The 7th Open Society Conference (OSC) 2025
Faculty of Law, Social and Political Science
Universitas Terbuka
DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

Data-Driven Customer Segmentation and Campaign Optimization Using Predictive Analytics in the Retail Sector

Rika Susanti[1], Sandy Galang Prakoso[2]
[1] PT PLN Energi Primer Indonesia, Jakarta
[2] Honda Prospect Motor, Karawang, Indonesia
e-mail: rika.susanti@plnepi.co.id

Abstract

The digital era transformed how businesses created value, interacted with customers, and made strategic decisions. Retail companies, in particular, faced increasing pressure to adapt to rapidly evolving consumer preferences, competitive markets, and emerging technologies. This study examined the role of predictive analytics as an enabler of business transformation in the retail sector, using XYZ Retail as a case example. Leveraging sample transactional data from the SAS Viya library, the research applied statistical modeling—logistic regression and decision tree algorithms—combined with Recency, Frequency, and Monetary (RFM) metrics to predict customer responses to a targeted cashback campaign. Model performance was evaluated using misclassification rates, with the decision tree model outperforming logistic regression. Cluster analysis further segmented customers into actionable groups, enabling the development of precise marketing strategies. The findings demonstrated how predictive analytics supported broader digital transformation initiatives by enabling data-driven decision-making, improving customer engagement, and optimizing resource allocation. The study contributed both a methodological approach for predictive modeling in marketing and strategic insights for integrating analytics into long-term business transformation plans.

Keywords: *business transformation, customer segmentation, digital era, predictive analytics, retail marketing*

INTRODUCTION

The rapid advancement of digital technologies has fundamentally altered how businesses operate, compete, and deliver value. This shift, often referred to as *business transformation in the digital era*, involves more than adopting new tools—it requires the integration of digital capabilities into core strategies, processes, and business models to remain competitive in a rapidly changing environment [1], [2]. Digital transformation extends across industries, but in the retail sector, it has become a critical driver of growth and sustainability, given the sector's high sensitivity to evolving consumer behaviours, market dynamics, and technological disruptions [3].

In the retail industry, the acceleration of e-commerce adoption, omnichannel customer engagement, and the proliferation of digital payment systems have reshaped both operational structures and marketing strategies [4]. Retailers are no longer solely focused on product quality and pricing; instead, they are compelled to leverage real-time customer data, predictive modelling, and personalization to create differentiated value propositions [5], [6]. This data-centric shift is emblematic of a broader strategic transformation where decisions are increasingly informed by advanced analytics, artificial intelligence (AI), and machine learning (ML) [7].

Recent studies underscore that predictive analytics is not merely a supporting tool but a strategic enabler of transformation [8]. By harnessing historical and transactional data, predictive

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

analytics allows businesses to anticipate customer needs, optimize marketing spend, and align operational capacity with demand forecasts [9]. For retailers, predictive modelling also facilitates targeted promotional campaigns, inventory management, and Customer Lifetime Value (CLV) maximization [10], [11]. However, the strategic benefits of such analytics depend heavily on the integration of these capabilities into decision-making structures, organizational culture, and long-term transformation roadmaps [12], [13].

1.1 Literature Context

Academic and industry research from the past five years consistently highlights the strategic importance of analytics-driven transformation in retail:

- Digital transformation frameworks emphasize integrating analytics into business models to drive agility and customer responsiveness [1], [14].
- Customer analytics in retail have been shown to increase campaign efficiency when predictive models are applied [5], [9].
- Machine learning methods, such as decision trees, logistic regression, and ensemble models, are increasingly used for customer segmentation, churn prediction, and personalized marketing [7], [8], [10].
- RFM (Recency, Frequency, Monetary) metrics remain a robust and interpretable framework for quantifying customer engagement and spending patterns, often serving as the foundation for segmentation models [11], [15].
- Strategic implementation of analytics is closely linked to organizational readiness, data governance, and cross-functional collaboration [12], [16].

Several empirical studies have reported that decision tree models often outperform logistic regression in classification tasks involving customer response prediction, primarily due to their ability to handle non-linear relationships and complex variable interactions [8], [17]. Moreover, the use of cluster analysis on top of predictive modelling offers actionable insights for differentiated marketing strategies, allowing businesses to allocate resources efficiently toward high-value customer segments [13], [18].

