similarity_Recommendation System of Food Package Using Apriori and FP-Growth Data Mining Methods

By Anthony Anggrawan

Recommendation System of Food Package Using Apriori and FP-Growth Data Mining Methods

Christofer Satria, Anthony Anggrawan* and Mayadi
Universitas Bumigora, Mataram, Indonesia
Email: {chris, anthony.anggrawan, mayadi.yadot}@universitasbumigora.ac.id

***Abstract*—Currently, the famous restaurant visited by many people is a roadside stall. Generally, the roadside stall sells multiple kinds of food, drink, and snacks. The problem is that roadside stalls have difficulty determining what food items are best-selling to be used as menu packages of choice from almost hundreds of menu items. That is why it needs data mining of roadside stall sales data to explore correlation information and sales transaction patterns for food items that most often become food pairs sold. Therefore, this study aims to analyze the frequency of the most item sets from data sales in food stalls using the Frequent Pattern Growth (FP-Growth) and Apriori data mining methods to recommend which foods/beverages are the best-selling menu packages. The research and development results show that with 980 transaction data with a minimum support value of 20% and a trust value of at least 50% for FP-Growth, it produces eight valid rules. For Apriori, it has five valid rules as a menu package recommendation. The results of the sales trial of the recommended menu package for two months showed that the total sales increased significantly up to 2.37 times greater than the previous sales**

***Index Terms*— Data Mining; Apriori; FP-Growth; Roadside Stall; Recommendation System; Food Package**

I. Introduction

A roadside stall is a place where certain people buy food and drinks. Each roadside stall has a different menu; in that menu, there must be a special menu from a roadside stall that aims to attract customers with the characteristic taste, delicacy, and enjoyment of the food/drink.

Generally, consumers return to visit due to the taste of the food/drink, the variety of dishes, and the food packages offered at food stalls [1]. Consequently, to remain competitive in retaining customers, food stalls must provide the favorite food consumers love [2]. Unfortunately, the roadside stall has trouble finding out which items menu consumers are most interested in using the manual method [3], even more so in making menu packages from menu items that customers like the most. Often customers do not get their favorite menu items because the ordered food menu items are sold out [3]. It means food stall managers need to know which menu items are often requested by customers and then determine the combination of menu item pairs that customers prefer to serve as food menu packages. The article in this study makes it happen by developing a system model that can recommend food menu packages containing pairs of food items that customers like the most using the data mining method.

The seller/manager does not have a definite ranking order or ensure which food items are the best-selling food items pairs or the most favorite food items pairs for customers. On the other hand, the favorite food menu packages available at the food stalls consisting of the customer's favorite food items will automatically increase customer satisfaction with the available menu packages and make the restaurant or food stall avoid losing customers. Because, after all, previous research confirms that there is a significant relationship between customer sustainability and customer satisfaction [4].

Foods that sell frequently indicate customer preferences for these foods. If the factors that are customer favorites for food and food packages are well available, it will increase productivity in food stalls [1]. Mining data on food sales increase the number of food sales [1]. Food recommendations in data mining generally refer to the customer's favored food items [5].

In other words, if the owner or manager of a restaurant or food stall does not know how to determine the contents of the package menu to be made based on the menu that is most often ordered by their customers, of course, it is unprofitable. Meanwhile, a food system that recommends entrepreneurs and customers based on the behavior of product users contributes to an increase in the proportion of success [5]. Therefore, referring to the background, this study aims to analyze the frequency of the most item sets from big data sales in food stalls using the Apriori and FP-Growth data mining methods to recommend the best-selling food items to serve as menu packages.

In the meantime, the availability of large amounts of data can reveal hidden patterns and correlations [6]. A theory becomes clear and indisputable when the data pattern is known [7]. Meanwhile, mining items with high usage have an essential role in data mining data, which helps make decisions and strategies [8]. Data mining has several well-performing methods for finding patterns, extracting knowledge, and predicting future outcomes [8] [9]. One of the essential functions of data mining is association mining [10]. Disclosure of association information and data patterns helps companies gain

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

broader insights and benefit from the competition [6]. Therefore, it is necessary to analyze big data or big transaction data by doing data mining. However, various studies have investigated data mining techniques because traditional algorithms cannot perform mining results as data mining algorithms [11]. Moreover, previous researchers recommend using data mining technology to get the efficiency of the relationship rules of mining results [12].

For this reason, the method used to process food sales transaction data becomes information using data mining technology. Data mining is a set of rules, processes, and algorithms developed to identify patterns and relationships from large data sets [13]. So in other words, data mining is a process of mining information from data sets using specific methods to recognize patterns and relationships from these data sets. Data mining plays an essential role in discovering hidden patterns from raw data in the database [14]. The extracted data set by data mining can provide crucial details implicitly [15]. The information generated by data mining can be helpful for various fields [15]. Data mining is used not only for obtaining transaction patterns that occur [8] but also for identifying hidden correlations and trends in data using specific methods [16]. The system model built using the data mining method has artificial intelligence [17]. Artificial intelligence is a breakthrough in the latest technology that is widely used in research [17][18]. In short, data mining is the process of extracting information from big data to find new patterns and build patterns [19].

There are many patterns in large data sets [20]. Recognizing patterns in data is valuable or invaluable; the best way is to perform data mining [20]. So no surprise if Ayyoubzadeh et al. (2020) ascertain that utilizing data mining methods on resource data can provide the desired benefits to the user. However, with advances in data mining technology, the need to make it easier for customers to choose recommended products from many existing data items can be realized [5]. Among the data mining activities is identifying things that have high-frequency transactions or often occur in transaction databases [8].

