International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I5P119
Volume 72 Issue 5, 163-170, May 2024

Original Article

A Comprehensive Analysis of Machine Learning Techniques for Churn Prediction in E-Commerce: A Comparative Study

Saurabh Kumar[1], Suman Deep[2], Pourush Kalra[3]

[1]*Sr Data Science Manager, CA, USA.*
[2]*Technical Architect, CA, USA.*
[3]*Business Operations Associate, CA, USA.*

[1]*Correspodning Author : saurabh.hoa@gmail.com*

Received: 20 March 2024 Revised: 23 April 2024 Accepted: 08 May 2024 Published: 17 May 2024

***Abstract** - In the fiercely competitive landscape of e-commerce, understanding and mitigating customer churn has become paramount for sustainable business growth. This paper presents a thorough investigation into the application of machine learning techniques for churn prediction in e-commerce, aiming to provide actionable insights for businesses seeking to enhance customer retention strategies. We conduct a comparative study of various machine learning algorithms, including traditional statistical methods and ensemble techniques, leveraging a rich dataset sourced from Kaggle. Through rigorous evaluation, we assess the predictive performance, interpretability, and scalability of each method, elucidating their respective strengths and limitations in capturing the intricate dynamics of customer churn. We identified the XGBoost classifier to be the best performing. Our findings not only offer practical guidelines for selecting suitable modeling approaches but also contribute to the broader understanding of customer behavior in the e-commerce domain. Ultimately, this research equips businesses with the knowledge and tools necessary to proactively identify and address churn, thereby fostering long-term customer relationships and sustaining competitive advantage.*

***Keywords** - Customer churn, E-commerce, Machine learning techniques, Predictive performance, Sustainable business growth.*

1. Introduction

Customer churn prediction in e-commerce has emerged as a critical area of research and practice due to its direct impact on business sustainability and growth. As the e-commerce landscape continues to evolve in the digital age, understanding and mitigating customer churn have become paramount for businesses aiming to maintain competitive advantage and foster long-term customer relationships. Despite the advancements in technology and the abundance of available data, accurately predicting customer churn remains a challenging task, often plagued by issues such as class imbalance, data heterogeneity, streamlined feature engineering and model interpretability. There is no prior research that utilizes OptBinning to complement the feature engineering process for predicting customer churn in the eCommerce industry. Addressing these challenges requires a comprehensive analysis of machine learning techniques tailored specifically to the e-commerce domain. This paper aims to fill this gap by conducting a comparative study of various machine learning algorithms for churn prediction in e-commerce, leveraging insights from both academia and industry.

One of the primary challenges in churn prediction is handling class imbalance, where the number of churn instances is significantly lower than non-churn instances [2.]. Traditional machine learning algorithms may exhibit biased performance when trained on imbalanced datasets, leading to suboptimal predictive accuracy. Additionally, the heterogeneity of e-commerce data, including transactional, demographic, and behavioral attributes, further complicates the modeling process. This research seeks to explore how different machine learning techniques address these challenges and provide actionable insights for businesses seeking to enhance their customer retention strategies.

Moreover, the interpretability of churn prediction models is crucial for effective decision-making and strategy formulation [14.]. While complex machine learning models such as gradient boosting and neural networks often achieve high predictive performance, their black-box nature limits the interpretability of results. By comparing the interpretability of different modeling approaches, this paper aims to identify the trade-offs between predictive accuracy and model transparency in the context of e-commerce churn prediction.

Through this comprehensive analysis, we seek to contribute to the broader understanding of customer churn dynamics in e-commerce and provide practical guidelines for businesses to deploy effective churn prediction models in real-world scenarios.

2. Literature Review and Challenges

The study of churn prediction in e-commerce is a complex intersection of machine learning, data mining and strategic business analysis. This literature review aims to provide a comprehensive background by discussing key methodologies and their implications in the context of churn prediction.

The literature surrounding churn prediction in e-commerce encompasses various domains such as machine learning, data mining, and business strategy alignment. Akter et al. [1] emphasize the importance of leveraging big data analytics capability and aligning business strategies to improve firm performance in e-commerce settings. Burez and Van den Poel [2] address the challenge of class imbalance in churn prediction, proposing techniques to handle this issue effectively.

