Assessing Gradient-Boosting Techniques for Customer Churn Prediction: A Statistical Perspective Using SMOTE-ENN and SMOTE-Tomek

Faroug A. Abdalla[1*], Ali Satty[2], Ashraf F. A. Mahmoud[1,4], Gamal Saad Mohamed Khamis[1], Zakariya M. S. Mohammed[2,3], Mohyaldein Salih[2], Abaker A Hassaballa[2,3], Ahmed M. A. Adam[2]

[1] Department of Computer Science, College of Science, Northern Border University, Arar 73213, Saudi Arabia
[2] Department of Mathematics, College of Science, Northern Border University, Arar 73213, Saudi Arabia
[3] Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia
[4] Translation, Authorship, and Publishing Center, Northern Border University, Arar 73213, Saudi Arabia

Corresponding Author Email: Faroug.abdalla@nbu.edu.sa

ABSTRACT

Customer churn prediction is vital for businesses, particularly in telecommunications, where retaining customers is key to revenue. A significant challenge is the imbalance in datasets, where most customers do not churn, leading to biased predictions from machine learning models. This study aims to tackle class imbalance using hybrid resampling techniques, namely SMOTE-Tomek and SMOTE-ENN, to balance the dataset and enhance churn prediction accuracy. Using a publicly available dataset from Kaggle, four gradient-boosting models, Gradient Boosting (GB), CatBoost, XGBoost, and XGBRF, were evaluated on both the original imbalanced dataset and the balanced datasets created by the resampling techniques. Performance was measured using several metrics. Among the evaluated models, GB achieved the highest performance (AUC = 0.992, F1-score = 0.958) under the SMOTE-ENN method, followed closely by CatBoost (AUC = 0.991, F1-score = 0.955). The findings highlight the effectiveness of SMOTE-ENN in improving model performance and offer insights into the best practices for handling imbalanced data in churn prediction tasks.

1. INTRODUCTION

Customer churn prediction has become a critical task for businesses, particularly in industries such as telecommunications, where retaining customers is essential for maintaining profitability [1]. Predicting which customers are likely to leave enables companies to implement targeted retention strategies, optimize marketing efforts, and reduce customer attrition. However, churn prediction often involves dealing with imbalanced datasets, where the number of customers who stay with the service far outweighs those who churn [2]. This imbalance poses a significant challenge in training machine learning models, as they tend to be biased toward predicting the majority class, leading to inaccurate predictions for the minority class (i.e., churned customers). Therefore, employing effective techniques for addressing class imbalance is crucial for improving the performance of churn prediction models [3].

To tackle the challenges of imbalanced datasets, this study employs a hybrid resampling approach that combines oversampling and undersampling techniques, specifically SMOTE-Tomek and SMOTE-ENN. The SMOTE-Tomek technique merges the Synthetic Minority Over-Sampling Technique (SMOTE) with Tomek Links to increase the representation of the minority class while eliminating majority class samples near the decision boundary, thereby enhancing

class separability and reducing noise [4, 5]. SMOTE-ENN, which integrates SMOTE with Edited Nearest Neighbor (ENN), further reduces misclassified instances, enhancing predictive accuracy [6]. Both techniques have improved classification performance, particularly in noisy or borderline instances, by balancing class distribution and minimizing overfitting [7, 8].

This study evaluates the performance of Gradient Boosting (GB), CatBoost, XGBoost, and XGBRF on imbalanced and balanced datasets. These gradient-boosting models effectively handle complex datasets, particularly for customer churn prediction [9]. XGBoost, in particular, has effectively identified at-risk customers in industries like telecommunications by managing missing values and large datasets through scalability and regularization [10]. CatBoost, on the other hand, excels in handling categorical data, improving churn prediction accuracy without extensive preprocessing, and has been demonstrated to outperform other models in capturing complex customer behavior patterns in telecom [11, 12]. Additionally, XGBRF, which combines Gradient Boosting and Random Forests, is known to mitigate overfitting and reduce variance in churn prediction tasks [13], enhancing model robustness and stability.

