A Hybrid Approach Using K-Means and Apriori Algorithms for Marketing Recommendation Systems in the Café Industry

Kezia Elice Yulianto[1], Indra Maryati[2*]

[1,2]Information Systems, School of Information Technology, Universitas Ciputra, Surabaya, East Java, Indonesia
E-mail: [1]kyulianto@student.ciputra.ac.id, [2*]indra.maryati@ciputra.ac.id

(Received: 10 Feb 2026, revised: 1 Mar 2026, accepted: 3 Mar 2026)

Abstract

Many food and beverage (F&B) businesses struggle to survive because of a lack of understanding of their market and customer behavior. Therefore, this study aims to analyze transaction data using K-Means Clustering and Apriori algorithms to identify customer purchasing patterns and provide actionable insights for a café. This research applies the CRISP-DM method, including six primary phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. K-Means reveals that products can be grouped into three main clusters, namely the low-demand products, premium products, and best-selling products. Furthermore, the association rule mining successfully indicates six meaningful correlations among products, with the strongest lift value of 5.22 was found between Piccolo Latte and Extra Oatmilk. These findings are transformed into actionable insights, such as recommendations for implementing bundling and up-selling. An interview with the business owner validates that the recommendations are considered applicable to the current business operations. This proves that the integration of K-Means and Apriori algorithms can be used to design effective data-driven marketing strategies for an F&B business.

Keywords: Apriori, CRISP-DM, F&B Transaction, K-Means Clustering, Marketing Strategy.

I. INTRODUCTION

The food and beverage (F&B) industry is one of the sectors that continues to grow and makes a significant contribution to the Indonesian economy. Increased income and changing lifestyles have led to a significant rise in demand for food and drink services [1]. According to data from the 2024 Food and Beverage Services Survey, the number of food and beverage services in Indonesia reached 4.85 million in 2023, with restaurants and eateries as the second largest group of businesses in the sector. However, the data shows that fewer than 35% of businesses survive for more than 10 years.

F&B businesses face various challenges, ranging from intense competition [2] to a rapidly changing and unpredictable business environment [3]. This makes it essential for business owners to be adaptable. One of the key indicators of this ability is understanding the market and clients and determining the right marketing strategy. Based on previous research [4], [5], several marketing strategies have been proven to be effective in increasing sales, such as bundling, up-selling, cross-selling, and offering discounts.

In the F&B industry, the integration of digital technology, including customer data analysis, can help businesses understand their customer and respond to changing trends [6]. Therefore, this study aims to provide actionable insights based on an analysis of the transaction data of an F&B business. The study combines two main algorithms: K-Means Clustering to group products and Apriori to identify frequently purchased menu combinations. This hybrid approach is implemented to ensure that the resulting insights are more actionable and deeper compared to a global analysis using each method in isolation, as proven in previous research [7].

K-Means Clustering is an algorithm that divides N-dimensional data into k sets (clusters) based on samples by reducing the spread of data within clusters [8]. In other words, K-Means Clustering is an algorithm that groups data into k groups based on their similarities, where the K-Means value is the average value of the represented group. In the industry, K-Means is often used for segmentation, such as customer segmentation and product segmentation, to decide which marketing strategy fits the needs [9].

Apriori is an association rule mining method that works by identifying frequent purchase patterns. This algorithm will first search for frequent itemset candidates that meet the minimum support value, then it will result in $X \rightarrow Y$ rules to show the correlation between products [10]. In the F&B industry, Apriori often helps in promotion strategy decisions, such as

bundling, up-selling, cross-selling, and product placement [11].

II. RESEARCH METHODOLOGY

This research applies the Cross-Industry Standard Process for Data Mining (CRISP-DM) method with K-Means Clustering and Apriori algorithms to analyze the transaction data of an F&B business (Figure 1). CRISP-DM consists of six correlated main phases, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment [12].

