Open Access ISSN: 3007-1437 (Online), ISSN: 3007-1429 (Print)

Alkadhim Journal for Computer Science (KJCS)

Journal Homepage: https://alkadhum-col.edu.iq/JKCEAS

Enhancing Telecom Customer Churn Prediction Using Voting Ensemble Learning with SMOTE and ADASYN

[1]Omar Shakir Hasan*, [2] Maan Y Anad Alsaleem, [3]Yahya Albugg

[1] Directorate of Educational Nineveh, Mosul, Iraq. Email: omarshakir06@gmail.com

[2] Directorate of Educational Nineveh, Mosul, Iraq. Email: maanyounis1983@gmail.com

[3]Northern Technical University, Mosul, Iraq. Email: dr.yahya.albugg@ntu.edu.iq

Article information

Article history:
Received: April, 26, 2026
Accepted: July, 07, 2026
Available online: September, 25, 2026

Keywords:
Customer Churn Prediction,
Machine Learning ,
Ensemble Learning,
SMOTE,
ADASYN.

Corresponding Author:
Omar Shakir Hasan
omarshakir06@gmail.com

DOI:
https://doi.org/10.61710/tasp1g44

Abstract

Customer churn prediction is an important task in the telecommunications sector, particularly when class imbalance affects the ability of machine learning models to identify churned customers. This study evaluates the performance of several machine learning algorithms using the Telco Customer Churn dataset, which contains 7,043 customer records and 21 features. Five classification algorithms, namely Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and CatBoost, were evaluated. A Voting Ensemble model combining Random Forest, XGBoost, and CatBoost was also developed. To address class imbalance, SMOTE and ADASYN oversampling techniques were applied to the training data. The results showed that ensemble-based models generally outperformed individual classifiers. The Voting Ensemble model combined with SMOTE achieved the best performance, obtaining 97% accuracy, 0.95 precision, 0.93 recall, 0.95 F1-score, and 0.96 ROC-AUC. The findings indicate that integrating ensemble learning with oversampling techniques can improve churn prediction performance in imbalanced telecom datasets.

1.Introduction

Predicting when customers will leave will enable businesses to prepare for losing those customers so they can take the necessary steps to prevent them from losing those customers—such as giving tailored service, sending targeted promotions or providing superior customer service. By being able to accurately predict whether or not a customer will leave, companies can develop better resource allocation practices, as well as improve customer loyalty and ultimately increase profitability over time.

For several years now, the telecom industry has created a wealth of demographic data, service usage data, billing data, and histories of customer interactions with their company. All of this cumulative structured information provides an opportunity for telcos to analyze this data using data mining and machine learning techniques to determine the factors that are associated with whether customers will leave or not.

Machine learning is now commonly used across the world for churn predictions due to its ability to discover intricate relationships within large amounts of data without having to adhere to any set of strict statistical assumptions. One type of machine learning algorithm that has shown to be quite successful in classification tasks where structured data is used is ensemble learning. An ensemble model consists of a group of learners that produce outputs from the same input data; by taking the outputs of numerous individual learners and combining them into one output from the whole ensemble, the overall prediction becomes more reliable and can be extended across a wider variety of circumstances. A number of machine learning algorithms are typically impacted by imbalanced classes in churn predictions (e.g. Where the number of customers who stay with a service provider is much larger than those that leave [7]. Class imbalance causes biased models in that the models are trained on the majority class, which leads to fewer accurate predictions of those customers who left the company. To help resolve this, resampling methods have been created to put additional examples in the training dataset that do not represent actual cases of customer churn; thus improving the performance of the classifier [8].

Recent literature cites use of ensemble methods, feature engineering methods, and data balancing methods in telecom customer churn prediction; however, there exists inconsistencies in the performance of each method across different datasets with varying types of learning configurations and how the various methods of oversampling affect the ensemble model's accuracy for prediction. In addition, there are limited studies that have analyzed the combined impact of Voting Ensemble models and synthetic oversampling methods through a standard experimental design. This research seeks to examine how certain machine learning algorithms (i.e., SVM, Decision Tree, Random Forest, XGBoost, and CatBoost) can provide classification for the Telco Customer Churn dataset. Furthermore, a Voting Ensemble model combining Random Forest, XGBoost, and CatBoost is developed and evaluated under three scenarios: the original dataset, SMOTE-balanced data, and ADASYN-balanced data. The objective is to examine the effect of ensemble learning and data balancing techniques on churn prediction performance using a common evaluation framework based on accuracy, precision, recall, F1-score, and ROC-AUC metrics.

2. Related Work

Machine learning approaches for telecom customer churn prediction can generally be categorized into three groups: traditional machine learning models, ensemble learning methods, and approaches that address class imbalance through data resampling techniques. Recent studies have investigated these directions individually or in combination to improve churn prediction performance.

Machine Learning has been extensively researched as a means to predict customer churn. There are many models developed by many researchers to predict churn as well as many methods to handle imbalanced datasets.

A study in [9] conducted a study on multiple telecommunications datasets including IBM Telco, Churn-in-Telecom, UCI Churn, and Orange Telecom. The study proposed a combined approach to improving churn predictions by using machine learning and deep learning models (XGBoost, Light GBM, LSTM, MLP, SVM) as well as feature engineering, stacking ensemble methods, and the use of SMOTE to handle class imbalance. The ensemble model achieved higher predictive performance than the individual classifiers evaluated in the study.