This research emphasizes assessing the performance of these gradient-boosting models in customer churn prediction, focusing on overcoming the challenges associated with

imbalanced datasets. The primary objective is to identify the most effective combination of models and resampling techniques for generating accurate and reliable churn predictions. The significance of this research lies in its potential to advance churn prediction methodologies by analyzing the impact of resampling techniques on model performance. The contribution of this paper is twofold: it offers valuable insights into the strengths and limitations of popular gradient-boosting models. It underscores the importance of advanced resampling techniques in improving prediction outcomes. Ultimately, this study seeks to enhance churn prediction accuracy, helping telecommunications companies develop more effective customer retention strategies and optimize customer relationship management (CRM). Moreover, the findings could serve as a framework for other industries dealing with similar challenges of class imbalance, enabling the development of robust predictive models across various domains.

2. METHODOLOGY

The process shown in Figure 1 starts with extracting a customer churn dataset from the telecommunications industry. This dataset then undergoes pre-processing, which involves cleaning the data and performing churn prediction with the original unbalanced dataset. Subsequently, customer churn prediction is conducted on the unbalanced dataset using various machine learning classifiers. Next, the dataset is balanced using hybrid methods that combine under-sampling and oversampling techniques. The classifiers are then applied to these balanced datasets. Finally, the classifiers' performance is evaluated using several metrics, enabling a comparison between the results obtained from the unbalanced and balanced datasets.

- **Data source**

This study utilizes a publicly available dataset from Kaggle's Call Detail Records file, comprising 7,043 customer records from a telecommunications firm. Each record contains 21 attributes that help forecast the likelihood of customer churn. Out of the total sample, 1,869 customers experienced churn, while 5,174 stayed with the company. Summarizes the characteristics of this dataset. Further details can be accessed at the study [14]. A summary of the study features is provided below.
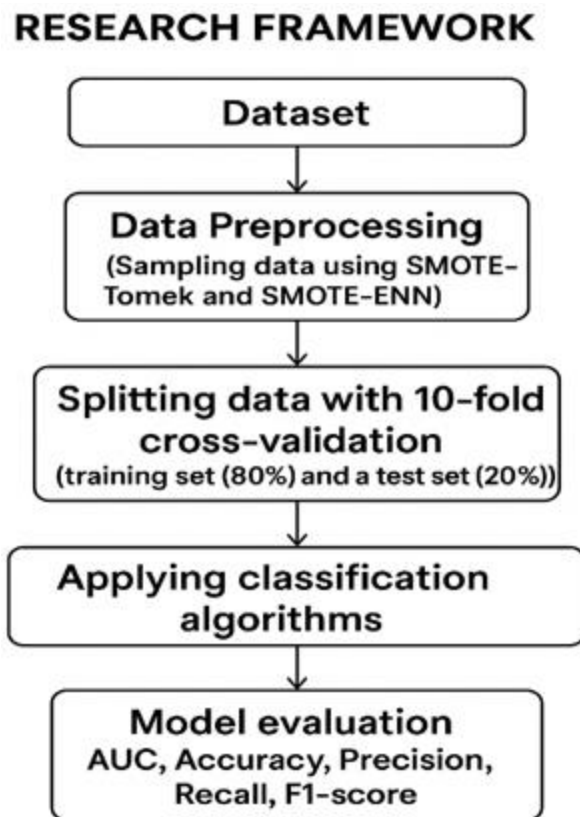
Figure 1. Workflow for data preprocessing, balancing, and model evaluation in predictive techniques

Target variable:
Churn (Yes/No) – Customer has stopped service or not.
Independent variables:
- CustomerID (String) – Unique identifier for the customer.
- Gender (Female/Male) – Customer gender.
- SeniorCitizen (0/1) – Indicates senior citizen status.
- Partner (Yes/No) – Customer has a partner.
- Dependents (Yes/No) – Customer has dependents.

- Tenure (Numeric) – Length of customer association in months.
- PhoneService (Yes/No) – Subscription status for phone service.
- MultipleLines (Yes/No) – Subscription status for multiple lines.
- InternetService (DSL/Fiber Optic/No) – Type of internet service provider.
- OnlineSecurity (Yes/No) – Online security service subscription status.

