Boosting Algorithms for Customer Reload Prediction: Optimizing Outcomes with XGBoost, AdaBoost, and CatBoost

Kusnaeni, Hartina Husain, Muhammad Rifki Nisardi, Wahyuni Ekasasmita, Nur Rahmi

Technology Bacharuddin Jusuf Habibie Institute of Technology, Parepare 91122, Indonesia
Kusnaeni25@ith.ac.id

Abstract. The telecommunications industry generates large volumes of data daily, creating significant opportunities for deeper analysis. One of the main challenges faced by companies in this sector is predicting customer reload behavior, where customers decide to top up or reload their services, which directly impacts a company's profitability. Accurate reload prediction is therefore crucial in helping companies develop effective retention strategies. This study evaluates three machine learning algorithms based on boosting: XGBoost, AdaBoost, and CatBoost, to build a customer reload prediction model. Boosting algorithms are known for their ability to iteratively correct prediction errors, improve model accuracy, and handle complex and imbalanced data. Through experiments conducted on customer data from the telecommunications industry, the results showed that XGBoost achieved an accuracy of 90.04%, AdaBoost 87.02%, and CatBoost outperformed with an accuracy of 88.09%. These findings demonstrate that XGBoost offers the most effective solution for predicting reload behavior, while XGBoost and AdaBoost also provide solid results. This research provides valuable insights for telecommunications companies in identifying customers likely to reload, as well as designing better, data-driven retention strategies.

Keywords: Machine Learning, Boosting Algorithms, Reload Behavior.

1 INTRODUCTION

In today's highly competitive telecommunications industry, prepaid services are a significant revenue driver for many companies. A crucial aspect of this model is understanding customer reload behavior—the frequency with which customers top up their prepaid accounts to continue accessing services [1]. Predicting customer reload patterns is pivotal for telecommunications companies, as it directly impacts revenue, marketing strategies, and customer retention. Despite its importance, most research in the field has focused predominantly on churn prediction, which analyzes the likelihood of a customer leaving the service, rather than on reload behavior [2]. This gap in research highlights the need for more comprehensive studies on reload prediction, particularly using advanced machine learning techniques. Accurately predicting customer reload behavior is a complex task influenced by numerous factors

M. Nohong et al. (eds.), *Proceedings of the 9th International Conference on Accounting, Management, and Economics 2024 (ICAME 2024)*, Advances in Economics, Business and Management Research 331,
https://doi.org/10.2991/978-94-6463-758-8_185

such as usage patterns, customer demographics, promotional campaigns, and even economic circumstances [3]. Traditional statistical models often fall short in addressing this complexity, especially when dealing with large, imbalanced datasets common in the telecommunications industry [4]. Recent advancements in machine learning, particularly boosting algorithms, provide promising solutions to this challenge [5]. Boosting algorithms like XGBoost, AdaBoost, and CatBoost are designed to improve predictive accuracy by iteratively focusing on data points that are difficult to classify, making them ideal for complex problems like reload prediction [6].

Machine learning has been applied extensively across various domains for predictive modeling. XGBoost has been widely recognized for its scalability and efficiency in handling large, complex datasets, particularly in fields such as finance and healthcare [7]. AdaBoost, known for its simplicity, has been effective in improving accuracy by weighting misclassified instances more heavily [8]. CatBoost, which specializes in handling categorical data, has demonstrated superior performance in tasks requiring the interpretation of intricate, non-linear relationships within the data [9]. While these algorithms have been applied to customer churn prediction in telecommunications [10], there remains a noticeable gap in their application for predicting customer reload behavior, marking this research as a critical and novel contribution to the field. Given the complexity of customer reload behavior, there is an increasing need to explore more sophisticated models that can capture the intricacies of customer interactions. The current study seeks to address this gap by investigating the performance of XGBoost, AdaBoost, and CatBoost in predicting customer reloads. In doing so, we aim to provide a comprehensive comparison of these algorithms and identify the key factors driving customer reload decisions, utilizing feature importance analysis. The growing body of literature on machine learning applications in telecommunications emphasizes the need for models that not only predict behavior but also provide interpretable insights into the factors influencing customer actions. As discussed by [11], the ability to understand why a model predicts certain outcomes is as important as the prediction itself, especially for actionable decision-making in business. By leveraging feature importance techniques inherent to boosting algorithms, this study aims to offer telecommunications companies insights into the variables most strongly associated with customer reloads, allowing for more targeted and efficient marketing strategies.

