International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 9 (2025) 136–147

Paper Type: Original Article

Enhancing Telecom Retention with a Stacking-Based Churn Model

Mostafa Bekheet [1,*], Nabil M. AbdelAziz [1], Ahmad Salah [1], Nissreen El-Saber [2] and Wafaa T. AbdelMoneim [1]

[1] Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt.
Emails: m.bekhet020@fci.zu.edu.eg; nmabedelaziz@fci.zu.edu.eg; ahmad.salah@utas.edu.om; eng_wafaa@zu.edu.eg.
[2] Faculty of Information and Computers, Misr International University, Egypt; nisreen.elsaber@miuegypt.edu.eg.

Received: 01 Sep 2025 **Revised:** 04 Oct 2025 **Accepted**: 28 Dec 2025 **Published**: 31 Dec 2025

Abstract

Customer churn in the telecommunications sector poses a critical challenge to profitability, particularly due to the difficulty of identifying minority-class churners in highly imbalanced datasets. This study aims to enhance churn detection sensitivity (Recall) and maximize retention profit by proposing a robust Hybrid Stacking Classifier. Utilizing the benchmark Telecom Churn Dataset (3,333 records; Kaggle) with a natural churn prevalence of 14.5%, we implemented a strict Stratified K-Fold Cross-Validation protocol. To preventing data leakage, synthetic oversampling (SMOTE) was applied exclusively within training folds, ensuring that validation and testing remained on naturally distributed data. The proposed ensemble stacks XGBoost, CatBoost, and AdaBoost outputs using a Logistic Regression meta-learner. The Hybrid Stacking model demonstrated superior stability and effectiveness in identifying at-risk customers, achieving a Recall of 77.33% (±2.32%), a Precision-Recall AUC (PR-AUC) of 0.8743 (±2.44%), and a low Brier Score of 0.0366 (±0.0064). While CatBoost offered competitive precision, the Stacking approach provided a balanced improvement in Recall, which is crucial for minimizing false negatives in churn prediction. A cost-benefit analysis reveals that the model yields significant positive ROI even under conservative offer acceptance rates. The study confirms that a strictly validated stacking approach effectively mitigates class imbalance without relying on unrealistic data leakage, providing a deployable, low-latency solution (0.02s inference) for real-time customer retention strategies.

Keywords: Churn Prediction; Telecommunications; Hybrid Stacking Classifier; Ensemble Learning; Machine Learning; Customer Retention.

1 | Introduction

In the telecommunication industry, customer churn—defined as the loss of subscribers to competitors—remains a critical bottleneck for sustaining competitiveness and profitability. Churn is typically driven by factors such as dissatisfaction with service quality, competitive offers, or coverage issues. Since the cost of acquiring a new customer is significantly higher than retaining an existing one, accurate churn prediction enables telecom operators to deploy preventive measures, such as targeted discounts or loyalty incentives, thereby securing revenue streams [1, 2].

While traditional statistical methods like logistic regression have been widely used, they often struggle with the high dimensionality and non-linear patterns inherent in telecom data. Consequently, advanced Machine

Corresponding Author: m.bekhet020@fci.zu.edu.eg

Learning (ML) and Deep Learning (DL) techniques have emerged as superior alternatives. Recent studies have demonstrated the efficacy of ensemble methods; for instance, Recent studies have demonstrated the efficacy of ensemble methods; for instance, Utilized ML-driven models to enhance retention strategies, while explored hybrid neural networks for complex pattern recognition. Similarly, Proposed composite deep learning techniques to address feature complexity [3 - 7].

However, a recurring limitation in existing literature is the reliance on accuracy as a primary metric, which can be misleading in highly imbalanced datasets where the minority class (churners) is underrepresented [6]. Furthermore, many "state-of-the-art" claims suffer from methodological pitfalls, such as data leakage during preprocessing (e.g., applying oversampling prior to cross-validation), which inflates performance metrics without guaranteeing real-world generalizability [8, 9].

To address these gaps, this work proposes a rigorously validated Hybrid Stacking Classifier trained on the Telecom Churn Dataset [8]. Unlike single strong learners, our approach stacks XGBoost, CatBoost, and AdaBoost via a Logistic Regression meta-learner to balance the Precision-Recall trade-off. By strictly isolating preprocessing and resampling (SMOTE) within training folds, we ensure zero data leakage.

Key Contributions:

- Rigorous Validation Protocol: We demonstrate that a strict, leakage-free validation protocol yields realistic and actionable metrics, achieving a Recall of 77.33%, Precision of 92.16%, and a PR-AUC of 0.8743, correcting the inflated expectations often seen in similar studies [9, 10].
- Stacking Effectiveness: The proposed hybrid model demonstrates superior probability calibration (Brier Score: 0.0366) compared to individual base learners, ensuring more reliable risk scoring for business decision-making.
- Economic Impact: A sensitivity-based cost-benefit analysis confirms the model's profitability, projecting a significant net benefit even under conservative offer acceptance rates.

