Unveiling the power of hybrid balancing techniques and ensemble stacked and blended classifiers for enhanced churn prediction

Kunal Gaikwad[1], Nabeela Berardinelli[1], and Nadeem Qazi[1]

University of East London,
University Way London,
E16 2RD United Kingdom.
gaikwadkunal00777@gmail.com
n.berardinelli@uel.ac.uk
n.qazi@uel.ac.uk

Abstract. For businesses, customer retention is crucial as it is more cost-effective than acquiring new customers. Identifying potential customer churn early allows for the development of effective retention strategies. With advancements in technology and data storage, machine learning has become a popular approach for predicting customer churn. To counteract data imbalance, researchers have utilized minority oversampling methods, particularly the Synthetic Minority Over-sampling Technique (SMOTE). Innovations in this area include hybrid techniques like SMOTE Tomek-Links and SMOTE ENN, which have shown effectiveness in data resampling. Traditional classifiers like Logistic Regression, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbors have been surpassed in performance by ensemble classifiers such as XgBoost, LightGBM, and CatBoost. Yet, there is limited research on the combination of SMOTE hybrid techniques with these advanced ensemble classifiers for churn prediction. This study aims to contribute to the field by integrating hybrid balancing techniques with ensemble classifiers and introducing new stacked and blended models. The findings reveal that a stacked model incorporating SMOTE ENN achieved impressive results: 96.46% accuracy, 97% F1 score, and 97.40% PR-AUC. This was closely followed by the CatBoost-SMOTE ENN model, which scored 95.32% in accuracy, 96% F1 score, and 96.50% PR-AUC. In contrast, ADASYN and standard SMOTE techniques did not significantly affect model performance.

Keywords: Churn Prediction, Data balancing, Ensemble stacking, Blending techniques, Hybrid SMOTE, SMOTE ENN, ADASYN, XgBoost, CatBoost, LightGBM, and Machine Learning,

1 Introduction

Customer churn, the phenomenon where a customer stops using a service or product, is a pivotal concern for businesses. Customer Churn Prediction (CCP) is the process of identifying customers at significant risk of churn based on their service usage data [5]. In the context of heightened competition due to market liberalization, CCP has gained importance. [18] highlighted that a 5% reduction in customer churn can lead to a profit increase of 25 to 85%. This is underscored by [26], who noted that acquiring new customers can be 5 to 6 times more expensive than retaining existing ones, underscoring the value of CCP for companies.

The telecommunications sector, in particular, has witnessed high churn rates, ranging from 10% to 60% in recent years [20]. The use of artificial intelligence in marketing and the development of predictive models has shown promising results [23]. [11] argue that focusing on a specific segment of customers likely to churn, rather than the entire customer base, is more efficient for retention strategies.

Data imbalance is a significant challenge in CCP, leading to suboptimal model performance due to inadequate representation of class differences [22]. Balancing techniques include data level, internal algorithmic, cost-sensitive approaches, and ensemble learning classifiers [2].

Various studies have explored churn prediction with mixed outcomes. [6] used Weighted Random Forest (WRF) but achieved limited success initially; improvements were noted when combined with sampling techniques, but the study did not extend to other machine learning models. Ensemble learning techniques like Random Forest have outperformed other classifiers such as Decision Trees, Naïve Bayes, and SVM in terms of sensitivity and accuracy, though these studies often did not address data imbalance [16]. Contrarily, [25] found that decision trees improved with oversampling and undersampling, while bagging and boosting techniques did not significantly enhance model performance.

This paper aims to develop a more effective churn prediction model, focusing on addressing data skewness, comparing the impact of encoding techniques on model performance, and employing ensemble techniques alongside stacking and blending approaches. The study will also explore optimal model architecture through hyperparameter tuning, assessed using various performance metrics. This comparative analysis will elucidate the effects of different balancing techniques on model performance, revealing the most effective combination of balancing technique and classifier.

2 Related work

In recent years, advancements in technology have led to the emergence of new data storage and analysis methods. For predicting customer churn in telecom companies, many approaches primarily utilize machine learning and data mining techniques [1]. Customer churn prediction is often treated as a supervised learning problem with binary classification. Researchers and data analysts are now concentrating on using customer data to develop models that predict churn events and enhance their predictive accuracy [20].

