**International Journal of Machine Learning**
P-ISSN: xxxx-xxxx
E-ISSN: xxxx-xxxx
Vol. 1, No. 1, June 2026, Page. 1-11
https://ijoml.com/
DOI: https://doi.org/10.52436/1.ijoml.1.1.id

# Evaluation of Undersampling and Oversampling Techniques in Term Deposit Prediction: A Gradient Boosting Approach

**Lasmedi Afuan*[1], Abdul Karim[2], Ipung Permadi[3]**

[1,3]Department of Informatics, Universitas Jenderal Soedirman, Indonesia
[2]Department of Artificial Intelligence Convergence, HallymUniversity, Chuncheon 24252, Republic of Korea

Email: [1]lasmedi.afuan@unsoed.ac.id

Received: Jan 9, 2026; Revised: Feb 28, 2026; Accepted: Feb 28, 2026; Published: June 1, 2026

**Abstract**

Time deposits play a pivotal role in maintaining banking liquidity, yet telemarketing campaigns designed to secure them are often inefficient due to low response rates and untargeted outreach. The primary challenge in predictive marketing modeling lies in extreme data class imbalance, which renders standard algorithms prone to bias and leads to a failure in detecting potential customers. This study aims to validate the effectiveness of Gradient Boosting models and empirically evaluate the impact of various resampling techniques in mitigating class distribution disparities. The applied methodology encompasses the utilization of XGBoost, LightGBM, and CatBoost algorithms on the UCI Bank Marketing dataset, integrated with Random Under-Sampling, Random Over-Sampling, SMOTENC, and Tomek Links strategies. Experimental results reveal a significant trade-off between sensitivity and precision, wherein LightGBM paired with Random Under-Sampling achieved the highest detection capability with a Recall of 88.28%. Concurrently, the combination of CatBoost with Random Over-Sampling demonstrated the optimal balance, attaining an F1-Score of 0.6040, a Recall of 81.95%, and an AUC-ROC value reaching 0.9326. These findings offer a strategic contribution to bank management in selecting analytic approaches aligned with business priorities, whether the focus is on operational cost efficiency or aggressive market penetration to optimize customer acquisition.

***Keywords:*** *Bank Marketing, Class Imbalance, Gradient Boosting, Resampling, Telemarketing.*



## 1. INTRODUCTION

Time deposits constitute a fundamental source of funding and liquidity, playing a critical role in maintaining financial stability and supporting a bank's credit distribution capacity [1], [2]. A bank's ability to secure and retain these deposits directly influences the future expansion of its lending business [3], [4]. To achieve these objectives, banks predominantly utilize telemarketing as a strategic method to promote deposit products to prospective customers [5], [6]. Nevertheless, telemarketing campaigns are frequently hindered by operational inefficiencies and low response rates [5]. Such untargeted outreach results in a significant drain on time, financial resources, and organizational energy [5], [6], while simultaneously causing customer dissatisfaction [3], [6].

To mitigate these inefficiencies and identify potential customers with greater precision, the implementation of Data Mining and Machine Learning (ML) has become indispensable [5], [6]. By analyzing historical customer data, ML algorithms can extract latent patterns and construct predictive models that assist banks in targeting their marketing efforts effectively [5], [7]. However, bank marketing datasets often contend with the issue of class imbalance [5], [6]. The data typically exhibits a skewed distribution where the vast majority of customers decline the offer, leaving only a small fraction of acceptances [5], [6]. This disparity causes standard classification models to exhibit bias toward the majority class, yielding high overall accuracy but failing to perform adequately on the

P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

minority class [8]. Consequently, crucial metrics for identifying potential customers, such as Recall and F1-Score, remain suboptimal [6].

In the current landscape, Gradient Boosting-based Ensemble Learning methods (such as XGBoost, LightGBM, and CatBoost) have emerged as the superior choice, delivering performance that surpasses classic single classifier models in tabular data classification tasks [1], [6], [9]. These algorithms aggregate multiple weak learners, specifically decision trees, in a sequential manner where each new tree attempts to correct the errors committed by its predecessor [10]. This mechanism inherently renders boosting effective in handling data that is difficult to classify [11]. Categorical Boosting (CatBoost), in particular, demonstrates a distinct advantage due to its ability to handle categorical features automatically without requiring complex preprocessing such as One-Hot Encoding [9], [12]. This capability mitigates target leakage and prediction shift, which are common issues in other boosting algorithms, thereby enhancing both accuracy and scalability on datasets rich in categorical variables [9], [12], [13].

