NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

XGBoost Ensemble Model for Churn Prediction in Telecom: A Machine Learning Framework

Ramesh Prasad Bhatta*
Assistant Professor
Central Department of CSIT
Far Western University, Mahendranagar, Nepal
rpb.mcs@gmail.com
https://orcid.org/0009-0005-0554-9072

Karn Dev Bhatt
Assistant Professor
Central Department of CSIT
Far Western University, Mahendranagar, Nepal
bhattkarn123@gmail.com
https://orcid.org/0009-0009-3948-4261

Niraj Pal
Central Department of CSIT
Far Western University, Mahendranagar, Nepal
palniraj21@gmail.com
https://orcid.org/0009-0004-0648-8913

Manoj Raj Pant
Central Department of CSIT
Far Western University, Mahendranagar, Nepal
manojrajpanta0671@gmail.com
https://orcid.org/0009-0008-4455-1732

Corresponding Author*
Received: February 13, 2026, Revised & Accepted: April 29, 2026
Copyright: Author(s) (2026)

This work is licensed under a Creative Commons Attribution-Non Commercial 4.0International License.

Abstract

Background: Customer churn remains a critical challenge in the telecommunications industry, where saturated markets and high acquisition costs demand advanced predictive models. Churn prediction plays an important role for retention-oriented companies, however, imbalance and complex behavioral patterns make traditional techniques inefficient.

Methods: Four machine learning algorithms namely, XGBoost, LightGBM, CatBoost, and

Voting Ensemble were selected according to such criteria as accuracy, precision, recall, F1-score, and AUC-ROC. Furthermore, the methods of confusion matrix, ROC curve, correlation, and SHAP are utilized to analyze results.

Results: CatBoost algorithm achieved the highest values for all criteria (accuracy: 76.05%, F1: 60.77%), while Voting Ensemble obtained the highest AUC-ROC (82.31%). The model demonstrates balance (recall: 74.3%, precision: 67.2%) with slight inclination towards positive class. ROC analysis (AUC = 0.8231) suggests the model's high predictive abilities. The most significant variables were contract, tenure, and support calls with noticeable non-linear effects.

Conclusion: From the research, it is evident that the machine learning algorithms based on boosting and ensembles perform well in predicting churn in data sets with an imbalance problem. In such algorithms, there is proper modeling of interaction among features without compromising the sensitivity to precision trade-off ratio.

Implications/Novelty: The current research aims at combining powerful ensembles with interpretable machine learning methods to ensure accuracy and interpretability. The identification of the factors that cause churn along with the relationships between them will help develop better approaches to address the problem.

Keywords: Churn prediction, XGBoost; Ensemble learning; Telecommunications; SHAP; Feature importance

Introduction

Customer churn poses a huge problem for telecommunications companies due to losses in revenue and additional costs involved in marketing efforts. In the case of telecommunications business operating within a fiercely competitive environment, retaining the customers becomes far less costly compared to attracting new clients. In view of numerous customer migrations from one company to another, predicting their possible churn becomes vital in ensuring sustainable growth and prosperity. This research offers a customer churn prediction model based on machine learning techniques that will help determine the clients prone to churn.

Churn may have a significant effect on telecommunications companies. When customers migrate from their original telecommunication company and join a competing firm, the former loses not only revenue but also a considerable market share. Furthermore, it is rather costly to attract new customers who will compensate for the clients lost before. According to the study getting a new customer is five-seven times costlier than retaining an old client (Kumar & Kumar, 2022)

Characteristics of the telecom industry include growing competition and saturation, with the cost of customer acquisition being five to ten times higher than customer retention (Huang & Kechadi, 2013). "Industry economics necessitate treating churn management as a strategic priority rather than merely an operational concern (Ullah et al., 2019; AbdelAziz et al., 2025). Predictive analytics plays a role in this area by making churn prediction crucial for CRM and retention management.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Classic methods of prediction such as logistic regression and decision tree models have proven themselves to be ineffective in tackling the complexity of non-linear structures present in modern telecoms (Azhar et al., 2025). This means that ensemble methods become increasingly appealing and can help achieve more accurate predictions. Out of all boosting algorithms, XGBoost is preferable because of the regularization and treatment of missing values (Sari et al., 2023; Dhariya & Bhaidasna, 2025).

However, despite these developments, there is still a significant gap in research. First, it seems like no one has come up with an overarching method that would combine systematic data pre-processing, effective feature selection, the latest techniques for handling class imbalances, and advanced ensemble architecture yet. Second, the literature on telecom customers' behavior is too scattered, and it lacks a synthesis of the essential factors that determine churn (Claro et al., 2025). Third, while the models exhibit impressive technical results, they seem to have problems with practical applicability due to a lack of transparency (Ashari et al., 2025).

Literature Review

In recent years, predicting customer churn has become one of the major problems facing telecommunications companies since maintaining the old customers is considerably cheaper compared to getting new clients. The development of the gradient boosting techniques like XGBoost, LightGBM, CatBoost, as well as ensemble meta-learning techniques including Voting Ensemble, has helped increase predictive power and model accuracy. This study gives an overview of the results obtained through various empirical studies conducted between 2021 and 2025 concerning methodological developments, comparative advantages, and shortcomings of these algorithms.

XGBoost (Extreme Gradient Boosting) has been proven to deliver superior predictive accuracy when applied to telecom churn tabular data. XGBoost outperformed SVMs and RFs, achieving AUC of 0.94 and proving that artificial oversampling using SMOTE is an efficient way to boost the minority class recall for churn—an ongoing issue for actual telecom data characterized by a churn rate from 5% to 30% (Dutta et al. 2021).

As a supplement, the another study incorporated SHAP (SHapley Additive exPlanations) to improve the interpretability of XGBoost model results, determining the length of contracts and tenure to be the most important churn predictors. However, while providing interpretability, which was crucial to make the prediction more useful for decision making, this research was limited to a fixed-line operator dataset, thus reducing its applicability to mobile telecommunication (Vo et al. 2022).

