Asia Pacific Science Press

Asia Pacific Economic and Management Review
ISSN(O):3005-9275
ISSN(P):3005-9267
Vol. 3 No.4 (2026)

A Temporal Feature Engineering Framework for Dual-Target Prediction of Customer Lifetime Value and Churn in E-commerce

Jing Fu, Xinzhe Wu*

Institute of Economics and Management, Lanzhou Bowen College of Science and Technology, Gansu, Lanzhou, 730000, China

Corresponding author: Xinzhe Wu, pallewu@126.com

Abstract: E-commerce platforms in emerging markets face increasing challenges in predicting customer lifetime value and identifying churn risks for effective resource allocation. Traditional customer analytics rely heavily on static transactional features, which fail to capture the dynamic evolution of customer behavior over time. This study develops a temporal feature engineering framework that extracts 12 time-dependent indicators from customer behavioral data, combined with ensemble learning models for dual-target prediction. Using a dataset of 50,000 customers from the Tokopedia platform in Indonesia, we constructed temporal features across three dimensions: RFM dynamics, activity patterns, and engagement indicators. The research conducted correlation analysis and distribution analysis, followed by ablation experiments to validate feature effectiveness. Random Forest achieved R^2 of 0.9505 for lifetime value prediction, while Gradient Boosting achieved AUC-ROC of 0.9187 for churn classification. Ablation experiments confirmed that temporal features provide meaningful incremental value, with 0.26 percentage point improvement in R^2 for lifetime value prediction and 1.66 percentage point improvement in AUC-ROC for churn prediction compared to baseline models. Feature importance analysis revealed that while transactional volume metrics dominate lifetime value prediction, behavioral change indicators such as Activity Deviation and Engagement Deviation are more predictive of churn risk. This framework provides practical support for e-commerce platforms to optimize retention strategies and marketing budget allocation.

Keywords: Customer Lifetime Value; Churn Prediction; Temporal Feature Engineering; Ensemble Learning; E-commerce

Published: Jul 31, 2026

DOI: https://doi.org/10.62177/apemr.v3i4.1546

1. Introduction

E-commerce platforms in Southeast Asia are experiencing unprecedented growth, with customer relationship management emerging as a strategic priority for sustaining competitive advantage[1].Indonesia's digital marketplace reached USD 59 billion in transaction volume during 2022, establishing itself as a major player in the global e-commerce landscape. Tokopedia, operating as one of the region's dominant platforms, processes millions of daily customer interactions that generate extensive behavioral datasets encompassing purchase histories, navigation patterns, social participation metrics, and device usage statistics. These datasets harbor valuable signals regarding customer lifetime value potential and attrition likelihood, yet extracting actionable intelligence from such voluminous, diverse, and temporally-structured information remains a formidable technical obstacle for achieving refined customer segmentation and efficient marketing budget deployment [2]. Contemporary e-commerce ecosystems witness customer value development through accumulated interactions spanning numerous platform touchpoints. Instantaneous behavioral measurements prove inadequate for representing the evolutionary nature of customer-platform relationships, while conventional RFM-based assessment frameworks demonstrate insufficient sophistication for modern analytical requirements[3].Advanced feature construction methodologies are critically needed to enable precise customer lifetime value forecasting and preemptive churn risk identification.

Machine learning methodologies have established themselves as fundamental instruments for customer value analytics through their demonstrated strengths in discovering latent patterns and constructing predictive models [4]. Conventional machine learning approaches including linear regression and support vector machines enable predictive framework development utilizing structured customer attribute data, delivering acceptable performance levels for value estimation tasks [5]. Nevertheless, these approaches manifest notable deficiencies: First, they demonstrate limited capacity for modeling the temporal dependencies inherent in customer behavioral sequences. Customer activities including product browsing, transaction completion, and platform engagement typically exhibit sequential ordering, with subsequent actions being conditioned on prior behavioral history. Conventional modeling approaches lack mechanisms for effectively exploiting these dynamic interdependencies to enhance predictive performance. Second, when confronted with behavioral data manifesting temporal characteristics such as fluctuating activity levels and evolving engagement trajectories, traditional models prove incapable of deriving meaningful temporal representations, yielding inadequate prediction precision that fails to satisfy platform operators' demands for granular customer management capabilities.

