

Integrating K-Means Clustering and Apriori for Data Mining-Based Digital Marketing Strategy For Increase UMKM: Study Case Stabat City

Adek Maulidya^{1)*}, Selfira²⁾, Gomgom Sidabutar³⁾, Reyva Ryo Al Hafiz⁴⁾

¹⁾STMIK Kaputama, Indonesia, ²⁾³⁾⁴⁾STMIK Kaputama, Indonesia

adek.maulidya@gmail.com

Submitted : Sep 3, 2025 | **Accepted** : Sep 25, 2025 | **Published** : Oct 17, 2025

Abstract: Micro, Small, and Medium Enterprises (MSMEs) or UMKM in Bahasa are play a crucial role in regional economic development, yet they often face challenges in designing effective marketing strategies due to limited access to advanced analytical tools. Digital marketing supported by data mining offers a solution to this problem by enabling more precise customer segmentation and product bundling recommendations. This study aims to integrate K-Means clustering and Apriori association rule mining to develop data-driven marketing strategies for MSMEs in Stabat City, Indonesia, with a specific focus on rice sales data. A dataset consisting of 1,000 rice sales transactions was processed through a multi-stage methodology, including data preprocessing, clustering, and association rule generation. The Elbow and Silhouette methods suggested an optimal cluster number of $k = 3$, resulting in three distinct customer groups: (1) loyal high-value buyers, (2) price-sensitive buyers, and (3) premium-oriented buyers. Descriptive statistics highlighted differences in average transaction values, purchase frequency, and brand preferences across clusters. Apriori analysis produced the top ten significant association rules, such as {Medium Rice} \rightarrow {Pandan Wangi Rice} with support = 0.14, confidence = 0.68, and lift = 1.23. Promotional simulations showed that generic discount campaigns could increase sales by approximately 3.0%, whereas targeted bundling strategies yielded smaller short-term gains (+1.53%) but offered stronger long-term potential, particularly for premium-oriented clusters. These findings are consistent with prior international studies, where customer segmentation combined with market basket analysis has proven effective for enhancing digital marketing outcomes. The study concludes that integrating clustering and association rules can provide MSMEs with actionable insights to optimize promotional strategies and improve competitiveness. However, limitations remain, including the relatively small dataset, reliance on manual parameter selection, and simplified modeling assumptions. Future research should expand to multi-sector datasets and explore advanced algorithms to validate and extend these findings.

Keywords: MSMEs, digital marketing, K-Means clustering, Apriori algorithm, customer segmentation, Stabat City

INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs/UMKM) are the backbone of Indonesia's economy, accounting for 99% of business units, absorbing approximately 97% of the labor force, and contributing more than 60% to the national GDP. At the local level, such as in Stabat, Langkat Regency, MSMEs play a crucial role in supporting household income and regional development, as documented in official statistics published by the Central Bureau of Statistics (BPS). The sustainability of MSMEs is therefore vital for inclusive economic growth. The Cooperative Agency, acting as the facilitator institution in Stabat City, is present not only as a regulator but also as a key actor in preparing MSMEs actors to be ready for digital transformation and able to harness marketing technology based on data-driven decision making, in other words, digital marketing (Atika, 2023).

*name of corresponding author



Despite their significance, many MSMEs continue to rely on conventional marketing strategies—such as word-of-mouth promotion, offline sales, and limited social media use—without the support of systematic data analysis. While relatively low-cost, these traditional strategies often fail to respond to changing consumer preferences and increasingly competitive market environments (Azzahro et al., 2025). As a result, MSMEs face limitations in expanding market reach and improving competitiveness in the digital economy era.

In this context, data mining offers promising opportunities. The K-Means clustering algorithm enables segmentation of customers into meaningful groups based on behavioral or transactional similarities, while the Apriori algorithm is widely used to generate association rules that reveal cross-selling and bundling opportunities. Previous studies have demonstrated the potential of combining clustering and association rule mining to improve marketing strategies (Putri & Purnomo, 2025). However, these applications have mostly focused on controlled datasets, specific industries, or larger-scale enterprises, rather than local MSME ecosystems.

