped into one class and those that are different are grouped in another class (Mardalius, 2017). K-Means clustering can group products into the most and least sales groups to find out which items are in demand and which are not (Metisen & Sari, 2015). FP-Growth algorithm is one of the algorithms in the association technique which is the development of the Apriori algorithm. FP-Growth is used to select the data set that appears most frequently in a dataset (Samuel, 2008). This algorithm uses a tree concept called the FP-Tree to find the frequency of item sets so that execution in the FP-Growth algorithm is faster than Apriori algorithm (Erwin, 2009). The application of the FP-Growth algorithm can be used to get recommendations for sellers (retailers) in providing sales packages for consumers (Abdullah, 2018).

Product recommendations with association rules using the Apriori or Fp-Growth algorithm are less accurate for large datasets. So it is necessary to group with K-Means so that the data is divided into smaller data groups (Nurlelasari, 2017). Therefore, to increase profits at the bakery shop, a combination of the K-Means Clustering algorithm and the FP-Growth algorithm is carried out. K-Means clustering is used to group bakery product sales data and is followed by the Fp-Growth algorithm to find patterns of association between each cluster. Based on the best-selling products (best sellers), you can find recommendations for making

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

package menus from the best-selling products (Abdullah, 2018). This menu package recommendation is expected to be an input for the store in implementing the recommended menu package sales strategy and the best-selling bakery products continue to be reproduced and products that are no longer in demand are no longer marketed.

2. Related Research

Kurniawan, Gata, & Wiyana (2018), apply the FP-Growth algorithm in extracting itemset patterns with large data to get recommendations for facial skin care products and help determine the position of products on sales shelves.

Lestari (2015), applies the use of the FP-Tree to determine the frequency of the itemset and FP-Growth to find the relationship between drugs that consumers often buy.

Syukra, Hidayat, Fauzi (2019), apply the K-Medoids algorithm to cluster sales data into smaller data segments and implement FP-Growth in each cluster to provide more accurate product recommendations to customers.

Kusrini (2015), performs grouping to determine the minimum stock with the K-Means algorithm on supermarket retail data to minimize losses.

Adriyendi (2016), compared the K-Means and Fuzzy C-Means algorithms on food production data and found K-Means computation time was faster than Fuzzy C-Means.

3. Research Method

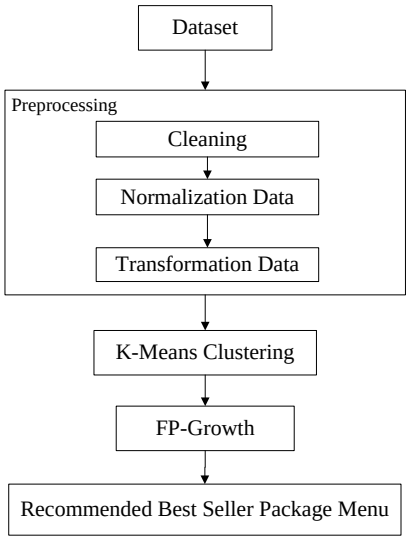

Fig 1. Research Framework

Figure 1 represents the research framework in this study. The online dataset is taken from retail sales recap at bakery stores. Then preprocessing is carried out first to remove null and double data (cleaning data), improve the table structure (normalization), and perform data transformation so that it can be processed in weka 3.8. Furthermore, the k-means clustering algorithm is implemented to get groups of product packages from each cluster and continues to implement the fp-growth algorithm to obtain recommendations based on the linkages between product items purchased simultaneously.

3.1 K-Means Clustering Algorithm

K-Means is one of the clustering algorithms where the algorithm is useful for grouping datasets into several classes/groups. The concept of K-Means is data that has similarities grouped in one class and differently grouped in another class (Mardalius, 2017). To calculate the distance from the data to the center of the cluster, the Euclidean formula is used:

$$d_{(Xj,Cj)} = \sqrt{\sum_{j=1}^{n}(Xj - Cj)^2} \quad (1)$$

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Description:
d= distance,
j= number centoid or data
C=centroid,
X=data

3.2 Fp-Growth Algorithm

The Fp-Growth algorithm is developed from the Apriori algorithm. The a priori algorithm requires a long computation time in the process because it uses the concept of candidate generation in determining frequent itemset, while the Fp-Growth algorithm uses the concept of tree formation in searching for itemset frequencies which makes this algorithm work faster (Syukra, Hidayat & Fauzi, 2019).

$$Support(X,Y) = \frac{\sum transactions\ X\ dan\ Y}{\sum all\ transactions} \text{ x } 100\ \% \quad (2)$$

$$Confidence(X \rightarrow Y) = \frac{\sum transactions\ containing\ X\ and\ Y}{\sum transactions\ containing\ X} \quad (3)$$

$$Lift = \frac{Confidence\ (X,Y)}{Confidence\ Benchmark\ (X,Y)} \quad (4)$$

3.3 Dataset

In this study using a sales dataset at a bakery shop from 2018 to 2019 with a total of 6017 data records. The data were obtained from an online dataset sourced from Kaggle.