In [10], researchers also used machine learning models in combination with data balancing strategies to predict telecom customer churn. Researchers reduced the number of attributes in their dataset from 21 to 16 through feature selection and tested several algorithms: Logistic Regression, SVM, KNN, Decision Tree, Random Forest, Naïve Bayes, and XGBoost. They then used SMOTE, ENN, oversampling, and under sampling to address class imbalance and K-fold cross-validation for consistency in evaluating their results. The researchers found that Random Forest produced the highest accuracy (90.90%), showing that feature selection and data balancing contributed to improving churn prediction performance. The results support the conclusion that data balancing techniques can positively affect model quality but that performance of other techniques is still dependent on the choice of classification algorithm and dataset features.

In [11], they describe the development of a stacking ensemble model for predicting customer churn in telecoms using the Churn-in-Telecom dataset found on Kaggle, programming their first level of predicted values

(predictions made with XGBoost) into a second level of prediction (predictions made with Logistic Regression, Decision Tree and Naïve Bayes classifiers) using a method called soft voting. Additionally, feature construction techniques were applied in order to create the best representation of the data characteristics to ensure accuracy in predictions. The resulting ensemble produced better predictions than either the first or second levels of prediction would have produced alone.

In [12], a DNN (Deep Neural Network) based customer churn prediction was evaluated using the IBM Telco Customer Churn dataset of which consists of 7,034 records and 21 variables, and achieved an accuracy of 82.26% and an F1 Score of 0.8209. All pre-processing steps (including: categorical encoding, imputation of missing values, outlier treatment based on IQR thresholds and standardization) have shown the effectiveness of this predictive modelling approach for being able to model complex behavior patterns of customers using deep learning methods.

Authors in [13] evaluated 16 different ML classifiers to predict Telecom customer churn. The classifiers included Logistic Regression, SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Decision Trees, Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Ada Boost, Bagging, Random Forests, Extra Trees, Gradient Boosting, Light GBM, CatBoost, and XGBoost. The authors ran their experiments with an 80/20 training test split and also using 10-fold cross-validation. The results indicate that CatBoost produced the best classification accuracy from all the classifiers tested in this study. The results from this study also suggested that there is generally a performance boost using boosting/ensemble techniques over traditional classifiers for prediction of telecom customer churn; however, the actual difference of performance varied depending on the data set and experimental setting.

Authors in [14] conducted a study using the Telco Customer Churn dataset from Kaggle, which consists of 7043 customer records and 21 attributes. After preprocessing and cleaning, 7010 records were left for analysis. Various classification techniques were then evaluated, including Decision Trees, Random Forests, SVM, and Logistic Regression. The Logistic Regression model produced the highest reported AUC of 0.852; thus indicating how useful the application of machine learning can be to detect factors related to customer churn.

In [15], the authors analyzed a telecom dataset consisting of 10,000 customer records using data mining techniques. Data preprocessing included missing value treatment, feature encoding, normalization, correlation analysis, PCA for feature selection, and SMOTE for class balancing. The authors used many techniques to identify customer groups that are likely to churn: K-means (KM), DBSCAN, Association Rule Mining, Isolation Forest, and LOF. The authors classified customers into three risk levels—low, medium, and high—and showed how data mining techniques can be useful for analyzing customer retention.

The authors in [16] proposed a framework for predicting telecom churn that integrated Genetic Algorithm (GA), XGBoost, and SHAP. ADASYN was used to deal with the class imbalance, and the authors compared multiple machine learning algorithms to develop the best model. The GA-XGBoost model achieved the best accuracy (93.1%) and AUC (95.61%), and the authors used SHAP to demonstrate how each feature contributes to the predicted churn.

Several limitations to current research on churn prediction exist. As noted, current studies have each evaluated one method of balancing samples, or composed of many individual methodologies of balancing samples without systematically comparing the results of different oversampling techniques with different ensemble models. In addition, direct comparisons between Voting Ensemble models and individual classifiers under identical experimental conditions remain limited. Therefore, this study evaluates the impact of SMOTE and ADASYN on both individual machine learning models and a Voting Ensemble framework using a common dataset and evaluation protocol.

3. Methodology

This study proposes a two-tiered stack. The base layer consists of a set of collective learning models, and voting takes place among these models in the second layer. Experiments were conducted using the original dataset as well as a balanced dataset. Figure (1) illustrates the pipeline of the proposed methodology.