According to Wu et al. (2018), mining items with frequent transactions from a collection of item set databases becomes a reference in business strategy [8]. Among several methods in data mining, one of the data mining methods is the Association Rule method, which is a data mining technique to find associational rules from item combinations. Association rules identify correlations between one item and another [21]. In other words, association rules imply correlations between the valuable data items of the data set [22]

The Association Rule method has several algorithm methods. However, according to Wojciechowski, Galecki, and Gawronek (2007), FP-Growth and Apriori are the most popular existing mining algorithms methods [23]. FP-Growth and Apriori are algorithms used for pattern mining related to transactional databases [24]. Meanwhile, according to Mohammed Al-Maolegi and Bassam Arkok (2014), FP-Growth and Apriori, help to extract the most frequent set of items from a large database [25]. Association rules help to find hidden knowledge. This study uses the Apriori and FP-Growth algorithms to determine the best-selling menu items as menu packages recommended for sale at roadside stalls. The recommended menu packages were tested on sales for two months to determine the effect on sales progress.

Data mining has a mining cycle or methodology that describes the project phases, tasks, and their respective relationships with duty [26]. There are three standard methodologies in the data mining process, namely KDD (Knowledge Discovery from Data), SEMMA (Sample, Explore, Modify, Model, and Assess), and CRISP-DM (Cross-Industry Standard Process for Data Mining) [12]. The CRISP-DM is a data mining methodology currently developing rapidly [13]. CRISP-DM plays a role in providing guidelines for extracting data in large datasets [13]. The standard data mining process used in this study is the Crisp-DM methodology.

The results of data mining analysis using the Apriori and FP-Growth methods on the developed model or the rules generated by the developed model will be used as recommendations for making menu packages based on menu items recommended by the data mining system. By using a support value of 20% and 50% confidence, it is expected to produce many rules and contain 2 item sets in each rule. The number of transactions that will be used in this study is relatively large (980 transactions) and will see the rules generated by the data mining system model, then become the basic recommendations for making menu packages for restaurants or food stalls. Or in other words, the developed data mining system model will extract the menu items that appear most frequently from many food transactions to be used as menu package recommendations. In short, this research specifically aims to develop a data mining system model to recommend menu packages containing customers' favorite food pairs to become mainstay menu packages sold by food stalls. So, in essence, this research's importance is uncovering hidden information (which is difficult to do manually) in getting the pairs of foodstuffs most often purchased together by customers.

The following writing organization of this paper is as follows: the second sub-section discusses the related works of the previous studies. The third sub-section described the research methodology, narrating the research data and methods used. The fourth sub-section discusses the results and discussion of the research. Finally, the conclusions of the research results and suggestions for further research are set out in the sub-section Conclusion.

II. RELATED WORKS

This subsection reviews several related works from recent scientific articles regarding their differences compared to the research in this article.

Patil et al. (2017) developed a computer application to help customers get food-based ordering recommendations based on the customer's profession, age, and gender. The difference between previous research and this research lies in the method used to determine the food items; the previous research used the Artificial Neural Network method [27]. Besides, the food items recommended in the previous study do not identify the best-selling food items and do not recommend food menu packages as researched in this article.

Jimmy Ming-Tai Wu, Justin Zhan, and Sanket Chobe (2018) investigated association mining rules for different associations between low and high-transaction item sets [8]. The similarity of previous research with the research in this article lies in identifying set items with a high frequency of transactions using the data mining algorithm to be used in business strategy. The difference is that the research in this article develops a system to identify the relationship between the best-selling transaction menu items and the making of a recommended menu package which was not carried out in previous studies. In addition, the research in this article builds a web-based application program to test the menu packages recommended by the data mining results that are built and to try out the sales of recommended menu packages that were not found in previous research.

Nur et al. (2019) examined the frequent food items in sales. The research method used in this previous research is the Apriori algorithm [3]. When comparing previous research to the research in this article, this previous study investigated the most frequent menu items ordered or interested by consumers. The difference also lies in the data mining method used in the study to determine the frequency of menu items ordered by consumers. The previous research uses the Apriori algorithm method, while this article uses the Apriori and FP-Growth algorithms. In addition, another difference is that the research in this article is research and development (R&D) to identify the best-selling menu items and determine the level of relationship to create a recommended menu package which was not carried out by a previous study.

Jaiswal (2019) built a diet plan recommendation system for each individual based on their needs. On the other hand, this previous study uses a tree-learning algorithm to create a healthy diet plan for specific individuals [14]. The difference lies in the data mining method used, namely the tree method. In contrast, the article in this study uses the Apriori algorithm and the FP-Growth algorithm and involves the algorithm tree. Furthermore, if previous research has focused on the recommended dietary diet, this article focuses on combining food items for the menu package.

Oiao & Luo (2020) designed an application system for banquet side dishes to choose healthy diet recommendations food for application users. The paper used the Apriori algorithm to build a banquet recommendation system [5]. The previous study has similarities with the article in this study, namely building a computer application system for food recommendations but using different data mining methods. Besides that, the difference also lies in the guidance given; in the previous study, the recommendations focused on the dietary level of the food served at the banquet, whereas in this article, the recommendations focused on the best-selling food items to be menu packages.

Anthony Anggrawan et al. (2021) built an intelligent system to predict drug users and the types of drugs used using the Forward Chaining and Certainty Factor methods [28]. However, unlike the research in this article, building an intelligent system that can determine the menu package consists of any food so that it sells well using the Apriori and the FP-Growth data mining methods. In a different year, Anthony Anggrawan, Mayadi, Christofer Satria & Lalu Ganda Rady Putra (2022) conducted R&D on an intelligent system for ranking scholarship recipients by applying the AHP (Analytical Hierarchy Process) and Moora (Multi-Objective Optimization method with Ratio Analysis method) data mining methods [29]. However, in contrast to the R&D in this article is to build an intelligent data mining system for recommendations for what food combinations to be recommended as menu packages to sell well with the Apriori algorithm and the FP-Growth algorithm.

