

Modeling Factors of Product Classification Proximity in Data Mining Applications

S.N.M.P. Simamora¹, M. R. Irwan²

¹Department of Informatics Engineering, Institut Digital Ekonomi LPKIA, Bandung, Indonesia

²Department of Information Systems, Institut Digital Ekonomi LPKIA, Bandung, Indonesia

ABSTRACT

In today's nano-technology era, as device sizes shrink, the volume of data generated has grown exponentially. This increase manifests in both the number and size of files. Real-world data corroborates this trend. Given the massive scale of data, efficient methods are required to extract and process information to meet specific needs and preferences. One promising approach is data mining, particularly in analyzing the relationships between frequently purchased products. For instance, in a convenience store, if a customer selects product Z1, what is the likelihood that they will also purchase product Z2? How does this pattern extend to subsequent purchases? By understanding these associations, retailers can optimize product placement, enhancing customer convenience and increasing sales. Furthermore, retailers can streamline procurement processes by focusing on products with strong association patterns, thereby reducing costs. This research employs data mining techniques, specifically the Apriori algorithm, to investigate product associations in a convenience store using data collected between May 1 and September 30, 2023. The CRISP-DM methodology guided the research. The findings reveal significant associations between products purchased by customers, as evidenced by the highest frequent itemsets and association rules. Additionally, the data underscores the need to classify consumer goods into food and beverage categories. The analysis indicates that the purchase of food items does not necessarily lead to the purchase of beverages, and vice versa. Consequently, it can be concluded that a customer's purchasing decisions are primarily driven by their immediate needs or preferences rather than established product associations or habitual behavior.

Keyword : Data Mining; CRISP; Algoritma Apriori; Frequent Itemset; Association Rules

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

S.N.M.P. Simamora
Department of Informatics Engineering
Institut Digital Ekonomi LPKIA
Jl. Soekarno Hatta No.456, Bandung 40266, Indonesia
Email : snmpsimamora@lpkia.ac.id

Article history:

Received Jun 9, 2026
Revised Aug 20, 2026
Accepted Sept 11, 2026

1. INTRODUCTION

Transactions using financial technology (fintech) applications have increased significantly, driven by a growing public preference to avoid the complexities of cash-based systems. This shift is further reinforced by government initiatives promoting cashless payments for all transactions (Ministry of Communication and Informatics, 2019), as they ensure every nominal value of incoming and outgoing funds is precisely recorded. This digital transparency directly supports system auditing processes and financial accountability. Consequently, it helps prevent embezzlement and financial manipulation in areas such as taxation, payroll, and corporate profit-and-loss reporting.

A single transaction type within a fintech application can generate millions of records stored daily. Consequently, a systematic method is required for data extraction to ensure that information can be processed according to specific needs and objectives. This challenge also applies to the retail sector, such as minimarkets, which typically offer a wide array of food and beverage products. There is often a strategic consideration regarding product placement on displays to ensure customers can easily locate items without confusion. Despite standard categorization across display aisles in most minimarkets, customers frequently encounter difficulties in finding specific products.

One approach to identifying customer purchasing patterns across various products in a minimarket is by sampling transaction data over a specific observation period. This process reveals discernible purchasing trends that serve as a foundational reference for future planning. Furthermore,

this method enables more efficient and effective procurement strategies; products that are seldom identified as proximity factors are deprioritized in the procurement cycle. Instead, priority is given to products with significant proximity factors. Consequently, operational costs are optimized, as the inventory consistently aligns with the items most frequently selected by customers.

Large and diverse datasets can be extracted from storage systems using data mining methods (Cao, L., et al. 2011)(Chen, H., et al. 2015). The general principles of data mining include the automatic discovery of patterns, the prediction of likely outcomes, the creation of actionable information, and a focus on large datasets and databases (Xu, L. et al. 2014)(Feng, Z. & Zhu, Y. 2016)(Wu, X. et al. 2014). Furthermore, the Apriori algorithm is utilized to identify frequency patterns within distributed data, enabling the extraction of association rules from transaction records (Wu, X. et al. 2014)(Liang, R. et al. 2023). An algorithm itself is defined as a series of systematic and structured steps designed to achieve a specific objective or solve a problem (Tian, M. et al. 2020)(Simamora, S.N.M.P. 2016)(Simamora, S.N.M.P. 2025). Various applications of data mining have been implemented across multiple fields, significantly assisting in the discovery of patterns within expansive and varied data distributions (Middelfart, M. et al. 2013)(Liang, R. et al. 2023)(Yin, Y., Long, L., & Deng, X. 2020). In the field of Information and Communication Technology (ICT), for instance, daily transactions can reach a frequency of; thus, specialized methods and techniques are required to retrieve target data according to specific requirements (Simamora, S.N.M.P. ed. 20 Juni 2023)(Martínez-Plumed, F. et al. 2021)(Simamora, S.N.M.P. ed. 16 Januari 2023). From these large datasets, a subset is used to discover a pattern, while the remaining data is employed to test the model, assuming that the simulation results are identical to one another (Guo, H. 2021)(Maarip, A. & Simamora, S.N.M.P. 2023).

