Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2695~2702
e-ISSN: 2548-6861

Knowledge Discovery on E-Commerce Customer Churn Using Interpretable Machine Learning: A Comparative Study of SHAP-Based Classifiers

Dhita Amanda Ardhani [1]*, Ken Ditha Tania [2]*
* Information Systems Department, Faculty of Computer Science,
Universitas Sriwijaya, Indonesia
09031282227042@student.unsri.ac.id [1], kenya.tania@gmail.com [2]

Article Info

Article history:

Received 2025-08-15
Revised 2025-09-05
Accepted 2025-09-10

Keyword:

Customer Churn,
Customer Retention,
E-Commerce,
Machine Learning,
SHAP.

ABSTRACT

Customer churn remains one of the most pressing issues in the e-commerce sector, as it directly erodes revenue and reduces customer lifetime value. This study proposes an interpretable machine learning approach designed not only to predict churn but also to uncover practical insights that can inform retention strategies. The analysis draws on a publicly available dataset containing customer behavior and transaction records. Data preparation involved handling missing values, applying label encoding, and addressing class imbalance with SMOTE. Five classification models—Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and Gradient Boosting—were trained on an 80:20 stratified split, with performance assessed through accuracy, precision, recall, F1-score, and AUC. Among these, XGBoost delivered the most consistent results, achieving 96% accuracy, 95% precision, 92% recall, and a near-perfect AUC of 0.999, followed closely by Random Forest. Logistic Regression produced the lowest AUC at 0.886. To ensure transparency in decision-making, SHAP (SHapley Additive exPlanations) was applied, revealing Tenure, Complain, and CashbackAmount as the most influential predictors. Longer customer relationships were linked to reduced churn risk, while frequent complaints and higher cashback usage indicated a greater likelihood of leaving. These findings contribute knowledge by blending robust predictive performance with interpretability, enabling e-commerce businesses to design more targeted and proactive customer retention measures.

I. INTRODUCTION

The increasing competitiveness of the e-commerce landscape, driven by rapid digital transformation, has positioned customer retention as a critical success factor for sustainable growth. As acquiring new users becomes more costly than retaining existing ones, predictive analytics using machine learning has emerged as a practical solution to proactively identify potential churners and reduce attrition rates [21], [22], [23]. Recent studies further emphasize the role of explainable and interpretable models in churn prediction, enabling businesses to convert raw behavioral data into actionable retention strategies [24], [26], [30]. Leveraging explainable ML not only boosts prediction performance but also bridges the gap between model output and business decision-making [22], [25], [28].

To address this issue, machine learning (ML) methods have been widely adopted. A bagging-based selective ensemble model was proposed to handle highly imbalanced and high-dimensional data, improving accuracy through hybrid feature selection and cost-sensitive classifiers [1]. Another study applied explainable artificial intelligence (XAI), integrating LightGBM, XGBoost, SHAP, and LIME to enhance both prediction and model interpretability in the telecom sector [2]. Adaptive learning frameworks that combine evolutionary computation and Naïve Bayes have also demonstrated promising results in dynamic environments [3].

A deep learning approach utilizing ConvLSTM and Grey Wolf Optimization with attention mechanisms was introduced to improve customer satisfaction prediction alongside churn classification [4]. In parallel, comparative studies on ML techniques like SVM, Random Forest, and Naïve Bayes have shown varying degrees of success across different telecom datasets [5]. Additional approaches explored inter-customer dependencies using graph-based churn modeling [6], while case-specific churn models have been developed for regional fixed wireless ISPs [7].

Supervised learning methods remain a common benchmark, with algorithms such as KNN, ANN, and RF frequently applied to telecom churn datasets [8]. Cloud service environments pose unique challenges, leading to the development of deep feature extraction techniques integrated with tuned Naïve Bayes classifiers and dimensionality reduction [9]. Moreover, recent research has focused on early-stage churn prediction using hybrid neural network architectures, especially in gaming platforms.

While these studies contribute significantly to the field, most focus solely on either prediction accuracy or model interpretability. There is limited research offering a unified framework that balances both goals using current explainable ML techniques. Moreover, the comparative evaluation of various classifiers with SHAP-based feature explanation in a standardized churn dataset remains underexplored.[10]

The novelty of this study lies in combining robust predictive modeling with knowledge discovery through SHAP-based interpretability. Unlike previous works that primarily emphasized achieving high accuracy, this research highlights how model explanations can reveal hidden behavioral patterns behind churn, thereby transforming predictive outputs into actionable business knowledge. This dual contribution strengthens both methodological rigor and managerial relevance, bridging the gap between machine learning performance and practical decision-making in customer retention.

Furthermore, this study opens potential avenues for future development. Churn prediction frameworks can be integrated with recommendation systems to provide personalized retention strategies, such as targeted promotions or adaptive service recovery plans. In addition, evaluating the model across multiple datasets from different e-commerce domains or real-time transactional environments would enhance the generalizability and robustness of the findings.