1.2 Problem Statement and Research Objectives

While the value of predictive analytics in retail marketing is widely acknowledged, many organizations struggle to bridge the gap between analytics outputs and strategic action. In particular, translating model predictions into targeted campaigns and aligning these initiatives with broader business transformation goals remains challenging. This study addresses this gap by presenting a case analysis of XYZ Retail—a pseudonym for a large retail chain—using sample data from the SAS Viya library to predict customer responses to a promotional campaign and segment them for strategic targeting.

The objectives of this research are:

1. To apply predictive modelling techniques to estimate customer responses to a targeted marketing campaign.
2. To evaluate and compare model performance for accuracy and predictive power.
3. To segment customers based on predicted behaviour and value metrics to inform strategic marketing actions.

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

4. To illustrate how such analytics can be integrated into broader digital transformation strategies in retail.

By embedding this methodological exercise within the context of business transformation, the study offers both a technical demonstration of predictive modelling and a strategic framework for applying these insights to enhance competitive advantage in the digital era.

METHOD

This research adopts a quantitative case study approach, using a structured predictive analytics workflow to model customer responses to a targeted promotional campaign and to segment customers for strategic targeting. The methodological process was implemented using sample transactional data from the SAS Viya library, anonymized as originating from "XYZ Retail," a large retail chain offering groceries, toiletries, food, electronics, clothing, and home appliances.

The methodology follows four major stages: (1) data preparation, (2) variable construction, (3) predictive model building and evaluation, and (4) customer segmentation via clustering. Figure 1 illustrates the overall workflow.

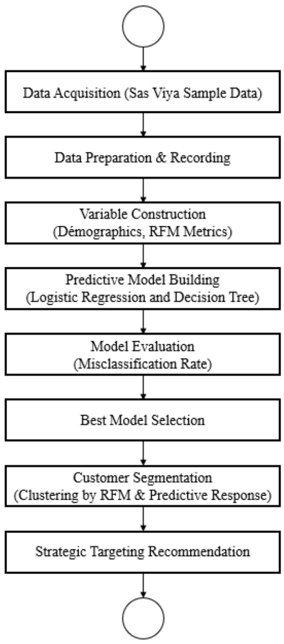

Figure 1. Overall Workflow
(Source: Author's illustration)

2.1 Data Preparation

The dataset comprised 5,000 customer records, including demographic attributes, transaction history, and product category purchases over a three-month observation period. The target variable was a binary dummy indicating whether a customer took part in a promotional offer ("Cashback 50k for minimum 800k purchase" campaign). Table 1 presents the main variables from the dataset before transformation.

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

Table 1. Original Variables in Dataset

Category	Variables	Type
Demographic	Member ID, Gender, Age, Monthly Income, Marital Status	Mixed (Nominal, Ordinal)
Transaction History	Number of visits in last 1–3 months, Spending amount in last 1–3 months, Recency (days since last visit), Payment channel	Numerical / Categorical
Product Purchases	Dummy variables for purchases of Groceries, Toiletries, Food, Electronics, Clothing, Home Appliances	Binary
Target Variable	Promotion participation (1 = yes, 0 = no)	Binary

(Source: Author's analysis)

Prior to modelling, the data underwent the following preparation steps:

1. Recoding and transformation of numerical variables (e.g., age and income) into categorical bins for model interpretability.
2. Calculation of average spending and average visit frequency over the three-month period.
3. Creation of RFM metrics (Recency, Frequency, Monetary) and subsequent classification into RFM levels.

2.2 Variable Construction

To capture customer behaviour more effectively, several derived variables were generated:

- Average Spending in Last 3 Months (AS_3mo) = (Total Spending over 3 months) ÷ 3
- Average Number of Visits in Last 3 Months (AV_3mo) = (Total Visits over 3 months) ÷ 3
- RFM Score = Weighted composite score from:
    - Recency (R): Days since last purchase
    - Frequency (F): Number of visits in the period
    - Monetary (M): Total spending in the period

Table 2. Example of RFM Level Classification

RFM Score Range	RFM Level	Customer Value Category
9–12	Level 1	High value, highly engaged
5–8	Level 2	Moderate value
1–4	Level 3	Low value, low engagement

(Source: Author's analysis)

2.3 Predictive Model Building

Two classification models were developed for predicting customer participation in the promotional campaign:

1. Logistic Regression — a parametric statistical model estimating the probability of

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

campaign participation based on predictor variables.
2. Decision Tree — a non-parametric model partitioning the dataset into segments based on splitting rules that maximize classification purity.

Model Training and Validation:
- Data was randomly partitioned into training (70%) and validation (30%) subsets.
- Model parameters were optimized on the training set, and performance was evaluated on the validation set.