Despite the robustness and adaptability of boosting algorithms, severe class imbalance remains a persistent challenge that necessitates data-level resampling techniques to equilibrate the class distribution [5], [6]. Popular oversampling methods, such as Synthetic Minority Over-sampling Technique (SMOTE), generate synthetic samples for the minority class [6]. Conversely, under sampling methods like Random Under-Sampling (RUS) and Tomek Links function by randomly removing samples from the majority class [13], [14]. However, the selection of the optimal resampling technique remains a subject of ongoing debate [15]. Synthetic methods like SMOTE may inadvertently introduce noise or overlapping samples, particularly near decision boundaries, which can potentially degrade precision [13], [16]. On the other hand, simpler under sampling methods such as RUS carry the risk of discarding valuable information from the majority class [13]. Therefore, the application of these methods requires careful evaluation to ensure the resulting models are both stable and unbiased.

Against this backdrop, this research aims to validate the superiority of Gradient Boosting models in predicting time deposit subscriptions and to empirically evaluate the impact of resampling techniques on the stability and performance of predictive models in addressing class imbalance.

## 2. METHOD

This chapter delineates the methodology adopted to address the challenge of banking customer prediction. The research workflow commences with data acquisition and comprehension, followed by data preprocessing techniques to ensure data quality. Given the uneven distribution of the target class, resampling strategies are integrated into the process prior to training the models using boosting-based algorithms. The specific procedural steps and the proposed framework for this study are illustrated in Figure 1.

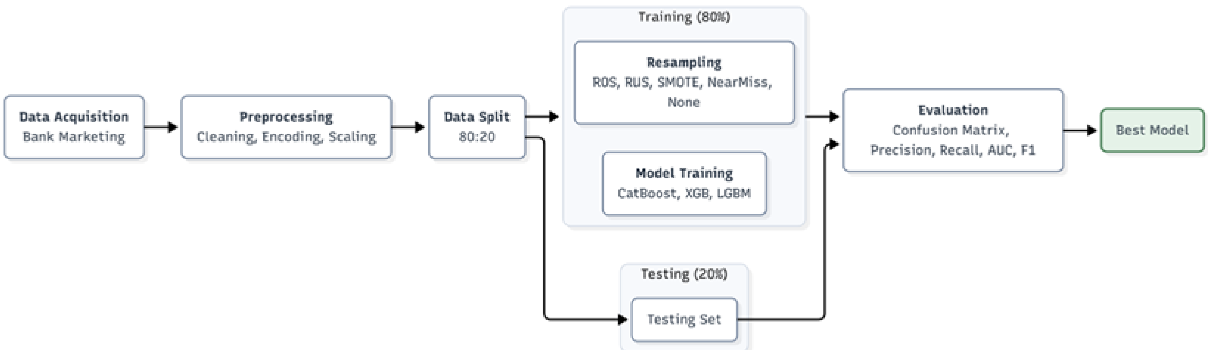


Figure 1. Flow of Research

P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

### 2.1. Data Source and Description

The data for this study is derived from the UCI Bank Marketing Dataset, a widely recognized and credible secondary source detailing the outcomes of direct telemarketing campaigns conducted by a Portuguese banking institution. Comprising 45,211 instances, the dataset captures the demographic profile, financial status, and campaign history of the bank's clients. The input features encompass a broad range of variables, including age, job, marital status, education, credit default status, average annual balance, housing loan status, personal loan status, contact communication type, day and month of the last contact, duration of the last contact, number of contacts performed during this campaign, number of days passed since the previous contact (pdays), number of previous contacts, and the outcome of the previous marketing campaign (poutcome). The primary target variable, denoted as 'y', is a binary indicator reflecting whether the client successfully subscribed to a term deposit.

Intrinsically, this dataset is characterized by a significant class imbalance. Existing literature confirms that the volume of customers who decline the deposit offer vastly exceeds those who accept it [17]. Subsequent studies utilizing this data have demonstrated that fewer than 12% of customers respond affirmatively to the subscription offer [6]. Addressing this extreme imbalance is a critical challenge that must be resolved to prevent the predictive model from exhibiting bias toward the majority class.

### 2.2. Data Preprocessing

The data preprocessing stage is pivotal for enhancing data quality and consistency prior to modeling [18]. The process commences with data cleaning, which encompasses the handling of missing values and outliers [17]. Within the context of this dataset, prior research indicates that categorical features require transformation into numerical representations via encoding methods [17], [19]. Specifically, nominal variables such as job type and marital status are converted using one-hot encoding, whereas binary features undergo label encoding [19]. Concurrently, numerical features including age, balance, and duration must be normalized or scaled utilizing the StandardScaler method to ensure a uniform value range across all features [17]. Such standardization is critical, as machine learning algorithms, particularly those based on boosting, exhibit high sensitivity to the scale of input features [4], [17].

Once the data preparation is complete, the dataset is partitioned into two subsets using a train-test split method, allocating 80% for training and 20% for testing. The selection of this 80:20 ratio aligns with standard practices widely adopted in machine learning research for classification tasks, including within the domain of financial prediction [17], [19], [20].

