Predictive Analytics for Customer Churn in Digital Banking Using Integrated Ensemble ML Approaches

Pradeep Rao Vennamaneni[1], Sri Prudhvi Adusumilli[2], Bose babu Ravula[3], Yudhveer Kandukuri[4], Sojy John[5], Venkata Phaneendra Reddi[6]

[1]JPMC - Senior Software Engineer, praovennamaneni@gmail.com
[2]Charter Communications, Sr. Software Developer, prudhvi2379@gmail.com
[3]John Wiley &Sons, Inc., Principal, bbravula@icloud.com
[4]Afiniti Inc., (Senior DBA and DevOps Engineer), veer.sqldba@gmail.com
[5]Recreational Equipment, Inc - Technology Lead, DevOps Engineer & Cloud Migration, sojy.john@gmail.com
[6]Atlantic Health Systems - Oracle Integrations Developer, v3nkata@gmail.com

Abstract
Customer churn prediction in digital banking has become a predictive analytics problem because retaining customers is less expensive than acquiring them. The increasing competition of digital financial services heightens the necessity of smart systems that can extract the customers who are likely to depart, but conventional statistical solutions can be inadequate in identifying the behavioral patterns that are not easily identified. The project aims to develop a reliable and robust machine-learning (ML) model that uses demographic data to predict client attrition, financial, and behavioral attributes based on a 10,000-record Kaggle digital banking customer churn dataset. Its methodology involves thorough preprocessing - removal of irrelevant identifiers, categorical encoding, feature standardization, and class balancing using SMOTE - followed by training four optimized models: Decision Tree, LightGBM, XGBoost, and an Ensemble Voting Classifier. The models that performed the best as ensemble classifiers (accuracy of 90% and ROC-AUC of 0.95%) were assessed using accuracy and precision, recall and F1-score, and ROC-AUC. When compared to the current models (Random Forest (77% accuracy), Logistic Regression (70.98% accuracy), and Naive Bayes (47.63% accuracy), the efficacy of the suggested ensemble technique demonstrates its superior capacity for generalization. It was found that the combination of multiple learners greatly minimizes false negatives, which are critical in avoiding loss of customers, and provides maximum discriminative powers relative to all other models that were tested. The work enhances a scaled and data-based churn prediction pipeline specific to the digital banking industry, with better accuracy, reduced error, and realistic applicability to real-time customer retention systems.

Keywords: Customer churn, Banking, Machine learning, Ensemble Voting classifier, SMOTE, Predictive Analytics.
DOI: 10.21590/ijtmh.10.04.43

Introduction
There is more rivalry in the financial services sector due to the rapid development of the industry, the need to adjust to new technologies, and the entry of FinTech firms into the conventional banking market. Currently, domestic commercial banks are dealing with numerous difficulties and significant and intricate adjustments[1]. Cloud computing and mobile Internet are examples of information technologies that are booming, capital-focused financial regulation is becoming more robust, and Interest rate marketization and financial disintermediation are processes that are accelerating gradually, which has caused banks'

interest margins to drastically shrink[2]. Additionally, third-party payment businesses are e-business entities, which leverage big data and Internet technologies, are moving into the conventional commercial bank business sector, undercutting the advantages of commercial banks' channels[3]. Consumer demand for financial spending is progressively rising in the meantime. Consumers are more concerned with experience, individualized service, diversity, and agility, which makes commercial banks even more competitive. The rivalry for clients, particularly high-quality clients, is one of the primary issues facing commercial banks[4]. Commercial banks must prevent customer attrition while attracting new clients since customer retention is closely correlated with earnings[5]. A major problem for firms is churn, which occurs when clients leave or move to another service provider. As a general rule, different sectors have different standards for categorizing churners[6][7].

The modern organization faces customer churn, which is an occurrence that leads to huge financial losses to a business when the customer leaves, as a challenge to retain customers [8][9]. Unsatisfied customers are often known to move to a different service provider when they consider the services offered by their current company unsatisfactory. There is already proof that machine learning (ML) methods have been effectively applied to identify client attrition. Under certain conditions, though, ML-based algorithms do not provide promising outcomes in regard to client churn detection [10]. The biggest analytical issue that was left was the fact that financial institutions were having a hard time keeping customers due to the growing competition and the fast-changing digital services. Despite the fact that conventional models like NB, RF, and LR have been extensively researched, the current models are characterized by critical research gaps such as, limited to address the issue of class imbalance and to represent features effectively, as well as, to project on different customer behaviours and to conduct the state-of-art research on the development of ensemble or hybrid models to enhance the stability of the predictions [11]. These constraints decrease the accuracy of churn detection systems particularly in the detection of high-risk customers hence lost retention opportunities and loss of revenues. The motivation factor of this work is to come up with a more precise, balanced and robust churn prediction model that is able to be sensitive to nonlinear behavioral pattern and reduce false negatives. Therefore, the proposed study aims to develop an ideal ML system, i.e., incorporating superior models to boost predictive accuracy and discriminative power, and eventually, offer a viable and evidence-based solution to proactive retention of customers in online banking. The following is the research contribution of this task:

- Creation of an Optimized Ensemble Framework: Designed and developed an effective machine-learning pipeline that included Decision Tree, LightGBM, XGBoost, and a soft-voting ensemble to reduce prediction mistakes and improve churn prediction accuracy.
- Improved Class Imbalance Handling: SMOTE-based oversampling was used to overcome the disparity in metrics between churned and non-churned consumers to improve forecast accuracy and model training.
- Extensive Evaluation and Comparison: One carried out a thorough performance evaluation on evaluated the suggested models using a range of parameters (ROC-AUC, F1-score, recall, accuracy, and precision) and contrasted them with prevailing methods.
- Reduction of False Negatives: It has been shown that the ensemble model is successful in removing false negatives, which is essential in determining at-risk customers and enhancing the retention strategy in real-world banking.

