# การวิเคราะห์ตระกร้าตลาดโดยใช้กฎความสัมพันธ์และแอปพลิเคชันการใช้งาน

# Market basket analysis using association rules and its applications

ศิริลักษณ์ แก้วศิริรุ่ง[1], ขนิษฐา หอมจันทร์[2], วรวิทย์ ฝั้นคำอ้าย[3], กัญญ์ณพัชญ์ พลเยี่ยม[4], จักรี ม่วงสาร[5], วราวุธ วิสมกา[6] และ นงนุช เกตุ้ย[7*]

Sirilux Kaewsirirung[1], Kanitha Homjun[2], Worawit Fankam-ai[3], Kannapat Polyiam[4], Jakkrit Muangsan[5], Warawut Wisomka[6] and Nongnuch Ketui[7*]

**Received**: 18 December 2023 ; **Revised**: 15 January 2024 ; **Accepted**: 1 February 2024

## บทคัดย่อ

ปัจจุบันบ่อยครั้งนักท่องเที่ยวในประเทศไทยต้องการเลือกซื้อผลิตภัณฑ์ท้องถิ่น เรียกว่า หนึ่งตำบลหนึ่งผลิตภัณฑ์ (OTOP) หลังจากท่องเที่ยว ซึ่งปัญหาหนึ่งของนักท่องเที่ยวที่มีความยากตัดสินใจว่าผลิตภัณฑ์ใดที่ขายดี หรือโปรโมชั่นใดที่เป็นทางเลือกในกลุ่มผลิตภัณฑ์มากมาย จากสาเหตุดังกล่าวจำเป็นต้องใช้การเรียนรู้กฎความสัมพันธ์ เพื่อค้นหาความสัมพันธ์ของข้อมูลและรูปแบบรายการขายของผลิตภัณฑ์หนึ่งตำบลหนึ่งผลิตภัณฑ์ที่มียอดความถี่ของคู่ผลิตภัณฑ์ที่ซื้อพร้อมกัน ในงานวิจัยนี้มุ่งเน้นการวิเคราะห์ด้วยความถี่ของการซื้อขายผลิตภัณฑ์หนึ่งตำบลหนึ่งผลิตภัณฑ์ และเปรียบเทียบประสิทธิภาพของเทคนิคเอฟพีโกรท และอัลกอริทึมอะพริโอริสำหรับการแนะนำผลิตภัณฑ์หนึ่งตำบลหนึ่งผลิตภัณฑ์ สำหรับการทดลองได้ใช้ชุดข้อมูลจำนวน 2 ชุด จากร้านบ้านถั่วลิสงและศูนย์จำหน่ายสินค้าโอทอป ระหว่างปี พ.ศ. 2559-2565 ที่มีจำนวน 200,000 รายการ การศึกษามุ่งเน้นการเปรียบเทียบประสิทธิภาพของเทคนิคเอฟพีโกรท และอัลกอริทึมอะพริโอริ พบว่าเวลาการประมวลผลของอัลกอริทึมอะพริโอริดีกว่าเทคนิคเอฟพีโกรท ในขณะที่ภาพรวมการประเมินความพึงพอใจของผู้ใช้งานของแอปพลิเคชันแนะนำผลิตภัณฑ์อยู่ในระดับดีมาก หรือเท่ากับ 4.72

**คำสำคัญ:** กฎความสัมพันธ์, อัลกอริทึมอะพริโอริ, เอฟพีโกรท, วิเคราะห์ตระกร้าตลาด, ระบบแนะนำ

## Abstract

Currently, many tourists in Thailand often discover a local product called One Tambon One Product (OTOP) during their travels. The problem is that tourists have difficulty determining which products are best-selling and suitable for use as their preferred item packages from the hundreds of available options. This is why association rule learning is needed to explore the correlation information and sales transaction patterns for OTOP items that are most frequently sold as product pairs. Our research aims to analyze the frequency of the most popular item sets from sales data in OTOP and to compare the performance of the Frequent Pattern Growth (FP-Growth) algorithm and the Apriori algorithm for OTOP recommendations. We used two datasets from Peanut House and the Nan OTOP Center, covering the years

