Jurnal Galaksi (Global Knowledge, Artificial Inteligent and Information System) E-ISSN: 3048-2399
Vol 3, No 1, May 2026 , pp. 1-7
https://doi.org/10.70103/galaksi.v3i1.119

Profit-Based HUI-Miner for Discovering Consumer Shopping Patterns in a Retail Store

I Putu Noven Hartawan[1*], I Gede Sudiantara[2], Ni Kadek Bumi Krismentari[3], Gede Rudiharta Pratama Giri[4], I Made Dwi Putra Asana[5]

[1*,2,3,4,5]*Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia*
[1*]*novenhartawan@instiki.ac.id*; [2]*sudiantara@instiki.ac.id*; [3]*kadek_bumi@instiki.ac.id*; [4]*rudigiri@instiki.ac.id*; [5]*dwiputraasana@instiki.ac.id*

ARTICLE INFO

Article history:
Received 2 March 2026
Revised 15 April 2026
Accepted 25 May 2026
Available Online 31 May 2026

Keywords:
HUI-Miner;
Profit-based;
Shopping patterns;
Consumer;
Retail store

ABSTRACT

Retail transaction logs often reveal products that are frequently purchased together, but frequency alone does not show their economic contribution. This study applies a profit-oriented High Utility Itemset approach using HUI-Miner to identify valuable shopping patterns at YSL Grocery Store. The analysis followed CRISP-DM on 266,394 sales records from 2022. After attribute selection and missing-value removal, 251,225 records containing transaction, item, quantity, and price were processed. Item utility was represented by quantity multiplied by selling price, while the mining stage used a minimum utility of IDR 5,000,000, minimum support of 2%, and minimum confidence of 10%. Positive associations were retained when lift exceeded 1. The procedure produced 248 frequent itemsets, 77 confidence-qualified rules, 64 positive-lift rules, and 23 rules that satisfied all criteria. The strongest association linked two Sedaap instant-noodle variants with a lift of 3.26. The findings also show that support is not proportional to total utility: some cross-category combinations generated substantially greater utility despite lower occurrence. Therefore, retail decisions should combine utility, support, confidence, and lift when prioritizing shelf placement, bundles, promotions, and stock.

1. Introduction

Retail transaction data record repeated purchasing decisions that can support shelf arrangement, replenishment, promotions, and product bundling. YSL Grocery Store accumulated 266,394 sales rows during 2022. At this scale, manual inspection is inefficient and encourages decisions based primarily on intuition. Data mining offers a systematic way to transform raw records into actionable patterns when transaction volume and relationships among attributes are difficult to examine conventionally (Shu & Ye, 2023; Solanki & Sharma, 2021).

Market basket analysis commonly applies Apriori, FP-Growth, or ECLAT to discover products that frequently occur together. These approaches have supported digital catalogs, sales strategies, product recommendations, and inventory planning (Nata et al., 2022; Rahmi & Mikola, 2021; Siregar et al., 2022). However, occurrence frequency does not necessarily represent economic importance. Low-priced products may dominate frequent itemsets, while less frequent combinations with greater monetary contribution can be overlooked.

High Utility Itemset (HUI) mining addresses this limitation by considering internal utility, such as quantity, and external utility, such as price or profit. HUI-Miner employs utility-lists to identify high-utility itemsets without generating a large candidate set (Fournier-Viger et al., 2019). Recent studies further show that the selected data structure, utility threshold, and search strategy directly affect computational efficiency and the number of discovered patterns (Yan et al., 2024; Nugroho & Gunawan, 2025).

This study applies a profit-oriented HUI-Miner procedure to YSL Grocery Store transactions and combines utility with support, confidence, and lift. The objective is to identify product combinations that contribute economic value, occur with adequate frequency, form meaningful directional rules, and indicate positive dependence. The research question is whether frequently purchased combinations are also the combinations with the highest total utility. The contribution is a layered selection procedure, parameter-sensitivity analysis, and an operational interpretation of the resulting rules for a local retail setting.

2. Literatur Review

Association rule mining represents a directional relationship as A → B. Support is the proportion of transactions containing both A and B, confidence is the proportion of transactions containing A that also contain B, and lift compares confidence with the overall probability of B. A lift greater than 1 indicates positive dependence. Market basket studies emphasize that these measures serve different purposes and should be interpreted together so that an apparently strong rule does not merely reflect a popular consequent (Alawadh & Barnawi, 2022; Wahyuningsih et al., 2023).