### 2.3. Resampling Strategies

To address the confirmed class imbalance within the dataset, resampling strategies were applied exclusively to the training data [17]. Restricting resampling to the training set is a recognized best practice to prevent data leakage, thereby ensuring the test data remains pristine and accurately representative of real-world application scenarios [21]. Four distinct resampling techniques were utilized to generate four modified versions of the training dataset. Designed to manipulate class distribution through various mechanisms, these techniques include Random Under-Sampling (RUS), Random Over-Sampling (ROS), the Synthetic Minority Over-sampling Technique for Nominal and Continuous features (SMOTENC), and Tomek Links (TL).

### 2.4. Classification Algorithms

This experiment centers on a comparative performance analysis of three modern boosting algorithms renowned for their efficacy in handling high-dimensional data and class imbalance like XGBoost, LightGBM, and CatBoost [22]. XGBoost is distinguished by its efficient and scalable

P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

gradient boosting framework, incorporating regularization techniques to rigorously control model complexity [23]. Complementing this, LightGBM, developed by Microsoft, is acclaimed for its exceptional training speed and memory efficiency achieved through a distinctive leaf-wise tree growth strategy [22]. Furthermore, CatBoost demonstrates specific advantages in the automatic processing of categorical features and employs ordered boosting to mitigate prediction shift and gradient bias, establishing it as a robust solution for mixed data types [22], [24]. Within the scope of this study, extensive hyperparameter tuning was intentionally omitted. Instead, each model was executed using default configurations to establish an objective performance baseline prior to the application of resampling techniques [4].

### 2.5. Experimental Setup

A total of 15 experimental scenarios were designed to rigorously analyze the performance of each algorithm under varying data conditions. Each of the three classification algorithms (XGBoost, LightGBM, and CatBoost) was trained across five distinct versions of the dataset: the original training data without resampling and four modified versions processed using different resampling techniques (RUS, ROS, SMOTENC, and Tomek Links). This combinatorial framework aims to facilitate a comprehensive evaluation, comparing not only the performance differences between algorithms but also the specific impact of each resampling strategy on the models' sensitivity in identifying the minority class.

### 2.6. Performance Evaluation Metrics

Model performance evaluation focuses rigorously on metrics most relevant to datasets exhibiting class imbalance. The assessment includes Precision, which quantifies the accuracy of positive predictions, and Recall (or sensitivity), which measures the model's aptitude for identifying all actual positive cases. Crucially, the F1-Score is utilized as the primary metric to determine the equilibrium of model performance by calculating the harmonic mean of Precision and Recall [18]. Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is computed to assess the model's discriminative ability across varying classification thresholds, serving as a vital indicator for domains defined by class imbalance [25].

## 3. RESULT

### 3.1. Data Analysis and Preprocessing

Initial exploration of the UCI Bank Marketing Dataset reveals a raw dataset comprising 45,211 entries and 17 primary features. An analysis of the class distribution for the target variable 'y' (TermDepositSubscribed) confirms an extreme imbalance; the majority class ('no') dominates with 39,922 samples (88.30%), while the minority class ('yes') accounts for merely 5,289 samples (11.70%). This significant disparity underscores the necessity for an evaluation strategy that looks beyond simple accuracy, prioritizing metrics sensitive to the minority class such as F1-Score and ROC-AUC. Furthermore, descriptive statistical analysis identified outliers within key numerical features, specifically age, balance, and duration. To mitigate this issue and maintain model stability, a percentile-based capping technique (clipped at the 5th and 95th percentiles) was implemented. This method effectively compressed the range of extreme values without discarding valuable information. The results of this outlier mitigation process are visualized in Figure 2.

Vol. 1, No. 1, June 2026, Page. 1-11
https://ijoml.com/
DOI: https://doi.org/10.52436/1.ijoml.1.1.id

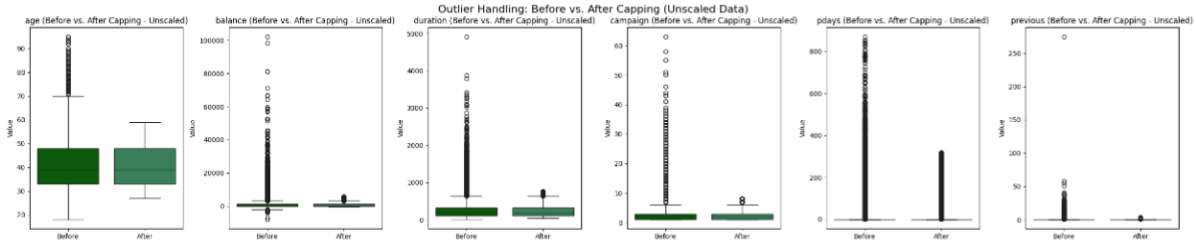


Figure 2. Visualization of Outlier Mitigation Using Percentile-Based Capping

Subsequently, data transformation was executed to align the input format with the requirements of machine learning algorithms. Nominal categorical features, such as job and marital status, were converted using One-Hot Encoding, which expanded the dataset's dimensionality from 17 to 45 columns following preprocessing. Numeric features underwent standardization via StandardScaler, resulting in value distributions with a mean approximating 0 and a standard deviation of 1, as evidenced by the post-processing statistical summary. Additionally, feature engineering successfully extracted novel insights, introducing features such as pdays_was_contacted to distinguish between new and previously contacted customers, and balance_per_age to capture financial capacity relative to the client's age. Finally, the cleaned dataset was partitioned using a stratified 80:20 split ratio, yielding 36,168 samples for training and 9,043 samples for testing. This stratification ensured that the proportion of the positive class remained consistent at approximately 11.7% across both subsets.

