Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Analysis of Beverage Sales Data Using the FP-Growth Algorithm at Sini Aja Cafe

Widisa Adi Kumara[1*], Rini Astuti[2], Willy Prihartono[3], Tati Suprapti[4]

[1,3,4]STMIK IKMI Cirebon
[2]STMIK LIKMI Bandung
Jl. Perjuangan N0.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131
widisaadik3@gmail.com[1*], riniastuti@likmi.ac.id[2], willy@ikmi.ac.id[3], tatisuprapti112004@gmail.com[4]

Abstract

The growth of information technology and data mining techniques has greatly helped analyze consumer purchasing behavior, particularly in marketing and inventory management. This study aims to uncover association patterns between products frequently bought by customers at Sini Aja Cafe and to measure these patterns' support and confidence values. The research uses Knowledge Discovery in Databases (KDD), including stages like data selection, preprocessing, transformation, applying the FP-Growth algorithm, and interpreting results. Data from 1,083 beverage sales transactions at Sini Aja Cafe from August 1 to October 31, 2024. The findings reveal five significant association rules when applying a minimum support of 0.1 (10%) and confidence of 0.3 (30%). Notably, if customers buy Red Velvet Oreo, there is a 56% chance they will also buy Thai Tea. Thai Tea sales dominate with a support value 0.557 (55.7%). The support values of the association rules range from 0.141, categorized as medium, and the confidence values range from 0.235, categorized as low. These findings offer valuable insights for the cafe owner to optimize operations, enhance customer satisfaction, and increase profits.

Keywords: *Sales Data, Association Rules, Frequent Itemset, FP-Growth, Knowledge Discovery in Databases (KDD)*

1. Introduction

Timely and precise decision-making is essential for formulating efficient company strategies in the digital age [1]. Data from every transaction documented in a company's database offers a more precise reference for decision-making than just reliance on intuition [2]. The progression of information technology has resulted in a growing volume of data production, especially within the retail sector, rendering the management of this extensive data a considerable difficulty [3]. Consequently, data mining methodologies, particularly association algorithms such as FP-Growth, are crucial for revealing transaction data trends that might enhance business decision-making [4].

The FP-Growth algorithm is one of the techniques in association data mining that is useful for finding frequent item sets that often appear in a dataset [5]. The selection of the FP-Growth algorithm is because this algorithm shows faster pattern formation compared to the Apriori algorithm in analyzing a dataset [6]. Additionally, the FP-Growth algorithm accurately identifies sold item patterns [7]. The FP-Growth algorithm uses Association Rules in Data Mining to process the sales dataset in its application [8]. This data processing uses the RapidMiner Studio tool version 10.1, which produces Support values to indicate how often certain items appear in the overall transaction data. Confidence is used to measure specific relationship patterns between items with the highest occurrence frequency.

The results of this research are expected to provide valuable insights for the owner of Sini Aja Cafe. By understanding customer purchasing patterns, the shop owner can design more effective sales strategies, improve operational efficiency, and create a more satisfying shopping experience for customers. These findings also support optimizing product placement, enhancing promotions for beverages that are frequently purchased together, and the development of bundling strategies or special discounts according to purchasing patterns. Moreover, this research is an important foundation for further development in sales data analysis and data mining applications in the retail business. These findings can enhance business competitiveness, provide a competitive edge, and assist shop owners in adjusting marketing and operational strategies to increase loyalty and attract new customers. Insights into purchasing patterns can be utilized to tailor more relevant offerings for customers, create more targeted promotional strategies, and optimize product stock according to market needs. Thus, the results of this research not only support business development but also make a significant contribution to the field of sales data analysis and open up opportunities for further research.

2. Research Methods

2.1. Data Mining

Data mining is the process of identifying patterns, correlations, and relevant information in huge databases by using statistical, mathematical, and computer tools. It entails using algorithms to uncover hidden correlations in data that may subsequently be used to guide decision-making, forecasting, and optimization. Classification, clustering, regression, and association rule mining are all important data mining techniques [9]. The objective is to convert raw data into actionable insights that improve corporate strategy and prediction accuracy.