Furthermore, this study addresses a significant gap in the current telecommunications research by providing practical insights specifically into customer reload prediction, an area that has received considerably less attention compared to churn prediction. While much of the existing research has concentrated on predicting customer churn, understanding and predicting reload behavior is equally critical, as it directly influences revenue streams and customer engagement. This underexplored area presents an opportunity to enhance decision-making processes and improve marketing effectiveness. By comparing three advanced boosting algorithms —XGBoost, AdaBoost, and CatBoost—this research offers a thorough evaluation of which model performs best in the context of customer reload prediction. Previous studies have applied these algorithms in related fields like customer churn prediction, but their application in the domain of reload behavior prediction is limited, leaving a clear research gap. Filling this gap not only contributes to academic literature but also provides telecommunications companies with actionable insights to refine their marketing strategies, optimize resource allocation, and ultimately improve customer retention.

In addition, this work sets a foundation for future studies to broaden the scope of machine learning applications to other aspects of customer behavior prediction, such as subscription renewals in industries that rely heavily on recurring payments or reloads. In summary, this paper aims to fill the identified research gap by evaluating the effectiveness of XGBoost, AdaBoost, and CatBoost in predicting customer reload behavior. It seeks to determine the most suitable algorithm for this task, while also providing insights into the key factors influencing reload behavior. This study thus offers both theoretical contributions to the academic literature and practical implications for telecommunications companies seeking to enhance their customer retention strategies.

2 LITERATURE REVIEW

The data used is customer data of PT. X which operates in the telecommunications sector which has 27 explanatory variables with the objective variable being output reload next month shown in Table 1.

Table 1. Table captions should be placed above the tables.

Column	Description
msisdn	Mobile Phone Number
current_tier	Tier Loyalty Customer
available_points	Points Loyalty Customer that available in their account
vlr_attached_p3d	Flag Attached in Network in Past 3 Days (1/0)
flag_arpu_90d	Classification of Customer based on Average Revenue in Past 90 Days
flag_arpu_last_30d	Classification of Customer based on Average Revenue in Past 30 Days
tenure_rgu	Tenure/Age of Customer in our service
rgu_flag	Flag of Customer that generating revenue in Past 30 Days
rld_30d	Have Reload in Past 30D (0/1)
rld_60d	Have Reload in Past 60D (0/1)
rld_90d	Have Reload in Past 90D (0/1)
rld_tot_30d	Reload Total in Past 0-30D
rld_tot_60d	Reload Total in Past 31-60D
rld_tot_90d	Reload Total in Past 61-90D
reload_p90d	Average Reload in 0-90D
tot_month_rld	Total Reload Month
denom_30d	Maximum Denom of Reload in Past 30D (0/1)

denom_60d	Maximum Denom of Reload in Past 60D (0/1)
denom_90d	Maximum Denom of Reload in Past 90D (0/1)
curr_balance	Balance that Customer's have in their account
active_pack	Have Package in current date (0/1)
Status	Card Status
n_days	Days to Grace Period
arpu_rld	Average Revenue Per User from 3 months
cust_flag	Output Type of Customer
rld_nm	Output Reload Next Month
rld_tot	Output Reload Denom that Customer will buy

3 METHOD

The method used in this study is the boosting algorithm which includes XGBoost, AdaBoost and CatBoost. Figure 1 shows the flow of this research method:

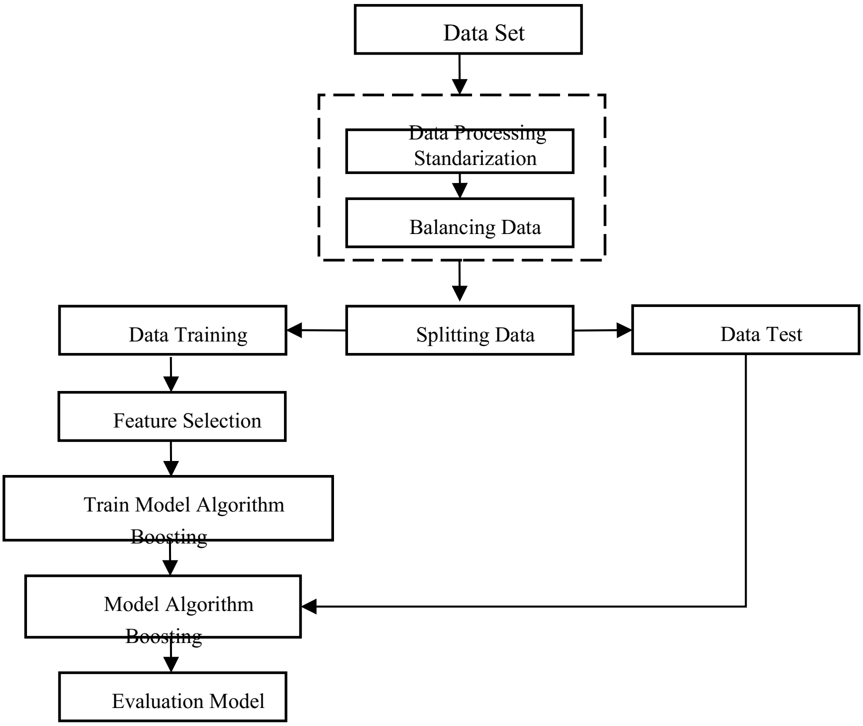

Fig. 1. FlowChart this research

3.1 Feature Selection

The goal of feature selection is to identify the most relevant features or subsets of features for the prediction model. Effective feature selection can help reduce model complexity, shorten training time, and improve prediction accuracy. Additionally, selecting appropriate features enhances the interpretability of the model. Feature selection techniques are categorized based on the availability of label information, falling into three groups: supervised, semi-supervised, and unsupervised methods [12].

3.2 XGBoost Algorithm

XGBoost, an abbreviation for eXtreme Gradient Boosting, was introduced by Chen and Guestrin. [13]. XGBoost is an optimized gradient tree boosting framework that incorporates algorithmic innovations such as approximate greedy search, parallel learning, and adjustable hyperparameters to enhance learning efficiency and control overfitting. XGBoost, as an ensemble learning algorithm, produces rankings of individual features using Gini Importance, similar to the mean decrease in impurity. This ranking can be used to efficiently narrow down the RFE search space [14]. The After fitting an XGBoost model, feature rankings based on weight and gain importance can be derived. The advantages of feature selection include improved interpretability, simplified models, reduced training time, and enhanced generalization, among other benefits [15].

3.3 AdaBoost Algorithm

The AdaBoost algorithm proposed in this study follows these steps [16]:

1. AdaBoost begins by randomly creating and assigning training and test subsets.
2. The model is trained iteratively by selecting a training set in each iteration.
3. Misclassified observations are given higher weights to increase their chances of being correctly classified in the next iteration.
4. After each iteration, the algorithm adjusts the classifier’s weight, using a fine-tuned Decision Tree classifier.
5. This process repeats until either all training data is classified correctly or a predefined maximum number of estimators is reached.
6. For classification, the model performs a "voting" among classifiers to determine the final output.

3.1 CatBoost Algorithm

CatBoost differs from other gradient boosting algorithms in several ways: it offers two boosting modes, plain and ordered; it can directly handle categorical features using ordered target statistics encoding; and it employs a symmetrical tree structure, which is less prone to overfitting [17]. The CatBoost algorithm has demonstrated superiority over other boosting methods in classifying imbalanced multi-class datasets. [18].