Practical Data-driven Solution to Digital Banking: It supplied a scalable and deployable churn prediction model that detects nonlinear customer behavior patterns and provides actionable implications to proactive retention campaigns to the business decision-making process.

The suggested structure of the paper is the following: Section II provide the examination of relevant research on churn prediction, Section III contains the specifics of the suggested approach, model construction, feature engineering, data pretreatment, and assessment metrics, Section IV includes the results, confusion matrix, and comparative analysis of the advantages and limitations of Future work and

the suggested approach, and Section V finishes the study and emphasizes the usefulness of the suggested approach in terms of client retention tactics.

Literature Review

Customer churn analysis and forecasting in the banking sector is something that is of concern nowadays since it is quite crucial that in order to predict which clients are about to cancel their membership with the telecommunications company, the telecommunications industries examine the behaviours of various customers. Thus, data mining and algorithm is significant in the current business environment as it would cost a company more to draw in rather than retaining existing clients. Therefore, while addressing these issues with customer messages, A. Xiong et al. (2019), This paper presents L-RBF, a simple and effective model for predicting customer attrition more precisely. The objective is to apply the Lasso Regression method (Lasso) to optimize Radial Basis Function Neural Networks (RBFs). Lasso is used to remove the characteristics that are linked to customer turnover when the customer features are first installed on edge terminals. Lastly, compared to the prior study, which used Boosting, RBF, and Logistic Regression (Log-R), L-RBF has a higher recall rate and better prediction categorization, according to experimental data[12].
The article that was conducted within this paper is Putta Hemalatha et al. (2019), which aims at creating a model that able to predict improved churn accuracy. They used the current SVM models and its kernel function in conjunction with artificial neural networks to make predictions. The outcomes were excellent in case the models interacted with each other. In this way, majority of the generalized innovations that have been highlighted in the writing to expand Neural networks, SVM, and other soft computing techniques are used in customer churn administration stages[13]. In this paper Abhishek Gaur et al. (2018), concentrates on the different ML algorithms through which it develops the effectiveness of the classification models, such as GBT, RF, SVM, and LR[14].
In this paper J. Hu et al. (2018), In the telecom industry, For predicting customer attrition, a product-based recurrent neural network (pRNN) is recommended. In the proposed approach, an RNN with long short-term memory units learns the customer data sequence pattern that varies over time, and a recurrent layer is introduced before the product's functionality to understand how features interact at a high level. RNN is applied to one of the telecommunication datasets in the real world, and the results of the experiment show that pRNN is significantly superior to other comparison models [15].
S. Bharadwaj et al. (2018), provide two models that reasonably accurately predict client attrition. The first model uses the LR model, a non-linear classifier, with the sigmoid acting as the activation function. Regularizing the model with a regularizing parameter of 0.01 improves its accuracy, yielding an accuracy of 87.52% on the test dataset. The second model is a complete MLP neural network with three hidden layers overlaid with the normalized input feature vector. It employs binary cross entropy as the loss function. It learns at a rate of 0.01. This assists in developing brand loyalty that consequently results into a self-sustaining network [16].
To solve this problem, Jin Xiao et al. (2018), Combine cost-sensitive learning with semi-supervised learning. A Metacost-based semi-supervised ensemble model is shown utilizing the ensemble technique random subspace. This model involves the following three steps: (1) create a new training set by transforming the class label of the original labelled training set using the Metacost approach; (2) label the unlabelled samples judiciously; (3) use the resultant training set to train several simple classifiers and use them to classify the test set's samples. The suggested model outperforms conventional supervised and semi-supervised models in terms of customer churn prediction performance, according to empirical findings on two customer churn prediction data sets[17].
This paper D. Do et al. (2017), In order to determine which subscribers are most likely to discontinue utilizing the services, a single ISP company's significantly skewed data served as the basis for the customer attrition prediction. Several models were developed using SMOTE oversampling to handle the difference between churners and non-churners in the predictive modeling, including AdaBoost, Extra Trees, KNN, Neural Network, and XGBoost. In terms of accuracy and recall, the XGBoost model proves to be superior

to the competition. The accuracy and recall of the model are 42.06 and 45.71 on the dataset with 2% churners and 98% non-churners, respectively [18].

Therefore A. Rodan et al. (2015), Developing the right churn prediction models may help the lack of retention initiatives in customer relationship management plans helps the service provider increase earnings. According to this study, telecom firms may predict customer attrition using an Echo State Network (ESN) and the SVM training methodology. Two datasets, one from a nearby service provider and the other from a publicly available internet source, were utilized to train and evaluate the proposed technique. The experiment's findings demonstrate that, for comparable customer churn prediction scenarios that have been employed in according to the literature, ESN with SVM readout performs better than other popular ML models[19].

The existing customer churn prediction literature demonstrates a definite diversion of the conventional machine-learned approach to more sophisticated hybrid, ensemble, and deep-learning approaches. The early literature involved the improvement of neural network architecture and feature selection, whereas more recent studies involved the use of hybrid models between SVM, ANN and boosting, among others, to predict better. Sequential customer behavior was well represented by deep-learning models, especially recurrent models. Other studies have also focused on dealing with noisy labels and class imbalance with semi-supervised learning and oversampling. In general, the available literature underscores that hybrid and combined ensemble models are always more effective than isolated algorithms that are more accurate, robust, and reliable in predicting churn in both the telecom and digital service sectors. Table I summarizes earlier research on predicting customer attrition