- OnlineBackup (Yes/No) – Customer subscribes to online data backup.
- DeviceProtection (Yes/No) – Subscription to device protection.
- TechSupport (Yes/No) – Availability of technical support subscription.
- StreamingTV (Yes/No) – TV streaming service subscription.
- StreamingMovies (Yes/No) – Movie streaming subscription.
- Contract (Month-to-month, one year, two years) – Contract length of customer subscription.
- Paperless Billing (Yes/No) – Preference for paperless billing.
- PaymentMethod (Bank transfer, Credit card, Electronic check, Mailed check) – Customer payment method preference.
- MonthlyCharges (Numeric) – Monthly billed charges.
- TotalCharges (Numeric) – Overall billed charges.

- **Data pre-processing**

A crucial part of machine learning is ensuring the dataset is clean, well-structured, and ready for accurate modeling. This study addressed the missing data problem by implementing multiple imputations, a robust technique that estimates and fills in missing values while maintaining the dataset's overall integrity. All string and categorical variables were converted to nominal data types. This key modification enhances the precision of machine learning classifiers by ensuring these variables are appropriately handled during the training phase. The dataset posed a significant challenge with class imbalance: 73.5% of the records were in the "customers not churned" category, while only 26.5% were in the "customers churned" category. This imbalance risks classifiers favoring the majority class, potentially misclassifying new data as "customers not churned." To address this, oversampling techniques were employed to balance the dataset and enhance model performance. Among the various resampling methods available (e.g., Random Under-Sampling, ADASYN), this study focused on SMOTE-Tomek and SMOTE-ENN, which combine oversampling with data-cleaning mechanisms. Prior research [7, 8] indicates that these hybrid approaches outperform simpler resampling techniques, particularly by reducing boundary noise and mitigating overfitting in imbalanced telecom datasets.

2.1 The SMOTE-Tomek technique

The SMOTE-Tomek method combines the Synthetic Minority Over-Sampling Technique (SMOTE) and Tomek Links to tackle class imbalance in datasets. This technique, proposed by Batista et al. [7], uses SMOTE to increase the number of minority class instances. It applies Tomek Links to refine the dataset by removing the majority class examples closest to the minority class. This dual approach boosts the minority class and enhances the class distinction, improving model performance.

2.2 The SMOTE-ENN technique

The SMOTE-ENN method, proposed by Batista et al. [7], addresses class imbalance by integrating the SMOTE algorithm with the Edited Nearest Neighbor (ENN) technique.

This fusion results in an improved sampling strategy that enhances the dataset's balance. The ENN technique filters out noisy and borderline samples by discarding points that do not align with the majority class in their K-nearest neighbors.

- **Cross-validation and model selection for accurate prediction**

The dataset was split into two random subsets: 80% for training and 20% for testing. The training subset was used to build and refine the predictive model, while the test subset was set aside to evaluate its performance on unseen data. A 10-fold cross-validation technique was applied during the training phase to enhance the model's reliability and stability. This process involved dividing the training data into 10 segments, one for validation, and the remaining for training. The procedure was repeated with different configurations to examine how variations in the training and testing splits affected the model's predictive power and ability to generalize. The final model was chosen based on its performance metrics, ensuring it delivered the highest possible accuracy on the test data. Four machine learning techniques were selected for constructing the predictive model. Each technique underwent thorough training on the training data and was subsequently evaluated using the test data to assess its performance.

- **Model parameter settings and tuning strategy**

To ensure fair and reproducible model comparison, all gradient-boosting algorithms were optimized using systematic hyperparameter tuning. The tuning process employed a grid search combined with 10-fold cross-validation on the training data, with the F1-score used as the primary performance metric for model selection. For the GB model, the optimal parameters were *learning_rate* = 0.1, *n_estimators* = 200, and *max_depth* = 3. The **CatBoost** model achieved best results with *iterations* = 500, *learning_rate* = 0.05, *depth* = 6, and *l2_leaf_reg* = 3. The XGBoost model performed optimally with *learning_rate* = 0.1, *n_estimators* = 300, *max_depth* = 4, and *subsample* = 0.8. Finally, the XGBRF model used *n_estimators* = 200, *max_depth* = 4, and *colsample_bytree* = 0.8. These parameter values were selected to balance model accuracy, generalization, and computational efficiency, ensuring that each algorithm was evaluated under its best-performing configuration.