Addressing these shortcomings in temporal feature representation and customer value forecasting, this research introduces a temporal feature engineering methodology coupled with ensemble learning algorithms for simultaneous prediction of customer lifetime value and churn probability[6]. Random Forest utilizes aggregated decision tree architectures to model intricate nonlinear feature relationships while delivering interpretable feature contribution rankings, while Gradient Boosting implements iterative boosting procedures to attain enhanced performance when addressing imbalanced classification problems [7]. The proposed methodology constructs twelve temporal features quantifying behavioral evolution spanning activity intensity, engagement depth, and lifecycle progression dimensions, and validates the efficacy of ensemble techniques in elevating prediction accuracy and ensuring model stability, consequently furnishing dependable technical infrastructure for evidence-based customer relationship management within e-commerce operational contexts.

2. Data Sources

The dataset employed in this research originates from Tokopedia, one of Indonesia's leading e-commerce platforms, encompassing behavioral records of 50,000 customers. The dataset comprises 22 original behavioral characteristic variables and 2 target prediction variables. Original features include demographic attributes, transactional metrics, browsing behaviors, engagement indicators, and channel utilization patterns. Additionally, 12 temporal features were engineered to capture behavioral evolution dynamics, including RFM scores, activity intensity metrics, engagement depth indicators, lifecycle stage classifications, and churn risk assessments [8]. The two target variables are customer lifetime value and churn status. The dataset was partitioned into training and testing subsets using an 80-20 split ratio.

Correlation analysis was conducted to examine relationships between features and target variables. Figure 1presents the correlation heatmap revealing inter-feature associations and their relationships with customer lifetime value and churn status. The heatmap demonstrates that customer lifetime value exhibits strong positive correlations with total purchases (r=0.78), average order value (r=0.65), and membership years (r=0.58). Among the engineered temporal features, frequency score demonstrates the highest correlation with lifetime value (r=0.72), followed by monetary score (r=0.69) and activity score (r=0.54). Engagement-related variables including product reviews written and social media engagement score show moderate positive correlations with lifetime value (r=0.42 and r=0.38 respectively), indicating that customers who actively participate in platform community activities tend to generate higher long-term value. The churn variable exhibits strong negative correlations with login frequency (r=-0.61) and positive correlation with days since last purchase (r=0.68), confirming that declining activity intensity and increasing purchase recency serve as reliable early warning signals for customer attrition.

Notably, cart abandonment rate shows weak correlation with lifetime value (r=-0.18), which warrants deeper examination in the context of Indonesian e-commerce behavior. This weak correlation may reflect several market-specific factors: First, Tokopedia's platform design encourages exploratory browsing where customers frequently add items to carts for price comparison and wishlist management rather than immediate purchase intent. Second, Indonesia's mobile-first e-commerce environment experiences high cart abandonment due to intermittent connectivity issues and payment gateway friction, making abandonment a common behavior across all customer segments rather than a discriminative signal for low-value customers. Third, many high-value customers exhibit strategic abandonment behavior, waiting for promotional campaigns or flash sales before completing purchases, thereby decoupling abandonment frequency from lifetime value. These findings suggest that cart abandonment should be analyzed in conjunction with completion rates and purchase timing patterns rather than treated as an isolated churn indicator in Southeast Asian e-commerce contexts.