[1] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: siriluxk@rmutl.ac.th
[2] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: kanithahomjun@rmutl.ac.th
[3] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: worawit@rmutl.ac.th
[4] คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: duangsamorn.d@rmutl.ac.th
[5] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: jakkri_mu63@live.rmutl.ac.th
[6] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: i-y@hotmail.com
[7] คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา น่าน Email: nongnuchketui@rmutl.ac.th
[1] Faculty of Science and Agricultural Technology, RMUTL Nan Email: siriluxk@rmutl.ac.th
[2] Faculty of Science and Agricultural Technology, RMUTL Nan Email: kanithahomjun@rmutl.ac.th
[3] Faculty of Science and Agricultural Technology, RMUTL Nan Email: worawit@rmutl.ac.th
[4] Faculty of Business and Liberal Art, RMUTL Nan Email: duangsamorn.d@rmutl.ac.th
[5] Faculty of Science and Agricultural Technology, RMUTL Nan Email: jakkri_mu63@live.rmutl.ac.th
[6] Faculty of Science and Agricultural Technology, RMUTL Nan Email: i-y@hotmail.com
[7] Faculty of Science and Agricultural Technology, RMUTL Nan Email: nongnuchketui@rmutl.ac.th
* Corresponding Author Email: nongnuchketui@rmutl.ac.th

2016 to 2022, with a total of 200,000 records for the experiment. This study aims to compare the performance of the Apriori and FP-Growth algorithms. The execution time of the Apriori algorithm outperformed that of the FP-Growth method. Overall, user satisfaction with the recommended system is rated very high, at 4.72.

**Keywords:** Association Rule, Apriori Algorithm, FP-Growth, Market Basket Analysis, Recommended System

## Introduction

Nan Province is located in the north of Thailand and is considered a non-principal town according to government policy. Many tourists like to visit and discover local products as souvenirs. These local products are known as One Tambon One Product (OTOP). The problem is that tourists have difficulty determining which Nan OTOP items are best-selling and suitable for use as their preferred item packages.

Market basket analysis is a data mining technique employed by merchants to enhance sales by gaining deeper insights into consumer purchasing behaviors. Data mining typically employs four techniques to generate descriptive and predictive capabilities: regression, association rule discovery, classification, and clustering. One popular application of data mining in recommender systems is finding association rules. Association rules have been successfully applied in various contexts, such as optimizing shelf arrangements in retail stores and determining promotional strategies.

Association rule learning can be applied to OTOP's market basket analysis to identify best-selling items and determine effective product promotion sets. A minimal degree of support and confidence is required as part of the association criteria. The initial stage in producing association rules involves applying a minimally supported frequent itemset to a given set of frequent items. Next, we can infer selection criteria for items from collections by setting a minimal level of certainty. Each item in the association rules is then linked together using the lift value. The rules take the form of "if A, then B," where "A" and "B" can represent specific items, values, words, etc. The association rule consists of two item sets:

1. Left-Hand Side (L) or antecedent
2. Right-Hand Side (R) or consequent

The reliability of a relationship can be evaluated based on support, confidence, and lift. Under the minimal support and minimum confidence assumptions, all association rules are examined during association rule learning. The process of mining association rules involves two stages: (Patil, M., & Patil, T., 2022) identifying the most commonly used groups of items as the join step and evaluating all the rules with threshold values to prune those that do not meet the minimum support and confidence criteria.

Market basket analysis uncovers customers' purchasing patterns by identifying significant associations among the products included in their shopping baskets (Hossain, M., Sattar, A.S., & Paul, M. K., 2019). The findings indicate that when analyzing the most popular items, it is feasible to achieve nearly identical frequent item sets and association rules in a shorter time compared to computing all items. Moreover, a comparison of the time taken between the FP-Growth and Apriori algorithms reveals that the former is more time-efficient.

Online shopping is gaining immense popularity in the modern virtual market, with customers making purchasing decisions based on their basic and relative needs. Shopkeepers play a significant role in influencing customers in the physical market. A recommendation engine serves as an automated shopkeeper, providing valuable suggestions. In this research (Tareq, S.U., Noor, M. H., & Bepery, C., 2020), the FP-Growth method was utilized due to its efficiency compared to the Apriori algorithm, which requires substantial time and memory resources to operate optimally.