- **Gradient-boosting techniques for customer churn prediction**

This study utilized four machine learning algorithms grounded in gradient-boosting techniques to develop the customer churn prediction model. These algorithms were carefully chosen based on several key considerations, including the nature of the churn prediction task, the dataset's unique characteristics, and the overarching goals. The gradient-boosting models selected, as described in Satty et al. [15], are well regarded for their efficiency and ability to manage complex datasets. Each algorithm brings distinct advantages that make it particularly well-suited for the complexities of churn prediction. These models detect intricate patterns within the data, addressing the everyday challenges of dataset imbalance and heterogeneity. Their capacity to improve prediction accuracy while reducing the risk of overfitting makes them highly effective for generating meaningful insights and ensuring robust predictive performance in this crucial research domain.

1) Gradient boosting (GB)

An ensemble learning algorithm creates models in sequence to minimize prediction errors. It utilizes weak learners, usually decision trees, and continuously refines them by adjusting a loss function [9]. Each new tree is designed to correct the errors of the previous one, leading to a more robust and accurate model. GB is particularly well-suited for structured datasets, where it can detect intricate patterns and complex interactions. Its ability to identify subtle, non-linear relationships between customer features and churn probability makes it a highly effective tool for churn prediction.

2) CatBoost

This Gradient Boosting library, developed by Yandex, was designed to address the challenges of working with categorical features in machine learning. It is particularly effective in outperforming other methods by simplifying the handling of categorical data without requiring extensive preprocessing [16]. With its distinctive algorithmic improvements, CatBoost reduces overfitting and enhances prediction accuracy, making it a highly competitive tool in various industries [17].

3) Extreme Gradient Boosting (XGBoost)

It is an optimized and scalable version of the Gradient Boosting framework, known for its efficiency and flexibility, making it a popular choice among data scientists. Incorporating advanced regularization techniques effectively prevents overfitting, making it versatile and reliable for various tasks, from time series forecasting to classification [18]. Its remarkable speed and accuracy set it apart, often outperforming traditional gradient-boosting methods.

4) Extreme Gradient Boosting Random Forest (XGBRF)

It is an enhanced version of XGBoost that integrates Random Forest principles with Gradient Boosting techniques, combining the strengths of both methods. While XGBoost focuses on sequentially building models to correct previous errors, XGBRF introduces an element of randomness to improve model diversity and overall performance [19]. This fusion leads to better accuracy, particularly in ensemble learning applications, making XGBRF a powerful tool for complex predictive tasks.

- **Comprehensive performance metrics for customer churn prediction models**

The evaluation of gradient-boosting techniques for predicting customer churn involved using a range of performance metrics to identify the most effective model. These included: 1) Area Under the Curve (AUC), which indicated each model's ability to distinguish between different churn outcomes, with higher values reflecting superior predictive power; 2) Classification Accuracy, which provided an initial assessment by showing the percentage of correct predictions made by the model; and 3) F1 Score, which was particularly valuable for balancing precision and recall, addressing the imbalanced nature of customer churn. Within the F1 Score, a) Precision assessed the accuracy of optimistic predictions, while b) Recall measured the model's sensitivity in identifying all instances of churn. Combining these metrics achieved a comprehensive evaluation of each model's performance, ensuring the selection of the most effective model for predicting customer churn.

3. RESULT

Table 1 shows the performance metrics for four gradient-boosting models used in customer churn prediction. CatBoost outperforms the other models with the highest AUC (0.844), Accuracy (0.801), and F1 Score (0.792), indicating its strong predictive power and good balance between precision and recall. GB follows closely with an AUC of 0.841, an Accuracy of 0.793, and an F1 Score of 0.786, showing solid performance but slightly behind CatBoost. XGBoost demonstrates a strong Recall (0.784) and a good F1 Score (0.778), but its AUC (0.823) and Accuracy (0.784) are slightly lower than those of CatBoost and GB. XGBRF, on the other hand, shows the weakest performance with the lowest Accuracy (0.735), Precision (0.540), and F1 Score (0.622), suggesting that it struggles with over-prediction of positive churn instances and lacks overall prediction reliability.