Within the concept of data mining, the term frequent itemset refers to recurring patterns that emerge from various datasets, such as transaction records, YouTube channel visits, or consumer access to a marketplace. A frequent itemset can further be defined as an itemset that appears in transactions with a frequency exceeding a predefined minimum support; from the resulting frequent itemsets, the number of products within each set can be identified through the length attribute (Xu, L. et al. 2014)(Wu, X. et al. 2014). The primary advantage of identifying frequent itemsets is the ability to predict customer behavior patterns. This reference point enables more accurate forecasting for procurement, which is expected to drive sales growth. Additionally, it facilitates purchasing analysis and the tracking of latent (hidden) buying patterns (Tian, M. et al. 2020)(Yin, Y., Long, L., & Deng, X. 2020). The Apriori Algorithm is commonly employed to identify these frequent itemsets (Chen, H., et al. 2015)(Tian, M. et al. 2020).

$$\text{support}(x,y) = P(xy) = \frac{\text{num}(xy)}{\text{num}(\text{All samples})} \quad (1)$$

$$\text{support}(X \Rightarrow Y) = P(X \cup Y) = \frac{\text{count}(X \cup Y)}{|D|} \quad (2)$$

In the field of data mining, association rules are defined as rules that describe the relationships between itemsets within distributed data, particularly in transactional activities. These are commonly identified in purchasing behavior, YouTube channel visits, or marketplace interactions. These rules indicate that if a specific set of items appears in a transaction, other items are also likely to appear within the same transaction. For instance, product becomes associated with because a pattern has formed based on historical transaction data. The core components of these rules include the antecedent, consequent, support, and confidence. Consequently, frequent itemsets serve as the fundamental basis for generating association rules.

$$\text{confidence}(x \Rightarrow y) = P(x|y) = \frac{P(xy)}{P(y)} \quad (3)$$

$$\text{confidence}(X \Rightarrow Y) = P(X|Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)} \quad (4)$$

Data mining also utilizes a parameter known as lift to measure the strength of the relationship between two or more itemsets. For instance, if the purchase of product is followed by the purchase of, the lift value indicates the relative strength of the association—determining, for example, whether the relationship is stronger than (Wu, X. et al. 2014)(Liang, R. et al. 2023)(Yin, Y., Long, L., & Deng, X. 2020).

$$\text{lift}(x \Rightarrow y) = \frac{P(x|y)}{P(x)} = \frac{\text{Confidence}(x \Rightarrow y)}{P(x)} \quad (5)$$

The lift metric employs a value of 1 as a benchmark to map the relationship between x and y . If the lift value is greater than 1, then $x \Rightarrow y$ is considered a valid and robust association rule. Conversely, a lift value of less than 1 indicates that the rule is invalid and weak. A lift value equal to 1 signifies that x and y are independent and share no significant relationship (Xu, L. et al. 2014)(Wu, X. et al. 2014).

2. RESEARCH METHOD

This study utilizes the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The stages consist of business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Martínez-Plumed, F. et al. 2021)(Ali, M. etc. 2018). Furthermore, all data processing is conducted using Jupyter.org, a cloud-based computing application.

The data samples for this study were obtained from a franchised minimarket outlet located in Bogor Regency, West Java. The observation period spanned from May 1, 2023, to September 30, 2023. The dataset involved in the testing process consists of 1,134 product types across 32,814 sales transactions, organized into a single transaction table containing 56,135 data rows.

Table 1. Statistik Data Awal

Data Awal		Produk	
Jumlah record	56135	Jumlah jenis produk	1134
Periode data	1 Mei 2023 sampai dengan 30 September 2023	Jumlah kategori produk	17
Jumlah hari transaksi	152		

Initial data collection commenced with the extraction of a dataset in .csv (comma-separated values) format, spanning a 153-day period. The data was categorized into 17 distinct product groups.

Table 2. Data Description

Attribute	Description
date	The date on which the transaction occurred
hour	Transaction timestamp
sales_id	A unique transaction identifier comprising the date and a sequential number, which resets daily starting from 1.
category	Product classification
product	Item designation
qty	Quantity of products sold per transaction

Each data row represents an individual product sold, and a single transaction (identified by sales_id) may comprise multiple rows, corresponding to the number of distinct product types purchased within that transaction. This structure is illustrated in the first 15 rows of the sample dataset presented in Figure 1.