3.4 Confusion Matrix

To assess the models, various performance metrics were derived from the confusion matrix. In the context of credit scoring, true positives and true negatives represent the counts of accurately classified default and non-default borrowers, respectively. Conversely, false negatives and false positives indicate the number of non-default and default borrowers incorrectly classified [19]. Accuracy reflects the ratio of correct predictions both positive and negative across the entire dataset and is calculated using Equation (1). Sensitivity, also known as recall, signifies the proportion of accurately identified positive cases out of all actual positive instances and is computed with Equation (2). Specificity assesses the ratio of correctly identified negative cases to all actual negative instances, determined by Equation (3). Precision evaluates the fraction of true positive predictions relative to all positive predictions made by the model and is computed using Equation (4). The F-score (or F1-score) represents the harmonic mean of precision and recall, providing a balanced measure between these two metrics, especially in situations where there is a class imbalance, and is calculated using Equation (5). This evaluation framework aids in understanding the model's accuracy in making predictions and its ability to correctly identify both positive and negative cases [20].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{3}$$

$$Precision = \frac{TP}{Tp + FP} \times 100\% \tag{4}$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{5}$$

4 RESULT AND DISCUSSION

4.1 Exploration Data

Figure 2 shows customer reload behavior across two loyalty tiers, "Healthy" and "Moderate". In both tiers, more customers did not reload compared to those who did. For the Healthy tier, around 300,000 customers did not reload, while 200,000 did. In the Moderate tier, 450,000 customers avoided reloading, with 200,000 reloading. This pattern shows that most customers in these tiers tend to avoid reloading, which may guide the company in developing strategies to encourage more reloads, particularly among higher loyalty customers.

Fig. 2. Customer reload behavior across two loyalty tiers, "Healthy" and "Moderate"

Figure 3 shows customer reload behavior across four types: Gold, Platinum, Red, and Silver. In all groups, most customers did not reload. For Gold customers, over 200,000 did not reload, while just above 100,000 did. Platinum customers had a smaller gap, with around 150,000 not reloading and just over 100,000 reloading. Red and Silver customers also showed a higher count of non-reloaders, following similar trends. Despite slight variations, the pattern is consistent across all types, with the majority avoiding reloads. This information can help the company design targeted strategies to boost reloads, especially in premium segments.

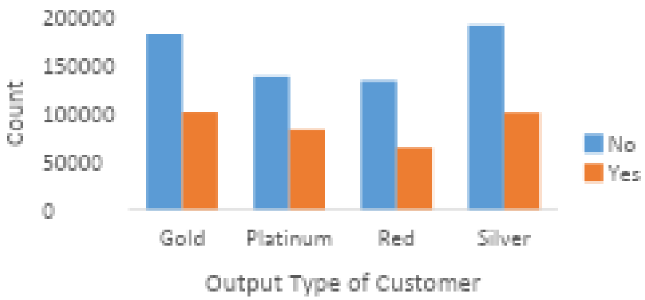

Fig. 3. Customer reload behavior across four types: Gold, Platinum, Red, and Silver.

Figure 4 shows customer reload behavior based on ARPU (Average Revenue Per User) across four segments: nvc (No Value Customers), lvc (Low Value Customers), mvc (Medium Value Customers), and hvc (High Value Customers). In the nvc group, reload activity is minimal, showing low engagement. The lvc segment has over 150,000 customers who did not reload, with fewer than 100,000 who did. The mvc group shows the largest counts, with nearly 400,000 not reloading and 250,000 reloading, indicating significant potential for increased engagement. The hvc segment also follows this trend, with 200,000 not reloading and around 100,000 reloading. Overall, higher ARPU customers show more reload activity, especially in the mvc and hvc groups, but a majority still refrains from reloading. This suggests opportunities for companies to boost reload behavior by focusing on medium and high-value segments.

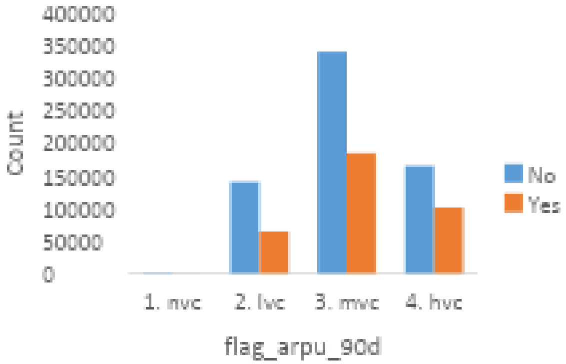

Fig. 4. Customer reload behavior based on ARPU (Average Revenue Per User) across four segments: nvc, lvc, mvc, and hvc.