4. Discussion

This study uses bakery shop sales transaction data with 1578 of 6017 data records for 2 years. The bakery shop sales data has attributes for Transaction Code, Transaction Time, Item Code, Item Name, and Price. This data will be processed to determine menu package recommendations to improve sales strategies. In table 1 is a dataset of bakery shop sales.

Table 1.
Bakery Shop Sales Data

No	Item Code	Transaction Time Date/Month/Year	Day	Item Name	Transaction Code	Price
1	1112	20/03/2018	Tuesday	Bread	1001	8000
2	1118	20/03/2018	Tuesday	Coffee	1001	15000
3	1126	20/03/2018	Tuesday	Mineral water	1001	5000
4	1115	20/03/2018	Tuesday	Jam	1001	15000
5	1138	20/03/2018	Tuesday	Cake	1001	45000
6	1148	20/03/2018	Tuesday	Brownies	1001	35000
7	1116	20/03/2018	Tuesday	Cookies	1001	30000
8	1145	20/03/2018	Tuesday	Sandwich	1001	14000
9	1122	20/03/2018	Tuesday	Tea	1001	10000
10	1129	20/03/2018	Tuesday	Juice	1001	10000
…	…	…	…	…	…	…
6017	1119	20/03/2019	Wednesday	Pastry	1578	16000

Table 2.
Determination Number of Clusters Based on SSE

Cluster	SSE value	Difference
2	107,027	
3	76,582	30,445
4	55,442	21,14

Determining the number of k (clusters) that will be used in the study is to take the largest SSE difference value. In table 2, it can be seen that the biggest difference in SSE values is at K = 3. So, the amount of data in this study is divided into 3 clusters. Menu grouping using 3 variables is Transaction Code, Item Code, Day, Price. Table 3 and Table 4 are the results of grouping product items. $Cluster_0$ produces 1750 product items with

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

the dominant item purchased is hotdog on Sundays with a price range of Rp. 10,000, $Cluster_1$ produces 2582 product items with the dominant item purchased is garlic breadstick on Sundays with a price range of Rp. 18,000, $Cluster_2$ produces 1685 product items with the dominant item purchased is milk on Monday with a price range of Rp. 15,000.

Table 3.
Sales Data Cluster Ratio

Cluster	0f Instance	Ratio%
$Cluster_0$	1750	33%
$Cluster_1$	2582	39%
$Cluster_2$	1685	28%

Table 4.
General Sales Data Cluster Model

Attribute	Full Data	$Cluster_0$	$Cluster_1$	$Cluster_2$
	6017	1750	2582	1685
Transaction Code	1021	1015	1033	1012
Item Code	1151	1125	1147	1123
Day	Sunday	Sunday	Sunday	Monday
Price	15000	10000	18000	15000

The process of applying the fp-growth algorithm to recommend bakery package menus begins with the formation of an fp-tree to determine frequent itemset from sales transaction data. Table 5 shows the results of the data that meet the minimum support for each cluster. Furthermore, the generate fp-tree in each cluster. Figure 2 is the generate fp-tree $Cluster_0$, Figure 3 is the generate fp-tree $Cluster_1$ and Figure 4 is the generate fp-tree $Cluster_2$, where each cluster takes 5 transactions as a sample to be presented in this paper.

Table 5.
Frequency Itemset of each cluster (*Min.Supp* = 20%)

$Cluster_0$			$Cluster_1$			$Cluster_2$		
Item Code	Frequency	*Support*	Item Code	Frequency	*Support*	Item Code	Frequency	*Support*
1112	168	57,14 %	1141	237	43,01 %	1115	218	73,15 %
1123	88	29,93 %	1147	298	54,08 %	1116	150	50,33 %
1125	177	60,2 %	1151	266	48,27 %	1123	231	77,52 %
1115	156	53,06 %	1138	176	31,94 %	1118	130	43,62 %
1116	117	39,8 %	1145	196	35,57 %	1126	174	58,39 %
…	…	…	…	…	…	…	…	…
1122	114	38,78 %	1135	128	23,23 %	1122	106	35,57 %
1126	128	43,54 %	1152	111	20,14 %	1125	137	45,97 %
1129	92	31,3 %	1148	217	39,38 %	1126	174	58,39 %