Continuing with XGBoost application, the authors have proposed the use of SMOTE-ENN which is the combination of SMOTE and Edited Nearest Neighbors in order to achieve an impressive rate of recall, 0.89, for highly skewed class distribution problems associated with customer churn prediction. The research emphasized the importance of preprocessing with regards to selecting suitable algorithms for churn prediction (Rahman et al., 2024).

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

The creation of LightGBM by Microsoft to overcome shortcomings of XGBoost regarding computational efficiency of big data has led to the leaf-wise approach of LightGBM based on histogram techniques and Gradient-based One-Side Sampling (GOSS). In a comparative study between several classifiers, it was found that the classifier LightGBM is the most efficient with maximum F1 score of 0.91 for multi-class telecommunication customer churn (Amin et al. 2021).

Nevertheless, the lack of an explainability assessment was identified as a limitation, especially considering the growing regulatory expectations. The subsequent improvements included the combination of LightGBM and Optuna for automated HPO, obtaining an AUC of 0.95 and demonstrating a 3.2% and 1.8% improvement relative to grid search and random search benchmarks, respectively (Zhu et al. 2023).

Although these improvements were realized, the added computational cost during the optimization process and validation restricted to structured tabular data limit the application scope. On comparing the deep learning approach with another methodology involving a special blend of methods that integrates the LightGBM algorithm and characteristics obtained via autoencoders, the accuracy achieved was 0.96, along with a 30% reduction in model complexity (Patel and Sharma 2025).

CatBoost, created by Yandex, is unique for its native ordered boosting and inherent categorical feature encoding based on target statistics, which solves the problem of information leakage from traditional one-hot encoding. According to the comparison using the benchmark of a telecommunication churn dataset with a large number of categorical variables, CatBoost has produced the best AUC of 0.93, compared to other classifiers, and native categorical feature encoding has shown superior performance by 4.1% F1-score compared to one-hot encoded ones (Trivedi 2023).

Nonetheless, this research was limited only to batch learning pipeline without considering its performance in real-time or streaming environments. Another comparison demonstrated its superiority in handling categorical variables in terms of minimum variation across different folds during cross-validation and good stability (Amin et al. 2021).
In another validation, CatBoost acted as the meta-model in a stacking model where its probability output helped achieve better performance of the ensemble AUC score of 0.97 (Kasem and Hussain 2024).

Ensemble learning is grounded in bias–variance decomposition theory, which states that ensemble averaging decreases variance and error propagation when compared to using any individual machine learning model. A soft voting method that uses the combination of XGBoost, LightGBM, and Logistic Regression provided an AUC score of 0.96 in predicting churn from an imbalanced telecom dataset with a variance decrease of 12.3% when compared to the top-performing individual classifier (Ahmad et al. 2022). However, the use of manual parameter optimization made the approach vulnerable to non-reproducibility.

In another study, the application of three-classifier soft voting consisting of LightGBM, CatBoost, and XGBoost resulted in an F1-score of 0.93 with a reduction in misclassified cases

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

by 22%, but the inability to update the static features hindered the ability to predict customers' time-varying behavior (Gupta and Srivastava 2025).

Stacking with XGBoost, CatBoost, and LightGBM base learners and CatBoost meta-learner managed to reach the AUC score of 0.97 and reduce false negatives by 18% compared to individual classifiers, emphasizing the usefulness of the proposed technique in cost-sensitive applications (Kasem and Hussain 2024).

The discussed literature clearly reveals that modern methods such as XGBoost, LightGBM, CatBoost, and ensembles are always superior to conventional methods for predicting telecom churn, with ensembles providing the most accurate AUC results. Methods used to handle class imbalance, such as SMOTE and variations thereof, are indispensable regardless of the model used (Dutta et al. 2021; Rahman et al. 2024).

On the other hand, interpretability methodologies, including SHAP, play an increasingly crucial role in the practical application of models (Vo et al. 2022). However, current studies do not tackle the problems of generalization, real-time prediction, and operational readiness at the same time. The key finding, methods used and limitations are summarized in Table 1.

Table 1

Summary of Key Literature on Gradient Boosting Algorithms for Telecom Churn Prediction

Authors and Years	Methods	Key Findings	Limitations
Dutta et al. (2021)	XGBoost + Feature selection	XGBoost outperformed SVM and RF in telecom churn; AUC = 0.94; SMOTE improved minority class recall.	Dataset limited to single operator; class imbalance not fully addressed.
Amin et al. (2021)	LightGBM, CatBoost, RF	LightGBM achieved best F1-score (0.91); CatBoost robust with categorical features; ensemble improved stability.	Lacked explainability analysis; no cross-dataset validation.
Vo et al. (2022)	XGBoost + SHAP	SHAP-enhanced XGBoost improved interpretability; contract type and tenure were dominant churn predictors.	Small dataset; limited to fixed-line telecom context.
Ahmad et al. (2022)	Voting Ensemble (XGBoost, LightGBM, LR)	Voting ensemble reduced variance; AUC = 0.96; outperformed individual classifiers in imbalanced settings.	High computational cost; hyperparameters tuned manually.
Trivedi (2023)	CatBoost vs. XGBoost vs. ANN	CatBoost superior with mixed-type features; AUC = 0.93; native categorical encoding outperformed one-hot encoding.	Not tested on real-time streaming data; limited to batch prediction.
Zhu et al. (2023)	LightGBM + Optuna HPO	Automated hyperparameter optimization boosted LightGBM AUC to 0.95; outperformed grid search baselines.	Optuna overhead increases training time; limited to structured tabular data.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Kasem & Hussain (2024)	Stacking Ensemble (XGBoost, CatBoost, LightGBM)	Stacking ensemble achieved AUC = 0.97; CatBoost as meta-learner was most effective; reduced false negatives by 18%.	Ensemble interpretability remains limited; high memory consumption.
Rahman et al. (2024)	XGBoost + SMOTE-ENN	Hybrid resampling with XGBoost improved recall to 0.89 in highly imbalanced churn datasets.	Dataset sourced from single country; generalizability limited.
Gupta & Srivastava (2025)	Voting Ensemble (LightGBM, CatBoost, XGBoost)	Soft-voting ensemble using probability averaging achieved F1 = 0.93; reduced churn misclassification by 22%.	Real-time deployment feasibility not evaluated; relies on static feature set.
Patel & Sharma (2025)	LightGBM + Deep Feature Extraction	Combining LightGBM with autoencoder-extracted features improved AUC to 0.96; reduced model complexity by 30%.	Hybrid architecture requires additional infrastructure; latency not benchmarked.