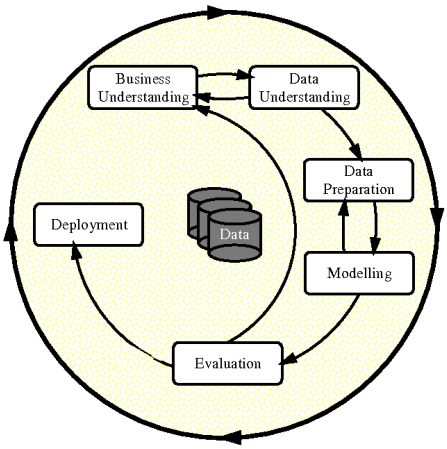


Figure 1. CRISP-DM Processes [12]

A. Business Understanding

The first phase of the research is business understanding, whose main purpose is to understand business conditions [13]. Thus, business needs and problems can be identified and used as the basis for the research to be conducted.

B. Data Understanding

The data understanding phase starts with data collection, followed by a process to understand data characteristics, identify data quality issues, find initial insights, and detect interesting data subsets to form hypotheses about hidden information [12]. This research also conducts an Exploratory Data Analysis (EDA) to explore data patterns, data distribution, and relationships between variables. This phase is closely related to the business understanding phase because problem formulation in data mining and project design requires an understanding of the existing data.

C. Data Preparation

Data preparation consists of feature selection, data integration, data cleaning, and data transformation. The feature selection process focuses on eliminating variables that do not represent product purchasing patterns. Meanwhile, the data integration process involves combining two primary datasets: transactional data and product data. This process is then followed by data cleaning to handle missing values and writing inconsistencies by removing unnecessary characters, such as new lines and excess white spaces. For data transformation, a standardization is conducted to minimize scale differences between variables, so that the clustering process is not dominated by any variable [14]. Finally, the data is transformed according to the required format of each algorithm, an aggregated format for the K-Means Clustering, and a basket format for the Apriori algorithm.

D. Modeling

At this stage, Python is used to create models with two main algorithms, namely K-Means and Apriori. The integration of K-Means and Apriori is to provide more varied association rules for each cluster that can optimize promotional strategies and improve competitiveness [7]. Firstly, K-Means groups data into several clusters according to the optimal number of clusters based on the elbow method. The purpose of the elbow method is to determine the best cluster by analyzing the percentage results of the number of clusters comparison in the form of elbow points [15]. To represent the total squared distance between each data point and the centroid of its assigned cluster, the within-cluster sum of squares (WCSS) of each value of K is calculated using Formula 1 [16]:

$$WCSS = \sum_{i=1}^{k} \sum_{x_j \in C_i} (x_j - \mu_i)^2 \quad (1)$$

where:

- C_i represents the i^{th} cluster;
- x_j is the data point within cluster C_i;
- μ_i denotes the centroid of cluster C_i;
- k represents the total number of clusters.

Afterwards, Apriori is used to find association rules in each cluster by testing the optimal minimum support. This value determines item combinations occurred frequently in the dataset, calculated using Formula 2 [17]:

$$Support(A) = \frac{Number\ of\ transactions\ containing\ A}{Total\ number\ of\ transactions} \times 100 \quad (2)$$

Only itemsets that meet the minimum support threshold are kept as frequent itemsets, but those failing to meet the criteria are pruned.

E. Evaluation

To measure the similarity of product characteristics within each cluster and the differences between clusters, the K-Means model was evaluated using the silhouette score. According to research [18], the silhouette score is a metric that shows the data similarity within groups, where the group with a silhouette score value closer to 1 has a higher clustering quality. The calculation of the silhouette score value is done based on Formula 3:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

DOI: 10.34148/teknika.v15i1.1461
CC BY SA

where:

- $a(i)$ is the average distance to other points in the same cluster;
- $b(i)$ is the average distance to points in the next nearest cluster.

In the Apriori model, evaluation based on the lift value was done, where the lift value is a value that shows the strength of the produced association rules. This value is calculated using Formula 4 [17]:

$$Lift(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A) \times Support(B)} \quad (4)$$

Rules with the lift value above 1 are considered strong because they show that the products appear together more often than they would appear randomly [19].