The following criteria and assumptions define the scope of the study site:

- a) Location Context: The minimarket is situated within an office district rather than a residential zone. Consequently, certain household staples—such as rice, sugar, and sliced bread—are not part of the product inventory.
- b) Operational Trends: A notable decrease in transaction volume occurs on Saturdays and Sundays, as most surrounding offices are closed.
- c) Operating Hours: The facility maintains standard operational hours from 07:00 to 22:00.

	date	hour	sales_id	category	product	qty
0	2023-05-01	7	20230501.0001	HEALTH CARE	VITACIMIN ORANGE STRIP 2S	2
1	2023-05-01	7	20230501.0001	HEALTH CARE	XON-CE TABLET VIT C 500MG 2S	1
2	2023-05-01	7	20230501.0002	DAIRY	DIAMOND F.M CHOCO TP 1 L	1
3	2023-05-01	7	20230501.0003	FOOD	TAC STEAMY HOT DOG 12 CM	1
4	2023-05-01	7	20230501.0004	DAIRY	FRISIAN FLG LIQ/D CH BTL 180ML	1
5	2023-05-01	7	20230501.0005	BAKERY	LB CHOCOLATE	1
6	2023-05-01	7	20230501.0005	PACKAGED BEVERAGE	GOLDA DOLCE LATTE 200ML	1
7	2023-05-01	7	20230501.0006	FOOD	TAC STEAMY HOT DOG 12 CM	1
8	2023-05-01	7	20230501.0007	DAIRY	ULTRA MILK CHOCO TP 250 ML	1
9	2023-05-01	8	20230501.0008	PACKAGED BEVERAGE	NU MILK TEA BTL 330 ML	1
10	2023-05-01	8	20230501.0009	DAIRY	HEAVENLY BLUSH GREEK 100 ML	1
11	2023-05-01	8	20230501.0009	PACKAGED BEVERAGE	PRISTINE 600 ML	1
12	2023-05-01	8	20230501.0010	PACKAGED BEVERAGE	CAP KAKI TIGA JERUK CAN 320 ML	1
13	2023-05-01	8	20230501.0010	PACKAGED BEVERAGE	LARUTAN CAP BADAK BTL 200ML	1
14	2023-05-01	8	20230501.0010	CONFECTIONARY	STREPSILS MENTHOL SACHET 12X8S	1

Fig 1. Sample of the initial 15 records

3. RESULTS AND DISCUSSION

The obtained data reveals several notable insights that enhance the understanding of the dataset's characteristics. As illustrated in Figure 2, there is no evident upward trend in the overall number of transactions. However, an upward tendency is observable specifically during the period from May 23 to June 24, 2023.

Analysis of the 17 sold product categories reveals that the Packaged Beverage category accounts for the highest transaction volume. As illustrated in Figure 3, this category significantly outperforms others, maintaining a substantial lead in total transactions.

Based on the simulation results obtained from jupyter.org, the daily transaction volume ranges from a minimum of 30 to a maximum of 393. The average daily transaction count is 215 (calculated from the total of 32,814 transactions). The monthly transaction distribution is illustrated in Figure 4.

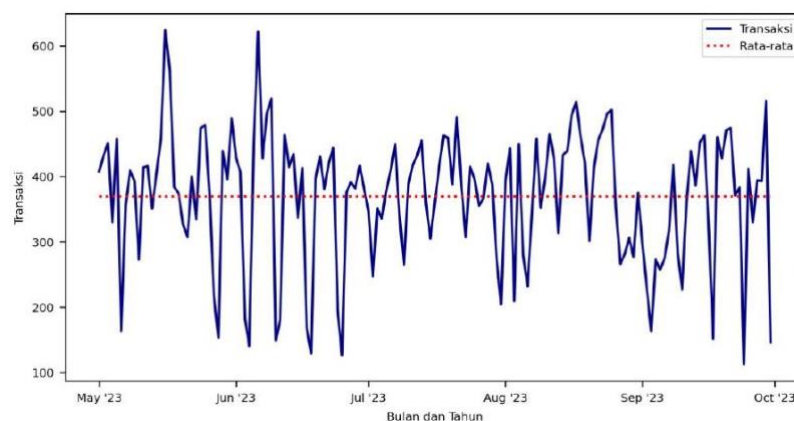


Fig 2. Daily transaction volume

Monthly transaction volumes show no significant increase; in fact, the data indicates a decline in transactions for September 2023. However, this decrease is not statistically significant when compared to the figures from June 2023.