Knowledge discovery involves extracting hidden patterns and insights from data, which can then be leveraged to develop accurate and informative predictive models [20]. In the context of churn analysis, this process enables businesses to uncover latent behavioral signals and design early intervention strategies to retain valuable customers. In this study, we present an interpretable churn prediction framework that evaluates the performance of five models it is Logistic Regression, Random Forest, XGBoost, SVM, and Gradient Boosting. By integrating SHAP, the framework not only highlights the most influential features at both global and individual levels but also offers practical insights to guide business strategies and strengthen model evaluation.

II. Methods

To develop an accurate and interpretable churn prediction model, this study follows a structured workflow that blends data-driven machine learning methods with post-hoc model interpretation. The process starts with data collection and preprocessing to ensure quality and consistency, followed by addressing class imbalance through SMOTE. Several classification algorithms particularly tree-based ensemble models are then trained and fine-tuned using stratified cross-validation and randomized hyperparameter search, as recommended in prior comparative studies [11], [12]. To improve the transparency of the model's predictions, SHAP is employed to explain feature contributions at both the global and individual levels. The use of explainable AI in this framework is consistent with earlier research underscoring its value in churn prediction, especially in sectors such as telecommunications and human resources [13], [14]. Additionally, profit-driven optimization and the use of Bayesian tuning for XGBoost have been shown to improve model performance in real-world churn scenarios [15]. Each step in the proposed framework is aligned with methodological recommendations from previous studies on churn prediction [1]–[15]. A flowchart illustrating the overall research framework is presented below.

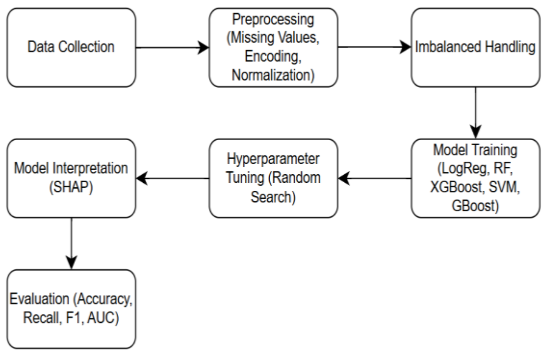

Figure 1. Flowchart Process

This study applies a supervised learning approach using five classifiers Logistic Regression, Random Forest, SVM, XGBoost, and Gradient Boosting due to their proven effectiveness in churn prediction [1], [3], [4], [5], [8], [11], [12]. Ensemble and boosting methods, especially XGBoost, have consistently shown superior performance across domains [11], [12], [15], [16], [17]. SHAP is employed to enhance model transparency by explaining feature contributions globally and locally, as recommended in recent XAI-based churn studies [2], [6], [13], [14]. To address class imbalance, SMOTE is applied during preprocessing [1], [9], with additional support from studies that emphasize its impact on improving minority class recall [13], [16], [18].

Several modifications were introduced to adapt existing techniques to the context of this study. While past studies have applied SHAP post-modeling, this research integrates SHAP throughout the analysis pipeline to guide feature understanding from the start [14]. This approach supports more transparent model development and facilitates earlier identification of influential variables [13], [14]. Additionally, hyperparameter tuning was optimized using Randomized Search CV, an efficient alternative to Grid Search, for identifying optimal model configurations while reducing computation time [4], [5], [15]. The study also standardizes the training and testing process using 10-fold stratified cross-validation, ensuring fair performance comparison across all classifiers [11], [12].

A. Data Collection

This study utilizes an open-source e-commerce churn dataset, which is publicly accessible through the Kaggle platform. The dataset consists of 5,630 customer records, with 950 churners (16.9%) and 4,680 non-churners (83.1%), along with various behavioral and demographic attributes. Each record includes information such as transaction frequency, app usage patterns, satisfaction scores, and preferred digital behaviors, which serve as features for churn prediction. An initial class distribution analysis shows that approximately 4,680 customers did not churn (label 0), while around 950 customers churned (label 1). This substantial disparity indicates a class imbalance problem in the dataset. Such imbalance has been shown to negatively affect classification performance, particularly by reducing the recall of the minority class [1], [5], [9], [21], [24], [27]. To address this challenge, the Synthetic Minority Oversampling Technique (SMOTE) was employed during preprocessing. SMOTE has been widely adopted in recent churn prediction studies for its ability to improve model generalization and sensitivity to minority classes [23], [25], [28].

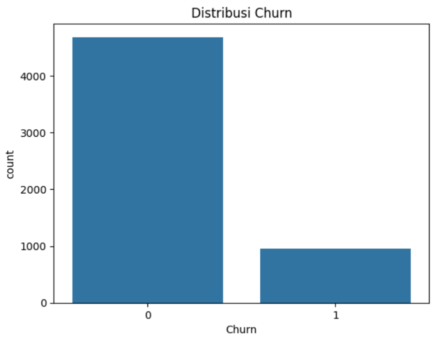

Figure 2. Churn Distribution

The features encompass a mix of numerical and categorical variables capturing customer activity, engagement, and satisfaction metrics.

