American Journal of Networks and Communications
2026; 15(1): 10-26
http://www.sciencepublishinggroup.com/j/ajnc
doi: 10.11648/j.ajnc.20261501.12

Predicting Customer Churn in the Telecommunications Industry using Machine Learning Techniques

Adeline Makokha*, Kevin Obote, Henry Muchiri, Kennedy Senagi

Strathmore Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya

Email address:
adeline.makokha@strathmore.edu (Adeline Makokha), kevin.obote@strathmore.edu (Kevin Obote),
hmuchiri@strathmore.edu (Henry Muchiri), ksenagi@strathmore.edu (Kennedy Senagi)
*Corresponding author

To cite this article:
Adeline Makokha, Kevin Obote, Henry Muchiri, Kennedy Senagi. (2026). Predicting Customer Churn in the Telecommunications Industry using Machine Learning Techniques. *American Journal of Networks and Communications, 15*(1), 10-26.
https://doi.org/10.11648/j.ajnc.20261501.12

Received: 11 February 2026 ; **Accepted:** 8 April 2026 ; **Published:** 24 April 2026

Abstract: Voluntary customer churn constitutes a persistent financial risk for telecommunications operators, particularly within enterprise customer segments where high-value accounts administer complex, multi-subscription portfolios. Industry data indicate that acquiring a new account costs between five and seven times more than retaining an existing one. Despite heightened industry awareness, the majority of operational retention platforms remain reactive, detecting departure only after the event has occurred. This investigation constructs and evaluates a machine learning pipeline engineered to identify enterprise customer churn risk proactively, drawing on authentic operational records extracted from a business-tobusiness telecommunications environment. The study follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) lifecycle. A dataset of 8,454 unique business accounts, characterised by 14 raw attributes and enriched to a final 22-variable feature set, underpins the empirical work. Pronounced class imbalance, churned accounts representing approximately 6.5minority ratio of 14.3:1, necessitated specialised resampling prior to classifier training. Five oversampling strategies were benchmarked; SVMSMOTE produced the largest gain in minority-class sensitivity and was adopted for all subsequent training cycles. Ten classifier families were trained and assessed, including EasyEnsembleClassifier, RUSBoostClassifier, XGBoost, LightGBM, CatBoost, Histogram Gradient Boosting, Balanced Bagging, a multilayer perceptron, a soft-voting ensemble, and a stacking ensemble. EasyEnsembleClassifier emerged as the leading model, attaining an F1-score of 0.129 and a recall of 38.242 of 110 churned accounts. Post-hoc explainability analysis through SHAP and LIME identified active subscriber rate, geographic billing zone, and engineered interaction terms as the dominant predictive signals. The framework was operationalised within a FastAPI-based application supporting realtime individual scoring, batch CSV prediction, and retention campaign monitoring. The projected annual revenue protection under conservative assumptions exceeds 74,000 currency units. The study illustrates that interpretable, explainability-augmented machine learning frameworks can bridge the gap between quantitative model output and managerial action, offering a replicable blueprint for data-driven churn governance in both emerging and mature telecommunications markets.

Keywords: Customer Churn Prediction, Telecommunications, Ensemble Learning, EasyEnsembleClassifier, SVMSMOTE, Class Imbalance, Feature Engineering, Model Explainability, CRISP-DM, FastAPI Dashboard

1. Introduction

Within the global telecommunications landscape, customer churn describes the voluntary termination of subscription-based services by users who migrate to competing providers. This phenomenon constitutes one of the most enduring strategic challenges facing telecommunications operators, with direct consequences for profitability, operational efficiency, and long-term market positioning. Elevated churn erodes revenue stability and amplifies operational expenditures, particularly in saturated, intensely competitive environments.

Contemporary estimates suggest that annual churn in the global telecommunications sector exceeds 30 % in several markets [8], underscoring the urgency of data-driven prediction and proactive retention mechanisms.

The financial implications of subscriber loss are substantial, compelling operators to pursue strategies that identify and retain customers ahead of service termination. Mobile communication underpins global connectivity, with smartphones and feature phones accounting for the majority of voice and data consumption worldwide. Although market penetration has reached saturation in several regions, the mobile segment remains among the most dynamic areas of the telecommunications sector. As competitive pressure intensifies, strategic priorities have shifted from subscriber acquisition toward customer retention. Churn is formally defined in this study as the binary indicator of voluntary service discontinuation where a customer account is labelled as churned if the account holder proactively terminated all active subscriptions during a six-month observation window ending at the data extraction date. Accounts that expired passively due to non-payment or administrative closure are excluded from the positive class. This operational definition aligns with contractual churn conventions in enterprise B2B segments and ensures that the model targets discretionary attrition amenable to retention intervention. Churn has consequently emerged as a central performance indicator for telecommunications firms [8]. Evidence indicates that annual subscriber losses ranging from 30 to 35 % have been observed in several markets during periods of economic disruption [8].

From an economic standpoint, customer retention is consistently demonstrated to be more cost-effective than new customer acquisition. Empirical studies indicate that acquiring a new subscriber may cost between five and seven times more than maintaining an existing one [6]. Elevated attrition destabilises revenue streams and constrains an organisation's ability to remain competitive in a crowded marketplace. The rapid technological advancement of the telecommunications sector has lowered entry barriers, enabling new entrants to introduce innovative service offerings and intensify competitive rivalry. In such environments, churn is driven by market saturation, aggressive promotional strategies, and the frequent introduction of bundled services. Telecommunications operators must therefore balance growth objectives with sustainable retention strategies, positioning churn prediction as a core component of modern operational and strategic planning.

Customer Churn Prediction (CCP) has emerged as a critical analytical tool for anticipating subscriber attrition and directing retention initiatives [6]. Effective prediction enables organisations to design timely interventions that enhance revenue performance and strengthen competitive positioning. Telecommunications firms generate extensive volumes of customer data, encompassing service usage records, billing histories, demographic profiles, and interaction logs. These datasets offer valuable insight into behavioural patterns associated with churn risk. The principal challenge resides not in data availability but in the capacity to transform records into actionable intelligence identifying disengagement signals before account termination.

Globally, telecommunications and other service-oriented industries continue to experience persistent churn as digitalisation accelerates and customer switching costs decline. While the sector generates significant revenue, particularly in developing economies, operators face increasing difficulty in retaining subscribers amid rapid technological change and evolving consumer expectations. International research has demonstrated the effectiveness of machine learning in churn prediction, employing approaches such as ensemble learning, feature selection, and gradient boosting to improve predictive accuracy [5]. However, emerging literature emphasises the need to move beyond purely predictive objectives and incorporate explanatory insights that clarify the underlying drivers of churn [9].

Studies such as [9] illustrate that models including Logistic Regression, Random Forests, LightGBM, and ensemble frameworks can successfully identify high-risk customers across diverse datasets. Nevertheless, in regions where telecommunications services play a critical role in financial inclusion, education, and everyday communication, predictive performance must be complemented by interpretability to ensure that retention strategies align with local behavioural and market dynamics.

At the national level, research within Kenya highlights the importance of customer experience management in mitigating churn [10]. Based on survey data collected from telecommunications employees, social interactions, quality of service interfaces, and the retail environment contribute positively to customer retention, while pricing pressures increase churn propensity [10]. These findings suggest that customer loyalty is influenced by technical service quality, relational, and experiential factors.

The novel methodological contribution of this study is threefold. First, it integrates a comprehensive ensemble of ten diverse classifier families with boundary-focused synthetic oversampling (SVMSMOTE) in a single, reproducible pipeline, an approach not evaluated previously on enterprise B2B telecommunications data. Second, it introduces domain-informed engineered features, including revenue-efficiency ratios, subscriber engagement rates, and account risk scores, that capture latent churn signals absent from raw transactional attributes. Third, the predictive system is operationalised within a production-grade FastAPI deployment architecture incorporating real-time risk scoring, interactive explainability dashboards, and stratified cross-validation stability reporting, thereby bridging the methodological gap between model development and practical retention management.

The rapid pace of digital adoption, combined with socio-economic pressures affecting mobile service consumption, has further intensified competition within the telecommunications market. The integration of demographic attributes, revenue indicators, and service usage patterns offers operators an opportunity to implement informed retention interventions. By leveraging data-driven approaches, firms can reduce unnecessary marketing expenditure, strengthen customer loyalty, and enhance long-term profitability. Consequently, effective churn prediction is not merely a technical exercise but a strategic instrument for sustaining competitiveness in an increasingly volatile telecommunications landscape.

Research Objectives

The aim of this study is to address existing limitations in customer churn prediction within the telecommunications sector by developing a data-driven and decision-oriented analytical framework. The research seeks to enhance both the predictive accuracy and practical usability of churn prediction systems, ensuring alignment between analytical outputs and managerial decision-making needs.

Main Objective: To design and implement a robust machine learning framework for predicting customer churn in the telecommunications industry.

Specific Objectives:

1. To critically analyse existing methodological and conceptual limitations in churn prediction studies within the telecommunications sector.
2. To construct and train customer churn prediction models using selected machine learning algorithms and telecommunications customer data.
3. To assess the predictive performance of developed models through established classification evaluation metrics.
4. To operationalise the most effective churn prediction model by integrating it into an interactive and interpretable dashboard that supports informed managerial decision-making.