F. Deployment

The deployment phase is the final phase, which focuses on converting the results from the K-Means and Apriori models into actionable insights, such as the planning of marketing strategies in the form of bundling and up-selling. These strategies are chosen based on research results that prove the high customer acceptance of product bundling and up-selling, even when they did not offer a price discount [5], [20]. To determine whether the recommendations are possible to be implemented, a follow-up evaluation is conducted through an interview with the business owner. This interview measures the suitability between the recommendations and the operational condition, as well as the potential and readiness of the business to implement the recommended strategies.

III. RESULTS AND DISCUSSION

This research generated knowledge in the form of a list of the most frequently sold menu items and combinations for the three clusters formed based on sales patterns.

A. Business Understanding

Currently, the subject of this study already uses digital technology, namely the Majoo Point of Sales (POS) system. This system is used to help record sales transactions, inventory management, and finance management. However, in terms of sales, the provided information is limited only to the form of a sales report for a certain period. Therefore, this study focuses on understanding sales patterns so that useful insights can be generated for future business development.

B. Data Understanding

In this research, 6 months of sales transaction data from a café is used, which was obtained from a POS system. A total of 5,973 transaction data consists of some variables that represent transaction information. To supplement this, product data from the same POS was also used, containing a total of 276 unique product items. A complete description of the variables from both datasets is presented in Table 1.

Table 1. Variable Description

No	Variable	Description	Source
1	Transaction ID	Unique identifier for each sales transaction	Transaction Data
2	Order Time	Timestamp recording when the customer's order was entered into the POS	
3	Payment Time	Timestamp indicating completion of payment process	
4	Outlet	Name of the specific branch or location where the transaction occurred	
5	Product	Purchased product(s) in a single transaction ID	
6	Order Type	Category of how an order is processed (dine-in, takeaway, other)	
7	Remaining Bills	Unpaid portion of the total bill after the transaction is complete	
8	Total Sales	Total transaction amount in a single transaction ID	
9	Payment Method	Payment method used by the customer to settle the bill	
10	Payment	User account responsible for handling the payment process in POS	
11	Order	User account responsible for recording customer's order in POS	
12	Product Name	The name of the product in the database	Product Data
13	Product Category	The classification of the product (e.g., Dessert & Bakery, Coffee, Side Dishes)	
14	Price	The price for a single item of the product	

To visualize the raw format of the variables defined in Table 1, Table 2 shows a sample of the initial transaction records, while Table 3 provides a sample of the product data before any preprocessing processes were applied.

The Exploratory Data Analysis (EDA) indicates that there are no missing values within the dataset, but the observation results show some redundant variables that do not directly contribute to determining client purchasing patterns, such as Outlet, Payment, and Order. Inconsistencies in the writing of product names were found in both datasets, such as variations in white spaces and new lines. In addition, some items recorded in the sales transaction data could not be matched with the existing product data, resulting in multiple rows with missing values, requiring cleaning to ensure analytical consistency. Furthermore, the best-selling items were found to contribute less than 10% of total transactions, indicating a long-tailed distribution where customer preferences are widely scattered. This shows the need to use a lower minimum support value in the Apriori model to avoid missing important patterns.

Table 2. Sample of Raw Transaction Data

Trans. ID	Order Time	Payment Time	Outlet	Product	Order Type	Rem. Bills	Total Sales	Payment Method	Payment	Order
CS/34/250601/0001	01-06-2025 10:34:58	01-06-2025 11:23:25	Cafe dAurelie IFI Thamrin	Chicken Tajine,Biscotti JAR	Lainnya	Rp0.00	Rp229,910.00	Kartu Debit/Kredit, EDC BCA	Cafe dAurelie	Cafe dAurelie
CS/34/250601/0002	01-06-2025 12:16:03	01-06-2025 12:18:10	Cafe dAurelie IFI Thamrin	Croque - Madame,Chocolate`Mousse,Hot Cappucino	Lainnya	Rp0.00	Rp192,288.00	Kartu Debit/Kredit, EDC BCA	Cafe dAurelie	Cafe dAurelie
CS/34/250601/0003	01-06-2025 13:27:13	01-06-2025 13:37:04	Cafe dAurelie IFI Thamrin	Ice Matcha Latte,Ice Cappucino	Lainnya	Rp0.00	Rp98,234.00	Kartu Debit/Kredit, EDC BCA	Cafe dAurelie	Cafe dAurelie
CS/34/250601/0004	01-06-2025 13:38:34	01-06-2025 13:38:47	Cafe dAurelie IFI Thamrin	Avocado Toast	Lainnya	Rp0.00	Rp103,459.00	Kartu Debit/Kredit, EDC BCA	Cafe dAurelie	Cafe dAurelie
CS/34/250602/0001	02-06-2025 07:59:34	02-06-2025 08:05:47	Cafe dAurelie IFI Thamrin	Ice Americano	Lainnya	Rp0.00	Rp32,919.00	Kartu Debit/Kredit, EDC BCA	Cafe dAurelie	Cafe dAurelie