Table 1: Summary of Previous Work on Customer Churn Prediction

Author/Year	**Technique/Model Used**	**Key Contribution**	**Strength/Finding**	**Advantage**	**Limitation/Future Work**
A. Xiong et al. (2019)	Lasso Regression + RBF Neural Network (L-RBF)	Automated feature selection and RBF topology generation	Higher recall and improved classification over LogR, RBF, Boosting	Reduces model complexity using Lasso; strong feature extraction	Performance depends on RBF kernel tuning; extension to large datasets recommended
Putta Hemalatha et al. (2019)	Hybrid SVM + ANN	Combined SVM kernel learning with ANN prediction	Improved accuracy compared to single models	Leverages strengths of SVM and ANN simultaneously	High computational cost; future work may involve deep learning hybrids
Abhishek Gaur et al. (2018)	GB Trees, RF, SVM, and LR	Comparative performance analysis of ML classifiers	Gradient Boosting achieves superior results	Simple yet effective baseline comparison	Limited deep-learning exploration; needs analysis on imbalanced datasets
J. Hu et al. (2018)	Product-based RNN (pRNN) with LSTM	Captures sequential customer behavior using feature interactions	Outperforms traditional ML methods on telecom data	Learns temporal patterns; adaptable for large volumes	Requires long training time; complex architecture; future enhancement in interpretability
Jin Xiao et al. (2018)	Semi-supervised Metacost +	Label correction +	Higher accuracy than supervised	Addresses class	Dependent on quality of

	Random Subspace Ensemble	unlabeled sample selection for improved classification	ensemble methods	imbalance and label noise	unlabeled data; needs scalability improvements
D. Do et al. (2017)	SMOTE + AdaBoost, Extra Trees, KNN, NN, XGBoost	Handling extreme class imbalance (2% churners)	XGBoost yields highest precision and recall	Effective imbalance handling using SMOTE	Synthetic data may introduce noise; future work could use advanced oversampling
A. Rodan et al. (2015)	Echo State Network (ESN) + SVM	Hybrid reservoir computing model for churn prediction	Outperforms classical ML models	Fast training due to ESN architecture	Limited generalization on non-temporal data; deeper ESN models suggested

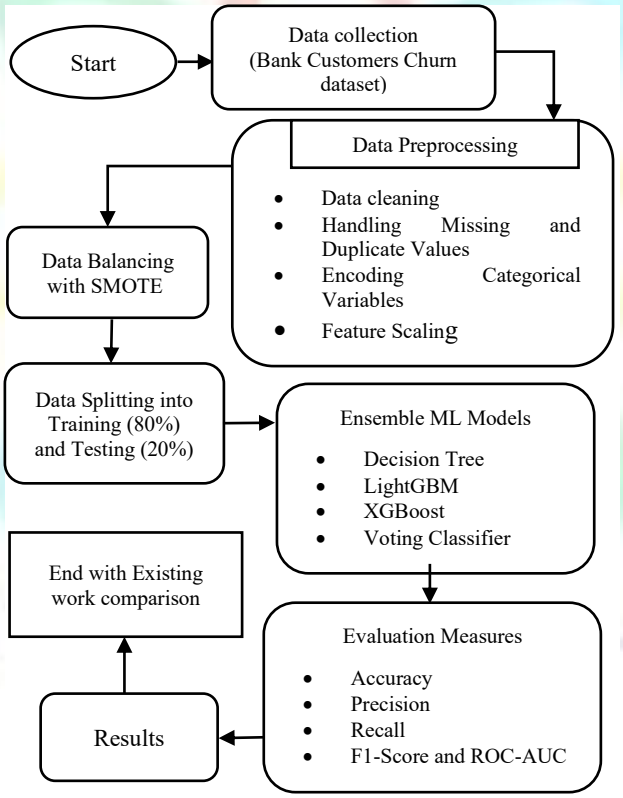


Figure 1: Proposed Flowchart for customer churn prediction using ML models

Methodology

This section details the entire end-to-end approach to churn prediction on the Bank Customers Churn dataset provided by Kaggle. Figure 1 below presents the workflow that involves preprocessing of data, exploratory analysis, balancing, feature scaling, development of model, and integration of ensembles.

Data Acquisition

This study utilizes data from the Bank Customers Churn Dataset, which includes 10,000 customers, on Kaggle records that include 14 attributes representing demographic, financial and behavioural data. Location, gender, age, tenure, credit score, amount of things, credit card status, level of activity, predicted income, and the objective Exited is one of these traits that shows whether or not a client churned. The data gives a thorough and well-rounded base on which to study churn behaviours and come up with predictive ML models on the digital banking applications.

The distribution of the classes in Figure 2 illustrates the extreme skewness of the Bank Customers Churn Dataset with 7,963 non-churned (79.6) and 2,037 churned (20.4) customers, which creates a ratio of about 4:1. This imbalance is common when churn prediction is involved, and the retention rates are inherently higher than the attrition rates. Nevertheless, these biased distributions are problematic to ML models, which are more likely to disregard the minority class in favor of the dominant one, which is the churned customers that are the main focus of retention strategies. This imbalance requires the use of SMOTE to make the training data balanced, such that models are equally trained on discriminative patterns of both classes and are not training on a few non-churn and fail to identify at-risk customers.