A major hurdle in achieving optimal performance in customer churn prediction is data imbalance [16]. Leevy et al. [10] highlighted that skewness in datasets can introduce bias into prediction models, leading to misclassifications. Kimura [8], referencing Lopez [14], discussed how misclassifications occur in imbalanced datasets, often because the majority class is correctly classified while many instances of the minority class are misclassified. This can happen either because the minority class is overlooked due to its smaller size or it is mistaken as noise. Addressing this skewness has thus become a focal point for researchers.

Ensemble techniques have been increasingly employed in recent years for constructing churn prediction models, thanks to their effective performance. These techniques, which involve combining outputs from multiple models, often use bagging and boosting methods [19]. Random Forest, a widely used bagging method, and boosting variations like Xgboost, LightGBM, and CatBoost, have gained prominence.

Research shows that Random Forest, as a bagging technique, achieves high accuracy and sensitivity, outperforming other classifiers such as Decision Tree, Naïve Bayes, and Support Vector Machine [16]; [21]. Ensemble methods, including bagging with neural networks and boosting approaches like XgBoost, have proven effective in churn prediction, with XgBoost exhibiting notable performance on datasets like IBM, despite its lower sensitivity compared to some algorithms [17].

LightGBM, another gradient boosting model, is recognized for its rapid training process and effectiveness with large datasets, surpassing XgBoost and Stochastic Gradient Boosting in speed and memory efficiency [15]; [13]. CatBoost, in contrast, has demonstrated higher accuracy and faster performance than other gradient boosting algorithms, particularly in balanced datasets [7].

From the literature review, the studies conducted have the following limitations:

- Very few researchers have implemented stacking and blending methods for customer churn prediction.
- No proper comparative study is being done for different balancing techniques.
- No two feature selection methods' outcomes are merged and the factors causing churn are not analyzed.

Based on the literature review, ensemble boosting algorithms like XGBoost, CatBoost, and LightGBM have shown superior performance compared to traditional classification algorithms. Newly emerged variations of ensemble methods, such as stacking and blending, have demonstrated their effectiveness in improving the performance of individually trained ensemble models. However, no studies have yet implemented these stacking and blending methods and compared them with the aforementioned boosting algorithms in the context of customer churn prediction using the IBM dataset. Unlike stacking, which utilizes an out-of-fold dataset to train the level 2 classifier, the blending method employs predictions made on a hold-out dataset, thereby preventing data leakage during training. Consequently, the impact of preventing data leakage will be assessed.

Churn prediction datasets are highly imbalanced, leading to a bias towards the majority class. Therefore, it is crucial to address this imbalance using an appropriate balancing method. Among the various data balancing methods reviewed, the data-level approach is widely adopted due to its simplicity and computational efficiency. A comparative study of models using ensemble techniques and their variations combined with the SMOTE balancing technique and its variants, namely ADASYN, SMOTE-Tomek links, and SMOTE-ENN, will be conducted with proper feature selection. This study has not been previously conducted.

3 Proposed method

Considering the importance of identifying customer churn this study will build a binary classification machine learning model to predict churners. Firstly, data will be cleaned to make it suitable for training purposes. Exploratory Data Analysis (EDA) will be carried out to understand and identify relevant features along with their impacts on churners. Owing to the detrimental effects of class imbalance on the model's performance, this study uses a data-level approach for resampling the minority class. Balancing techniques like SMOTE and its hybrid variations are used to calculate their impact on the model performance by comparative study. The novelty of the study is to build classification models using ensemble techniques and their variations, namely blending and stacking by combining them with resampling and data encoding techniques. The models built will be evaluated based on Accuracy, F1 Score and PR-AUC metrics.

The dataset is an open-source customer churn dataset from the telecommunications industry, provided by IBM. It's available on both the IBM community platform and Kaggle. It encompasses the records of 7,043 customers in California, detailing their engagement with the company's services, including whether they have subscribed, stayed, or left. It includes 21 features with information about the customers and their utilized services. The final column serves as a label, distinguishing churners (labeled 'Yes') from non-churners (labeled 'No').

This dataset is particularly suitable for addressing the issue of class imbalance in churn prediction, a topic that has not been extensively explored with stacking and blending techniques in model building. It contains 18 features of object data type, two integer, and one float. The majority of the features are categorical and require encoding. All columns, except 'Total Charges', are free from null values. Of the 7,043 records, 1,869 are churners and 5,174 non-churners, making up approximately 26.53% and 73.47% of the dataset, respectively, indicating a significant imbalance.