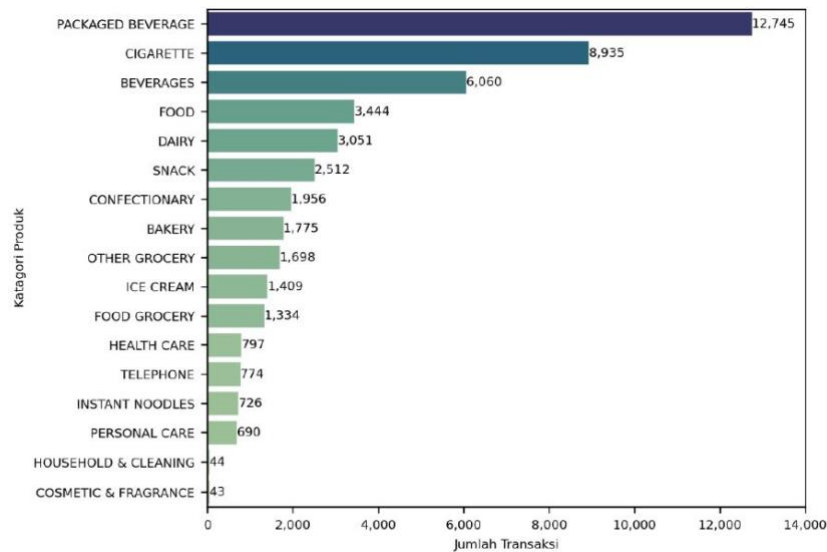


Fig 3. Transaction Volume by Category

The analysis subsequently focuses on the total and average transaction volume by day, as illustrated in Figure 5. Furthermore, the products with the highest transaction frequencies are presented in Figure 6. Within the beverage category, AQUA AIR MINERAL BTL 600 ML recorded the highest number of transactions, while INDOMIE GORENG SPECIAL BAG 84GR emerged as the leading product in the food category.