Frequency-based methods have produced useful retail recommendations across different transaction contexts. Apriori has been used to reveal purchase patterns in grocery transactions (Rahmi & Mikola, 2021; Siregar et al., 2022), while ECLAT has supported product recommendations from historical sales data (Widyan & Rozi, 2021). A hybrid Apriori implementation also demonstrated that lower support and confidence thresholds increase the number of resulting rules (Asana et al., 2021). FP-Growth and comparative studies similarly show that algorithm choice influences memory use and the resulting pattern set (Ayu et al., 2021; Hery & Widjaja, 2024).

These frequent-pattern approaches remain oriented toward co-occurrence. Visual analytics and big-data platforms extend transaction analysis to segmentation, personalization, and retail decision support (Bakhrun et al., 2025; Shivampeta, 2025), but they do not automatically distinguish popular low-value combinations from less frequent high-value combinations. This distinction is important when the managerial objective involves revenue contribution, stock value, or margin rather than frequency alone.

Utility mining evaluates patterns according to economic contribution. Nonredundant high-utility association rules can reduce the result set and improve interpretability (Mai et al., 2020). Profit-support association mining likewise shows that including profit can reduce recommendation volume while increasing business relevance (Dogan, 2023). Within HUI mining, recent methods reduce costly join operations, use alternative search structures, or optimize the search process; higher utility thresholds consistently produce fewer itemsets (Yan et al., 2024; Nugroho & Gunawan, 2025). A recent retail study using Binary Particle Swarm Optimization also combined support, confidence, and lift to identify meaningful product relationships (Krismentari et al., 2026).

The identified gap is the limited integration of HUI-Miner with support, confidence, and lift on a large local grocery dataset, followed by an explicit comparison between frequency and economic contribution. Research on AI-driven data mining stresses that analytical results must remain interpretable and connected to organizational decisions (Mohammed et al., 2024). Accordingly, this study reports not only the number of patterns but also their sensitivity to threshold changes, the relationship between support and total utility, and their implications for shelf placement, promotions, bundling, and stock priorities.

3. Research Method

The study followed CRISP-DM through business understanding, data understanding, data preparation, modeling, evaluation, and deployment interpretation. Business understanding defined the store's need to distinguish patterns that are merely frequent from patterns with greater economic contribution. The secondary dataset consisted of 266,394 sales rows from 2022 and included transaction code, date, product, quantity, price, discount, category, and operator information.

During data preparation, the retained fields were transaction identifier, product, quantity, and selling price. Rows with missing product names were removed, leaving 251,225 rows. Item utility within a transaction was calculated as quantity multiplied by selling price. Because purchase cost and net margin were not included in the final model, this value is treated as a proxy for economic contribution rather than verified net profit.

Transaction Utility (TU) was calculated as the sum of all item utilities in a transaction. Transaction-Weighted Utility (TWU) for an item was calculated as the sum of TU values for every transaction containing that item. Items with TWU below the selected threshold were pruned. The remaining items were organized into utility-lists containing transaction identifiers, itemset utility, and remaining utility. The primary model used a minimum utility of IDR 5,000,000. The complete processing sequence is presented in Fig. 1.

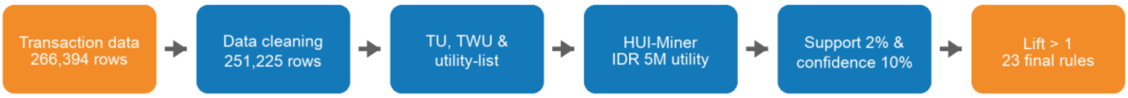

Fig 1. Research processing workflow

After high-utility itemsets were generated, transactions containing more than 25 distinct products were selected for association analysis to focus on baskets with rich combinations; this stage contained 57,765 rows. Products were transformed into binary transaction representations. Frequent itemsets were generated with minimum support of 2%, directional rules were formed with minimum confidence of 10%, and rules were retained when lift exceeded 1.

Evaluation was conducted at two levels. First, sequential filtering recorded the number of outputs after each threshold. Second, sensitivity analysis compared minimum utility values of IDR 5–20 million, support values of 2–5%, and confidence values of 10–50%. The final interpretation compared support with total utility and identified rules that may inform product placement, cross-promotion, bundling, and inventory. The thresholds and row counts reported here allow the analytical sequence to be reproduced on the same transaction structure.