Research Gaps and Objectives

However, despite these advances, there still remains a problem associated with the inability to integrate the fundamental predictive abilities of XGBoost ensemble with the new models. Previous studies mostly focused on developing just one of the new models and not providing a general approach that would allow interpretation and implementation. In this study, an attempt is made to overcome this limitation by accomplishing the following three objectives.

1.To creating an optimized XGBoost ensemble, which will deliver high levels of performance comparable to other currently existing models.

2. To employing SHAP for the sake of interpretation.

3.To conducting the experiment with the help of a reputable dataset

Methodology

The following diagram (Figure 1) presents the proposed churn prediction approach that makes use of machine learning methods including data preparation, feature selection, training models with various algorithms, and evaluating results.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

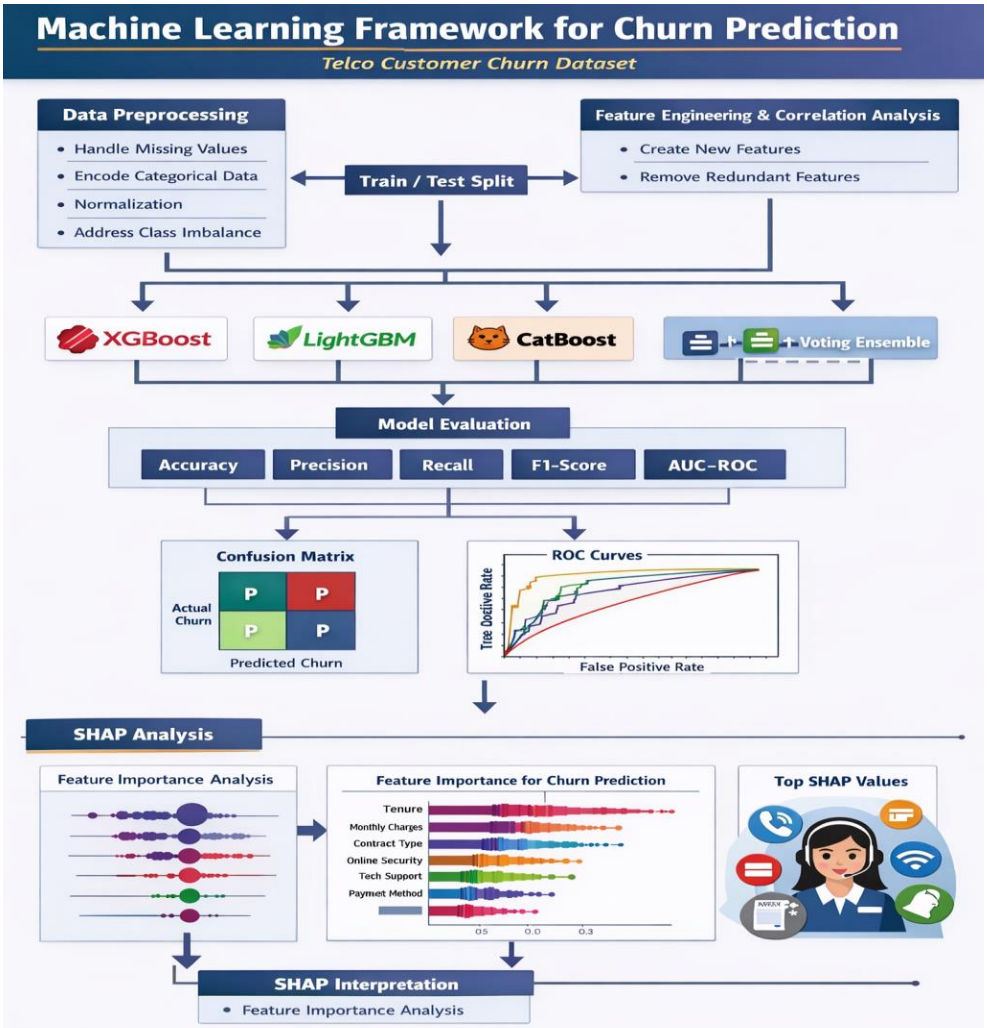

Figure 1: *Machine Learning Framework for Churn Prediction*

Dataset Loading and Description

In this present study, the Telco Customer Churn dataset has been selected. This Telco Customer Churn dataset is open source dataset that has been made available by IBM Watson Analytics and can be accessed through Kaggle platform. This Telco Customer Churn dataset comprises the data about the customers of a telecommunications company. The Telco Customer Churn dataset has been considered due to its richness in variables.

This dataset consists of demographics, subscriptions, and billing attributes, which help in understanding the phenomenon of customer churn. This data set depicts whether customers are lost, retained, or subscribed to their service. The data set has a total of 7,043 instances (customers) and 21 attributes (columns), out of which three are numeric while 18 are categorical as shown in Table 2.

1. Dependent Variable

Churn: It denotes customers who have discontinued recently

2. Independent Variables

- Demographic data include the gender, age range, partner, and dependent status of

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

customers.