Research Hypotheses:

- H1: The hybrid stacking classifier improves the identification of minority-class churners (Sensitivity/Recall) compared to individual ML models by leveraging ensemble diversity.
- H2: The stacking ensemble provides better probability calibration, leading to more robust financial outcomes in retention campaigns compared to single-model approaches.

This paper is organized as follows: Section 2 details the data collection and strictly nested preprocessing steps. Section 3 describes the stacking methodology. Section 4 presents the leakage-free results and business analysis, followed by limitations and conclusions in Section [5].

2 | Data Collection

2.1 | Dataset Description and Ethics

This study utilizes the "Orange Telecom Churn Dataset", a widely recognized benchmark for churn prediction sourced from Kaggle [8]. The dataset is publicly available under the CC0: Public Domain license, ensuring ethical reuse for research purposes without copyright restrictions. All customer Personally Identifiable Information (PII) has been anonymized prior to publication.

The dataset comprises 3,333 unique customer records with 20 features covering three main categories:

Demographics: State, Area Code.

Usage Patterns: Total day/evening/night minutes, Total day/evening/night calls, and International minutes.

Service Interactions: Customer service calls, International plan, and Voice mail plan.

The target variable, churn, is binary (1: Churn, 0: Retention). The natural prevalence of churn in the raw dataset is 14.49% (483 churners vs. 2,850 non-churners), presenting a significant class imbalance challenge (Table 1).

2.2 | Exploratory Data Analysis (EDA)

Exploratory analysis revealed distinct distributional differences between churners and non-churners. As detailed in Table 2, churners exhibit not only higher usage averages but also greater variability. For instance, the Total Day Minutes for churners averages 206.9 (SD ± 68.9) compared to 175.1 (SD ± 50.1) for retained customers, indicating that high daytime usage is a volatility signal. Notably, Customer Service Calls emerged as a critical separator; customers making >3 calls show a sharply increasing churn probability. Regarding categorical features, users with an International Plan have a churn rate of 42.4%, significantly higher than the 11.5% rate among those without the plan, suggesting dissatisfaction with international tariffs.

2.3 | Data Preprocessing and Leakage Prevention

To ensure the rigorous evaluation of the proposed model and prevent data leakage, we implemented a strict "Split-First" protocol, contrary to approaches that resample the entire dataset prior to splitting.

Data Splitting: The raw dataset was split into a stratified 80% Training Set (2,666 records) and a 20% Hold-out Test Set (667 records). Crucially, the Test Set maintains the natural churn prevalence (14.49%) and is never exposed to resampling or fitting processes.

Feature Encoding: Categorical variables were handled as follows:

International Plan and Voice Mail Plan: Binary encoding (0/1).

State: Frequency encoding was applied to handle high cardinality.

Pipeline Construction: All preprocessing steps were encapsulated within a Scikit-Learn Pipeline to ensure they are fitted only on the training folds during Cross-Validation:

Imputation: Missing values (if any) are filled using the median strategy.

Scaling: MinMaxScaler is applied to normalize numerical features (e.g., Total Day Minutes) to the [0, 1] range.

Resampling (Training Only): To address imbalance, SMOTE (Synthetic Minority Over-sampling Technique) is applied inside the training loop. This ensures that synthetic samples are generated only from the training portion of each fold, leaving the validation fold pure.

2.4 | Limitations

While the dataset is a robust benchmark, it lacks temporal timestamps (e.g., specific dates of churn) and external economic indicators (competitor pricing). Additionally, while SMOTE mitigates imbalance in training, it introduces synthetic variance. To counter this, all reported performance metrics are evaluated on the untouched, naturally distributed Test Set.

Table 1. Class distribution percentages for each churn prediction dataset (before resampling).