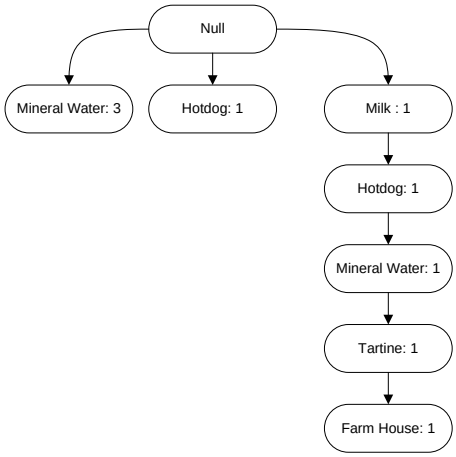

Fig 2. Generate FP-Tree $Cluster_0$

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

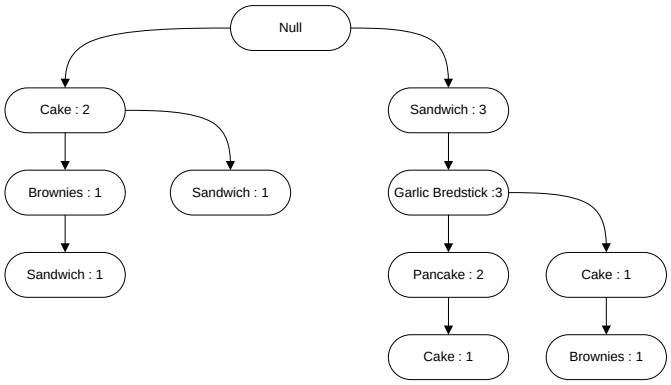

Fig 3. Generate FP-Tree Cluster$_1$

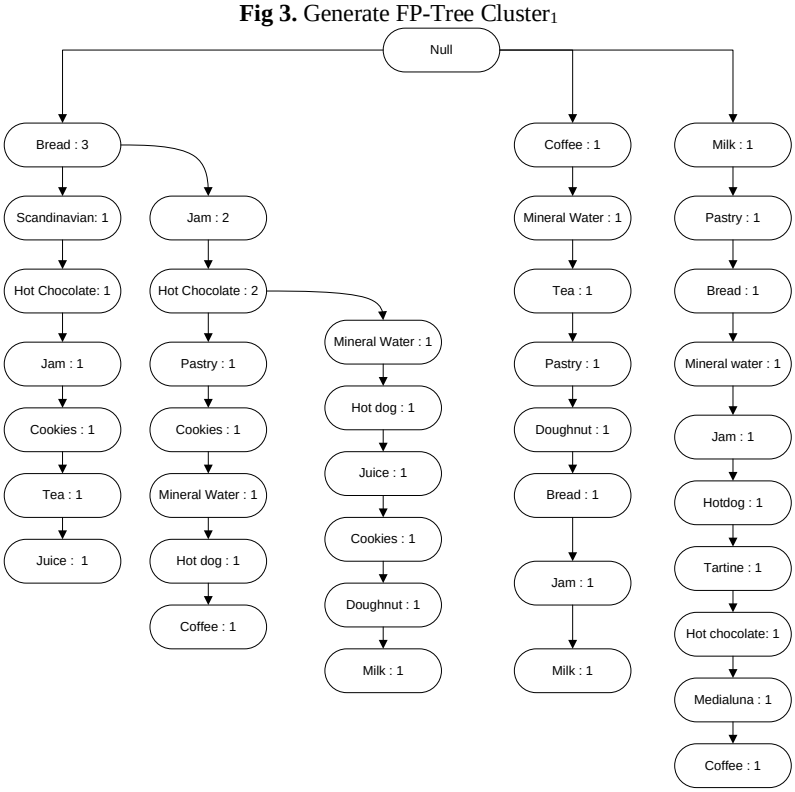

Fig 4. Generate FP-Tree Cluster$_2$

In table 6, the results of the association rule obtained in each cluster meet minimum support of 20% and minimum confidence of 60% with lift > 1. Where in Cluster$_0$ 4 rules are formed, Cluster$_1$ 6 rules and Cluster$_2$ 3 rules with an average the confidence value of each itemset is above 70% for the bakery products that are most often purchased simultaneously. The results of the analysis show the product itemset that so that the results of this association rule from each cluster can be used as a recommendation to create a menu package.