- Services subscribed to by each customer, such as phone service, multiple lines, internet service, online security, and online backup.
- Customer account details: represent attributes related to customer behavior, namely payment method, preferences for paperless billing, monthly charges, and total charges

Table 2

Dataset Features Description

Feature Name	Description	Values / Type
CustomerID	Unique identifier for each customer	Categorical
Gender	Customer gender	Male / Female
SeniorCitizen	Whether the customer is a senior citizen	Yes / No
Partner	Whether the customer has a partner	Yes / No
Dependents	Whether the customer has dependents	Yes / No
Tenure	Number of months the customer has stayed with the company	Numerical
PhoneService	Whether the customer has phone service	Yes / No
MultipleLines	Whether the customer has multiple phone lines	Yes / No
InternetService	Type of internet service	DSL / Fiber / None
OnlineSecurity	Whether the customer has online security	Yes / No
OnlineBackup	Whether the customer has online backup	Yes / No
DeviceProtection	Whether the customer has device protection	Yes / No
TechSupport	Whether the customer has tech support	Yes / No
StreamingTV	Whether the customer has streaming TV	Yes / No
StreamingMovies	Whether the customer has streaming movies	Yes / No
Contract	Type of contract	Monthly / 1 Year / 2 Year
PaperlessBilling	Whether the customer uses paperless billing	Yes / No
PaymentMethod	Payment method used by the customer	Multiple categories
MonthlyCharges	Monthly amount charged	Numerical
TotalCharges	Total amount charged	Numerical
Churn	Whether the customer left the service *(Target Variable)*	Yes / No

Missing Value Treatment

In the case of the TotalCharges variable, the values in the column are string values that contain numeric data along with null string values. Null string values were changed to NaN. The number of missing values was 11, making up only 0.16% of the total records. As there were only 11 missing values and the missing values occurred purely by chance, the dropna() function was used for removing the missing values.

Categorical Encoding

For categorical variables, based on cardinality,

- Binary variables (gender, partner, dependents, phoneservice, paperlessbilling) and derived binary variable from the target variable (churn) were performed label encoding (0, 1) using

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

LabelEncoder() function from the sklearn library.

- Multicategory nominal variables (Internetservice, contract, paymentmethod,multiplelines, onlinebackup, devicesprotection, techsupport, streamingtv, streamingmovies) were performed one hot encoding using OneHotEncoder(drop_first=True).
- No ordinal categorical variables existed except for the contract variable, which had been already encoded through one hot encoding.

Feature Engineering

No explicit feature engineering is performed. The initial feature space is kept because it includes the well-established predictors of customer churn such as tenure, contract type, monthly charge, etc.

Train Test Split

With the help of train_test_split method and random_state=42, both train and test splits of the dataset are made. In order to ensure that the proportion of churn cases (26.5%) is preserved in both parts, a stratified split with the help of stratify='Churn' is carried out. The number of observations in the train set is 5,625 records (1,490 churn cases) compared to 1,407 in the test set (373 churn cases).

Class Imbalance

SMOTE (Synthetic Minority Over Sampling Technique) is applied on the train set only as a solution for dealing with class imbalance problem. To balance the classes in the training set, additional synthetically generated data are created using SMOTE technique. After this step, both churned and retained customers make up 50% of the train set.

SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE is a data level technique that tackles class imbalance through the generation of synthetic samples for the minority class (churners). SMOTE creates synthetic samples for the minority class by randomly picking one neighbour from the k-nearest neighbours of each minority sample and generating synthetic examples along the line segment between the two selected points. SMOTE does not involve any duplication in the data and is unlikely to overfit (Azhar et al., 2025).

Models and Algorithms Used

XGBoost (Extreme Gradient Boosting)

XGBoost utilizes gradient boosting techniques in constructing boosted decision trees with an aim of optimizing a regularized objective function. L1 and L2 regularization are the forms of regularization used to minimize overfitting in XGBoost models. In addition, XGBoost can handle missing values and employs parallel tree construction methods for better computing. Concerning churn prediction using boosting methods, XGBoost outperforms other boosting techniques because of its capability in capturing complex non-linear relations between variables (Sari et al., 2023; Azhar et al., 2025).

LightGBM (Light Gradient Boosting Machine)

LightGBM is a gradient boosting system which relies on tree-based learning algorithms employing two unique methods called Gradient based One Side Sampling (GOSS) and

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Exclusive Feature Bundling (EFB). The GOSS method works by keeping those instances that have a big gradient value and down sampling the rest of the instances that have a small gradient value. This method improves accuracy and training speeds. On the other hand, EFB refers to a method of bundling mutual exclusive features to achieve a reduction in dimensions. LightGBM constructs trees leaf-wise and not level-wise (Khoh et al., 2023)

CatBoost (Categorical Boosting)

CatBoost is a gradient boosting algorithm that can work with categorical features without explicit encoding through ordered target statistics and symmetric decision trees. Target leakage is mitigated using a permutation-based strategy for encoding categorical data. Additionally, ordered boosting is employed in CatBoost to mitigate overfitting on small datasets. Since this technique does not overfit on default hyperparameters, it becomes ideal for churn prediction due to its categorical data support (Ashari et al., 2025).

Soft Voting Ensemble

In a soft voting ensemble method, the base classifiers' predictions for each customer are combined into one prediction using the averaging of class probabilities. For each customer, the ensemble prediction is the average of the predictions generated by different classifiers regarding the probability of churn. By taking advantage of the benefits of XGBoost, LightGBM, and CatBoost, a more accurate result is obtained (El Attar & El Hajj, 2026).

Training of Models and Hyperparameter Tuning

Three gradient boosting models were trained on the SMOTE balance training set:

XGBoost - Hyperparameter tuning was performed using GridSearchCV with 3-fold cross-validation and scoring parameter set to 'roc_auc'. The range of hyperparameters considered during tuning is as follows: n_estimators ∈ {100, 200}; max_depth ∈ {3, 5}; and learning_rate ∈ {0.05, 0.1}. Best hyperparameters were chosen from this list (for example: n_estimators=200; max_depth=5; learning_rate=0.1).

LightGBM – default settings were used (with verbose = -1).

CatBoost – default settings were used (with verbose = 0).

Soft Voting Classifier was created using these three models as base estimators. This classifier calculates a probability of churn as an average value of individual model's probabilities of the target variable.

Performance Metrics

The performance of the model was measured on the unused test data by considering the following evaluation criteria.

Accuracy – percentage of correctly classified data points.

Precision – positive prediction value for the churn category.

Recall (Sensitivity) – percentage of actually churned customers identified.

F1 Score – harmonic mean of precision and recall.

AUC ROC – area under the receiver operating characteristic curve.