Incorporating data mining techniques, companies can analyze transaction data to uncover consumer buying patterns. In this research (Aldino, A.A., Pratiwi, E.D., *et al*., 2021), the authors employed association rule mining, commonly known as market basket analysis, for processing transaction data using RapidMiner. The study compares the performance of the FP-Growth and Apriori algorithms in this context. Based on the comparison results, it can be inferred that the FP-Growth algorithm outperforms the Apriori algorithm.

A study (Fadillah, A. R., Yulita, *et al.*, 2021) used the FP-Growth association algorithm to evaluate transaction data and identify the best parameters for cross-selling and upselling coffee. The association guidelines can assist organizations in implementing effective cross-selling and upselling techniques.

Knowledge discovery in databases (KDD) and data mining (DM) are processes for extracting useful information from large datasets using tools such as statistical analysis, machine learning, and database management systems (Goebel, V., 2014). The following are the three steps involved in data mining:

**1. Preprocessing** involves obtaining the right data and organizing it in a useful way. The raw dataset needs to have noise removed, and the remaining relational dataset will be used to build the actual model.

**2. Modeling** means creating a representation using the selected data. The appropriate approach is chosen after exploring the problem, and the model is then tested with new data.

**3. Postprocessing** entails putting the model into operation in the real world via an application programming interface.

Purba *et al*. (Fatoni, C. S., Utami, E., & Wibowo, F. W., 2018) proposed an Android app for shoe recommendations using the Apriori algorithm. The system collects user-shoe interaction data to generate association rules, which are used to recommend shoes to users based on their past interactions. The system is user-friendly and provides relevant recommendations.

PK Singh, M. Sinha, S. Das, and P. Choudhury (Singh, P.K., Sinha, M., *et al.*, 2020) propose a method to improve the accuracy of item-based collaborative filtering by using the Bhattacharyya coefficient to measure similarity between items. The authors also consider the categorical attributes of items when calculating similarity.

Both the Apriori and FP-Growth algorithms are well-known techniques for finding such recurrent patterns. This paper studies the performance evaluation of these two techniques based on the Nan OTOP Center dataset. We compare the execution times of the two algorithms and identify the factors influencing performance. The association rules generated by the algorithms are used to analyze Nan OTOP's market basket, and they can also inform OTOP promotions.

## Data Mining

### Association Rule Learning

Association rule learning is a machine learning technique that uses rules to find novel correlations in large datasets (Liu, B., Hsu, W., & Ma, Y., 1998). To produce the association rules, one must first identify the frequent itemsets—groups of items that appear frequently in transactions. Then, using support, confidence, and lift values, the association rules that emerge from these groups can be uncovered. Let LLL denote an itemset, L→RL \right arrow RL→R an associative rule, and TTT a collection of database transactions.

1. **Support** measures the frequency of occurrence of a set in the dataset.

$$sup(L) = \frac{freq(L)}{freq(T)} \tag{1}$$

$$sup(R) = \frac{freq(R)}{freq(T)} \tag{2}$$

Here, freq(L)\text{freq}(L)freq(L) represents the number of times item LLL (the left side) appears in all transactions. The frequency of occurrence of item RRR (the right side) in all transactions, denoted by freq(T) \text{freq}(T)freq(T), is referred to as freq(R)\text {freq}(R)freq(R).

2. **Confidence** indicates the likelihood that the association rule is true based on its historical occurrence in many instances.

$$conf(L \to R) = \frac{sup(L, R)}{sup(L)} \tag{3}$$

Here, sup(L,R)\text{sup}(L, R)sup(L,R) represents the frequency with which the itemsets LLL and RRR occur together.

3. **Lift** compares the observed support to the expected support under the assumption that LLL and RRR are independent.

$$lift(L \rightarrow R) = \frac{sup(L,R)}{sup(L)\ x\ \sup(R)} \tag{4}$$

Here, sup(L)\text{sup}(L)sup(L) is the relative frequency of item LLL (the left-hand side), and freq(T)\text{freq}(T)freq(T) is the total frequency of all transactions. Similarly, sup(R)\text{sup}(R)sup(R) represents the relative frequency of item RRR (the right-hand side) in relation to the entire transaction set freq(T)\text{freq}(T)freq(T).