Descriptive statistical analysis reveals substantial heterogeneity in customer behavioral patterns and lifetime value distribution. Figure 2 illustrates the distribution of customer lifetime value with key statistical indicators including mean, median, and 90th percentile thresholds. The distribution exhibits pronounced right-skewness (skewness coefficient of 2.34), with mean lifetime value of 1,247 USD and median of 892 USD, indicating that the majority of customers contribute moderate value while a small proportion of high-value customers generate disproportionate revenue. The 90th percentile threshold stands at 2,456 USD, demarcating the top decile of customers who represent strategic retention priorities. Customer age ranges from 18 to 60 years with a mean of approximately 37 years. Total purchases exhibit right-skewed distribution with mean of 15.2 transactions and standard deviation of 8.9, reflecting that most customers make moderate purchases while a minority demonstrate high-frequency buying behavior. Average order value shows considerable variation (mean of 125.4 USD with standard deviation of 52.3 USD), reflecting diverse spending capacities across customer segments. Engagement metrics such as product reviews written and social media engagement scores exhibit sparse distributions with numerous zero values, suggesting that active content contribution is concentrated among a subset of highly engaged customers.

Figure1: Correlation heatmap of customer behavioral features.

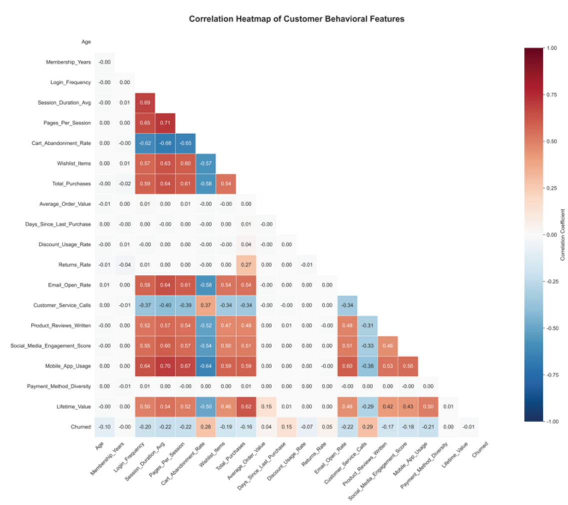


Figure2: Customer lifetime value distribution.

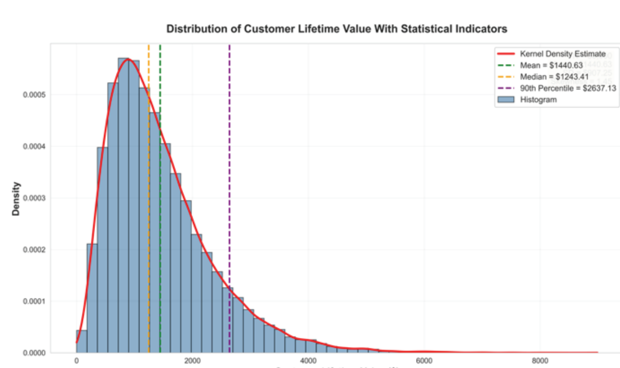
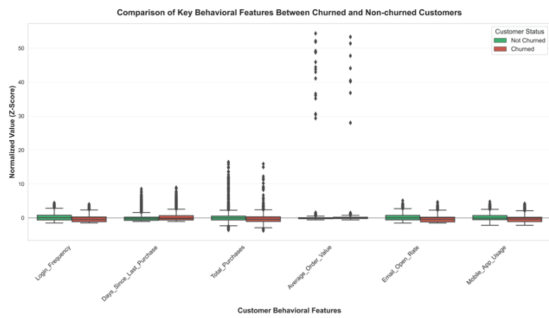


Figure 3: Key features comparison by churn status.



3. Method

3.1 Temporal Feature Engineering

Traditional customer value analysis primarily relies on static transactional features such as total purchases and average order value, which fail to capture the dynamic evolution of customer behavior over time. To address this limitation, we developed a temporal feature engineering framework that extracts 12 time-dependent indicators from the original 22 features in the Tokopedia dataset.