SMOTE randomly oversampled minority classes from a count of 1863 to 4880. SMOTE-Tomek Link is a hybrid technique which is the combination of SMOTE which generates synthetic records of minority class and Tomek-link's ability to identify and eliminate the data that has Tomek links with the majority class.

The difference between the functioning of SMOTE Tomek link and SMOTE ENN is that the latter handles both the majority and minority classes. It balances

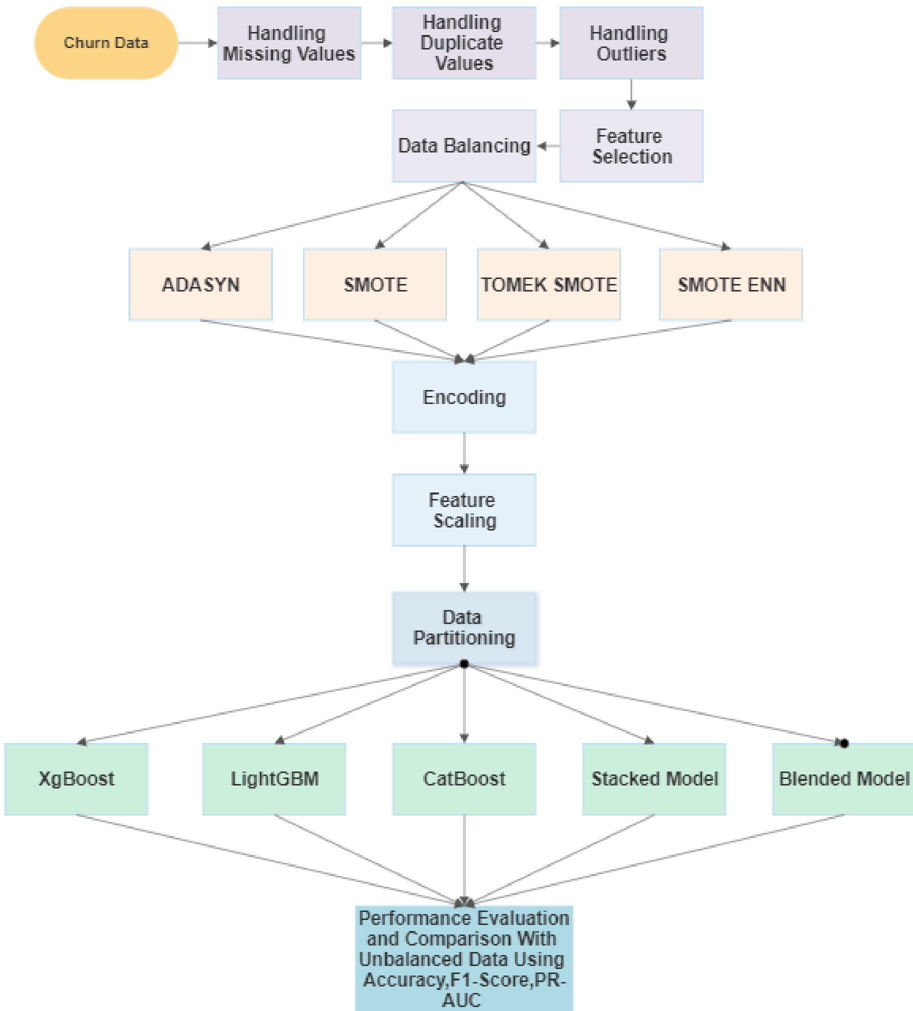

Fig. 1. Schematic Diagram of Proposed Methodology

the data by increasing the instances of minority class thus oversampling it using SMOTE and reducing the instances of majority class thus undersampling it [12]. Therefore, SMOTE ENN is more aggressive in balancing the dataset and has only 5907 records after rebalancing. Table 1 compares the class count for imbalanced and balanced datasets to assess and determine the effectiveness of various techniques employed to handle class imbalance in this study.

Data	Churners (0)	Non-churners (1)	Total count
Imbalanced	1836 (27.33%)	4880 (72.67%)	6716
SMOTE	4880 (50.00%)	4880 (50.00%)	9760
ADASYN	4880 (49.87%)	4904 (50.13%)	9784
SMOTE Tomek	4604 (48.54%)	4880 (51.46%)	9484
SMOTE ENN	2714 (45.94%)	3193 (54.06%)	5907

Table 1. Class count for imbalanced and balanced data.

