Journal of Emerging Computer Technologies
Research Article
Received: *2025-01-21* | ***Reviewing:*** *2025-01-21 & 2025-04-27*| ***Accepted:*** *2025-04-28* | ***Online:*** *2025-05-07* | ***Issue Date:*** *2025-12-31*
Doi: 10.57020/ject.1623937

Customer Churn Prediction Using Machine Learning Techniques: Awash Bank Wolaita Sodo Region

Abel Mekuria MOLLA
Dilla University
Dilla, Ethiopia
0009-0004-2937-9939
abel.mekuria@du.edu.et
(Corresponding Author)

Mohammed Abebe YIMER
Arba Minch University
Arba Minch,Ethiopia
0000-0003-0622-4841
mohammed.abebe@amu.edu.et

Yared Dereje WOLDEHANA
Dilla University
Dilla, Ethiopia
0009-0009-4644-4364
yared.dereje@du.edu.et

***Abstract*— Customer churn prediction refers to the procedure of identifying customers who are highly likely to terminate their service subscription based on their utilization. Being able to predict a customer who is likely to churn is essential for solving business problems. The banking industry in Ethiopia currently has millions of users, making it challenging to analyze and anticipate customer attrition. There are diverse researches conducted in this particular domain. The primary challenges encountered in the majority of the prior investigations were associated with the selection of suitable technique for achieving data balancing, the predicaments revolving around the choice of a technique for handling missing values, the excessive dependence of the model on a singular attribute, and various others. The aim of this research is to develop a machine-learning model that can predict customer churn. The dataset utilized for this investigation comprises 50,987 entries encompassing 11 attributes, which were collected from Awash Bank Wolaita Sodo region. Among these, 31,619 represent active accounts, while the remaining 19,368 pertain to closed (churn) accounts. To achieve balance within the dataset, a SMOTE-ENN method is employed, while an extraction tree classifier is employed for important feature selection. This research used an experimental research approach, and eight model are tested, including Extreme Gradient Boosting (XGBoost), random forest, Light Gradient-Boosting Machine (LightGBM), decision tree, Convolutional Neural Network (CNN), Gradient Boosting Machine (GBM), Deep Neural Network (DNN), and Multilayer Perceptron (MLP). Model performance is evaluated using accuracy, f1-score, recall, and precision. Experimental results show random forest model outperformed other models with an overall accuracy of 99.14% and recall, precision and f1-score of 99%.**

Keywords*— *Customer churn, Customer segmentation, Machine learning, Prediction, Survival analysis

I. INTRODUCTION

The banking industry is a pivotal financial institution that offers diverse services to both individuals and organizations. Its significance is particularly noteworthy for developing nations such as Ethiopia, as it fosters a culture of saving within local communities. Furthermore, it plays an indispensable role in providing credit and other financial services to customers. The predominant services provided by the banking industry include checking and savings accounts, credit cards, mortgages, investment products, and other related services. The financial institution endeavors to maintain its current customer base and attract new potential customers over an extended period of time. In a highly competitive banking sector, customers possess the ability to exercise discretion amongst various service providers by virtue of engaging in a certain action. If customers encounter inadequate service from their financial institution, there is a potential threat of terminating their relationship, consequently giving rise to the phenomenon of customer churn, which is a significant concern for most banks. The term "churn" refers to the act of terminating or discontinuing the utilization of a product or service provided by a business entity [1].

Financial institutions have come to realize that maintaining customer relationships is pivotal to their success. Customer churn prediction is an area of machine learning that focuses on predicting whether a customer will stay with or leave a business. Churn prediction algorithms identify customers who are likely to discontinue using a service or product. This is a major concern for product providers, as many churning customers not only diminish revenue but also adversely affect a company's reputation [2]. The issue of customer churn is a challenge that the banking industry encounters. Consequently, the Ethiopian banking sector is similarly impacted by customer churn. The primary emphasis of this study is to center on forecasting customer churn for Awash Bank, one of the pioneering private banks in Ethiopia[3].