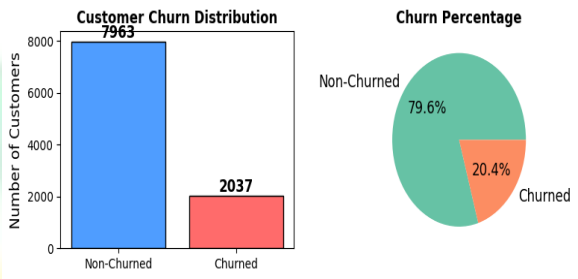


Figure 2: Data Distribution graphs of Bank Customers Churn Dataset

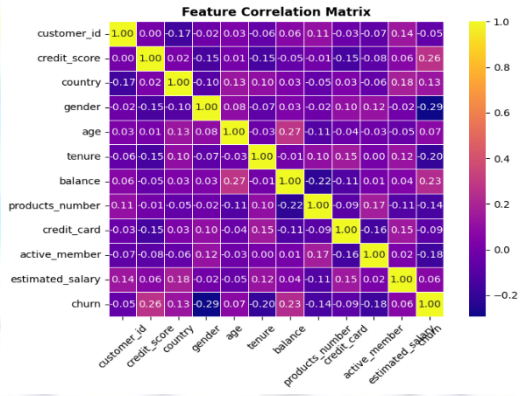


Figure 3: Correlation feature heatmap of Bank Customers Churn Dataset

Figure 3 shows the feature correlation matrix which shows the key relationships between customer attributes and the churn behavior in the Bank Customers Churn Dataset. The heatmap reveals that a number of features show significant correlation with the target (churn) variable, the age having a moderate positive relationship (0.27), meaning that older customers tend to be more likely to churn, and the balance with which customers tend to churn has a negative correlation (-0.22). Also, active member feature shows a negative correlation of -0.16 with churn which validates the fact that engaged customers are better retained. Interestingly, products number has positive correlation (0.09) with churn which could be the case that the multiple products customer might experience complexity or dissatisfaction. The correlation coefficient also shows that there is a problem of multicollinearity with customer id having very low correlation with all features, which confirms the fact that they are irrelevant identifiers and need to be dropped.

Data Preprocessing

The pre-processing of data is a step towards better results before the ensemble machine learning models are developed. The data preprocessing methods used in the churn dataset of bank customers are:

Data cleaning: As part of the preprocessing stage, data cleaning commenced with the process of excluding unnecessary features: CustomerId, RowNumber, and Surname as they don't offer any insights on churn prediction.

Handling Missing and Duplicate Values: Null values and duplicate values were not worked on and as the dataset had no null values; only duplicate entries were examined and removed where necessary.

Encoding Categorical Variables: Categorical fields such as Geography were transformed using One-Hot Encoding, while Gender was label-encoded to convert them into numerical forms suitable for ML algorithms.

Feature Scaling: StandardScaler was used to standardize numerical features such as CreditScore, Age, Balance and EstimatedSalary to have a uniform distribution of features and increase the model performance. For efficient churn prediction, systematic pretreatment workflow guaranteed clean, consistent, and model-ready data.

Data Balancing

The data was imbalanced with a large percentage of non-churned customers being represented in the dataset compared to the consumers who left, it might cause ML models to be biased against the dominant class. To oversample the minority group, synthetic samples were created utilizing feature space similarities between the present minority cases using a technique known as the Synthetic Minority Oversampling Technique (SMOTE)[20]. This approach effectively increased the representation of churned customers, resulting in a balanced dataset that improved the fairness, stability, and predictive accuracy of the classification models. To generate new samples, the SMOTE algorithm interpolates between existing examples of the minority class and their nearest neighbors. The mathematical representation of this process is Equation. (1):

$$x_{new} = x_{minority} + \lambda \cdot \left(x_{neighbor} - x_{minority}\right) \qquad (1)$$

The minority class instance is represented by $xminorby$, its k-nearest neighbours are represented by $xneihborh$, and λ is a random number in the interval [0, 1]. The minority class is more represented since the new synthetic instance x_{new} is located along the line segment that connects $x_{minority}$ and $x_{neighbor}$.

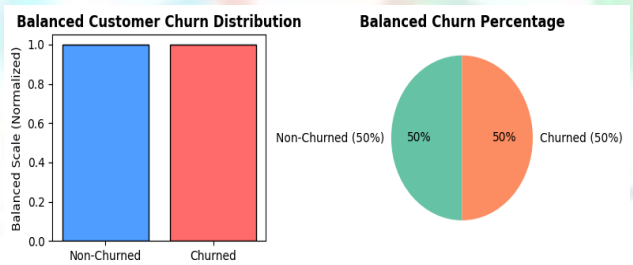


Figure 4: After data balanicng graphs of Bank Customers Churn Dataset

This method has served to overcome class imbalance by increasing up the number of churned customers to a balanced dataset of 50% of non-churned and 50% churned customers, as illustrated in Figure 4. Having a balanced dataset (normalized between 0-1), both classes have equal representation in the training of ML models, which does not favor the majority class and makes the model discover discriminatory patterns that can be used to identify at-risk customers. The synthetic oversampling generates 5,926 new churned customer records, which makes the total number of 7,963 instances per class, which is a strong basis to train predictive models without the risk of overfitting that can be generated by simple random oversampling.

Data Splitting

The processing and balancing dataset were then split into the test and training sets to evaluate the models' ability to generalize. Model testing used 20% of the data, and model training used the remaining 80%. The stratified splitting was used to make both classes (churn and non-churn) equally represented in the assessment and training settings to prevent prejudice against the majority class.

Ensemble Machine Learning Approaches

In a number of fields, ensemble learning techniques are being extensively researched in an effort to solve real-world problems. The precise forecasting, identification, evaluation, and prediction of several churn circumstances have significantly improved because to these models. Four machine learning methods based on ensemble learning were used in the study to investigate the goal of forecasting client attrition. Four machine learning models were developed to forecast customer attrition: Decision Tree, LightGBM, XGBoost, and an Ensemble Voting Classifier. To get the greatest prediction outcomes, each model was trained and refined through hyperparameter tweaking. The following algorithms were used in the experiment.

Decision Tree

In decision theory, DT learning is a fundamental technique. These trees are made up of a top knot which is the root and all knots are intertwined by the branches. Using the DT model as the baseline classifier to predict churn, the dataset is iteratively split into homogeneous groups according to the values of several variables. It uses impurity reduction to choose the best feature to split on, most often the Gini Impurity[21]. The impurity for a node is calculated using the formula Equation (2):

$$Gini = 1 - \sum_{i=1}^{C} p_i^2 \quad (2)$$

where p_i^2 represents the number of classes, denoted by C, and the probability of class i within the node. A purer node is indicated by a lower Gini value, which directs the tree toward the split that should be made at each level. The Decision Tree can be used with this criterion to capture important behavioral patterns associated with churn, but can also be subject to overfitting, which in this research is tamed with the help of tuned hyperparameters (minimum sample per split, minimum sample per leaf, and maximum depth). The Decision Tree classifier was a baseline model in predicting churn since it is easily understandable and interpretable. To optimize performance, Hyperparameters were optimized with the help of GridSearchCV. The most suitable setting was max_depth = 5, min_samples_split = 50, min_samples-leaf = 20, and the criterion = gini. This arrangement served to inhibit overfitting and at the same time statistically significant decision boundaries in the data leading to balanced performance in terms of accuracy.