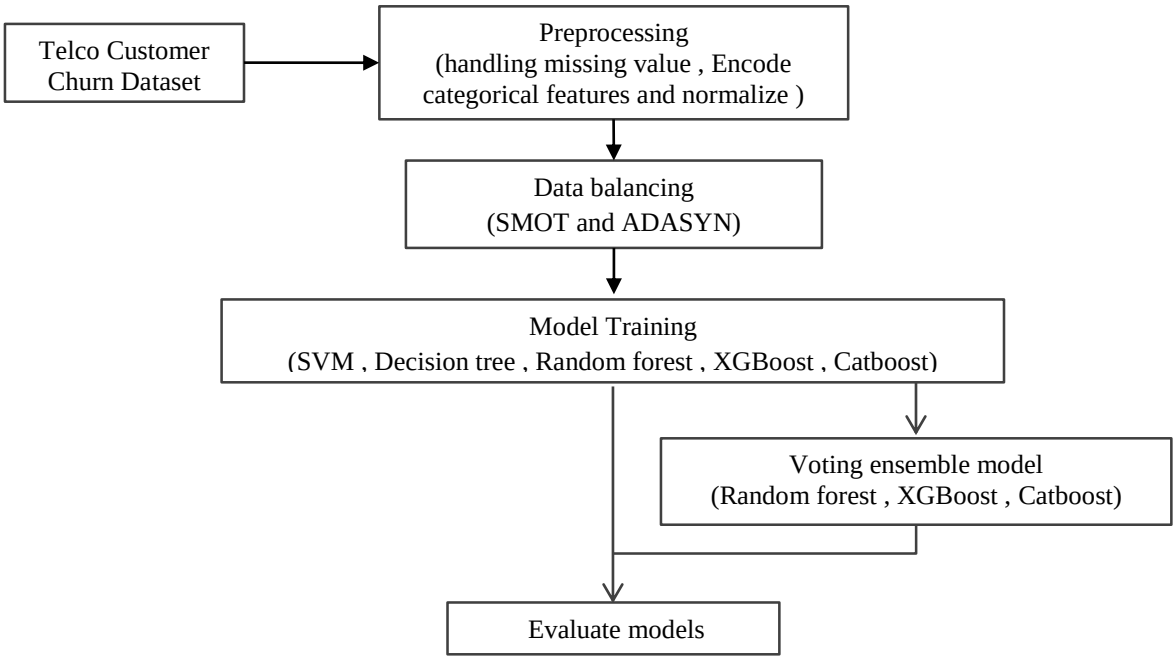


Figure (1): Proposed Methodology

3.1. Dataset Description

The Telco Customer Churn dataset is used in this study to analyze customers from a telecommunication company. The dataset is made up of 7,043 row records, and 21 attributes describe various aspects of the customers like demographic information, account information, services they have signed up for, as well as their billing details. The dataset is moderately unbalanced with around 26% of the customers having churned and 74% of them are still active[16].

3.2 Data Preprocessing

Several data cleansing procedures were undertaken before providing the machine learning algorithms with pre-processed data to ensure that it was in a state that would provide reliable information from which machine learning models can be trained. The first data cleaning procedure involved eliminating an attribute called "customerID" because that attribute was a unique identifier for every customer without any statistical relationship to the attributes in the dataset and, therefore, does not provide any useful information in predicting whether or not a customer will churn. The second data cleaning procedure involved checking the dataset for any missing or incomplete values. Missing values can introduce bias into predictive models or create instability within the model unless they are resolved appropriately. Third, some of the dataset's attributes were categorical (for example, gender, contract type, internet service type, payment method). Therefore, these attributes needed to be converted to numerical values through the use of a technique called One-Hot Encoding. Fourth, the target variable "Churn" originally contained categorical values ("Yes" and "No"), which needed to be transformed into a numerical format in order to conduct classification modeling. Finally, in order to validate the generalization ability of the models, the dataset was divided into training and testing datasets using a standard 80/20 train/test split. To avoid data leakage, all resampling procedures were applied only to the training subset after data splitting, while the testing subset remained unchanged throughout the experiments.

3.3 Machine Learning Algorithms

In this paper, we've examined a variety of machine learning algorithms to determine how effective they are when predicting customer churn within the telecommunications industry. . Some of the different algorithms used in this study include support vector machines (SVM), decision trees, random forests, XGBoosts and CatBoosts. Each of the algorithms have their own unique advantages for performing classification tasks. The support vector machine

(SVM) is capable of working with a large number of dimensions (high-dimensional space) to capture intricate relationships amongst different dimensions with kernel functions[17]. Decision trees are interpretable in their structure, By using a single decision tree, overfitting may occur because the complex nature of the decision tree will be difficult to manage, so ensemble learning methods like random forests to overcome the issue of overfitting [18]. Random forests reduce variance in making predictions by combining multiple decision trees, and therefore, provide a more stable by randomly selecting samples of both data and features for each tree in the forest[19]. Similarly, XGBoost and CatBoost to iteratively reduce the error of the predictions. and minimizing prediction bias[20]

These algorithms were selected to represent different classification paradigms. SVM was included as a margin-based classifier capable of handling high-dimensional feature spaces. Decision Tree was used as an interpretable baseline model. Random Forest represents bagging-based ensemble learning, while XGBoost and CatBoost represent boosting-based ensemble techniques that have shown strong performance on structured tabular datasets. In addition, a Voting Ensemble Model was also used in this study . Voting Ensemble Models will take advantage of ensemble learning, which aggregates the predictions from several single or individual model(s), and produces a final prediction that has a larger level of accuracy and robustness[21]. In this study's , the Ensemble Model includes three of the highest performing prediction algorithms and gives an independent prediction for every specific input data point provided. Lastly, the final classification of that input data point will determine the final classification of that input by taking the majority vote from the vote from each individual model, where the class receiving the most votes among the individual models will be returned as the final classification for that input[22]. Voting Ensembles will help to decrease the variance among the models while improving predictive accuracy by leveraging the strengths from several different models[23].

To ensure reproducibility of the experiments, the main configuration parameters used for the classification models, ensemble framework, and resampling techniques are summarized in Table 1. Unless otherwise specified, default parameter settings provided by the corresponding machine learning libraries were used.

The ensemble combines Random Forest, XGBoost, and CatBoost classifiers. These models were selected because they achieved strong individual performance and employ different ensemble learning strategies. The final prediction is determined through majority voting, where the class receiving the highest number of votes is selected as the final output. The objective of this configuration is to reduce model-specific prediction errors and improve classification robustness.