### 3.2. Baseline Experiment Results

In this initial experimental phase, three gradient boosting algorithms—CatBoost, XGBoost, and LightGBM—were trained utilizing the preprocessed original dataset prior to any class imbalance intervention. The primary objective was to establish a performance baseline, serving as a benchmark for evaluating the effectiveness of the subsequent resampling strategies. Evaluations were conducted on unseen test data using a comprehensive set of performance metrics.

Table 1. Performance Metrics of Models on Original Data

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC | AUC-PR |
|---|---|---|---|---|---|---|
| CatBoost | 0.9067 | 0.6379 | 0.4679 | 0.5398 | 0.9316 | 0.6183 |
| XGBoost | 0.9058 | 0.631 | 0.4688 | 0.538 | 0.9261 | 0.5982 |
| LightGBM | 0.9077 | 0.6389 | 0.4849 | 0.5513 | 0.9315 | 0.6205 |

As presented in Table 1, all three models demonstrated exceptionally high Accuracy rates, ranging from 90.5% to 90.7%. These figures, however, can be misleading as they are heavily skewed by the prevalence of the majority class. More critical insights are revealed through the Recall and F1-Score metrics. LightGBM achieved the superior baseline performance, recording an F1-Score of 0.5513 and a Recall of 0.4849, slightly outperforming both CatBoost and XGBoost. Nevertheless, with Recall values remaining below 50%, the baseline models failed to detect more than half of the actual potential customers likely to subscribe to a deposit. This confirms the hypothesis that, in the absence of imbalance handling, algorithms exhibit a bias toward the majority class, resulting in a high volume of False Negatives.

**International Journal of Machine Learning** Vol. 1, No. 1, June 2026, Page. 1-11
P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

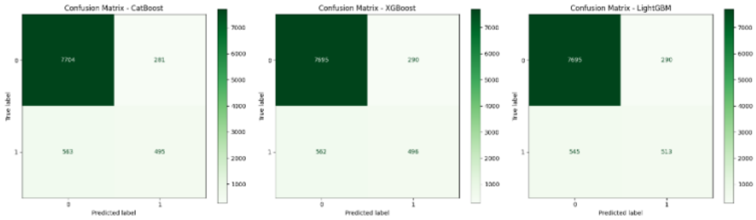


Figure 3. Baseline Models Confusion Matrix Visualization

Referring to Figure 3, it is evident that while the models are highly proficient in predicting class '0' (No), there is a substantial rate of misclassification for class '1' (Yes). For instance, in the case of LightGBM, despite successfully predicting a portion of positive cases, a significant number of potential customers were missed and incorrectly classified as 'No'. These results provide a strong justification for the implementation of resampling methods in the subsequent phase, aiming to enhance model sensitivity toward the minority class while maintaining a reasonable level of precision.

### 3.3. Resampling Experiment Results

A comprehensive evaluation of twelve testing scenarios, involving combinations of three algorithms (CatBoost, XGBoost, LightGBM) and four resampling techniques, reveals a significant trade-off between Precision and Recall metrics. Table 2 presents a summary of model performance on the test data (unseen during training) based on key evaluation metrics.

Table 2. Performance Metrics of Models with Resampling Techniques

| Resampling | Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC | AUC-PR |
|---|---|---|---|---|---|---|---|
| SMOTENC | CatBoost | 0.9066 | 0.6081 | 0.5662 | 0.5864 | 0.9284 | 0.6152 |
| | XGBoost | 0.8987 | 0.5656 | 0.5784 | 0.572 | 0.9204 | 0.5791 |
| | LightGBM | 0.8996 | 0.5628 | 0.6352 | 0.5968 | 0.9257 | 0.5853 |
| ROS | CatBoost | 0.8743 | 0.4782 | 0.8195 | 0.604 | 0.9291 | 0.6103 |
| | XGBoost | 0.8759 | 0.4815 | 0.7883 | 0.5978 | 0.9232 | 0.6013 |
| | LightGBM | 0.8561 | 0.4421 | 0.8762 | 0.5876 | 0.9316 | 0.6175 |
| RUS | CatBoost | 0.847 | 0.4252 | 0.8752 | 0.5723 | 0.9279 | 0.5963 |
| | XGBoost | 0.8428 | 0.4179 | 0.8762 | 0.5659 | 0.924 | 0.5709 |
| | LightGBM | 0.8445 | 0.4215 | 0.8828 | 0.5706 | 0.9266 | 0.5862 |
| Tomek Links | CatBoost | 0.9085 | 0.6282 | 0.535 | 0.5778 | 0.9318 | 0.6237 |
| | XGBoost | 0.9061 | 0.615 | 0.5284 | 0.5684 | 0.9282 | 0.5996 |
| | LightGBM | 0.9066 | 0.612 | 0.5501 | 0.5794 | 0.9326 | 0.6177 |