### Apriori Algorithm

The Apriori algorithm, first presented by R. Agrawal and R. Srikant in 1994 (Yuan, X., 2017), is considered a pioneering and fundamental approach for identifying frequent itemsets. The Apriori algorithm employs a two-step approach to explore candidates: (1) generating all frequent itemsets, where an itemset is considered frequent if its occurrence surpasses a predefined minimum support threshold, and (2) creating (k+1)(k+1)(k+1)-itemsets based on the frequent kkk-itemsets that have already been explored, retaining only the frequent (k+1)(k+1)(k+1)-itemsets. This effectively applies a priori pruning to eliminate infrequent (k+1)(k+1)(k+1)-itemsets.

### FP-Growth Algorithm

The FP-Growth algorithm uses a divide-and-conquer strategy to mine common database entries. It involves two database scans. In the initial scan, FP-Growth creates the F-List, which is a list of frequent items sorted by frequency in descending order. The second scan compresses the database into an FP-tree (Li, H., Wang, Y., Zhang, D., *et al*., 2008). Once the FP-tree is constructed, FP-Growth recursively mines the tree to identify frequent itemsets. It does so by building a conditional FP-tree for each item whose support is greater than a predefined threshold. This recursive process involves constructing and searching trees to uncover frequent itemsets, effectively converting the problem of finding frequent itemsets into a tree-based search problem.

## Methodology

In this research, we evaluate the performance of the Apriori and FP-Growth algorithms in conducting a market basket analysis for OTOP products, focusing on execution time and the number of generated rules. The methodology for this experiment is structured around the data mining process, which includes preprocessing, modeling, and postprocessing, as described below.

**1. Preprocessing**

The datasets from Nan OTOP are used in the experiment and are processed through three sub-processes: data cleaning, data selection, and data transformation. The preprocessing steps are as follows:

● **Data Cleaning:** The Nan OTOP dataset is collected from two sources: (1) the NAN OTOP Center (577,861 records from 2016 to 2022) and (2) the Peanut House (77,758 records from 2019 to 2022). The details of the datasets are shown in Table 1.

**Table 1** Statistics of Each Dataset

| Detail | Nan OTOP Center (records) | Peanut House (records) |
|---|---|---|
| Product | 3,841 | 569 |
| Sale transactions | 212,507 | 24,946 |
| Sale Item Transactions | 577,861 | 77,758 |

● **Data Selection:** Nine features are available for analysis: PAYMENT-NO, PAYMENT-DATE, PRODUCT-ID, PRODUCT-NAME, QUANTITY, CATEGORY-NAME, STORE-ID, and STORE-NAME. For our analysis, we selected three features: PAYMENT-NO (which defines the sale transaction), PAYMENT-DATE (which indicates the sale date), and PRODUCT-ID (which identifies the product). Since association rule learning is based on the relationships between items, we eliminated transactions that contained only a single item. Additionally, PAYMENT-NO and PRODUCT-NAME were replaced with TRANSACTION-ID and ITEMS, respectively.

● **Data Transformation:** The dataset is transformed into JSON file format. One-to-one transactions are converted into multiple-item sales transactions. For example, if both transactions with PAYMENT-NO (199379) and PAYMENT-NO (199380) contain the same PRODUCT-ID (101472), both PAYMENT-NOs are removed. However, if transaction PAYMENT-NO (199381) is associated with PRODUCT-ID (101588) and PRODUCT-ID (101589), this transaction is retained for model creation.

**2. Modeling**

After data processing, we compare the performance of two algorithms: Apriori and FP-Growth, focusing on execution time and the number of generated association rules. The modeling processes are described as follows:

● **Analysis:** Now that the dataset is prepared for analysis, we will determine the likelihood that consumers who purchased item L from the Nan OTOP products also purchased item R from the same group. This information is essential for the Peanut House and Nan OTOP Center to decide whether to reorder new items from the same group (item R).

● **Formalization:** To assist with the analysis, we will look for pairwise association rules in the dataset and calculate the confidence conf(itemL → itemR)\ text{conf}(item L \rightarrow item R)conf(itemL → itemR) of the item pairs, as described in the preceding equation (Equation 3). We will only consider instances where item L → item R occurs a minimum of MIN-COUNT times, and the calculated confidence value exceeds the THRESHOLD. If the association rules extracted from the Nan OTOP dataset include item L → item R, then the Nan OTOP Center will contemplate reordering it. We have effectively completed the two phases of association rule mining and are now ready to implement the algorithms.