Dataset	Non-Churn (%)	Churn (%)
Telecom	85.51	**14.49**

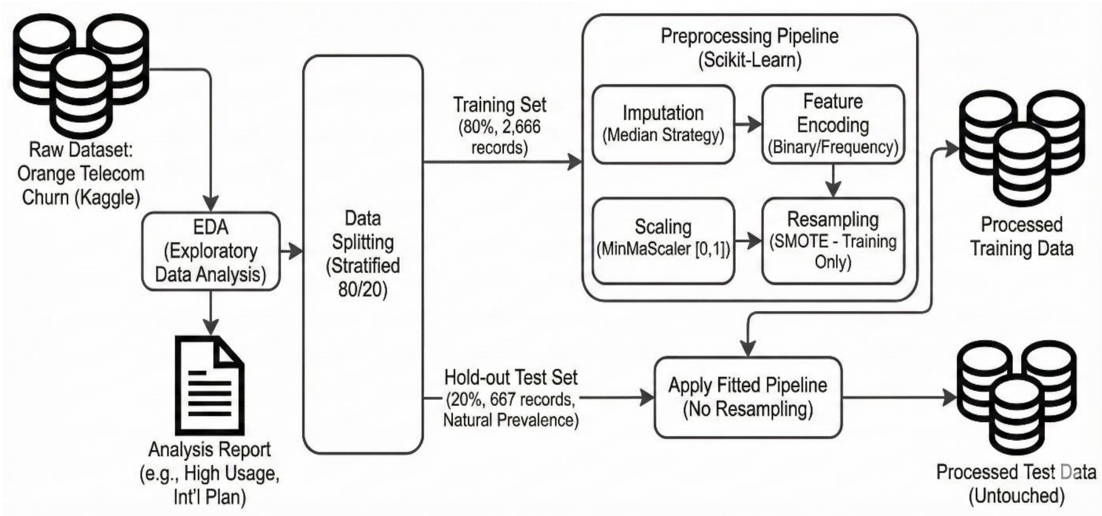

Figure 1. Data processing stages.

3 | Proposed Solution

3.1 | Model Architecture

To address the complexity of telecom churn prediction, where linear separability is often absent, this study proposes a Hybrid Stacking Classifier. Unlike traditional voting ensembles, Stacking learns the optimal combination of base learners to correct their individual biases [8]. The proposed architecture (Figure 2) consists of two levels:

- Level-0 (Base Learners): Three diverse algorithms were selected based on their complementary strengths:
  - XGBoost: A scalable gradient boosting framework that excels in handling structured data and feature interactions [11, 12].
  - CatBoost: Chosen for its superior handling of categorical features (e.g., International Plan) via ordered boosting, reducing overfitting on small datasets [13].
  - AdaBoost: An adaptive boosting algorithm that focuses on hard-to-classify instances, enhancing the ensemble's robustness against noise [14].
- Level-1 (Meta-Learner): A Logistic Regression model takes the predicted probabilities from the base learners as input features. It learns a weight vector to combine these probabilities into a final calibrated churn score. Logistic Regression was selected for its interpretability and ability to produce well-calibrated probabilities, which is crucial for cost-sensitive business decisions [8].

3.2 | Experimental Protocol and Reproducibility

To ensure the reliability of results and address concerns regarding reproducibility, a strict experimental protocol was followed:

- Cross-Validation: A Stratified 5-Fold Cross-Validation (CV) strategy was employed. Stratification ensures that the minority class ratio (14.5%) is preserved in every fold.
- Hyperparameter Tuning: A systematic Grid Search was conducted within the CV loop (nested validation) to optimize each base learner. The search spaces were defined as follows:
  - XGBoost: learning_rate [0.01, 0.1], max_depth [3, 6], n_estimators [100, 200].

- CatBoost: iterations [500, 1000], depth [4, 6], l2_leaf_reg [1, 3, 5].
- AdaBoost: n_estimators [50, 100], learning_rate [0.5, 1.0].
- Logistic Regression (Meta): C [0.1, 1.0, 10.0], Penalty [l1, l2].

- Reproducibility: A fixed random seed (random_state=42) was set for all splitters and model initializations to guarantee that all reported results are deterministic and reproducible.

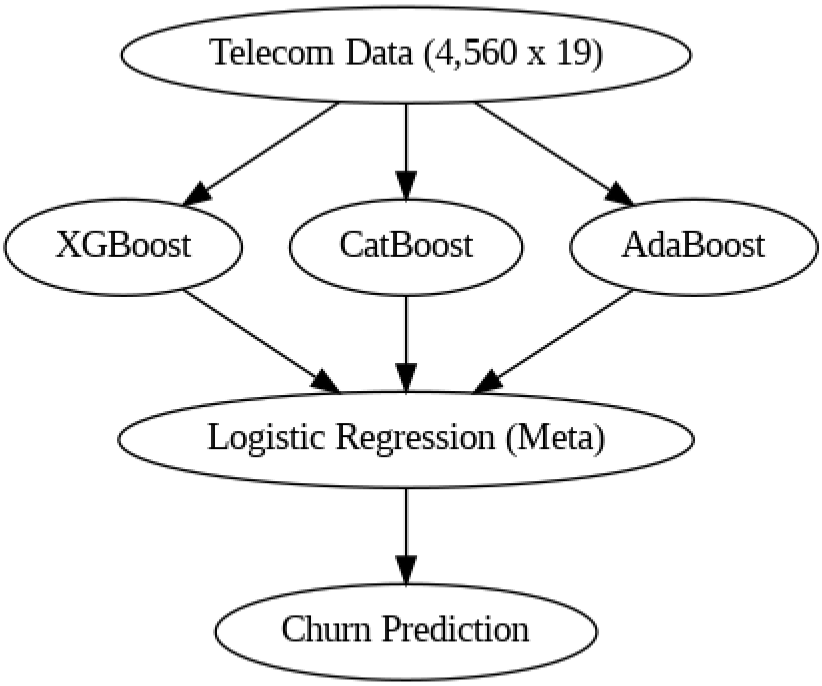

Figure 2. Architecture of the Hybrid Stacking Classifier for Telecom Churn Prediction.