Table 1. Model performance metrics for customer churn prediction based on original data

Model	AUC	Accuracy	F1 Score	Precision	Recall
CatBoost	0.844	0.801	0.792	0.790	0.801
GB	0.841	0.793	0.786	0.783	0.793
XGBRF	0.827	0.735	0.622	0.540	0.735
XGBoost	0.823	0.784	0.778	0.775	0.784

Table 2. Model performance metrics for customer churn prediction using SMOTE-Tomek

Model	AUC	Accuracy	F1 Score	Precision	Recall
GB	0.953	0.876	0.876	0.876	0.876
CatBoost	0.953	0.875	0.875	0.876	0.875
XGBoost	0.949	0.871	0.871	0.872	0.871
XGBRF	0.906	0.823	0.823	0.823	0.823

Table 3. Model performance metrics for customer churn prediction using SMOTE ENN

Model	AUC	Accuracy	F1 Score	Precision	Recall
XGBoost	0.992	0.957	0.957	0.957	0.957
GB	0.992	0.958	0.958	0.958	0.958
CatBoost	0.991	0.955	0.955	0.955	0.955
XGBRF	0.972	0.933	0.933	0.934	0.933

Table 2 compares the gradient-boosting models used in this study with the SMOTE-Tomek technique based on their performance metrics. GB and CatBoost achieve the highest AUC of 0.953, demonstrating strong discriminative power in distinguishing between churn and non-churn instances. These models also have identical scores across Accuracy (0.876), F1 Score (0.876), Precision (0.876), and Recall (0.875), indicating they perform equally well in balancing predictions and handling both false positives and false negatives. XGBoost follows with a slightly lower AUC of 0.949, while its Accuracy (0.871), F1 Score (0.871), Precision (0.872), and Recall (0.871) are still strong but not as high as GB and CatBoost. XXGBRF, on the other hand, performs the weakest, with the lowest AUC of 0.906 and lower scores across all metrics (Accuracy, F1 Score, Precision, and Recall of 0.823), indicating that it is less effective in predicting customer churn compared to the other models.

Table 3 presents performance metrics for the gradient-boosting models evaluated using SMOTE ENN metrics. XGBoost and GB achieve the highest AUC of 0.992, indicating exceptional discriminative power and ability to distinguish between churn and non-churn instances. These models also show very similar performance across all metrics, with Accuracy, F1 Score, Precision, and recall all around 0.957 for XGBoost and 0.958 for GB, reflecting a well-balanced performance in predicting customer churn without significant bias toward false positives or false negatives. CatBoost follows closely with an AUC of 0.991 and similar Accuracy (0.955), F1 Score (0.955), Precision (0.955), and Recall (0.955), performing slightly lower than XGBoost and GB but still demonstrating strong overall effectiveness. XGBRF has a notably lower AUC of 0.972, with Accuracy, F1 Score, Precision, and recall values around 0.933 to 0.934, suggesting that it performs somewhat weaker than the other models in terms of both prediction accuracy and its ability to balance false positives and negatives.

The study findings reveal that GB and CatBoost are the top-performing gradient-boosting models for customer churn prediction across all preprocessing techniques. However, GB is the best model overall, especially when using the SMOTE ENN technique. CatBoost follows closely, with slightly lower metrics but still delivering strong performance. When using SMOTE-Tomek, both GB and CatBoost show equal performance, indicating a very balanced model performance. Regarding handling imbalanced data, SMOTE ENN outperforms SMOTE-Tomek, as it yields higher AUC and overall more consistent performance across the models, particularly for GB and XGBoost. XGBoost is a strong contender but consistently ranks second to GB and CatBoost in both preprocessing techniques. XGBRF is the weakest performer across all models and methods. Therefore, GB is the best gradient-boosting model, and SMOTE ENN is the most effective technique for handling imbalanced data.

To confirm that the observed differences among the gradient-boosting models were statistically significant and not due to random variation, a one-way repeated-measures ANOVA was conducted across all five evaluation metrics (AUC, Accuracy, F1-score, Precision, and Recall). When the overall ANOVA indicated significant effects, pairwise *t*-tests with Bonferroni correction were applied for post-hoc comparisons. The results revealed that GB and CatBoost significantly outperformed XGBRF ($p < 0.01$) and XGBoost ($p < 0.05$) in terms of both AUC and F1-score. These findings demonstrate that the superior performance of GB and CatBoost is statistically robust, reinforcing their reliability as the most effective models for customer churn prediction.