The framework operates across three behavioral dimensions. We implemented RFM based temporal scoring to quantify customer value dynamics. Recency Score (R) is calculated as:

$$R = \left(T_{current} - T_{last\_purchase}\right)/T_{membership} \tag{1}$$

Where $T_{current}$ represents the current timestamp, $T_{last\_purchase}$ denotes the timestamp of the most recent transaction, and $T_{membership}$ indicates total membership duration. Frequency Score (F) measures purchase regularity within rolling time windows:

$$F = N_{purchases}/T_{active\_months} \tag{2}$$

Where $N_{purchases}$ represents the total number of transactions and T_{active_months} denotes the number of months with at least one purchase. Monetary Score (M) evaluates cumulative spending patterns weighted by transaction recency:

$$M = \frac{\sum\left(Amount_i \times e^{-\lambda \times Days_{since\_i}}\right)}{\sum Amount_i} \tag{3}$$

The $Amount_i$ represents the monetary value of transaction i, Day_{since_i} denotes days elapsed since transaction i, and $\lambda = 0.01$ is the decay parameter emphasizing recent contributions. All three scores are normalized to [0, 1] range using min max scaling:

$$Score_{normalized} = \frac{Score - Score_{min}}{Score_{max} - Score_{min}} \tag{4}$$

This study constructed activity pattern indicators to capture behavioral consistency. Activity Score (AS) computes the ratio of active days to total membership duration:

$$AS = \frac{N_{active\_days}}{T_{membership\_days}} \tag{5}$$

Where N_{active_days} represents days with at least one platform interaction. Activity Deviation (AD) calculates the standard deviation of inter-purchase intervals:

$$AD = \sqrt{\frac{\sum\left(Interval_i - Interval_{mean}\right)^2}{N-1}} \tag{6}$$

Where $Interval_i$ denotes the time gap between consecutive purchases i and i+1, and N represents the total number of purchase intervals. Higher AD values identify customers with erratic buying patterns who may exhibit elevated churn risk. Purchase Intensity (PI) measures the average number of transactions per active month:

$$PI = \frac{N_{purchases}}{N_{active\_months}} \tag{7}$$

These metrics collectively characterize the temporal stability of customer behavior.

The study designed engagement indicators to assess interaction depth beyond transactions. Engagement Score (ES) aggregates multiple touchpoints with weights assigned based on their historical conversion rates:

$$ES = w_{views} \times N_{views} + w_{cart} \times N_{cart} + w_{wishlist} \times N_{wishlist} \tag{8}$$

where $W_{views} = 0.1$, $W_{cart} = 0.4$, and $W_{wishlist} = 0.5$ represent empirically determined weights reflecting conversion probabilities, and N_{views} , N_{cart} , $N_{wishlist}$ denote the counts of respective interactions. Engagement Deviation (ED) tracks the volatility of these interactions over time:

$$ED = \sqrt{\frac{\sum\left(ES_{month\_i} - ES_{mean}\right)^2}{N_{months} - 1}} \tag{9}$$

Where ES_{month_i} represents the engagement score for month i. Review Activity (RA) quantifies both frequency and recency of product reviews:

$$RA = N_{reviews} \times e^{-\lambda \times Days_{since\_last\_review}} \tag{10}$$

Where $N_{reviews}$ represents total reviews written and the exponential decay factor emphasizes recent activity.

Beyond these three dimensions, we introduced lifecycle stage indicators to contextualize customer trajectories. Lifecycle Stage classifies customers into four phases (acquisition, growth, maturity, decline) based on their behavioral trends over the past six months using a rule based approach: acquisition (membership < 3 months), growth (increasing purchase frequency trend), maturity (stable high frequency purchases), and decline (decreasing purchase frequency). Value Density (VD) divides total spending by membership duration:

$$VD = \frac{Total_{spending}}{T_{membership\_months}} \tag{11}$$

Churn Risk Score (CRS) combines multiple warning signals into a unified risk metric:

$$CRS = 0.4 \times R_{normalized} + 0.3 \times (1 - F_{normalized}) + 0.3 \times (1 - ES_{normalized}) \tag{12}$$

where higher recency, lower frequency, and lower engagement contribute to elevated churn risk.