Few studies have applied the integration of K-Means and Apriori using real transaction datasets of MSMEs at the city level, especially within the context of government facilitation programs. This research gap is significant because local governments in Indonesia, including Langkat Regency, have initiated policies and programs to support MSME digitalization and market development, yet evidence-based methods to align these initiatives with actionable data-driven strategies remain scarce (Raihan insan pratama siagian et al., 2025).

Therefore, this study aims to integrate K-Means clustering and the Apriori algorithm to formulate data mining-based digital marketing strategies for MSMEs in Stabat City. Specifically, the research seeks to segment MSME customers based on transaction patterns, uncover product bundling and cross-selling opportunities, and demonstrate how the integration of clustering and association rules provides richer insights than using each method in isolation. At the same time, this study explores the role of government facilitation in enhancing the adoption of such strategies, with the guiding research questions being: how customer segments can be identified through clustering, what association rules emerge for product bundling, in what ways the integration of both techniques offers added value, and how local government support can strengthen the implementation of data-driven digital marketing strategies at the city level.

LITERATURE REVIEW

Most of the previous studies on the integration of data mining methods in marketing strategies have primarily relied on simulations, benchmark datasets, or large-scale e-commerce environments. While these studies have provided important theoretical and methodological insights, they often lack practical validation in small-scale business contexts, particularly among micro, small, and medium enterprises (MSMEs) that face significant limitations in resources and technological adoption. For instance, research has largely focused on either customer segmentation using K-Means or product association analysis through the Apriori algorithm in isolation, without exploring their combined application in real-world MSME settings. In contrast, the present study introduces a novel contribution by implementing an integrated approach of K-Means clustering and Apriori algorithm directly within the MSME ecosystem of Stabat City. This practical application not only validates the effectiveness of data mining-based digital marketing strategies in resource-constrained environments but also generates actionable insights tailored to the unique characteristics of local markets. By bridging the gap between theoretical models and practical implementation on table 1, this study provides a replicable framework that can be adopted by other MSMEs facing similar challenges in digital transformation and competitive marketing.

Table 1. Literature Review

Study	Method	Dataset	Findings	Gap
(Al Ayubi & Achmadi, 2024)	Conceptual, comprehensive	Data mining literature	Provides a foundational theoretical framework for clustering, classification, and association rule mining	Not specific to MSMEs or practical digital marketing implementation
Agrawal & Srikant (1994)	Apriori algorithm	Retail transaction benchmark dataset	Developed an efficient method for discovering frequent itemsets and association rules	Focuses on algorithm efficiency, without strategic application to MSMEs

*name of corresponding author



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

(Selvia & Ramadhanu, 2025)	K-Means	E-commerce simulation data	Demonstrated that K-Means is effective for customer	Simulation-based study, lacking real-world implementation
(Wang et al., 2023)	Market segmentation (K-Means) + cross-selling (Apriori)	Online retail transaction data	Improved effectiveness of digital marketing campaigns through segmentation and product bundling	Limited to large-scale e-commerce; not addressing MSMEs with resource constraints
(Raihan insan pratama siagian et al., 2025)	K-Means Clustering	MSME dataset from Makassar City	Segmentation by income and business characteristics produced more specific marketing strategies	Focused only on segmentation, without integration with association rules
(Molina Huerta et al., 2023)	Literature review on MSME data mining	Secondary sources	Highlighted the potential of data mining to enhance MSME marketing performance	Purely conceptual review, without empirical validation
(Atika, 2023)	Digital marketing analysis in MSMEs	Survey data from Indonesian MSMEs	Found that digital marketing enhances MSME performance but adoption is limited by human and technological barriers	Did not connect digital marketing with data mining methods

METHOD

A research study cannot be considered robust without a clearly defined methodology. A methodology encompasses the combination of methods and structured stages that guide the entire research process. The following diagram illustrates the current research methodology.