Encoding: The categorical features were encoded using One Hot Encoding. In the dataset, there are some categorical features with binary classes which could be mapped to 1 or 0. But this may add bias to the class mapped as 1. To deal with this problem one hot encoding created separate columns for each label. It creates 'n' number of columns for the features having 'n' number of labels. Standardization: Considering the difference between the values of the 'monthly charges' and 'total charges' with the other encoded features, these two features were scaled. This prevents bias towards the higher values as the algorithms used for classification consider the distance between the points for classification. The 'StandardScaler' function is used for scaling.

The following classifiers will be explored XGBoost, LightGBM, CatBoost, Stacking and Blending approach.

Considering imbalanced data, accuracy alone is not an appropriate approach for performance evaluation as the result will be biased towards the majority class and would give higher accuracy [4].

In the case of churn prediction, the model must have a low False Negative (FN) and high True Positive (TP) values. Further, the performance will be evaluated using Accuracy, F-measure, recall, precision and PR AUC

4 Experimental results

In the case of imbalanced data, the accuracy of models built using stacked ensembles was the lowest at 77.29%, slightly lower than that of the blended ensemble models at 78.27%. However, these figures improved significantly upon data balancing, with the stacked and blended models surpassing the performance of individual boosting models like LightGBM and CatBoost. Notably, LightGBM and CatBoost demonstrated higher accuracy with imbalanced data, except when combined with the SMOTE ENN technique, where all models achieved their peak accuracy. Among these, the stacked model with SMOTE ENN emerged as the most accurate, closely followed by CatBoost.

In summary, the application of hybrid balancing techniques notably enhanced the accuracy of all models. The stacked and blended models particularly excelled, outshining other boosting algorithms. It's important to note that the data imbalance did not significantly impact the accuracy of these models, with XgBoost showing similar results to the blended models.

However, given the imbalanced nature of the data, there's a risk of the models being biased towards the majority class, potentially inflating the accuracy. This suggests that relying solely on the accuracy metric may not provide a complete assessment of a model's performance. To address this limitation and avoid the accuracy paradox, other metrics such as the F1-score and PR-AUC were employed for a more comprehensive evaluation.

Balancing Technique	XgBoost	LightGBM	CatBoost	Stacked	Blended
Unbalanced data	79.71	80.89	81.38	77.29	78.72
ADA SYN	78.67	76.83	74.37	79.31	79.89
SMOTE	78.79	77.93	78.07	79.38	79.77
Tomek Link	81.68	80.76	79.73	83.63	83.57
SMOTE ENN	94.92	93.89	95.32	96.46	93.96

Table 2. Accuracy comparison of models

In the context of an imbalanced dataset, all models initially displayed low F1 scores, with the stacked model notably at just 54%. This low score reflected poor precision, indicating the model's inefficiency in accurately predicting the churn class, and low recall, signifying its inability to correctly predict across the entire dataset. However, the F1 scores for all models showed significant improvement after data balancing.

Among the balancing techniques, the combination of SMOTE ENN with a stacked model emerged as the most effective, closely followed by the integration of SMOTE Tomek-links. In contrast, ADASYN and standard SMOTE produced similar impacts on the F1 scores across all models, without notable enhancements.

When considering individual boosting algorithms, XgBoost stood out with superior F1 scores. LightGBM and CatBoost yielded comparable results, but XgBoost was closer in performance to the stacked and blended models, particularly in handling the imbalanced dataset. Therefore, for achieving a better F1 score, hybrid balancing techniques, especially SMOTE ENN in conjunction with a stacked model, are recommended. In terms of F1 score, XgBoost demonstrated a clear advantage over LightGBM and CatBoost.

For imbalanced data stacked and blended models failed to exceed the value obtained by other ensemble models. Catboost outperformed all the other models for the imbalanced dataset.