2.2. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) is extracting useful knowledge from large datasets through steps. These include data collection, preprocessing (cleaning and preparing the data), transformation (converting data into a suitable format), data mining (applying algorithms to find patterns), and interpretation and evaluation (analyzing and validating the results). KDD is widely used in various fields, such as business and healthcare, to uncover insights that aid decision-making, predictions, and optimization [10]. The stages of KDD used in this research are shown below in Figure 1, and they include the following steps:

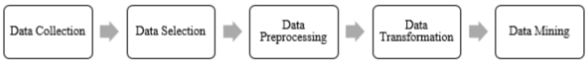

Figure 1: Knowledge Discovery in Databases (KDD)

2.3. FP-Growth Algorithm

FP-Growth (Frequent Pattern Growth) is an efficient algorithm used in association rule mining to identify frequent itemsets in large datasets. In contrast to the Apriori technique, FP-Growth circumvents the generation of candidate itemsets by creating a compact data structure known as the FP-tree, which retains item frequency data [11]. The algorithm then recursively mines frequent itemsets by dividing the dataset into smaller conditional databases, reducing computational complexity and memory usage. In association rule mining, association rules describe relationships between items, These rules are evaluated using key metrics: support, which is calculated as :

$$Support(X) = \frac{Number\ of\ transactions\ containing\ itemset\ X}{Total\ number\ of\ transactions}$$

Support shows the frequency of occurrence of an item or combination of items in the dataset and confidence, which is calculated as :

$$= Confidence(A \rightarrow B) = \frac{Number\ of\ transactions\ containing\ both\ A \text{ and } B}{Number\ of\ transactions\ containing\ A}$$

Confidence measures how strong the relationship is between two items that often appear together in transactions.

3. Implementation

3.1. Data Collection

The collection of beverage sales data for Sini Aja Cafe was obtained from the Qasir application at Sini Aja Cafe, covering beverage sales transactions over a three-month period, from August 2024 to October 2024. There are 1083 transactions including transaction numbers, transaction dates, product names, and the prices of each product. The data used in this study is secondary data. Table 1 below shows the sales dataset:

Table 1: Sales Dataset

Transaction Number	Transaction Date	Product Name	Price
TRX1	2024-08-01	Mochachino	Rp.10000
TRX1	2024-08-01	Berry Coffe Latte	Rp.12000
TRX1	2024-08-01	Chocholatte Oreo	Rp.12000
TRX1	2024-08-01	Thai Tea	Rp.8000
....	……..	………	……..
TRX1083	2024-10-31	Red Velvet Oreo	Rp.12000

3.2. Data Selection

This stage begins with importing the Sini Aja Cafe Sales Dataset into the Rapidminer Studio application, importing the dataset into the Local Repository, and then adding the Retrieve operator. Next, open the dataset using the Retrieve operator as shown in the Figure 2 below :

Figure 2: Retrieve Operator in Rapidminer

The next step is to convert the initial dataset into a format that can be processed by the FP-Growth algorithm using the Pivot operator. The function of this operator is to change the long table data format (where each row represents an individual item in a transaction) into a wide table format, where each row represents a complete transaction along with the list of products purchased together. Figure 3 shows the Pivot operator:

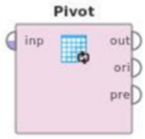

Figure 3: Pivot Operator in Rapidminer

Below are the details of the Parameters table from the pivot operator used, as shown in Table 2:

Table 2: Pivot Parameter

No	Parameter	Value
1	Group by Attributes	Transaction Number
2	Column Grouping Attribute	Transaction Date
3	Aggregation Attributes	
	Aggregation Attribute	Product Name
	Aggregation Function	Count

3.3. Data Preprocessing

In this preprocessing stage, the initial step taken is to assign roles to the dataset, which aims to ensure that each attribute has a suitable function for analysis. The use of the Set Role operator is useful for preprocessing and determining the role of the beverage sales data at Sini Aja Cafe. The following Figure 4 is the Set Role operator.