LightGBM

The computational efficiency of the gradient boosting framework LightGBM is well-known. To speed up training and lower data dimensionality, it makes use of techniques like One-Side Gradient-Based Sampling and Select Feature Bundling[22]. Equation is the representation of the iterative updating process (3):

$$F_{t+1}(X) = F_t(X) + \eta \sum_{i=1}^{N} \gamma_i h_i(X) \quad (3)$$

where γ_i are the loss function's gradients in relation to the predictions $h_i(X)$ at iteration t, $F_t(X)$ is the model at iteration t, and η is the learning rate. In this case, Xvolume, Xtrend, and Xexog guarantee that the LightGBM model makes use of thorough data on sales dynamics and external influences. LightGBM was selected because it is efficient with large tabular data and can be trained in a short time. The following hyperparameters were chosen after the randomized search was performed to optimize them: Subsample = 0.8, max-depth = 7, learning-rate = 0.05, n-estimators = 200, num-leaves = 31, and colsample-bytree = 0.8." The settings enabled the model to effectively learn complicated trends in the data with generalization, which led to an improved predictive accuracy and stability.

XGBoost

A regularized learning framework is used by the eXtreme Gradient Boosting (XGBoost) technique to optimize model performance and processing speed[23]. XGBoost, a robust gradient boosting technique, is

renowned for its capacity to manage multicollinearity and produce accurate prediction results. In order to manage the model's complexity and lessen overfitting, this approach incorporates a regularisation term into the goal function. The objective function's general form is as follows: Equation.(4):

$$Obj = \sum_{i=1}^{n} L\,(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \quad (4)$$

where Ω penalises the model's complexity and L is the difference between the predicted $\hat{y}_i$ and actual y_i outcomes is measured using a differentiable loss function. According to their definitions, the variables n and k are the feature or component index (k) and the total number of observations (n). The XGBoost model Xvolume, Xtrend, and Xexog as inputs to optimize predictions by balancing model complexity with generalization. The model was updated using grid search and cross-validation. Max_depth = 6, learning rate = 0.1, n_estimators = 250, subsample = 0.9, colsample bytree = 0.8, and gamma = 0 were among the hyperparameters that performed well. These values were sure that the underlying patterns of churn were not overfitted in the model, which led to high accuracy.

Soft Voting

Soft voting uses many algorithms with opposing strategies to estimate the class probability and increase forecast accuracy. The key classifier is given a higher weight, and the model's predicted probabilities are added up to determine the highest category. A mathematical illustration of the soft-voting classifier prediction is shown below: Equation.(5):

$$\hat{Y} = arg\max_{i} \sum_{j=1}^{m} w_i p_{ij} \quad (5)$$

where p_{ij} is the likelihood that the classifier correctly predict a certain class, w_i is the weight associated with the classifier prediction, and argmax is the function that generates a maximum. High-confidence models are given greater weight during the soft voting weight assignment procedure, which is determined by the classifier's significance and accuracy.

Ensemble Voting Classifier

Ensemble Voting Classifier utilized soft voting to combine Decision Tree, LightGBM and XGBoost. The above-described hyper parameters were used before integration in each model. This ensemble drew on the strengths of individual models, bias, and variance were minimized, and performance was the greatest between the models. In online banking, the ensemble approach has proven to be the most successful in forecasting client attrition.

Model Evaluation Process

The systematic evaluation was done after training the models and this was done to evaluate how well they predict the testing set. Representing the many facets of performance using a range of performance measures (ROC-AUC, f1, rec, acc, and pre), the evaluation was carried out to determine how well each model tracked the churning clients. Several statistical variables are utilized to characterise a model's generalisation capacity in classification tasks. The confusion matrix is a highly helpful tool for characterizing a model's capacity to categorize data. True negatives (TN), false positives (FP), false negatives (FN), and true positives (TP) may all be distinguished using the confusion matrix. The key categorization metrics are constructed using these confusion matrix data.

Accuracy: is a broad indicator of the performance of the model, which shows the percentage of each case's correct predictions. It is considered using the Equation.(6):

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+TN+FN}} \quad (6)$$

Precision: measures the precision of optimistic forecasts. It displays the proportion of TP situations that the model deems positive. The calculation formula is Equation (7):

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (7)$$

Recall: measures the accuracy with which all pertinent (positive) examples are identified by the model. In situations when missing positive cases are crucial, this is crucial. It is calculated as Equation.(8):

$$Recall = \frac{\text{TP}}{TP+FN} \qquad (8)$$

F1 score: represents the accuracy and recall harmonic mean. It provides a balance between memory and accuracy and is helpful in situations when the distribution of classes is uneven. The formula for the F1 score is Equation.(9):

$$F1 - Score = 2 \times \frac{Precision \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (9)$$

ROC: To show how effectively binary classifiers function, the true positive rate (TPR) and false positive rate (FPR) at different threshold values are displayed via the ROC curve. For different categorization levels, the ROC curve clearly illustrates the trade-off between specificity and sensitivity (recall). The classifier's performance is represented by the AUC using the ROC curve. AUC is a scalar statistic that provides a thorough evaluation of classifier performance across all classification criteria and depicts the whole two-dimensional region beneath the ROC curve.