Table 3. Sample of Raw Product Data

No	Product Name	Product Category	Price
1	Afternoon Hot Cappucino	NEW PROMO	59,000
2	Afternoon HoT Matcha	NEW PROMO	59,000
3	AHMAD TEA	TEA	40,000
4	all breakfast bundling	Bundling Breakfast	99,000
5	Almond Croisant New	Dessert & Bakery	35,000

C. Data Preparation

To transform the raw data into a structured format suitable for the hybrid model, this stage involves the following steps:

1. Feature Selection and Data Integration
   The selection process began with removing columns that did not contribute to purchasing patterns, leaving three key variables, namely Transaction ID, Order Time, and Product Name. The transaction data was then combined with the product data using a left join method based on the product name. This process was executed to enrich the transaction records with the product prices, which are essential for the K-Means Clustering. During this phase, a feature engineering process was done by breaking down the Order Time column into three features (Order Month, Order Day, and Order Hour) to determine its impact on customer purchase patterns. However, the analysis shows that the peak activity occurred during the day and was almost always consistent every day, indicating no significant patterns that would add value to the segmentation. Therefore, these time features were excluded from the final dataset to keep the model focused on purchasing patterns and product associations.
2. Data Cleaning
   Following the data integration process, some rows of data were found to have missing values in the price column. These null values were caused by product naming inconsistencies between the two datasets or by products that did not exist in the product data. To address this problem, excess new lines and white spaces were removed from the product names. After that, any records that remained null (containing products that were missing from the product data) were deleted. This was considered safe as there were only 102 transaction data records excluded, so it did not affect the analysis results significantly.
3. Data Transformation
   The final step of data preparation is to transform the cleaned data to meet the specific requirements of both the K-Means and Apriori algorithms. For clustering purposes, transaction data was aggregated based on the product name, resulting in three numerical variables for each product: frequency (the total number of transactions), total revenue (the cumulative sales value), and price (the unit price of the product). These features represent the sales characteristics of each product, which can be used as the basis for the clustering process. The sample of aggregated data can be seen in Table 4.

Table 4. Sample of Aggregated Data

No	Product Name	Freq.	Total Revenue	Price
1	AHMAD TEA	179	7,160,000	40,000
2	Afternoon HoT Matcha	5	295,000	59,000
3	Afternoon Hot Cappucino	3	177,000	59,000
4	Almond Croisant New	12	420,000	35,000
5	Almond Croisant	160	35000	35,000

To overcome the large variations in scales between variables, numerical data were standardized using

StandardScaler, preventing variables with larger numerical ranges from dominating the model. The results of this standardization are presented in Table 5.

Table 5. Sample of Standardized Data

No	Product Name	Freq.	Total Revenue	Price
1	AHMAD TEA	1.239489	0.778929	-0.503009
2	Afternoon HoT Matcha	-0.544602	-0.572665	-0.125686
3	Afternoon Hot Cappucino	-0.565108	-0.595897	-0.125686
4	Almond Croisant New	-0.472828	-0.548054	-0.602304
5	Almond Croisant	1.044674	0.471793	-0.602304

Meanwhile, for the Apriori model, the data was converted into a basket transaction format in the form of a binary matrix, where each column represents a product, and each row represents a transaction [11]. A value of 1 indicates that the item appears in a transaction, while a value of 0 indicates that the item was not purchased in a transaction. Table 6 below displays a sample of the transformed basket data.