Prashant Dixit, Harish Nagar, and Sarvottam Dixit (2022) researched to predict student performance based on data mining [30]. The difference between this previous research and the research in this article is in the data mining method used and the research objectives. Previous research used the CBR (case-based reasoning) method or did not use the Apriori and FP-Growth data mining methods as the data mining methods used in the research in this article. In addition, previous research has focused on predicting student performance, while the research in this article focuses on recommending which foods are the best-selling menu packages.

Table 1 shows a comparison between this article compared to previous related works. Referring to the description of the latest related previous works (see also Table 1), the article's essence in this research is research that has not been studied by other researchers before. In short, this article's main strength is the R&D (Research and Development) study that identified the best-selling menu item associations for making menu package recommendations. Another advantage of this research is building a web-based innovative application to replace manual work (identifying the best-selling menu item associations from each food menu sold). In short, the contribution of the results of this study is not only to build a system model of menu package recommendation but also to produce a web-based intelligent application. The web-based innovative application can generate/recommend favorite menu packages from the best-selling food items for consideration for roadside stall sellers to serve as menu packages for food offered to the customer.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research By	Type of Research	Method Used		Build Apps	Field trial	Research Object	Description
		Apriori	FP-Growth				
Patil et al. (2017) [27]	Design system	No	No	Yes	No	To help customers get food-based ordering recommendations based on the customer's profession, age, and gender.	This previous study used the Artificial Neural Network method. However, previous research does not recommend the best-selling foods and does not recommend menu packages like our research.
Jimmy Ming-Tai Wu, Justin Zhan, and Sanket Chobe (2018) [8]	Design system	No	Yes	No	No	Previous research proposed a method to find different association rules for combinations of item sets with low transaction frequency and high transaction frequency items.	This previous study did not develop a system to identify the relationship between the best seller transaction menu items in recommending menu packages as this manuscript researched.
Nur et al. (2019) [3]	Design system	Yes	No	No	No	This previous study aims to help restaurant managers find out which menus are most in demand by buyers.	This previous study did not develop a system to identify the relationship between best-selling transaction menu items in recommending menu packages.
Jaiswal (2019) [14]	Design system	No	No	No	No	This study recommends an application to help everyone to control their diet.	This previous research paper uses the Tree method to find out healthy food with monitored calories and nutrients.
Oiao & Luo (2020) [5]	Design system	Yes	No	No	No	This previous research designed banquet intelligent diet recommendation system.	This previous study did not develop an approach to identify the relationship between best-selling transaction menu items in recommending menu packages.
Anthony Anggrawan et al. (2021) [28]	Design system	No	No	Yes	No	This previous study developed an intelligent system to predict drug users and the types of drugs used by users.	This previous study used the Forward Chaining and Certainty Factor methods in predicting drug users and the types of drugs used by users.
Anthony Anggrawan et al. (2022) [29]	Design system	No	No	Yes	No	This previous study built a recommendation system for scholarship recipients.	This previous study applied the AHP and Moora methods in recommending student recipients.
Prashant Dixit, Harish Nagar, and Sarvottam Dixit (2022) [30]	Design system	No	No	No	No	This previous study aimed to develop a system to predict student performance.	This previous study used the CBR (case-based reasoning) method in predicting student performance.
Our/this research	Experimental (R & D)	Yes	Yes	Yes	Yes	To analyze the frequency of the most item sets from big data sales in food stalls using the Frequent Pattern Growth (FP-Growth) and Apriori data mining methods to recommend which foods/beverages are the best-selling as menu packages	The total sales of food and beverages sold at roadside stalls increased by 2.37 times more than the total sales before implementing sales based on the menu packages recommended by the Apriori and FP-Growth methods.

III. METHODOLOGY

This research is a case study conducted at a roadside stall called *Narmada* food stall. The food stall is one of the roadside stalls in *Narmada* County, West Nusa Tenggara Province, Indonesia. The food stall sells food that provides many items that each buyer can choose freely according to their preferences. There are as many as 70 categories of food ingredients available or offered to buyers. In this study, the big data used for the data mining process is sales transaction data from sales data from June to July 2021 (or approximately sales data 5 months before the sales trial from November 10, 2021, to January 9, 2022).

The data mining methodology used in this study is CRISP-DM; CRISP-DM is a standardized data mining. CRISP-DM consists of a six-stage process [31]; as shown in Figure 1. There are various computer programming languages for building application programs [32]. Java and PHP Hypertext Preprocessor is the most popular high-level programming languages for building a website or mobile-based application programs. In developing application programs for any programming problem, cognitive skills and mastery of programming languages are required [32][33][34]. Creating an application program requires skills or expertise to make it happen. The web-based data mining intelligent application system developed in this study uses the PHP programming language with the FP-Growth algorithm data mining method.

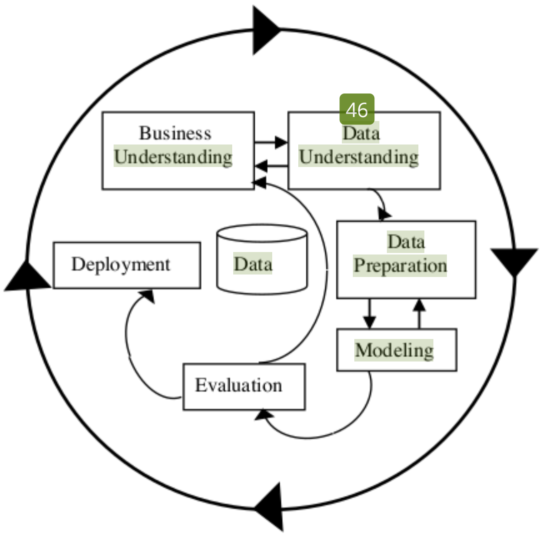

FIGURE 1. CRISP-DM Data Mining Stages [26]