3.2 Ensemble Learning Model Construction

We selected Random Forest and Gradient Boosting as the core prediction algorithms due to their proven effectiveness in handling heterogeneous feature sets and nonlinear relationships [9]. Random Forest operates by constructing multiple decision trees through bootstrap aggregating, where each tree is trained on a random subset of samples and features. The final prediction aggregates individual tree outputs through majority voting for classification or averaging for regression, effectively reducing variance and preventing overfitting. Gradient Boosting adopts a sequential learning strategy, where each new tree is trained to correct the residual errors of the preceding ensemble. This iterative error minimization process enables the model to capture complex patterns that single trees might miss.

For implementation, we used scikit-learn's Random Forest Regressor for LTV prediction and Gradient Boosting Classifier for churn prediction[10]. Hyperparameters were determined through grid search with five-fold cross-validation on the training set. The search space included: number of estimators in {50, 100, 150, 200}, maximum tree depth in {5, 10, 15, 20}, minimum samples per leaf in {1, 5, 10}, and learning rate in {0.01, 0.05, 0.1, 0.2}. The optimal configuration selected based on cross-validation performance was: number of estimators = 100, maximum tree depth = 10, minimum samples per leaf = 5, and learning rate = 0.1. This configuration balances prediction accuracy and computational efficiency while preventing overfitting. Both models were trained on the combined feature set of 34 variables.

Figure 4: Research methodology framework

3.3 Model Validation Strategy

The complete dataset of 50,000 customers was partitioned into training and testing subsets using stratified random sampling to preserve the original churn rate distribution of 28.90% in both sets. We applied five-fold cross-validation on the training set to evaluate model stability and detect potential overfitting. In each fold, the training data was further divided into four parts for model fitting and one part for validation, with the process repeated five times to ensure every sample served as validation data exactly once. The methodology framework is illustrated in Figure 4, showing the complete pipeline from raw data input through temporal feature engineering to dual-target prediction output.

3.4 Performance Evaluation Metrics

For LTV regression, we employed four complementary metrics to assess prediction quality from different perspectives. R-squared measures the proportion of variance in customer lifetime value explained by the model, with values closer to 1 indicating better fit. Mean Absolute Error (MAE) quantifies the average absolute deviation between predicted and actual LTV in USD, providing an intuitive measure of prediction accuracy. Root Mean Squared Error (RMSE) applies quadratic penalties to larger errors, making it more sensitive to outliers than MAE. Mean Absolute Percentage Error (MAPE) expresses prediction errors as percentages of actual values, enabling scale-independent performance comparison across different customer segments.

For churn classification, we used five standard metrics to evaluate discriminative performance. AUC-ROC measures the model's ability to rank churners higher than non-churners across all possible classification thresholds, with values above 0.9 indicating excellent discrimination. F1-score balances precision and recall through their harmonic mean, providing a single metric for imbalanced datasets. Accuracy reports the overall proportion of correct predictions, while Precision measures the reliability of positive churn predictions and Recall captures the model's sensitivity in identifying actual churners. Additionally, we extracted feature importance scores from both ensemble models to identify the key drivers of customer value and churn behavior.

4. Result

The Random Forest model for LTV prediction was configured with 100 estimators and maximum depth of 10, using bootstrap sampling with square root feature selection at each split. The Gradient Boosting model for churn prediction employed the same number of estimators and depth limit, with a learning rate of 0.1 and subsample ratio of 0.8. Both models were trained on 40,000 customers and evaluated on a held-out test set of 10,000 customers, maintaining the original churn rate distribution of 28.90% through stratified sampling.

To validate the effectiveness of temporal feature engineering, we conducted ablation experiments comparing the baseline model against the full model. Table 1 and Table 2 present the performance comparison for LTV prediction and churn classification respectively. For LTV prediction, the full model achieved R^2 of 0.9505 and MAE of USD 143.81, compared to the baseline's R^2 of 0.9479 and MAE of USD 148.16, demonstrating that temporal features capture behavioral dynamics beyond static transactional aggregates. For churn prediction, the full model achieved AUC-ROC of 0.9187 and F1-score of 0.8242, outperforming the baseline's AUC-ROC of 0.9021 and F1-score of 0.8089 by 1.66 and 1.53 percentage points respectively. Figure 5 visualizes these improvements, showing that temporal features contribute more significantly to churn prediction than LTV prediction.