Balancing Technique	XgBoost	LightGBM	CatBoost	Stacked	Blended
Unbalanced data	60.00	62.00	61.00	54.00	56.00
ADA SYN	80.00	78.00	76.00	80.00	81.00
SMOTE	79.00	79.00	79.00	80.00	80.00
Tomek Link	83.00	82.00	81.00	85.00	84.00
SMOTE ENN	95.00	94.00	96.00	97.00	94.00

Table 3. F1 score comparison of models

Comparing the effect of balancing techniques on PR-AUC value SMOTE ENN outperformed all other models followed by SMOTE Tomek-links. ADASYN and SMOTE did not show considerable differences in their PR-AUC values for all the models.

The PR-AUC score was highest for the stacked model and the blended model for all the balancing techniques. XgBoost obtained similar results. Clearly, SMOTE hybrid techniques, SMOTE Tomek-links and SMOTE ENN helped achieve higher PR-AUC values however the latter outperformed the former.

To sum up, Stacked, blended and XgBoost models showed better performance for PR-AUC when combined with SMOTE ENN.

Balancing Technique	XgBoost	LightGBM	CatBoost	Stacked	Blended
Unbalanced data	66.00	67.60	67.90	61.10	63.70
ADA SYN	83.60	82.40	80.90	84.00	84.70
SMOTE	83.70	83.10	83.20	84.20	84.40
Tomek Link	87.00	86.30	85.70	88.30	88.10
SMOTE ENN	96.20	95.40	96.50	97.40	95.90

Table 4. PR-AUC comparison of models

4.1 Comparison of models' performance over imbalanced and balanced dataset

Table 5 compares the performances of all the models trained over imbalanced and balanced datasets. For the models trained over the imbalanced dataset, the accuracy of all the models was comparatively higher than F1-score and PR-AUC due to the accuracy paradox. However, these values increased significantly after balancing the dataset. A major change in the F1 score was seen for the balanced data.

5 Discussion

Churn prediction, a supervised binary classification problem, has seen promising results with the application of machine learning. Recently, newer ensemble

Classifier	Before Balancing			After Balancing		
	Accuracy	F1-Score	PR-AUC	Accuracy	F1-Score	PR-AUC
XgBoost	79.71	60.00	66.00	94.94	95.00	96.20
LightGBM	80.89	62.00	67.60	93.89	94.00	95.40
CatBoost	81.38	61.00	67.90	95.32	96.00	96.50
Stacked	77.29	54.00	61.10	96.46	97.00	97.40
Blended	78.72	56.00	63.70	93.96	94.00	95.90

Table 5. Comparison of performance for imbalanced and balanced data

classifiers like LightGBM, CatBoost, and XgBoost have demonstrated superior performance compared to traditional algorithms, yet their application remains limited in studies [8]. To address dataset imbalance, data-level resampling techniques, particularly hybrid SMOTE methods such as SMOTE Tomek-links and SMOTE ENN, have enhanced model performance [26].

Regarding the effect of data imbalance on model accuracy, only SMOTE ENN showed a notable improvement after balancing. XgBoost, in particular, exhibited higher accuracy with balancing techniques compared to LightGBM and CatBoost, corroborating findings from [8]. However, Liu's study [13] indicates that LightGBM, when combined with ADASYN, outperformed other algorithms with SMOTE. Given that other metrics for imbalanced data were suboptimal, accuracy alone is not a sufficient measure of performance.

In terms of F1 score and PR-AUC, stacked and blended models surpassed individual boosting algorithms, reducing bias from weaker learners. Stacking, where the second-level classifier uses the first level's predictions as inputs, significantly improved performance. Models trained with SMOTE and ADASYN were less effective than those using hybrid SMOTE techniques. SMOTE ENN was particularly effective, simultaneously oversampling the minority class and undersampling the majority class.

The F1 score is crucial in imbalanced datasets, balancing precision and recall trade-offs. The study found that the highest F1 scores were achieved with SMOTE ENN combined with stacked algorithms, with CatBoost also performing well with SMOTE ENN. This is supported by [8], who found that hybrid SMOTE techniques enhanced performance in several classifiers, though XgBoost fared better with standard SMOTE.

Given class imbalances, PR-AUC is a more reliable metric than AUC-ROC, as it focuses on the positive class (churn) and provides a more accurate reflection of precision and recall. A higher PR-AUC value indicates lower rates of false positives and negatives, crucial for accurately identifying churners. SMOTE ENN, when used with all proposed classifiers, achieved high PR-AUC values, with XgBoost showing comparable results, thus making it a reliable choice for predictions in imbalanced data scenarios.