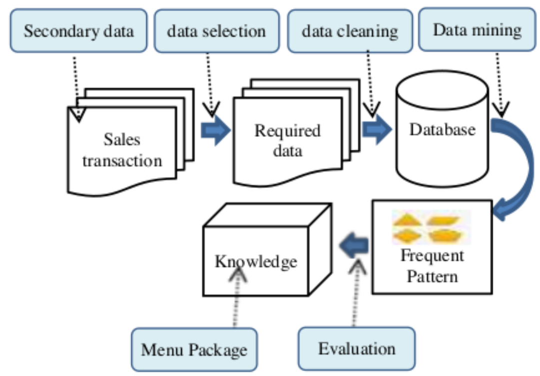

FIGURE 2.Mining Process of Frequent Food Sales Item Set

At the Business understanding stage, activities are carried out to determine sales transaction data at the roadside stall. Sales transaction data that is used as a sample is sales transaction data for two months with a total number of transactions as many as 980 transactions. The data mining construction carried out produces several rules or patterns based on the frequency of transactions from each set of items which are then used as recommendations for making menu packages for roadside stall owners. The recommended menu from the built system consists of 2 types of menu items.

The next stage, the data understanding stage, collects data and prepares to evaluate data requirements. The data collected is from sales transaction data for each menu item in the roadside stall. Data collection from secondary sources is from cash sales records. Sales receipts consist of several attributes of date, item name, item quantity, price, and total price. There are menu item data that are not used in the data mining process in research. Menu items that are not used during the data mining process are food items provided after ordering from consumers and complimentary food items from each menu item ordered from consumers at the roadside stall. Menu items that are not used during the data mining process include grilled fish rice, chili sauce, etc. So the data cleaning process is needed first before the data enters the database (please see Figure 2).

Table 2 is an example of the contents of a sales cash note at the *Narmada* food stall, consisting of several menu items. Because a menu item in this transaction is not used during the data mining process, it is necessary to do data cleaning to adjust to the type of item used in this study.

TABLE II. TRANSACTION DETAILS

Date	Item Name	Number of Items	Price (In Rupiah)	Sub-Total (In Rupiah)
18/06/2021	Iced tea	2	12,000	24,000
	Hot Orange Tea	5	12,500	62,500
	Lombok Coffee	2	11,300	22,600
	Mineral water	4	3,000	12,000
	French fries	3	12,000	36,000
	Grilled sausage	5	13.000	65,000
	Toast	2	13,000	26,000
	Indomie gravy	5	23,000	115,000

Table 3 is the result of selecting the types of items used later in the data mining process. In this study, 980 used transactions were cleaned, which were temporarily stored in an Excel file. After the data is collected, the data needs to be identified, selected, and then built into the data preparation stage's desired format. At this stage, the process that is carried out is to combine data into a more precise form using binomial data (see Table 4); binomial data has a value of only "1" and "0". It means that if the item/code column is "1", then the transaction contains an item that appears. Meanwhile, if the item/code column is "0" the item does not appear in the transaction.

TABLE III. TRANSACTION DATA ITEMS 18/06/2021

Transaction	Item	Item Code
T200618001	Iced tea	Mmc30
	Hot Orange Tea	Mjh28
	Lombok Coffee	Mlk35
	Mineral water	Mmw41
	Warm orange	Mwo45
	Grilled sausage	Csb013
	Toast	Crb008
	Indomie gravy	Mgr01

TABLE IV. DATA TRANSFORMATION

Date	Id Transaction	Item Code						
		Mmc30	Mjh28	…	Mkl34	…	Mam39	Mgr01
18/06/2021	T200618001	1	1	…	0	…	0	0
	T200618002	0	1	…	0	…	0	0
	T200618003	0	0	…	0	…	0	1
	…	…	…	…	…	…	...	…
07/07/2020	T200808002	1	0	…	1	…	0	0
	T200808002	0	1	…	0	…	1	1
	…	…	…	…	…	…	…….	…….
29/07/2020	T201118001	0	1	…	0	…	0	0
	T201118002	1	1	…	0	…	0	0

The Modeling stage is the stage of applying the algorithms used to find, identify and display transaction patterns. The data is selected according to the type of data to be used in the mining process. The data mining technique used is an association technique with two methods: the FP-Growth algorithm and the Apriori algorithm. Mining item sets with Apriori aims to find buying patterns from its customers and FP-Growth. FP-Growth uses tree construction to find the most frequently occurring data sets.

In the Apriori algorithm, the process to get frequent item sets includes: joining and pruning processes. In the joining process, each item is combined with other items until no more combinations are formed. While in the pruning process, the results of the items that have been combined in the previous process are trimmed using the minimum support specified by the user. The disadvantage of the Apriori algorithm is that in performing frequent item set searches, it must scan the database repeatedly for each item combination. As a result, it takes a lot of time to scan the database; In addition, the Apriori process requires a large candidate generation to get a combination of items from the database.

The FP-Growth algorithm is a development of the Apriori algorithm. The FP-Growth algorithm corrects the shortcomings of the Apriori algorithm. In the Apriori algorithm, it is necessary to generate candidates to get frequent item sets. In contrast, the FP-Growth algorithm generates candidate tree-based concepts in the frequent item set search. It is what causes the FP-Growth algorithm to be faster than the Apriori algorithm

At the modeling stage, the rules/purchase patterns are obtained based on the transactions that have been used. The association rules in data mining are carried out in a two-step process: (1) looking for a set of items that frequently occur to determine the minimum support and (2) generating strong association rule from the item set to meet the minimum support and minimum confidence. The support equation is a parameter used to determine data mining patterns to find statistically significant patterns, as shown in equation 1. At the same time, the confidence equation is a measure that shows the relationship between two items conditionally (based on certain conditions), as shown in equation 2.

$$\text{Support }(A,B) = P\,(A \cap B) = \frac{\Sigma \text{ Transaction containing A and B}}{\Sigma \text{ Transaction}} \quad (1)$$

$$\text{Confidence} = P\,(A|B) = \frac{\Sigma \text{ Transaction containing A and B}}{\Sigma \text{ Transaction containing A}} \quad (2)$$

Support (A,B) or P (A∩B) is the support value of 2 items or is the number of transactions containing A and B divided by the total transactions. Confidence or = P (A B) is a measure of the accuracy of a rule, namely the presentation of transactions containing A and B.