4. DISCUSSION

The results of this study highlight the importance of selecting the appropriate machine learning model and effectively handling class imbalance in customer churn prediction. Our findings demonstrate that gradient-boosting models, particularly Gradient Boosting (GB) and CatBoost, significantly outperform other models, such as XGBoost and XGBRF, with GB being the best performer overall, especially when using the SMOTE-ENN resampling technique.

- **Performance of CatBoost and GB**

In this study, CatBoost outperformed other models in the original dataset. These findings align with those reported by Hasan et al. [12], where CatBoost demonstrated superior performance in customer churn prediction tasks due to its efficient handling of categorical variables and explainability with SHAP values. The model's ability to handle categorical features without extensive preprocessing makes it attractive for real-world applications like customer churn prediction. Similarly, Gradient Boosting (GB) closely followed CatBoost in performance, demonstrating high AUC, Accuracy, and F1 Scores. GB's strength is well-documented in the literature [20], mainly due to its ability to combine the strengths of weak learners (decision trees) to minimize overfitting and provide high predictive power. These findings support the continued use of GB for churn prediction tasks where model performance and interpretability are both important.

- **Role of preprocessing techniques: SMOTE-Tomek vs. SMOTE-ENN**

The preprocessing techniques, SMOTE-Tomek and SMOTE-ENN, played a pivotal role in enhancing model performance, mainly when dealing with imbalanced datasets. Table 3 shows that both CatBoost and GB significantly improved when SMOTE-Tomek was applied, with these models achieving the highest AUC of 0.953 and balanced scores across Accuracy, F1, Precision, and Recall. The application of SMOTE-Tomek, which combines SMOTE oversampling with Tomek links to clean the decision boundary, has effectively reduced the noise and imbalance in churn datasets, leading to more robust models [4].

However, the SMOTE-ENN technique, which combines SMOTE with Edited Nearest Neighbors (ENN), demonstrated even more significant improvements in model performance, particularly for GB and XGBoost, as shown in Table 3. This combination further enhanced the discriminative power of the models, with both GB and XGBoost achieving a perfect AUC of 0.992. The importance of SMOTE-ENN in improving the performance of machine learning models for imbalanced classification problems has been widely recognized in recent research [7], particularly in scenarios like customer churn prediction, where accurately identifying both positive and negative churn instances is critical.

- **XGBoost and XGBRF performance**

While XGBoost also performed reasonably well, its results were consistently slightly lower than those of CatBoost and GB, particularly in the original dataset and with SMOTE-Tomek. XGBoost's performance is often attributed to its

effective regularization techniques and parallelization capabilities, which reduce overfitting while speeding up training [10]. However, as observed in the results, XGBoost's performance can still be sensitive to hyperparameter tuning, especially compared to CatBoost, which often requires less tuning.

On the other hand, XGBRF (XGBoost with Random Forest) underperformed in all evaluations. XGBRF's lower performance could be attributed to the randomization of its tree-building process, which may reduce its ability to learn the underlying patterns in churn data effectively. This aligns with recent research indicating that ensemble methods such as XGBRF can be effective in specific applications and may not always perform as well as pure gradient-boosting methods like CatBoost or GB in tasks requiring fine-grained model calibration [21].

5. CONCLUSION

In conclusion, this study highlighted the importance of addressing class imbalance in customer churn prediction, particularly in industries like telecommunications, where customer retention directly impacts revenue. By leveraging hybrid resampling techniques, specifically SMOTE-Tomek and SMOTE-ENN, the study effectively mitigated the challenge of imbalanced datasets, improving predictive accuracy. These findings significantly impact the application of machine learning models in customer churn prediction. The results suggest that CatBoost and GB are the most reliable choices for such tasks, particularly when coupled with SMOTE-ENN for handling class imbalance. The strong performance of GB highlights its versatility and robustness in different preprocessing environments, making it a strong candidate for real-time applications in industries such as telecommunications, banking, and retail. The results further suggested that utilizing these advanced resampling methods in combination with powerful gradient-boosting models offered a promising approach to better handle class imbalance and optimize churn prediction tasks.