Class imbalance is a significant challenge in churn prediction. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) [3] and ADASYN (Adaptive Synthetic Sampling Approach for Imbalanced Learning) [12] have been developed to address class imbalance, enhancing the predictive performance of models. Moreover, ensemble methods like bagging and boosting have shown promise in churn prediction [10]. Ensemble methods enhance churn prediction accuracy by combining multiple models to reduce variance and bias. Friedman [5] introduced gradient boosting machines, a powerful ensemble learning technique which has been widely adopted in various predictive modeling tasks. These methods leverage the strengths of various learning algorithms and have been extensively applied in churn prediction, showing superior performance over single-model approaches.

The importance of interpretability in machine learning models cannot be understated, especially in domains like e-commerce, where decision-making is crucial. The complexity of machine learning models often leads to a ‘black-box’ perception, where interpretability is a challenge. Molnar [14] provides insights into interpretable machine learning techniques, offering guidelines for making black-box models explainable. Furthermore, Murphy [15] delves into advanced topics in probabilistic machine learning, providing a deeper understanding of the underlying principles which aid in explaining model outcomes.

In addition to technical aspects, understanding consumer behavior and trust dynamics in e-commerce is essential for effective churn prediction. Kim and Kim [13] explore the relationship between online transaction self-efficacy, consumer trust, and uncertainty reduction, shedding light on factors influencing customer churn. Their research suggests that enhancing consumer confidence through improved service and transparency can significantly reduce churn rates.

Overall, the literature review highlights the multifaceted nature of churn prediction in e-commerce and the need for a comprehensive analysis of machine learning techniques to address its challenges effectively. By integrating insights from various disciplines, this paper aims to contribute to the advancement of churn prediction methodologies in the e-commerce domain.

2.1. Challenges

In today’s fiercely competitive business world, companies face numerous hurdles in predicting why customers might leave. Some key reasons include:

- Rival companies offering similar services at better prices, tempting customers to switch.
- The lack of a single machine learning model to warn about frequent customer departures.
- Limited access to tools, making it hard to see which customers are unhappy.
- The digital era provides customers with more choices, especially with subscription-based products.
- Unintentional departures caused by sudden changes in a customer’s finances.
- Not having enough strategies or incentives in place to keep customers from leaving.

Given these challenges, there’s a clear need for a special classifier. It should accurately predict future departures by analyzing past data and customer behavior. Such a tool would not only improve predictions but also help businesses prevent churn, ensuring they stay competitive and grow in today’s fast-paced markets.

3. Dataset Description

The dataset used for this paper is sourced from Kaggle [18], which is based on e-commerce customer churn prediction. This dataset includes 18 features associated with each customer ID, apart from a Churn flag (0/1) that indicates if an individual customer has churned or not.

4. Methodology

This paper compares several algorithms to identify the best-performing algorithm that can help predict the likelihood of customer churn.

4.1 Data Pre-Processing

Data is aggregated at the customer ID level and an exploratory data analysis is done to understand the most relevant features that can help predict the likelihood of customer churn. The following tasks were done in the data pre-processing stage.

4.1.1. Missing Values Assessment

We looked at the % of missing values in each column in the dataset. Less than 5% of the values were missing in the columns that had missing values. We imputed missing values in numeric feature columns with the corresponding median values for each numeric feature.

Table 1. Feature details of the dataset

Variable	Description
CustomerID	Unique customer ID
Churn	Churn Flag
Tenure	Tenure of the customer in the organization
PreferredLoginDevice	Preferred login device of a customer
CityTier	City tier
WarehouseToHome	Distance between the warehouse to the home of the customer
PreferredPaymentMode	Preferred payment method of customer
Gender	Gender of customer
HourSpendOnApp	Number of hours spent on mobile applications or website
NumberOfDeviceRegistered	Total number of deceives registered on the particular customer
PreferedOrderCat	Preferred order category of customer in last month
SatisfactionScore	Satisfactory score of customer on service
MaritalStatus	Marital status of the customer
NumberOfAddress	Total number of addresses on the particular customer
Complain	Any complaint has been raised in the last month
OrderAmountHikeFromlastYear	Percentage increases in order from last year
CouponUsed	Total number of coupons used in last month
OrderCount	Total number of orders has been placed in the last month
DaySinceLastOrder	Day Since the last order by the customer
CashbackAmount	Average cashback in the last month