Table 6. Sample of Basket Data

Trans. ID	AHMAD TEA	Afternoon HoT Matcha	Afternoon Hot Cappucino	…	Extra Oatmilk
CS/34/250601/0001	0	0	0	…	0
CS/34/250601/0002	0	0	0	…	0
CS/34/250601/0003	0	0	0	…	0
CS/34/250601/0004	0	0	0	…	0
CS/34/250602/0004	0	0	0	…	1
…	…	…	…	…	…
CS/34/251121/0026	0	0	0	…	1
CS/34/251121/0028	0	0	0	…	0
CS/34/251121/0029	0	0	1	…	0
CS/34/251121/0032	0	0	0	…	1
CS/34/251121/0033	0	0	0	…	0

D. Modeling

The modeling phase in this study implemented a hybrid approach, where K-Means Clustering was first implemented to create product segmentations, followed by the Apriori algorithm to identify purchase patterns through association rules. This hybrid approach was chosen to provide more specific and actionable insights about the purchasing behaviors.

To determine the optimal number of clusters, the elbow method was applied to both the unscaled and scaled data. As shown in Figure 2, the WCSS values of the unscaled data were extremely high, reaching a value of 10^{15}. This indicates that the calculations were potentially dominated by a variable that has a wide range of values, so the model is less sensitive to smaller features. In comparison, Figure 3 illustrates the WCSS values of the scaled data, where each feature contributes equally to the distance calculation.

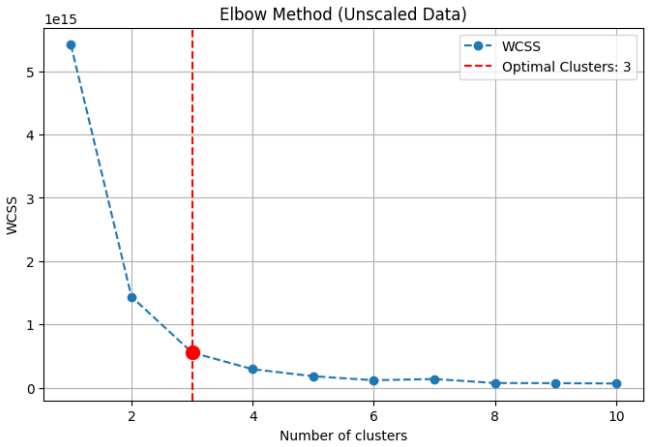


Figure 2. Elbow Curve of Unscaled Data

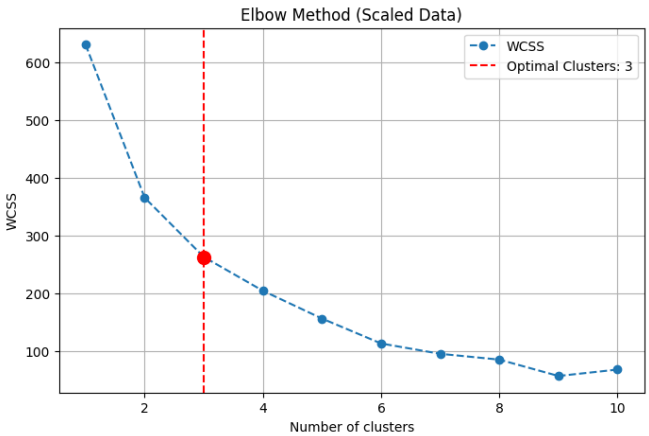


Figure 3. Elbow Curve of Scaled Data

As the K value grows from 1 to 3, there is a significant reduction in the WCSS value. Once the value hits three clusters, the average reduction value becomes smaller. This leads to the selection of the K = 3 value as the optimal number of clusters for the K-Means model to group the products into each cluster, as can be seen in Table 7.