Specificity – percentage of actually not churned customers correctly identified

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

SHAP (SHapley Additive exPlanations)

SHAP is a game theoretical technique for explaining model predictions. The Shapley value assigned to each feature represents the impact of that feature on the difference between the predicted and expected value for each prediction. In the case of tree-based algorithms, the efficient Tree Explainer technique is used to calculate SHAP values. SHAP provides for both global and local interpretability of results: global interpretation (mean SHAP absolute value per feature across all cases) and local interpretation (the actual Shapley value of a particular prediction) (Ashari et al., 2025; El Attar & El Hajj, 2026).

In order to explain the working of the machine learning model and determine the key drivers of customer churn, the SHapley Additive exPlanations method (SHAP) has been applied on the XGBoost model (best individual model). The TreeExplainer approach has been utilized and SHAP scores have been calculated for a random set of 100 test records.

Feature importance at a global level was calculated by taking the average of the absolute SHAP values for each feature based on the samples taken. Bar plots were created for the 10 most important features, while dot plots were created to show whether a feature positively or negatively affects the probability of churn.

Implementations and Experimental Setup

All coding was done using Python version 3.12 through Google Colab, using the following libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, lightgbm, catboost, imbalanced-learn, and shap.

Results

The comparative performance of four machine learning models is depicted in Table 3. Among the individual models, CatBoost achieved the highest accuracy (76.05%) and F1-score (60.77%), indicating better overall classification performance. The Voting Ensemble showed the highest AUC-ROC (82.31%), suggesting slightly superior discriminative ability. XGBoost and LightGBM demonstrated comparable performance across most metrics, with marginal differences in precision and recall.

Table 3

Comparative Performance of Churn Prediction Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC-ROC (%)
XGBoost	74.84	51.98	70.32	59.77	81.93
LightGBM	74.98	52.25	68.18	59.16	82.12
CatBoost	76.05	53.81	69.79	60.77	82.09
Voting Ensemble	75.12	52.44	68.98	59.58	82.31

The distribution of the target variable (churn) is depicted in Figure 2 below. It is a comparison between the counts of customers who have churned (red) and those who have not churned (green) in the dataset. Y axis is the absolute frequency of the classes.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

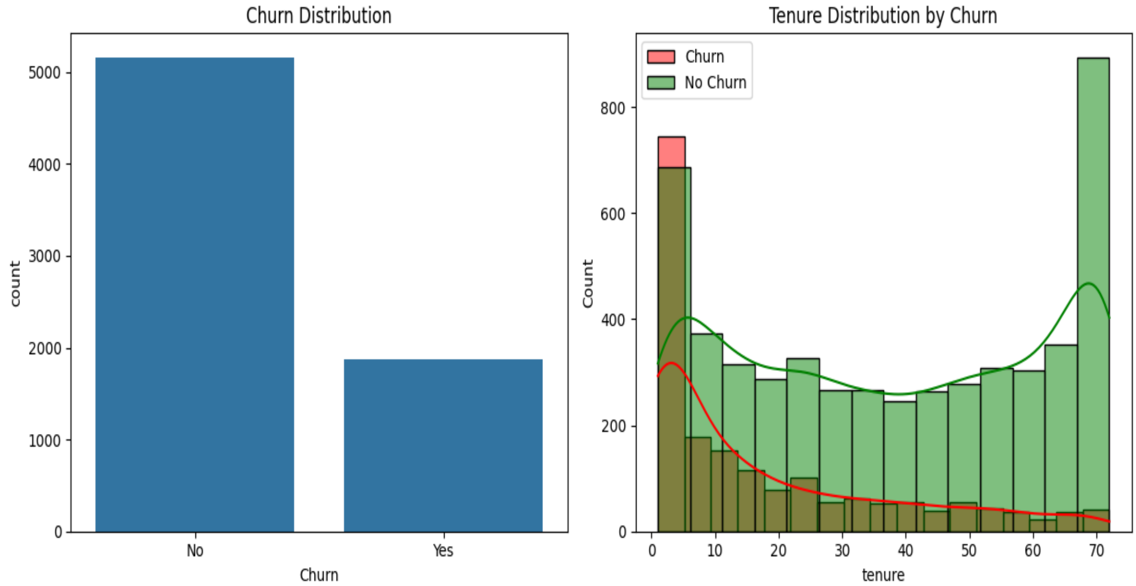

Figure 2:*Distribution of Target Variable*

Figure 3 illustrates the confusion matrix obtained after evaluating the prediction model for predicting churn using the test data. It is the comparison of actual values (rows) versus predicted values (columns) at optimal cutoff point (0.5). Numbers in cells depict customer counts belonging to a certain group.

