Eduvest – Journal of Universal Studies
Volume 6 Number 7, July, 2026
p- ISSN 2775-3735- e-ISSN 2775-3727

Predictive Segmentation for Retail CRM: An EIV Framework with Machine Learning and Causal Experimentation

Fandy Akhmad Ariyadi*, Raden Bagus Fajriya Hakim
Universitas Islam Indonesia, Indonesia
Email: fandy.ariyadi@students.ac.id *, hakimf@uii.ac.id

ABSTRACT

Background: Identifying high-conversion customers is a key challenge in retail CRM. Traditional segmentation methods often fail to accurately predict customer behavior, leading to suboptimal targeting strategies. This study proposed a predictive segmentation framework based on Engagement, Intent, and Value (EIV) to enhance conversion rates in the Indonesian retail sector. **Objective**: This study aimed to evaluate the effectiveness of the EIV-based segmentation framework through offline benchmarking and online causal experimentation, particularly in terms of conversion uplift and CRM performance. Methods: Five algorithmic models—Logistic Regression, Random Forest, Gradient Boosting Machines (GBM), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN)—were trained using omnichannel behavioral data. Model performance was evaluated using Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC). The framework was further validated through a large-scale field experiment involving 59 email campaigns across 15 e-commerce platforms. **Results**: The GBM model demonstrated the most robust predictive performance and was selected for deployment. Experimental results showed that the EIV-based predictive segmentation consistently produced statistically significant conversion uplift compared to traditional operational segmentation. **Conclusion**: This study highlighted the effectiveness of a unified, propensity-based segmentation approach in improving CRM strategies. By integrating offline predictive modeling with online causal validation, the framework demonstrated how predictive segmentation can enhance CRM outcomes in real-world omnichannel environments.

Keywords:
Predictive Segmentation
Machine Learning
Customer Relationship Management (CRM)
Causal Field Experimentation
Email Conversion Uplift

INTRODUCTION

The Indonesian retail sector has evolved into a digitally driven ecosystem characterized by a sophisticated omnichannel landscape. In 2022, retail sales reached IDR 1,526.2 trillion, driven by a middle-class population of 275 million and internet penetration exceeding 70% (RI, 2023). As e-commerce platforms such as Tokopedia and Shopee reshape consumer expectations, retailers have shifted toward integrated online–offline strategies to maintain competitiveness (Hosmer et al., 2013; Loh & Ren, 2024; Verhoef et al., 2015).

However, this digital transformation introduces significant challenges, particularly fragmented behavioral signals and high shopping cart abandonment rates (Chopra et al., 2024).

For large-scale lifestyle retailers such as MAP Group—operating more than 3,000 outlets and 28 e-commerce sites—the primary challenge is no longer data collection but the precision of customer engagement. While email remains a cost-efficient communication channel, its effectiveness is often limited by traditional segmentation models (e.g., “active on site” or “recent buyer”), which operate in silos and fail to capture a holistic view of customers (Li et al., 2025a; Ma et al., 2024; Zhu et al., 2023).

Current research indicates a shift in retail analytics away from heuristic segmentation based on static data and rule-based filters. These conventional approaches often fail to capture the latent complexity of modern consumer behavior (Gao & Yang, 2022; Hajek & Phung, 2024; Li et al., 2025b). To address these limitations, there has been a transition toward machine learning models that integrate both static attributes and dynamic behavioral signals. Such methods enable the identification of nonlinear patterns and key performance drivers, providing deeper insights into conversion behavior compared to traditional heuristic approaches (Izuzquiza et al., 2026). Despite these advancements, a significant gap remains in the literature regarding empirical evidence of the causal uplift these models generate in live retail environments. Specifically, limited evidence exists demonstrating their ability to increase conversion rates compared to traditional segmentation methods (Chaudhuri et al., 2021).