Based on Table 2, a consistent pattern is observed where oversampling (ROS) and undersampling (RUS) techniques drastically improve the model's ability to detect the minority class (customers subscribing to deposits). This improvement is demonstrated by Recall values surging into the 78% to 88% range. Specifically, LightGBM combined with RUS and ROS recorded the highest Recall rates of 88.28% and 87.62% respectively. This increase in sensitivity is crucial within the context of bank marketing because failing to detect potential customers (False Negatives) is more detrimental than prediction errors regarding uninterested customers. However, this gain in Recall comes at the cost of overall Precision and Accuracy as the models tend to predict more False Positives. Conversely, the Tomek Links (TL) technique maintains high overall accuracy (above 90%) and the best Precision (reaching 62.82% with CatBoost) but fails to improve Recall significantly, which remains in the 53% to 55% range. This indicates that TL is more effective at clarifying decision boundaries rather than fundamentally addressing class bias issues.

P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

To validate the discriminative capability of the models in greater depth, a visual analysis was conducted using Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves.

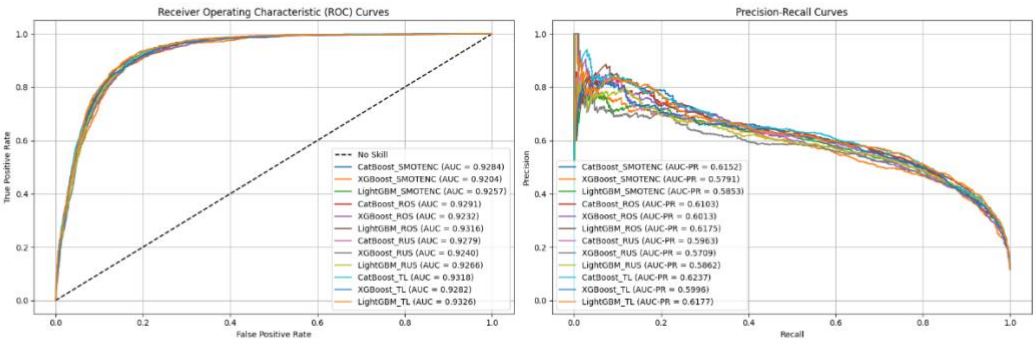


Figure 4. ROC and Precision-Recall Curves

Based on Figure 4, the ROC curves indicate that all models possess excellent performance in separating positive and negative classes in general, with AUC-ROC values consistently remaining above 0.92 for almost all variations. The LightGBM model trained with Tomek Links (LightGBM_TL) and Random Over-Sampling (LightGBM_ROS) recorded the highest AUC-ROC values of 0.9326 and 0.9316 respectively. Nevertheless, on imbalanced datasets, the Precision-Recall (PR) curve provides a more realistic depiction of model performance regarding the minority class. As seen in the PR graph in Figure 4, CatBoost with the Tomek Links strategy (CatBoost_TL) produced the highest AUC-PR value of 0.6237, followed by LightGBM_ROS (0.6175) and CatBoost_SMOTENC (0.6152). This indicates that while aggressive resampling techniques like RUS increase Recall, those models lose too much precision compared to models based on TL or SMOTENC which maintain a better balance between precision and sensitivity.