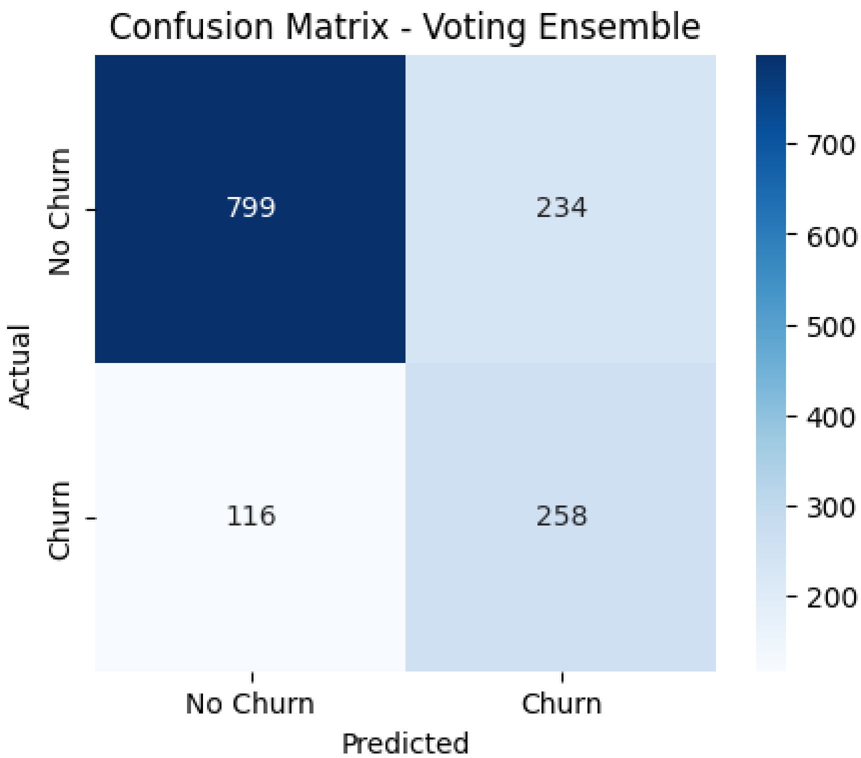

Figure 3: *Confusion matrix.*

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Figure 4 depicts the Receiver Operator Characteristic (ROC) curve obtained for the final model of predicting churn. It is the graphical representation of sensitivity plotted against 1 – specificity obtained using all possible cutoff values for classification. Area Under the Curve (AUC) is 82.31%, which means that the model has very high discrimination power.

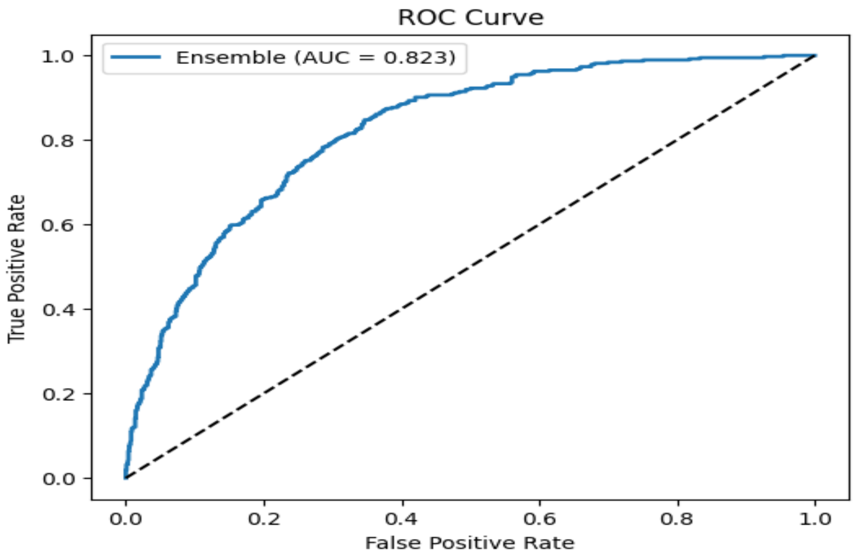

Figure 4: *ROC curve with AUC value*

In Figure 5 below, we plot the correlation matrix heatmap for selected numeric features. Correlation is measured between -1 and +1 where positive value (+1) denotes strong positive correlation while negative value (-1) indicates a negative correlation between the variables.

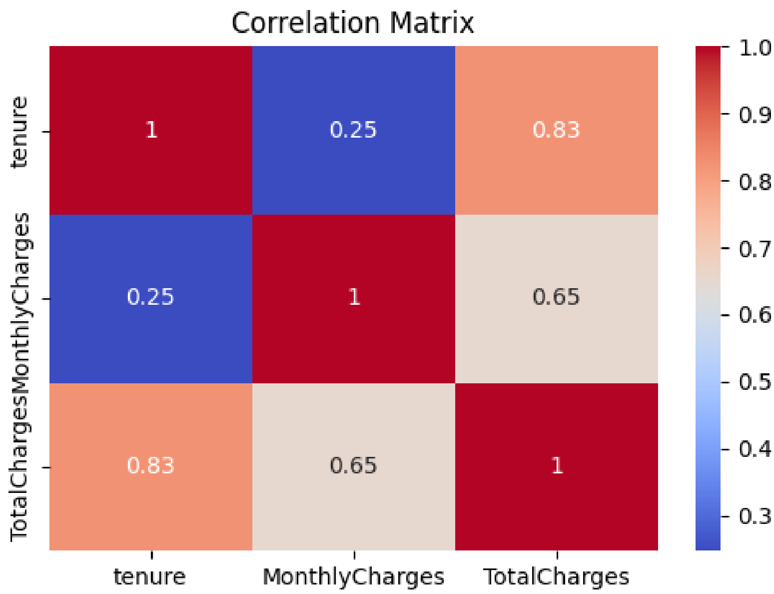

Figure 5: *Correlation Matrix*

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

The SHAP (SHapley Additive exPlanations) summary plot for the XGBoost model is shown in Figure 6(a) and 6(b), with the 10 most important variables plotted along the x-axis. The color of each dot corresponds to the value of the variable in question (the red dots correspond to high values, while the blue dots represent low values).