TABLE 1
FEATURES

No	Features	Data Type	Description
1	OrderCount	Numerical	Number of orders placed by the customer
2	HourSpendOnApp	Numerical	Average hours per week spent in the app
3	Complain	Categorical	Whether the customer has made a complaint (Yes/No)
4	CashbackAmount	Numerical	Total cashback amount received
5	Tenure	Numerical	Duration of customer relationship (months)
6	SatisfactionScore	Numerical	Survey-based customer satisfaction score (1–5)
7	PreferredLoginDevice	Categorical	Primary device used to log in
8	Gender	Categorical	Customer's gender
9	PaymentMethod	Categorical	Preferred payment method
10	NumberOfAddress	Numerical	Total number of shipping addresses linked
11	OrderAmountHikeFromLastYear	Numerical	Increase in order amount compared to the previous year
12	CouponUsed	Categorical	Whether the customer used a coupon
13	NumberOfDeviceRegistered	Numerical	Total devices registered to the account
14	WarehouseToHome	Numerical	Average delivery distance from warehouse to home (km)
15	CityTier	Categorical	Tier classification of the customer's city (1st, 2nd, 3rd)
16	MaritalStatus	Categorical	Customer's marital status
17	PreferedOrderCat	Categorical	Preferred product category for orders
18	DaySinceLastOrder	Numerical	Number of days since the customer last ordered
19	Churn	Binary/Target	Churn label: 0 = retained, 1 = churned

B. Data Preprocessing

The preprocessing stage included handling missing values via statistical imputation (mean for numerical, mode for categorical), encoding categorical features using label or one-hot encoding as appropriate, and normalizing continuous variables with Min-Max scaling to enhance model performance [5], [6], [8], [9], [10]. These steps align with preprocessing standards in churn prediction literature, particularly for improving classifier convergence and interpretability [16], [17], [18], [19].

C. Handling Class Imbalance with SMOTE

Class imbalance is a common issue in churn prediction that can lead to biased learning. This study applies SMOTE to generate synthetic samples of the minority class, improving recall without duplicating records [1], [6], [9]. Prior studies have shown that SMOTE, particularly when combined with ensemble methods, enhances classification performance [5], [16], [17], [18].

D. Model Development and Training

Five classification algorithms were selected for comparative analysis: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), XGBoost, and Gradient Boosting Classifier (GBC). To ensure fair and reliable evaluation, all models were trained and tested using stratified 10-fold cross-validation, which maintains consistent class proportions across folds. This validation approach helps reduce bias and variance, making the performance comparison across models more robust. The choice of algorithms was guided by their demonstrated effectiveness in churn prediction: tree-based models such as RF and XGBoost have repeatedly shown superior performance, while LR remains a dependable baseline [3], [4], [5], [16], [17], [19].

TABLE 2
ALGORITHM

No.	Algorithm	Description
1.	Logistic Regression	A basic statistical model for binary classification, used as a baseline.
2.	Random Forest	A tree-based ensemble algorithm that builds multiple decision trees randomly.
3.	XGBoost	A fast and accurate boosting model, highly effective for churn prediction.
4.	Support Vector Machine (SVM)	A margin-based classification algorithm that performs well on high-dimensional data.
5.	Gradient Boosting	A tree-based boosting technique that improves accuracy by gradually minimizing residuals.

E. Hyperparameter Tuning

To improve model performance, hyperparameter tuning was carried out using Randomized Search Cross-Validation. This method efficiently explores a wide range of parameter combinations while reducing computation time compared to an exhaustive grid search. The tuning process focused on key parameters such as the number of estimators, maximum tree depth, and learning rate for ensemble models (Random Forest, Gradient Boosting, and XGBoost), as well as the regularization coefficient (C) and kernel type for SVM. This optimization ensured that each classifier was evaluated under its best-performing configuration. The tuning process focused on model-specific parameters such as the number of estimators and tree depth for RF and XGBoost, and the regularization coefficient for SVM [4], [5], [8]. This strategy is consistent with findings in [4] and [8], which emphasize the importance of parameter optimization in enhancing classifier performance.

F. Model Interpretation using SHAP

After training, SHapley Additive exPlanations (SHAP) was applied to interpret the model's predictions. SHAP is a model-agnostic method grounded in cooperative game theory, assigning each feature a contribution score that explains both overall model behavior and individual predictions [2], [6], [7]. To visualize these insights, SHAP summary plots were generated to highlight the most influential features in the dataset, while SHAP force plots were created for selected customers to illustrate the reasoning behind their predicted churn probabilities. By incorporating SHAP, this study addresses the limitations of earlier works that relied solely on traditional feature importance measures, as noted in [2] and [6].