Figure 4: Set Role Operator in Rapidminer

Next, the process of assigning the Transaction Number attribute as the transaction ID using the Set Role operator. Details of the Set Role operator parameters can be seen in Table 4.2 below:

Table 3: Set Role Parameter

No	Parameter	Target Role
1	Attribute Name	Transaction Number
2	Target Role	ID

After the assignment of the transaction ID above, the data contains missing values. To address this issue, the application of the Missing Value Replacement operator is necessary. The function of this operator is to replace missing values with zero (0). The operator Replace missing value operator is shown in the Figure 5 below:

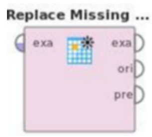

Figure 5: Replace Missing Value Operator in Rapidminer

The Replace Missing Values operator used has parameters as shown in Table 4 below:

Table 4: Replace Missing Value Parameter

No	Parameter	Value
1	Attribute filter type	all
2	Default	Zero
3	Columns	

The result of processing the Replace Missing Value operator can be seen in Table 5 below:

Table 5: Result of Processing the Replace Missing Value operator

Transaction Number	count(Product Name)_Anggur Yakult	…….	count(Product Name)_Vanilla Black Cookies
TRX1	0	…	0
TRX2	0	…	0

TRX3	0	…	0
TRX4	0	…	0
TRX5	0	…	0
TRX6	0	…	0
TRX7	0	…	0
TRX8	1	…	0
TRX9	0	…	0
TRX10	0	…	0
…….	…….	…	…….
TRX1083	0	……	0

3.4. Data Transformation

At this stage, it is necessary to convert the data from Numeric to Binominal so that all data can be adjusted to the FP-Growth algorithm format. Figure 6 shows the appearance of the Numerical to Binominal operator:

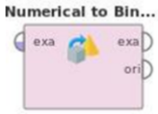

Figure 6: Numerical to Binomical Operator in Rapidminer

Details of the Numerical to Binominal operator parameters are in Table 6 below:

Table 6: Numerical to Binominal Parameter

No	Parameter	Value
1	Attribute Filter Type	All

Based on Figure 6 above, the Product Name attribute must be corrected to remove irrelevant words such as count (Product Name). The removal of irrelevant words was done on 28 Product Name attributes using the Rename operator, as shown in Figure 7:

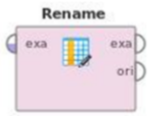

Figure 7: Operator Rename in Rapidminer

The details of the Rename operator parameters can be seen in Table 7 below:

Table 7: Rename Parameter

No	Parameter	Value
1	Rename Attributes	
2	OldNames	count(Product Name)_Anggur Yakult count(Product Name)_Berry Coffee Latte count(Product Name)_Blackcurren Tea count(Product Name)_Choco Hazelnut count(Product Name)_Chocolatte Coffee count(Product Name)_Chocolatte Milk count(Product Name)_Chocolatte Oreo count(Product Name)_Chocolatte Puding count(Product Name)_Chocolatte Thai Tea count(Product Name)_Coffee Milo count(Product Name)_Creamy Manggo count(Product Name)_Green Tea count(Product Name)_Kopi Susu count(Product Name)_Kopi Susu Gula Aren count(Product Name)_Lechie Yakult

		count(Product Name)_Lemon Tea
		count(Product Name)_Manggo Jelly
		count(Product Name)_Manggo Yakult
		count(Product Name)_Matcha Latte
		count(Product Name)_Milo Dino
		count(Product Name)_Moccachino
		count(Nama Pr Product Name oduk)_Oreo Black Coockies
		count(Product Name)_Red Velvet
		count(Product Name)_Red Velvet Oreo
		count(Product Name)_Strawberry Yakult
		count(Product Name)_Taro
		count(Product Name)_Thai Tea
		count(Product Name)_Vanilla Black Cookies
3	NewNames	Anggur Yakult
		Berry Coffee Latte
		Blackcurren Tea
		Choco Hazelnut
		Chocolatte Coffee
		Chocolatte Milk
		Chocolatte Oreo
		Chocolatte Puding
		Chocolatte Thai Tea
		Coffee Milo
		Creamy Manggo
		Green Tea
		Kopi Susu
		Kopi Susu Gula Aren
		Lechie Yakult
		Lemon Tea
		Manggo Jelly
		Manggo Yakult
		Matcha Latte
		Milo Dino
		Moccachino
		Oreo Black Coockies
		Red Velvet
		Red Velvet Oreo
		Strawberry Yakult
		Taro
		Thai Tea
		Vanilla Black Cookies