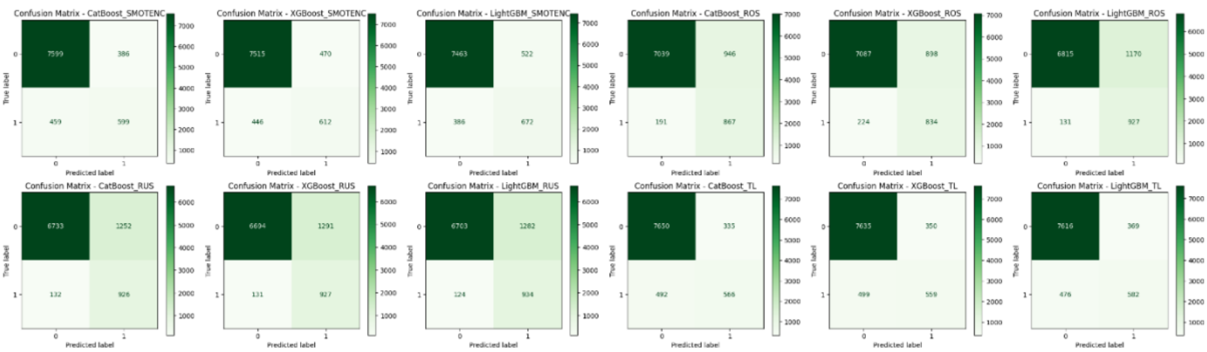


Figure 5. Confusion Matrix Analysis

Based on Figure 5, a clear shift in prediction behavior between strategies is visible. Models utilizing RUS (the third row in the matrix) successfully identify the majority of potential customers (high True Positives) but also misclassify a large number of non-potential customers as potential. In contrast, models with Tomek Links are highly conservative; they produce few errors on the negative class but miss nearly half of the target potential customers. Among all scenarios, CatBoost with the Random Over-Sampling (ROS) strategy and SMOTENC offer the most compelling compromise from a business perspective. CatBoost_ROS achieves the highest F1-Score (0.6040) with excellent Recall (81.95%), meaning this model is capable of capturing the majority of the market target while maintaining an acceptable level of accuracy.

## 4. DISCUSSION

This research underscores that addressing class imbalance is a critical step that significantly influences predictive model performance, a finding consistent with prior literature on imbalanced class studies [4], [13], [14], [17]. While oversampling and undersampling techniques have proven effective in enhancing metrics such as Recall and F1-Score, no single solution is universally optimal for all data scenarios. The validation process demonstrates that performance improvements inevitably involve an inherent trade-off between Precision and Recall.

An analysis of the employed resampling mechanisms elucidates the specific impact of each approach. The RUS method may lead to a decline in Precision due to the risk of information loss resulting from the random removal of majority data [13], [24]. Conversely, oversampling techniques such as SMOTE and ROS tend to bolster Recall or sensitivity by expanding the decision boundaries of the minority class through the generation of synthetic samples [17], [22]. However, this synthetic generation introduces risks that warrant caution, specifically the introduction of noise or overlapping samples which can degrade data quality [25], [26]. Meanwhile, cleaning methods such as Tomek Links maintain high Precision by removing noisy or ambiguous samples at class borders [27], thereby clarifying the boundaries between classes [28].

Empirically, the managerial implications of this study can be mapped into three distinct strategic priorities. First, in scenarios where institutions prioritize Cost-Efficiency, the CatBoost algorithm combined with Tomek Links emerges as the optimal choice. This combination yields the highest precision and effectively minimizes False Positives. Consequently, banks can avoid wasting time and agent resources on contacting uninterested customers, thereby increasing the return on investment (ROI) per call. Conversely, if an institution adopts an aggressive Market Penetration strategy where the objective is to capture as many potential customers as possible regardless of contact costs, LightGBM paired with Random Under-Sampling (RUS) or Random Over-Sampling (ROS) is the primary recommendation. The high Recall values in these models guarantee that the vast majority of interested customers are not overlooked by the system. Finally, for a Balanced Growth approach, LightGBM with SMOTENC or CatBoost with ROS offers the best compromise by maintaining a stable F1-Score around 0.60, effectively balancing new customer acquisition with operational efficiency.

Despite the promising performance of the developed models, several limitations must be considered when interpreting the results. The experiment relies heavily on the contact duration feature, which is often unknown prior to the call being made. High dependency on this feature can lead to "look-ahead" bias if not handled carefully during real-time implementation. Additionally, the use of oversampling techniques such as ROS and SMOTE carries the potential for overfitting on training data if the model memorizes duplicated data rather than learning general patterns, although this risk was partially mitigated by the built-in regularization mechanisms of the XGBoost and CatBoost algorithms. Future research is advised to explore cost-sensitive learning techniques as an alternative to resampling and to conduct model validation across different time periods (out-of-time validation) to test model robustness against dynamic changes in customer behavior.

## 5. CONCLUSION

This research confirms that the application of resampling techniques to bank marketing datasets significantly impacts the performance of boosting algorithms. Comparative evaluations demonstrate an inherent trade-off between sensitivity (Recall) and Precision across all experimental scenarios. Oversampling techniques (ROS and SMOTENC) and undersampling (RUS) proved effective in enhancing the models' capacity to detect potential customers, whereas the Tomek Links (TL) technique excelled in preserving high levels of Precision and overall accuracy.

P-ISSN: xxxx-xxxx https://ijoml.com/
E-ISSN: xxxx-xxxx DOI: https://doi.org/10.52436/1.ijoml.1.1.id

From a practical perspective, the selection of the optimal model must be aligned with the operational strategic priorities of the bank. For a strategy focused on cost efficiency, the combination of CatBoost and Tomek Links is recommended due to its superior capability in minimizing false positive predictions. Conversely, for a market penetration strategy prioritizing customer acquisition volume, LightGBM paired with Random Over-Sampling (ROS) emerges as the optimal choice owing to its high sensitivity. Furthermore, for institutions targeting balanced growth between acquisition and efficiency, CatBoost with ROS offers the best performance stability via an optimal F1-Score.