4. Result and Discussions

Cleaning removed 15,169 rows, or 5.69% of the initial dataset. At the IDR 5,000,000 minimum-utility threshold, TWU screening retained 2,858 items. Utility-list calculation then produced 57 qualifying one-itemsets and 1,270 qualifying two-itemsets. During association analysis, 2% minimum support generated 248 frequent itemsets, 10% minimum confidence retained 77 rules, and lift greater than 1 retained 64 rules. Matching every criterion produced 23 final rules. Table 1 summarizes the successive processing and selection results.

Table 1. Processing and modelling summary

Stage	Result	Selection basis
Initial dataset	266,394 rows	2022 sales data
Clean dataset	251,225 rows	Nonempty item name
TWU screening	2,858 items	Utility ≥ IDR 5,000,000

Stage	Result	Selection basis
Utility-list 1-itemset	57 items	Total utility threshold
Utility-list 2-itemset	1,270 combinations	Total utility threshold
Frequent itemset	248 itemsets	Support ≥ 2%
Association rules	77 rules	Confidence ≥ 10%
Positive association	64 rules	Lift > 1
Final rules	23 rules	All criteria

Sensitivity analysis showed a monotonic pattern: increasing a threshold reduced the number of outputs. Minimum utility of IDR 5 million produced 1,270 combinations, while IDR 15 million and IDR 20 million each retained only 15. Minimum support of 2% produced 248 itemsets and decreased to 30 at 5% support. Similarly, 10% confidence produced 77 rules, whereas 50% confidence produced only 14. Table 2 presents these changes. The result is consistent with utility-mining principles in which thresholds control search-space size and pattern selectivity (Yan et al., 2024; Nugroho & Gunawan, 2025).

Table 2. Parameter sensitivity

Parameter	Threshold	Number of results
Minimum utility	IDR 5,000,000	1,270 combinations
	IDR 10,000,000	146 combinations
	IDR 15,000,000	15 combinations
	IDR 20,000,000	15 combinations
Minimum support	2%	248 itemsets
	3%	105 itemsets
	4%	52 itemsets
	5%	30 itemsets
Minimum confidence	10%	77 rules
	20%	48 rules
	30%	37 rules
	40%	22 rules
	50%	14 rules

The highest-lift rule was Sedaap Soto → Sedaap Goreng, with lift of 3.26, confidence of 58.73%, support of 6.88%, and combined utility of IDR 5,430,200. The reverse rule had identical support and lift but confidence of 38.14%, showing that directional strength depends on antecedent prevalence. Indomie Kari → Indomie Soto produced lift of 2.85, while Yakult → Bear Brand produced utility of IDR 14,344,000 with lift of 1.90. Representative rules meeting all criteria are listed in Table 3.

Table 3. Representative rules satisfying all criteria

Rule (A → B)	Utility (IDR)	Sup. (%)	Conf. (%)	Lift
Sedaap Soto → Sedaap Goreng	5,430,200	6.88	58.73	3.26
Sedaap Goreng → Sedaap Soto	5,430,200	6.88	38.14	3.26
Indomie Kari → Indomie Soto	5,931,000	5.39	38.67	2.85
Indomie Soto → Indomie Kari	5,931,000	5.39	39.73	2.85
Indomie Goreng → Indomie Soto	5,708,300	9.20	28.61	2.11
Indomie Soto → Indomie Goreng	5,708,300	9.20	67.81	2.11
Yakult → Bear Brand	14,344,000	2.51	17.88	1.90
Nice Facial → Indomie Kari	36,718,200	3.16	17.26	1.24

Comparison of all 23 rules showed that support was not proportional to total utility. The Indomie Soto and Indomie Goreng pair had the highest support, 9.20%, but generated utility of IDR 5,708,300. In contrast, Indomie Kari and Nice Facial Spack had only 3.16% support but generated utility of IDR 36,718,200. This result demonstrates a limitation of frequency-based market basket analysis: popular products do not necessarily form the combinations with the greatest economic contribution. The finding agrees with profit-support mining, which treats profit as a separate selection dimension (Dogan, 2023), and with retail assortment research that considers product value in addition to occurrence (Kiani, 2020). Fig. 2 compares support and total utility across the final rules.

Operationally, rules with high lift and adequate support may guide adjacent shelf placement or cross-promotion, particularly among instant-noodle variants. Rules with high utility but lower support are more suitable for selective bundles, cashier recommendations, or limited promotional tests. Frequently occurring products with strong economic contribution should receive stock-availability priority, while cross-category rules may be used to expand basket value. Retail visualization and analytics can support repeated monitoring of these interventions (Bakhrun et al., 2025; Shivampeta, 2025).