The Ethiopian banking industry presently accommodates a vast number of users, thereby posing a formidable challenge in the analysis and prediction of consumer attrition. Notably, the industry experiences a considerable degree of customer turnover, attributable to its rapid expansion and clients' discontent with the service. Regrettably, the banking industry invests minimal effort in anticipating the reasons for clients' account closures, as it prioritizes the acquisition of new clients over retaining existing ones. At Awash Bank, a significant number of customers have initiated the closure of their accounts. As an illustration, within the past five years, a staggering sum exceeding 21,000 customers have made the decision to terminate their accounts solely within the confines of the Awash Bank Wolaita Sodo region, encompassing a total of 44 branches. Presently, in order to mitigate the issue of customer churn, Awash Bank has adopted the dormant account system, which effectively identifies customers who have failed to execute any transactions within the preceding half-year duration. Thus, this system is instrumental in revealing those customers who have abstained from

Cite (APA): Mekuria, A.M., Yimer, M.A., & Woldehana, Y. D. (2025). Customer Churn Prediction Using Machine Learning Techniques: Awash Bank Wolaita Sodo Region. *Journal of Emerging Computer Technologies*, 5(1), 36-46. Doi: 10.57020/ject.1623937

conducting any financial activities in the past six months. However, there remains an absence of a predictive system that forewarns of impending customer churn.

There has been numerous machine learning- and deep learning-based studies conducted that pertain to the prediction of customer churn, employing various algorithms. These investigations have exhibited primary deficiencies, such utilization of unbalanced data for model development, selection of a technique to address the issue of missing values, and the process of feature selection. In the majority of the studies conducted[4, 5], the model significantly relies on a single attribute. Unfortunately, many researchers have failed to implement any techniques to tackle this issue. Moreover, the majority of local research have solely focused on commercial bank of Ethiopia (CBE), disregarding private banks and other areas.

II. Related Works

A number of studies have attempted to investigate customer churn. Among these, Gebremeskel [6] attempted to estimate customer turnover for CBE using data mining techniques such as J48, LR, and bagging. The investigator employed a total of nine attributes for the purpose of this investigation. Within these attributes, one of them happens to be the current balance. The issue with this particular attribute lies in the fact that the current balance of a closed account consistently remains at zero. Hence, it is only active accounts that possess a current balance. This observation reveals that this attribute exhibits a significantly strong correlation with the class churn, consequently resulting in the model's dependence solely on the value of this attribute. Additionally, the dataset lacked demographic information about the customers also. Furthermore, Gebremeskel[6] employed a support vector machine (SVM), K-nearest neighbor (KNN), Deep Neural Network (DNN), Random Forest (RF), and logistic regression (LR) to forecast CBE customer turnover. This investigation bears significance to the proposed study, and the study employed a substantial quantity of data for the purpose of model development. However, owing to the highly imbalanced nature of the dataset, consisting of 159,570 instances for the non-churn category and 44,591 instances for the churn category, a strategy was implemented by the researcher to address this issue. Specifically, the SMOTE oversampling technique was utilized, which resulted in the transformation of both categories to 159,570 instances. Consequently, within the realm of the churn category, a total of 114,979 erroneous records were identified, indicating a predominant presence of false data points within the dataset under examination. As a result, the overall performance of the model was negatively affected. A study [7] used SVM, KNN, LR, and Naïve Bayes to forecast customer turnover for CBE. Aside from the Commercial Bank of Ethiopia. This research bears relevance to the proposed research, whereby 54,623 records with 34 attributes were utilized for model development purposes. Notably, two primary gaps are observed in this study. Initially, the handling of categorical attributes was found to be inadequate. The researcher refrained from employing encoding techniques such as label encoding or one-hot encoding, opting instead for a manual approach to convert categorical values into numeric counterparts. This methodology introduces the potential for human error. Furthermore, the researcher neglected to employ techniques for handling missing values, such as imputation or others. Instead, a total of 420 records were simply removed from the dataset. This course of action is ill-advised due to the researcher's limited data availability. The study [8] used SVM, KNN, RF, Naive Bayes, and DNN to forecast client turnover for Lion Insurance. This study a total of nine characteristics were employed in order to forecast customer attrition for the insurance provider, Lion Insurance. Additionally, an extra-tree classifier was utilized by the investigator to ascertain the significance of the features. The graph depicting feature importance unequivocally demonstrates that the model is heavily reliant on a sole attribute, namely the premium attribute, and the researcher neglected to employ any methodologies to address this matter. Additionally, the dataset solely comprises two years' worth of customer data. The study [9] used 204,161 observations with eleven attributes. The other study [6] used 13,172 customer records with nine characteristics and a dataset containing 628,634 transactions; however, this dataset did not include consumer demographic data. Furthermore, the study [7] used five years of CBE customer data, which contained 54,623 customer records with 34 factors. The most current Lion Insurance research [8] used two years of customer data acquired from one Lion Insurance and included 12,007 records with nine features.