This study addressed this gap by proposing an EIV-based framework (Engagement, Intent, and Value). By benchmarking five algorithmic families and validating results through a large-scale field experiment across 15 e-commerce platforms, this research provided an end-to-end methodology that bridges offline predictive performance with real-world conversion uplift.

METHOD

Research Design

This study proposed an EIV-based predictive segmentation framework in which customer behavior was represented through three integrated dimensions: Engagement (email interaction signals), Intent (website activity and browsing behavior), and Value (transaction history based on RFM metrics). The framework followed a four-stage analytical pipeline consisting of data engineering, feature construction, predictive modeling, and marketing experimentation. The overall process is illustrated in Figure 2.

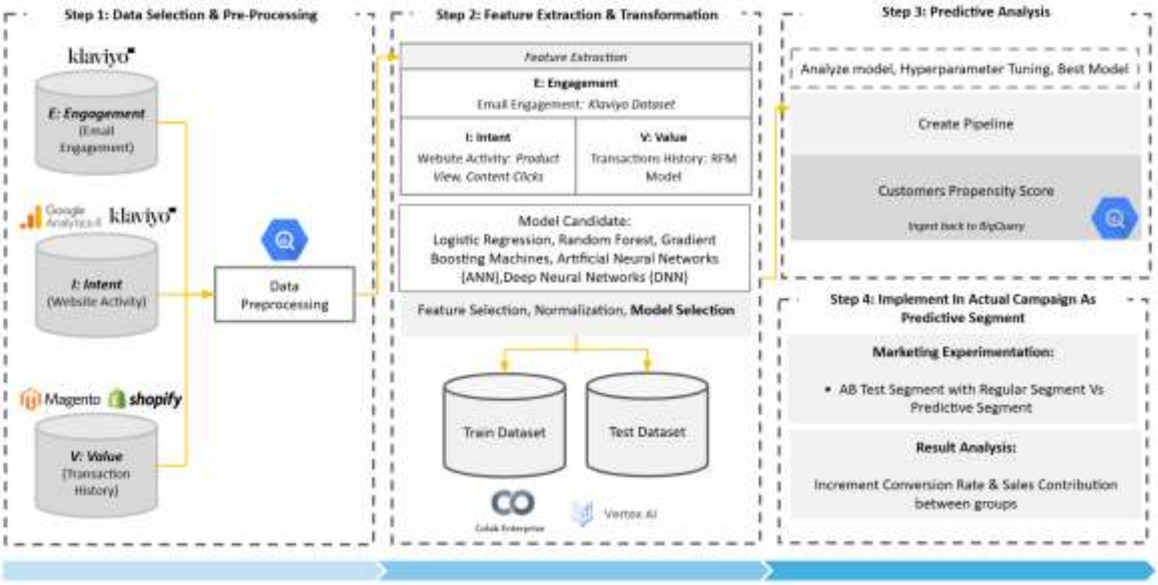

Fig. 2. An EIV-Based Predictive Segmentation Framework

Population, Sample, and Sampling Method

The study population comprises approximately 3,847,758 unique registered email addresses within a consolidated retail ecosystem consisting of ten brand-specific websites (e.g., Aldoshoes.co.id, Fitflop.co.id, and FootLocker.id). Inclusion in the population requires full omnichannel traceability across three behavioral data streams: website activity, email interaction, and transaction history, spanning from 1 January 2023 to 31 March 2025. Customer records are acquired through integrated digital and physical touchpoints, including website sign-ups, newsletter subscriptions, and in-store self-registrations.

The analytical sample was derived via non-probability purposive sampling based on data completeness and operational eligibility. Customers were selected if they possessed integrated data from Google Analytics 4 (site interactions), Klaviyo (email engagement), and Magento/Shopify (transactional outcomes), and had interacted with the platform during the experimental period from 5 March to 5 April 2025. Within this timeframe, 59 email campaigns were executed across 15 websites using standardized segmentation logic. Following extraction, cleaning, and deduplication, the final sample consists of 160,997 customer-level records.