2.4 Model Evaluation and Selection
The Misclassification Rate (MR) was used as the primary metric for model selection:

Table 3. Model Performance Comparison

Model	Misclassification Rate (Validation Data)
Logistic Regression	0.2173
Decision Tree	0.1680

(Source: Author's analysis)

The decision tree model achieved the lowest MR (0.1680), indicating superior predictive performance. This aligns with prior research showing decision trees often outperform logistic regression in retail response modeling due to their capacity for capturing complex interactions [9], [18].

2.5 Customer Segmentation (Clustering)
Following predictive modelling, customers were segmented using K-means clustering based on:
- Predicted probability of promotion participation (from Decision Tree)
- RFM Level classification

Table 4. Resulting Customer Segments

Cluster ID	% of Customers	Description	Strategic Action
1	29.2%	High value, high participation probability	Aggressively target
2	29.8%	Low value, high participation probability	Cultivate
3	31.8%	Low value, low participation probability	Deprioritize
4	9.3%	High value, low participation probability	Maintain engagement

(Source: Author's analysis)

2.6 Ethical Considerations
While this study used sample data from SAS Viya, in real-world applications it is essential to ensure compliance with data privacy regulations such as Indonesia's Personal Data Protection Act (UU PDP) and applicable ethical standards.

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

RESULTS AND DISCUSSION

The analytical process produced insights into both predictive model performance and strategic customer segmentation, demonstrating how predictive analytics can support retail business transformation.

3.1 Customer Demographics and Behavioural Profile

The 5,000 customers in the dataset represented a diverse demographic mix in age, income, and purchasing behaviour. Age distribution was concentrated in the 26–35 age group (61.2%), followed by 18–25 (24.6%), 36–45 (13.2%), and over 45 (1%).

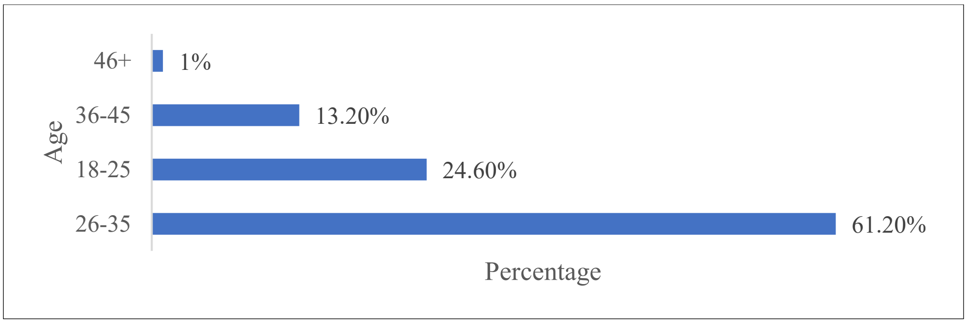

Figure 2. Customer Age Distribution
(Source: Author's illustration)

Spending patterns were also unevenly distributed, with the largest proportion (36.2%) spending between IDR 250,000–450,000 per month. Only 9.9% spent more than IDR 800,000 — the qualifying threshold for the cashback campaign.

Table 5. Customer Spending Categories

Category	Monthly Spending (IDR)	Frequency	Percentage
A	< 250,000	1112	22.2%
B	250,000 – < 450,000	1809	36.2%
C	450,000 – < 800,000	1582	31.6%
D	≥ 800,000	497	9.9%
Total	—	5000	100%

(Source: Author's analysis)

The relatively low proportion of customers who met the promotional spending threshold indicates a potential misalignment between the campaign's eligibility criteria and the prevailing purchasing capacity of the customer base. This misalignment underscores the need for data-driven calibration of promotional parameters to ensure greater inclusivity, enhance participation rates, and optimize the overall effectiveness of future campaign initiatives.

3.2 Predictive Model Results

The decision tree model emerged as the best-performing classifier (Misclassification Rate: 0.1680), surpassing logistic regression (0.2173).