Results & Discussion

In this section provide the results of experiments using suggested models to predict banking turnover. This work shows that the ensemble voting method is stronger in its predictive performance than about individual machine-learning models, which have a greater degree of stability and generalization to predict customers who are at risk of churn. A Windows computer with an Intel i7 CPU and 16GB of RAM was utilized for the experiments, and the main development environment was Jupyter Notebook, Scikit-learn, LightGBM, XGBoost, and Python. Effective combination of the merits of decision trees and boosting algorithms is achieved by the ensemble model that leads to stronger churn classification. Relative to the current literature, the given approach complies with the current trends that demonstrate that hybrid and ensemble methods demonstrate a consistent higher degree of success than the traditional independent methods for forecasting churn in online banking.

Table 2: Proposed ML models performance for Churn prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC–AUC**
Decision Tree	81	80	83	81	0.85
LightGBM	83	81	84	83	0.89
XGBoost	84	86	85	86	0.92
Ensemble Voting	90	89	90	90	0.95

Table 2 is the summary of the suggested ML algorithms' effectiveness in churn prediction, where the basic and the advanced algorithms have a clear increase in performance. According to the results, the baseline Decision Tree model has a ROC-AUC of 0.85, 81% acc, 80% pre, 83% rec, and 81% F1. LightGBM demonstrated the advantages of gradient boosting by improving the outcomes with 83% acc, 81% pre, 84% rec, 83% F1, and 0.89 ROC-AUC. Further improvement was made on XGBoost, With an F1 of 86, rec of 85, acc of 84%, pre of 86%, and ROC-AUC of 0.92, it demonstrates a high capacity for learning complex patterns. The combination of Decision Tree, LightGBM, and XGBoost significantly improves predictive performance, as seen by the Ensemble Voting Classifier's 90% accuracy, 89% precision, 90% recall, 90% F1-score, and ROC-AUC of 0.95. This advancement supports ensemble learning's effectiveness as a reliable technique for successfully identifying churn in retail banking.

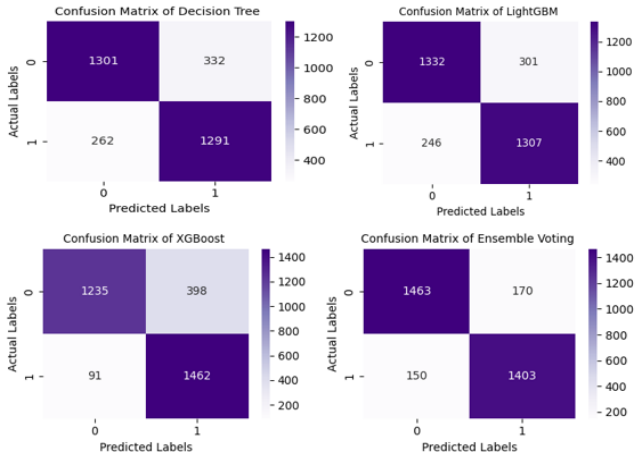


Figure 5: Confusion Matrix of Proposed Models for Churn Prediction

Figure 5 displays the four models' confusion matrices were based on the models that were put into practice, which are evaluated using the test dataset and include Decision Tree, LightGBM, XGBoost, and Ensemble Voting Classifier. The Decision Tree model identified 1,301 non-churned (TN) and 1,291 churned customers (TP) properly, but with 332 false positives and 262 false negatives, it shows that it performed moderately. When LightGBM was used, classification was computed using 1,332 true negatives and 1,307 true positives, and errors were kept down to 301 false positives and 246 false negatives. XGBoost predicted churning aggressively, producing 1,462 TP and 1,235 TN with just 91 FN and 398 FP, which is biased towards detecting at-risk customers. The Ensemble Voting Classifier's ideal balance was 1,463 TN and 1,403 TP with fewer FP (170) and FN (150). The fact that the ensemble can minimize FN is especially useful in churn prediction, since customers who are not at-risk would be lost to retention, and in a banking application, it is the most ideal solution to deploy.

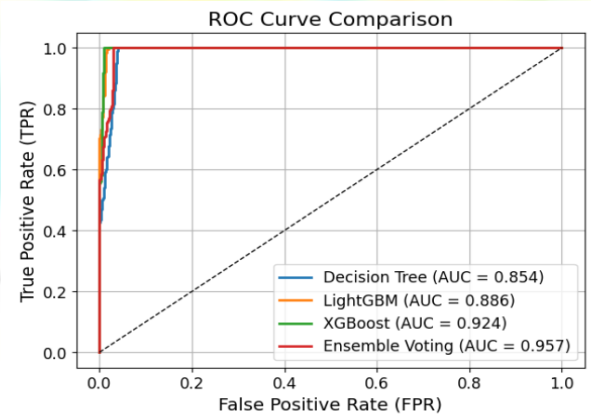


Figure 6: ROC Curve of Proposed Models for Churn Prediction

Figure 6 displays the ROC curves for each of the four models at various categorization levels. The AUC is the main performance indicator. Although not awful enough, the DT's AUC of 0.854 is still below that of the more sophisticated models. LightGBM rose to a new AUC of 0.886 and XGBoost was outstanding with a 0.924, which indicated its execution of nonlinear complex styles. The Ensemble Voting Classifier had the greatest AUC of 0.957 which is a great discriminative ability since it uses the strengths of the individual models by voting on them. The diagonal dashed line is random classification (AUC = 0.5), and the wide distance between the model curves and this baseline proves their practical values, and the close-to-perfect AUC of 0.957 of the ensembles makes it the best option in churning customers predictor in digital banking.