Table 1. Experimental Configuration and Model Parameters

Model	Parameters
SVM	Kernel = RBF, C = 1.0, Gamma = scale
Decision Tree	Criterion = Gini, Max Depth = Default
Random Forest	n_estimators = 100, Criterion = Gini, Random State = 42
XGBoost	n_estimators = 100, Learning Rate = 0.1, Max Depth = 6, Random State = 42
CatBoost	Iterations = 100, Learning Rate = 0.1, Depth = 6, Verbose = False
Voting Ensemble	Base Models = Random Forest, XGBoost, CatBoost; Voting = Hard Voting
SMOTE	k_neighbors = 5, Sampling Strategy = Auto
ADASYN	n_neighbors = 5, Sampling Strategy = Auto

3.4 Handling Imbalanced Data Using SMOTE and ADASYN

Customer churn prediction has a major hurdle with respect to class imbalance. That is, many real-world datasets (such as the Telco Customer Churn dataset) have far more customers who stay with the company than there are customers who leave the company. Because of this class imbalance, machine learning models can suffer from improper training since most algorithms will learn better from majority class customers than minority class customers and therefore the model can be highly accurate with regard to all customers while performing poorly with regard to the minority class (those who left the company). To address this issue, two of the most commonly

used methods of oversampling were used in this study to balance out the class distribution of the dataset: SMOTE (Synthetic Minority Oversampling Technique) and ADASYN (Adaptive Synthetic Sampling). By generating synthetic samples of customers (minority class) who left the company, these methods enable machine learning models to better understand the churn pattern amongst the minority class. For both SMOTE and ADASYN, the number of nearest neighbors was set to k = 5. Resampling was applied only to the training data until a balanced class distribution was obtained. The same configuration was used across all experiments to ensure a fair comparison between balancing techniques.

3.4.1.SMOTE (Synthetic Minority Oversampling Technique)

To balance the number of instances in a dataset (the class distribution) so that both classes will have an approximately balanced number of instances, synthetic samples can be created for minority classes by interpolating between two existing Instance classes[27]. Rather than creating duplicates of the data points present in your database, you can apply the synthetic data generation method (i.e., SMOTE) to the original data by selecting the records in your class that are in the minority and then selecting their k nearest Neighbors in your feature space. By doing this, you can create a diverse set of synthetic data points, reducing the possibility of creating a biased model through sample oversample methods[28].

3.4.2.ADASYN (Adaptive Synthetic Sampling)

ADASYN is an expansion of the SMOTE technique which aims to assist in the creation of (synthetic) new observations (or samples) that match those observations in the underlying dataset that are particularly challenging to learn in the minority class. Many of the synthetic observations created by SMOTE are generated uniformly across all minority samples regardless of how difficult or easy they are to learn[29]. In contrast, the ADASYN algorithm takes the level of difficulty experienced while learning a sample and adapts the amount of synthetic data that is created for that sample accordingly[30]. Therefore, if a certain portion of the minority class is located in an area with fewer samples than others or if the boundary of classification within that minority class is more difficult to learn, more synthetic data will be generated in that area as compared to those areas of the minority class that are easier to learn[31]. This adaptive approach to generating synthetic data will assist the training algorithms by placing stronger emphasis on the harder-to-learn samples and reducing bias towards those samples of the majority class[32]. Thus, by using the ADASYN technique, models can improve the likelihood of correctly identifying churn customers when trained with a severely imbalanced data set. This paper implemented both the SMOTE and ADASYN methods of synthetic data generation when training the prediction performance models from the training data set and explored how each technique contributed to prediction performance improvements for each model.

3.5.Evaluation Metrics

In order to evaluate the performance of the classification models, we used several standard metrics such as accuracy, precision, recall and F1-score; all of which are derived from the confusion matrix that contains true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN)[33][34].

- Accuracy: Accuracy evaluates the total ratio of classified instances that were classified correctly out of all samples. In essence.
- $$\text{Accuracy} = \frac{\text{TP+TN}}{\text{Total}} \ldots\ldots\ldots\ldots (1)$$

  TP (True Positive): Correctly predicted positive cases
  TN (True Negative): Correctly predicted negative cases
  Total: Total number of samples (TP + TN + FP + FN)

- Precision :Precision measures the proportion of correctly predicted positive cases among all instances predicted as positive. It evaluates the model's ability to avoid false positive errors.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \ldots\ldots\ldots\ldots (2)$$

- Recall: Recall, also known as sensitivity, measures the proportion of actual positive cases that are correctly identified by the model. It evaluates the model's ability to detect positive instances. A high recall value indicates that the model successfully captures most of the true positive cases, minimizing false negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \ldots\ldots\ldots\ldots (3)$$

- F1-Score:The F1-score is the harmonic mean of precision and recall. It provides a balanced measure that considers both false positives and false negatives.

$$\text{F1} - \text{score} = \frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}} \quad \ldots\ldots\ldots\ldots (4)$$

- Receiver Operating Characteristic (ROC) Curve :The Receiver Operating Characteristic (ROC) curve is a graphical evaluation tool used to assess the diagnostic ability of a binary classification model. It illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across different classification thresholds.

  1. True Positive Rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \quad \ldots\ldots\ldots\ldots (5)$$

  2. False Positive Rate (FPR)

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \quad \ldots\ldots\ldots\ldots (6)$$

Since churn prediction is an imbalanced classification problem, Recall and F1-score are particularly important for evaluating model effectiveness.

3.6. Feature Importance Analysis

XGBoost was utilized to conduct a feature importance analysis to assist in identifying factors that affect customer churn. This technique provides insight into what variables drive customer behavior. Variables having high feature importance will have the greatest impact on the outcome of the predictive model. As shown in Figure 1, the top 15 most important features identified by the XGBoost model are as follows: Monthly Charges are significantly more important in relation to predicting customer churn than any other variable (highest importance). The next most important predictor of customer churn is tenure.