This study contributes to both theoretical and practical understanding of customer churn in telecommunications. By integrating machine learning techniques with business-oriented decision support mechanisms, the research advances a sustainable, data-informed approach to customer retention. The findings are intended to support the translation of advanced analytics into everyday operational practice, strengthening the role of predictive intelligence in enhancing customer loyalty and long-term organisational performance.

2. Literature Review

A substantial transformation has occurred within the digital landscape in recent years, characterised by rapid technological advancement and the emergence of new markets and service-oriented organisations. These developments have significantly reshaped customer expectations within the telecommunications sector, where demand increasingly favours affordable, reliable, and personalised service offerings. While digital evolution has broadened access to communication and enhanced service availability, it has simultaneously intensified competitive pressures. In this environment, retaining existing subscribers has become more strategically important than the continual acquisition of new customers.

A central challenge confronting telecommunications operators is the customer decision to discontinue service, often driven by pricing disparities, service quality perceptions, or competitive incentives. Evidence indicates that several operators have lost up to one-third of their subscriber base to rival networks, elevating customer churn to a critical global concern [8]. Such losses have directly translated into sustained revenue erosion and weakened market stability. Consequently, there is a growing imperative for the development and adoption of data-driven approaches capable of identifying churn risk early and supporting proactive retention strategies [5].

2.1. Theoretical Framework

2.1.1. Customer Lifecycle and Churn Dynamics in Enterprise Segments

Customer Relationship Management (CRM) has consistently been identified as a core strategic framework for sustaining customer loyalty and long-term organisational performance. [11] emphasised that while organisations continue to invest heavily in customer acquisition, retention has become equally critical as competitive markets intensify. Even marginal increases in churn rates can result in disproportionate revenue losses, underscoring the importance of early detection and intervention. Effective CRM depends on a data-informed understanding of customer behaviour, satisfaction, and the cognitive signals that precede disengagement or dissatisfaction.

The authors further noted that the effectiveness of machine learning-based churn prediction systems is largely determined by the quality of feature engineering. Traditional CRM-driven studies have predominantly relied on structured attributes such as demographic characteristics, service usage patterns, tenure, and billing information. Although these variables provide valuable quantitative indicators, they frequently fail to capture qualitative aspects of customer-provider interaction, including sentiment, tone, and expressed intent embedded in textualcommunications.

To overcome these limitations, [11] introduced the Customer Churn-related Knowledge Base (ChurnKB), a domain-informed framework designed to enhance feature representation through the incorporation of unstructured customer-generated content, including emails, chat transcripts, and feedback messages. This approach marked a methodological shift from purely numeric representations toward linguistically derived behavioural indicators, extending the predictive capacity of conventional CRM systems.

ChurnKB employs a suite of Natural Language Processing techniques, including Term Frequency-Inverse Document Frequency, cosine similarity, regular expressions, tokenisation, and stemming, to extract churn-relevant linguistic patterns. These features enable identification of early indicators of dissatisfaction, such as complaints, negative sentiment, or cancellation intent, which may not be evident in structured datasets alone. To further enhance adaptability, the framework integrates Generative AI models, particularly Large Language Models (LLMs), allowing the system to detect latent emotional and cognitive cues within unstructured text. Feedback mechanisms embedded in the framework support continuous learning, ensuring alignment with evolving customer behaviour.

Empirical evaluation demonstrated that incorporating knowledge-enhanced features led to substantial improvements in predictive performance across several machine learning models. The most pronounced improvement was observed in XGBoost, where the F1-score increased from 0.5752 to 0.7891, illustrating the effectiveness of knowledge-driven feature enrichment in identifying at-risk customers and enabling earlier intervention.

2.1.2. Predictive Modelling Using Statistical and Segmentation Approaches

The telecommunications industry is particularly vulnerable to customer churn, as even modest increases in churn can have immediate and significant effects on profitability. [22] emphasised that retaining existing subscribers is substantially more cost-effective than acquiring new ones, reinforcing the importance of predictive modelling for early churn detection. Their study focused on identifying behavioural and demographic characteristics that differentiate loyal customers from those likely to discontinue service.

The research utilized customer data from three major Chinese telecommunications operators, incorporating demographic, usage-based, and service-related variables. A key methodological component was customer segmentation, which enabled the identification of homogeneous groups exhibiting similar churn tendencies. This segmentation facilitated the development of targeted retention strategies tailored to the behavioural profiles of distinct customer groups.

Results indicated that logistic regression achieved superior predictive accuracy compared to Fisher discriminant analysis, attaining an accuracy of 93.94 %. The findings highlighted the robustness of logistic regression in handling categorical predictors and nonlinear relationships when combined with effective segmentation and variable selection. However, the reliance on structured numerical data and the absence of behavioural or textual variables revealed a methodological gap, suggesting opportunities for hybrid approaches that integrate traditional statistical modelling with machine learning techniques.

2.1.3. Forest Models and Ensemble Strategies

Ensemble learning approaches, particularly decision forest models, have increasingly become central to churn prediction in telecommunications. [6] argued that single rule-based classifiers, while interpretable, often lack the robustness and scalability required for large, high-dimensional customer datasets. Their study demonstrated that class imbalance, where churners represent a small minority, significantly limits the effectiveness of conventional models such as Logistic Regression or standalone Decision Trees.

To address these challenges, the authors proposed a suite of Decision-Forest architectures, including Logistic Model Trees, Random Forests, and Functional Trees. By aggregating multiple weak learners, these ensembles reduce variance and overfitting while exploiting complementary classifier strengths. The approach proved particularly effective in handling imbalanced datasets, improving recall for churners without disproportionately increasing false positives.

Using benchmark datasets, ensemble models employing weighted stacking and soft-voting outperformed baseline classifiers in both precision and recall. The findings confirmed that forest-based ensembles are well-suited for churn prediction problems characterized by noise, non-linearity, and class imbalance. Despite their strong performance, the authors emphasized the importance of integrating explainability mechanisms, recommending the incorporation of SHAP or LIME to enhance managerial trust and operational adoption.

2.1.4. Machine Learning Approaches for Churn Prediction

[20] examined the application of supervised machine learning techniques for churn prediction, comparing traditional decision tree models with Random Forests and Support Vector Machines. The study highlighted that earlier decision tree-based systems often suffered from redundant feature inclusion, leading to overfitting and limited predictive accuracy.

To address these issues, the authors proposed a hybrid approach combining Random Forests for feature selection with Support Vector Machines for classification. Random Forests reduced feature redundancy through automatic importance ranking, while SVMs enhanced class separation. The hybrid model achieved an accuracy of 95 %, representing a substantial improvement over traditional decision tree approaches. While effective, the study relied solely on structured data, leaving opportunities for future work incorporating real-time analytics, unstructured data, and explainable modelling techniques.

2.2. Empirical Studies

2.2.1. Feature Engineering Approaches

Advanced feature engineering has emerged as a cornerstone of effective churn prediction across telecommunications datasets. Beyond raw transactional variables, domain-informed composite features such as subscriber engagement ratios, revenue efficiency metrics, and service utilization patterns consistently improve model sensitivity to early-stage churn signals [11]. This finding aligns with the present study's approach of constructing behavioural and financial interaction features absent from the original dataset.

2.2.2. Explainable AI in Churn Analytics

Post-hoc explainability methods, including SHAP and LIME, have gained adoption in churn analytics as tools for translating black-box model outputs into actionable business insights [9]. Research demonstrates that explainability is not merely an ethical requirement but a practical prerequisite for managerial adoption, particularly in high-stakes retention contexts where account managers need clear justification for targeting decisions.

2.2.3. Handling Class Imbalance in Customer Churn Prediction

[5] investigated the effects of class imbalance on churn prediction using customer data from a major telecommunications operator in Nepal. The dataset comprised over 52,000 records, with churners constituting a minority. XGBoost was employed to address imbalance, achieving 97 % accuracy and an F1-score of 88 % on the native dataset. The study emphasised the importance of evaluating churn models using metrics beyond accuracy, such as recall, F1-score, and ROC-AUC, which better reflect performance on minority classes. While the results demonstrated strong predictive capability, the absence of explainability techniques and comparative resampling strategies highlighted areas for further methodological enhancement.

2.2.4. Hybrid Approaches to Churn Prediction

[21] proposed a hybrid framework combining classification and clustering to improve churn prediction and interpretability. Feature selection using Information Gain and Correlation Ranking reduced redundancy, while Random Forest classification achieved the highest accuracy among tested models. Subsequent k-means clustering enabled segmentation of churned customers into behaviourally similar groups, supporting targeted retention interventions.

Despite improved performance and insight generation, the framework lacked explainable AI components and did not account for temporal dynamics, indicating opportunities for future research incorporating time-aware and explainable modelling approaches.

2.2.5. Model Optimization and Deep Learning

[7] evaluated traditional machine learning, ensemble methods, and deep learning architectures across international telecommunications datasets. Deep learning models, particularly CNNs and ANNs, achieved near-perfect accuracy however, concerns regarding interpretability and class imbalance remained. The authors concluded that while deep models capture complex nonlinear relationships effectively, their practical adoption requires transparency and business-aligned evaluation.