To safeguard internal validity and prevent cross-campaign interference, a D+6 quarantine window was implemented. Any customer targeted by a campaign was placed in a six-day cool-down period during which they received no additional email communications. This controlled mechanism prevents attribution inflation and ensures that conversion behavior is mapped accurately to specific stimuli, maintaining the integrity of the causal interpretation (Breiman, 2001; Lemon & Verhoef, 2016; Setyarini et al., 2024; Zhu et al., 2023).

Data Analysis Method

This study employs a structured analytical framework integrating offline model development with online experimental validation. The offline phase focuses on the development, training, and comparative evaluation of multiple machine learning models—including Logistic Regression, Random Forest, Gradient Boosting, and Deep Neural Networks—to predict customer purchase propensity. Model performance is rigorously assessed using statistical metrics to ensure predictive accuracy and consistency in identifying high-conversion candidates.

Subsequently, the champion model is deployed within a real-world marketing experimental framework to validate operational effectiveness. In this online stage, a predictive segment generated by the model is benchmarked against the retailer's baseline regular segmentation through live email campaigns. This dual-stage design ensures that the analytical framework produces a model that is both statistically robust and practically impactful in driving higher conversion rates under actual campaign conditions.

RESULT AND DISCUSSION

EIV Framework

The results of this study demonstrate that a unified propensity model can be effectively constructed by consolidating these heterogeneous data sources within a centralized analytics pipeline.

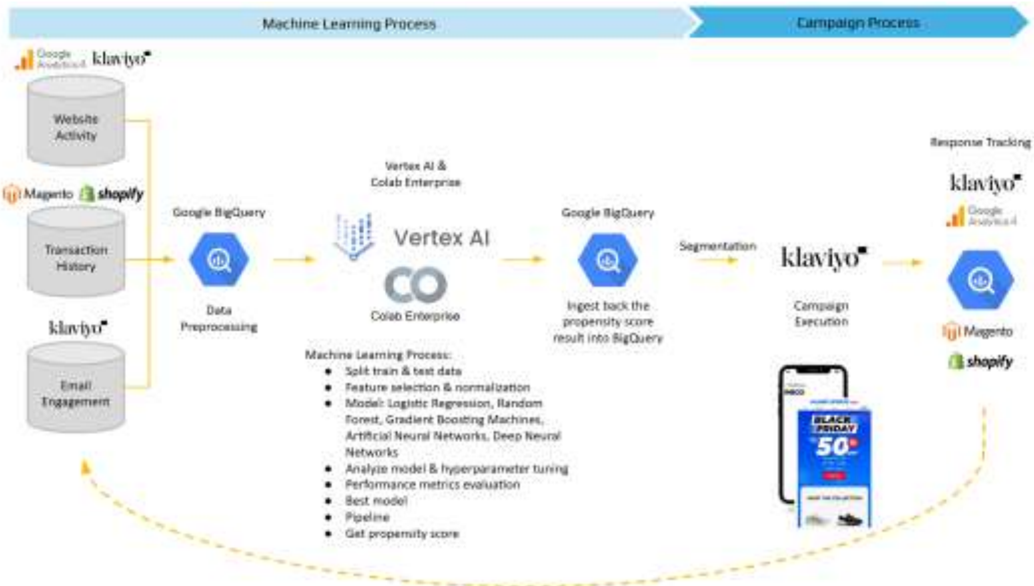

Fig 3. Unified Propensity Model from Omnichannel Behavioral Data

As illustrated in the system architecture (Fig. 3), behavioral signals originating from website interactions, email campaigns, and transactional systems are ingested into a common data warehouse (Google BigQuery). Feature engineering and preprocessing are applied to transform raw behavioral logs into structured predictors, including recency, frequency, monetary indicators, and engagement intensity metrics. These features are subsequently used to train and evaluate multiple predictive models within a standardized machine-learning workflow before propensity scores are fed back into the campaign execution platform.