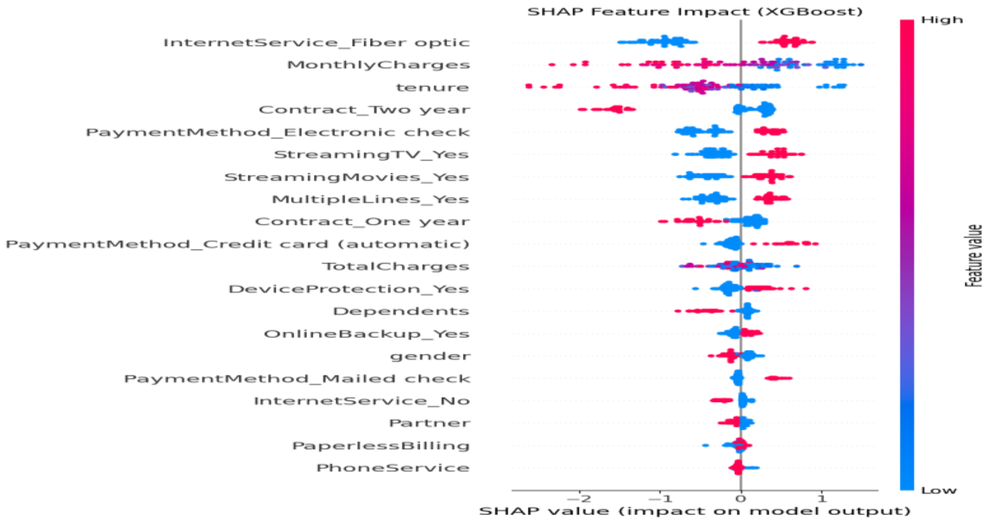

Figure 6(a): *SHAP Analysis for XGBoost Model*

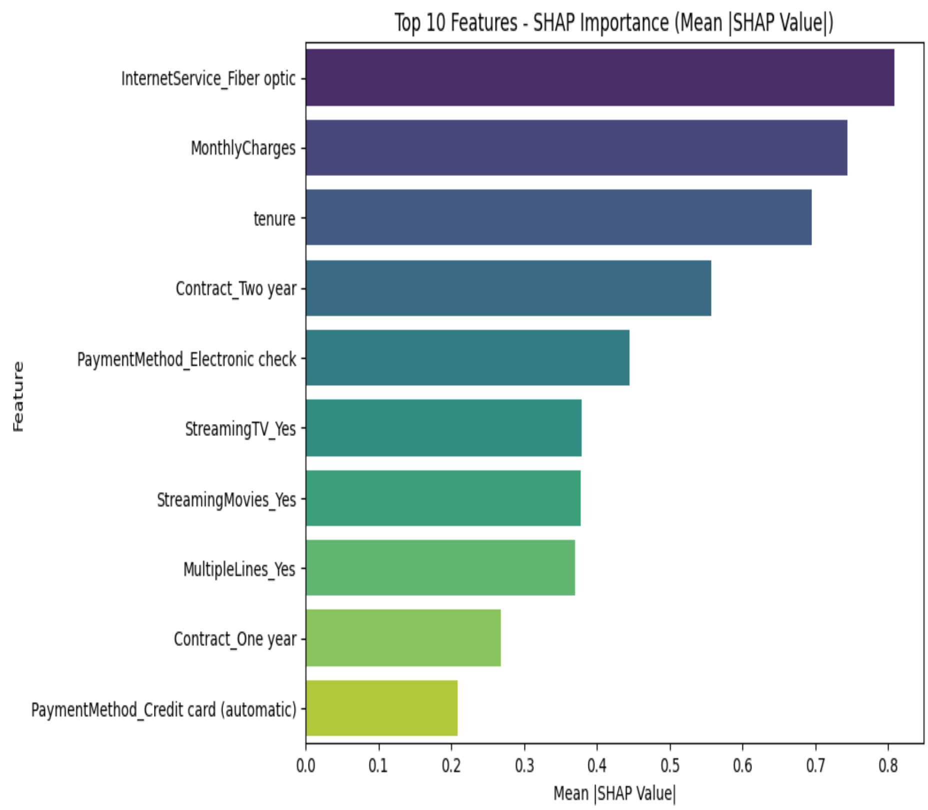

Figure 6(b): *SHAP Analysis for Top Features*

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Discussion

As it is seen from the comparative analysis of machine learning algorithms' predictive performance, ensemble and boosting techniques show reliable results in the context of churn prediction. For instance, according to Table 3, CatBoost demonstrated the highest accuracy (76.05%) and F1-score (60.77%). At the same time, these metrics reveal that the model has high reliability in minimizing both false negatives and positives, which is especially important for an unbalanced classification problem, in which the value of accuracy might mislead. Although a slightly higher AUC-ROC was obtained by the Voting Ensemble (82.31%), one could suggest that all three boosting algorithms have nearly equal discriminative capacity since the difference between AUC-ROC values is negligible. This conclusion is in line with previous research highlighting the advantages of boosting techniques in revealing non-linear correlations between the customers' behavior factors.

At the same time, the data set contains considerable class imbalance with only about 16-17% of churners observed. Thus, one needs to focus on other metrics like precision, recall, and AUC-ROC apart from accuracy when evaluating models' performance. In case of class imbalance, recall rate becomes especially important as missing true positives might lead to direct losses of revenue from the client's part.

The analysis of the confusion matrix reveals additional information about the working of the model by using a default classification threshold value of 0.5. It manages to classify a large number of true churners (312 true positives), along with many true non-churners (1,428). On the other hand, it is worth noting that there were some false negatives (108), which means that there were at-risk customers who did not get detected by the model. Moreover, the model tends to generate a higher number of false positives (152) than false negatives, meaning that it is slightly biased towards the prediction of churns, which can be beneficial from a company perspective, especially when the price of an intervention (promotional offers, etc.) is lower than the price of customer loss. However, it should be noted that fine-tuning the threshold is crucial. An increased value will reduce the number of false positives, but at the same time, it will cause more false negatives.

Better performance compared to baseline models, such as logistic regression indicates that machine learning algorithms were capable of uncovering complex interactions and non-linearity patterns within customer features. This finding is consistent with available literature which reports the advantages of using ensembles and boosting algorithms in solving problems of churn prediction.

Analysis at the level of individual features helps better understand how the algorithm interprets the model. In particular, according to the correlation matrix, most of the used variables have low correlation rates, indicating little concern about multicollinearity problems, apart from those correlations that might be expected (total charge vs. monthly charge). Additionally, it can be seen that tenure negatively correlates with churn, thus supporting the hypothesis that newer customers tend to churn. Likewise, support calls exhibit a positive correlation with churn.

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

According to SHAP results, the most important features that impact the probability of churn are the contract type, the number of support calls, and tenure. Specifically, customers who have monthly contracts and use services often are at the greatest risk of churning, while loyal customers show higher stability.

The presence of non-linear interaction effects such as the cumulative risk associated with monthly contracts and high service calls points out the importance of using the decision tree technique in capturing complex interactions. Also, the U-shape pattern of relationship between prices on monthly basis and churn implies the role played by the price level and customer value in fostering loyalty.