● **Algorithms:** The association rules from the Apriori algorithm are generated based on the frequency with which item L and item R occur together. The lift value is calculated by taking the ratio of the frequency of item L to the total number of transactions (T) and the ratio of the frequency of item R to the total number of transactions (T). Meanwhile, a list of frequent items is sorted by their frequency in descending order, and the database is compressed into the FP-tree. The FP-Growth algorithm then recursively mines the tree to identify frequent itemsets until the support value exceeds a predefined threshold. We use the generated rules in the format itemL → itemRitem L \rightarrow item RitemL → itemR. Thus, the condition for the rule is that the lift value must be greater than 1, and the output is sorted by confidence value in descending order.

**3. Post-Processing**

In this experiment, 2,200 association rules were generated, and we need to refine these rules based on the following criteria:

● Set the minimum support and minimum confidence thresholds to ensure the frequency of the itemsets is sufficiently high. If multiple rules have the same maximum confidence, the lift value will be used as a tiebreaker.

● Eliminate ordered products that appear on the shelf, and select only the left-hand side of products that have corresponding sales transactions.

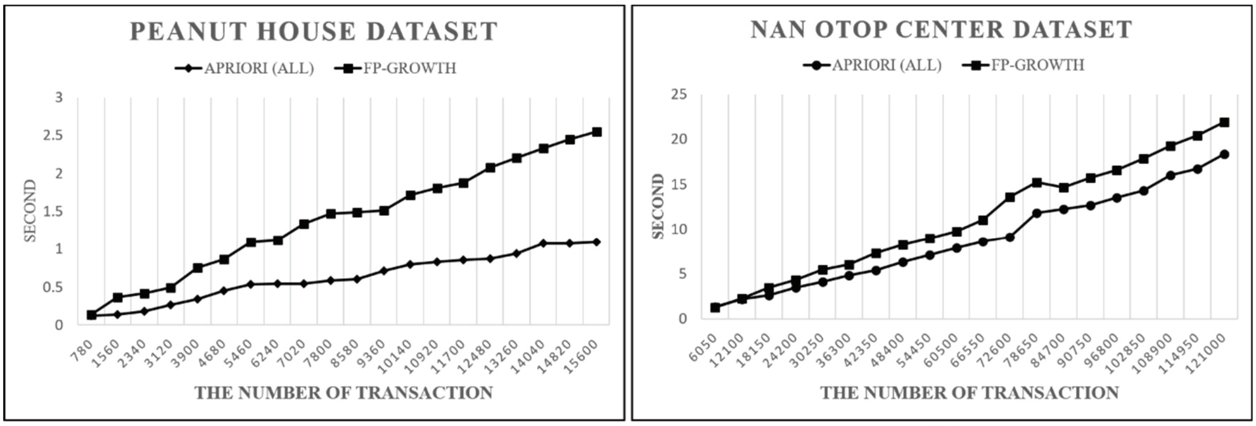
● Use the permutation method (Ketui, N., Wisomka, W., & Homjun, K., 2019) to generate candidate rules and identify the most significant rules.

**Experimental Results and Discussion**

In our experiment, we analyzed the execution time of two algorithms and studied the number of rules generated. Figures 1 (a) and (b) show the execution times of the algorithms based on the Peanut House dataset and the Nan OTOP Center dataset. The Apriori algorithm (ALL) was executed without imposing any conditions on minimum support or minimum confidence (MIN-SUP and MINCONF $\geq$ 0).