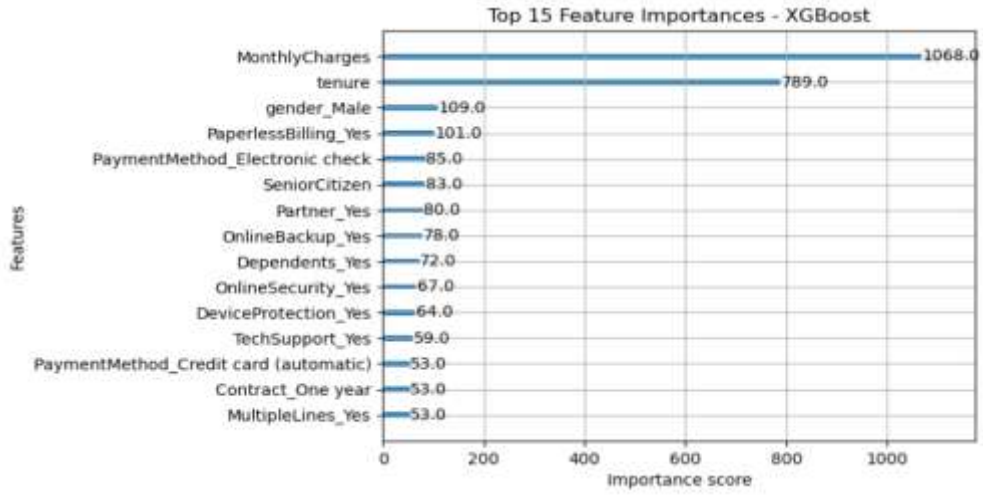


Figure (2): Top 15 Feature Importance obtained using the XGBoost model

4. Results

Table 2 displays the comparison of performance for various machine learning algorithms based on precision, recall, and F1-score metrics for each class as well as overall accuracy scores. Random Forest and CatBoost had a better performance than both SVM and Decision Tree individually. The models combined in the Voting Ensemble along with XGBoost, Random Forest, and CatBoost presented superior performance than the individual models. Additionally, /further enhancements in predicting performance were observed as a result of implementing data balancing techniques. The Voting Ensemble using the ADASYN method achieved a 92% accuracy rating; however, the SMOTE method provided the best results, resulting in a 97% accuracy rating with excellent precision, recall, and F1-scores on both classes. Comparison charts for precision, recall, F1-score, and accuracy for each of the individual models and the ensemble models under the SMOTE and ADASYN methods are presented in Figures 3 - 8.

Table (2): Performance Comparison of Machine Learning Models

Model	Class 0 Precision	Class 1 Precision	Class 0 Recall	Class 1 Recall	Class 0 F1	Class 1 F1	Accuracy
SVM-Original	0.85	0.69	0.83	0.66	0.82	0.66	0.88
SVM_SMOTE	0.89	0.78	0.88	0.74	0.88	0.76	0.90
SVM_ADASYN	0.88	0.76	0.86	0.75	0.87	0.75	0.89
DecisionTree_Original	0.82	0.67	0.80	0.67	0.83	0.73	0.87
DecisionTree_SMOTE	0.86	0.75	0.85	0.74	0.86	0.74	0.89
DecisionTree_ADASYN	0.85	0.74	0.84	0.73	0.85	0.73	0.88
Random Forest_Original	0.84	0.72	0.85	0.73	0.86	0.74	0.89
Random Forest_SMOTE	0.91	0.81	0.92	0.83	0.89	0.80	0.93
RandomForest_ADASYN	0.90	0.79	0.90	0.79	0.88	0.78	0.92
XGBoost_Original	0.87	0.74	0.86	0.75	0.87	0.76	0.88
XGBoost_SMOTE	0.92	0.85	0.93	0.88	0.91	0.84	0.94
XGBoost_ADASYN	0.91	0.83	0.91	0.87	0.90	0.82	0.93
CatBoost_Original	0.88	0.76	0.87	0.77	0.88	0.77	0.89
CatBoost-SMOTE	0.93	0.90	0.93	0.89	0.93	0.90	0.95
CatBoost_ADASYN	0.92	0.89	0.92	0.88	0.92	0.88	0.94
Voting Ensemble_Original	0.91	0.83	0.88	0.83	0.90	0.81	0.89
VotingEnsemble_ADASYN	0.93	0.89	0.93	0.90	0.93	0.89	0.92
Voting Ensemble_SMOTE	0.96	0.95	0.96	0.93	0.97	0.95	0.97

The results indicate that ensemble-based methods generally achieved higher performance than individual classifiers. Among the baseline models, CatBoost achieved the highest accuracy of 95% after applying SMOTE, followed by XGBoost with 94%. The Voting Ensemble combined with SMOTE achieved the best overall performance with an accuracy of 97%, indicating that combining complementary ensemble classifiers can improve churn detection performance. The improvement was particularly evident in the minority churn class, where the F1-score increased to 0.95.

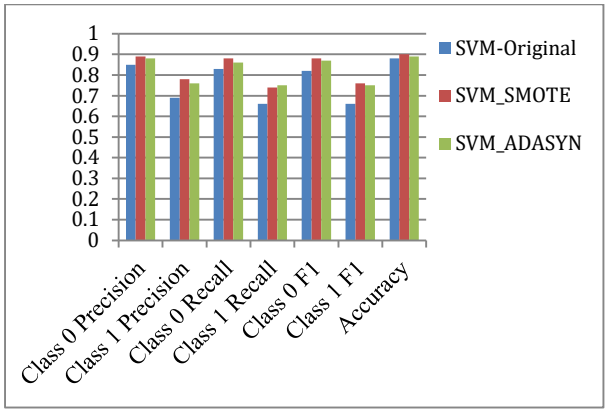


Figure (3): Performance Comparison of SVM Model Model Using Original Data, SMOTE, and ADASYN

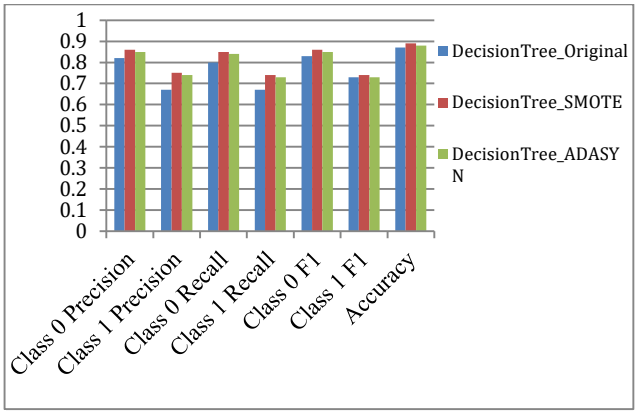


Figure (4): Performance Comparison of Decision Tree model Using Original Data, SMOTE, and ADASYN