Table 3: Comparative analysis of customers Churn Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC–AUC**
Random Forest[24]	77	-	76	-	-
Logistic Regression[25]	70.98	49	70	57.7	49.6
Naïve Bayes[25]	47.63	71.5	47.3	45.6	62.9
Decision Tree	81	80	83	81	0.85
LightGBM	83	81	84	83	0.89
XGBoost	84	86	85	86	0.92
Ensemble Voting	90	89	90	90	0.95

Table 3 indicates the relative performance of several machine-learning methods that differ significantly from both traditional and sophisticated ensemble algorithms in their ability to forecast customer attrition. Random Forest, previously used models also had lower precision (76) with 77% accuracy and no F1-score or ROC-AUC, with 70.98 percent accuracy, 49% precision, 70% recall, 57.7% F1-score, and 49.6% ROC-AUC with Logistic Regression. Meanwhile, the decision tree, LightGBM, and XGBoost reached 81%, 83%, and 84% accuracy, 82%, 83%, and 85% precision, 81%, 83% and 84% recall, and 82%, 83% and 84 percent F1-score with good support of 0.92 ROC-AUC. With 90% accuracy, 89% precision, 90% recall, 90% F1-score, and the highest 0.95 ROC-AUC, the Ensemble Voting model outperformed all other models in terms of predictive stability and performance. This clearly shows that combining many strong learners improves performance. In general, the findings obtained indicate unequivocal improvement in ensemble techniques to predict churn as opposed to traditional machine-learning methods.

Advantages, Limitations, and Future Work

The churn-prediction framework that is proposed has a number of significant benefits, such as a full end-to-end machine-learning pipeline, successful control of class imbalance via SMOTE, and high predictive performance of the Ensemble Voting Classifier that minimizes false negatives and false positives, which is significant when maximizing customer retention in digital banking. The fact that it incorporates some of the latest models such as LightGBM and XGBoost allows the system to identify the complex behavioural patterns that are usually ignored by traditional methods. Nonetheless, the research has some limitations which include; it is based on a single Kaggle dataset, a small number of features and the research lacks real time transactional/behavioural data which could be used to enhance predictive power. Moreover, the models are also sensitive to data drifts, as the models are not dynamic and thus, they have to be retrained periodically. Further research in this task can be done by introducing more sophisticated features of feature-selection algorithms like RFE, Boruta, and Mutual Information to obtain more accurate predictors and minimize noise. To enhance generalizability, the model can be trained using more broad and varied datasets of telecoms, e-commerce, and insurance industries. Both behavioural and temporal pattern learning could be improved by developing hybrid models comprising machine learning and deep learning (e.g., XGBoost-LSTM models or autoencoder-ensemble models). Besides, it is possible to do cross-industry cross-industry transfer learning and domain adaptation to add the suggested churn-prediction approach to the rest of the customer-centric applications.

Conclusion

This research paper concerning presents a comprehensive ML model that can be used to identify high-risk clients and accurately assess their likelihood of leaving, addressing the critical issue of client attrition in online banking. This paper has shown that a systematic ML pipeline, as well as proper preprocessing and class balancing, can yield substantial customer churn prediction in online banking. The best accuracy of 90%, based on high precision recall balance, and the largest ROCAUC of any model implemented, confirmed the best capability of the Ensemble Voting a classifier to detect clients who could be about to leave. The ensemble model's ability to lower FN and FP is further confirmed by the ROC analysis and confusion matrix, making it very usable in practical retention tactics. It is also evident upon comparative analysis that the offered approach is superior to the traditional ones like LR, NB, and RF, and the superiority of hybrid and ensemble-based learning methods is apparent. It is suggested that, in order to enhance the model going forward, real-time behavioral and transactional data be used, and deep learning networks like LSTM or Transformer networks be tested, use XAI to enhance transparency, and test the model in a live banking setting to observe the dynamics of real customers.

References

1. K. G. M. Karvana, S. Yazid, A. Syalim, and P. Mursanto, “Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry,” in *2019 International Workshop on Big Data and Information Security, IWBIS 2019*, 2019. doi: 10.1109/IWBIS.2019.8935884.
2. P. Godinho, J. Dias, and P. Torres, “An application of data mining methods to the analysis of bank customer profitability and buying behavior,” in *Data Analysis and Applications 1: Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*, 2019. doi: 10.1002/9781119597568.ch16.
3. S. Cui and N. Ding, “Customer churn prediction using improved FCM algorithm,” in *2017 3rd International Conference on Information Management, ICIM 2017*, 2017. doi: 10.1109/INFOMAN.2017.7950357.
4. S. Achouche, U. B. Yalamanchi, and N. Raveendran, “Method, apparatus, and computer-readable medium for performing a data exchange on a data exchange framework,” 2019
5. B. He, Y. Shi, Q. Wan, and X. Zhao, “Prediction of customer attrition of commercial banks based on SVM model,” in *Procedia Computer Science*, 2014. doi: 10.1016/j.procs.2014.05.286.
6. U. Prasad and S. Madhavi, “Prediction of Churn Behavior of Bank Customers,” *Bus. Intell. J.*, 2012.
7. S. Garg, “Predictive Analytics and Auto Remediation using Artificial Inteligence and Machine learning in Cloud Computing Operations,” *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 2, 2019, doi: 10.5281/zenodo.15362327.
8. M. S. Basarslan and I. D. Argun, “Classification of a bank data set on various data mining platforms,” in *2018 Electric Electronics, Computer Science, Biomedical Engineerings’ Meeting, EBBT 2018*, 2018. doi: 10.1109/EBBT.2018.8391441.
9. V. M. L. G. Nerella, “Automated cross-platform database migration and high availability implementation,” *Turkish J. Comput. Math. Educ.*, vol. 9, no. 2, pp. 823–835, Jul. 2018, doi: 10.61841/turcomat.v9i2.15284.
10. A. Mishra and U. S. Reddy, “A comparative study of customer churn prediction in telecom industry using ensemble based classifiers,” *Proc. Int. Conf. Inven. Comput. Informatics, ICICI 2017*, no. January, pp. 721–725, 2018, doi: 10.1109/ICICI.2017.8365230.
11. Y. Li and B. Wang, “A Study on Customer Churn of Commercial Banks Based on Learning from Label Proportions,” in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, Nov. 2018, pp. 1241–1247. doi: 10.1109/ICDMW.2018.00177.
12. A. Xiong, Y. You, and L. Long, “L-RBF: A Customer Churn Prediction Model Based on Lasso + RBF,” in *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing*