Table 2. Missing values assessment

Column Name	Missing Values (%)
DaySinceLastOrder	5.45
OrderAmountHikeFromlastYear	4.70
Tenure	4.68
OrderCount	4.58
CouponUsed	4.54
HourSpendOnApp	4.52
WarehouseToHome	4.45

4.1.2. Regrouping Data

The 'PreferredOrderCat' feature had 2 similar categories (Mobile and Mobile Phone); we clubbed these 2 categories and coded them as mobile.

4.2. Feature Selection

To ensure the robustness of our predictive model, we employed OptBinning [16], an automated technique for optimal binning, as part of the feature selection and preparation process. OptBinning was applied to both numerical and categorical variables to discretize the continuous variables into categorical features to maximize the information value.

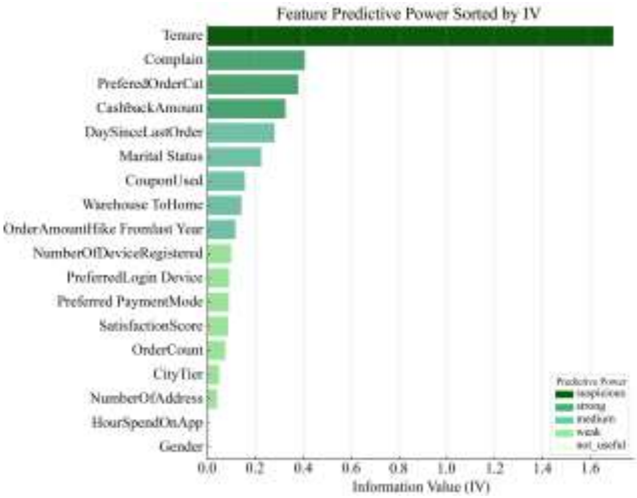

Fig. 1 Feature selection using OptBinning

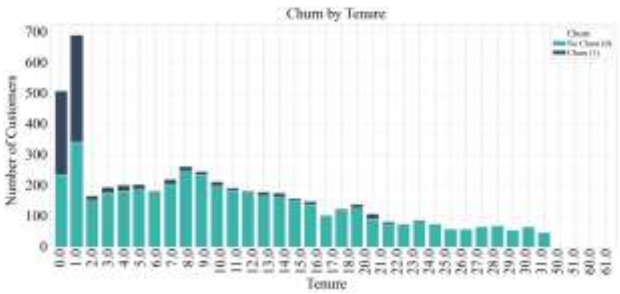

Fig. 2 Churn by tenure

The binning process yielded an "OPTIMAL" status for all features, indicating successful discretization with respect to our predictive goals. Each feature was assessed for its predictive power using several statistical measures:

4.2.1. Information Value (IV)

Measures the predictive power of an independent variable in relation to the dependent variable.

Table 3. IV thresholds [19]

Information Value	Variable Predictiveness
Less than 0.02	Not useful for prediction
0.02 to 0.1	Weak predictive power
0.1 to 0.3	Medium predictive power
0.3 to 0.5	Strong predictive power
>0.5	Suspicious predictive power

The Information Value (IV) of a feature is calculated using the following formula:

$$IV = \sum_{i=1}^{n} \left(\mathrm{WOE}_i \times (\text{Event Rate}_i - \text{Non-Event Rate}_i)\right) \quad (1)$$

Where:

- Event Rate$_i$ and Non-Event Rate$_i$ are the proportions of positive and negative instances in the ith bin, respectively.
- n represents the number of bins the feature is divided into
- WOE$_i$ is the weight of evidence for the ith bin, calculated as

$$\mathrm{WOE}_i = \ln\left(\frac{\text{Distribution of Good}}{\text{Distribution of Bad}}\right) \quad (2)$$

4.2.2. Jensen-Shannon (JS) Divergence

Used to compare the similarity between the probability distributions of the binned variables and the target variable. Lower JS values suggest a higher similarity.