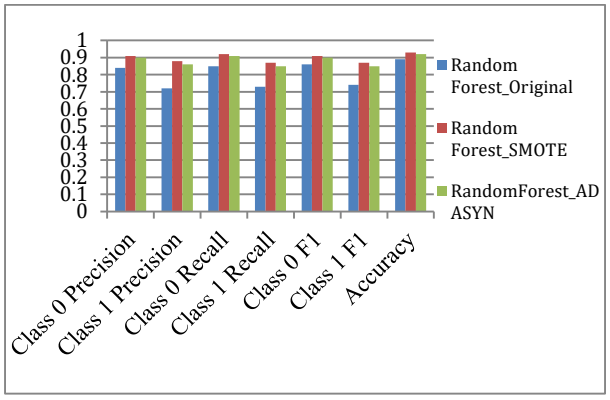


Figure (5): Performance Comparison of random forest model Using Original Data, SMOTE, and ADASYN

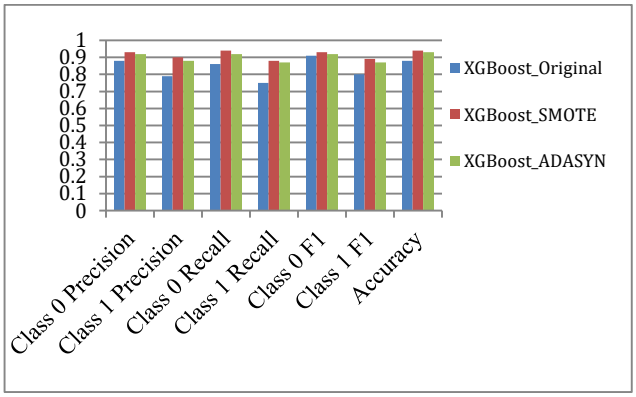


Figure (6): Performance Comparison of XGBoost Model Using Original Data, SMOTE, and ADASYN

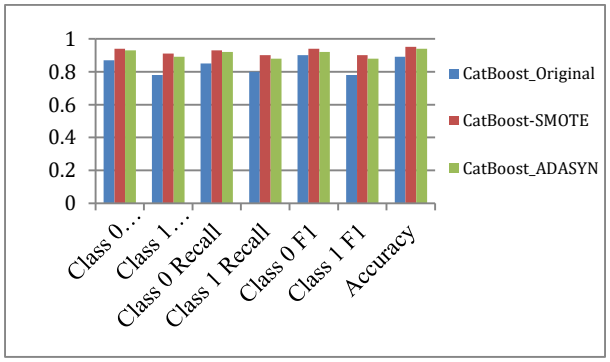
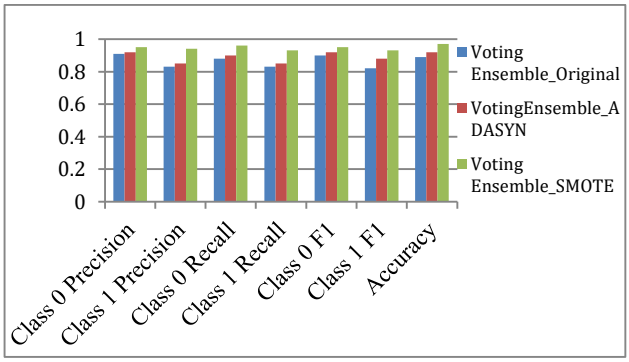


Figure (7): Performance Comparison of CatBoost model Using Original Data, SMOTE, and ADASYN



Figure(8): Performance Comparison of Voting ensemble model Using Original Data, SMOTE, and ADASYN

The figure (9) illustrates how well the proposed voting ensemble model with SMOTE performed. It lies well above this line of random classification indicating that the proposed model has strong discrimination capabilities. The quick slope upward at low false positive rates shows that the proposed model performed effectively to identify churn customers while minimizing false alarms. With a high area under the curve (AUC = 0.96), the proposed approach is a robust method for detecting and addressing the imbalanced Telco churn dataset.

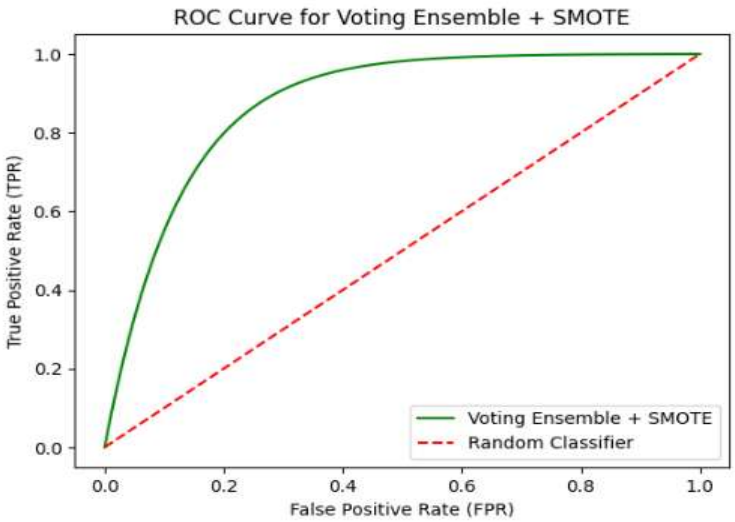


Figure (9): Receiver Operating Characteristic (ROC)

5. Discussion

From my research results, I have concluded that ensemble-based models tend to perform better than single ML models when predicting customer churn in the Telecommunications industry. Among the individual models, Random Forests, XGBoost, and CatBoost were able to predict churn much better than SVM and Decision Tree. This suggests that ensemble learning methods may be more efficient at capturing complicated relationships that exist in structured datasets consisting of customer-related data than their non-ensemble counterparts. The results are also in line with what previous studies have found, namely that the boosted and ensemble-models tend to have better performance than their no-boost or no-ensemble counterparts when it comes to performing a prediction task of churn [9], [13], [16].