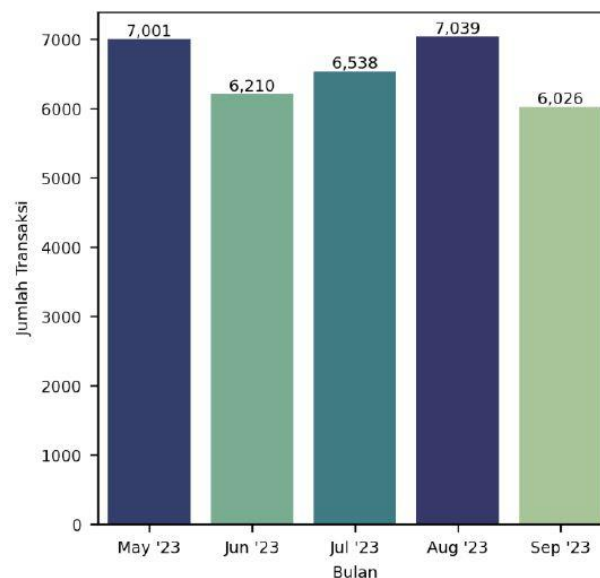


Fig 4. Monthly Transaction Volume

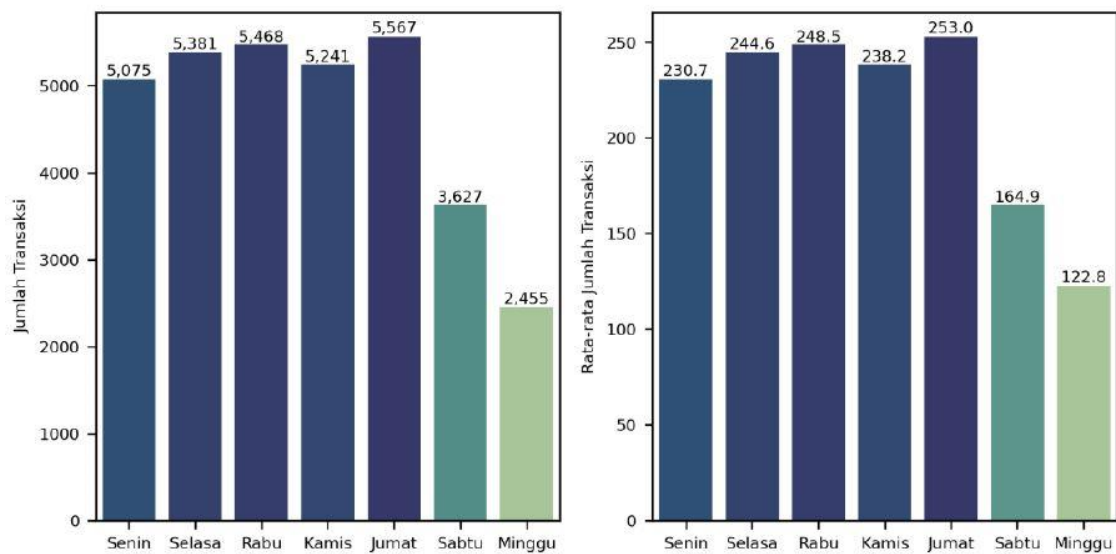


Fig 5. Total and Average Transactions by Day

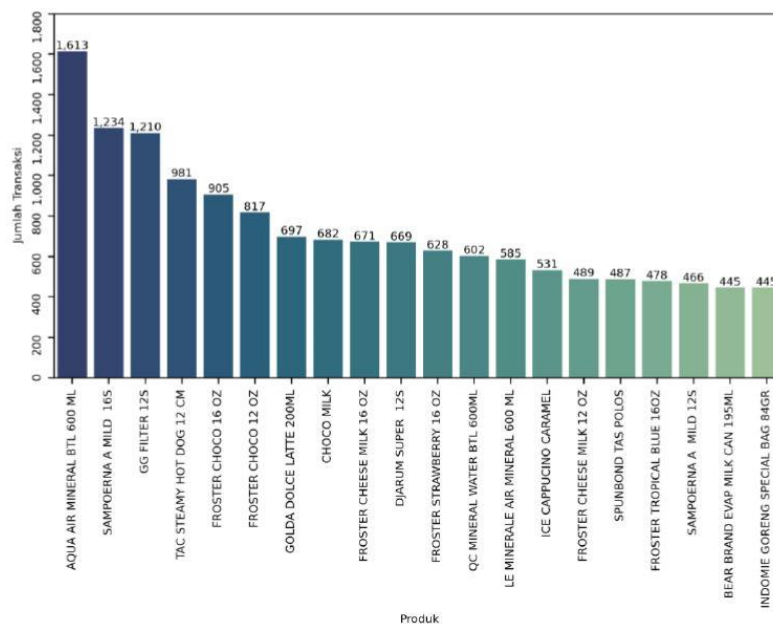


Fig 6. Products with the Highest Transaction Frequency

The data is categorized into three distinct groups: the primary dataset, the non-food and non-beverage subset, and the food and beverage subset. This classification was implemented to achieve more specific modeling and to enhance the accuracy of the results.

The primary dataset encompasses the complete initial data obtained. The non-food and non-beverage subset comprises products that are neither edible nor drinkable; this subset is classified into six categories: confectionery, other grocery, healthcare, personal care, household & cleaning, and cosmetics & fragrance. Furthermore, the food and beverage subset includes edible and drinkable products categorized into eight groups: bakery, dairy, food, food grocery, ice cream, instant noodles, packaged beverages, and snacks.

Subsequently, the Apriori Algorithm was applied, which requires input data in a tidset format with boolean encoding. Consequently, a matrix transformation was implemented: each product serves as a column, and the sales_id serves as a row, as illustrated in Figure 7. The association modeling in this study utilizes the Apriori Algorithm and association rules from the Python mlxtend library.

product	5 DAYS CROISSANT CHOCOLATE	5 DAYS CROISSANT MIXED BERRIES	5 DAYS CROISSANT PEANUT BUTTER	ABC KACANG HIJAU 250ML	ABC KECAP ASIN 133ML	ABC KECAP MANIS 135ML	ABC KOPI SUSU BOX 5X32 GR	ABC SAMBAL ASLI 135ML	ABC SELERA PEDAS NAMPOL 80GR	ABC SQUASH ORANGE BTL 460ML	...	WAKU WAKU CHEWY GRAPE 50ML	WAKU WAKU CHOCOLOOP 50ML	YAKULT BANDED PACK 5X65ML
sales_id														
20230501.0005	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230501.0009	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230501.0010	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230501.0016	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230501.0020	False	False	False	False	False	False	False	False	False	False	...	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20230930.0041	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230930.0050	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230930.0051	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230930.0094	False	False	False	False	False	False	False	False	False	False	...	False	False	False
20230930.0099	False	False	False	False	False	False	False	False	False	False	...	False	False	False

Fig 7. Tidset Format with Boolean Encoding

In the developed model, two parameters are defined: minimum support and minimum confidence. The minimum support parameter serves as the lower bound for itemset support; any itemset with a support value below this threshold is excluded from subsequent stages. Similarly, the minimum confidence parameter acts as the lower threshold for the confidence values generated during the association rule mining process.

For all datasets and subsets, the minimum confidence parameter is set at 0.8 to ensure that the generated association rules are both valid and robust. To determine the optimal minimum support value, an iterative approach is employed to identify a sufficient number of association rules at the specified confidence level. This involves testing various minimum support thresholds until the desired quantity of association rules is achieved.

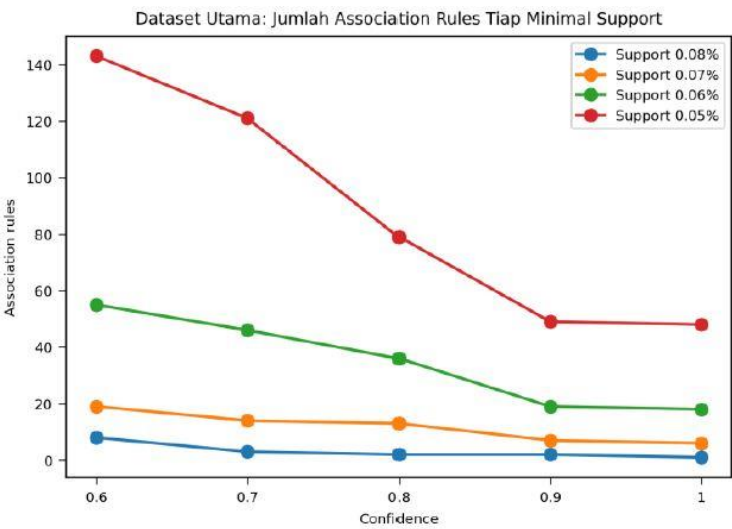


Fig 8. Primary Dataset Minimum Support Threshold

Based on Figure 8, it is observed that setting a minimum support threshold that is too high, such as 0.08% or 0.07%, results in very few or no association rules being identified. Therefore, a minimum support parameter of 0.06% was applied to the primary dataset. For the food and beverage subset, the minimum support parameter was set at 0.09% (0.0009), as illustrated in Figure 9.