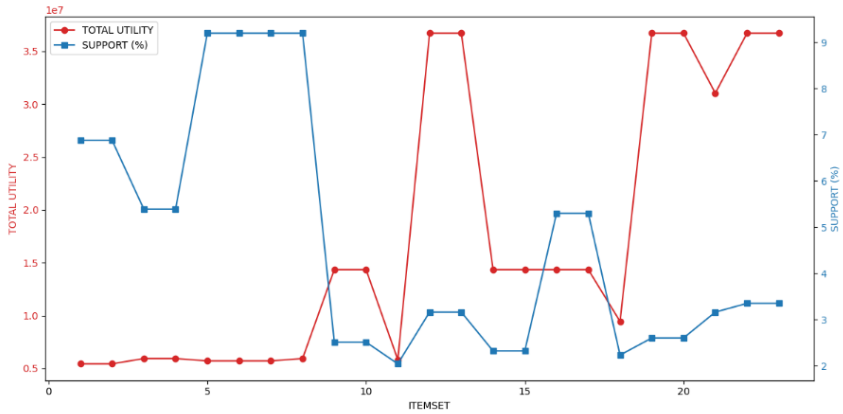

Fig 2. Total utility and support across the 23 final rules

The findings must be interpreted with three limitations. First, selling price was used as the utility proxy, so the model does not separate net margin, discounts, procurement cost, or inventory expense. Second, selecting transactions with more than 25 distinct products changed the association-analysis base and may overrepresent customers with larger baskets. Third, the study did not report execution time, memory consumption, or field-experiment validation. Therefore, the claims are limited to the relevance of the discovered patterns and do not establish computational superiority or a verified increase in sales.

5. Conclusion

This study aimed to identify product combinations that are both economically valuable and behaviorally meaningful and to determine whether frequently purchased combinations also yield the highest total utility. The objective was achieved: applying HUI-Miner to 251,225 cleaned rows and jointly applying a minimum utility of IDR 5,000,000, support of 2%, confidence of 10%, and lift greater than 1 produced 23 final rules. The findings directly answer the research question: purchase frequency was not proportional to economic contribution. The combination of Indomie Soto and Indomie Goreng recorded the highest support (9.20%) but generated utility of IDR 5,708,300, whereas the combination of Indomie Kari and Nice Facial generated IDR 36,718,200 at only 3.16% support. In addition, the directional rule from Sedaap Soto to Sedaap Goreng showed the strongest dependence (lift 3.26), and sensitivity analysis confirmed that higher utility, support, and confidence thresholds progressively reduced the number of retained patterns. Thus, the research objective of identifying high-utility, sufficiently frequent, directionally meaningful, and positively associated shopping patterns was fulfilled. Retail decisions on shelf placement, bundling, promotions, and stock should therefore combine utility, support, confidence, and lift. Future research should use net margin as external utility, evaluate segmented thresholds, compare algorithmic performance, and validate the 23 rules through controlled experiments across multiple periods and stores.

Acknowledgment

The authors thank the Institute of Business and Technology Indonesia, through the Directorate of Research, Community Service, and Innovation, for supporting the 2026 INSTIKI Research and Development Program. The authors also thank the management of YSL Grocery Store for providing the transaction data used in this study.

References

Alawadh, M. M., & Barnawi, A. M. (2022). A survey on methods and applications of intelligent market basket analysis based on association rule. Journal on Big Data, 4(1), 1-25. https://doi.org/10.32604/jbd.2022.021744

Asana, I. M. D. P., Ginantra, N. L. W. S. R., Parwita, W. G. S., Krismentari, N. K. B., & Meinarni, N. P. S. (2021). Penerapan Apriori hybrid pada transaksi penjualan. Jurnal Media Informatika Budidarma, 5(4), 1624-1632. https://doi.org/10.30865/mib.v5i4.3350

Ayu, A. A., Windarto, A. P., & Suhendro, D. (2021). Implementasi data mining dengan metode FP-Growth terhadap data penjualan barang sebagai strategi penjualan pada CV. A & A Copier. RESOLUSI: Rekayasa Teknik Informatika dan Informasi, 2(2), 67-75.