In addition to the individual predictions produced by Random Forest, XGBoost and CatBoost, utilizing a Voting Ensemble model also increased the overall classification output. By employing three different types of learning mechanisms (Random Forest, XGboost and Catboost), combining their predictions will reduce the amount of model-specific error in predicting customer churn and also improve stable-ness between all the classifiers. These results suggest that combining the predictions of several ensemble-based classifiers will provide a more reliable means of identifying customers at risk of churning than a single standard ML classifier would alone.

Experiments have shown how essential it is to deal with a class imbalance problem. Both methods - SMOTE and ADASYN - improved the performance of the classifiers tested, but most classifiers gave better predictions using SMOTE than ADASYN in terms of accuracy, precision, recall, and F1 score. One possible explanation for this is that SMOTE generates synthetic observations throughout the minority class's feature space more equally than does ADASYN. This could produce more consistent decision boundaries than ADASYN because the ADASYN method produces synthetic observations predominantly in a small area close to the boundary between classes, which can increase susceptibility to local noise and will negatively impact generalization ability in some instances. The ROC-AUC scores support the use of the methods discussed above. The Voting Ensemble method combined with SMOTE produced the best AUC result of all models tested, suggesting a very strong ability to correctly classify customers who did and did not churn at various thresholds. This is significant because datasets containing information about customers who churn and do not churn will usually be very imbalanced; thus, it would be prudent to use measures that do not depend on the threshold used to assess model performance. The feature importance analysis found that Monthly Charges and tenure are among the most significant factors in predicting customer churn, which is consistent with empirical evidence from the telecommunications industry indicating that long-term customers usually have a lower likelihood of churning than those who pay higher monthly service fees and are therefore more likely to terminate their accounts. Accordingly, Monthly Charges and tenure represent valuable determinants of retention strategies for those customers who do churn.

Despite these positive results, there are limitations to this work. First, the experiments used one benchmark dataset, thus limiting generalizability of results beyond the dataset to other telecommunications. Second, only traditional machine learning and ensemble learning methods were effectively tested so far; therefore, future research will examine the use of deep learning architectures, explainable artificial intelligence (XAI) techniques, and additional feature engineering techniques to potentially improve churn prediction accuracy and model interpretability.

6. Conclusion

This paper investigates the application of various machine learning techniques including SVM, Decision Tree, Random Forest, XGBoost, and CatBoost on a dataset of telecommunications customers (Telco Customer Churn Dataset). The results of the analysis indicate that 'ensemble' based models were more accurate than 'traditional' machine learning models. The Voting Ensemble model which combines XGBoost, Random Forest and Cat Boost performed better than all three models considered. Additionally, implementation of two methods of dealing with class imbalance - ADASYN and SMOTE - resulted in greatly improved predictive ability for the models developed. The maximum predictive ability was attained by using Voting Ensemble combined with SMOTE .These findings suggest that combining ensemble learning with methods for dealing with imbalance is necessary for successful viabilities of churn prediction models. Future work could consider the use of deep-learning models and feature selection techniques; however, an emphasis on providing interpretability through Explainable AI processes (like SHAP) also could improve the predictive ability and subsequent use of churn prediction models.

Conflict of Interest: The authors declare that there are no conflicts of interest.

References

[1] P. Wachwanakijkul, S. Junsiritrakhoon, N. Kantanantha, G. Narayanamurthy, and P. Jarumaneeroj, “Data-driven approaches to predicting customer churn in a non-contractual car-sharing company,” *Transportation Research Interdisciplinary Perspectives*, vol. 33, p. 101600, 2025, doi: 10.1016/j.trip.2025.101600.

[2] H. Shoja, E. Sabet, and S. S. Sadat, “Customer churn prediction system using machine learning: A case study ROSHAN Telecom-Afghanistan,” *International Journal of Integrated Science and Technology*, vol. 4, pp. 123–137, 2026, doi: 10.59890/ijist.v4i2.287.

[3] T. Zhang, S. Moro, and R. Ramos, “A data-driven approach to improve customer churn prediction based on telecom customer segmentation,” *Future Internet*, vol. 14, pp. 1–19, 2022, doi: 10.3390/fi14030094.

[4] L. Theodorakopoulos, A. Theodoropoulou, and C. Klavdianos, “Big data analytics and AI for consumer behavior in digital marketing: Applications, synthetic and dark data, and future directions,” *Big Data and Cognitive Computing*, vol. 10, p. 46, 2026, doi: 10.3390/bdcc10020046.

[5] M. Martinović, K. Dokic, and D. Pudić, “Comparative analysis of machine learning models for predicting innovation outcomes: An applied AI approach,” *Applied Sciences*, vol. 15, p. 3636, 2025, doi: 10.3390/app15073636.

[6] Y. Rimal, N. Sharma, A. Alsadoon, and S. K. Abbas, “A comparative analysis of ensemble AutoML machine learning prediction accuracy of STEM student grade prediction: A multi-class classification perspective,” *Multimedia Tools and Applications*, vol. 84, pp. 38343–38369, 2025, doi: 10.1007/s11042-024-20554-8.