In realizing the FP-Growth algorithm, the three stages of the process are as follows: (a) The generation phase of the Conditional Pattern Base is a sub-database that contains the path of the prefix and suffix pattern. The conditional pattern base generation is obtained through the previously built FP-Tree. (b) The generation stage of each item's Conditional FP-Tree Support count in each conditional pattern base is summed. Each item with a support count greater than the minimum support count will be generated with a conditional FP-Tree. (c) Frequent item set search stage if the Conditional FP-Tree is a single path, then the frequent item set is obtained by combining items for each conditional FP-Tree. If it is not a single path, then the FP-Growth generation is done recursively.

Rule correlation is measured by the value of support and confidence and the correlation between two sets of items. The correlation size to get the closeness of the relationship between entities in this study uses equation 3.

$$\text{Lift}(A,B) = \frac{P(AUB)}{P(A).P(B)} \quad (3)$$

Lift (A,B) represents the correlation between A and B. The P(AUB) value is the confidence value of item set A with B. P(A) is the number of transactions containing A. P(B) is the number of transactions containing B. So P(A) P(B) is the number of transactions containing A multiplied by the number of transactions containing B in total transaction. Suppose the result of the calculation in the formula is less than one. In that case, A's correlation is negatively correlated with item set B, which means there is no relationship between them. On the other hand, if the result obtained is more than 1, then A and B are positively correlated. Meanwhile, if the result is equal to 1, then A and B are independent.

IV. RESULT AND DISCUSSION

4.1 Result of Rule Item Set with FP-Growth

In searching for menu packages in this study, 980 transactions were used as sample data in the calculation process. The item name is changed to an item code which is then used as data for the calculation process on FP-Growth. The item code simplifies the calculation process and provides sufficient space because the item code has multiple digits. In other words, using the item name of each existing item will be difficult to process the system because the name of each item with long digits of various item names is relatively large.

Support and confidence tests were carried out with several minimum support and confidence values to get support and confidence values, resulting in more rules and providing menu package recommendations. The first test of support and confidence uses a minimum support value of 20% and minimum confidence of 60%. The second test of support and confidence uses a minimum support value of 30% and minimum confidence of 70%. Meanwhile, the third support and confident test use a minimum support value of 30% and minimum confidence of 60%. Finally the fourth support and confidence test have a minimum value of 20% support and 50% confidence. The four support and confidence test results or rules generated with FP-Growth are shown in Table 5. The minimum value of 20% support and 50% confidence produces the most valid rules, namely eight rules, so these rules are used for making menu items.

TABLE V. RULE COMPARISON RESULTS

No.	Minimum Support	Minimum Confidence	Average Lift Ratio	Generated rules	Valid rules
1	20%	60%	1.06	16	2
2	30%	70%	1.20	3	3
3	30%	60%	1.11	16	5
4	20%	50%	1.20	57	8

Table 6 is the rule used in the recommendation for making menu item packages. There are two recommended menu packages with three sets of items and six recommended menu packages with two sets of items. So the number of menu packages formed is eight menu packages.

TABLE VI. RULE DENGAN 8 RULE VALID

No	Rule
1	If it's orange juice, then french fries
2	If it's grilled sausage, then buy orange juice and french fries
3	If the buyer buys Lombok coffee, the buyer also buys French fries
4	If it's grilled sausage, then orange juice
5	If orange juice, then french fries
6	If Avocado juice, then french fries
7	If it's grilled sausage, then Lombok Coffee
8	If Potato, Grilled Sausage, then Avocado Juice

So the model chosen is a model with a support value of 20% and a confidence value of 50% because it is an association model that produces the most valid rules (ie eight rules). It is also the reason the 20% support value and 50% confidence value are used as the association model for Apriori.

4.2 Search and Result of Rule Item Set with Apriori

Apriori testing process uses a support value of 20%, a confidence value of 50%, and a number of transactions of 980. The test results produce five rules with two rules with three items and three rules with two items, as shown in Table 7.

TABLE VII. RULE RESULTS WITH APRIORI

No	Role		
1	Avocado Juice (Mja 009)	French Fries (Ckg001)	
2	Orange Juice (Mjj012)	French Fries (Ckg001)	Grilled Sausage (Csb013)
3	Avocado juice (Mja 009)	French Fries (Ckg001)	
4	Lombok Coffee (Mkl034)	French Fries (Ckg001)	Grilled Sausage (Csb013)
5	Grilled Sausage (Csb013)	French Fries (Ckg001)	

Table 7 is the result of the Apriori process. The resulting rule helps set menu packages. The resulting menu package is then tested for sales for two months to see the number of sales development.

4.3 Evaluation of menu package recommendations

The evaluation process in making the menu package uses the results with minimum support of 20% and minimum confidence of 50% with FP-Growth, as shown in Table 10, resulting in a lift ratio above 1 with 57 rules and eight valid rules. The eight applicable rules are used as menu packages that are promoted (presented as menu packages that are sold) to customers. The trial process is carried out for two months, from November 10, 2021, to January 9, 2022.