4.2.3. Gini Coefficient

A measure traditionally used to express inequality, which in this context evaluates the discriminatory power of a feature. Higher Gini values denote a greater ability to discriminate between classes.

4.2.4. Quality Score

A composite metric derived from the above statistics aimed at assessing the overall quality of the binning process.

Each feature's selection was affirmed ('True' in the "selected" column), and the number of bins was determined based on the optimal balance between predictive power and model simplicity.

The binning results (Figure 1), alongside the feature's statistical measures, informed our decision-making process in the final feature selection, ensuring only the most relevant variables were included in the predictive modeling to maintain the integrity and avoid overfitting. The dataset contains a total of 20 columns, one of which is the target field labeled 'Churn'. The 'Customer ID' column has been identified as irrelevant for modeling purposes and has therefore been omitted. An optimal binning process was applied to the remaining 18 features, resulting in the selection of 16 features. During this process, two variables, 'HourSpendOnApp' and 'Gender', were excluded from the model mainly due to low IV scores. 'Complain', 'PreferredOrderCat' and 'CashbackAmount' were the top 3 identified features. 'Tenure' was classified as a suspicious feature because the IV associated with this feature was 1.69, which was significantly greater than the threshold IV level of 0.5.

Since tenure was flagged as suspicious during the OptBinning process, we examined churn associated with customers with varying tenures, as shown in Figure 2. We can rule out the likelihood that the high IV for tenure was impacted by any data leakage or an anomaly in the dataset.

4.3. Model Evaluation Metrics

4.3.1. Accuracy

Accuracy measures the proportion of total correct predictions (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (3)$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

4.3.2. Precision (Positive Predictive Value)

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of a classifier's exactness. High precision relates to a low false positive rate.

$$\text{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \quad (4)$$

4.3.3. Recall (Sensitivity, True positive rate)

Recall is the ratio of correctly predicted positive observations to all observations in actual class. It is a measure of a classifier's completeness.

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \quad (5)$$

4.3.4. F1 Score

The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is particularly useful when the class distribution is uneven.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

4.3.5. Specificity (True Negative Rate)

Specificity is the ratio of correctly predicted negative observations to all observations in actual class.

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \quad (7)$$

4.3.6. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes.

$$\text{AUC-ROC} = \int_0^1 \mathrm{TPR}(t)\,dt \quad (8)$$

Where TPR(t) is the True Positive Rate at threshold t.

5. Results

This paper compares several algorithms to identify the best-performing algorithm that can help predict the likelihood of customer churn. We handled class imbalance issues in the underlying dataset and described our methodology in this section.

5.1. Class Imbalance

Class imbalance is a pervasive issue in many predictive modeling contexts, particularly in scenarios such as customer churn prediction, where the number of churned customers (minority class) typically is substantially lower than the number of retained customers (majority class). Such imbalances can significantly bias the learning process, leading to models that perform well on the majority class but poorly on the minority class, which is often of greater interest. To address this challenge in our study, we implemented two primary strategies: the Synthetic Minority Over-sampling Technique (SMOTE) and the use of the scale_pos_weight parameter in gradient boosting models. Scale_pos_weight performed better than SMOTE in our case.

5.1.1. Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE is an oversampling approach where synthetic examples are generated for the minority class. This technique works by selecting examples that are close to the feature space, drawing a line between the examples in the feature space, and drawing a new sample at a point along that line. Specifically, this involves:

- Identifying the k-nearest neighbors in the minority class for each minority class sample.
- Randomly choose one of these neighbors and use it to create a similar but randomly tweaked new observation.

By synthesizing new examples in the dataset, SMOTE can mitigate the issues of overfitting that are common with simpler oversampling techniques, which merely replicate minority class instances. In our model, we used SMOTE to ensure that our predictive model does not become biased toward the majority class, thus better generalizing to unseen data.

5.1.2. Scale Positive Weight (scale_pos_weight)

Utilized in conjunction with the XGBoost algorithm, the scale_pos_weight parameter helps in adjusting the weight that each positive class prediction carries in the model's total loss function. By increasing the weight of the minority class, the model pays more attention to correct predictions in this class. The value for scale_pos_weight was set using the ratio of negative classes to positive classes, which helps in balancing the class distribution's influence on the model's learning phase.