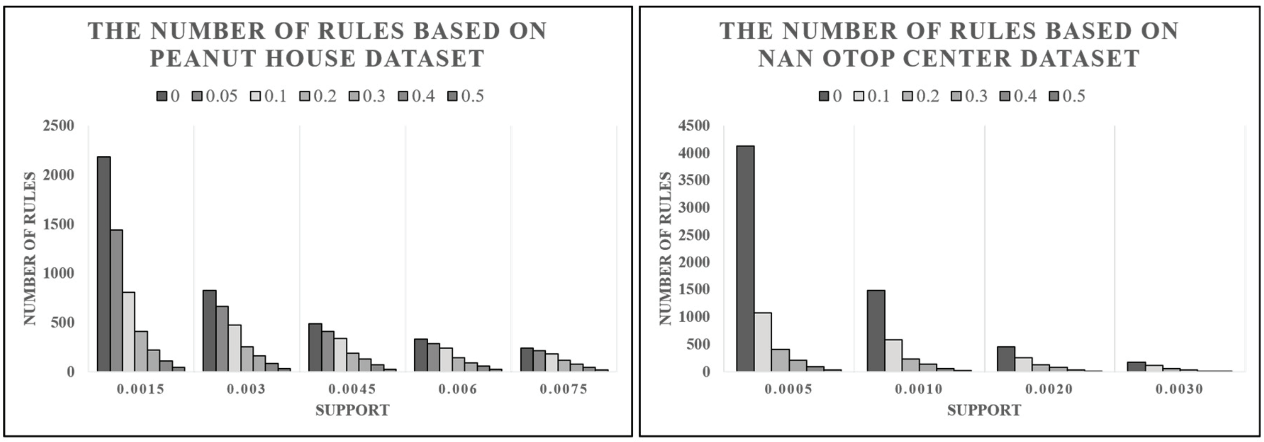
For the Peanut House dataset, the transactions ranged from 780 to 15,600 records. We found that the execution time of the Apriori algorithm outperformed that of the FP-Growth algorithm (see Figure 1 (a)). In contrast, when we considered the Nan OTOP Center dataset, which included approximately 6,000 to 120,000 transactions, the FP-Growth algorithm exhibited longer execution times than the Apriori algorithm (Figure 1 (d)).

As shown in Figure 1, the Apriori algorithm is suitable for both OTOP datasets, even though they differ significantly. This effectiveness can be attributed to the simplicity of the Apriori algorithm, which operates based on the frequency of product sets.




(a) (b)

**Figure 1** The execution times of peanut house (a) and Nan OTOP (b) datasets.




(c) (d)

**Figure 2** A Comparison of Association Rules Generation.

On the other hand, the number of rules is considered. We found that both datasets generated an equal number of rules using two methods: the Apriori and FP-growth algorithms (Figure 2(c)). In the peanut house dataset, the support value on the x-axis ranges from 0.0015 to 0.0075, while the y-axis indicates the number of rules. The highest number of rules in this instance is 2,200, with the confidence value close to 0. This bar chart displays five confidence values ranging from 0 to 0.5. Figure 2(d) shows the number of rules generated from the Nan OTOP center dataset. At a support value of 0.0030, the minimum number of rules is generated. Although the dataset contains many products, the frequency of occurring transactions is very important for generating rules. Following our experiment, the generated rules will be stored in the cloud and selected based on conditions such as minimum support (MIN-SUP $\geq$ 0.001) and minimum confidence (MIN-CONF $\geq$ 0.3), as well as a permutation method for reducing the related products as the rules.

**A Background Process with API**

After gathering the rules, we implement a process using an application programming interface (API) for the Nan OTOP recommendation system. Our API is designed to receive new datasets and create new rules. Its functions can check the number of transactions, the user status, and the item recommendations in the TOP-10 (ranked by confidence value). We will use this API to prepare the Nan OTOP recommendation system in the future.

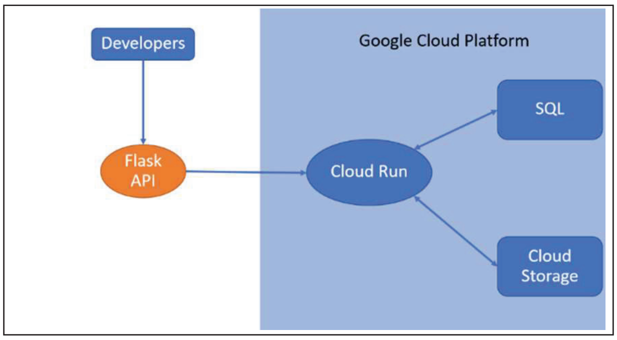



**Figure 3** Process of the API.

In Figure 3, the API runs on Cloud Run and has connections between SQL (DATABASE) and Cloud Storage within the Google Cloud Platform. Cloud Run operates similarly to a Docker container, with an image created from Cloud Build and then deployed to Cloud Run. The image is stored in the Container Registry. To connect Cloud Run and SQL (MySQL) within the Google Cloud Platform, a VPC (Virtual Private Cloud) network must be used, providing a private IP for SQL to serve as the host. After that, we created an application to call the API. An example of calling the API to request a list of product recommendations involves specifying the PRODUCT-NAME or PRODUCT-ID based on the data uploaded to create the rules. The number of rules received depends on the limit value set during the rule creation process.