3.5. Data Mining

After the pre-processing stage is complete, the next step is the Data Mining process on the dataset to find associations between products based on the minimum support value. This process uses the FP-Growth operator. Figure 8 shows the FP-Growth operator interface.

Figure 8: FP-Growth Operator in Rapidminer

In the data mining process using the FP-Growth operator, a selection is made with a minimum support of 0.1, as shown in the following parameter Table 8:

Table 8: FP-Growth Parameter

No	Parameter	Value
1	Input format	Items in dummy coded columns

2	Positive value	
3	Min requirement	Support
4	Min support	0.1
5	Min items per itemsets	1
6	Max items per itemsets	0

Here are the results of the Frequent Item Sets from the above process, displayed in Table 9 below:

Table 9: Result Frequent item sets

Size	Support	Item 1	Item 2
1	0.557	Thai Tea	
1	0.288	Red Velvet Oreo	
1	0.094	Kopi Susu Gula Aren	
1	0.091	Lemon Tea	
1	0.083	Green Tea	
1	0.081	Strawberry Yakult	
1	0.077	Manggo Yakult	
1	0.076	Lechie Yakult	
1	0.071	Blackcurren Tea	
1	0.070	Anggur Yakult	
1	0.068	Kopi Susu	
1	0.067	Creamy Manggo	
1	0.066	Moccachino	
1	0.061	Manggo Jelly	
1	0.053	Chocolatte Thai Tea	
1	0.048	Milo Dino	
1	0.047	Berry Coffee Latte	
1	0.047	Choco Hazelnut	
1	0.045	Chocolatte Coffee	
1	0.043	Coffee Milo	
1	0.043	Matcha Latte	
1	0.041	Oreo Black Coockies	
1	0.039	Chocolatte Puding	
1	0.038	Vanilla Black Cookies	
1	0.033	Chocolatte Milk	
1	0.033	Chocolatte Oreo	
1	0.031	Red Velvet	
1	0.030	Taro	
2	0.163	Thai Tea	Red Velvet Oreo
2	0.027	Thai Tea	Kopi Susu Gula Aren
2	0.026	Thai Tea	Lemon Tea
2	0.025	Thai Tea	Green Tea
2	0.024	Thai Tea	Strawberry Yakult
2	0.028	Thai Tea	Manggo Yakult
2	0.031	Thai Tea	Lechie Yakult
2	0.022	Thai Tea	Blackcurren Tea
2	0.027	Thai Tea	Anggur Yakult
2	0.022	Thai Tea	Creamy Manggo

The next important step is finding the dataset's confidence value if the support value has been found. The Create Association Rule Operator is chosen to perform this task because this operator generates rules in the form of IF-THEN, which describes the relationships between products in transactions. The display of the Create Association Rule operator is shown in Figure 9 :

Figure 9: Create Association Rules Operator in Rapidminer

The parameters of the Create Association Rules operator can be seen in Table 10 below:

Table 10: Create Association Rules Parameter

No	Parameter	Value
1	Criterion	Confidence
2	Min Confidence	0.3
3	Gain Tetha	2.0
4	Laplace K	1.0

The result of the Create Association Rules operator produces five association rules between products along with their support and confidence values, as shown in Table 11 below:

Table 11: Association Rules Results

No	Premises	Conclusion	Support	Confidence
1	Red Velvet Oreo	Thai Tea	0.163	0.564
2	Lechie Yakult	Thai Tea	0.031	0.415
3	Anggur Yakult	Thai Tea	0.027	0.382
4	Mango Yakult	Thai Tea	0.028	0.361
5	Creamy Manggo	Thai Tea	0.022	0.329

Figure 10 below displays the design of the entire research process:

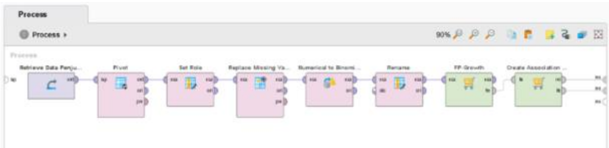

Figure 10: Design of the entire research process

This study's data consists of 1,083 sales transaction data obtained from the Qasir Point of Sale (POS) application at Sini Aja Cafe over three months from August 2024 to October 2024. This data undergoes several processing stages before being used for association analysis.