Figure 5 shows a clear trend of increasing customer retention (represented by the orange bars) with longer subscription duration (indicated by the "tot_month_rld" variable). This suggests that customers who have been subscribed for a longer period are more likely to renew their subscriptions. To further investigate this finding, it would be beneficial to explore additional factors, such as customer demographics, pricing plans, and promotional activities, that may contribute to this observed pattern.

Fig. 5. Trend of increasing customer retention with longer subscription duration

4.2 Correlation Among Variables

Figure 6 shows the correlation between variables used in the customer churn prediction model. Variables like total reload in 30, 60, and 90 days (`rld_tot_30d`, `rld_tot_60d`, `rld_tot_90d`) exhibit strong correlations with each other and with variables such as average revenue (`arpu_rld`) and total reload (`reload_p90d`). This indicates that customer reload behavior is consistent over time, and the more frequently customers reload, the higher their generated revenue.On the other hand, variables like loyalty points (`available_points`) and active package status (`active_pack`) show weaker correlations with reload-related variables, suggesting that their impact on customer reload behavior is relatively minor. Overall, the strong correlations between reload and revenue variables provide valuable insights for identifying customers at risk of churn.

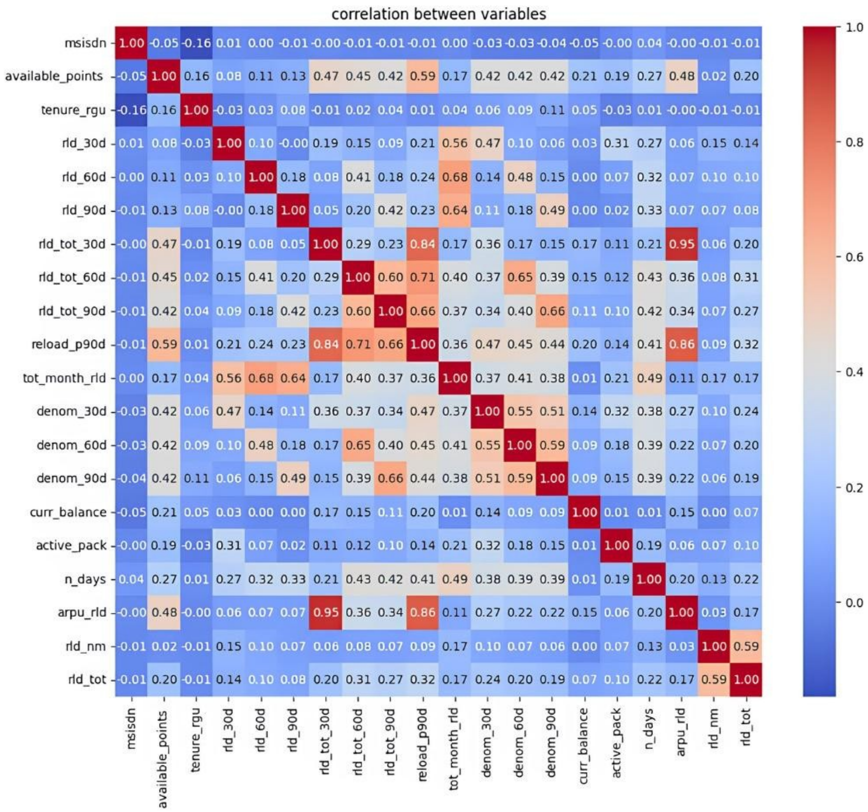

Fig. 6. Correlation between variables used in the customer churn prediction model

4.3 Feature importance from algorithm Boosting

The feature importance analysis using the XGBoost model revealed that the tenure_rgu feature has the highest influence on predicting customer reload behavior, with an importance score of approximately 0.5 as presented in Figure 7. This indicates that the duration of service usage significantly affects reload decisions. The rld_tot_30dfeature, representing total reloads in the past 30 days, also plays a crucial role with an importance score of around 0.3, highlighting the significance of recent reload frequency. Other features like reload_p90d and rld_tot_60d contribute as well, though with lower scores. Features such as n_days, denom_30d, and curr_balance have smaller but relevant impacts. Overall, the model emphasizes the importance of customer tenure and recent reload activity in predicting reload likelihood, underscoring the need to consider both long-term usage and recent interactions for accurate behavior predictions.