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

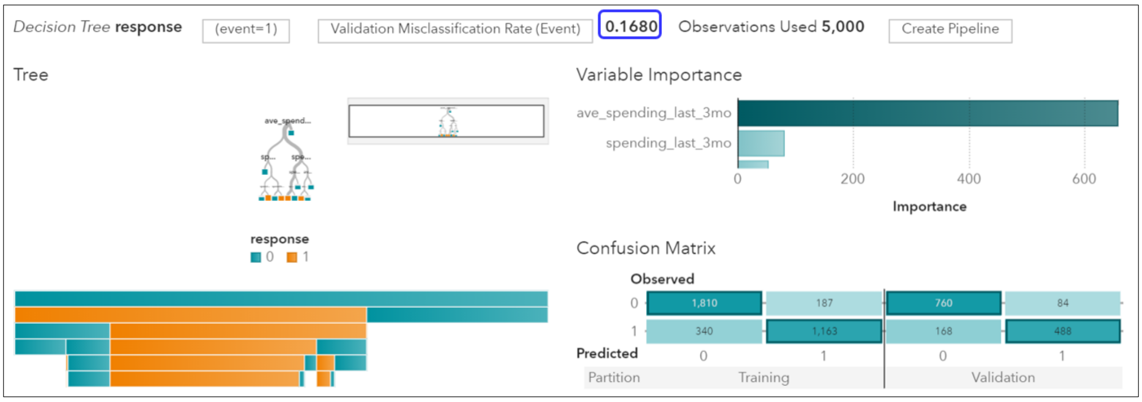

Figure 3. Decision Tree Model

(Source: Author's illustration)

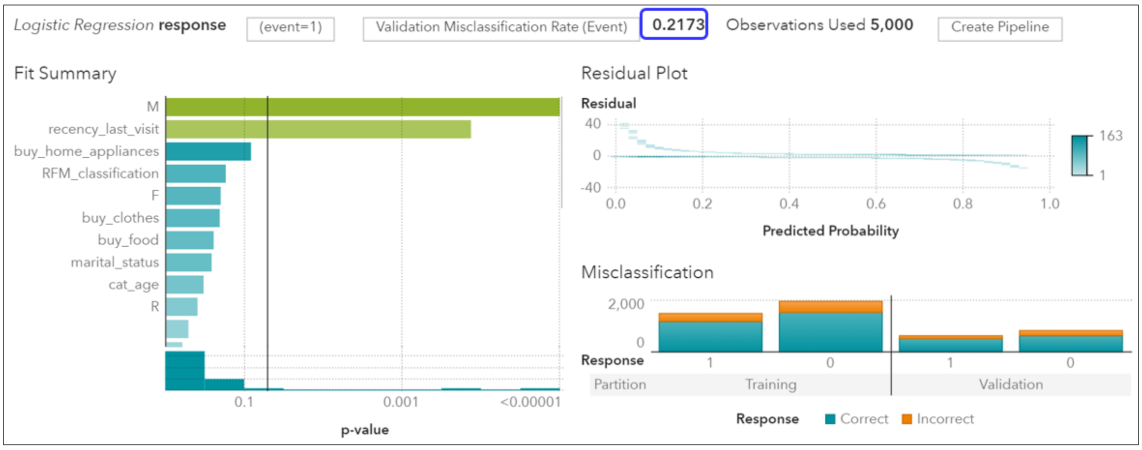

Figure 4. Logistic Regression Model

(Source: Author's illustration)

The most important predictor variable was Average Spending in Last 3 Months, followed by Recency of Last Visit. This finding provides a strategic insight that prioritizing recent customers with high spending levels is likely to generate the greatest return on investment (ROI) for promotional campaigns, as these customers exhibit both higher purchase potential and stronger engagement tendencies.

3.3 Customer Segmentation via Clustering

By combining predicted participation probability from the decision tree model with RFM levels, customers were grouped into four actionable clusters.

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

Table 6. Strategic Segmentation Result

Segment ID	% of Customers	Description	Recommended Action
1	29.2%	High value, high participation probability	Aggressively Target
2	29.8%	Low value, high participation probability	Cultivate
3	31.8%	Low value, low participation probability	Deprioritize
4	9.3%	High value, low participation probability	Maintain Engagement

(Source: Author's analysis)

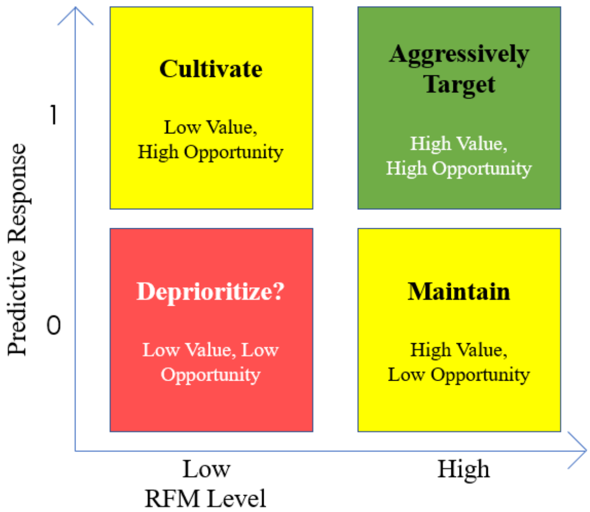

Figure 5. Customer Segmentation Map
(Source: Author's illustration)