This study concludes that addressing data imbalance constitutes a crucial strategic decision within business analytics. Future research is advised to extend model validation using out-of-time data and to explore alternative predictive features beyond contact duration to enhance the robustness of model implementation in real-world settings.

## CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

## ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all individuals and institutions who have contributed to the research process and the preparation of this article, including colleagues, reviewers, and editors, whose valuable suggestions and constructive feedback have helped improve the quality of this work.

## REFERENCES

[1] Q. Yu, “Enhancing Bank Term Deposit Predictions: A Machine Learning Approach with CatBoost and SHAP,” *Applied and Computational Engineering*, vol. 120, no. 1, pp. 171–180, Jan. 2025, doi: 10.54254/2755-2721/2025.19485.

[2] Z. Yang, Y. Zhang, and L. Yu, “Predicting Bank Users’ Time Deposits Based on LSTM-Stacked Modeling,” *Acadlore Transactions on AI and Machine Learning*, vol. 3, no. 3, pp. 172–182, Sep. 2024, doi: 10.56578/ataiml030304.

[3] Y. Feng, Y. Yin, D. Wang, and L. Dhamotharan, “A dynamic ensemble selection method for bank telemarketing sales prediction,” *J Bus Res*, vol. 139, pp. 368–382, Feb. 2022, doi: 10.1016/j.jbusres.2021.09.067.

[4] M. Abbasi, S. López Flórez, A. Shahraki, A. Taherkordi, J. Prieto, and J. M. Corchado, “Class Imbalance in Network Traffic Classification: An Adaptive Weight Ensemble-of-Ensemble Learning Method,” *IEEE Access*, vol. 13, pp. 26171–26192, 2025, doi: 10.1109/ACCESS.2025.3538170.

[5] S. Kaisar, M. M. Rashid, A. Chowdhury, S. S. Shafin, J. Kamruzzaman, and A. Diro, “Enhancing Telemarketing Success Using Ensemble-Based Online Machine Learning,” *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 294–314, Jun. 2024, doi: 10.26599/BDMA.2023.9020041.

[6] L. Afuan and R. R. Isnanto, “A comparative study of machine learning algorithms for fall detection in technology-based healthcare system: analyzing SVM, KNN, decision tree, random forest, LSTM, and CNN,” in *E3S Web of Conferences*, EDP Sciences, Jan. 2025. doi: 10.1051/e3sconf/202560503051.

[7] M. A. T. Rony, Md. M. Hassan, E. Ahmed, A. Karim, S. Azam, and D. S. A. A. Reza, “Identifying Long-Term Deposit Customers: A Machine Learning Approach,” in *2021 2nd International Informatics and Software Engineering Conference (IISEC)*, IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/IISEC54230.2021.9672452.

[8] D. K. Katakam, “Enhancing Customer Subscription Prediction in Bank Telemarketing Using Deep Learning and Ensemble Model,” *Int J Res Appl Sci Eng Technol*, vol. 13, no. 8, pp. 335–340, Aug. 2025, doi: 10.22214/ijraset.2025.73563.

[9] Q. Zhu, X. Wang, N. Zhang, Y. Xu, and Y. He, “Industrial Imbalanced Fault Diagnosis Method Based on Borderline SMOTE Integrated with NPE and CatBoost,” in *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)*, IEEE, Aug. 2022, pp. 612–617. doi: 10.1109/DDCLS55054.2022.9858431.

[10] D. Rohmayani, C. A. Sugianto, R. S. Perdana, and M. M. Nafea, “Improving Extreme Gradient Boosting Model for Heart Disease Prediction Using SMOTE for Class Imbalance,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 1717–1728, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4753.

[11] O. Lyashevska, F. Malone, E. MacCarthy, J. Fiehler, J.-H. Buhk, and L. Morris, “Class imbalance in gradient boosting classification algorithms: Application to experimental stroke data,” *Stat Methods Med Res*, vol. 30, no. 3, pp. 916–925, Mar. 2021, doi: 10.1177/0962280220980484.

[12] Q. Wang, L. Li, B. Jiang, Z. Lu, J. Liu, and S. Jian, “Malicious Domain Detection Based on K-means and SMOTE,” in *Computational Science – ICCS 2020*, V. V Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, Eds., Cham: Springer International Publishing, 2020, pp. 468–481.

[13] S. Yu, V. Chang, G. Huỳnh, V. Jesus, and J. Luo, “Advanced Supervised Machine Learning Algorithms in Credit Card Fraud Detection,” in *Proceedings of the 7th International Conference on Finance, Economics, Management and IT Business*, SCITEPRESS - Science and Technology Publications, 2025, pp. 126–138. doi: 10.5220/0013485400003956.