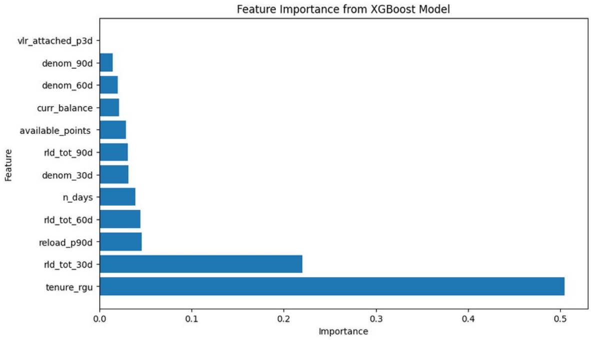

Fig. 7. Feature Importance from XGBoost Model.

The feature importance analysis from the AdaBoost model indicates that rld_tot_30d (total reloads in the last 30 days) is the most significant predictor, with an importance score over 0.25, highlighting the importance of recent reload frequency as presented in Figure 8. Available_points (customer points) and tenure_rgu (customer tenure) also play substantial roles, each scoring around 0.15, reflecting the customer's rewards status and loyalty. Other features like denom_30d (nominal reload amount in the last 30 days), n_days (active days), and curr_balance (current balance) have moderate importance scores near 0.10, providing insights into recent activity and financial status. Features such as rld_tot_90d and reload_p90d contribute less significantly compared to more recent activity, while vlr_attached_p3d (3-day attachment value) is the least influential. Overall, the AdaBoost model emphasizes the importance of recent reload behavior, customer loyalty, and rewards status in predicting future reload decisions

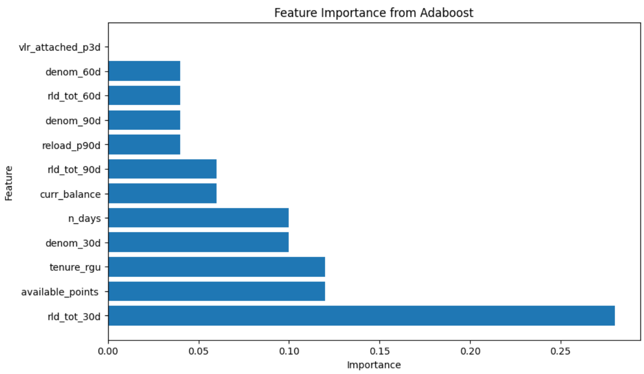

Fig. 8. Feature importance from Adaboost

The feature importance analysis from the CatBoost model reveals that tenure_rgu is the most significant feature, with an importance score close to 70, indicating that the length of customer tenure is the key factor in predicting reload behavior as presented in Figure 9. The model relies heavily on this feature, suggesting that long-term customers exhibit more predictable behavior. Other features, such as available_points and denom_30d (nominal reload in the last 30 days), have much lower importance scores, while n_days (active days) and rld_tot_30d (total reloads in the last 30 days) also contribute less significantly. Features like curr_balance, rld_tot_60d, and vlr_attached_p3d show minimal importance. Overall, the CatBoost model highlights that customer loyalty and tenure are the primary drivers of reload behavior, with recent activity and financial status playing a minor role.