G. Model Evaluation and Comparison

The performance of all models was assessed using multiple classification metrics, including Accuracy, Precision, Recall, F1-Score, and the Area Under the ROC Curve (AUC). Using a combination of these metrics offers a well-rounded view of each model's ability to correctly identify both churn and non-churn cases, which is particularly important when dealing with imbalanced datasets [3], [5], [8]. The results were then compared to determine which model achieved the best trade-off between predictive performance and interpretability. This multi-metric evaluation approach is consistent with the recommendations in [3] and [9], which caution against relying solely on accuracy in churn detection tasks.

H. Data Analysis Techniques

All modeling, evaluation, and visualization were carried out in Python (v3.10) using libraries such as scikit-learn, xgboost, imbalanced-learn, and shap. These tools offer the flexibility, scalability, and reproducibility required for effective churn modeling. The experimental pipeline was structured to support both comparative analysis and model interpretability, providing an advantage over the traditional black-box approaches commonly used in earlier studies [2], [4], [6], [10].

III. RESULT AND DISCUSSION

This study assessed the performance of five classification models Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, and XGBoost for predicting customer churn using an e-commerce dataset. Model evaluation employed Accuracy, Precision, Recall, F1-Score, and AUC, with stratified 10-fold cross-validation to ensure consistent class distribution. SHAP was also applied to explain model predictions at both the global and individual levels.

The overall results are presented in Table 1. XGBoost achieved the highest F1-Score and AUC, demonstrating a strong balance between precision and recall an especially important aspect when working with imbalanced data. Random Forest and Gradient Boosting also delivered competitive results, while Logistic Regression and SVM obtained reasonable accuracy but lower recall.

TABLE 3
MODEL PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.88	0.82	0.72	0.76	0.886
Random Forest	0.96	0.97	0.89	0.93	0.999
XGBoost	0.96	0.95	0.92	0.93	0.999
SVM (Balanced)	0.90	0.64	0.89	0.74	0.88
Gradient Boost	0.92	0.87	0.63	0.73	0.91

The overall results are presented in Table 3. XGBoost achieved the highest F1-Score and AUC, demonstrating a strong balance between precision and recall—an especially important aspect when working with imbalanced data. Random Forest also performed consistently well with very high accuracy and precision, showing its robustness in handling nonlinear patterns. Gradient Boosting produced good precision but suffered from lower recall, which makes it less effective in identifying churners. Logistic Regression, while achieving a solid accuracy of 88%, showed relatively weak recall, confirming its limitations in handling complex churn behavior. Meanwhile, SVM performed reasonably, with high recall but low precision, resulting in more false positives. This indicates that while SVM can capture many churners, it also misclassifies a considerable number of retained customers.

Overall, the comparative analysis shows that ensemble-based methods (Random Forest and XGBoost) provide the best trade-off between accuracy, recall, and interpretability, while baseline models such as Logistic Regression still serve as useful benchmarks for evaluating more advanced approaches.

Based on the evaluation results, Random Forest and XGBoost emerged as the top-performing models for churn detection. Logistic Regression achieved 88% accuracy, with a precision of 0.82, recall of 0.72, and an F1-score of 0.76—reflecting solid overall performance but limited sensitivity, consistent with earlier findings on linear models [1], [3].

Random Forest reached 96% accuracy, 0.97 precision, 0.89 recall, and an F1-score of 0.93. XGBoost matched this accuracy but delivered slightly higher recall (0.92), reinforcing its ability to identify churners effectively in imbalanced datasets [1], [4], [5], [15], [16]. These results align with prior research demonstrating the robustness of boosting methods in handling complex churn prediction tasks [17].

With class weighting and standardization, SVM produced a strong recall (0.89) but relatively low precision (0.64), leading to more false positives—a pattern also noted in previous churn studies [8], [19]. Gradient Boosting recorded 92% accuracy and high precision (0.87) but weaker recall (0.63), which limits its suitability for churn detection.

Overall, XGBoost provided the best balance between precision and recall, echoing earlier studies that emphasize its adaptability and high accuracy across various domains [5], [9], [17]. Combined with SHAP analysis, the model offers actionable insights into critical churn drivers, such as declining customer engagement or low satisfaction, enabling businesses to design proactive and targeted retention strategies [2], [6], [7], [10], [18].