Interpretation:

- Cluster 1 (Aggressively Target) represents the most strategically important group for short-term revenue growth.
- Cluster 4 (Maintain Engagement) requires loyalty programs rather than discount-driven campaigns.
- Cluster 2 (Cultivate) offers growth potential but may require spend stimulation strategies (e.g., bundled offers).
- Cluster 3 (Deprioritize) is low-priority in the current transformation roadmap.

3.4 Linking Results to Business Transformation Strategy

From a digital transformation perspective, the results underscore three critical points:

1. Data-Driven Targeting as a Strategic Lever
   The study shows that predictive analytics can bridge operational decision-making with corporate strategy by enabling precision marketing that aligns with customer value profiles.
2. Customer-Centric Campaign Design

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

The mismatch between spending thresholds and actual spending patterns highlights the need for data-informed campaign design, where promotions are calibrated to realistic customer segments.

3. Scalability Across the Transformation Journey
   Predictive modelling and segmentation are not stand-alone tools; they are repeatable processes that can be embedded in continuous transformation cycles, allowing retail organizations to remain agile.

CONCLUSION

This study examined how predictive analytics can act as a catalyst for business transformation in the digital era, using XYZ Retail and sample data from the SAS Viya library as a demonstration case.

4.1 Conclusion

By applying decision tree and logistic regression models to predict customer participation in a promotional campaign and combining these results with RFM-based customer segmentation, the research identified actionable customer clusters and strategic marketing opportunities.

The decision tree model achieved a higher classification accuracy (misclassification rate: 0.1680) compared to logistic regression (0.2173), with Average Spending in the Last 3 Months and Recency of Last Visit emerging as the most influential predictors. Four strategic customer segments were identified: Aggressively Target, Cultivate, Maintain Engagement, and Deprioritize — each requiring distinct engagement strategies.

From the perspective of Indonesia's retail sector, these findings highlight the growing need for data-driven decision-making in an environment where:

- Digital adoption is rapidly increasing, accelerated by smartphone penetration and e-commerce growth.
- Consumer purchasing power remains uneven, requiring tailored campaign thresholds and localized product bundling.
- Regulatory frameworks, such as Indonesia's Personal Data Protection Act (UU PDP), demand secure and compliant customer data handling in all analytics-driven activities.

The misalignment observed between the campaign spending threshold and actual spending behaviours reinforces a common challenge in Indonesian retail: the tendency to replicate global promotional strategies without sufficient local market calibration. This underscores the importance of data-informed localization in digital transformation initiatives.

4.2 Managerial and Strategic Recommendations

1. Embed Predictive Analytics in Core Marketing Processes
   Retailers in Indonesia should institutionalize predictive modelling for campaign planning, ensuring that promotions are designed based on actual spending and behavioural patterns rather than generalized assumptions.
2. Localize Promotional Thresholds and Product Bundles
   Given that only 9.9% of customers exceeded the IDR 800,000 monthly spend threshold,

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

future campaigns should either adjust the qualifying spend or offer product bundles that encourage incremental purchases without alienating lower-spending segments.

3. Prioritize High-Value Segments for Digital Engagement
   Focus digital marketing investments on Aggressively Target and Maintain Engagement clusters, leveraging personalized offers, loyalty rewards, and exclusive digital experiences.
4. Integrate Analytics with Omnichannel Strategies
   Predictive insights should inform both online and offline touchpoints, ensuring a seamless and consistent customer experience across e-commerce platforms, mobile apps, and physical stores.
5. Strengthen Data Governance and Compliance
6. As predictive analytics becomes central to business transformation, compliance with UU PDP and secure data handling protocols must be embedded into analytics workflows to maintain consumer trust.

4.3 Future Research Directions

While this study used sample data from SAS Viya, future research could:

- Incorporate real transactional data from Indonesian retailers to validate model applicability in live environments.
- Explore hybrid modelling techniques (e.g., ensemble methods combining decision trees and neural networks) for improved predictive performance.
- Assess the long-term ROI of predictive analytics investments in Indonesia's retail sector, including their impact on customer lifetime value and brand loyalty.