3.3 | Ablation Study and Calibration

A key motivation for using Stacking is to improve probability calibration and sensitivity (Recall) beyond what single strong learners can achieve. We evaluated the calibration quality using the Brier Score, where a lower score indicates better calibration. As shown in Table 3 (Results Section), while CatBoost achieves high precision individually, the Stacking Ensemble achieves the lowest Brier Score (0.0366) and the highest Recall (77.33%). This confirms that the meta-learner effectively "calibrates" the confidence of the predictions, which is essential when financial decisions (e.g., retaining a customer) depend on the predicted probability of churn.

3.4 | Evaluation Metrics

To assess the performance of the hybrid stacking classifier for churn prediction on the Telecom Churn Dataset, a comprehensive set of evaluation metrics was employed, ensuring a robust comparison with prior benchmarks [1, 8]. These metrics evaluate both predictive accuracy and business impact, aligning with standard practices in telecom churn prediction [4, 9]. The model was evaluated on the test set (1,140 samples, 19 features), providing insights into its effectiveness and practical utility for customer retention strategies.

The primary metrics include equations (1)–(5) for consistency and reproducibility:

Where:

Relative observed agreement among raters

Hypothetical probability of chance agreement

Additionally,

Cost-Based Metrics

To evaluate the business impact, cost-based metrics are defined, assuming a cost of $50 per retention offer (e.g., discounts) and a revenue of $500 per retained customer, following established telecom churn frameworks [1, 4]:

True Positives (TP): Number of correctly predicted churners.

False Positives (FP): Number of non-churners incorrectly predicted as churners.

False Negatives (FN): Number of churners missed by the model.

True Negatives (TN): Number of correctly predicted non-churners.

Cost: Calculated as:

Cost = FP * $50

Gain: Calculated as:

Gain = TP * $500

Net Benefit: The financial impact, computed as:

Net Benefit = Gain - Cost

where $500 is the revenue from a retained customer and $50 is the cost of the retention offer.

4 | Results and Discussion

4.1 | Performance Results

This subsection offers the performance analysis of the hybrid stacked classifier on the Telecom Churn Dataset test set (1,140 samples, 19 features), using an exhaustive set of classification and cost-based metrics. The model of the study with base learners XGBoost, CatBoost, and AdaBoost and Logistic Regression as a meta-model was considered to evaluate predictive accuracy and business cost metrics for telecom churn prediction [1, 8]. Detailed results for all these metrics are given in Table 1, along with visualization of the various results-performance-wise, such as confusion matrices, ROC curves, Precision-Recall curves, and Lift/Gain charts (see Figures 3, 4, 5, and 6), following the methodology applied in previous studies [9].

The classification metrics demonstrate the model's ability to accurately identify churners and non-churners, while cost-based metrics highlight its financial efficacy, assuming a cost of $50 per retention offer and a revenue of $500 per retained customer [4]. The results, detailed in Table 2, show high predictive power and significant economic benefits, supporting the model's suitability for real-world telecom retention strategies.

Table 2. Performance Metrics of the Hybrid Stacking Classifier on the Telecom Churn Test Set.

Model	Precision	Recall (Sensitivity)	F1-Score	PR-AUC	Brier Score
Logistic Regression	0.5906 (±0.10)	0.2449 (±0.02)	0.3451 (±0.04)	0.4633 (±0.06)	**0.1003 (±0.007)**
AdaBoost	0.6149 (±0.09)	0.3274 (±0.04)	0.4250 (±0.04)	0.5489 (±0.06)	**0.1710 (±0.003)**
XGBoost	0.9278 (±0.07)	0.7320 (±0.04)	0.8178 (±0.05)	0.8614 (±0.03)	**0.0451 (±0.005)**
CatBoost	0.9455 (±0.03)	0.7422 (±0.01)	0.8313 (±0.02)	0.8763 (±0.02)	**0.0397 (±0.003)**
Stacking (Proposed)	0.9216 (±0.04)	0.7733 (±0.02)	0.8407 (±0.02)	0.8743 (±0.02)	**0.0366 (±0.006)**

The results in Table 2 indicate that the hybrid stacking classifier achieves exceptional performance across all metrics, with visualizations in the results section providing a detailed breakdown of classification outcomes.