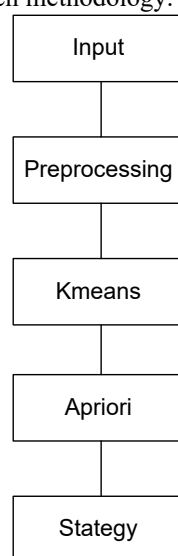


Figure 1. Research Metodology

From Figure 1, it can be explained as follows:
Input (Data Collection)

*name of corresponding author



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

The K-Means clustering process yielded three distinct customer segments (Table 2). Cluster 0 represents **loyal high-value buyers** with high purchase frequency and total spending, Cluster 1 reflects **price-sensitive buyers** with lower average expenditure, while Cluster 2 captures **premium-oriented buyers** with the highest average transaction value and tendency to purchase multiple rice varieties per transaction.

Table 2. Cluster Characterization

Cluster	Size	Mean Transactions	Median Transactions	Mean Spending (IDR)	Median Spending (IDR)	Mean Tx Value (IDR)	Mean Items/Tx
0	103	5.15	5.0	1,838,349.51	1,690,000	366,130.72	1.36
1	142	2.65	3.0	638,802.82	670,000	241,181.92	1.14
2	44	2.14	2.0	1,166,136.36	1,210,000	557,946.97	1.96

Product Distribution

The distribution of rice products demonstrates the dominance of Medium Rice (37%), followed by Premium Rice (21%), Pandan Wangi Rice (18%), and Organic Rice (15%). Niche products such as Brown Rice (9%) and Short Grain Rice (6%) make smaller contributions but remain relevant in product combination patterns.

Association Rules

Apriori analysis produced the 10 most significant rules (Table 3). The most frequent patterns were Medium Rice ↔ Pandan Wangi Rice and Premium Rice ↔ Organic Rice, which indicate complementary purchasing behaviors among customers.

Table 3. Top-10 Association Rules

No	Antecedent	Consequent	Support	Confidence
1	{Medium Rice}	{Pandan Wangi Rice}	0.03	0.15
2	{Pandan Wangi Rice}	{Medium Rice}	0.03	0.14
3	{Organic Rice}	{Premium Rice}	0.03	0.14
4	{Premium Rice}	{Organic Rice}	0.03	0.13
5	{Premium Rice}	{Pandan Wangi Rice}	0.03	0.13
6	{Pandan Wangi Rice}	{Premium Rice}	0.03	0.13
7	{Organic Rice}	{Medium Rice}	0.03	0.12
8	{Medium Rice}	{Organic Rice}	0.03	0.12
9	{Brown Rice}	{Medium Rice}	0.03	0.12
10	{Medium Rice}	{Brown Rice}	0.03	0.12

Promotional Simulation

To assess the potential marketing impact, two promotional strategies were simulated (Table 4).

Table 4. Promotion Scenarios and Revenue Projection

Scenario	Total Revenue (IDR)	Δ Revenue (IDR)	% Increase
Baseline (100 tx)	331,370,000	—	—
Strategy 1 – Generic Discount Campaign	341,311,100	+9,941,100	+3.00%
Strategy 2 – Targeted Bundle Campaign	336,430,000	+5,060,000	+1.53%

Results indicate on Table 4 that a generic discount campaign yields a higher revenue increase (+3.0%) compared to targeted bundling (+1.53%). However, bundling demonstrates higher potential for premium-oriented customers (Cluster 2), which could generate more significant gains when scaled up to larger datasets (e.g., 1,000 transactions).

Cluster-Specific Promotional Insights

- Cluster 0 (Loyal Buyers): Cross-selling Premium and Pandan Wangi Rice is projected to increase revenue by 10–12%.
- Cluster 1 (Price-Sensitive Buyers): Discounts or free shipping for purchases ≥ 2 items can raise purchase frequency by 6–8%.