Table 6.
Results of the Association Rules for Each Cluster (*Min. Supp* = 20% ; *Min. Conf* = 60%)

Amount C	The Results of Each Cluster Rules	*Confidence*	*Lift*	*Support*
Cluster$_0$	Hotdog, Pancake ➔ Milk, Garlic Breadstick	0,92	1,25	0,2
	Garlic Breadstick, Hotdog ➔ Chicken Sand, Pancake	0,92	1,5	0,2
	Garlic Breadstick, Pancake ➔ Chicken Sand, Hotdog	0,92	1,34	0,2
	Garlic Breadstick, Hotdog ➔ Milk, Pancake	0,85	1,39	0,2
Cluster$_1$	Hotdog, Pancake ➔ Milk, Garlic Breadstick	0,92	1,25	0,2
	Garlic Breadstick, Hotdog ➔ Chicken Sand, Pancake	0,92	1,5	0,2
	Garlic Breadstick, Pancake ➔ Chicken Sand, Hotdog	0,92	1,34	0,2
	Garlic Breadstick, Hotdog ➔ Milk, Pancake	0,85	1,39	0,2
	Hotdog, Pancake ➔ Chicken Sand, Garlic Breadstick	0,85	1,16	0,2
	Pancake, Chicken Sand ➔ Garlic Breadstick, Hotdog	0,79	1,53	0,2

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

$Cluster_2$	Hotdog, Pancake ➔ Milk, Garlic Breadstick	0,92	1,25	0,2
	Garlic Breadstick, Hotdog ➔ Milk, Pancake	0,85	1,39	0,2
	Garlic Breadstick, Pastry ➔ Milk	0,79	1,34	0,2

5. Conclusion

The resulting menu package recommendations in $cluster_0$ and $cluster_1$ with a combination of 4 itemsets, namely that most customers who buy hotdogs and pancakes will definitely buy milk and garlic breadstick with a confidence of 92% and a lift value of 1.25. Then, customers who buy garlic breadsticks and hot dogs will also buy chicken sand and pancakes with a confidence of 92% and a lift value of 1.5. In $cluster_2$, customers who buy garlic breadstick and pastry will definitely buy milk with a confidence of 79% and a lift value of 1.34.

6. References

[1] E. A. Pribadi, M. N. Alam, Effendy, "Analysis of Revenue Athousehold Bread Industry of Aisyah Bakery in Palu", e-J. Agrotekbis 5 (4) : 466 - 471, Agustus 2017.
[2] D. M. D. U. Putra, B. Setio, "Penerapan Data Mining Pada Penjualan Tan's Bakery Menggunakan Algoritma Apriori. Jurnal Teknologi Informasi Dan Komputer", Volume 3, Nomor 2, 164–174, Oktober 2017.
[3] S. U. Bagus, "Implementasi Algoritma Fp-Growth Pada Sistem Penjualan Toko Roti Untuk Meningkatkan Penjualan Dan Meminimalisir Kerugian", 2018.
[4] D. T. Larose, "Discovering Knowledge In Data An Introduction to Data Mining", ISBN 9786468600, 2005.
[5] Mardalius, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan", ISBN 9786025096105, 2017.
[6] M. M. Benri & L. S. Herlina, "Analysis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila", ISSN 1858 – 2680, Jurnal Media Infotama Vol. 11 No. 2, September 2015.
[7] S. David, "Penerapan Stuktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset", 2008.
[8] Erwin, "Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth" , Jurnal Generic Vol. 4 No. 2, ISSN 1907-4093, Juli 2009.
[9] A. Abdullah, "Rekomendasi Paket Produk Guna Meningkatkan Penjualan Dengan Metode FP-Growth" , Vol. 4 No. 1, ISSN: 2621-038X, Online ISSN: 2477-698X , Juni 2018.
[10] E .Nurlelasari, "Algoritma Asosiasi k-Means dan FP-Growth untuk Analisis Keranjang Pasar pada Penjualan Produk Alumunium", Vol.1 No.2, ISSN : 2548-3331,179-186, Juni 2017
[11] K. Sigit, G Windu, & W. Hari, "Analisis Algoritma FP-Growth Untuk Rekomendasi Produk Pada Data Retail Penjualan Produk Kosmetik", Seminar Nasional Teknologi Informasi dan Komunikasi, ISSN: 2089-9815, Maret 2018.
[12] D. L. Yuyun, "Penerapan Data Mining Menggunakan Algoritma FP-Tree dan FP-Growth Pada Data Transaksi Penjualan Obat", Seminar Nasional Teknologi Informasi dan Komunikasi, ISBN 976-602-19837-9-9, 2015.
[13] S. Imaduddin, H. Assad, & M. Z. Fauzi, "Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations" , Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM) Vol 2, No.2, pp. 107 – 115 p-ISS N: 2614-3372 | e-ISSN: 2614-6150x`, September 2019.
[14] K. Kusrini, "The Third Information Systems International Conference Grouping of Retail Items by Using K-Means Clustering", Procedia Computer Science 72 (2015) 495 – 502, 2015.
[15] Andriyendi, "Prediksi Clustering, Calculation And Classification Fruit And Vegetable Consumption", 2016.
[16] Mardalius, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan" ,2017.