At the Data Selection stage, the imported dataset from the file Sini Aja Cafe Dataset is entered into the RapidMiner Studio application. This dataset is transformed using the Pivot operator, which changes the format from a long table to a wide one so that each row represents a complete transaction and the list of products purchased together. This process results in 28 product attributes, which are then further processed with the Set Role operator to make the Transaction Number attribute the transaction ID.

At this preprocessing stage, missing values in the transaction data are handled using the Replace Missing Values operator, which replaces missing values with zero. (0). Next, the Numerical to Binominal operator converts numerical data into binominal format to align with the FP-Growth algorithm. In the Data Mining stage, the FP-Growth algorithm finds association patterns between products by setting a minimum support of 0.1. The support value is calculated using the formula Support (X) = (Number of Transactions Containing Itemset X) / (Total Number of Transactions), which measures how often the product combination appears in all transactions. In this study, the Thai Tea and Red Velvet Oreo item has a support value of 0.163, which means this combination appears in 16.3% of the total transactions analyzed, indicating the significance of this association pattern.

At the final stage, the association analysis uses a minimum confidence value of 0.5. Confidence is calculated using the formula confidence (A → B) = Number of transactions containing both A and B / Number of transactions containing A, where A → B is the rule that connects item A to item B. This confidence association analysis is conducted using the Create Association Rule operator. These association rules describe the relationships between products as IF-THEN rules. The analysis results show that there are several association rules with varying confidence levels, such as:

- If consumers buy Red Velvet Oreo, there is a 56.4% chance that they will also buy Thai Tea (confidence: 0.564), with an item occurrence rate of 16.3%.
- If consumers buy Lechie Yakult, there is a 41.5% chance that they will also buy Thai Tea (confidence: 0.415), with an item occurrence rate of 3.1%.
- If consumers buy Anggur Yakult, there is a 38.2% chance that they will also buy Thai Tea (confidence: 0.382), with an item occurrence rate of 2.7%.
- If consumers buy Anggur Yakult, there is a 36.1% chance that they will also buy Thai Tea (confidence: 0.361), with an item occurrence rate of 2.8%.
- If consumers buy Anggur Yakult, there is a 32.9% chance that they will also buy Thai Tea (confidence: 0.382), with an item occurrence rate of 2.2%.

Thai tea is the main product, often combined with various other products. Products such as Red Velvet Oreo, Lechie Yakult, and Anggur Yakult have a significant relationship with Thai Tea. Thus, they can be categorized as complementary products. Calculations were made for the range of support and confidence values for each of the five association patterns found. The analysis results show that the support values are in the low range, while the confidence values are in the medium range. Association patterns with low support values are usually considered less significant [12], indicating that the relationship between these products is not strong enough. This may be due to the limited data collection period, which was only three months.

4. Conclusion

In Based on the research conducted, several conclusions were drawn as follows:

1. Using minimum support of 0.1 and minimum confidence of 0.3, successfully generated five association rules in the beverage sales data of Sini Aja Cafe. However, the associations found tend to be weak, as seen from the support category, which falls within the range of 0.062 to 0.112, and the relatively low confidence value, which is less than 40%. The five identified association rules are as follows: the best rule is Red Velvet Oreo → Thai Tea with a support value of 0.163 and confidence of 0.564, followed

by the second rule, Lechie Yakult → Thai Tea with a support value of 0.031 and confidence of 0.415, then the third rule, Anggur Yakult → Thai Tea with a support value of 0.027 and confidence of 0.382, followed by the fourth rule, Mango Yakult → Thai Tea with a support value of 0.028 and confidence of 0.361, and the last rule is Creamy Mango → Thai Tea with a support value of 0.022 and confidence of 0.329.