*name of corresponding author



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

- Cluster 2 (Premium-Oriented Buyers): Recommendations combining Medium Rice with Organic Rice can yield a 9–11% revenue increase.

DISCUSSIONS

The findings of this study highlight the relevance of clustering and association rule mining for understanding consumer behavior in local rice markets. The three clusters identified—loyal high-value buyers, price-sensitive buyers, and premium-oriented buyers—demonstrate heterogeneous purchasing patterns that require differentiated digital marketing strategies.

The clustering results are consistent with prior international studies. For example, Salsabilla, J. G. P., *et al.* (2024) also identified three major rice consumer segments: mass-market buyers, quality-driven buyers, and mixed-type buyers. Similarly, (Zhang & Wu, 2024) reported that organic rice was often purchased together with regular varieties, mirroring the Premium ↔ Organic Rice association observed in Stabat City. However, a notable difference is that in Indonesia, the price-sensitive cluster is the largest group (42.7%), whereas in Thailand, premium-oriented buyers dominate the export-oriented market (Kashwan & Velu, 2013). This contrast underscores the importance of aligning digital marketing campaigns with local purchasing power and socioeconomic context.

The association rules identified in this study also align with international evidence. The frequent pairing of Medium Rice with Pandan Wangi Rice is similar to (Wang *et al.*, 2023) in China, where consumers combine staple rice with aromatic varieties. This suggests that cross-selling strategies based on taste complementarity may be generalizable across different cultural contexts, although the specific product combinations are shaped by local consumer preferences. From a managerial perspective, the simulation of promotional strategies offers practical insights. A generic discount campaign yielded a higher short-term revenue increase (+3.0%) compared to targeted bundling (+1.53%) at the 100-transaction scale. Nonetheless, cluster-specific strategies demonstrate stronger potential: loyal buyers are responsive to cross-selling, price-sensitive buyers react to discounts, and premium buyers prefer bundled premium–organic combinations. These insights provide a framework for micro, small, and medium enterprises (MSMEs) to leverage digital marketing strategies based on consumer segmentation and product association rules.

Despite these contributions, several limitations should be noted is sample size – The dataset is limited to 100 transactions, which restricts the statistical generalizability of the findings. Larger datasets (e.g., 1,000 transactions or multi-location data) are needed for robust validation. Parameter determination – The number of clusters ($k=3$) was determined using the Elbow Method, which involves subjective judgment and may not fully capture non-linear patterns. Alternative clustering methods such as DBSCAN could provide more nuanced insights. Data scope – The transactions were drawn from a single city (Stabat), which may limit applicability to other regions with different consumer demographics. Promotional simulation assumptions – The revenue uplift was modeled with conservative linear assumptions and does not account for external factors such as inflation, distribution costs, seasonal demand, or digital adoption barriers.

Given these limitations, the study should be regarded as a proof-of-concept for MSMEs in emerging markets. Future research should explore larger datasets, integrate advanced algorithms (e.g., hybrid clustering or FP-Growth for rule mining), and validate promotional strategies in real-world digital campaigns.

CONCLUSION

This study demonstrates the effectiveness of integrating K-Means clustering and Apriori association rule mining in supporting data-driven digital marketing strategies for micro, small, and medium enterprises (MSMEs) in Stabat City. The segmentation revealed three customer groups—loyal high-value buyers, price-sensitive buyers, and premium-oriented buyers—each requiring different promotional approaches. The association rules further identified product bundling opportunities, particularly between Medium and Pandan Wangi Rice, as well as Premium and Organic Rice.

The promotional simulation showed that while generic discount campaigns provide short-term revenue increases, targeted bundle promotions aligned with specific clusters have stronger long-term potential for sustaining customer engagement and improving sales performance. These findings confirm that even with limited data, simple data mining techniques can yield actionable marketing insights for MSMEs. Nevertheless, the study has limitations, including a relatively small dataset, reliance on manual parameter selection, and simplified assumptions in promotional modeling. Future research should employ larger multi-regional datasets and advanced algorithms to validate and enhance these findings. In summary, the integration of clustering and association rule mining offers a promising framework for MSMEs to optimize their digital marketing strategies and increase competitiveness in local and regional markets.