This architecture operationalizes the theoretical engagement–intent–conversion pathway by translating observable customer behaviors into quantifiable signals of latent purchase intent. Importantly, the results of the feature importance analysis confirm that exposure- and behavior-related variables—such as email receipt frequency, recency since last transaction, and RFM-based segmentation—play a dominant role in predicting conversion. This finding validates the conceptual assumption that omnichannel behavioral data can be meaningfully unified to model purchase propensity in a real-world CRM environment.

Model Performance Results

Table 4.2 reports model performance after hyperparameter optimization. Overall, tuning led to measurable performance improvements, although the magnitude and direction of gains varied across models.

Table 2. Performance Metrics After Hyperparameter Tuning

Performance Metrics:	PPV	PPV_ rank	Recall	Recall _rank	ROC-AUC	ROC-AUC _rank	MCC	MCC_ rank	aggregate _rank
Gradient Boosting Machines	0.749294	1	0.590882	2	0.932520	1	0.617788	1	5
Artificial Neural Networks (ANN)	0.675389	4	0.499121	3	0.901132	3	0.523189	3	13
Random Forest	0.488433	5	0.823977	1	0.918751	2	0.556981	2	10
Deep Neural Networks (DNN)	0.676524	3	0.477206	4	0.895482	4	0.510871	4	15
Logistic Regression	0.705683	2	0.414743	5	0.883083	5	0.487257	5	17

Gradient Boosting Machines retained their position as the top-ranked model, achieving further improvements in recall, ROC–AUC, and MCC, and maintaining the lowest aggregate rank score. This confirms the robustness of GBM and its suitability for conversion prediction tasks in imbalanced CRM datasets.

Random Forest exhibited the most notable change following tuning, with a substantial increase in recall, indicating improved sensitivity to conversion events. However, this improvement was accompanied by a marked decline in precision, highlighting a stronger bias toward positive predictions and reinforcing the need for multi-metric evaluation.

ANN models showed modest but consistent gains across most metrics, while Logistic Regression remained relatively stable, suggesting limited sensitivity to hyperparameter adjustments.

In contrast, the DNN model did not experience performance gains after tuning; its ROC–AUC declined considerably despite moderate recall levels. This outcome suggests potential overfitting or instability under the given data constraints, underscoring the importance of balancing model complexity with dataset characteristics.

Model Selection Justification

The final model selection is grounded in academic rigor and operational feasibility, balancing the challenges of imbalanced conversion data with the requirements of large-scale retail analytics. Across all evaluation stages, Gradient Boosting Machines (GBM) consistently achieved the superior aggregate rank, driven by its robust performance on ROC–AUC and Matthews Correlation Coefficient (MCC) metrics while maintaining competitive precision and recall. Methodologically, GBM's iterative error-correction capability allows it to capture complex, non-linear interactions between website activity, email engagement, and transaction history more effectively than Deep Neural Networks (DNN), which exhibited unstable discrimination and potential overfitting post-tuning. Beyond predictive accuracy, GBM offers significant operational advantages, including higher computational efficiency for frequent retraining and greater model interpretability through feature importance analysis—a critical factor for stakeholder trust in CRM environments. Given the dynamic nature of streaming data from BigQuery, this study adopts an adaptive selection framework; by utilizing imbalance-aware metrics and rank-based aggregation, the pipeline ensures that the optimal model can be systematically re-identified as customer behavioral regimes and data scales evolve over time.

Feature Importance

To elucidate the predictive logic of the model, a permutation importance analysis was conducted, measuring the mean decrease in ROC-AUC when feature values are randomly shuffled. This model-agnostic approach identifies the variables most critical to distinguishing converted from non-converted customers.