DOI: 10.34148/teknika.v15i1.1461

Table 7. Cluster Summary

Cluster	Label	Purchase Frequency (Mean [Min-Max])	Product Price (Mean [Min-Max])	Total Revenue (Mean [Min-Max])
0	Low-demand	30.41 (1-226)	52,817 (500-125,000)	1,27M (1,000-7,16M)
1	Premium	69.93 (1-163)	146,196 (59,000-485,000)	7,93M (148,500-18,5M)
2	Best-selling	344.47 (193-542)	53,666 (15,000-115,000)	15,8M (6,04M-25,28M)

After the clustering process is done, the Apriori algorithm was applied to understand the relationship between products in each cluster. The minimum support value initially used was 0.1, but no frequent itemset was found. This was because the largest number of product sales only reached 542 times, which was less than 10% of total sales. Therefore, the minimum support value was then gradually lowered to 0.01, which produced enough frequent itemsets: 41, 24, and 23 itemsets for cluster 0, 1, and 2. Table 8 shows the frequent itemsets with more than one product for each cluster.

Table 8. Frequent Itemsets

No	Itemsets	Support	Cluster
1	(ekstra Oatmilk, Picollo Latte)	0.011575	Cluster 0
2	(Emerald Green, Eye Opener)	0.017053	Cluster 1
3	(Canele, Assorted Madeleine)	0.015637	Cluster 2
4	(Hot Cappucino, Sparkling Water)	0.014577	
5	(Hot Cappucino, Hot Cafe Latte)	0.014312	
6	(Hot Americano, Hot Cappucino)	0.013517	
7	(Assorted Madeleine, Hot Cafe Latte)	0.011132	
8	(Mineral Water | Pristine, Poulet Roti Au Thyme)	0.010602	
9	(Canele, Hot Americano)	0.010337	
10	(Canele, Hot Cappucino)	0.010072	

From these frequent itemsets, six association rules were generated, including various product categories, starting from coffee, non-coffee, fresh juice, bakery, and main course. Here are those six association rules:

1. Picollo Latte → ekstra Oatmilk
2. Emerald Green → Eye Opener
3. Canele → Assorted Madeleine
4. Assorted Madeleine → Hot Cafe Latte
5. Hot Cappucino → Sparkling Water
6. Poulet Roti Au Thyme → Mineral Water

E. Evaluation

For the K-Means model, a silhouette score of 0.73 was obtained without data scaling, while a score of 0.55 was obtained after scaling. Despite the decrease in silhouette score, the scaled model was chosen because it was able to produce a more balanced and representative product segmentation for all variables used. To verify this hypothesis, an experiment was done by applying the Apriori algorithm to both models. The results show zero association rules produced by the unscaled model, even at low support thresholds, while the standardized model successfully identifies six meaningful association rules. This validates that the standardized model can generate more actionable insights than the raw model.

On the other hand, an evaluation for the Apriori model was performed based on the lift value of each association rule, as shown in Table 9.

Table 9. Lift Association Rule Values

No	Antecedents	Consequents	Lift
1	(ekstra Oatmilk)	(Picollo Latte)	5.224189
2	(Picollo Latte)	(ekstra Oatmilk)	5.224189
3	(Emerald Green)	(Eye Opener)	1.837093
4	(Eye Opener)	(Emerald Green)	1.837093
5	(Assorted Madeleine)	(Canele)	1.740137
6	(Canele)	(Assorted Madeleine)	1.740137
7	(Hot Cappucino)	(Sparkling Water)	1.621895
8	(Sparkling Water)	(Hot Cappucino)	1.621895
9	(Poulet Roti Au Thyme)	(Mineral Water | Pristine)	1.523121
10	(Mineral Water | Pristine)	(Poulet Roti Au Thyme)	1.523121
11	(Hot Cafe Latte)	(Assorted Madeleine)	1.247548
12	(Assorted Madeleine)	(Hot Cafe Latte)	1.247548

From the result above, it can be known that all the association rules have a lift value greater than 1. This indicates that the purchase of a product in a transaction increases the likelihood of purchasing other products compared to random conditions. Therefore, these association rules can be considered for the implementation of bundling and up-selling strategies.

F. Deployment

The deployment phase aims to translate the modeling results into actionable insights for the café. The K-Means and Apriori algorithms identified six product relationship patterns that have a high potential for implementation in the form of bundling and up-selling. To evaluate whether the analysis results are applicable and in line with the café's operational

DOI: 10.34148/teknika.v15i1.1461
CC BY SA

conditions, this research conducted an interview with the café owner, without implementing direct experiments or A/B testing. It is important to note that further assessment is required before implementing these strategies on a larger scale. In addition, promotional constraints, including maximum discount limits and operational considerations, may affect the scope of execution.