2. The highest support value is found in Thai Tea and Red Velvet Oreo products, indicating that the combination of these two products appears at 0.163 or 16.3% of the total transactions conducted. Meanwhile, the highest confidence value of 0.564 in the rule Red Velvet Oreo > Thai Tea indicates that when consumers buy Red Velvet Oreo, there is a 56.4% chance they will also buy Thai Tea. Thus, the relationship between Thai Tea and Red Velvet Oreo shows a significant purchasing pattern in the sales data, with Red Velvet Oreo having a strong tendency to encourage the purchase of Thai Tea.
3. Thai Tea is the most frequently purchased beverage product by consumers, with the highest support value of 0.557, indicating that this product appears in 55.7% of total transactions. Besides Thai Tea, accompanying products such as Red Velvet Oreo and Lechie Yakult show a significant correlation, indicating that consumers purchase various products in a single transaction.

References

[1] K. Munthe, T. Syahputra, and R. Mahyuni, “Analisa Algoritma FP-Growth Pada Pola Pemesanan Makanan Dan Minuman Konsumen,” vol. 3, no. September, pp. 677–684, 2024.

[2] R. L. Najmi, M. Irsyad, F. Insani, and A. Nazir, “Analisis Pola Asosiasi Data Transaksi Penjualan Minuman Menggunakan Algoritma FP-Growth dan Eclat,” vol. 5, no. 1, pp. 0–7, 2023, doi: 10.47065/bits.v5i1.3592.

[3] F. Achmad *et al.*, “ANALISA POLA TRANSAKSI PEMBELIAN KONSUMEN PADA TOKO RITEL KESEHATAN MENGGUNAKAN ALGORITMA FP-GROWTH,” vol. 7, no. 1, pp. 168–175, 2023.

[4] A. Sofialaela, N. Rahaningsih, and I. Ali, “OPTIMASI POLA PENJUALAN DI TOKO KELONTONG SUMBER REJEKI BANDUNG MELALUI ANALISIS ASOSIASI DENGAN ALGORITMA FP-GROWTH,” vol. 8, no. 2, pp. 1916–1920, 2024.

[5] A. R. Wibowo and A. Jananto, “Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan Ritel,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 200, 2020, doi: 10.35585/inspir.v10i2.2585.

[6] Y. Syach Putra, R. Kurniawan, and Y. Arie Wijaya, “Penerapan Data Mining Menggunakan Algoritma Fp-Growth Pada Data Penjualan Sembako,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 561–567, 2024, doi: 10.36040/jati.v8i1.8391.

[7] H. Maulidiya and A. Jananto, “Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako,” *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.

[8] K. T. Wijaya and I. Pratama, “Penerapan Algoritma FP-Growth Untuk Analisis Data Transaksi Penjualan Di Internet Learning Cafe Kaliurang,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 4, pp. 642–651, 2022, doi: 10.32672/jnkti.v5i4.4585.

[9] D. D. Darmansah and N. W. Wardani, “Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 105–117, 2021, doi: 10.35957/jatisi.v8i1.590.

[10] M. Defriani and I. J. Jaelani, “Recognition of Regional Traditional House in Indonesia Using Convolutional Neural Network (CNN) Method,” *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, pp. 104–115, 2022, doi: 10.47709/cnahpc.v4i2.1562.

[11] P. Aufia, I. Lestari, and N. Cahyani, “Journal of Computer Networks , Architecture and High Performance Computing Application Of The Association Rule Method Based On Book Borrowing Patterns In Bojonegoro Regional Libraries Journal of Computer Networks , Architecture and High Performance Comput,” vol. 5, no. 2, pp. 751–759, 2023.

[12] B. Anwar, P. T. Kejuruan, F. Teknik, and U. N. Padang, “Application of the FP-Growth Method to Determine Drug Sales Patterns,” vol. 8, no. 1, pp. 405–414, 2023.