Despite its strengths, this study has several limitations. Although gradient-boosting models demonstrated high predictive accuracy, they are computationally intensive and may be sensitive to noisy or incomplete customer records. Their interpretability also remains limited compared to simpler models, which can hinder practical insights and decision-making. Future work will focus on addressing these limitations by enhancing model explainability through SHAP values, applying cost-sensitive optimization, and testing robustness under varying data-quality conditions. Further research could also explore advanced hyperparameter tuning methods, such as Bayesian optimization or genetic algorithms, and experiment with ensemble strategies like stacking or blending to further improve predictive performance in churn prediction tasks.

REFERENCES

[1] Burez, J., Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36: 4626-4636. https://doi.org/10.1016/j.eswa.2008.05.027

[2] Salunkhe, U.R., Mali, S.N. (2018). A hybrid approach for class imbalance problem in customer churn prediction: A novel extension to under-sampling. International Journal of Intelligent Systems and Applications, 10(5): 71-81. https://doi.org/10.5815/ijisa.2018.05.08

[3] Wagh, S.K., Andhale, A.A., Wagh, K.S., Pansare, J.R., Ambadekar, S.P., Gawande, S.H. (2024). Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization, 14: 100342. https://doi.org/10.1016/j.rico.2023.100342

[4] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321-357. https://doi.org/10.1613/jair.953

[5] Tomek, I. (1976). An experiment with the edited nearest-neighbor rule for deriving the nearest-neighbor decision rule. IEEE Transactions on Systems, Man, and Cybernetics, 6(6): 448-452. https://doi.org/10.1109/TSMC.1976.4309523

[6] Wilson, D.R. (1972). Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, 2(3): 408-421. https://doi.org/10.1109/TSMC.1972.4309137

[7] Batista, G.E.A.P.A., Prati, R.C., Monard, M.C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6(1): 20-29. https://doi.org/10.1145/1007730.1007735

[8] He, H., Garcia, E.A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9): 1263-1284. https://doi.org/10.1109/TKDE.2008.239

[9] Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5): 1189-1232. https://doi.org/10.1214/aos/1013203451

[10] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Jeju Island, Republic of Korea, pp. 785-794. https://doi.org/10.1145/2939672.2939785

[11] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1706.09516

[12] Hasan, B., Shaikh, S.A., Khaliq, A., Nadeem, G. (2024). Data-driven decision-making: Accurate customer churn prediction with Cat-Boost. The Asian Bulletin of Big Data Management, 4(2): 239-250. https://doi.org/10.62019/abbdm.v4i02.175

[13] Ke, G.L., Meng, Q., Finley, T., Wang, T.F., Chen, W., Ma, W.D., Ye, Q.W., Liu, T.Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

[14] Telco Customer Churn Dataset. (2024). Telco customer churn dataset. Kaggle. https://www.kaggle.com/blastchar/telco-customer-churn.

[15] Satty, A., Khamis, G.S.M., Mohammed, Z.M., Mahmoud, A.F., Abdalla, F.A., Salih, M., Gumma, E.A. (2025). Statistical insights into machine learning models for predicting under-five mortality: An analysis from multiple indicator cluster survey (MICS). IEEE Access, 13: 45312-45320. https://doi.org/10.1109/ACCESS.2025.3549097

[16] Hancock, J.T., Khoshgoftaar, T.M. (2020). CatBoost for big data: An interdisciplinary review. Journal of Big Data, 7(1): 94. https://doi.org/10.1186/s40537-020-00369-8

[17] Bentéjac, C., Csörgo, A., Martínez-Muñoz, G. (2019). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54: 1937-1967. https://doi.org/10.1007/s10462-020-09896-5

[18] Ali, Z.A., Abduljabbar, Z.H., Tahir, H.A., Sallow, A.B., Almufti, S.M. (2023). eXtreme gradient boosting algorithm with machine learning: A review. Academic Journal of Nawroz University, 12(2): 320-334. https://doi.org/10.25007/ajnu.v12n2a1612

[19] Chen, T., Contributors. (2024). XGBoost random forest mode. XGBoost Documentation. https://xgboost.readthedocs.io/en/stable/tutorials/rf.html.

[20] Li, Y., Yan, K. (2025). Predicting bank credit customers churn based on machine learning and interpretability analysis. Data Science in Finance and Economics, 5(1): 19-34. https://doi.org/10.3934/DSFE.2025002

[21] Imani, M., Beikmohammadi, A., Arabnia, H.R. (2025). Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. Technologies, 13(3): 88. https://doi.org/10.3390/technologies13030088