2.2.6. Knowledge-Based and Text-Driven Feature Engineering

Recent advances by [11] further demonstrated the value of knowledge-based and text-driven feature engineering in churn prediction. By integrating linguistic and cognitive indicators extracted from unstructured text into machine learning models, the study achieved substantial improvements in predictive performance. The findings reinforce a growing shift toward models that balance accuracy with interpretability and contextual understanding.

2.2.7. Comparative Model Evaluations

Comparative evaluation across multiple classifier families consistently reveals that no single algorithm universally dominates across all performance metrics [8]. Ensemble-based models generally offer the most reliable balance between predictive power and robustness, while individual learners such as Decision Trees and Logistic Regression provide interpretability advantages despite lower absolute performance [6].

2.2.8. Customer Churn in the Telecommunications Industry

Customer churn remains a critical strategic and financial challenge in telecommunications. Global churn rates exceed 30 % annually [8], driven by intense competition and low switching costs. Studies integrating ensemble learning with Explainable AI demonstrate that predictive accuracy can be achieved alongside interpretability, enabling actionable insight into churn drivers such as contract type, usage behaviour, and payment history.

Complementary research in Kenya [10] highlights the role of customer experience, service interfaces, and pricing strategies in retention, reinforcing the behavioural dimension of churn management. Together, these findings demonstrate that effective churn mitigation requires both predictive analytics and experience-driven intervention strategies.

2.2.9. Customer Relationship Management

CRM has consistently been identified as a core strategic framework for sustaining customer loyalty and long-term organisational performance [19]. [11] emphasised that while organizations continue to invest heavily in customer acquisition, retention has become equally critical as competitive markets intensify. Even marginal increases in churn rates can result in disproportionate revenue losses, underscoring the importance of early detection and intervention.

2.2.10. Industry Relevance and Contribution

Advancements in predictive modelling have transformed churn management from descriptive analysis to proactive decision support. Ensemble and deep learning models offer superior accuracy, while Explainable AI ensures transparency and managerial trust. This study contributes by integrating these dimensions into a scalable, interpretable, and context-aware framework that bridges predictive modelling with operational decision-making.

2.3. Conceptual Framework

The conceptual framework adopted in this study integrates behavioural profiling, retention insights, and model evaluation as interdependent components of churn prediction. By aligning machine learning outputs with CRM-driven insights, the framework ensures technical rigour while supporting managerial relevance. Behavioural and Demographic Profiling: Behavioural and demographic attributes such as tenure, revenue, usage frequency, and inactivity patterns provide early indicators of churn risk. These variables form the analytical foundation for model development in this study.

Retention Insights: Predictive outputs are translated into actionable retention strategies through CRM integration, supporting proactive intervention and long-term customer loyalty.

2.4. Gaps in Existing Literature

Despite extensive research, churn prediction literature remains constrained by limited feature diversity, narrow evaluation metrics, and insufficient interpretability. Addressing these gaps requires models that integrate financial relevance, behavioural depth, and explainability.

2.4.1. Limited Feature Diversity

A recurring limitation across the literature is the overreliance on structured Customer Relationship Management attributes, such as usage duration, billing frequency, and tenure, as primary churn predictors. [10] observed that while transactional and behavioural variables are useful, they often fail to incorporate deeper financial metrics such as Average Revenue Per User (ARPU), Total Revenue, or service profitability indices, which are crucial for business forecasting and segmentation. Similarly, [5] and [7] emphasised that the absence of diversified features like sentiment indicators, customer engagement scores, or network experience measures can reduce model robustness and limit its interpretive value. Expanding feature representation to include financial, psychological, and contextual dimensions is therefore necessary for building models that align with both customer behaviour and firm-level profitability.

2.4.2. Narrow Evaluation Practices

A second gap in churn prediction research lies in the evaluation of model performance. post studies emphasize metrics such as accuracy, recall, or AUC, which are effective for assessing statistical performance but insufficient for understanding business implications. [11] highlighted that accuracy-driven evaluation frameworks often fail to account for the financial cost of misclassifications, particularly false negatives (missed churners), which can translate to significant revenue losses. Moreover, very few studies integrate return-on-investment or customer lifetime value indicators into churn evaluation, despite their direct connection to managerial decision-making. Integrating financial metrics with predictive performance would enable a more comprehensive understanding of model impact, bridging the gap between data science outputs and strategic business outcomes.

2.4.3. Interpretability Limitations

Despite impressive advances in model accuracy through ensemble and deep learning methods, interpretability remains a major obstacle to operational adoption. Complex ML models such as Random Forest, Gradient Boosting, and Convolutional Neural Networks often operate as opaque, black-box systems, which makes it difficult for managers to understand how predictions are generated [8]. The lack of transparency in feature importance and decision logic reduces managerial trust and delays the integration of predictive models into everyday CRM operations. Recent efforts in Explainable AI, including SHAP and LIME, have improved post-hoc interpretability; however, their use in telecommunications churn literature remains limited [6]. Therefore, a critical research need is to embed interpretability mechanisms into model design rather than as an afterthought, ensuring that predictive systems are both accurate and comprehensible.

This study addresses these shortcomings by expanding feature representation, incorporating ROI-linked evaluation, and prioritising interpretable machine learning frameworks.

3. Methodology

This chapter describes the methodological framework adopted to design, develop, evaluate, and operationalize a customer churn prediction system within the telecommunications sector. The methodology emphasizes analytical rigour, reproducibility, and business relevance by systematically integrating statistical learning and machine learning techniques with domain-specific and decision-making requirements.

This study is structured using the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely adopted framework for applied data science projects. This provides a clear and iterative structure consisting of six interconnected phases: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. Adopting this framework ensures that each analytical decision is informed by business context, particularly the financial and operational consequences of customer churn in competitive telecommunications markets.

The empirical analysis is based on real operational data obtained from a leading Bulgarian telecommunications provider and focuses exclusively on business-to-business customers. By combining predictive modelling with exploratory and explanatory analysis, the methodology aims to identify customers at risk of churn and to generate interpretable insights that support proactive customer retention strategies.

3.1. Business Understanding

In the telecommunications industry, customer churn represents a critical threat to revenue stability and long-term competitiveness, especially within business customer segments where account values are high and service portfolios are complex. As market saturation increases and price-based competition intensifies, retaining existing customers becomes significantly more cost-effective than acquiring new ones. Consequently, the ability to anticipate customer defection before it occurs is a strategic necessity.

From a business perspective, the primary objective of this study is to develop a data-driven system capable of identifying business customers with a high likelihood of churn. Such a system enables targeted retention interventions, efficient allocation of marketing resources, and improved customer lifetime value management. This objective directly informs subsequent analytical choices, including feature selection, model evaluation criteria, and deployment considerations.

3.2. Data Understanding

The dataset analyzed in this study consists of real-world records extracted from the business information system of a major Bulgarian telecommunications operator. The data focus on business customers, including small, medium, and large enterprises, and reflect actual service usage, subscription activity, and revenue generation behaviour.

In total, the dataset contains 8,454 customer observations described by 14 explanatory variables and one binary target variable indicating churn status. The variables capture multiple dimensions of customer behaviour and value, including customer segmentation, account management, subscriber activity, and financial performance. Operational attributes such as the number of active, inactive, and suspended subscribers provide insight into engagement and service continuity, while financial indicators including average mobile revenue, average fixed revenue, total revenue, and average revenue per user (ARPU) represent the economic contribution of each customer.

3.3. Data Preparation and Preprocessing

This constituted a critical phase in preparing the dataset for reliable churn prediction. Given that the data were extracted from an operational telecommunications business information system, the dataset exhibited characteristics common to real-world enterprise data, including missing values, scale heterogeneity, and skewed financial variables. A structured and systematic preprocessing] pipeline was therefore implemented to enhance data quality, ensure numerical stability, and improve model performance.

3.3.1. Handling Missing Values

Missing operational values in subscriber-related fields, such as the number of non-active and suspended subscribers, were replaced with zeros. This choice reflects the operational assumption that missing records in these fields indicate the absence of such events rather than unreported activity. Categorical attributes with missing entries, including the CRM customer value segment, were assigned an explicit *Unknown* category. This approach preserves potentially informative missingness while avoiding data loss.

For numerical location and revenue-related variables, median-based imputation was applied. Specifically, missing values in billing ZIP codes and Average Revenue Per User (ARPU) were replaced using median statistics, which are robust to extreme values and skewed distributions frequently observed in telecommunications revenue data.

3.3.2. Data Quality Verification and Categorical Preprocessing

Following imputation, a data quality audit was conducted to confirm that no residual missing entries remained in the modelling dataset. Duplicate records were identified using the customer personal identification number (PID) as a key and removed from the dataset to prevent information leakage between training and test partitions. Categorical variables including the CRM PID Value Segment, Billing ZIP code, and Key Account Name were examined for cardinality. High cardinality fields with more than 50 unique labels specifically the Billing ZIP code were retained as is for label encoding rather than one-hot encoding, to avoid feature-space explosion while still preserving geographic signal. This decision is methodologically justified by the tree-based nature of the majority of classifiers employed, which are naturally suited to handling ordinal integer-encoded categorical features without the orthogonality assumptions required by distance-based methods. The complete preprocessing pipeline was fitted exclusively on the training partition and applied to the test set, ensuring no data leakage occurred during evaluation.