TABLE VIII. EVALUATION RESULTS BASED ON THE FP-GROWTH METHOD

No	Rule	Order Quantity
1	If it's orange juice, then french fries	1250
2	If it's grilled sausage, then buy orange juice and french fries	1585
3	If the buyer buys Lombok coffee than buys French Fries also buys Orange Juice	1020
4	If it's grilled sausage, then orange juice	523
5	If Avocado juice, then french fries	205
6	If it's grilled sausage, then Lombok Coffee	201
7	If Potato, Grilled Sausage, then Avocado Juice	203

TABLE IX. EVALUATION RESULTS BASED ON THE APRIORI METHOD

No	Rule	Order Quantity
1	Orange Juice, French Fries, and Grilled Sausage	1585
2	Avocado Juice and French Fries	205
3	Lombok Coffee, French Fries, and Orange Juice	1020
4	Grilled Sausage and French Fries	876

Tables 8 and 9 are the results of trial sales of food and beverage packages using the Apriori and FP-Growth rules. The test results with FP-Growth show that there are three items with three roles and two items with four roles. The total sales of packages with three items of goods were 2808. Meanwhile, the total sales of packages with two items of goods were 1979 packages. The test results with a priori show that there are three items with two roles and two items with two roles. Total sales of packages of 3 items of goods are 2605, and total sales of packages of 2 items of goods are 1081. The number of rules generated by FP-Growth algorithm is seven rules. In comparison, the number of rules generated by the Apriori algorithm is four rules. The combination of the two rules produces eight different rules. This means there are rules that are twins from the rules resulting from the two methods, namely: If it's grilled sausage, then buy orange juice and french fries; If the buyer buys Lombok coffee, then buys French Fries, also buys Orange Juice; and If Avocado juice, then french fries. Of the four rules generated by the Apriori algorithm, it can be said that almost all of them are represented by the rules generated by the FP-Growth algorithm. This concludes that the FP-Growth algorithm is superior to the Apriori algorithm for obtaining association rules. Table 10 is the test result based on the combination of rules from both Apriori and FP-Growth methods.

TABLE X. EVALUATION RESULTS BASED ON A COMBINATION OF BOTH APRIORI AND FP-GROWTH METHODS

No	Rule	Number of Rules	Order Quantity
1	If Potato, Grilled Sausage, then Avocado Juice	3	203
2	If it's grilled sausage, then buy orange juice and french fries	3	1585
3	If it's Lombok Coffee and French Fries than Orange Juice	3	1020
4	If it's grilled sausage, then orange juice	2	523
5	If orange juice, then french fries	2	1250
6	If Avocado juice, then french fries	2	205
7	If it's grilled sausage, then Lombok Coffee	2	201
8	Grilled Sausage and French Fries	2	876
Total Purchase			5863

Table 11 shows the difference in the increase in total sales before applying the data mining method recommendations and after implementing the data mining recommendations.

TABLE XI. COMPARISON OF SALES RESULTS BEFORE AND AFTER APPLYING THE RESULTS OF THE MENU PACKAGE RECOMMENDATIONS

No	Sales for two months			Percentage of increase in sales
	Food/beverage items	Before	After	
1	Orange juice	1500	3743	149,53
2	French fries	1700	4445	161,47
3	Grilled sausage	760	2309	203,82
4	Lombok Coffee	904	1221	35,07
5	Avocado juice	176	205	16,48
	Total	5040	11923	136,57

The sales testing results on eight valid food/beverage menu package rules can increase sales significantly up to 2.37 times the previous total sales. In short, the contribution of the results of this study is not only to build a menu package recommendation system but also to produce a web-based intelligent application for the implementation of selling menu packages at roadside food stalls. The implementation of sales at roadside food vendors by applying eight valid rules used as package menus shows an increase in the total actual sales results by 136.57% (See Table 11 and Figure 3).

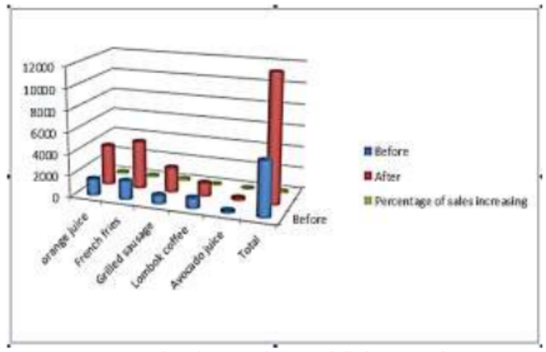

FIGURE 3. The Total Actual Sales Results

V. CONCLUSION

The study results found that applying the FP-Growth and Apriori methods to recommend food and beverage menu packages that customers frequently purchase increased the total sales of food and beverages. The results of this study have positive and significant implications for roadside food and beverage sellers or restaurants. Using data mining methods to get menu package recommendations from food and drinks sold at roadside stalls increases sales of food and beverages up to 2.37 times the previous total sales. The novelty of this study is that the recommendation system model proposed in recommending food and beverage packages by applying two data mining methods, FP-Growth and Apriori, has never been done by other researchers.

The drawback of this study is that not all items from restaurant food were selected as food and beverage items used in the study (because of data cleansing), which needs to be developed in future studies. In addition, it is necessary to conduct further research on the frequency of sales of goods using data mining or other artificial intelligence methods and conduct comparative research using various data mining or artificial intelligence methods. Also, in future research, it is necessary to conduct studies to increase the computation time of the Apriori and FP-Growth algorithms on larger data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and third.

REFERENCES

[1] T. Gayathri, "Data mining of restaurant review using WEKA," *Int. J. Eng. Tech.*, vol. 4, no. 3, pp. 642–645, 2018.

[2] W. K. Liu, Y. S. Lee, and L. M. Hung, "The interrelationships among service quality, customer satisfaction, and customer loyalty: Examination of the fast-food industry," *J. Foodserv. Bus. Res.*, vol. 20, no. 2, pp. 146–162, 2017.

[3] Y. Nur, A. Triayudi, and I. Diana, "Implementation of Data Mining to Predict Food Sales Rate Method using Apriori," *Int. J. Comput. Appl.*, vol. 178, no. 35, pp. 22–28, 2019.

[4] S. Abraham, "A study on Sustainability and customer satisfaction," *Intercont. J. Mark. Res. Rev.*, vol. 1, no. 5, pp. 16–22, 2013.