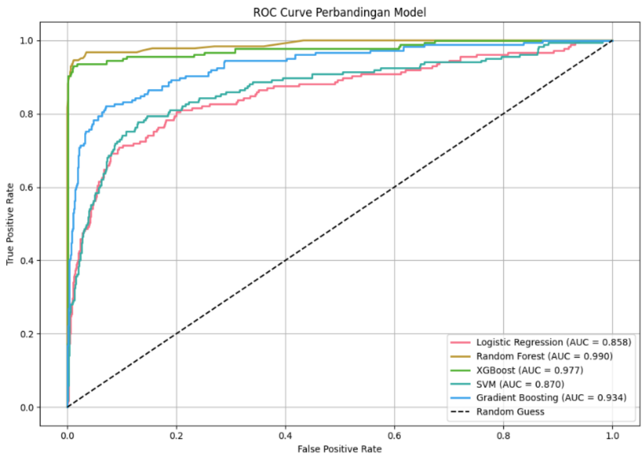

Figure 3. ROC Curve

The ROC Curve analysis further confirms the strength of ensemble models in churn classification. Both Random Forest and XGBoost achieved AUC scores close to 1.0, indicating excellent capability in distinguishing between churn and non-churn customers [1], [4], [5], [16]. This finding is consistent with prior studies that highlight the robustness of ensemble methods in handling imbalanced and nonlinear datasets [17]. Gradient Boosting and SVM also attained strong AUC values (>0.90), though slightly lower than the top-performing models. Logistic Regression recorded a lower AUC (~0.88), reflecting reduced sensitivity in detecting the minority class—an observation also reported in earlier churn research [3], [8], [19]. Overall, these results underscore the superiority of ensemble-based models such as XGBoost and Random Forest, which deliver both high accuracy and strong generalization for real-world applications [15], [17], [18].

Feature importance analysis was conducted using SHAP to evaluate the global contribution of each input variable in churn prediction. As shown in Figure 4, the SHAP summary plot for the XGBoost model identifies Tenure, Complain, and CashbackAmount as the most influential features, followed by SatisfactionScore and OrderAmountHikeFromLastYear. These results align with prior studies that emphasize the role of customer tenure, dissatisfaction signals, and reward patterns in churn behavior [16], [18]. Loyalty duration and satisfaction level are consistently linked to churn likelihood, while complaints and incentive-related features often signal disengagement. Such insights support explainable AI frameworks that enable actionable, domain-specific retention strategies [2], [6], [10], [19].

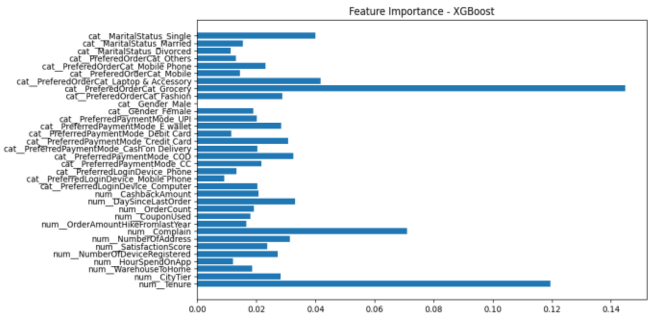

Figure 4. Figure 4. SHAP Summary Plot for XGBoost (Top 20 Features)

Figure 5 presents a simplified SHAP visualization, displaying the top 20 features ranked by contribution weight. Tenure and Complain remain the most dominant, reaffirming that loyalty duration and negative service experiences are strong churn indicators [16], [18]. Other behavioral attributes such as NumberOfAddress, CashbackAmount, and SatisfactionScore also contribute meaningfully, though with lower relative impact. These results suggest that churn is influenced by a mix of temporal, behavioral, and experiential factors, consistent with recent findings in churn prediction research [17], [19].

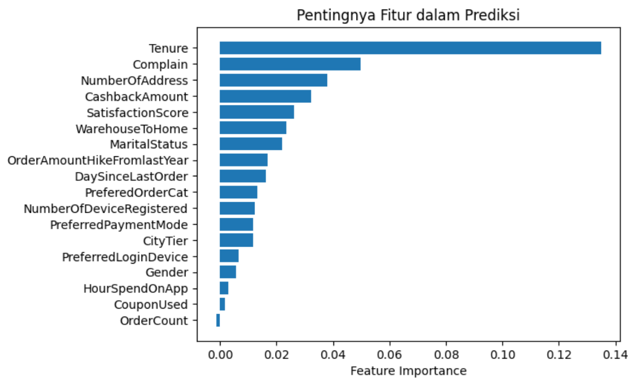

Figure 5. Feature Importance

To assess the overall influence of each feature on the model's predictions, a SHAP summary plot was generated, as shown in Figure 5. This visualization displays the distribution of SHAP values for each feature, indicating both the magnitude and the direction of their impact on churn outcomes. Each dot represents a single observation, with color denoting the original feature value red for high values and blue for low values. This enables an intuitive interpretation of how a feature's value influences the prediction: whether it increases the likelihood of churn (positive SHAP value) or retention (negative SHAP value). By combining a global overview with local variability, the summary plot supports transparent and interpretable decision-making.