DECLARATION OF GENERATIVE AI

During the preparation of this work, the author(s) used ChatGPT (OpenAI) to assist in structuring and minor language editing. The content, analysis, and interpretations are based on the author(s)' own research and dataset (sample data from SAS Viya library), and the author(s) take full responsibility for the accuracy and validity of the final manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Universitas Terbuka for providing the academic platform for this research presentation, and SAS Institute for providing sample datasets via the SAS Viya library that facilitated the demonstration of predictive analytics techniques in this study. Appreciation is also extended to colleagues and mentors who provided valuable feedback during the development of this work.

REFERENCES

[1] M. Feki and I. Boughzala, "The Shape of Digital Transformation: A Systematic Literature Review," 2015. [Online]. Available: http://aisel.aisnet.org/mcis2015

[2] K. S. R. Warner and M. Wäger, "Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal," *Long Range Plann*, vol. 52, no. 3, pp. 326–349, Jun. 2019, doi: 10.1016/j.lrp.2018.12.001.

[3] J. Reis, M. Amorim, N. Melão, and P. Matos, "Digital transformation: A literature review and guidelines for future research," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 411–421. doi: 10.1007/978-3-319-77703-0_41.

DOI: 10.33830/osc.v3i1.7914 / ISSN: 3032-2227

[4] Y. J. Cai and C. K. Y. Lo, "Omni-channel management in the new retailing era: A systematic review and future research agenda," *Int J Prod Econ*, vol. 229, Nov. 2020, doi: 10.1016/j.ijpe.2020.107729.

[5] M. Wedel, P. K. Kannan, and R. J. Tyser, "Marketing Analytics for Data-Rich Environments Marketing Analytics for Data-Rich Environments Marketing Analytics for Data-Rich Environments," 2016.

[6] F. Chad, "The Role of Organizational Readiness in the Successful Adoption of AI in Retail Customer Relationship Management (CRM)," 2025. [Online]. Available: https://www.researchgate.net/publication/391875020

[7] F. Zivic, A. K. Malisic, N. Grujovic, B. Stojanovic, and M. Ivanovic, "Materials informatics: A review of AI and machine learning tools, platforms, data repositories, and applications to architectured porous materials," *Mater Today Commun*, vol. 48, p. 113525, Sep. 2025, doi: 10.1016/j.mtcomm.2025.113525.

[8] C. Brackmann, M. Hütsch, and T. Wulfert, "Identifying Application Areas for Machine Learning in the Retail Sector: A Literature Review and Interview Study," *SN Comput Sci*, vol. 4, no. 5, Sep. 2023, doi: 10.1007/s42979-023-01888-w.

[9] X. (Shane) Wang, J. H. (Joseph) Ryoo, N. Bendle, and P. K. Kopalle, "The role of machine learning analytics and metrics in retailing research," *Journal of Retailing*, vol. 97, no. 4, pp. 658–675, Dec. 2021, doi: 10.1016/j.jretai.2020.12.001.

[10] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *Journal of Retailing and Consumer Services*, vol. 61, Jul. 2021, doi: 10.1016/j.jretconser.2021.102566.

[11] A. Lutfi *et al.*, "Drivers and impact of big data analytic adoption in the retail industry: A quantitative investigation applying structural equation modeling," *Journal of Retailing and Consumer Services*, vol. 70, Jan. 2023, doi: 10.1016/j.jretconser.2022.103129.

[12] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy Artificial Intelligence," *Gov Inf Q*, vol. 37, no. 3, Jul. 2020, doi: 10.1016/j.giq.2020.101493.

[13] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.

[14] E. T. Bradlow, M. Gangwar, P. Kopalle, and S. Voleti, "The Role of Big Data and Predictive Analytics in Retailing," *Journal of Retailing*, vol. 93, no. 1, pp. 79–95, Mar. 2017, doi: 10.1016/j.jretai.2016.12.004.

[15] M. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, Sep. 2023, doi: 10.1007/s10257-023-00640-4.

[16] R. P. Silva, C. Saraiva, and H. S. Mamede, "Assessment of organizational readiness for digital transformation in SMEs," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 362–369. doi: 10.1016/j.procs.2022.08.044.

[17] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *Journal of Marketing Analytics*, vol. 11, no. 4, pp. 677–692, Dec. 2023, doi: 10.1057/s41270-023-00235-5.

[18] B. Madhiraju, S. Reddy, and G. Sasikala, "CUSTOMER SEGMENTATION USING RFM ANALYSIS," vol. 12, no. 7, 2024, doi: 10.36713/epra2012.