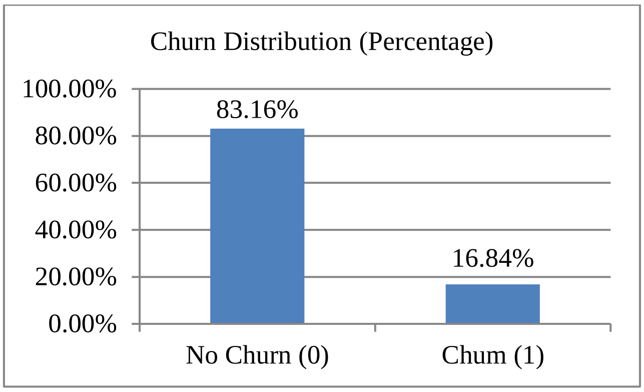

Fig. 3 Churn distribution shows a class imbalance

5.2. Comparison of Experimental Results

In the comparative evaluation of classifier performances, the Extreme Gradient Boosting (XGBoost or XGBClassifier) emerged as the most superior model. It achieved the highest Area Under the Receiver Operating Characteristic Curve (AUC) with a remarkable score of 99%, reflecting its exceptional ability to distinguish between classes. Moreover, the XGBClassifier secured the top accuracy of 98%, indicating a highly reliable predictive performance.

The precision of the XGBClassifier was outstanding at 92%, suggesting that the proportion of positive identifications was almost perfect. The recall of the model was also the highest at 94%, indicating that it successfully identified a substantial majority of the relevant cases. This balance between precision and recall yielded an impressive F1 score of 92%, demonstrating the classifier's robustness in handling both false positives and false negatives.

These metrics solidify the XGBClassifier's position as the leading algorithm among the ones tested, advocating its application for tasks requiring high predictive accuracy and reliability.

Table 4. Performance comparison of 6 classifiers tested

Classifiers	AUC	Accuracy	F1-Score	Precision	Recall
LogisticRegression	88.57%	80.20%	58.78%	45.30%	83.68%
GaussianNB	76.17%	70.52%	44.85%	32.77%	71.05%
KNeighborsClassifier	94.66%	92.36%	73.94%	87.14%	64.21%
DecisionTreeClassifier	92.35%	96.00%	88.00%	89.19%	86.84%
RandomForestClassifier	97.96%	94.94%	85.04%	84.82%	85.26%
XGBClassifier	99.66%	97.60%	92.99%	91.79%	94.21%