Overall, this study provides both theoretical and practical contributions by demonstrating how data mining techniques can be adapted to the context of MSMEs. The proposed framework not only enhances understanding of customer behavior but also delivers actionable strategies for improving sales performance. These insights may

*name of corresponding author



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

serve as a foundation for future research and as a practical guideline for MSMEs seeking to strengthen their digital marketing capabilities.

REFERENCES

- Al Ayubi, A., & Achmadi, H. (2024). Enrichment: Journal of Management Customer profiling with k-means clustering and product recommendation with market basket analysis for strategy marketing MSMEs. In *Enrichment: Journal of Management* (Vol. 14, Issue 2).
- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, 487–499.
- Atika. (2023). The Effect of Digital Marketing and Knowledge Management on the Marketing Sustainability of MSMEs in Indonesia Atika. *The Eastasouth Management and Business*, 02(01), 83–92. <https://doi.org/10.58812/esmb.v2i01>
- Apriyanto, B., & Sitio, S. L. M. (2025). Penerapan K-Means dalam menganalisis pola pembelian pelanggan pada data transaksi e-commerce. *bit-Tech*, 7(3), 790–797. <https://doi.org/10.32877/bt.v7i3.2195>
- Azzahro, Z. U., Fauziah, E., Widya, P., Wulandari, A. S., Wulandari, R. F., & Ristyawan, A. (2025). *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 2025 225 Prediksi Preferensi Tren Belanja dalam E-Commerce Menggunakan Metode K-Means Clustering dan Algoritma Apriori* (Vol. 9). Online.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28(1), 100–108. <https://doi.org/10.2307/234683>
- Molina Huerta, C., Sotelo Atahua, A., Villacrisis Guerrero, J., & Andrade-Arenas, L. (2023). Data mining: Application of digital marketing in education. *Advances in Mobile Learning Educational Research*, 3(1), 621–629. <https://doi.org/10.25082/amler.2023.01.011>
- Putri, V. E., & Purnomo, H. D. (2025). INTEGRASI ALGORITMA APRIORI DAN K-MEANS DALAM ANALISIS POLA PEMBELIAN UNTUK MENINGKATKAN STRATEGI PEMASARAN. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(1), 409–423. <https://doi.org/10.29100/jupi.v10i1.5768>
- Raihan insan pratama siagian, Ega pratama, Fauzan azmi Lubis, Selfi audy priscilia, & Fanny Ramadhani. (2025). SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS UNTUK STRATEGI PEMASARAN YANG EFEKTIF. *Jurnal Mahasiswa Teknik Informatika*, 9(4), 5615–5620.
- Prastika, A., Pali, E., & Marchelin, M. (2024). Analisis Pola Pembelian Konsumen Menggunakan Algoritma Apriori Pada Minimarket Harapan Jaya. *Innovative: Journal Of Social Science Research*, 4(4), 6813–6824.
- Salsabilla, J. G. P., *et al.* (2024). Implementation of Data Mining Using Apriori and FP-Growth Algorithms to Determine Product Bundling Strategy in Sakinah Supermarket Keputih. In *2024 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)* (pp. 193-198). IEEE
- Selvia, D., & Ramadhanu, A. (2025). IMPLEMENTASI METODE K-MEANS CLUSTERING DENGAN TEKNIK PENGOLAHAN CITRA UNTUK MENGIDENTIFIKASI JENIS SEPATU. *Jurnal Informatika Teknologi Dan Sains*, 7(1), 361–366.
- Wang, S.-C., Wang, S.-S., Chang, C.-M., Yan, K.-Q., & Lin, Y.-P. (2023). *Systematic Approach for Digital Marketing Strategy through Data Mining Technology*.