Table 6 shows the comparison between the accuracy and F1 score of this study with the results obtained by the other studies conducted previously. It can be clearly seen that the stacked model trained using the hybrid SMOTE ENN

	Source	Method	Accuracy	F1 Score
Proposed study		SMOTE ENN with stacked model	96.46%	97.00%
Paper 2	Fujo, Subramanian, and Khder (2022)	Random oversampling with Deep BP ANN	88.12%	88.68%
Paper 3	[8]	SMOTE Tomek Links with LightGBM	77.10%	61.90%
Paper 4	[3]	Random oversampling using Logistic Regression	76.20%	77.10%
Paper 5	[17]	XgBoost	79.80%	58.20%
Paper 6	[9]	Adaboost	81.70%	-
Paper 7	[24]	SMOTE using Adaboost	77.19%	63.11%

Table 6. Comparison of results with other studies.

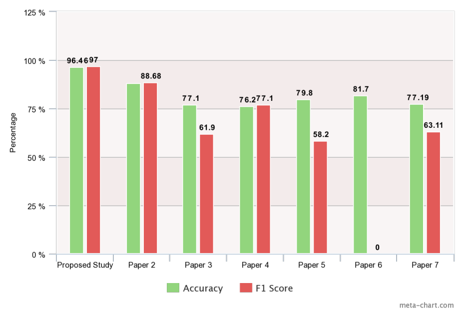

Fig. 2. Graphical comparison of the results with other studies

balancing technique showed significantly high accuracy and F1 score compared to other studies. In this case, machine learning classifiers outperformed neural networks combined with random oversampling. Thus, it can be concluded that stacking the individual algorithms together helps improve the model's performance. Also, hybrid SMOTE balancing techniques outperformed random oversampling or SMOTE techniques.

The analysis of the results yields several important insights: Data balancing significantly enhanced the performance of the models, with hybrid SMOTE techniques showing considerable improvements over ADASYN and standard SMOTE. In this context, both stacked and blended ensemble models outperformed individual ensemble algorithms. LightGBM and CatBoost were less affected by imbalanced data, but XgBoost stood out, surpassing both in terms of F1-score and PR-AUC. However, when LightGBM and CatBoost were paired with ADASYN and SMOTE, they achieved lower metric values. Overall, the exceptional performance was most notable in XgBoost, stacked, and blended models, especially when combined with the SMOTE ENN balancing technique.

6 Conclusion

Addressing the crucial problem of customer churn, which has a direct impact on a company's revenue and profits, this study emphasizes the importance of developing accurate predictive models. This is particularly essential since retaining existing customers is more cost-effective than acquiring new ones. By employing machine learning techniques, the study aimed to predict customer churn, while also delving into the issue of data imbalance and its influence on the predictive accuracy of the models. The research involved a thorough comparison of various models, including those based on boosting ensemble techniques like XgBoost, Light GBM, and Cat Boost, as well as a model utilizing a two-level stacking and blending technique. This comparison was focused on understanding the models' performance with balanced versus imbalanced data, using different oversampling methods. The study concluded that while the accuracy of these models on imbalanced data was close to that on balanced data, other evaluation metrics were significantly impacted by data imbalance. Therefore, relying solely on accuracy for model selection was deemed unrealistic. Among the balancing techniques examined, such as ADA SYN, SMOTE, and its hybrid methods (SMOTE ENN and SMOTE Tomek), SMOTE ENN proved to be the most effective. It notably enhanced all evaluation metrics. Additionally, stacking traditional machine learning algorithms with XgBoost at the second level of the stack showed improved results on balanced data across all four balancing techniques, with the highest accuracy observed in combination with the SMOTE ENN technique. Overall, the study filled research gaps in churn prediction by proposing suitable balancing techniques and identifying the most appropriate predictive models.