[5] X. Oiao and J. Luo, "Design Method of Banquet Intelligent Side Dishes System Based on Data Mining and Correlation Analysis," in *International Conference on Software Quality, Reliability, and Security, QRS-C 2020*, 2020, pp. 434–437.

[6] S. Sagiroglu and D. Sinanc, "Big data: A systematic review," in *International Conference on Collaboration Technologies and Systems (CTS)*, 2013, vol. 558, pp. 501–506.

[7] E. Begoli and J. Horey, "Design principles for effective knowledge discovery from big data," in *2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture*, 2012, pp. 215–218.

[8] J. M. T. Wu, J. Zhan, and S. Chobe, "Mining association rules for low-frequency itemsets," *PLoS One*, vol. 13, no. 7, pp. 1–24, 2018.

[9] C. Bellinger, M. S. Mohomed Jabbar, O. Zaïane, and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC Public Health*, vol. 17, no. 1, pp. 1–19, 2017.

[10] Z. Zhao, Z. Jian, G. S. Gaba, R. Alroobaea, M. Masud, and S. Rubaiee, "An improved association rule mining algorithm for large data," *J. Intell. Syst.*, vol. 30, no. 1, pp. 750–762, 2021.

[11] C. Fu, X. Wang, L. Zhang, and L. Qiao, "Mining algorithm for association rules in big data based on Hadoop," in *AIP Conference Proceedings*, 2018, vol. 1955, no. April, pp. 1–6.

[12] Y. H. Choa, J. K. Kimb, and S. H. Kim, "A personalized recommender system based on web usage mining and decision tree induction," *Expert Syst. Appl.*, vol. 23, pp. 329–342, 2020.

[13] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, pp. 1–43, 2020.

[14] V. Jaiswal, "A new approach for recommending healthy diet using predictive data mining algorithm," *Int. J. Res. Anal. Rev.*, vol. 6, no. 1, pp. 58–66, 2019.

[15] S. M. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, and S. R. N. Kalhori, "Predicting COVID-19 Incidence Using Data Mining Techniques: A case study of Pakistan," *JMIR Public Heal. Surveill.*, vol. 11, no. 4, pp. 168–184, 2020.

[16] A. S. Bozkir and E. A. Sezer, "Predicting food demand in food courts by decision tree approaches," *Procedia Comput. Sci.*, vol. 3, no. 2011, pp. 759–763, 2011.

[17] A. Anggrawan, C. Satria, B. K. Triwijoyo, and R. Rismayati, "Machine Learning to Predict the Treatment Status of Covid-19 Patients," *J. Adv. Inf. Technol.*, vol. 13, 2022.

[18] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[19] S. Arora, “Data Mining Vs. Machine Learning: What Is the Difference?," *Simplilearn Solut.*, vol. Juni, pp. 1–9, 2021.

[20] G. Smith, "Data mining fool's gold," *J. Inf. Technol.*, vol. 35, no. 3, pp. 182–194, 2020.

[21] O. M. Epifania, P. Anselmi, and E. Robusto, "DscoreApp: A Shiny Web Application for the Computation of the Implicit Association Test D-Score," *Front. Psychol.*, vol. 10, no. January, pp. 1–10, 2020.

[22] W. A. C. Smink, J. P. Fox, E. T. K. Sang, A. M. Sools, G. J. Westerhof, and B. P. Veldkamp, "Understanding therapeutic change process research through multilevel modeling and text mining," *Front. Psychol.*, vol. 10, no. MAY, pp. 1–14, 2019.

[23] M. Wojciechowski, K. Galecki, and K. Gawronek, "Three strategies for concurrent processing of frequent itemset queries using FP-growth," in *International Workshop on Knowledge Discovery in Inductive Databases*, 2007, vol. 4747 LNCS, no. 1, pp. 240–258.

[24] K. Garg and D. Kumar, "Comparing the Performance of Frequent Pattern Mining Algorithms," *Int. J. Comput. Appl.*, vol. 69, no. 25, pp. 21–28, 2013.

[25] M. Al-Maolegi and B. Arkok, "An Improved Apriori Algorithm for Association Rules," *Int. J. Nat. Lang. Comput.*, vol. 3, no. 1, pp. 21–29, 2014.

[26] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[27] P. D. Patil, N. P. Patil, P. Y. Geete, A. V Nikat, and P. S. Kulkarni, "Food Recommendation System," *Int. J. Sci. Technol. Eng.*, vol. 3, no. 12, pp. 15–19, 2017.

[28] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[29] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[30] P. Dixit, H. Nagar, and S. Dixit, "Student performance prediction using case based reasoning knowledge base system (CBR-KBS) based data mining," *Int. J. Inf. Educ. Technol.*, vol. 12, no. 1, pp. 30–35, 2022.

[31] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[32] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[33] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar

Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.

[34] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

similarity_Recommendation System of Food Package Using Apriori and FP-Growth Data Mining Methods

ORIGINALITY REPORT

15%
SIMILARITY INDEX

PRIMARY SOURCES

1	media.neliti.com Internet	114 words — 2%
2	thesai.org Internet	103 words — 2%
3	www.ijiet.org Internet	59 words — 1%
4	journal.universitasbumigora.ac.id Internet	40 words — 1%
5	Islamiyah, Putri Lestari Ginting, Nataniel Dengen, Medi Taruk. "Comparison of Priori and FP-Growth Algorithms in Determining Association Rules", 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2019 Crossref	38 words — 1%
6	jurnal.narotama.ac.id Internet	34 words — 1%
7	Ines Bouzouita. "Efficient Generic Association Rules Based Classifier Approach", Lecture Notes in Computer Science, 2008 Crossref	30 words — < 1%