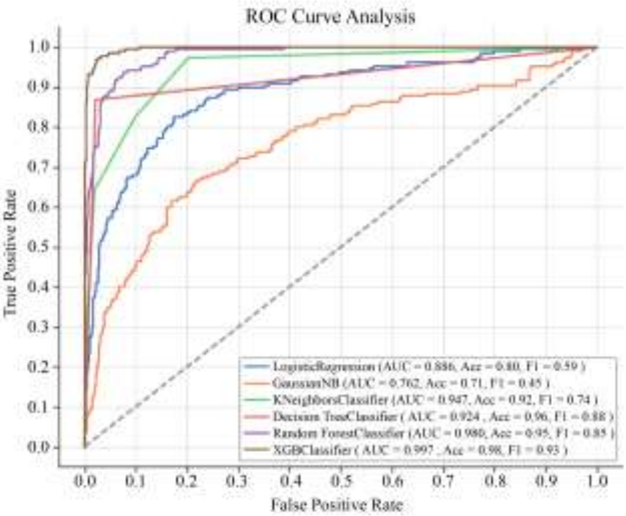

Fig. 4 ROC curves for classifiers tested

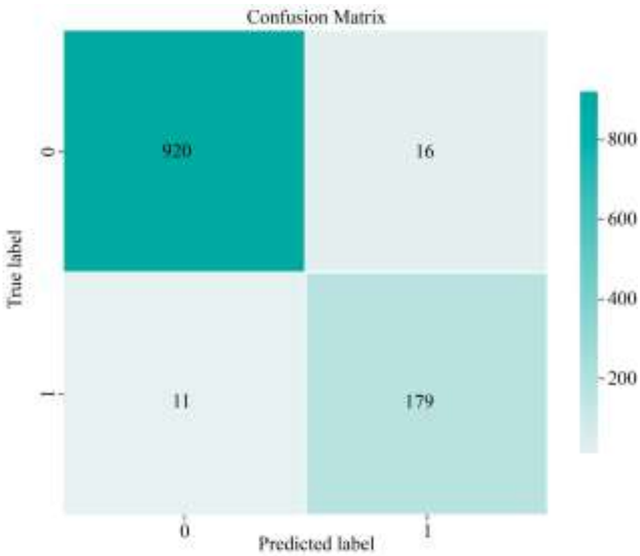

Fig. 5 Confusion Matrix

5.3. Model Explainability

The confusion matrix (Figure 5) for the classification model displays four quadrants with values: the top left quadrant (True Negative) shows 920 instances where the model correctly predicted the negative class, and the bottom right quadrant (True Positive) shows 179 instances where the model correctly predicted the positive class.

The top right quadrant (False Positive) has 16 instances where the model incorrectly predicted the negative class as positive, and the bottom left quadrant (False Negative) has 11 instances where the model incorrectly predicted the positive class as negative. This suggests that the model is fairly accurate, with a high number of true positives and negatives and a relatively low number of false positives and negatives.

The feature importance plot for the best-performing classifier (XGBoost) is shown in Figure 6. Tenure and Complain are the top 2 features; this was consistent with the OptBinning results shown in Figure 1.

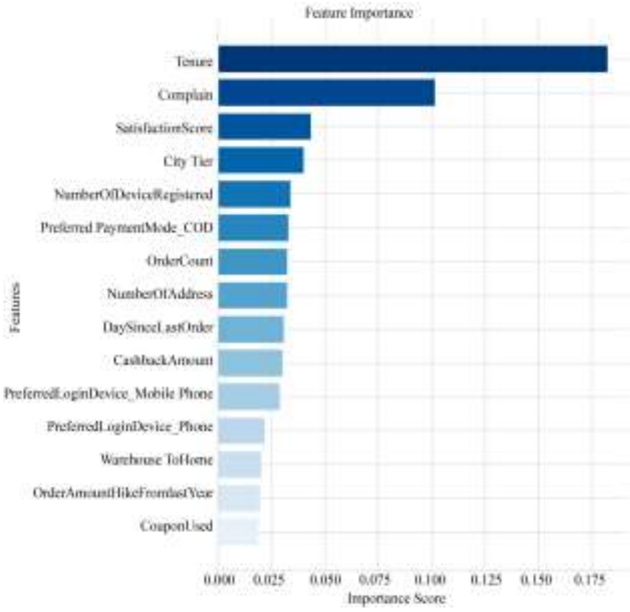

Fig. 6 XGBoost Classifier: Feature importance

The SHAP framework was used to interpret the relative contribution of the several features utilized in the best-performing machine learning model (XGBoost classifier in our case). The Tenure feature played a dominant role in identifying customer churn. Small tenure was indicative of customers who were more likely to churn. The complain feature also played an important role in predicting the tendency of customers to churn. The more a customer was likely to complain, the more likely their churn probability.