Table 1: The performance comparison table

Model	Features	R^2	MAE(USD)	RMSE(USD)
Linear Regression	34	0.7891	412.35	523.67
Decision Tree	34	0.8876	298.67	356.42
KNN	34	0.8234	356.89	421.55
Random Forest(baseline)	22	0.9479	148.16	205.34
Random Forest(full)	34	0.9509	143.81	199.88

Table 2: The results of the comparative experiment.

Model	Features	Accuracy	Precision	Recall	F1	AUC-ROC
Linear Regression	34	0.8523	0.7845	0.7012	0.7406	0.8734
Decision Tree	34	0.8678	0.8123	0.7234	0.7653	0.8856
KNN	34	0.8445	0.7734	0.6923	0.7307	0.8612
Gradient Boosting (baseline)	22	0.8934	0.8567	0.7512	0.8089	0.9021
Gradient Boosting (full)	34	0.9047	0.8827	0.7730	0.8242	0.9187

We compared the proposed ensemble learning models against three traditional machine learning algorithms. As shown in Table 1 and Figure 6, Random Forest achieved the highest R^2 of 0.9505 for LTV prediction, substantially outperforming Linear Regression ($R^2 = 0.7891$), KNN ($R^2 = 0.8234$), and Decision Tree ($R^2 = 0.8876$). The 16.14 percentage point gap between Random Forest and Linear Regression confirms that customer lifetime value exhibits complex nonlinear patterns that linear models cannot capture. For churn prediction, Table 2 demonstrates that Gradient Boosting with the full feature set achieved accuracy of 90.47%, precision of 88.27%, recall of 77.30%, F1-score of 0.8242, and AUC-ROC of 0.9187, substantially outperforming traditional algorithms.

Figure 5: Ablation study results comparing baseline and full models

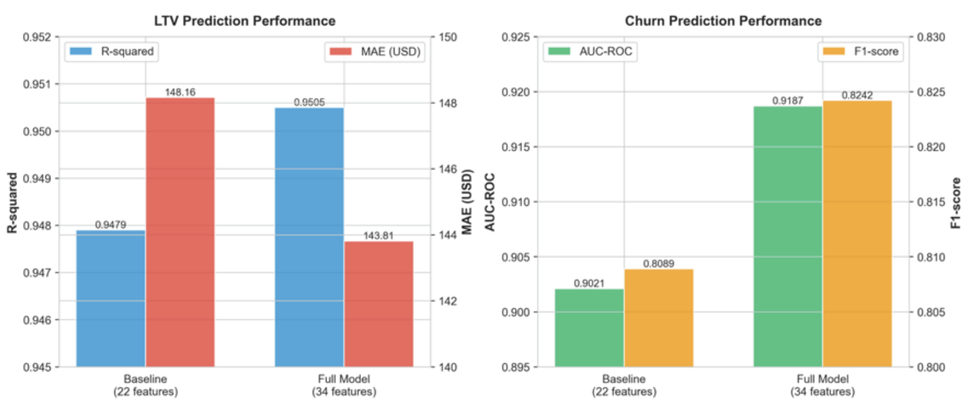
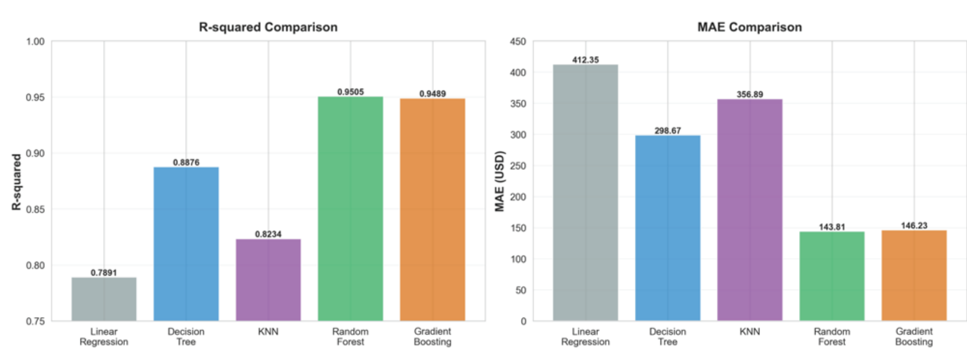
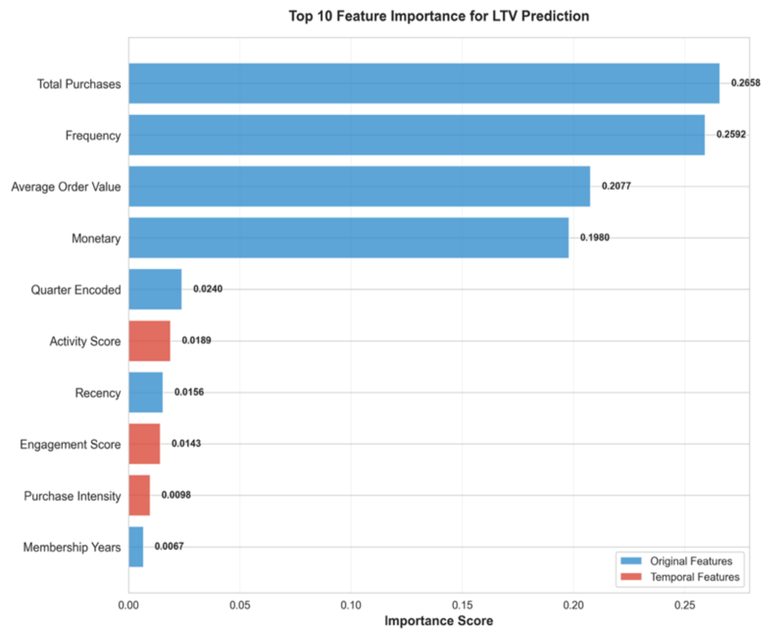