Subsequently, frequent itemsets are identified by invoking the Apriori function, beginning with the application of the previously defined minimum support and minimum confidence parameters. The implementation using the Python library is detailed in Figure 10. For each dataset and subset, the simulation results for frequent itemsets (Food Category) are presented in Figure 11. Furthermore, the top five association rules and the overall distribution for the primary dataset are illustrated in Figure 12 and Figure 13, respectively.

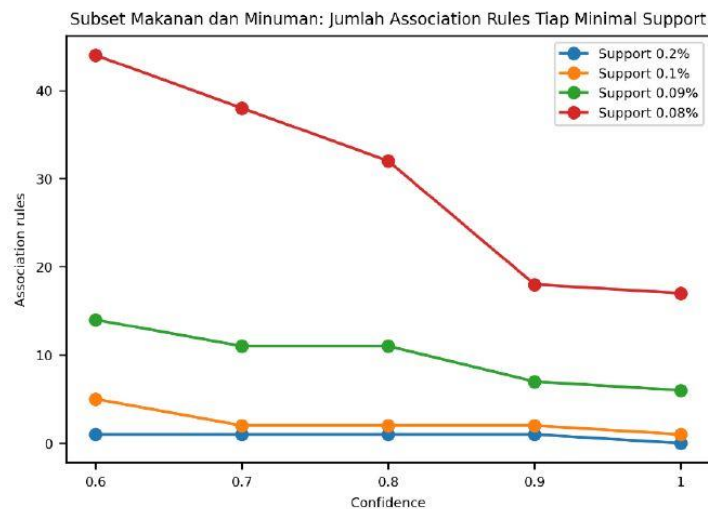


Fig 9. Minimum Support for the Food and Beverage Subset

```

1. n_min_support_mak = 0.0012
2. n_min_confidence_mak = 0.8
3.
4. # Cari frequent itemset
5. freq_itemsets = apriori(makanan_pv, min_support=n_min_support_mak,
6.                           use_colnames=True)

```

Fig 10. Procedure for executing the Apriori function

support	itemsets	length
0.032784	(FROSTER CHOCO 16 OZ, FROSTER CHEESE MILK 16 OZ)	2
0.022085	(FROSTER TROPICAL BLUE 16OZ, FROSTER STRAWBERR...	2
0.016097	(FROSTER STRAWBERRY 16 OZ, FROSTER CHOCO 16 OZ)	2
0.011386	(FROSTER TROPICAL BLUE 16OZ, FROSTER CHOCO 16 OZ)	2
0.010208	(INDOMIE GORENG SPECIAL BAG 84GR, BEAR BRAND E...	2

Fig 11. Primary Dataset: Five Highest Frequent Itemsets

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(BIMOLI KLASIK POUCH 2LT)	(INDOMIE GORENG SPECIAL BAG 84GR)	0.010699	0.023066	0.010012	0.93578	40.569042
(LB ALMOND BROWNIE, TAC - TUNA SANDWICH)	(LB MARBLE CAKE)	0.001080	0.010797	0.001080	1.00000	92.618182
(TAC BROWN BREAD TUNA CHEESE, LB MARBLE CAKE)	(TAC - TUNA SANDWICH)	0.000785	0.004024	0.000785	1.00000	248.467805
(LB CHEESE CAKE 85 GR, LB FUDGY BROWNIE)	(LB MARBLE CAKE)	0.000785	0.010797	0.000785	1.00000	92.618182
(ONIGIRI TAC TUNA MAYO, LB FUDGY BROWNIE)	(LB MARBLE CAKE)	0.000785	0.010797	0.000785	1.00000	92.618182