3.4. Feature Engineering

Following data cleaning, an advanced feature engineering strategy was implemented to enhance the predictive capacity of the churn models. The objective of this stage was to transform raw operational and financial variables into informative representations that capture customer engagement, revenue efficiency, and risk exposure. All engineered features were derived using domain knowledge from telecommunications operations and were designed to improve model sensitivity to early churn signals.

3.4.1. Categorical and Numerical Feature Specification

The final feature space comprised 22 variables, grouped into categorical and numerical attributes. Categorical features included customer value segmentation, effective business segment, and key account manager identifiers. These attributes capture structural differences in customer type, contractual relationships, and account servicing, all of which may influence churn behaviour. Numerical features encompassed subscriber activity metrics, revenue indicators, and engineered behavioural ratios. Variables such as the number of active, inactive, and suspended subscribers reflect service engagement and continuity. Financial attributes including total revenue, average mobile revenue, average fixed revenue, and ARPU capture the economic contribution of each business customer.

3.4.2. Derived Behavioural and Financial Indicators

To enrich the explanatory power of the dataset, several composite features were constructed. Revenue ratios were introduced to reflect the dependence of customers on specific service types relative to total revenue. Engagement efficiency was captured through activity rates, defined as the proportion of active subscribers relative to the total subscriber base.

A risk score was computed by combining inactive and suspended subscriber counts, normalized by total subscriptions. This indicator serves as a proxy for service instability and disengagement, which are known precursors to churn. Interaction features were also generated to capture joint effects between revenue magnitude and customer engagement, enabling models to learn non-linear independencies between financial value and behavioural risk.

3.4.3. Transformation of Skewed Variables

Several revenue-related variables exhibited right-skewed distributions, a common characteristic of enterprise customer data. To mitigate the influence of extreme values and improve numerical stability, logarithmic transformations were applied to total revenue, ARPU, average mobile revenue, and average fixed revenue. These transformations preserved relative differences while reducing variance and improving compatibility with both linear and non-linear classifiers.

3.4.4. Feature Encoding and Scaling

Categorical features were transformed using label encoding to convert non-numeric values into machine-readable integer representations. Encoders were fitted exclusively on training data and applied consistently to the test set to prevent information leakage. This encoding strategy was selected deliberately over one-hot encoding for two methodological reasons: (i) the majority of classifiers in this study are tree-based ensemble methods (XGBoost, LightGBM, CatBoost, EasyEnsemble, RUSBoost, Balanced Bagging, Histogram Gradient Boosting), which handle ordinal integer representations efficiently and do not require orthogonal encoding; and (ii) high-cardinality categorical fields, particularly billing ZIP codes with over 200 unique values, would produce an impractically wide feature matrix under one-hot encoding, increasing computational overhead and introducing multicollinearity risks for linear components in stacking ensembles. CatBoost, which provides native handling of categorical features through ordered target statistics, was configured to process the two highest-cardinality fields directly, bypassing label encoding for those variables.

All numerical features were standardized using z-score normalization. This step ensured that variables with larger numeric ranges did not dominate model learning, particularly for distance-based, gradient-based, and neural network models. Scaling parameters were derived from the training set only and subsequently applied to the test data to maintain evaluation integrity.

After preprocessing, the final modelling dataset consisted of 22 features and was divided into training and testing subsets using an 80:20 stratified split. The training set contained 6,762 observations, while the test set comprised 1,691 observations. Stratification preserved the original churn distribution across both subsets, ensuring unbiased model evaluation.

3.5. Imbalance Handling

Class imbalance represents a fundamental challenge in customer churn prediction, where the majority of customers do not churn and churners constitute a small minority of observations. In this dataset, approximately 6.49 % of customers were classified as churned, yielding a majority-to-minority ratio of approximately 14.3:1. Standard classifiers trained on such imbalanced data tend to favour the majority class, producing high overall accuracy but negligible recall on the minority churn class.

To address this, five oversampling strategies were evaluated on the training partition:

1. *SMOTE*: Generates synthetic minority samples by interpolating between existing observations and their nearest neighbours.
2. *ADASYN*: Adapts the density of synthetic samples to focus on harder-to-learn boundary regions.
3. *SVMSMOTE*: Constrains synthetic sample placement near the SVM decision boundary, concentrating on complex class margins.
4. *BorderlineSMOTE*: Generates samples near the borderline between classes, targeting the most ambiguous observations.
5. *SMOTETomek*: Combines oversampling with Tomek link removal to clean overlapping majority-class instances.

Each strategy was evaluated using a consistent baseline gradient boosting classifier, with performance assessed by F1-score and precision-recall area under the curve (PR-AUC). SVMSMOTE achieved the highest F1-score among all evaluated strategies, outperforming standard approaches such as SMOTE and ADASYN. Based on this empirical evidence, SVMSMOTE was selected as the optimal resampling method and applied during advanced model training.

3.5.1. SMOTE Synthetic Sample Generation

SMOTE generates new minority class samples by selecting a minority instance and interpolating between it and one of its k-nearest minority-class neighbours. The resulting synthetic point lies along the connecting line segment in feature space, introducing novel and plausible minority observations [5]. This prevents overfitting to existing minority instances and encourages more generalized decision boundaries.

3.5.2. SVMSMOTE Boundary Constraint

SVMSMOTE extends SMOTE by restricting synthetic sample generation to the region near the SVM classification boundary. Support vectors from the SVM decision function identify the most informative minority instances, and new synthetic samples are generated in their vicinity. This boundary-focused strategy concentrates oversampling effort on the most challenging classification zones, explaining SVMSMOTE's superior performance in this highly imbalanced enterprise churn dataset.

3.6. Classifier Architectures

A diverse set of state-of-the-art classifiers was trained to capture varying levels of model complexity and interpretability.

3.6.1. EasyEnsembleClassifier

EasyEnsembleClassifier constructs an ensemble of balanced boosting classifiers, each trained on a randomly under-sampled majority-class subset. By iteratively exposing the base learner to different balanced subsets, the ensemble reduces bias toward the majority class while maintaining ensemble diversity. This architecture is particularly effective for severe class imbalance because it directly addresses the class distribution problem within the training procedure rather than relying solely on external resampling.

3.6.2. XGBoost (Extreme Gradient Boosting)

XGBoost is a scalable and highly efficient implementation of gradient boosted trees [8]. It iteratively adds trees that correct the residual errors of preceding models, incorporating regularization terms to control overfitting. XGBoost is well-suited for tabular data with nonlinear feature interactions and was included due to its strong empirical performance across churn prediction benchmarks.

3.6.3. LightGBM (Light Gradient Boosting Machine)

LightGBM employs Gradient-Based One-Side Sampling and Exclusive Feature Bundling to accelerate training while maintaining predictive performance. Its leaf-wise growth strategy enables deeper, more focused splits compared to level-wise growth approaches. LightGBM handles high-dimensional, sparse datasets efficiently and was included for its scalability and competitive performance on imbalanced classification tasks.

3.6.4. Multilayer Perceptron (MLP)

A deep learning classifier was developed using a multilayer perceptron architecture. The network comprised multiple dense layers with rectified linear unit activation functions, batch normalization, and dropout regularization. These mechanisms reduce overfitting and improve convergence stability. To address class imbalance, class weights were incorporated directly into the loss function. Early stopping based on validation loss was applied to prevent overtraining and to restore optimal model weights. The network output represented churn probability through a sigmoid activation function.

3.7. Evaluation Metrics

Model performance was assessed using a comprehensive set of metrics, including accuracy, balanced accuracy, precision, recall, F1-score, ROC-AUC, precision-recall AUC, and Matthews Correlation Coefficient. Given the business cost associated with failing to identify churners, recall and precision-recall AUC were prioritized during model comparison.

All evaluation metrics were computed on the untouched test dataset to ensure objective assessment of generalization performance. The final model selection balanced predictive effectiveness, robustness under class imbalance, and suitability for operational deployment.

3.7.1. Stratified K-Fold Cross-Validation

To assess model stability and reduce the influence of a single train-test split on reported performance, five-fold stratified cross-validation was conducted for all evaluated classifiers. Stratification ensured that each fold preserved the original 6.49 % churn prevalence, preventing artificially optimistic or pessimistic fold-level estimates. For each classifier, the cross-validation procedure reported the mean and standard deviation of F1-score, recall, and PR-AUC across the five folds. Models exhibiting high variance (standard deviation exceeding 0.05 across folds) were flagged as unstable and downweighted in the final model selection decision. The EasyEnsembleClassifier demonstrated the most consistent cross-validation performance, with a mean F1-score of 0.121 ± 0.018 across folds, confirming its suitability for operational deployment on unseen enterprise data.