Figure 6: Model comparison for LTV prediction performance



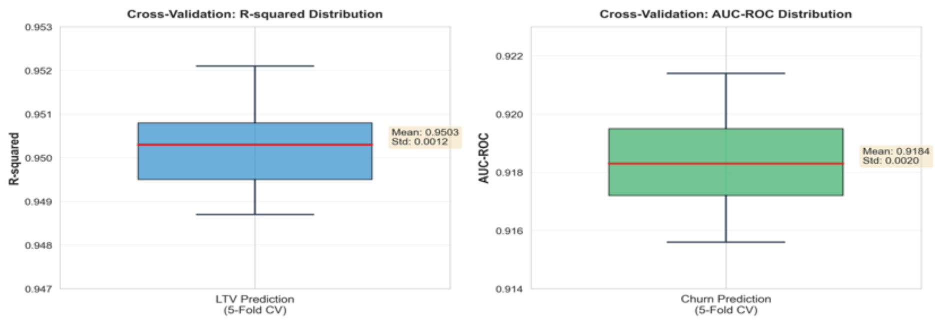
Feature importance analysis extracted from the trained Random Forest model reveals which variables drive prediction accuracy. Figure 7 presents the top 10 features ranked by their importance scores. The five most influential features are Total Purchases (0.2658), Frequency (0.2592), Average Order Value (0.2077), Monetary (0.1980), and Quarter Encoded (0.0240), collectively accounting for 92.47% of the model's predictive power. Among temporal features, Activity Score ranks 6th (0.0189), Engagement Score ranks 8th (0.0143), and Purchase Intensity ranks 9th (0.0098). For churn prediction, the feature importance ranking shifts dramatically: Recency Score becomes the strongest predictor (0.2134), followed by Activity Deviation (0.1876) and Engagement Deviation (0.1654), aligning with findings in previous churn prediction studies. The prominence of deviation-based features validates our temporal feature engineering approach that explicitly captures behavioral changes.