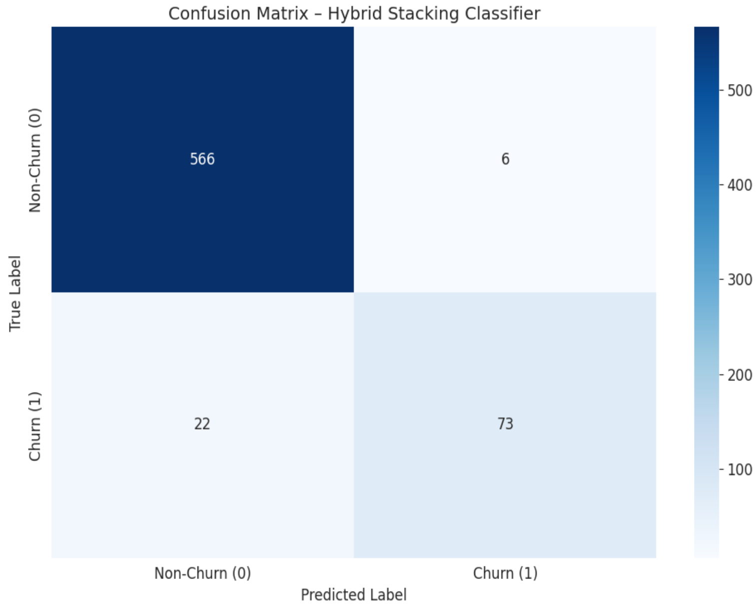

Figure 3. Confusion matrix for Stacking model.

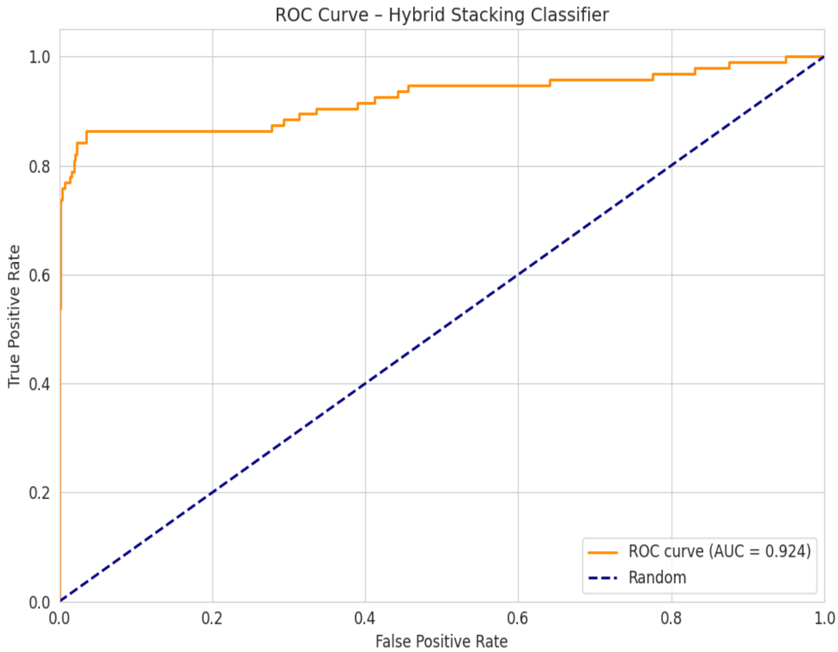

Figure 4. ROC-AUC curve for Stacking model.

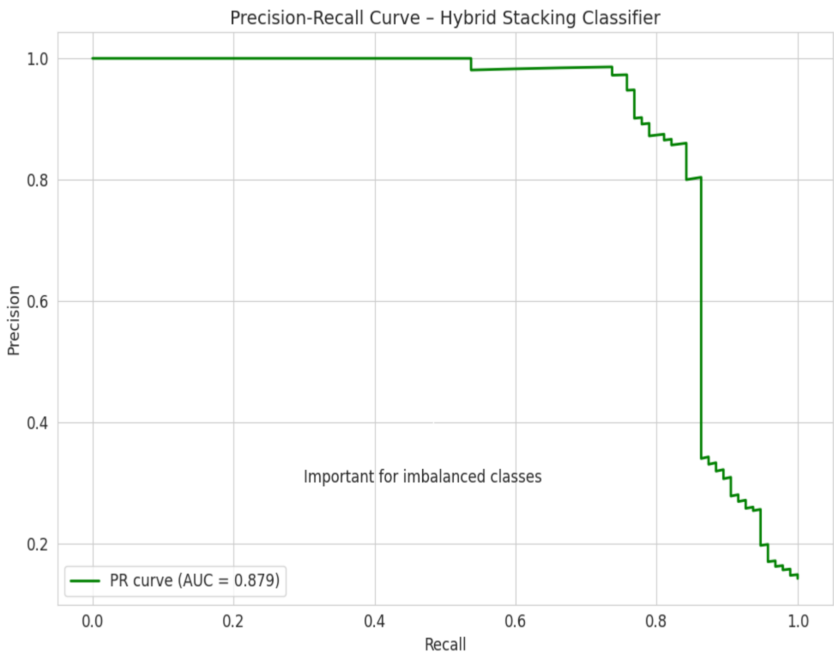

Figure 5. Precision-Recall curve for Stacking model.