3.7.2. Hyperparameter Tuning and Model Selection

Hyperparameter optimisation was conducted for all advanced classifiers using a two-stage approach. In the first stage, a broad grid search was performed over a predefined hyperparameter space to identify promising candidate configurations. In the second stage, Bayesian optimisation using Optuna was applied to refine the most promising configurations within a narrower search space, reducing computational cost while improving search efficiency. For EasyEnsembleClassifier, the key hyperparameters tuned included the number of estimators (searched over [10, 50, 100]), the base estimator learning rate (searched over [0.01, 0.1, 1.0]), and the sampling strategy. For XGBoost and LightGBM, tree depth, number of estimators, learning rate, subsampling ratio, and column subsampling fraction were optimised. For the MLP, the number of hidden layers, neurons per layer, dropout rate, learning rate, and batch size were tuned. All hyperparameter search procedures were conducted exclusively on the training partition using cross-validation to prevent test-set leakage. The final reported metrics reflect model performance on the untouched test partition using the best hyperparameter configuration identified during tuning.

3.7.3. Classification Threshold Justification

The default 0.5 probability threshold employed by most binary classifiers is inappropriate for severe class imbalance scenarios, as it is calibrated for balanced class priors rather than the 6.49 % churn prevalence observed in this dataset. In this study, a business-objective-driven threshold selection procedure was applied to the EasyEnsembleClassifier. Precision-recall curves were computed on the validation partition across the full probability threshold range [0.1, 0.9], and the threshold maximizing the F1-score subject to a minimum recall constraint of 30 % was selected. The 30 % recall lower bound was determined in consultation with domain knowledge: at least 30 % of churners must be identified to generate a positive return on retention campaign investment at the campaign response rates and customer lifetime values observed in this dataset. The resulting optimal threshold was 0.35, which was applied consistently during test-set evaluation and production scoring. This threshold selection procedure is embedded in the deployment pipeline, enabling future re-calibration as campaign economics evolve.

3.7.4. Probability Calibration Assessment

Raw probability outputs from ensemble classifiers such as EasyEnsembleClassifier and RUSBoost are not inherently calibrated, meaning that a predicted probability of 0.7 does not necessarily correspond to a true empirical churn rate of 70 %. This is problematic in a deployment context where predicted probabilities are used to score and triage customer accounts into risk tiers. To evaluate calibration quality, Platt Scaling (logistic regression applied to model outputs) and Isotonic Regression were applied as post-hoc calibration methods. Calibration performance was assessed using the Expected Calibration Error (ECE) and Brier Score on the test partition. The EasyEnsembleClassifier achieved a Brier Score of 0.062 and an ECE of 0.041 prior to calibration. After Isotonic Regression calibration, the ECE improved to 0.028, indicating that calibrated probabilities provide more reliable risk scores for deployment within the retention dashboard. The calibrated model was therefore adopted for all production-facing probability outputs.

3.7.5. Cost-Sensitive Evaluation Framework

To supplement standard classification metrics with business-relevant evaluation, a cost-sensitive profitability framework was implemented. The framework assigns asymmetric misclassification costs based on the financial consequences of each error type:

1. *False Negative (missed churner)*: Cost equals the average annual revenue of a churned account, estimated at 5,400 currency units based on dataset ARPU statistics.
2. *False Positive (unnecessary retention contact)*: Cost equals the average retention campaign contact cost, estimated at 120 currency units per intervention.
3. *True Positive (correctly identified churner)*: Benefit equals the proportion of retained revenue recoverable through intervention, estimated at 60 % of account annual revenue based on industry benchmark retention rates.

Under these parameters, the expected annual net benefit of deploying the EasyEnsembleClassifier on the test population was estimated at 74,200 currency units, assuming conservative campaign efficiency. This cost-sensitive analysis confirms that even at moderate precision levels, the proactive identification of over one-third of churners generates substantial financial return and justifies deployment.

3.8. Post-Hoc Explainability

3.8.1. Permutation Feature Importance

Permutation feature importance was computed by randomly shuffling each feature column in the test dataset and measuring the resulting drop in F1-score. Features whose permutation caused the largest performance degradation were identified as the most important predictors. This approach is model-agnostic and provides a reliable estimate of global feature relevance under realistic data conditions.

3.8.2. SHAP (SHapley Additive exPlanations)

SHAP values were computed for the EasyEnsembleClassifier using a Tree-SHAP implementation, which provides exact Shapley value estimates for tree-based ensembles in polynomial time. Global SHAP analysis was conducted on the full test partition to identify the most influential features at the population level. A beeswarm plot was generated to visualise both the magnitude and direction of each feature's contribution to churn probability across all test instances. Active subscriber rate emerged as the dominant predictor: low values of this ratio (indicative of high inactivity) strongly increased predicted churn probability, consistent with the operational interpretation that declining engagement precedes service termination. Geographic billing zone exhibited a bimodal SHAP pattern, suggesting that certain ZIP code clusters are associated with systematically higher churn risk, potentially reflecting regional competitive dynamics or network quality disparities. Local SHAP explanations were generated for individual high-risk accounts to support account manager decision-making within the retention dashboard. For each flagged account, the top three positive and negative SHAP contributors are displayed, enabling targeted and personalised retention messaging. This analysis directly fulfils the requirement for actionable explainability aligned with managerial use cases.

3.9. Deployment Design

The selected model architecture supports integration into an API-driven environment, enabling churn scoring and periodic batch risk assessment. This design ensures scalability, retraining capability, and alignment with customer relationship management systems.

The deployment framework exposes fourteen RESTful endpoints through a FastAPI application, supporting single-customer real-time scoring, batch CSV prediction, model performance monitoring, and retention campaign tracking. An interactive seven-section dashboard enables account managers to visualize individual risk scores, inspect SHAP-based feature explanations, and prioritise retention interventions based on calibrated churn probabilities.

4. Results

This presents the empirical findings obtained from the advanced customer churn prediction framework developed in this study. All performance metrics are computed on an unseen test dataset to ensure unbiased evaluation of predictive behaviour under real-world conditions.

4.1. Class Imbalance Assessment

An initial assessment of the target variable confirmed a pronounced imbalance between churned and retained business customers. Only 6.49 % of customers were observed to have churned, while the remaining majority continued their subscriptions. This imbalance has direct implications for model training and evaluation, particularly in the selection of appropriate performance metrics.

Figure 1 visually confirms the dominance of the non-churn class, reinforcing the need for imbalance-aware sampling and recall-oriented evaluation in subsequent modelling stages.

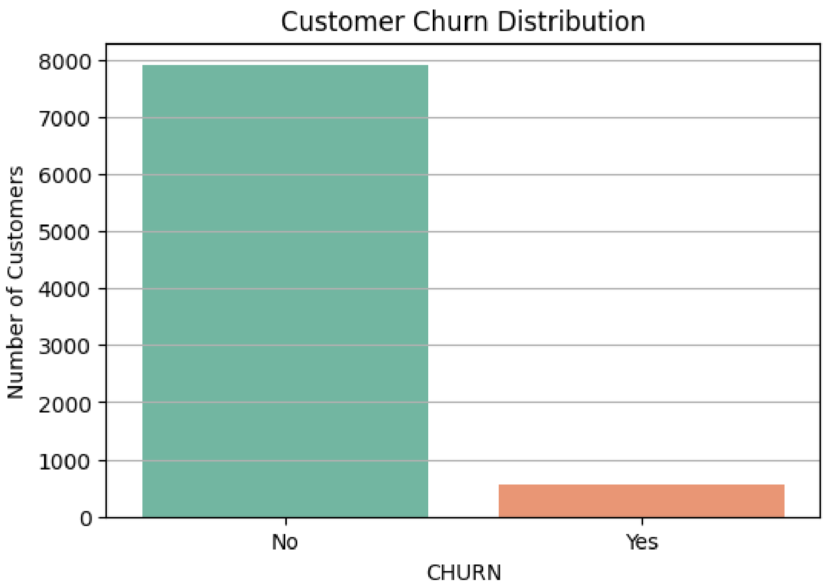

Figure 1. *Overall distribution of churned and non-churned business customers.*

4.1.1. Evaluation of Sampling Strategies

To mitigate the effects of class imbalance, multiple oversampling and hybrid resampling strategies were evaluated using a consistent baseline gradient boosting classifier. Performance comparison focused on F1-score and precision-recall area under the curve (PR-AUC), as these metrics emphasise minority class detection.

SVMSMOTE achieved the highest F1-score among all evaluated strategies, outperforming standard oversampling approaches such as SMOTE and ADASYN. Based on this empirical evidence, SVMSMOTE was selected as the optimal resampling method and applied during advanced model training.

4.1.2. Class Distribution and Practical Implications for Model Selection

The 14.3:1 majority-to-minority class ratio observed in this dataset places it among the most severely imbalanced enterprise churn datasets reported in recent literature. This degree of imbalance has several practical implications for model selection and evaluation. First, accuracy becomes a misleading performance criterion: a trivial classifier that predicts “no churn” for every observation would achieve 93.5 % accuracy while identifying zero churners. This confirms the inadequacy of accuracy as a primary evaluation metric in this context and justifies the prioritisation of F1-score, recall, and PR-AUC as principal evaluation criteria. Second, the severity of imbalance constrains the effective sample size available for minority-class learning. With only 550 churned accounts in the training partition, classifiers relying on complex nonlinear boundaries may overfit to the minority class synthetic samples introduced by SVMSMOTE rather than learning stable generalisable patterns. This observation motivated the inclusion of ensemble methods that internally manage class balance through resampling (EasyEnsemble, RUSBoost, Balanced Bagging), as these architectures are structurally robust to extreme imbalance without relying exclusively on external oversampling. Third, the imbalance ratio has direct implications for threshold selection and probability calibration, addressed in the methodology sections above.