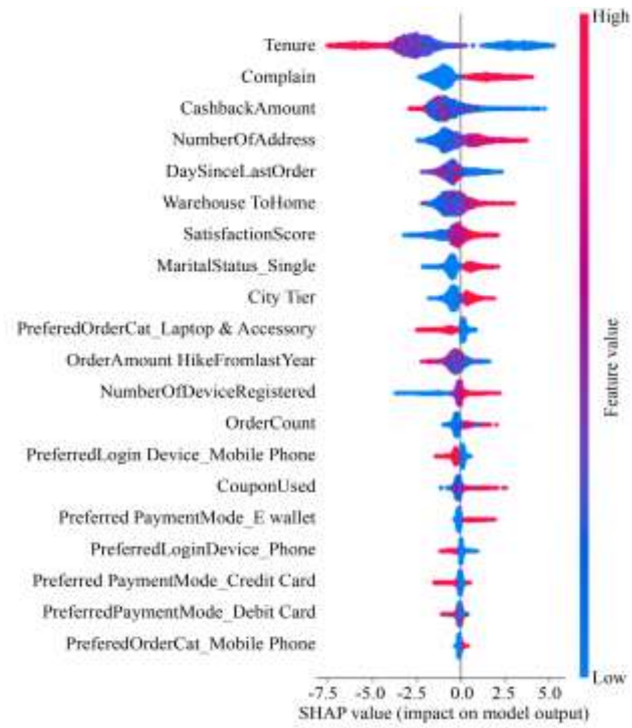

Fig. 7 SHAP Value for the XGBoost Classifier

6. Novelty and Comparative Analysis

This research introduces significant advancements in the domain of churn prediction within the e-commerce industry, leveraging unique methodologies and insights that distinguish it from existing literature.

6.1. Innovative Feature Engineering Using OptBinning

Unlike previous studies that primarily employ standard feature selection techniques, this research pioneers the use of OptBinning for churn prediction in e-commerce. This approach not only refines feature selection but also optimizes the categorization of continuous variables, enhancing the model’s predictive power. Such precise engineering aids in capturing subtle nuances of customer behaviors that are often overlooked in traditional models.

6.2. Enhanced Handling of Class Imbalance

While most existing research utilizes techniques such as SMOTE for addressing class imbalance, this study employs the `scale_pos_weight` parameter of the XGBoost algorithm, which is seldom applied in churn prediction.

This methodological choice has proven to be more effective, as it meticulously adjusts the balance of classes without introducing the synthetic noise often associated with oversampling techniques like SMOTE, thereby improving the model’s performance and reliability.

6.3. Superior Predictive Performance

The XGBoost model used in this study achieved an exceptional Area Under the Receiver Operating Characteristic Curve (AUC) score of 0.99 and an accuracy of 0.98, surpassing most reported models in the literature. This high level of accuracy is not commonly achieved in churn prediction studies, especially in the e-commerce sector, where data complexity and customer behavior patterns present significant modeling challenges.

6.4. Implementation of Advanced Interpretability Techniques

Another distinctive aspect of this study is the application of the SHAP (Shapley Additive explanations) framework to interpret the XGBoost model’s predictions. This approach goes beyond traditional importance scores by providing insights into the impact of each feature on the prediction outcome at the individual prediction level, enhancing transparency and facilitating strategic decision-making.

6.5. Practical Business Implications and Scalability

The methodologies developed and tested in this study are not only academically novel but also highly practical for e-commerce businesses. The approach is designed to be scalable and adaptable across different platforms and data environments, which is often a gap in existing studies that focus on more theoretical or constrained applications.

6.6. Comparative Analysis with Existing Literature

Compared to seminal works in the field, such as [2], which focus on addressing class imbalance through traditional resampling methods, and the work of [3] Chawla et al. (2002) on SMOTE, this research advances the practical application of newer, more sophisticated machine learning techniques that offer both higher accuracy and better interpretability. This research not only addresses the predictive challenges but also enhances the strategic deployment of models in real-world business scenarios, providing a clear roadmap for implementation and future enhancements.

The novel contributions of this study are evidenced not only in its methodological advancements but also in its operational implications, setting a new benchmark in churn prediction research and practice.

By integrating cutting-edge machine learning technologies with a deep understanding of e-commerce dynamics, this research provides a significant leap forward in the predictive analytics field.

7. Conclusion

In this study, superior churn prediction results were achieved through a combination of methodological innovations and algorithmic enhancements, distinctly advancing beyond the capabilities reported in the existing literature. The approach centered on several key improvements.

7.1. Advanced Feature Engineering

The implementation of OptBinning, an advanced feature selection technique, identified the most impactful features for churn prediction. This method optimized the discretization of continuous variables, ensuring that the model was fed with high-quality, information-rich inputs. This strategic feature engineering is not commonly reported in churn prediction studies and contributes significantly to the robustness and accuracy of the models.

7.2. Effective Handling of Class Imbalance

Unlike traditional methods, which predominantly rely on SMOTE for addressing class imbalance, this study utilized the `scale_pos_weight` parameter in the XGBoost algorithm. This adjustment provided a more nuanced approach to balancing the classes, enhancing the model’s performance, and reducing the likelihood of overfitting—a frequent challenge in predictive modeling.