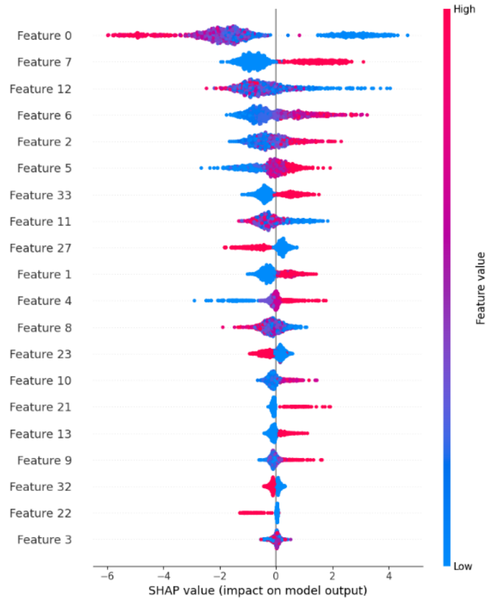

Figure 6. SHAP Value

The SHAP summary plot (Figure 6) identifies Tenure, NumberOfDeviceRegistered, and Complain as the most influential features in predicting churn. Longer tenure (red) was associated with strongly negative SHAP values, indicating a lower churn risk among loyal customers. In contrast, frequent complaints contributed positively to churn probability, highlighting the impact of negative customer experiences [16], [18]. Lower engagement levels—such as reduced *HourSpendOnApp*—were also linked to a higher likelihood of churn. In addition, categorical variables like *PreferredOrderCat* and *PreferredPaymentMode* ranked among the top contributors, suggesting that digital preferences may shape churn behavior. These findings are consistent with prior research that underscores loyalty duration, dissatisfaction signals, and behavioral patterns as key churn drivers [17], [19].

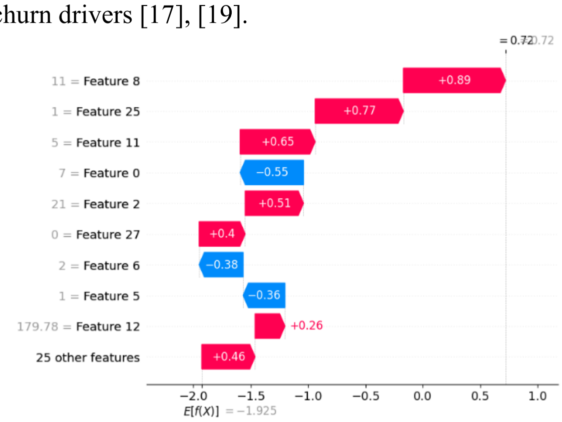

Figure 7. SHAP waterfall plot

The SHAP analysis not only identified the most influential features in predicting churn but also provided insights that can be directly translated into managerial strategies. Several practical implications are summarized as follows:

A. Tenure

Longer customer tenure was associated with lower churn risk. This indicates that loyal customers can be retained more effectively by implementing *loyalty programs*, exclusive membership benefits, or long-term engagement initiatives.

B. Complain

A high number of complaints consistently increased churn probability. Firms should therefore allocate *priority customer service* or establish specialized support channels to handle dissatisfied customers more promptly.

C. CashbackAmount

Despite receiving substantial cashback, some customers still exhibited churn tendencies. This suggests that *monetary incentives alone are insufficient* as retention tools. Companies may need to complement cashback strategies with non-monetary approaches, such as personalized recommendations or improved user experience.

D. HourSpendOnApp

Low engagement levels, such as reduced time spent on the application, emerged as early indicators of churn. Targeted *re-engagement strategies*, including push notifications, personalized offers, and adaptive marketing campaigns, should be deployed to address this risk.

These implications highlight how interpretable machine learning results can extend beyond statistical performance to generate actionable business knowledge. By integrating SHAP-based insights into retention strategies, e-commerce platforms can design proactive interventions that directly address customer needs and behaviors.

To interpret an individual churn prediction, a SHAP waterfall plot was used to track how each feature shifts the model output from the base value (E[f(x)] = –1.925) to the final prediction (f(x) = 0.72). The strongest positive contributors were *PreferredOrderCat* (+0.89), *CashbackAmount* (+0.77), and *NumberOfAddress* (+0.65), suggesting that certain purchase patterns and accumulated rewards may signal disengagement [16], [17].

In contrast, *Tenure* (–0.55), *HourSpendOnApp* (–0.38), and *Gender* (–0.36) reduced the likelihood of churn, highlighting the protective influence of long-term engagement and active app usage. *Complain* contributed moderately (+0.26), indicating that dissatisfaction can play a more nuanced role in individual cases. These personalized explanations demonstrate how SHAP improves both trust and actionability in churn prediction, supporting its use in real-world decision-making systems [2], [10], [18], [19].

In addition to evaluating model performance, this study uncovers behavioral patterns that can be translated into actionable knowledge. SHAP-based interpretations reveal several key insights. For example, longer customer tenure is strongly linked to a lower churn risk, while frequent complaints and higher cashback usage are associated with an increased likelihood of churn. This suggests that dissatisfied customers may still leave despite receiving monetary incentives. Lower app engagement—such as reduced *HourSpendOnApp*—also emerges as an early indicator of disengagement.

These patterns contribute to knowledge discovery by revealing latent relationships between customer behavior and churn that might be overlooked in traditional analyses. Such insights can inform early-warning systems, customer segmentation strategies, and targeted retention campaigns. This perspective aligns with previous research emphasizing the importance of knowledge discovery in supporting real-world decision-making beyond achieving high predictive accuracy.