**Nan OTOP's Recommendation Application**

In this work, we have developed the OTOP Market Basket application based on data mining principles, selecting appropriate rules using established mathematical methods, and identifying groups of products that reveal interesting relationships through association rule discovery. The software and hardware utilized in our implementation are as follows: The web-based application was developed using the PHP (Personal Home Page) programming language. The original examinations and statistical results were stored in a MySQL database, which was designed for the association rules. All programming was done using a Python editor. The maximum related products will be displayed in the Top-10 when we click on each product.

We can access the application from the main page, which features eleven categories of Nan OTOP products, such as local cloth, agricultural food, drinks, souvenirs, etc. Figure 4 illustrates the details of Nan OTOP's products on the left-hand side of the rules. In this case, the product is chewing mulberry. Each product entry includes a picture, name, price, category, access time, and the owner's contact information. When we click on an interesting product on the website, the detailed product information is shown on the right-hand side (Figure 4). The recommended products are displayed below on the same page (Figure 5).

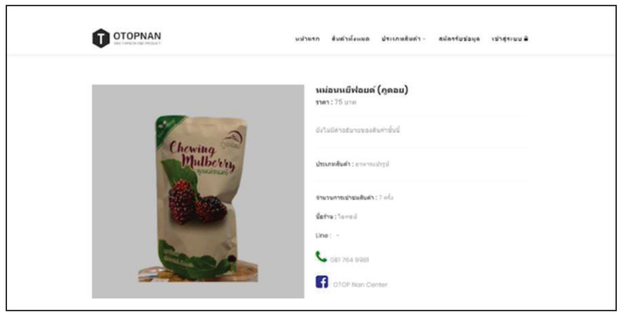



**Figure 4** Detailed representation of a Nan OTOP's product.

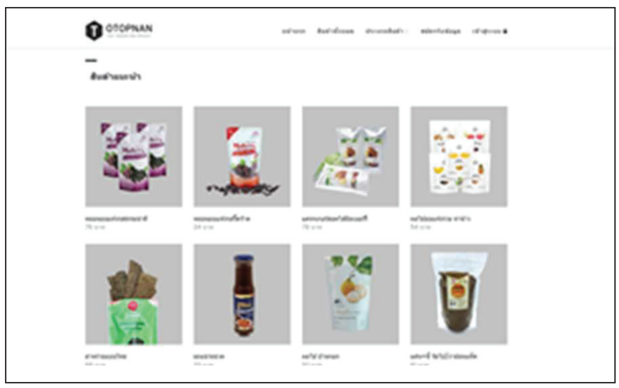



**Figure 5** Recommended products selected from association rules.

The relationships between the left-hand side products are shown in Figure 6. The most recommended products are mulberry agriproducts and dried fruits, reflecting the actual situation in OTOP shops. In our system, the related products are determined in the Top-10 based on the conditions of our experiment.

**User Satisfaction**

Nan OTOP's Recommendation Application was evaluated using 100 examples and 10 questions. A 1-to-5 satisfaction scale was employed to measure user satisfaction or dissatisfaction with functionality, performance, and usability. The users' responses are shown in Table 2.

The average overall user satisfaction is 4.72 (very satisfied), with a standard deviation of 0.59. Users reported that the website clearly displays information and provides recommended results based on interesting products. The website is designed similarly to e-commerce platforms, making it easy to access and use.

**Table 2** User Satisfaction with Nan OTOP's Recommendation Application.

| Satisfaction | $\bar{X}$ | S.D. | Scale |
|---|---|---|---|
| Display clearly | 4.8 | 0.40 | Very satisfied |
| Enough information | 4.2 | 0.75 | Satisfied |
| Completeness data | 4.6 | 0.80 | Very satisfied |
| Fast to retrieve data | 4.8 | 0.40 | Very satisfied |
| Easy to navigate | 4.91 | 0.29 | Very satisfied |
| Access to the recommended product | 4.60 | 0.49 | Very satisfied |
| Number of products available for shopping | 4.60 | 0.49 | Very satisfied |
| Saving time for shopping | 4.91 | 0.29 | Very satisfied |
| Web design: color, font, picture, etc. | 4.91 | 0.29 | Very satisfied |
| Works on various browsers such as Chrome, Safari, Firefox. | 4.89 | 0.32 | Very satisfied |
| **Overall** | **4.72** | **0.59** | **Very satisfied** |

Users who responded to the questionnaire suggested that adding more of the newest Nan OTOP products to the corpus would keep the related promotions up-to-date for entrepreneurs and assist customers in their decision-making. For example, if a tourist buys dried peanuts, the recommended products might include cookies made from peanut ingredients. In another case, if sun-dried wampee is selected, other dried fruit products would be suggested. Thus, the recommendation system would help entrepreneurs set promotions during festivals and support customers in making product decisions.