Table 3. Summary of feature importance

Features	Importance Mean	Importance Standard Deviation
count_email_received	0.2445778000	0.0036030000
day_since_last_trx	0.0896037100	0.0012420000
RFM_segment_ordinal	0.0745004100	0.0007830000
RFM_segment	0.0684614900	0.0014610000
channel_code	0.0321048400	0.0006800000
F	0.0220537400	0.0004010000
R	0.0095647380	0.0004740000
M	0.0073752740	0.0001490000
count_site_activity	0.0049570790	0.0001510000
count_soft_bounced	0.0022933810	0.0002490000
count_coupon_received	0.0012298110	0.0001060000
count_trx	0.0004055632	0.0000570000
count_hard_bounced	0.0003347511	0.0000660000
count_email_opened	0.0002016672	0.0000690000
count_email_opened_clicked	0.0001983350	0.0000770000
count_content_clicked	0.0001796328	0.0000560000
net_sales	0.0000390374	0.0000160000
count_product_viewed	0.0000053118	0.0000020000
count_automation_trx	0.0000042379	0.0000020000
count_campaign_unsubscribed	0.0000002855	0.0001070000
count_campaign_spam_marked	-0.0000000080	0.0000000000
count_chat_delivered	-0.0000000080	0.0000000000

The results indicate that behavioral exposure dominates the predictive structure, led by count_email_received (0.2446), which underscores the primary role of communication frequency and reinforcement in driving conversion. Lifecycle and recency signals follow as secondary drivers, with day_since_last_trx (0.0896) and RFM-based segments (0.0745) confirming that recent transactional history remains a vital contextual anchor for intent. Interestingly, deeper engagement metrics—such as email clicks and product views—exhibited minimal incremental importance, suggesting these actions represent exploratory behavior rather than decisive purchase intent within the observed data. Ultimately, these findings align with the theoretical premise that digital conversion is driven by the synergy of interaction intensity and structural customer state, providing a robust foundation for the live operational experiments detailed in the following section.

Online Experiments Result

Table 4 reports the incremental conversion rate uplift of predictive segmentation compared to regular segmentation for each SBU.

Table 4. Incremental Conversion Rate Uplift of Predictive Segmentation Across SBU

SBU	Conversion Rate Regular Segment	Conversion Rate Predictive Segment	Increment
Fashion Footwear	0.02%	0.16%	742%
Kids	0.02%	0.62%	2918%
Sports	0.14%	0.80%	453%

Table 4 demonstrates substantial relative improvements in conversion performance across all SBUs when predictive segmentation is applied. The Kids SBU exhibits the most pronounced uplift, with the conversion rate increasing from 0.02% in the regular segment to 0.62% in the predictive segment—corresponding to a 2918% relative increase. This exceptionally high uplift reflects the strong ability of the predictive model to identify highly responsive customers within a traditionally low-conversion segment.

Similarly, Fashion Footwear shows a significant improvement, where the conversion rate rises from 0.02% to 0.16%, yielding a 742% uplift. Although the absolute conversion rate remains modest, the relative gain indicates a meaningful enhancement in targeting efficiency. In the Sports SBU, the conversion rate improves from 0.14% to 0.80%, representing a 453% increase, which is notable given the already higher baseline conversion level in this category.

Overall, these results confirm that predictive segmentation consistently outperforms regular segmentation across heterogeneous business units, delivering not only higher absolute conversion rates but also substantial relative gains. The magnitude of uplift varies by SBU, suggesting differences in customer behavior dynamics and baseline responsiveness; however, the direction of impact remains uniformly positive. This reinforces the practical value of deploying predictive models in real-world campaign environments, particularly for improving conversion efficiency under constrained reach conditions.

CONCLUSION

This research established a validated framework for customer propensity modeling by integrating omnichannel behavioral data—specifically website activity, email engagement, and transaction history—within a centralized digital marketing ecosystem. The study demonstrated that consolidating heterogeneous behavioral signals into a unified analytical pipeline effectively operationalized the engagement–intent–conversion pathway. Using Google BigQuery and Vertex AI, the framework identified communication frequency, time since last transaction, and RFM-based lifecycle positioning as the most robust predictors of short-term conversion, confirming that purchase behavior is a cumulative outcome of measurable digital interactions rather than a stochastic event.