[7] N. M. AbdelAziz *et al.*, “A comprehensive evaluation of machine learning and deep learning models for churn prediction,” *Information*, vol. 16, no. 7, p. 537, 2025, doi: 10.3390/info16070537.

[8] C. Kaope and Y. Pristyanto, “The effect of class imbalance handling on datasets toward classification algorithm performance,” *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, pp. 227–238, 2023, doi: 10.30812/matrik.v22i2.2515.

[9] P. Magadum and M. A. Shaikhsurab, “Enhancing customer churn prediction in telecommunications: An adaptive ensemble learning approach,” 2024, doi: 10.13140/RG.2.2.24573.78569.

[10] A. Bhatnagar and S. Srivastava, “Customer churn prediction: A machine learning approach with data balancing for telecom industry,” *International Journal of Computing*, pp. 9–18, 2025, doi: 10.47839/ijc.24.1.3873.

[11] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Applied Sciences*, vol. 11, p. 4742, 2021, doi: 10.3390/app11114742.
[12] X. Chen *et al.*, "A comprehensive analysis of churn prediction in telecommunications using machine learning," *arXiv preprint arXiv:2509.22654*, 2025.
[13] A. Bhatnagar, "Customer churn prediction using machine learning approach: A comprehensive study," *Journal of Information Systems Engineering and Management*, vol. 10, pp. 80–92, 2025, doi: 10.52783/jisem.v10i25s.3944.
[14] Y. Wei, "Telco customer churn prediction," *Highlights in Science, Engineering and Technology*, vol. 92, pp. 218–226, 2024, doi: 10.54097/84bmrd32.
[15] D. Gupta, N. Patel, and A. Shri, "Enhancing telecom customer retention through data mining-based churn prediction," *International Journal of Scientific Research in Science and Technology*, vol. 11, pp. 566–572, 2024, doi: 10.32628/IJSRST24122223.
[16] K. Peng and Y. Peng, "Research on telecom customer churn prediction based on GA-XGBoost and SHAP," *Journal of Computer and Communications*, vol. 10, pp. 107–120, 2022, doi: 10.4236/jcc.2022.1011008.
[17] A. Kumar and D. Mishra, "Improving support vector machine using modified kernel function," *International Journal of Scientific Research and Modern Technology*, vol. 4, pp. 1–5, 2025, doi: 10.38124/ijsrmt.v4i5.501.
[18] I. Mienye and N. Jere, "A survey of decision trees: Concepts, algorithms, and applications," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3416838.
[19] H. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylonian Journal of Machine Learning*, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.
[20] K. İleri, "Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimized with PSO," *International Journal of Machine Learning and Cybernetics*, vol. 16, pp. 6937–6956, 2025, doi: 10.1007/s13042-025-02654-5.
[21] S. Ganie, P. D. Pramanik, and Z. Zhao, "Ensemble learning with explainable AI for improved heart disease prediction," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-97547-6.
[22] H. Weko and H. Suparwito, "SVM and ensemble majority voting algorithm on sentiment analysis of using ChatGPT in education," *International Journal of Applied Sciences and Smart Technologies*, vol. 7, pp. 451–468, 2025, doi: 10.24071/ijasst.v7i2.12680.
[23] S. Ganie, P. D. Pramanik, and Z. Zhao, "Ensemble learning with explainable AI for improved heart disease prediction," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-97547-6.
[24] A. Omari *et al.*, "A predictive analytics approach to improve telecom's customer retention," *Frontiers in Artificial Intelligence*, vol. 8, 2025, doi: 10.3389/frai.2025.1600357.
[25] S. J. Haddadi *et al.*, "Customer churn prediction in imbalanced datasets with resampling methods: A comparative study," *Expert Systems with Applications*, vol. 246, p. 123086, 2024, doi: 10.1016/j.eswa.2023.123086.
[26] N. M. S. Neethu and V. C. S. Vinod, "A review of unlabeled and imbalanced data challenges in machine learning: Strategies and solutions," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 15, 2025, doi: 10.1002/widm.70043.
[27] D. Elreedy, A. Atiya, and F. Kamalov, "A theoretical distribution analysis of SMOTE for imbalanced learning," *Machine Learning*, vol. 113, 2023, doi: 10.1007/s10994-022-06296-4.
[28] Y. Li *et al.*, "An improved SMOTE algorithm for enhanced imbalanced data classification," *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-09506-w.
[29] S. F. Taskiran *et al.*, "A comprehensive evaluation of oversampling techniques for enhancing text classification performance," *Scientific Reports*, vol. 15, p. 21631, 2025, doi: 10.1038/s41598-025-05791-7.
[30] A. Taskeen, S. U. R. Khan, and A. Mashkoor, "An adaptive synthetic sampling and batch generation-oriented hybrid approach," *Soft Computing*, vol. 28, pp. 13595–13614, 2024, doi: 10.1007/s00500-024-10378-x.
[31] D. Salifu *et al.*, "Data augmentation and machine learning algorithms for multi-class imbalanced morphometrics data," *Heliyon*, vol. 11, p. e42214, 2025, doi: 10.1016/j.heliyon.2025.e42214.
[32] X. Gao *et al.*, "A comprehensive survey on imbalanced data learning," *Frontiers of Computer Science*, vol. 20, 2026, doi: 10.1007/s11704-025-50274-7.
[33] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, 2024, doi: 10.1038/s41598-024-56706-x.

[34] K. Sujon *et al.*, "Accuracy, precision, recall, F1-score, or MCC? Empirical evidence from advanced statistics, ML, and XAI," *Journal of Big Data*, vol. 12, 2025, doi: 10.1186/s40537-025-01313-4.