Fig 12. Primary Dataset: Five Highest-Ranking Association Rules

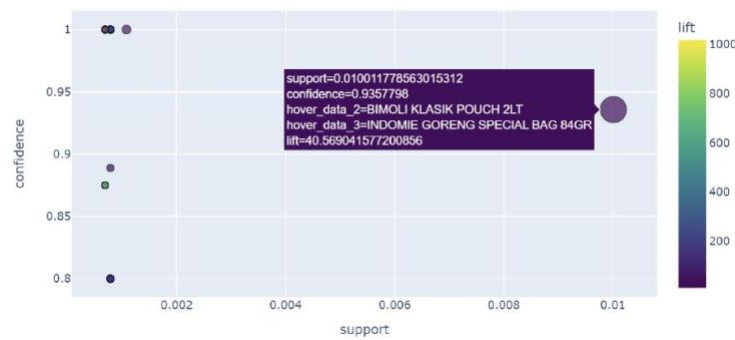


Fig 13. Primary Dataset: Distribution of Association Rules

Based on the simulation results for the primary dataset, a significant association pattern emerges: customers purchasing BIMOLI KLASIK POUCH 2LT are highly likely to also purchase INDOMIE GORENG SPECIAL BAG 84GR. This relationship is supported by a support value of 1%, a confidence level of 93.6%, and a lift ratio of 40.6. The lift ratio obtained from the jupyter.org simulation indicates a robust correlation between these two products, suggesting a strong tendency for them to be purchased concurrently. Furthermore, as illustrated in Figure 13, the distribution data shows that the itemset {BIMOLI KLASIK POUCH 2LT, INDOMIE GORENG SPECIAL BAG 84GR} significantly outperforms all other itemsets. This indicates that this specific itemset is the only one of substantial interest within the primary dataset.

support	itemsets	length
0.012148	(INDOMIE GORENG SPECIAL BAG 84GR, BEAR BRAND E...	2
0.011914	(INDOMIE GORENG SPECIAL BAG 84GR, BIMOLI KLASI...	2
0.008060	(5 DAYS CROISSANT MIXED BERRIES, 5 DAYS CROISS...	2
0.007359	(SOSIS TUSUK BOLADO, SOSIS TUSUK JUMBO HOT)	2
0.006424	(STMY HOT DOG HOT&SPCY, TAC STEAMY HOT DOG 12 CM)	2

Fig 14. Food and Beverage Subset: Five Highest-Ranking Frequent Itemsets

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(BIMOLI KLASIK POUCH 2LT)	(INDOMIE GORENG SPECIAL BAG 84GR)	0.012732	0.027450	0.011914	0.93578	34.090260
(LB ALMOND BROWNIE, TAC - TUNA SANDWICH)	(LB MARBLE CAKE)	0.001285	0.012849	0.001285	1.00000	77.827273
(TAC BROWN BREAD TUNA CHEESE, LB MARBLE CAKE)	(TAC - TUNA SANDWICH)	0.000934	0.004789	0.000934	1.00000	208.804878
(LB CHEESE CAKE 85 GR, LB FUDGY BROWNIE)	(LB MARBLE CAKE)	0.000934	0.012849	0.000934	1.00000	77.827273
(ONIGIRI TAC TUNA MAYO, LB FUDGY BROWNIE)	(LB MARBLE CAKE)	0.000934	0.012849	0.000934	1.00000	77.827273

Fig 15. Food and Beverage Subset: Five Highest-Ranking Association Rules

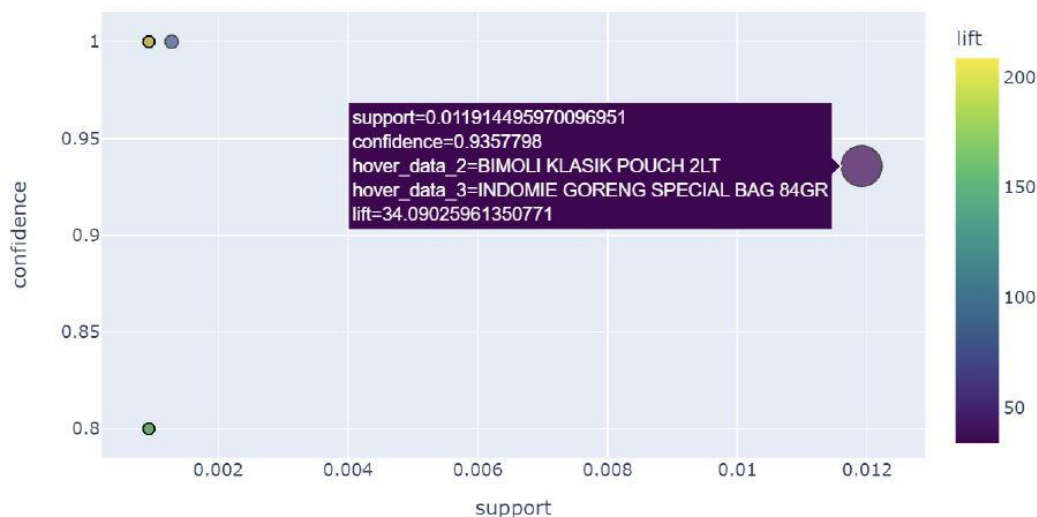


Fig 16. Food and Beverage Subset: Association Rule Distribution

As illustrated in Figures 14, 15, and 16, the simulation results for the food and beverage subset do not differ significantly from the results of the primary dataset. This is because the products within the strongest itemsets of the primary dataset also belong to the food and beverage categories. Furthermore, the analysis reveals a relationship indicating that customers who purchase BIMOLI KLASIK POUCH 2LT are highly likely to also purchase INDOMIE GORENG SPECIAL BAG 84GR, with a support value of 1.2%, confidence of 93.6%, and a lift ratio of 34.1.