## Conclusion

In this work, we aimed to support the economy of Nan tourism by enhancing the Nan OTOP recommendation system. The challenge is that tourists often struggle to determine which of the hundreds of product items are the most popular for use in item packages. We utilized two algorithms for association rule learning: Apriori and FP-growth, to evaluate their performance. Association rule mining algorithms can be divided into three major categories: (1) frequent itemset mining, (2) sequential pattern mining, and (3) structured pattern mining.

In our comparison of the two algorithms, we evaluated their execution time and the number of rules generated using two large datasets from the peanut house and Nan OTOP center. Our investigation showed that the Apriori algorithm outperforms the FP-growth technique. During the rule generation experiment, both methods produced an equal number of rules. The recommendation system can utilize these associated rules to determine related products.

In future work, we will explore additional data mining techniques to improve performance and increase the number of association rules generated.

## Acknowledgment

This research was funded by the Research and Development Institute (RDI) of Rajamangala University of Technology Lanna, Thailand, under the Fundamental Fund Project Number 2565FFFP062. We would also like to thank the Artificial Intelligence Association of Thailand (AIAT) for providing academic resources.

## References

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.

Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207-216).

Al-Maolegi, M., & Arkok, B. (2014). *An improved Apriori algorithm for association rules. arXiv preprint arXiv:1403.3948*.

Fadillah, A. R., Yulita, I. N., Pradana, A., & Suryani, M. (2021, October). Data Mining Implementation Using Frequent Pattern Growth on Transaction Data for Determining Cross-selling and Up-selling (Case Study: Cascara Coffee). In *2021 International Conference on Artificial Intelligence and Big Data Analytics* (pp. 1-6). IEEE.

Fatoni, C. S., Utami, E., & Wibowo, F. W. (2018, December). Online store product recommendation system uses apriori method. In *Journal of Physics: Conference Series* (Vol. 1140, No. 1, p. 012034). IOP Publishing.

Goebel, V. (2014). *Knowledge Discovery in Databases (KDD) and Data Mining (DM)*. University of Oslo. Department of Informatics.

Hossain, M., Sattar, A. S., & Paul, M. K. (2019, December). Market basket analysis using apriori and FP growth algorithm. In *2019 22nd international conference on computer and information technology (ICCIT)* (pp. 1-6). IEEE.

Ketui, N., Wisomka, W., & Homjun, K. (2019). Association rule mining with permutation for estimating students' performance and its smart education system. *Journal of Computers, 30*(2), 93-102.

Liu, B., Hsu, W., & Ma, Y. (1998, August). Integrating classification and association rule mining. In *Proceedings of the fourth international conference on knowledge discovery and data mining* (pp. 80-86).

Li, H., Wang, Y., Zhang, D., Zhang, M., & Chang, E. Y. (2008, October). Pfp: parallel fp-growth for query recommendation. In *Proceedings of the 2008 ACM conference on Recommender systems* (pp. 107-114).

Patil, M., & Patil, T. (2022). *Apriori Algorithm against Fp Growth Algorithm: A Comparative Study of Data Mining Algorithms*. Available at SSRN 4113695.

Singh, P. K., Sinha, M., Das, S., & Choudhury, P. (2020). Enhancing recommendation accuracy of item-based collaborative filtering using Bhattacharyya coefficient and most similar item. *Applied Intelligence*, *50*, 4708-4731.

Tareq, S. U., Noor, M. H., & Bepery, C. (2020). Framework of dynamic recommendation system for e-shopping. *International Journal of Information Technology*, *12*(1), 135-140.

Yuan, X. (2017, March). An improved Apriori algorithm for mining association rules. In *AIP conference proceedings* (Vol. 1820, No. 1). AIP Publishing.