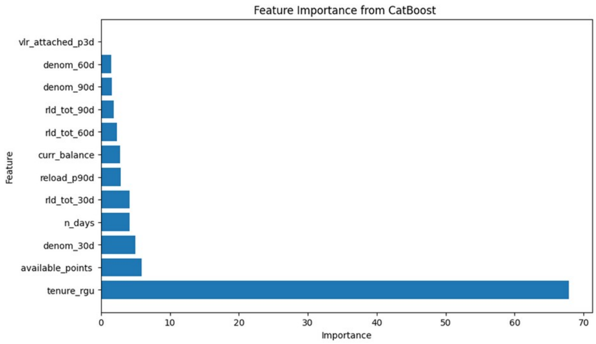

Fig. 9. Feature Importance from CatBoost

4.4 Evaluating Algorithm Boosting

The Table 2 presents a performance comparison of three boosting algorithms: XGBoost, AdaBoost, and CatBoost, evaluated using key metrics such as F-Score, Recall, Precision, Specificity, Sensitivity, and Accuracy. From the results, it is clear that XGBoost stands out as the best-performing algorithm across most metrics. It achieves the highest F-Score of 0.87, indicating a well-balanced performance between precision and recall, and has the strongest precision at 93.44%, showing that its predictions for customer reload are highly accurate. Its recall score of 0.88 suggests that XGBoost captures the majority of customers who actually reload, making it an effective model for identifying positive cases. Moreover, with a sensitivity of 0.93, it demonstrates an excellent ability to correctly detect true reload cases. Overall, XGBoost's accuracy of 90.04% positions it as the top model for customer reload prediction.

Table 2. Table captions should be placed above the tables.

Algorithm Boosting	F-Score	Precision	Specifity	Sensivity	Accuracy

XGBoost	0.87	93.44	0.73	0.93	90.04
AdaBoost	0.68	87.02	0.69	0.87	87.02
CatBoost	0.79	90.08	0.71	0.90	88.09

In contrast, AdaBoost, while still effective, exhibits a lower F-Score of 0.68, indicating less balance between precision and recall. AdaBoost achieves a recall of 0.87, comparable to the other models, but its precision of 87.02% is noticeably lower than that of XGBoost. AdaBoost also struggles slightly more with identifying non-reload customers, as indicated by its lower specificity of 0.69. However, despite these limitations, AdaBoost maintains a solid accuracy of 87.02%, showing that it remains a competitive option, though not as robust as XGBoost.CatBoost, the third algorithm evaluated, performs well across all metrics, though it does not surpass XGBoost. With a precision of 90.08%, CatBoost demonstrates strong accuracy in predicting customer reloads. Its recall is consistent with the other models at 0.87, and its sensitivity is high at 0.90, which means that it is effective at identifying true reload cases. CatBoost's overall accuracy of 88.09% places it as a strong competitor, though slightly behind XGBoost in terms of performance balance and general accuracy.

5 CONCLUSION

Boosting algorithms (XGBoost, AdaBoost, and CatBoost) have proven effective in classifying customer reload behavior in the telecommunications industry. With the ability to present feature importance, these algorithms help identify key factors influencing reload decisions. Performance comparisons indicate XGBoost as the most suitable algorithm, given its superior accuracy, precision, recall, and F1 score. This study emphasizes the importance of applying data mining to design effective prediction systems, assisting telecommunications companies in resource allocation and enhancing customer retention.However, this study has several limitations. The dataset used is restricted to one telecommunications provider, which may not reflect the broader variability in customer behavior. Additionally, the high computational burden of boosting algorithms may limit their use in real-time prediction systems. Future research should aim to expand the data scope, optimize algorithms for computational efficiency, and explore advanced techniques such as deep learning, particularly recurrent neural networks (RNN), to capture complex non-linear relationships. Approaches like rough set theory and uncertainty quantification should also be considered to enhance prediction reliability. This research demonstrates the effectiveness of boosting algorithms in predicting customer reload behavior, but further studies are needed to address these limitations and explore more advanced techniques to improve accuracy and system efficiency, which will benefit telecommunications companies in designing better retention strategies.