IV. CONCLUSION

This study demonstrates that ensemble learning models particularly XGBoost and Random Forest deliver the most reliable and accurate performance in predicting customer churn within the e-commerce sector.By integrating SHAP for model interpretability, the research not only achieves high predictive capability but also provides transparent reasoning behind each prediction, addressing a major gap in many previous studies. The findings reveal that features related to loyalty duration, complaints, cashback, and digital engagement are key churn indicators, offering valuable insights for targeted retention strategies. The impact of this research lies in its practical applicability: businesses can adopt the proposed framework to identify high-risk customers and implement data-driven interventions, ultimately reducing churn rates and improving customer lifetime value. Furthermore, these results provide new business knowledge on customer behavior that can be utilized to prevent churn more effectively, in line with recent studies that integrate knowledge discovery frameworks for actionable outcomes. Nevertheless, this study has certain limitations. The analysis was conducted using a single publicly available dataset, which may not fully capture the diversity of customer behaviors across different e-commerce platforms. In addition, the framework has not yet been evaluated in real-time operational settings, which could affect its applicability in dynamic business environments.

Future research should therefore explore the integration of churn prediction frameworks with recommendation systems, enabling more personalized and adaptive retention strategies. Testing the proposed approach on multi-domain or real-time datasets is also recommended to enhance the generalizability and robustness of the findings, ensuring broader applicability across industries.

REFERENCES

[1] B. Zhu, C. Qian, S. vanden Broucke, J. Xiao, and Y. Li, “A bagging-based selective ensemble model for churn prediction on imbalanced data,” *Expert Syst Appl*, vol. 227, Oct. 2023, doi: 10.1016/j.eswa.2023.120223.

[2] D. Asif, M. S. Arif, and A. Mukheimer, “A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry,” *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.104629.

[3] A. Amin, A. Adnan, and S. Anwar, “An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes,” *Appl Soft Comput*, vol. 137, Apr. 2023, doi: 10.1016/j.asoc.2023.110103.

[4] H. Zhang and W. Zhang, “Application of GWO-attention-ConvLSTM model in customer churn prediction and satisfaction analysis in customer relationship management,” *Heliyon*, vol. 10, no. 17, Sep. 2024, doi: 10.1016/j.heliyon.2024.e37229.

[5] R. Krishna, D. Jayanthi, D. S. Shylu Sam, K. Kavitha, N. K. Maurya, and T. Benil, “Application of machine learning techniques for churn prediction in the telecom business,” *Results in Engineering*, vol. 24, Dec. 2024, doi: 10.1016/j.rineng.2024.103165.

[6] K. Ljubičić, A. Merćep, and Z. Kostanjčar, “Churn prediction methods based on mutual customer interdependence,” *J Comput Sci*, vol. 67, Mar. 2023, doi: 10.1016/j.jocs.2022.101940.

[7] K. A. Pflughoeft, N. T. Butz, and A. Corbley, “Customer churn prediction for fixed wireless access: The case of a regional internet service provider,” *Telecomm Policy*, vol. 49, no. 4, May 2025, doi: 10.1016/j.telpol.2025.102929.

[8] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, “Customer churn prediction in telecom sector using machine learning techniques,” *Results in Control and Optimization*, vol. 14, Mar. 2024, doi: 10.1016/j.rico.2023.100342.

[9] S. Arockia Panimalar and A. Krishnakumar, “Customer churn prediction model in cloud environment using DFE-WUNB: ANN deep feature extraction with Weight Updated Tuned Naïve Bayes classification with Block-Jacobi SVD dimensionality reduction,” *Eng Appl Artif Intell*, vol. 126, Nov. 2023, doi: 10.1016/j.engappai.2023.107015.

[10] H. D. Hoang and N. T. Cam, “Do they like your game? Early-stage churn prediction using a two-phase neural network system,” *Eng Appl Artif Intell*, vol. 144, Mar. 2025, doi: 10.1016/j.engappai.2025.110102.

[11] F. E. Usman-Hamza *et al.*, “Empirical analysis of tree-based classification models for customer churn prediction,” *Sci Afr*, vol. 23, Mar. 2024, doi: 10.1016/j.sciaf.2023.e02054.

[12] P. Boozary, S. Sheykhan, H. GhorbanTanhaei, and C. Magazzino, “Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction,” *International Journal of Information Management Data Insights*, vol. 5, no. 1, Jun. 2025, doi: 10.1016/j.jjimei.2025.100331.

[13] H. Habiba Shabbir, M. Hamza Farooq, A. Zafar, B. Ayesha Akram, T. Waheed, and M. Aslam, “Enhancing employee churn prediction with weibull time-to-event modeling,” *Journal of Engineering Research (Kuwait)*, 2025, doi: 10.1016/j.jer.2025.03.009.