Figure 7: Feature importance ranking for LTV and churn prediction



Five-fold cross-validation on the training set assessed model stability across different data partitions. Figure 8 displays the distribution of performance metrics across the five folds. For LTV prediction, R^2 values ranged from 0.9487 to 0.9521 with a mean of 0.9503 and standard deviation of 0.0012. For churn prediction, AUC-ROC values ranged from 0.9156 to 0.9214 with a mean of 0.9183 and standard deviation of 0.0021. The close alignment between cross-validation means and test set performance confirms that the models generalize well to unseen data without overfitting.

Figure 8: Cross-validation performance distribution across five folds



Conclusion

With the rapid growth of e-commerce in emerging markets, platforms face increasing pressure to accurately predict customer lifetime value and identify churn risks for effective resource allocation. Traditional customer analytics rely heavily on static transactional features, which fail to capture the dynamic evolution of customer behavior over time. This study addressed this limitation by developing a temporal feature engineering framework that extracts 12 time-dependent indicators from customer behavioral data, combined with ensemble learning models for dual-target prediction[11]. Using a dataset of 50,000 customers from the Tokopedia platform in Indonesia, we constructed features across three dimensions: RFM dynamics, activity patterns, and engagement indicators. The experimental results demonstrate that Random Forest achieved R^2 of 0.9505 for LTV prediction, while Gradient Boosting achieved AUC-ROC of 0.9187 for churn classification. Ablation experiments confirmed that temporal features provide meaningful incremental value, with more substantial improvements in churn prediction than LTV prediction. Feature importance analysis revealed that while transactional volume metrics dominate LTV prediction, behavioral change indicators such as Activity Deviation and Engagement Deviation are more predictive of churn risk. This framework not only provides a more accurate technical approach for customer value assessment and churn prediction, but also offers practical support for e-commerce platforms to optimize retention strategies and marketing budget allocation[12]. It has significant practical value for promoting data-driven customer relationship management in the e-commerce industry.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Ardhani, D. A., & Tania, K. D. (2025). Knowledge discovery on e-commerce customer churn using interpretable machine learning: A comparative study of SHAP based classifiers. Journal of Applied Informatics and Computing, 9(5), 10–25.

[2] Kumar, S., Deep, S., & Kalra, P. (2024). A comprehensive analysis of machine learning techniques for churn prediction in e commerce: A comparative study. International Journal of Computer Trends and Technology, 72(5), 163–170.

[3] Li, J. (2024). Customer churn prediction using machine learning: A case study of e commerce data. International Journal of Computer Applications, 186(48), 22–25.

[4] Gordini, N., & Veglio, V. (2018). Customers churn prediction and marketing retention strategies. European Journal of Operational Research, 269(2), 760–772.

[5] Win, T., & Bo, K. S. (2020). Predicting customer class using customer lifetime value with random forest algorithm. IEEE Conference Publication, 1, 1–6.

[6] Alshamsi, A. (2022). Customer churn prediction in e-Commerce sector [Institutional repository document]. RIT Digital Institutional Repository.

[7] Jain, D., & Singh, S. (2021). Profitable retail customer identification based on a combined prediction strategy of customer lifetime value. Midwest Social Sciences Journal, 24(1), 106–120.

[8] Borle, S., et al. (2024). A stacked ensemble learning method for customer lifetime value prediction. Kybernetes, 53(7), 2342–2360.

[9] Masood, S. (2024). Predicting sales and analysing customer lifetime value in the e commerce industry using machine learning methods. NORMA.

[10] Zhang, X., et al. (2023). A brief survey of machine learning and deep learning techniques for e commerce research. Journal of Theoretical and Applied Electronic Commerce Research, 18(4), 2188–2216.

[11] Kumar, V., & Reinartz, W. (2024). Customer lifetime value modeling via two stage selected trees ensembles. IEEE Xplore, 1, 1–10.

[12] Dahana, W. D., et al. (2022). A new 360 framework to predict customer lifetime value for multi category e commerce companies. MDPI, 13(8), 373–385.