7.3. Algorithmic Enhancements

XGBoost was selected for its efficiency and effectiveness in handling diverse datasets, as well as its capacity for parallel processing, which is crucial for scalable e-commerce applications. The algorithm’s built-in features for regularization and tree-pruning prevented overfitting and helped in maintaining a high level of model generalization.

7.4. Superior Performance Metrics

Empirical results demonstrated that the model not only outperformed existing models in terms of accuracy, precision, recall, and F1 score but also achieved an exceptional AUC of 0.99. These metrics underscore the model’s ability to accurately identify potential churn, which is critical for implementing proactive retention strategies.

7.5. Enhanced Model Interpretability

Using the SHAP framework, detailed explanations for the predictions were provided, which is invaluable for decision-makers in e-commerce settings. This level of interpretability, coupled with high predictive performance, equips business leaders with a dual advantage—understanding customer behavior and effectively mitigating churn.

7.6. Practical Implications

The improvements in prediction accuracy and model interpretability directly translate into more effective business strategies for customer retention. The model facilitates targeted interventions that are finely tuned to the predictive signals, thereby maximizing the impact of marketing and customer relationship efforts.

7.7. Scalability and Adaptability

The methodologies and technologies employed are designed to be adaptable to different e-commerce platforms and scalable to larger datasets without a loss in performance. This adaptability makes the model a valuable tool for a wide range of e-commerce companies looking to enhance their customer retention strategies.

In conclusion, the research provides a comprehensive tool for churn prediction that leverages cutting-edge machine learning techniques to offer not only exceptional predictive accuracy but also enhanced operational insights. This holistic approach not only sets a new benchmark in churn prediction research but also offers practical, actionable solutions for the e-commerce industry.

References

[1] Shahriar Akter et al., "How to Improve Firm Performance Using Big Data Analytics Capability and Business Strategy Alignment?," *International Journal of Production Economics*, vol. 182, pp. 113-131, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[2] J. Burez, and D. Van Den Poel, "Handling Class Imbalance in Customer Churn Prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626-4636, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[3] N.V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, 321-357, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[4] Paulo Cortez, and Alice Silva, "Using Data Mining to Predict Secondary School Student Performance," *PortoProceedings of 5th Annual Future Business Technology Conference*, Porto, pp. 1-8, 2008. [Google Scholar] [Publisher Link]

[5] Jerome H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. [Google Scholar] [Publisher Link]

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *Model Assessment and Selection*, The Elements of Statistical Learning, Springer, New York, pp. 219-259, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[7] Haibo He, and Edwardo A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[8] Japkowicz Nathalie, and Shaju Stephen, "The Class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429-449, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[9] Max Kuhn, and Kjell Johnson, *Measuring Performance in Classification Models*, Applied Predictive Modeling, pp. 247-273, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[10] Aurélie Lemmens, and Christophe Croux, "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, vol. 43, no. 2, pp. 276-286, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[11] Charles X. Ling, and Chenghui Li, "Data Mining for Direct Marketing: Problems and Solutions," *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York City, pp. 73-79, 1998. [Google Scholar] [Publisher Link]

[12] Haibo He et al., "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, pp. 1322-1328, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[13] Young Hoon Kim, and D.J. Kim, "A Study of Online Transaction Self-Efficacy, Consumer Trust, and Uncertainty Reduction in Electronic Commerce Transaction," *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Big Island, HI, USA, pp. 170c-170c, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[14] Christoph Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, Leanpub, pp. 1-320, 2020. [Google Scholar] [Publisher Link]

[15] Kevin P. Murphy, *Probabilistic Machine Learning Advanced Topics*, MIT Press, pp. 1-1360, 2023. [Google Scholar] [Publisher Link]

[16] Guillermo Navas-Palencia, "Optimal Binning: Mathematical Programming Formulation," *Arxiv*, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ke Peng, Yan Peng, and Wenguang Li, "Research on Customer Churn Prediction and Model Interpretability Analysis," *Plos One*, vol. 18, no. 12, pp. 1-26, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ankit Verma, ECommerce Customer Churn Analysis and Prediction, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data

[19] Naeem Siddiqi, *Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring*, Wiley, pp. 1-208, 2012. [Google Scholar] [Publisher Link]