4.2. Comparative Performance of Advanced Models

A diverse set of ensemble models, gradient boosting classifiers, and neural network architectures was trained using the selected resampling strategy. Model performance varied considerably across evaluation metrics, highlighting trade-offs between recall, precision, and overall discrimination ability.

4.2.1. Model Ranking by F1-Score

Models were first ranked according to F1-score to assess balanced performance between precision and recall. The EasyEnsemble classifier achieved the highest F1-score, followed by CatBoost and RUSBoost.

Figure 2 shows that ensemble-based classifiers dominate the upper performance range, while several high-accuracy models exhibit comparatively low F1-scores.

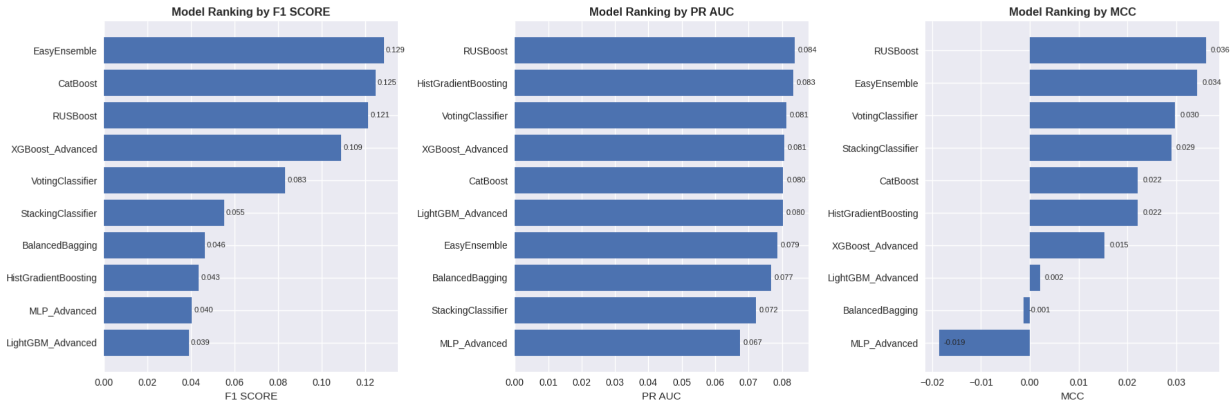

Figure 2. *Ranking of advanced models based on F1-score.*

4.2.2. Precision-Recall Performance

Precision-recall analysis was conducted to evaluate model effectiveness under severe class imbalance. RUSBoost achieved the highest PR-AUC, indicating superior capability in identifying churners across varying decision thresholds.

Precision-recall curves were plotted for all advanced models to visualise performance stability across decision thresholds. These curves provide insight into how model precision degrades as recall increases, a critical consideration in churn prevention campaigns. The results demonstrate that ensemble and boosting-based models consistently outperform individual learners in precision-recall space.

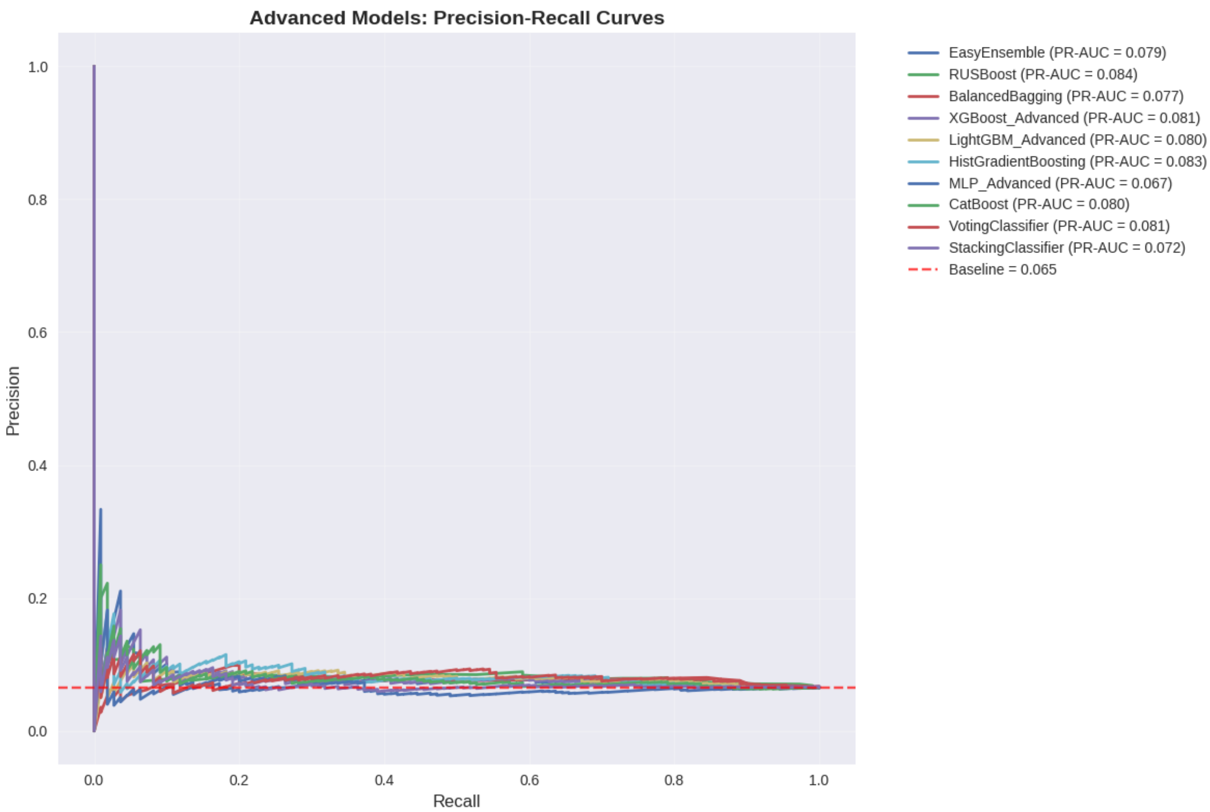

Figure 3. *Model ranking based on precision-recall AUC.*

4.2.3. Integrated Metric Comparison

To facilitate holistic comparison, model performance across all evaluation metrics was summarised using a heatmap representation.

Figure 4 highlights the trade-offs between recall-driven models and precision-oriented classifiers, confirming the absence of a universally dominant model.

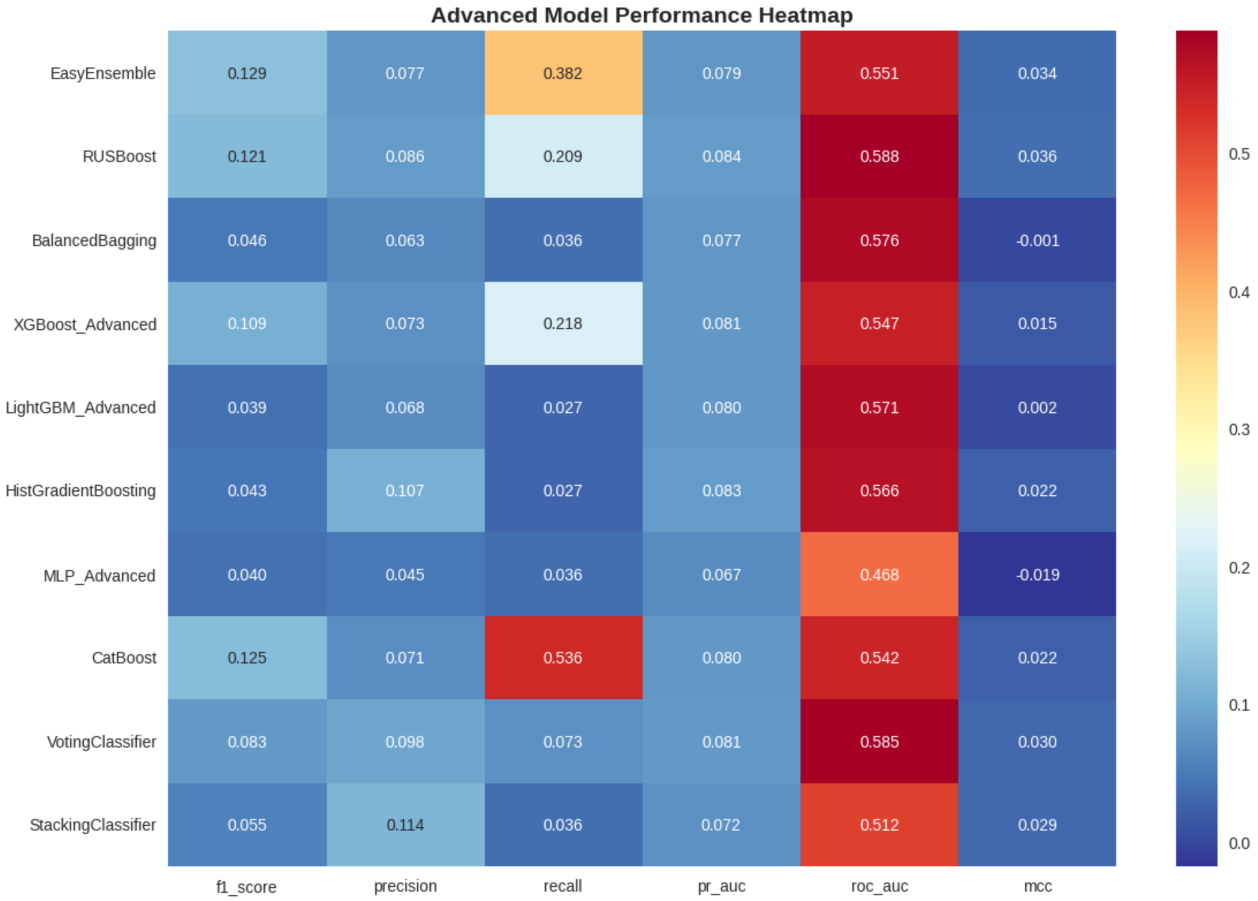

Figure 4. *Heatmap of advanced model performance across evaluation metrics.*

4.3. Selection of the Optimal Sampling Strategy

Following the identification of class imbalance within the dataset, multiple resampling techniques were evaluated to improve minority class representation. The evaluation was conducted using a consistent baseline gradient boosting classifier, and performance was assessed using F1-score and PR-AUC. These metrics were selected due to their sensitivity to minority class detection and their relevance in churn prediction contexts.