Bakhrun, A., Maghfyra, Y., Putri, R. N., & Larassati, D. A. (2025). Data visualization to analyze consumer behavior for strategic business decision making in the retail industry: Walmart case study. CSRID Journal, 17(3), 354-371. https://doi.org/10.22303/csrid-.17.3.2025.354-371

Dogan, O. (2023). A recommendation system in e-commerce with profit-support fuzzy association rule mining (P-FARM). Journal of Theoretical and Applied Electronic Commerce Research, 18(2), 831-847. https://doi.org/10.3390/jtaer18020043

Fournier-Viger, P., Lin, J. C.-W., Truong-Chi, T., & Nkambou, R. (Eds.). (2019). High-Utility Pattern Mining: Theory, Algorithms and Applications. Springer. https://doi.org/10.1007/978-3-030-04921-8

Hery, & Widjaja, A. E. (2024). Analysis of Apriori and FP-Growth algorithms for market basket insights: A case study of The Bread Basket bakery sales. Journal of Digital Market and Digital Currency, 1(1), 63-83. https://doi.org/10.47738/jdmdc.v1i1.2

Kiani, G. H. (2020). Determining profitable products in the retail market with consideration of cash limitation and exhibition periods. Journal of Retailing and Consumer Services, 55, 102079. https://doi.org/10.1016/j.jretconser.2020.102079

Krismentari, N. K. B., Asana, I. M. D. P., Herdiana, I. K., Anandita, I. B. G., & Syarifah, N. F. (2026). A binary particle swarm optimization-based association rule mining model for retail transactions. Krisnadana Journal, 5(2), 280-293. https://doi.org/10.58982/krisnadana.v5i2.1064

Mai, T., Nguyen, L. T. T., Vo, B., Yun, U., & Hong, T.-P. (2020). Efficient algorithm for mining non-redundant high-utility association rules. Sensors, 20(4), 1078. https://doi.org/10.3390/s20041078

Mohammed, Y. F., Ibrahim, A. A., Hamdi, M. I., Abdullah, S. K., & Hussein, Z. K. (2024). A comprehensive review of AI-driven data mining techniques. Al-Noor Journal for Information Technology and Cyber Security, 1, 49-58. https://doi.org/10.69513/jnfit.v1.i0.a3

Nata, G. N. M., Bagiarta, I. G. N. N., Antarajaya, I. N. S., & Effendy, M. (2022). Frequent pattern untuk katalog digital. NARATIF: Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika, 4(1).

Nugroho, R. T., & Gunawan, R. (2025). Pencarian high utility itemset pada dataset YooChoose Buys. Prosiding KONSTELASI, 2(1), 64-73.

Rahmi, A. N., & Mikola, Y. A. (2021). Implementasi algoritma Apriori untuk menentukan pola pembelian pada customer: Studi kasus Toko Bakoel Sembako. Information System Journal, 4(1), 14-19. https://doi.org/10.24076/infosjournal.2021v4i1.561

Shivampeta, S. (2025). Leveraging big data for personalized retail engagement: A rewards analytics platform architecture. Sarcouncil Journal of Engineering and Computer Sciences, 4(8), 671-679. https://doi.org/10.5281/zenodo.16895491

Siregar, D. R. M., Sonata, F., & Yetri, M. (2022). Analisa pola belanja untuk meningkatkan omset penjualan menggunakan algoritma Apriori. Jurnal Sistem Informasi TGD, 1(6), 876-887. https://ojs.trigunadharma.ac.id/index.php/jsi/article/view/6488

Shu, X., & Ye, Y. (2023). Knowledge discovery: Methods from data mining and machine learning. Social Science Research, 110, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817

Solanki, M. S., & Sharma, A. (2021). A review of data mining techniques and its applications. International Journal of Innovative Research in Computer Science & Technology, 9(6), 100-104. https://doi.org/10.55524/ijircst.2021.9.6.23

Wahyuningsih, R., Suharsono, A., & Iriawan, N. (2023). Comparison of market basket analysis method using Apriori algorithm, FP-Growth and ECLAT: Supermarket X transaction data for 2021. Business and Finance Journal, 8(2), 192-201.

Widyan, A., & Rozi, A. F. (2021). Analisis rekomendasi produk menggunakan algoritma ECLAT berdasarkan riwayat data penjualan PT XYZ. Jurnal Teknologi dan Sistem Informasi Bisnis, 3(2), 395-411. https://doi.org/10.47233/jteksis.v3i2.296

Yan, Y., Niu, X., Zhang, Z., Fournier-Viger, P., Ye, L., & Min, F. (2024). Efficient high utility itemset mining without the join operation. Information Sciences, 665, 120392. https://doi.org/10.1016/j.ins.2024.120392