The results of the interview session confirmed that the product combinations generated by the system are very realistic to be implemented as bundling and up-selling strategies. These actions are considered in line with the café's operational conditions, where the café has implemented some promotions in the form of bundling and packages. The café owner mentioned that each promotion was determined based on the considerations of profit margin to ensure that each package still provides enough profit for the business. The café owner also planned to arrange an internal meeting as a follow-up to this validation, discussing the implementation of the new strategies generated by the system.

For further development, this research proposes the integration of actionable insights from the model into the POS system that has been used by the café. This focuses on the implementation of a recommendation engine while also considering profit margin as a priority weight in determining product packages. By integrating these, the POS system is enabled to suggest product combinations that are financially profitable, optimized based on the profit margin parameters set by the owner.

IV. CONCLUSION

This research aims to optimize sales strategies in a café based on its transaction data over 6 months. The CRISP-DM method was applied in the analysis of product segmentation and customer purchasing patterns using the K-Means Clustering and Apriori algorithms. The clustering process successfully grouped the products into three clusters with a silhouette score of 0.55, namely the low-demand products, premium products, and best-selling products. This segmentation provides a clear understanding of the available products and can be used as the basis for the customer purchasing patterns analysis.

Next, the Apriori algorithm was applied to each cluster to generate frequent itemsets that formed several association rules. Overall, the highest product association was found between Extra Oatmilk and Piccolo Latte, with a lift value of 5.22. Furthermore, high correlations were also found in five other association rules, namely Emerald Green Juice → Eye Opener Juice, Assorted Madeleine → Canele, Hot Cappucino → Sparkling Water, Poulet Roti Au Thyme → Mineral Water | Pristine, and Hot Cafe Latte → Assorted Madeleine, with lift values ranging from 1.2 to 1.8. A lift value greater than 1 in all association rules shows that there is a strong association between these products.

In the final stage of this research, the results were validated through an interview with the café owner. The interview results showed that the generated rules are applicable for product bundling or up-selling strategy and are considered in line with the current operational conditions of the café. This research proposes the integration of actionable insights into the POS system used by the café while involving profit margin calculation to provide financially profitable product combinations.

Overall, this study proves that the hybrid approach using K-Means and Apriori algorithms can be used to design effective data-driven strategies for an F&B business. For further research, it is recommended to conduct direct trials or experiments to determine the impact of the marketing strategies implemented on sales and transaction values, as well as customer responses to these promotions. Thus, business actors are expected to conduct a continuous evaluation to improve business quality and maintain good relationships with customers.

REFERENCES

[1] Badan Pusat Statistik Indonesia, "Statistik Penyediaan Makanan dan Minuman 2023," 2024.

[2] A. G. Henryanto, H. Hanifah, T. S. Kaihatu, R. K. Hardjono, M. Cahyadin, and W. Hartono, "Digital Platform Capability and Strategic Agility: Uncovering Strategies to Strengthen Indonesian F&B SMEs Competitiveness," *Indonesian Journal of Business and Entrepreneurship*, vol. 11, no. 3, p. 685, Sep. 2025, doi: 10.17358/ijbe.11.3.685.

[3] M. Isa, A. Praswati, and S. Zulaekah, "Factors Driving Sustainability of Food and Beverage Industries," *Jurnal Ilmiah Bidang Ilmu Ekonomi*, vol. 19, no. 2, pp. 162–171, 2024, [Online]. Available: http://journal.umpo.ac.id/index.php/ekuilibrium

[4] G. Yılmaz Benk, B. Badur, and S. Mardikyan, "A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable Artificial Intelligence," *Information (Switzerland)*, vol. 13, no. 8, Aug. 2022, doi: 10.3390/info13080373.

[5] K. A. Carroll, A. Samek, and L. Zepeda, "Consumer Preference for Food Bundles under Cognitive Load: A Grocery Shopping Experiment," *Foods*, vol. 11, no. 7, Apr. 2022, doi: 10.3390/foods11070973.