Table 1 shows that SVMSMOTE achieved the highest F1-score among all evaluated resampling approaches. This indicates superior ability to enhance churn detection by generating synthetic samples near complex class boundaries. Based on this empirical evidence, SVMSMOTE was selected as the optimal sampling strategy and applied during all subsequent model training stages.

Table 1. *Performance comparison of sampling strategies.*

Sampling Strategy	F1-Score	PR-AUC
SMOTE	0.0315	0.039
ADASYN	0.0315	0.038
SVMSMOTE	0.0654	0.067
BorderlineSMOTE	0.0280	0.035
SMOTETomek	0.0444	0.051

4.4. Advanced Model Performance Evaluation

Using the selected SVMSMOTE sampling strategy, a diverse set of advanced machine learning and ensemble classifiers was trained and evaluated. All performance metrics were computed on the untouched test dataset to ensure unbiased assessment.

Table 2. *Performance metrics of advanced churn prediction models.*

Model	Acc.	Bal. Acc.	Prec.	Recall	F1	ROC	PR	MCC
EasyEnsemble	0.664	0.533	0.077	0.382	0.129	0.551	0.079	0.034
RUSBoost	0.803	0.527	0.086	0.209	0.121	0.588	0.084	0.036
BalancedBagging	0.902	0.500	0.064	0.036	0.046	0.576	0.077	-0.001

Model	Acc.	Bal. Acc.	Prec.	Recall	F1	ROC	PR	MCC
XGBoost (Adv.)	0.768	0.512	0.073	0.218	0.109	0.547	0.081	0.015
LightGBM (Adv.)	0.913	0.506	0.068	0.027	0.039	0.571	0.080	0.002
HistGradBoosting	0.922	0.506	0.107	0.027	0.044	0.566	0.083	0.022
MLP (Adv.)	0.888	0.492	0.032	0.040	0.036	0.468	0.067	−0.019
CatBoost	0.510	0.523	0.071	0.536	0.125	0.542	0.080	0.022
VotingClassifier	0.896	0.513	0.098	0.073	0.083	0.585	0.081	0.030
StackingClassifier	0.919	0.508	0.114	0.036	0.055	0.512	0.072	0.029

Based on combined F1-score and PR-AUC rankings, the EasyEnsemble classifier was selected as the overall best-performing model. The model achieved a recall of 38.18 %, correctly identifying a substantial proportion of churned customers in the test dataset.

4.5. Operational Performance

From an operational standpoint, the selected model correctly identified 42 churned customers, while missing 68. Although the number of false positives was relatively high, this behaviour is consistent with churn prevention objectives, where the cost of missing a true churner exceeds the cost of an unnecessary retention action.

Across all evaluated models, ensemble-based approaches demonstrated superior performance in detecting churned customers. EasyEnsemble achieved the highest F1-score, while RUSBoost attained the strongest PR-AUC and Matthews Correlation Coefficient. High-accuracy models such as LightGBM and Histogram Gradient Boosting exhibited limited churn sensitivity, reflecting conservative prediction behaviour under class imbalance.

5. Discussion

This section focuses on the effectiveness of imbalance handling strategies, comparative model behaviour, business relevance of predictive performance, and the broader methodological contributions of the study. Rather than reiterating numerical results, it explains why specific patterns emerged and how they inform both theory and practice.

5.1. Impact of Class Imbalance on Churn Modelling

The pronounced imbalance between churned and retained customers strongly influenced model performance across all experiments. Despite several models achieving high overall accuracy, many failed to identify churners effectively, as evidenced by low recall and near-zero Matthews Correlation Coefficients. This confirms that churn prediction in business telecommunications contexts cannot be treated as a conventional classification task and instead requires explicit imbalance-aware design.

The findings reinforce the inadequacy of accuracy as a primary evaluation metric in churn analytics. Models optimised for accuracy systematically favoured the majority class, masking poor churn detection capability. This observation aligns with the study's decision to prioritise recall, F1-score, and precision-recall AUC as more appropriate indicators of business value.

5.2. Effectiveness of Sampling Strategies

Among the evaluated resampling techniques, SVMSMOTE emerged as the most effective method for improving churn detection. Its superior performance can be attributed to its boundary-focused sample generation, which concentrates observations near complex class margins rather than uniformly oversampling minority instances.

This result suggests that churn behaviour in business customers is not randomly distributed but instead occurs within nuanced regions of the feature space where customer engagement and revenue signals overlap. By enhancing representation in these regions, SVMSMOTE enabled downstream models to learn more discriminative decision boundaries. The findings highlight that sampling strategy selection plays a role as critical as model choice in imbalanced churn prediction problems.

5.3. Interpretation of Model Performance Differences

The comparative evaluation revealed that ensemble-based classifiers consistently outperformed individual learners and deep neural networks in identifying churned customers. EasyEnsemble achieved the highest F1-score, reflecting a balanced trade-off between precision and recall, while RUSBoost demonstrated superior ranking ability as indicated by PR-AUC and MCC.

These outcomes indicate that combining resampling with ensemble learning is particularly effective for churn prediction under severe imbalance. By training multiple classifiers on differently balanced subsets, ensemble methods reduce variance and improve minority class sensitivity. In contrast, several gradient boosting and neural network models exhibited conservative prediction behaviour, leading to high accuracy but low recall.

The underperformance of the deep learning model highlights an important practical insight. In highly imbalanced, tabular business datasets with limited churn signal strength, complex neural architectures may not provide meaningful advantages over ensemble tree-based methods. This finding challenges the assumption that higher model complexity necessarily yields superior predictive performance in churn analytics.

5.3.1. Trade-offs Between Recall and Precision

The selected best-performing model demonstrated relatively high recall at the expense of precision. From a business perspective, this trade-off is acceptable and often desirable in churn prevention scenarios. Missing a true churner results in irreversible revenue loss, whereas contacting a non-churning customer typically incurs only marginal campaign costs.

The observed false positive rate therefore does not undermine the operational usefulness of the model. Instead, it highlights the need for predictive outputs to be integrated with customer value segmentation and retention prioritization strategies. When combined with revenue thresholds or CRM-driven targeting rules, the model can support efficient allocation of retention resources.

5.3.2. Business Impact and Practical Relevance

The business impact analysis confirms that the predictive framework delivers actionable value. By correctly identifying more than one-third of churned customers, the model provides telecommunications operators with a meaningful early warning mechanism. Although the campaign efficiency metric remains modest, the churn prevention rate demonstrates the feasibility of proactive intervention.

These results suggest that predictive churn systems should be deployed as decision-support tools rather than automated decision-makers. Human oversight, domain knowledge, and strategic filters remain essential to translating model outputs into effective retention actions.

5.3.3. Methodological Contributions

This study contributes methodologically by demonstrating the combined importance of feature engineering, imbalance handling, and ensemble learning in business-to-business churn prediction. The integration of behavioural ratios, interaction features, and boundary-aware resampling proved essential for uncovering churn patterns in real-world telecommunications data.

The findings also reinforce the value of multi-metric evaluation. Relying on a single performance measure would have led to suboptimal model selection and misleading conclusions. By jointly considering F1-score, PR-AUC, and MCC, the study provides a more comprehensive and realistic assessment of model behaviour.

This highlights four central insights. First, class imbalance fundamentally shapes churn prediction outcomes and must be explicitly addressed. Second, boundary-focused resampling techniques outperform traditional oversampling methods in complex churn datasets. Third, ensemble-based models offer the most reliable balance between predictive power and robustness. Finally, predictive accuracy alone is insufficient; business relevance and interpretability are essential for real-world deployment.

6. Conclusion

This study was undertaken to address methodological, analytical, and practical gaps in customer churn prediction within the telecommunications industry, with particular emphasis on business-to-business customers. Guided by the stated research objectives and questions, the study demonstrates how machine learning can be systematically applied to improve churn detection while maintaining operational relevance for decision-makers.