Comparative analysis revealed that Gradient Boosting Machines (GBM) outperformed high-complexity architectures such as Deep Neural Networks (DNN) in this empirical context. While DNNs achieved high recall, they exhibited lower precision and greater instability under class imbalance. In contrast, the GBM model consistently delivered superior performance across ROC–AUC and Matthews Correlation Coefficient (MCC) metrics. Beyond statistical performance, the deployment of predictive segments in live email campaigns yielded substantial operational gains, with conversion rate uplifts reaching 519% at the company level. These findings provide strong empirical evidence that data-driven predictive segmentation enables more efficient allocation of marketing resources compared to traditional baseline segmentation approaches. Ultimately, this research bridges the gap between machine learning methodology and real-world retail analytics, offering an adaptive decision-support system that remains relevant as customer behavioral patterns evolve over time.

REFERENCE

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*, *149*. https://doi.org/10.1016/j.dss.2021.113622

Chopra, I. P., Jebarajakirthy, C., Jain, T., & Maseeh, H. I. (2024). Electronic shopping cart abandonment: What do we know and where should we be heading? *Electronic Markets*, *34*(1). https://doi.org/10.1007/s12525-024-00697-0

Gao, H., & Yang, Y. (2022). *Multi-Head Online Learning for Delayed Feedback Modeling*. http://arxiv.org/abs/2205.12406

Hajek, P., & Phung, R. (2024). Deep learning-based forecasting of customer conversion in omnichannel retailing. *International Journal of Information Management Data Insights*, *4*(1). https://doi.org/10.1016/j.jjimei.2024.100215

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118548387

Izuzquiza, A., Patricio, M. A., Cuadrado-Gallego, J. J., Berlanga, A., & Molina, J. M. (2026). Integrating static and dynamic hierarchical clustering and its application to retail segmentation. *Computers in Industry*, *174*. https://doi.org/10.1016/j.compind.2025.104410

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Li, X., Chen, Y., & Liu, Z. (2025a). Neural Network Architectures for Predicting Long-term Customer Value. *Journal of Retailing and Consumer Services*, *82*.

https://doi.org/10.1016/j.jretconser.2024.104012
Li, X., Chen, Y., & Liu, Z. (2025b). Neural Network Architectures for Predicting Long-term Customer Value in Digital Ecosystems. *Journal of Retailing and Consumer Services*, *82*. https://doi.org/10.1016/j.jretconser.2024.104012
Loh, W. W., & Ren, D. (2024). The Incremental Propensity Score Approach for Diversity Science. *Advances in Methods and Practices in Psychological Science*, *7*(2). https://doi.org/10.1177/25152459241240681
Ma, L., Sun, B., & Yang, H. (2024). Adaptive Gradient Boosting Machines for Dynamic Customer Behavior Analysis. *Decision Support Systems*, *176*. https://doi.org/10.1016/j.dss.2023.114065
RI, K. P. (2023). *Buku Profil Industri Ritel 2023*. https://ditjenpen.kemendag.go.id/storage/publikasi/5QqkD1tZ9rzn9yPSIBW7RkiyYreg0Gbs6CG2xmyl.pdf
Setyarini, D. A., Gayatri, A. A. M. D., Aditya, C. S. K., & Chandranegara, D. R. (2024). Stroke Prediction with Enhanced Gradient Boosting Classifier and Hyperparameter Tuning. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *23*(2). https://doi.org/10.30812/matrik.v23i2.3555
Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From Multi-Channel Retailing to Omni-Channel Retailing. Introduction to the Special Issue on Multi-Channel Retailing. *Journal of Retailing*, *91*(2), 174–181. https://doi.org/10.1016/j.jretai.2015.02.005
Zhu, Y., Zhang, J., & Wang, S. (2023). An optimized Gradient Boosting Decision Tree model for customer purchase prediction in e-commerce. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.119044