[14] F. Nasir, A. A. Ahmed, M. S. Kiraz, I. Yevseyeva, and M. Saif, “Data-Driven Decision-Making for Bank Target Marketing Using Supervised Learning Classifiers on Imbalanced Big Data,” *Computers, Materials & Continua*, vol. 81, no. 1, pp. 1703–1728, 2024, doi: 10.32604/cmc.2024.055192.

[15] M. Rai, B. Ghimire, N. Uprety, and R. Shrestha, “A Comparative Analysis of Data Balancing Techniques in Machine Learning for Enhancing Bank Loan Default Risk Predictions,” *International Journal of Research Publications*, vol. 155, no. 1, Aug. 2024, doi: 10.47119/IJRP1001551820247097.

[16] K. Swain *et al.*, “Enhancing Stroke Prediction Using LightGBM With SMOTE-ENN and Fine-Tuning: A Comprehensive Analysis,” *Cureus Journal of Computer Science*, Dec. 2024, doi: 10.7759/s44389-024-02268-y.

[17] D. Z. Abidin, M. Rosario, A. Sadikin, N. Nurhadi, and J. Jasmir, “Improving Term Deposit Customer Prediction Using Support Vector Machine with SMOTE and Hyperparameter Tuning in Bank Marketing Campaigns,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 3, pp. 1267–1278, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4585.

[18] A. Aldelemy and Raed A. Abd-Alhameed, “Binary Classification of Customer’s Online Purchasing Behavior Using Machine Learning,” *Journal of Techniques*, vol. 5, no. 2, pp. 163–186, Jun. 2023, doi: 10.51173/jt.v5i2.1226.

[19] E. R. D. Aditya and D. K. Y. Satria, “Optimizing Bank Marketing Strategies Through Analysis Using Lightgbm,” *CoreID Journal*, vol. 1, no. 2, pp. 58–65, Jul. 2023, doi: 10.60005/coreid.v1i2.11.

[20] F. Aziz, M. Maulida, J. Jafar, N. Shahnyb, N. Nasir, and A. Ampauleng, “Development of an Optimized Ensemble Least Squares Model for Identifying Potential Deposit Customers,” *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 6, no. 1, pp. 48–59, Dec. 2024, doi: 10.37385/jaets.v6i1.5974.

[21] H.-J. Park, Y.-S. Koo, H.-Y. Yang, Y.-S. Han, and C.-S. Nam, “Study on Data Preprocessing for Machine Learning Based on Semiconductor Manufacturing Processes,” *Sensors*, vol. 24, no. 17, p. 5461, Aug. 2024, doi: 10.3390/s24175461.

[22] M. R. Anshari *et al.*, “Performance Comparison of AdaBoost, LightGBM, and CatBoost for Parkinson’s Disease Classification Using ADASYN Balancing,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3495–3508, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.4726.

[23] J. B. G. Brito *et al.*, “A framework to improve churn prediction performance in retail banking,” *Financial Innovation*, vol. 10, no. 1, p. 17, Jan. 2024, doi: 10.1186/s40854-023-00558-3.

Vol. 1, No. 1, June 2026, Page. 1-11
https://ijoml.com/
DOI: https://doi.org/10.52436/1.ijoml.1.1.id

[24] G. Li, Y. Wu, Y. Bai, and W. Zhang, "ReMAHA–CatBoost: Addressing Imbalanced Data in Traffic Accident Prediction Tasks," *Applied Sciences*, vol. 13, no. 24, p. 13123, Dec. 2023, doi: 10.3390/app132413123.

[25] N. A. Siagian, S. P. Sipayung, A. Rikki, and N. Marbun, "Integrating SMOTE with XGBoost for Robust Classification on Imbalanced Datasets: A Dual-Domain Evaluation," *sinkron*, vol. 9, no. 3, pp. 1094–1107, Jul. 2025, doi: 10.33395/sinkron.v9i3.15029.

[26] L.-H. Li, R. Ahmad, R. Tanone, and A. K. Sharma, "STB: synthetic minority oversampling technique for tree-boosting models for imbalanced datasets of intrusion detection systems," *PeerJ Comput Sci*, vol. 9, p. e1580, Nov. 2023, doi: 10.7717/peerj-cs.1580.

[27] Y. Li and Z. Wu, "Prediction of Customers' Subscription to Time Deposits Based on SMOTEENN-XGBoost Model," in *WHICEB 2022 Proceedings*, 2022, p. 61. [Online]. Available: https://aisel.aisnet.org/whiceb2022/61

[28] P. Kallappagol, S. R. Kawale, and S. Holaychi, "Early-Stage Diabetes Mellitus Risk Prediction Using Light Gradient-Boosting Machine With Synthetic Minority Oversampling Technique Combined With Edited Nearest Neighbors on a Novel Southern Indian Clinical Dataset," *Cureus Journal of Computer Science*, Sep. 2025, doi: 10.7759/s44389-025-06580-z.