References

1. Ahmad, A.K., Jafar, A., Aljoumaa, K.: Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data **6**(1), 1–24 (2019)
2. Ali, H., Salleh, M.M., Saedudin, R., Hussain, K., Mushtaq, M.F.: Imbalance class problems in data mining: A review. Indonesian Journal of Electrical Engineering and Computer Science **14**(3), 1560–1571 (2019)
3. Beeharry, Y., Tsokizep Fokone, R.: Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. Concurrency and Computation: Practice and Experience **34**(4), e6627 (2022)
4. Bekkar, M., Djemaa, H.K., Alitouche, T.A.: Evaluation measures for models assessment over imbalanced data sets. J Inf Eng Appl **3**(10) (2013)
5. De, S., Prabu, P., Paulose, J.: Effective ml techniques to predict customer churn. In: 2021 Third international conference on inventive research in computing applications (ICIRCA). pp. 895–902 (2021)
6. Effendy, V., Baizal, Z.A.: Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest. In: 2014 2nd International Conference on Information and Communication Technology (ICoICT). pp. 325–330 (2014)
7. Hancock, J., Khoshgoftaar, T.M.: Medicare fraud detection using catboost. In: 2020 IEEE 21st international conference on information reuse and integration for data science (IRI). pp. 97–103 (2020)
8. Kimura, T.: Customer churn prediction with hybrid resampling and ensemble learning. Journal of Management Information Decision Sciences **25**(1) (2022)
9. Lalwani, P., Mishra, M.K., Chadha, J.S., Sethi, P.: Customer churn prediction system: a machine learning approach. Computing pp. 1–24 (2022)
10. Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A., Seliya, N.: A survey on addressing high-class imbalance in big data. Journal of Big Data **5**(1), 1–30 (2018)

11. Lemmens, A., Gupta, S.: Managing churn to maximize profits. Marketing Science **39**(5), 956–973 (2020)
12. Lin, M., Zhu, X., Hua, T., Tang, X., Tu, G., Chen, X.: Detection of ionospheric scintillation based on xgboost model improved by smote-enn technique. Remote Sensing **13**(13), 2577 (2021)
13. Liu, J., Gao, Y., Hu, F.: A fast network intrusion detection system using adaptive synthetic oversampling and lightgbm. Computers Security **106**, 102289 (2021)
14. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information sciences **250**, 113–141 (2013)
15. Machado, M.R., Karray, S., de Sousa, I.T.: Lightgbm: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In: 2019 14th International Conference on Computer Science Education (ICCSE). pp. 1111–1116 (2019)
16. Mishra, A., Reddy, U.S.: A comparative study of customer churn prediction in telecom industry using ensemble based classifiers. In: 2017 International conference on inventive computing and informatics (ICICI). pp. 721–725 (2017)
17. Pamina, J., Raja, B., SathyaBama, S., Sruthi, M.S., VJ, A.: An effective classifier for predicting churn in telecommunication. Jour of Adv Research in Dynamical Control Systems **11** (2019)
18. Reichheld, F.F., Sasser, W.E.: Zero defections: quality comes to services. Harvard Business Review **68**(5), 105–111 (1990)
19. Sabbeh, S.F.: Machine-learning techniques for customer retention: A comparative study. International Journal of advanced computer Science and applications **9**(2) (2018)
20. Sharma, T., Gupta, P., Nigam, V., Goel, M.: Customer churn prediction in telecommunications using gradient boosted trees. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019, Volume 2. pp. 235–246. Springer Singapore (2020)
21. Singh, M., Singh, S., Seen, N., Kaushal, S., Kumar, H.: Comparison of learning techniques for prediction of customer churn in telecommunication. In: 2018 28th International Telecommunication Networks and Applications Conference (ITNAC). pp. 1–5. IEEE (2018)
22. Veganzones, D., Séverin, E.: An investigation of bankruptcy prediction in imbalanced datasets. Decision Support Systems **112**, 111–124 (2018)
23. Wierenga, B.: Marketing and artificial intelligence: Great opportunities, reluctant partners. In: Marketing intelligent systems using soft computing: Managerial and research applications, pp. 1–8. Springer Berlin Heidelberg (2010)
24. Wu, S., Yau, W.C., Ong, T.S., Chong, S.C.: Integrated churn prediction and customer segmentation framework for telco business. IEEE Access **9**, 62118–62136 (2021)
25. Yap, B.W., Rani, K.A., Rahman, H.A.A., Fong, S., Khairudin, Z., Abdullah, N.N.: An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In: Proceedings of the first international conference on advanced data and information engineering (DaEng-2013). pp. 13–22. Springer Singapore (2014)
26. Zhu, B., Baesens, B., vanden Broucke, S.K.: An empirical comparison of techniques for the class imbalance problem in churn prediction. Information sciences **408**, 84–89 (2017)