Additionally, there are a lot of researchers who use various methodologies to anticipate client attrition. Among these, Rahman and Kumar [10] work used four machine learning algorithms to forecast customer turnover in banking using a Kaggle dataset. The dataset utilized in this study displays a significant lack of balance, as it consists of 7,963 instances of non-churn and merely 2,037 instances of churn. In order to address this issue of imbalance, the researcher opted to employ SMOT oversampling, a technique that generates an excessive quantity of fabricated data points. Consequently, this approach introduces a notable degree of falsified information into the dataset. Moreover, the researcher adopted balance as an attribute of interest for this study; however, it must be acknowledged that the churned customers exhibit a balance of zero, rendering it unsuitable to utilize this particular attribute in the context of this study. Furthermore, IEEE Communications Society [11] compares the SVM model to ANN, decision trees (DT), LR, and Naïve Bayes classifiers to predict customer turnover for China Construction Bank (CCB) VIP users. However, the data collection for this investigation was restricted solely to one branch, with an exclusive focus on VIP customers. The dataset encompasses a mere eight-month timeline from April 2007 to December, and the dataset employed for this study comprises a limited amount of data. Likewise, Dalmia et al. [12] seeks to predict client attrition in banking using two supervised machine learning algorithms, KNN and XG Boost. Another attempt, He et al. [13] , used SVM to predict client attrition for a Chinese commercial bank. When we look at the performance of the model in most research investigations, the deep neural network is the model that outperforms the rest. DNN outperforms the other four machine learning algorithms in the study [9], with an accuracy of 79.32%. The study by [8] is another study that indicates deep learning systems outperform. With 97.04% accuracy, DNN surpassed the other machine learning models. Gebremeskel [6] determined that J48 was the best model, with an accuracy of 94.8%. The KNN model was chosen as the best model for the study [7], with an accuracy of 99.91%.

A. Gap Analysis

There are numerous studies that have been conducted pertaining to the prediction of customer Numerous studies have been conducted on the prediction of customer churn, employing various machine learning and deep learning models. These are the primary deficiencies in the prior investigations. The first concerns the selection of a dataset-balancing technique. Numerous researchers have employed the SMOTE oversampling method. This approach oversamples the minority classes to match the majority ones, but it is only applicable if the disparity between the class data is not substantial. In cases where the dataset is highly imbalanced, this can generate a considerable number of synthetic data. Furthermore, some researchers have also utilized an imbalanced dataset for model development. Consequently, the model performs well for classes that possess a sizable number of records. The second issue revolves around the selection of a technique to address missing values. Certain researchers possess a limited dataset and thus eliminate the row that contains the missing value. The third challenge arises when converting categorical values to numerical values. Certain researchers have employed manual methods for this conversion instead of utilizing alternative techniques such as label encoding, one-hot encoding, binary encoding, or others. The fourth concern relates to feature selection. In most studies, the model heavily relies on a single attribute. Regrettably, most researchers have neglected to implement any techniques to address this matter. The fifth drawback is majority of previous studies, the current balance attribute was selected. However, it is illogical to use this attribute since the current balance of a churned customer is zero, and only active accounts possess a current balance.