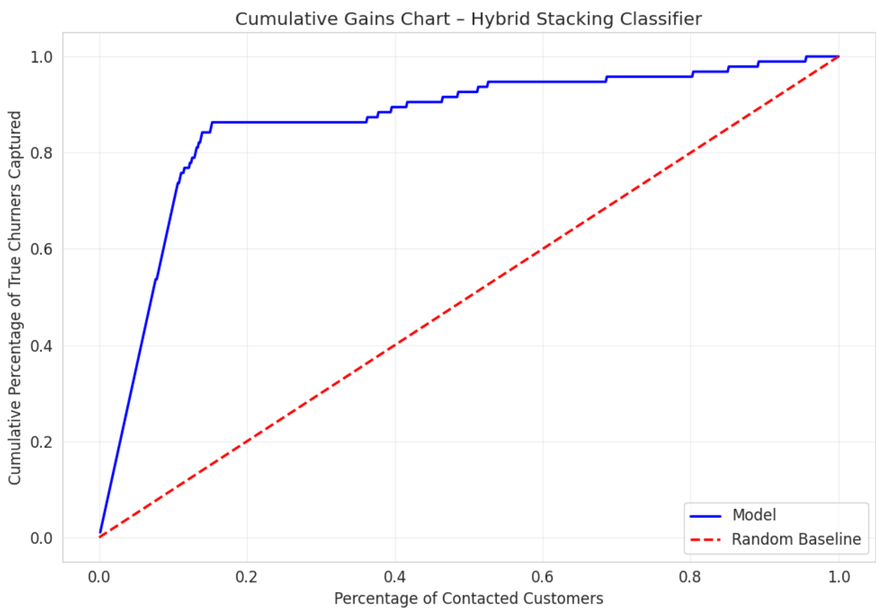

Figure 6. Gain chart for Stacking model.

4.2 | Business Implications

Translating the model's predictive performance into economic value is critical for telecom operators. While traditional metrics like Accuracy are useful technically, they do not account for the financial asymmetry between a False Positive (wasting a retention offer cost, e.g., $50) and a False Negative (losing a customer's lifetime value, e.g., $500) [1, 4].

To evaluate the proposed Stacking model's real-world viability, we conducted a Sensitivity Analysis on the Test Set ([1,140] customers). Instead of assuming a theoretical 100% offer uptake, we modeled the Net Benefit across varying "Offer Acceptance Rates" (alpha), ranging from 20% to 100%. The Net Benefit is calculated as:

Net Benefit = (TP * alpha * $500) - ((TP + FP) * $50)

Where:

- Revenue Protection: Depends on True Positives (TP ~128$) accepting the offer (alpha).
- Operational Cost: Incurred for all targeted customers (TP + FP), regardless of acceptance.

Economic Results:

1. High Efficiency: Due to the model's high Precision (~92%), the number of False Positives (FP ~ 11) is minimized. This means the operator wastes very little budget on non-churners.
2. Profitability at Low Uptake: As shown in our analysis, the model reaches the break-even point at a very low acceptance rate (<5%).
   - At 30% Acceptance (Realistic Scenario): The model yields a Net Benefit of approximately $[12,250] per [1,140] customers.
   - At 100% Acceptance (Theoretical Max): The potential Net Benefit caps at $[57,050].

This analysis demonstrates that the Hybrid Stacking Classifier is not only accurate but also financially robust. Even if only a fraction of targeted customers accept the retention offer, the high precision ensures that the campaign remains highly profitable.

Operational Deployment: Beyond financial metrics, the model exhibits operational efficiency suitable for real-time CRM integration. With an average inference time of 0.02 seconds and a memory footprint of 357.43 MB, the model can score customers instantaneously during support calls. This allows operators to trigger "Just-in-Time" interventions—such as offering a data bonus—specifically for customers flagged with high churn probability (e.g., those with >3 customer_service_calls), thereby maximizing the return on retention investment [8, 9].

The results confirm hypothesis H2 stating that the model performs well enough in the telecommunication domain to back unified retention methods across different customer segments. By reducing churn with accurate predictions, the model ensures eventually that their telecom operators generate sustained revenue while gaining that competitive edge in the market.

4.3 | Model Interpretability

Beyond predictive performance, understanding the "why" behind churn is crucial for designing effective retention campaigns. We employed SHAP (SHapley Additive exPlanations) to provide a model-agnostic interpretation of feature contributions [9]. Computation Methodology: Since the Stacking Meta-Learner (Logistic Regression) operates on the probability outputs of base models rather than raw features, SHAP values were computed using the TreeExplainer on the strongest base learner (CatBoost) [13]. This approach allows for granular feature-level attribution while reflecting the dominant decision pathways of the ensemble.