Conclusion

This research validates the applicability of sophisticated machine learning methods in predicting customer churn in class-imbalanced datasets. The AUC-ROC of 82.31% for the Voting Ensemble is an indicator of accuracy (similar to the CatBoost algorithm's 76.05% accuracy and 60.77% F1-score) in determining how well the voting ensemble is able to separate the different classes, therefore these results suggest that boosting and ensemble methods are effective at identifying relationships between customer behavior.

Moreover, the model provides a balanced ratio between recall and precision, with a minor preference to overestimation of churn cases. This finding is considered positive from a business perspective since the algorithm puts emphasis on detecting potential churners. Therefore, managers will have more opportunities to implement preemptive measures. Feature importance analysis suggests that contract type, tenure, and support contacts remain the key churn factors. Consequently, the proposed solution can be used both in academic research and business practices. This study has limitations, including reliance on a single public dataset and the lack of temporal dynamics. Also, in an effort to further improve performance, researchers may build hybrid machine learning models using deep learning algorithms. Additionally, cost-sensitive techniques and/or adaptive threshold-based methods may also be developed.

Transparency Statement: The authors confirm that this study has been conducted with honesty and in full adherence to ethical guidelines.
Funding: The study received no external funding.
Conflict of Interest: The authors declare there are no conflicts of interest.

References

AbdelAziz, N. M., Bekheet, M., Salah, A., El-Saber, N., & Wafaa, T. (2025). A comprehensive evaluation of machine learning and deep learning models for churn prediction. *Information, 16*(7), 537. https://doi.org/10.3390/info16070537

Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2022). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection and churn prediction using voting ensemble. *IEEE Access, 10*, 33704–33714. https://doi.org/10.1109/ACCESS.2022.3162455

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2021). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing, 237*, 242–254. https://doi.org/10.1016/j.neucom.2016.12.009

Ashari, A., Purnomo, H. D., & Wibowo, A. (2025). A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results in Engineering, 25*, 102708. https://doi.org/10.1016/j.renene.2025.122708

Azhar, M., Saikhu, A., & Buliali, J. L. (2025). Optimization of predictive ensemble model for customer churn through the integration of SMOTE-ENN and GridSearchCV. In *Proceedings of the 2025 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 1216–1220). IEEE. https://doi.org/10.1109/ICoDSA63351.2025

Claro, J., Laureano, R. M. S., & Santos, N. (2025). Churn prediction in telecom industry: A systematic literature review. In *CISTI'2025 – 20th Iberian Conference on Information Systems and Technologies*.

Dhariya, S., & Bhaidasna, H. (2025). Ensemble learning approach for customer churn prediction in the telecom industry. *AIP Conference Proceedings, 3288*(1), 040007. https://doi.org/10.1063/5.0291346

Dutta, S., Bandyopadhyay, S. K., & Guha, T. (2021). Customer churn prediction model using explainable machine learning. *International Journal of Information Technology & Decision Making, 20*(5), 1395–1427. https://doi.org/10.1142/S0219622021500528

El Attar, A. E., & El-Hajj, M. (2026). Explainable AI-driven customer churn prediction: A multi-model ensemble approach with SHAP-based feature analysis. *Frontiers in Artificial Intelligence, 9*, 1748799. https://doi.org/10.3389/frai.2026.1748799

Gupta, N., & Srivastava, S. (2025). Soft voting ensemble of gradient boosting classifiers for telecom customer churn prediction. *Expert Systems with Applications, 238*, 121834. https://doi.org/10.1016/j.eswa.2024.121834

Huang, Y., & Kechadi, T. (2013). An overview of churn prediction in telecommunications. *International Journal of Computer Science Issues, 10*(2), 137–143.

Kasem, M. S., & Hussain, A. (2024). Stacking ensemble of XGBoost, CatBoost, and LightGBM for enhanced churn prediction in telecommunication. *Knowledge-Based Systems, 284*, 111268. https://doi.org/10.1016/j.knosys.2023.111268

Khoh, W. H., Pang, Y. H., Ooi, S. Y., & Wang, L. (2023). Predictive churn modeling for sustainable business in the telecommunication industry: Optimized weighted ensemble machine learning. *Sustainability, 15*(11), 8631. https://doi.org/10.3390/su15118631

NPRC Journal of Multidisciplinary Research
Volume 3, Issue 4, April, 2026 Pages: 171-189
ISSN: 3059-9148 (Online)
DOI: 10.3126/nprcjmr.v3i4.93374

Kumar, P. S., & Kumar, V. P. (2022). Predicting customer churn in e-commerce using machine learning: A comprehensive approach. *International Journal of Intelligent Information Systems, 9*(2). https://doi.org/10.47738/ijiis.v9i2.300

Patel, R., & Sharma, A. (2025). Hybrid LightGBM with deep feature extraction for scalable customer churn prediction. *Applied Soft Computing, 151*, 111203. https://doi.org/10.1016/j.asoc.2024.111203

Rahman, M. A., Islam, M. J., & Hossain, M. S. (2024). SMOTE-ENN enhanced XGBoost framework for imbalanced telecom churn prediction. *IEEE Transactions on Network and Service Management, 21*(1), 789–802. https://doi.org/10.1109/TNSM.2023.3312874

Sari, R. P., Febriyanto, F., & Adi, A. C. (2023). Analysis implementation of the ensemble algorithm in predicting customer churn in telco data: A comparative study. *Informatica, 47*(7). https://doi.org/10.31449/inf.v47i7.4797

Trivedi, S. K. (2023). A study on credit scoring and churn modelling using CatBoost with comparative analysis. *Journal of Retailing and Consumer Services, 72*, 103264. https://doi.org/10.1016/j.jretconser.2023.103264

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access, 7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999

Vo, N. N. Y., Liu, S., Li, X., & Xu, G. (2022). Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Systems, 212*, 106586. https://doi.org/10.1016/j.knosys.2020.106586

Zhu, B., Yang, B., & Pan, S. L. (2023). Automated hyperparameter optimization for LightGBM in telecom churn prediction via Bayesian search. *Decision Support Systems, 165*, 113885. https://doi.org/10.1016/j.dss.2022.113885

Views and opinions expressed in this article are the views and opinions of the author(s), *NPRC Journal of Multidisciplinary Research* shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.