(CPSCom) and IEEE Smart Data (SmartData), IEEE, Jul. 2019, pp. 621–626. doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00121.
13. P. Hemalatha and G. M. Amalanathan, "A Hybrid Classification Approach for Customer Churn Prediction using Supervised Learning Methods: Banking Sector," in *Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019*, 2019. doi: 10.1109/ViTECoN.2019.8899692.
14. A. Gaur and R. Dubey, "Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques," in *2018 International Conference on Advanced Computation and Telecommunication, ICACAT 2018*, 2018. doi: 10.1109/ICACAT.2018.8933783.
15. J. Hu *et al.*, "PRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector," in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 2018. doi: 10.1109/BigData.2018.8622094.
16. S. Bharadwaj, B. S. Anil, A. Pahargarh, A. Pahargarh, P. S. Gowra, and S. Kumar, "Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron(MLP)," in *Proceedings of the 2nd International Conference on Green Computing and Internet of Things, ICGCIoT 2018*, 2018. doi: 10.1109/ICGCIoT.2018.8752982.
17. J. Xiao, L. Huang, and L. Xie, "Cost-Sensitive Semi-Supervised Ensemble Model for Customer Churn Prediction," in *2018 15th International Conference on Service Systems and Service Management, ICSSSM 2018*, 2018. doi: 10.1109/ICSSSM.2018.8465099.
18. D. Do, P. Huynh, P. Vo, and T. Vu, "Customer churn prediction in an internet service provider," in *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2017. doi: 10.1109/BigData.2017.8258400.
19. A. Rodan and H. Faris, "Echo State Network with SVM-readout for customer churn prediction," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2015*, 2015. doi: 10.1109/AEECT.2015.7360579.
20. D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," *Inf. Sci. (Ny).*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.
21. S. Takieddine and F. K. A. Baidoo, "An exploratory analysis of internet banking adoption using decision tree induction," *Int. J. Electron. Financ.*, vol. 8, no. 1, p. 1, 2014, doi: 10.1504/IJEF.2014.063996.
22. Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., Ragireddy, U. K., & Addepalli, P. V. (2023). Data-Driven Approaches for Forecasting Cost Overruns in Infrastructure Projects. *European Journal of Technology*, *7*(5), 23-34.
23. Dhuli, S., Prabu, M., Muzaffer, H. S., Uday Kumar, R., & Prasanth Varma, A. (2023). A Survey on Critical Success Factors Influencing Project Success Across Multiple Industries. *J Contemp Edu Theo Artific Intel: JCETAI-106*.
24. Ragireddy, U. K., Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., & Syed, M. H. (2023). An Overview of Different Automation Strategies Used in Cloud-Based CI/CD Pipelines for Software Deployment. *International Journal of Technology, Management and Humanities*, *9*(01), 116-127.
25. Addepalli, P. V., Bandaru, S. R., Shyam, D., Manoharan, P., Syed, M. H., & Ragireddy, U. K. (2023). Performance Analysis of Machine Learning Models for Cybersecurity Risk Classification in IT Projects. *Universal Library of Engineering Technology*, (Issue).
26. Mylavarapu, K. B., Pilli, J., Jannu, P. K., Mohammad, J. A., & Panchali, S. H. (2023). Artificial Intelligence for Continuous Quality Assurance in Cloud-Native Applications: A Review. *European Journal of Technology*, *7*(5), 35-49.
27. Jannu, P. K., Mohammad, J. A., Panchali, S. H., Mylavarapu, K. B., & Pilli, J. (2023). A Survey on Energy-Efficient Microarchitecture Design Strategies for Mobile CPUs and GPUs: Techniques

and Future Directions. *International Journal of Technology, Management and Humanities*, *9*(04), 137-151.

28. Taluri, R. (2021). Cloud-Native Architectures for Enterprise Financial Data Management, Analytics, and Regulatory Reporting Compliance. International Journal of Emerging Trends in Computer Science and Information Technology, 2(2), 101-111. https://doi.org/10.63282/3050-9246.IJETCSIT-V2I2P112
29. Taluri, R. (2022). Cloud Data Engineering Strategies for Large-Scale Financial Data Integration and Intelligent Corporate Performance Reporting. International Journal of Emerging Research in Engineering and Technology, 3(4), 176-188. https://doi.org/10.63282/3050-922X.IJERET-V3I4P119
30. Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).
31. Kavuluri, H. V. R., Pathani, A. K., Dasari, A., Chilakapati, V. K., Keshireddy, S. R., & Nagumotu, V. T. (2023). Cloud Data-Center Network Verification: Approaches, Algorithms and Toolchains. *American Journal of Computing and Engineering*, *6*(2), 18-32.
32. Chilakapati, V. K., Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., & Dasari, A. (2023). A Survey on 5G/6G Radio Access Networks: Technologies, Protocols, and Performance Metrics. *International Journal of Technology, Management and Humanities*, *9*(02), 79-93.
33. Keshireddy, S. R., Nagumotu, V. T., Kavuluri, H. V. R., Pathani, A. K., Dasari, A., & Chilakapati, V. K. (2023). An AI-Based Framework for Detecting IoT Botnets Through Network Traffic Analysis and Modeling. *Universal Library of Engineering Technology*, (Issue).
34. G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017.
35. F. Climent, A. Momparler, and P. Carmona, "Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach," *J. Bus. Res.*, vol. 101, pp. 885–896, Aug. 2019, doi: 10.1016/j.jbusres.2018.11.015.
36. N. Kumar and C. Naik, "Comparative Analysis of Machine Learning Algorithms for their Effectiveness in Churn Prediction in the Telecom Industry," *Int. Res. J. Eng. Technol.*, vol. 04, no. 08, pp. 1–5, 2017.
37. I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.