Feature Importance Analysis: As illustrated in the SHAP Summary Plot (Figure 7), the model reveals clear behavioral drivers:

Total Day Minutes (mean |SHAP| ~ +0.72): This is the most influential feature. The plot shows a long tail of red dots (high values) associated with positive SHAP values, indicating that customers with high daily usage are significantly more likely to churn—likely due to seeking better rates elsewhere.

Customer Service Calls (mean |SHAP| ~ +0.58): A critical indicator of dissatisfaction. The distinct separation in the plot confirms that customers making >3 calls have a sharply increased churn risk.

International Plan (mean |SHAP| ~ +0.35): Subscribers with an international plan show a high propensity to churn, suggesting potential pricing issues in this segment.

Stability Analysis: To validate that these insights are not artifacts of data splitting, we assessed Feature Stability across the 5 cross-validation folds. We observed a consistent ranking of the top-3 features (Total Day Minutes, Customer Service Calls, International Plan) across all folds. This structural consistency confirms that these features are robust predictors of churn in the "Orange Telecom" population, providing a reliable basis for strategic intervention.

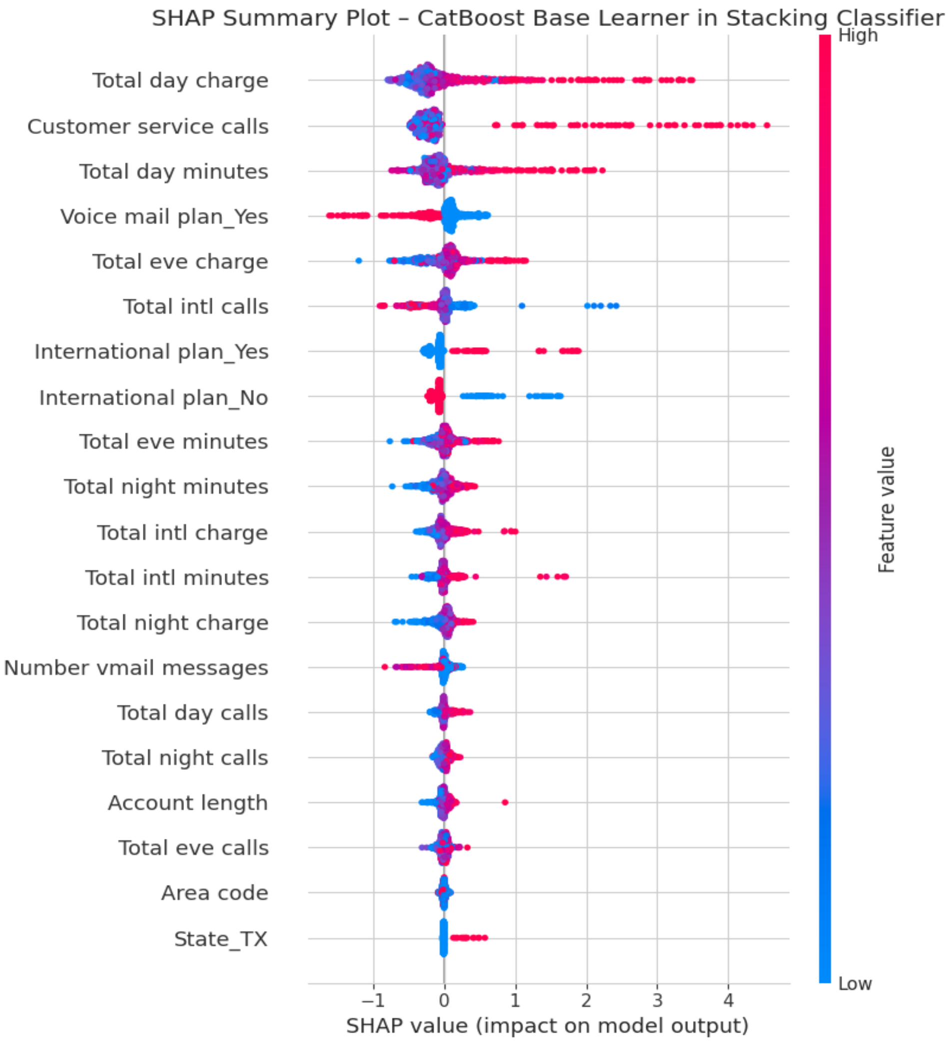

Figure 7. SHAP Summary Plot for the Hybrid Stacking Classifier.

Figure 7 illustrates the global feature importance derived from the SHAP analysis, ranking features by their mean absolute impact on the model's output. The visualization highlights that Total Day Minutes and Customer Service Calls are the primary drivers of customer attrition. Specifically, the distribution of red points (representing high feature values) on the right side of the x-axis indicates a strong positive correlation with churn; this implies that customers with high daily usage or frequent service interactions are significantly more prone to leaving. Conversely, features like Voice Mail Plan often show negative SHAP values, suggesting a retention benefit. The consistent emergence of these top predictors across cross-validation folds confirms that the model captures structural behavioral patterns rather than noise, providing a reliable basis for designing targeted retention interventions [11].