The association results indicate that customer purchasing behavior deviates from conventional assumptions. For instance, the simulation results raise questions regarding the correlation between instant noodle and cooking oil purchases, as instant noodles are typically more associated with eggs or sausages. This suggests that product purchases are primarily driven by the customer's immediate needs at that time. Observation of the post-COVID-19 pandemic period reveals that the public generally prioritized stockpiling instant food and cooking oil.

4. CONCLUSION

Simulation results conducted via jupyter.org demonstrate an association between BIMOLI KLASIK POUCH 2LT and INDOMIE GORENG SPECIAL BAG 84GR. Based on the sampled data from May 1, 2023, to September 30, 2023, customers who purchased BIMOLI KLASIK POUCH 2LT also purchased INDOMIE GORENG SPECIAL BAG 84GR. Consequently, retailers may strategically position these two products in close proximity within their shelf displays.

The test results demonstrate the successful identification of customer purchasing patterns within the parameters of frequent itemsets and association rules. This was achieved through the application of data mining using the Apriori Algorithm based on the Python mlxtend library.

A higher lift ratio signifies a stronger resulting association. Furthermore, based on the testing results, the modeling demonstrates that the application of data mining aligns with the objectives established for this study; specifically, it has successfully modeled the proximity factors of product classifications through data mining techniques.

ACKNOWLEDGEMENTS

We express our sincere gratitude to the Institut Digital Ekonomi LPKIA Bandung for their support throughout this study, specifically for providing the necessary network infrastructure and computer facilities. Their contribution was instrumental in ensuring the research was conducted successfully and efficiently.

REFERENCES

- Cao, L., et al. (2011). Combined Mining: Discovering Informative Knowledge in Complex Data. *IEEE Transactions on Systems, Man, and Cybernetics*, Volume: 41, Issue: 3.
- Chen, H., et al. (2015). A Decision-Theoretic Rough Set Approach for Dynamic Data Mining. *IEEE Transactions on Fuzzy Systems*, Volume: 23, Issue: 6.
- Xu, L. et al. (2014). Information Security in Big Data: Privacy and Data Mining. *IEEE Access*, Volume: 2.
- Feng, Z. & Zhu, Y. (2016). A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access*, Volume: 4.
- Wu, X. et al. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, Volume: 26, Issue: 1.
- Tian, M. et al. (2020). Data Dependence Analysis for Defects Data of Relay Protection Devices Based on Apriori Algorithm. *IEEE Access*, Volume: 8.
- Liang, R. et al. (2023). Exploring the Fusion Potentials of Data Visualization and Data Analytics in the Process of Mining Digitalization. *IEEE Access*, Volume: 11.
- Simamora, S.N.M.P. (2016). *Modul Belajar Praktis Algoritma dan Pemrograman*. Penerbit Deepublish, Yogyakarta. ISBN:978-602-401-318-9.
- Middelfart, M. et al. (2013). Efficient Sentinel Mining Using Bitmaps on Modern Processors. *IEEE Transactions on Knowledge and Data Engineering*, Volume: 25, Issue: 10.
- Liang, R. et al. (2023). Exploring the Fusion Potentials of Data Visualization and Data Analytics in the Process of Mining Digitalization. *IEEE Access*, Volume: 11.
- Yin, Y., Long, L., & Deng, X. (2020). Dynamic Data Mining of Sensor Data. *IEEE Access*, Volume: 8.
- Simamora, S.N.M.P. (ed. 20 Juni 2023). *Terapan Data-Science dalam Sistem Telekomunikasi*. Koran LintasTungkal. Edisi 20 Juni.
- Martínez-Plumed, F. et al. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, Volume: 33, Issue: 8.
- Simamora, S.N.M.P. (ed. 16 Januari 2023). *Model Pohon Relasi Untuk Mendeteksi Jalur Kedekatan Hubungan Antar-Entitas*. Koran Monitor. tgl.16 Januari.
- Guo., H. (2021). Research on Web Data Mining Based on Topic Crawler. *Journal of Web Engineering*. Volume: 20, Issue: 4.
- Maarip, A. & Simamora, S.N.M.P. (2023). Selektivitas Pemilihan Komputer dengan Pendekatan Pohon Keputusan. *Jurnal Teknik Informatika dan Sistem Informasi (TEFSIN)* Vol. 1 No. 2. Fakultas Teknik Universitas Nahdlatul Ulama Sumatera Barat. ISSN: 2987-6362.
- Ali, M. et al. (2018). A Data-Driven Knowledge Acquisition System: An End-to-End Knowledge Engineering Process for Generating Production Rules". *IEEE Access*. Volume: 6.
- Simamora, S.N.M.P. (2025). *Belajar Algoritma dan Pemrograman Dengan JavaScript*. Penerbit Uwais Inspirasi Indonesia, Kabupaten Ponorogo. ISBN:978-623-133-643-9.