References

1. Shukla, V., Prashar, S., Pandiya, B.: Is price a significant predictor of the churn behavior during the global pandemic? A predictive modeling on the telecom industry. J. Revenue Pricing Manag. 21(4), 470–483 (2022). https://doi.org/10.1057/s41272-021-00367-2
2. Puspitasari, A., Masruroh, N.A.: Reducing Customer Churn for XL Axiata Prepaid: Factors and Strategies. J. Ind. Eng. Educ. 2(1), 83–107 (2024).
3. Borah, S.B., Prakhya, S., Sharma, A.: Leveraging service recovery strategies to reduce customer churn in an emerging market. J. Acad. Mark. Sci. 48(5), 848–868 (2020). https://doi.org/10.1007/s11747-019-00634-0
4. Adeniran, I.A., Efunniyi, C.P., Osundare, O.S., Abhulimen, A.O.: Implementing machine learning techniques for customer retention and churn prediction in telecommunications. Comput. Sci. IT Res. J. 5(8), 2011–2025 (2024). https://doi.org/10.51594/csitrj.v5i8.1489
5. Lukita, C., Bakti, L.D., Rusilowati, U., Sutarman, A., Rahardja, U.: Predictive and Analytics using Data Mining and Machine Learning for Customer Churn Prediction. J. Appl. Data Sci. 4(4), 454–465 (2023). https://doi.org/10.47738/jads.v4i4.131
6. AlShourbaji, I., Helian, N., Sun, Y., Hussien, A.G., Abualigah, L., Elnaim, B.: An efficient churn prediction model using gradient boosting machine and metaheuristic optimization. Sci. Rep. 13(1), 1–18 (2023). https://doi.org/10.1038/s41598-023-41093-6
7. Xia, Y., Jiang, S., Meng, L., Ju, X.: XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring. Systems 12(7), 1–26 (2024). https://doi.org/10.3390/systems12070254
8. Wijaya, N., Diqi, M., Mustiadi, I.: AdaBoost Classification for Predicting Residential Habitation Status in Mount Merapi Post-Rehabilitation, Eruption. J. Comput. Sci. Inf. Technol. (CoSciTech) 4(2), 429–436 (2023). https://doi.org/10.37859/coscitech.v4i2.5141
9. Maulana, A., Afidh, R.P.F., Maulydia, N.B., Idroes, G.M., Rahimah, S.: Predicting Obesity Levels with High Accuracy: Insights from a CatBoost Machine Learning Model. Infolitika J. Data Sci. 2(1), 17–27 (2024). https://doi.org/10.60084/ijds.v2i1.195
10. Ogbonna, O.J., Aimufua, G.I.O., Abdullahi, M.U., Abubakar, S.: Churn Prediction in Telecommunication Industry: A Comparative Analysis of Boosting Algorithms. Dutse J. Pure Appl. Sci. 10(1b), 331–349 (2024). https://doi.org/10.4314/dujopas.v10i1b.33
11. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B.: New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. Eur. J. Oper. Res. 218(1), 211–229 (2012). https://doi.org/10.1016/j.ejor.2011.09.031
12. Miao, J., Niu, L.: A Survey on Feature Selection. Procedia Comput. Sci. 91(ITQM), 919–926 (2016). https://doi.org/10.1016/j.procs.2016.07.111
13. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. pp. 785–794 (2016). https://doi.org/10.1145/2939672.2939785
14. Zhang, Y., Liu, J., Shen, W.: A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. Appl. Sci. 12(17), 1–19 (2022). https://doi.org/10.3390/app12178654
15. Nadya, F.E.P., Ferdiansyah, M.F.I., Nastiti, V.R.S., Aditya, C.S.K.: Implementation of Feature Selection Strategies to Enhance Classification Using XGBoost and Decision Tree. Sci. J. Informatics 11(1), 18–194 (2024). https://doi.org/10.15294/sji.v11i1.48145
16. Sevinç, E.: An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. Comput. Ind. Eng. 165(January) (2022). https://doi.org/10.1016/j.cie.2021.107912
17. Aldania, A.N.A., Soleh, A.M., Notodiputro, K.A.: A Comparative Study of CatBoost and Double Random Forest for Multi-class Classification. J. RESTI (Rekayasa Sist. dan Teknol. Informasi) 7(1), 129–137 (2023). https://doi.org/10.29207/resti.v7i1.4766

18. Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., Asadpour, M.: Boosting methods for multi-class imbalanced data classification: an experimental review. J. Big Data 7(1) (2020). https://doi.org/10.1186/s40537-020-00349-y
19. Yotsawat, W., Wattuya, P., Srivihok, A., Info, A.: Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization. Int. J. Electr. Comput. Eng. 11(6), 5477–5487 (2021). https://doi.org/10.11591/ijece.v11i6.pp5477-5487
20. Mokheleli, T., Museba, T.: Machine Learning Approach for Credit Score Predictions. J. Inf. Syst. Informatics 5(2), 497–517 (2023). https://doi.org/10.51519/journalisi.v5i2.487