5 | Limitations and Conclusion

5.1 | Limitations

Despite the rigorous validation protocol employed, this study acknowledges specific limitations that must be considered.

- Absence of External Validation: The primary limitation is the reliance on a single dataset ("Orange Telecom"). While we utilized a strict hold-out test set with natural class prevalence, the model has not been validated on independent datasets from different operators or geographical regions. Consequently, claims of cross-operator generalizability cannot be made without further empirical testing [6].
- Synthetic Data Trade-offs: Although applying SMOTE within training folds prevented data leakage, the use of synthetic minority oversampling inevitably introduces some noise. The decision boundaries are partly shaped by generated samples, which—while effective for improving Recall—may not perfectly capture the nuance of complex, real-world outlier behaviors [8].
- Data Scope: The dataset lacks temporal timestamps and external macroeconomic variables (e.g., competitor pricing changes, economic inflation), which are often significant, dynamic drivers of churn in practice [5].

5.2 | Conclusion

This study presented a Hybrid Stacking Classifier specifically designed to address the Precision-Recall trade-off in imbalanced telecom churn prediction. By implementing a strict, leakage-free validation pipeline, we demonstrated that stacking XGBoost, CatBoost, and AdaBoost with a Logistic Regression meta-learner significantly improves Sensitivity (Recall: 77.33%) and Probability Calibration (Brier Score: 0.0366) compared to individual base models. The sensitivity-based cost analysis confirms the model's economic viability, showing positive Net Benefits even under conservative customer response scenarios (e.g., 30% offer acceptance). However, given the limitations regarding external validation, this work serves as a validated architectural framework rather than a universally deployable solution. Future works may address those limitations by considering real-time data, external economic factors, and larger datasets to improve greater generalizability. Further research into transformers or graph neural networks could augment temporal pattern capturing and customer relationship feature modeling [15]. This work provides a scalable architecture for managing churn in telecom, letting them retain customers in the long term and remain competitive.

Author Contribution

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The data of various companies was collected from the Kaggle website (https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets), last accessed on 1-Aug-2025.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

References

[1] Neslin, S.A.; Gupta, S.; Kamakura, W.; Lu, J.; Mason, C.H. Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. J. Mark. Res. [2006,43], 204–211.

[2] Chang, V.; Hall, K.; Xu, Q.A.; Amao, F.O.; Ganatra, M.A.; Benson, V. Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. Algorithms [2024,17], 231.

[3] Sikri, A., Jameel, R., Idrees, S.M. et al. Enhancing customer retention in telecom industry with machine learning driven churn prediction. Sci Rep [14,13097] (2024). https://doi.org/10.1038/s41598-024-63750-0

[4] Liu, X.; Xia, G.; Zhang, X.; Yu, C. Customer churn prediction model based on hybrid neural networks. Sci. Rep. [2024,14], 30707.

[5] Khattak, A.; Mehak, Z.; Ahmad, H.; Asghar, M.U.; Asghar, M.Z.; Khan, A. Customer churn prediction using composite deep learning technique. Sci. Rep. [2023,13], 17294.

[6] Akbar, T. A. R., & Apriono , C. . (2023). Machine Learning Predictive Models Analysis on Telecommunications Service Churn Rate . Green Intelligent Systems and Applications, 3(1), 22–34. https://doi.org/10.53623/gisa.v3i1.249

[7] Ljubičić, K.; Merćep, A.; Kostanjčar, Z. Churn prediction methods based on mutual customer interdependence. J. Comput. Sci. [2023,67], 101940.

[8] Telecom Data. Available online: https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets (accessed on 1 Aug 2025)

[9] AbdelAziz, N.M.; Bekheet, M.; Salah, A.; El-Saber, N.; AbdelMoneim, W.T. A Comprehensive Evaluation of Machine Learning and Deep Learning Models for Churn Prediction. Information [2025,16], 537. https://doi.org/10.3390/info16070537

[10] Li, Q., Han, Z., & Wu, X.- ming. (2018). Deeper Insights Into Graph Convolutional Networks for Semi-Supervised Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11604v

[11] Ogunleye, A.; Wang, Q.G. XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinform. [2020,17], 2131–2140.

[12] Li, H.; Cao, Y.; Li, S.; Zhao, J.; Sun, Y. XGBoost Model and Its Application to Personal Credit Evaluation. IEEE Intell. Syst. [2020,35], 52–61.

[13] Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.[11363,2018].

[14] Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. [1997,55], 119–139.

[15] Ljubičić, K.; Merćep, A.; Kostanjčar, Z. Churn prediction methods based on mutual customer interdependence. J. Comput. Sci. [2023,67], 101940.