[14] S. S. Poudel, S. Pokharel, and M. Timilsina, “Explaining customer churn prediction in telecom industry using tabular machine learning models,” *Machine Learning with Applications*, vol. 17, p. 100567, Sep. 2024, doi: 10.1016/j.mlwa.2024.100567.

[15] Z. Liu, P. Jiang, K. W. De Bock, J. Wang, L. Zhang, and X. Niu, “Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction,” *Technol Forecast Soc Change*, vol. 198, Jan. 2024, doi: 10.1016/j.techfore.2023.122945.

[16] A. De Caigny, K. W. De Bock, and S. Verboven, “Hybrid black-box classification for customer churn prediction with segmented interpretability analysis,” *Decis Support Syst*, vol. 181, Jun. 2024, doi: 10.1016/j.dss.2024.114217.

[17] P. Jiang, Z. Liu, L. Zhang, and J. Wang, “Hybrid model for profit-driven churn prediction based on cost minimization and return maximization,” *Expert Syst Appl*, vol. 228, Oct. 2023, doi: 10.1016/j.eswa.2023.120354.

[18] A. L. D. Loureiro, V. L. Miguéis, Á. Costa, and M. Ferreira, “Improving customer retention in taxi industry using travel data analytics: A churn prediction study,” *Journal of Retailing and Consumer Services*, vol. 85, Jul. 2025, doi: 10.1016/j.jretconser.2025.104288.

[19] J. Sanchez Ramirez, K. Coussement, A. De Caigny, D. F. Benoit, and E. Guliyev, “Incorporating usage data for B2B churn prediction modeling,” *Industrial Marketing Management*, vol. 120, pp. 191–205, Jul. 2024, doi: 10.1016/j.indmarman.2024.05.008.

[20] N. A. Sofiah, K. D. Tania, A. Meiriza and A. Wedhasmara, "A Comparative Assessment SARIMA and LSTM Models for the Gurugram Air Quality Index's Knowledge Discovery," 2024 International Conference on Electrical Engineering and Computer Science (ICECOS), Indonesia, 2024, pp. 26-31, doi: 10.1109/ICECOS63900.2024.10791243.

[21] J. Shobana and C. G. Gangadhar, “E-commerce customer churn prevention using machine learning-based business intelligence strategy,” *Measurement*, vol. 270, Jan. 2023, Art. no. 110998. doi: 10.1016/j.measurement.2023.110998.

[22] I. Boukrouh and A. Azmani, “Explainable machine learning models applied to predicting customer churn for e-commerce,” *International Journal of Artificial Intelligence (IJAI)*, vol. 14, no. 1, pp. 286–297, Feb. 2025. doi: 10.11591/ijai.v14.i1.pp286-297.

[23] S. Kumar, S. Deep, and P. Kalra, “A comprehensive analysis of machine learning techniques for churn prediction in e-commerce: A comparative study,” *International Journal of Computer Trends and Technology (IJCTT)*, vol. 72, no. 5, pp. 163–170, May 2024. doi: 10.14445/22312803/IJCTT-V72I5P119.

[24] J. Maan and H. Maan, “Customer churn prediction model using explainable machine learning,” *arXiv preprint arXiv:2303.00960*, Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.00960

[25] J. Li, “Customer churn prediction using machine learning: A case study of e-commerce data,” *International Journal of Computer Applications*, vol. 186, no. 48, pp. 1–6, Nov. 2024. doi: 10.5120/ijca2024924140.

[26] O. S. Owolabi, A. T. Adepoju, and A. A. Ajayi, “Comparative analysis of machine learning models for customer churn prediction in the U.S. banking and financial services: Economic impact and industry-specific insights,” *Journal of Data Analysis and Information Processing*, vol. 12, pp. 388–418, 2024. doi: 10.4236/jdaip.2024.123021.

[27] A. Almahadeen, “Evaluating machine learning techniques for predicting customer churn in e-commerce sector,” *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 6, pp. 439–450, 2024. [Online]. Available: https://www.aasmr.org/liss/onlinefirst/Vol11/No.6/Vol.11.No.6.27.pdf

[28] S. Baghla and G. Gupta, “Performance evaluation of various classification techniques for customer churn prediction in e-commerce,” *Microprocessors and Microsystems*, vol. 101, Art. no. 104689, Apr. 2023. doi: 10.1016/j.micpro.2023.104689.

[29] D. Y. C. Wang, L. A. Jordanger, and J. C.-W. Lin, “Explainability of highly associated fuzzy churn patterns in binary classification,” *arXiv preprint arXiv:2410.15827*, Oct. 2024. [Online]. Available: https://arxiv.org/abs/2410.15827

[30] H. Ren, “Machine learning-based prediction of customer churn risk in e-commerce,” in *Proc. Int. Conf. on Business Intelligence and Big Data (BIBD)*, Chengdu, China, Oct. 2024, pp. 55–60. doi: 10.1109/BIBD.2024.9932147.