[6] Jandrianto Patandianan, "Strategi Manajemen Pengelolaan Peluang Bisnis F&B Sebagai Upaya Mengatasi Permasalahan Ekonomi Keluarga dengan Pemanfaatan Teknologi Digital," *Jurnal Ilmiah Dan Karya Mahasiswa*, vol. 2, no. 2, pp. 199–216, Mar. 2024, doi: 10.54066/jikma.v2i2.1710.

[7] A. M. Maulidya, S. Selfira, G. Sidabutar, and R. R. Al Hafiz, "Integrating K-Means Clustering and Apriori for Data Mining-Based Digital Marketing Strategy For Increasing UMKM: Study Case Stabat City," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 4, pp. 3190–3196, Oct. 2025, doi: 10.33395/sinkron.v9i4.15299.

[8] J. Macqueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Jan. 1967.

[9] F. A. Maresti, W. I. Rahayu, M. B. B. C. Lustin, and T. H. Pakpahan, "Implementasi K-Means Untuk Melakukan Segmentasi Produk Berdasarkan Data Transaksi Retail," *Jurnal Ilmiah Sains dan Teknologi*, vol. 9, no. 1, Feb. 2025.

[10] S. Marselina, J. Haerul Jaman, and D. Ely Kurniawan, "Sales Analysis Using Apriori Algorithm in Data Mining Application on Food and Beverage (F&B) Transactions," *Journal of Applied Informatics and Computing (JAIC)*, vol. 7, no. 2, pp. 2548–6861, Jul. 2023, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[11] A. Ivana and I. Maryati, "Implementation of Apriori Algorithm in Identifying Purchase Relationships at Bluder Cokro Pakuwon Mall," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 2, p. 556, 2025, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[12] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining."

[13] H. Ar-Rasyid, A. Fitriansyah, J. Sari, and I. Ramadhan, "Sistem Rekomendasi Penjualan Cross-Selling Potensial Dengan Pendekatan Data Mining," *Jurnal Rekayasa Informasi Swadharma (JRIS)*, Jan. 2025.

[14] Amiruddin and R. Ishak, "Optimalisasi Seleksi Atribut K-Means Menggunakan Correlation Matrix pada Clustering Penyakit Pasien," *Jambura Journal of Electrical and Electronics Engineering*, Jul. 2025.

[15] H. Supriyanto, M. Al Hafidz, A. C. Puspitaningrum, R. A. P. Firmansyah, and R. Zuhdi, "Klasterisasi Data Obat Farmasi Berdasarkan Jumlah Persediaan Dengan Menggunakan Metode K-Means," *Teknika*, vol. 13, no. 3, pp. 361–369, Sep. 2024, doi: 10.34148/teknika.v13i3.987.

[16] R. Ishak, Nurmawanti, and Amiruddin, "Optimasi K-Means pada Clustering Penyakit Ibu Hamil Menggunakan Random Forest Optimization of K-Means in Disease Clustering of Pregnant Women Using Random Forest," *Jambura Journal of Electrical and Electronics Engineering*, p. 41, Jan. 2024, doi: https://doi.org/10.37905/jjeee.v7i1.28374.

[17] L. E. Dendy and R. Tanamal, "Association Rule Mining for Truck Body Damage Pattern Analysis Using Apriori and CRISP-DM," 2026. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[18] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator IPM Dengan K-Medoids," *Jurnal Gaussian*, vol. 13, no. 1, pp. 13–24, Aug. 2024, doi: 10.14710/j.gauss.13.1.13-24.

[19] I. Wayan Supriana and L. Arida Ayu Rahning Putri, "Implementasi Algoritma Apriori sebagai Association Rule Learning untuk Mengidentifikasi Pola Item Dataset Penjualan," *Jurnal Buana Informatika*, Apr. 2025.

[20] H. Indra, S. Saragi, and L. Putriana, "Analysis of the Effect of Up Selling and Cross Selling Sales Methods on the Brand Image of KFX UB Jaya I Pharmacy and its Impact on Purchasing Decisions," *International Journal of Science and Society*, vol. 6, no. 1, p. 2024, 2024, [Online]. Available: http://ijsoc.goacademica.com