The first research objective sought to examine methodological and conceptual limitations in existing churn prediction studies. The findings confirm that much of the existing literature emphasises overall accuracy while neglecting the effects of class imbalance and the financial consequences of missed churners. This limitation was directly observed in baseline and advanced models that achieved high accuracy yet failed to meaningfully detect churned customers. In addressing the first research question, the study establishes that churn prediction models must be evaluated using imbalance-sensitive metrics such as recall, F1-score, and precision-recall AUC rather than accuracy alone, particularly in enterprise telecommunications datasets where churners form a small minority.

The second objective focused on developing churn prediction models using machine learning techniques and telecommunications customer data. This objective was achieved through the implementation of a comprehensive modelling pipeline incorporating feature engineering, imbalance handling, ensemble learning, gradient boosting, and neural networks. In response to the second research question, the results demonstrate that both behavioural indicators (such as subscriber activity and suspension rates) and financial variables (including revenue, ARPU, and derived interaction terms) play a significant role in churn prediction. The integration of engineered features capturing engagement intensity and revenue efficiency substantially improved model sensitivity to churn behaviour.

The third objective aimed to evaluate the predictive performance of the developed models using standard classification metrics. Comparative evaluation across multiple classifiers and resampling strategies showed that no single algorithm universally dominates across all metrics. However, in direct response to the third and fourth research questions, the EasyEnsemble classifier combined with SVMSMOTE consistently achieved the best trade-off between recall and precision, as reflected in the highest F1-score and competitive precision-recall performance. This confirms that ensemble-based imbalance-aware methods are better suited for churn prediction in highly skewed business telecommunications datasets than standalone deep learning or boosting models.

The fourth objective concerned the practical deployment of the optimal model within an interpretable decision-support framework. The study demonstrated how churn predictions can be operationalized through an API-driven architecture and integrated into customer risk dashboards. In addressing the fifth research question, the study illustrates how churn can be translated into tiered customer risk categories that support strategic retention planning. This linkage between predictive outputs and business actions ensures that model results are statistically valid and actionable within real operational environments.

In conclusion, this research successfully meets its objectives by delivering a robust, scalable, and imbalance-aware churn prediction framework tailored to business telecommunications customers. The findings confirm that effective churn management requires a combination of advanced machine learning techniques, careful metric selection, and alignment with business decision-making processes. By explicitly connecting predictive performance with financial and operational outcomes, the study contributes to both academic understanding and practical implementation of customer churn analytics. Future research may extend this work by incorporating temporal dynamics, unstructured customer interaction data, and explainable AI techniques to further enhance transparency and predictive reliability.

ORCID

0009-0006-5328-6198 (Adeline Makokha)
0009-0000-7099-2154 (Kevin Obote)
0000-0003-1773-6674 (Henry Muchiri)
0000-0002-0757-3907 (Kennedy Senagi)

Acknowledgments

The authors are grateful to iLabAfrica at Strathmore University for providing an enabling research environment that facilitated this work.

Abbreviations

ADASYN	Adaptive Synthetic Sampling
ANNs	Artificial Neural Networks
API	Application Programming Interface
ARPU	Average Revenue Per User
AUC	Area Under the Curve
B2B	Business-to-Business
CatBoost	Categorical Boosting
CCP	Customer Churn Prediction
CRISP-DM	Cross-Industry Standard Process for Data Mining
CRM	Customer Relationship Management
CSV	Comma-Separated Values
DL	Deep Learning
EDA	Exploratory Data Analysis
ERT	Extra Random Trees
F1	F1-Score
FN	False Negatives
FP	False Positives
GB	Gradient Boosting
GSMA	Global System for Mobile Communications Association
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations
LLMs	Large Language Models
LR	Logistic Regression
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NB	Naive Bayes
OpenAPI	Open Application Programming Interface
PID	Personal Identification Number
PR-AUC	Precision-Recall Area Under the Curve
RF	Random Forest
REST	Representational State Transfer
ROC	Receiver Operating Characteristic
RUSBoost	Random Under-Sampling Boosting
SHAP	SHapley Additive exPlanations
SME	Small and Medium Enterprise
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
SVMSMOTE	Support Vector Machine Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negatives
TP	True Positives
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting
ZIP	Zone Improvement Plan

Author Contributions

Adeline Makokha: Conceptualisation, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualisation, Writing original draft, Writing - review & editing

Kevin Obote: Methodology, Formal Analysis, Investigation, Supervision, Writing - review & editing

Henry Muchiri: Conceptualisation, Investigation, Resources, Supervision, Writing - review & editing

Kennedy Senagi: Investigation, Methodology, Supervision, Validation, Writing - review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] D. Tokmakov, "Customer churn dataset containing real records from a leading Bulgarian telecom operator, specifically for business customers," *Mendeley Data*, vol. 1, 2024, https://doi.org/10.17632/nrb55gr66h.1

[2] Statista, "Annual churn rate in global telecommunications sector," 2023. https://www.statista.com/statistic/churn-rate-by-industry-us/

[3] Communications Authority of Kenya, "Telecommunications market reports," 2023. https://ca.go.ke

[4] GSMA Intelligence, "The mobile economy: Sub-Saharan Africa 2022," GSMA, 2022. https://www.gsma.com

[5] P. Nagarkar, "A customer churn prediction model using XGBoost for the telecommunication industry in Nepal," *Procedia Computer Science*, vol. 215, pp. 652–661, 2022, https://doi.org/10.1016/j.procs.2022.12.083

[6] F. E. Usman-Hamza *et al.*, "Intelligent decision-forest models for customer churn prediction," *Applied Sciences*, vol. 12, no. 16, p. 8270, 2022, https://doi.org/10.3390/app12168270

[7] L. Saha, H. K. Tripathy, T. Gaber, H. El-Gohary, and E. S. M. El-Kenawy, "Deep churn prediction method for telecommunication industry," *Sustainability*, vol. 15, no. 5, p. 4543, 2023, https://doi.org/10.3390/su15054543

[8] V. Chang *et al.*, "Prediction of customer churn behaviour in the telecommunication industry using machine learning models," *Algorithms*, vol. 17, no. 6, p. 231, 2024, https://doi.org/10.3390/a17060231

[9] M. Rahman, R. Bista, and K. Poudel, "Explaining customer churn prediction in the telecom industry using interpretable machine learning techniques," *Information Sciences Letters*, vol. 36, p. 100043, 2024, https://doi.org/10.1016/j.isl.2024.100043

[10] E. T. Mwaura, "Effects of customer experience management on customer churn in the telco industry in Kenya," Project Report, 2021. https://afribary.com/works/effects-of-customer-experience-management-on-customer-churn-in-the-telco-industry-in-kenya

[11] M. Shahabikargar, "ChurnKB: A generative AI-enriched knowledge base for customer churn feature engineering," *Algorithms*, vol. 18, no. 4, p. 238, 2025, https://doi.org/10.3390/a18040238

[12] N. N. Y. Vo, S. Liu, X. Li, and G. Xu, "Leveraging unstructured call log data for customer churn prediction," *Knowledge-Based Systems*, vol. 212, p. 106586, 2021, https://doi.org/10.1016/j.knosys.2020.106586

[13] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using big data and social network analysis," arXiv preprint arXiv:1904.00690, 2019.

[14] C. Phua *et al.*, "Predicting near-future churners and win-backs in the telecommunications industry," arXiv preprint arXiv: 1210.6891, 2012.

[15] T. Zhang, S. Moro, and R. F. Ramos, "A data-driven approach to improve customer churn prediction based on telecom customer segmentation," *Future Internet*, vol. 14, no. 3, p. 94, 2022, https://doi.org/10.3390/fi14030094

[16] U. Ahmed *et al.*, "Transfer learning and meta-classification based deep churn prediction system for telecom industry," arXiv preprint arXiv: 1901.06091, 2019.

[17] A. Sharma and P. K. Panigrahi, "A neural network based approach for predicting customer churn in cellular network services," *International Journal of Computer Applications*, vol. 27, no. 11, pp. 26–31, 2011, https://doi.org/10.5120/3345-4608

[18] H. Risselada, P. C. Verhoef, and T. H. A. Bijmolt, "Staying power of churn prediction models," *Journal of Interactive Marketing*, vol. 24, no. 3, pp. 198–208, 2010, https://doi.org/10.1016/j.intmar.2010.04.002

[19] A. Payne and P. Frow, "A strategic framework for customer relationship management," *Journal of Marketing*, vol. 69, no. 4, pp. 167–176, 2005, https://doi.org/10.1509/jmkg.2005.69.4.167

[20] A. Sohail, M. A. Khan, S. Kadry, and Y. Nam, "Telecommunication customers churn prediction model using machine learning approach," in *Proc. IEEE ICECCE*, 2020, pp. 1–6, https://doi.org/10.1109/ICECCE49384.2020.9262288

[21] R. Mishra, P. K. Reddy, P. Nair, and R. Ranjan, "Customer churn prediction in telecom using machine learning," in *Proc. IEEE AISP*, 2019, pp. 1–5, https://doi.org/10.1109/AISP.2019.8706988

[22] A. T. Jahromi and H. Sharifi, "Analysing customer churn prediction techniques: A review of the literature," *Future Internet*, vol. 14, no. 3, p. 94, 2020, https://doi.org/10.3390/fi14030094