8	Rinna Rachmatika, Kecitaan Harefa. "Analysis of Determination of Strategy Promotion using Apriori Algorithm", Journal of Physics: Conference Series, 2020 Crossref	30 words — < 1%
9	Vijayalakshmi, M., and B. Menezes. "Using Data Mining to Identify a Consistent Set of Experts for Times Series Sales Forecasting", Artificial Intelligence and Applications, 2010. Crossref	29 words — < 1%
10	ejurnal.seminar-id.com Internet	28 words — < 1%
11	Larose, . "Association Rules", Discovering Knowledge in Data, 2014. Crossref	26 words — < 1%
12	J Chandra, K R Dewi. "Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method", IOP Conference Series: Materials Science and Engineering, 2019 Crossref	22 words — < 1%
13	mobt3ath.com Internet	22 words — < 1%
14	www.mdpi.com Internet	21 words — < 1%
15	Harpreet Singh, Parminder Kaur. "chapter 9 Website link Structure Optimization", IGI Global, 2016 Crossref	19 words — < 1%
16	Natalia Distefano, Salvatore Leonardi, Giulia Pulvirenti, Richard Romano, Erwin Boer, Ellie	19 words — < 1%

Wooldridge. "Mining of the association rules between driver electrodermal activity and speed variation in different road intersections", IATSS Research, 2021
Crossref

17 Oloan Sihombing, Veronica T. D. Simamora, Vivian P. Leo, Petrus Kristian P. Marpaung. "Implementation of The Apriori Algorithm Method to Analyze the Pattern of Motorcycle Sales at PT. Alpha Scorpio", Paradigma - Jurnal Komputer dan Informatika, 2022 — 19 words — < 1%
Crossref

18 eprints.undip.ac.id — 19 words — < 1%
Internet

19 Krzysztof J. Cios. "Unsupervised Learning: Association Rules", Data Mining, 2007 — 17 words — < 1%
Crossref

20 dokumen.pub — 17 words — < 1%
Internet

21 www.zora.uzh.ch — 16 words — < 1%
Internet

22 Tarannum Shaila Zaman, Tingting Yu. "Extracting implicit programming rules: comparing static and dynamic approaches", Proceedings of the 7th International Workshop on Software Mining - SoftwareMining 2018, 2018 — 15 words — < 1%
Crossref

23 Phindika Phindika, Yopi Handrianto, Sulaeman Hadi Sukmana. "Application of medical equipment procurement data mining using the Apriori method", SinkrOn, 2021 — 14 words — < 1%
Crossref

Lecture Notes in Computer Science, 2012.

24	Crossref	13 words — < 1%
25	Advances in Intelligent Systems and Computing, 2016. Crossref	12 words — < 1%
26	Altuntas, S.. "Facility layout using weighted association rule-based data mining algorithms: Evaluation with simulation", Expert Systems With Applications, 201201 Crossref	11 words — < 1%
27	e-tarjome.com Internet	11 words — < 1%
28	Korczak, Jerzy, and Piotr Skrzypczak. "FP-Growth in Discovery of Customer Patterns", Lecture Notes in Business Information Processing, 2012. Crossref	10 words — < 1%
29	thescipub.com Internet	10 words — < 1%
30	B. Subbulakshmi, C. Deisy. "An adaptive algorithm for frequent pattern mining over data streams using diffset strategy", International Journal of Business Information Systems, 2019 Crossref	9 words — < 1%
31	Dejan Lukic, Robert Cep, Jovan Vukman, Aco Antic, Mica Djurdjev, Mijodrag Milosevic. "Multi-Criteria Selection of the Optimal Parameters for High-Speed Machining of Aluminum Alloy Al7075 Thin-Walled Parts", Metals, 2020 Crossref	9 words — < 1%
32	W Septiani, I A Marie, D Sugiarto, L Hakim. "The Pattern Failure Analysis of Sulfuric Acid Production	9 words — < 1%

Process with the Association Rules Algorithm Apriori", IOP Conference Series: Materials Science and Engineering, 2019
Crossref

33	eprints.utas.edu.au Internet	9 words — < 1%
34	iocscience.org Internet	9 words — < 1%
35	www.semanticscholar.org Internet	9 words — < 1%
36	"Encyclopedia of Database Systems", Springer Science and Business Media LLC, 2018 Crossref	8 words — < 1%
37	R. Ivancsy, S. Juhasz, F. Kovacs. "Performance prediction for association rule mining algorithms", Second IEEE International Conference on Computational Cybernetics, 2004. ICCC 2004., 2004 Crossref	8 words — < 1%
38	ijair.id Internet	8 words — < 1%
39	lib.um.ac.id Internet	8 words — < 1%
40	ojs.stmikpringsewu.ac.id Internet	8 words — < 1%
41	www.polyvyanyy.com Internet	8 words — < 1%
42	www.sciencegate.app Internet	8 words — < 1%

43	Jimmy Ming-Tai Wu, Justin Zhan, Sanket Chobe. "Mining Association rules for Low-Frequency itemsets", PLOS ONE, 2018 Crossref	7 words — < 1%
44	Lasmedi Afuan, Ahmad Ashari, Yohanes Suyanto. "Query Expansion in Information Retrieval using Frequent Pattern (FP) Growth Algorithm for Frequent Itemset Search and Association Rules Mining", International Journal of Advanced Computer Science and Applications, 2019 Crossref	7 words — < 1%
45	infor.seaninstitute.org Internet	7 words — < 1%
46	scar Marbn, Gonzalo Mariscal, Javier Segovi. "Chapter 1 A Data Mining & Knowledge Discovery Process Model", IntechOpen, 2009 Crossref	7 words — < 1%
47	"Data Science and Security", Springer Science and Business Media LLC, 2021 Crossref	6 words — < 1%
48	Ahmad Ari Aldino, Evi Dwi Pratiwi, Setiawansyah, Sanriomi Sintaro, Ade Dwi Putra. "Comparison Of Market Basket Analysis To Determine Consumer Purchasing Patterns Using Fp-Growth And Apriori Algorithm", 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2021 Crossref	6 words — < 1%