III. Materials and Methods

A. Data Collection Method

A combination of primary and secondary data collection methods was employed. The primary data collection methods involved conducting interviews with domain experts, including customer service officers, quality assurance officers, and directors in the Awash Bank Wolaita Sodo region. The region is responsible for managing the information of 44 branches in various zones and special Woredas. Furthermore, the researcher acquired data from Awash Bank Wolaita Sodo, which contained details regarding active customers and a report on customers who had closed their accounts.

B. Dataset Description

The dataset was gathered from the Awash Bank Wolaita Sodo region, which encompasses forty-four branches across ten zones and one special woreda. The researcher obtained information on active and closed account customers separately. The active account dataset consists of 32,373 records, while the closed (churn) account dataset comprises contains 20,446 detail records of customers who closed their Awash bank account between January 1, 2018, and December 31, 2022. Overall, the research collected 52,819 customer records, spanning fourteen attributes, including account number, name, sex, telephone number, date of birth, civil status, product type, branch, opening and closing dates, ATM card, current balance, mobile banking, and internet banking. For active accounts, instead of indicating the date of closure, the database contains information regarding the current balance.

C. Data Preprocessing

1. Data Standardization and Transformation

The dataset consists of 18 attributes, encompassing derived attributes such as tenure, age, location, and churn (target class), out of which 7 have been excluded, and the remaining 11 are employed for model development. Among the 11 attributes, four of them possess yes and no values, namely ATM, mobile banking, internet banking, and churn, which are converted to 1 and 0, respectively. Age has numerical values, which are normalized using a standard scalar (Z-score), while the other six possess categorical values, namely sex, tenure, civil status, product type, branch, and location. These categorical attributes were converted to numerical values using label encoding.

2. Dataset Preparation for Model Development

After concluding the necessary preprocessing steps, the resultant dataset comprises ten attributes, including, sex, age, civil status, location, branch, product type, tenure, ATM, Mobile banking, Internet Banking, and Churn. Ultimately culminating in a final dataset that consisted of 50,987 records. Furthermore, within this final dataset, 31,619 records were assigned to the "non-churn" class, while 19,368 were assigned to the "churn" class. To tackle the problem of an imbalanced dataset, a method of SMOTE-ENN sampling was employed. After employing this method to achieve class balance within the dataset, the resulting dataset encompasses a total of 23,954 data entries for non-churn, while the remaining 22,739 data entries correspond to churn class.

3. Feature Importance

The utilization of an extra-tree classifier model is employed to evaluate the significance of features. The Extra Trees Classifier is selected for predicting customer churn because it can deal efficiently with high-dimensional data and is resilient to overfitting, which is important in finding intricate patterns in customer behavior. Moreover, its quick training time and simplicity of interpretation using feature importance scores make it well-suited to comprehend which variables have the largest impact on churn. The Extra Tree Classifier, through computation of the average impurity decrease caused by each feature across the ensemble of decision trees, estimates the relevance of each feature. The results, depicted in Figure 1, exhibit a distinct attribute arrangement based on respective relevance scores in descending order. A high importance score or weight of an attribute indicates its utmost significance, whereas a lower importance score implies insignificance. Tenure carries the greatest importance, with an importance score greater than 0.25, whereas age and branch have an importance score greater than 0.2.

According to the Extra Tree classifier output, the features that have the greatest significance are tenure, age, and branch. Meanwhile, the features' location, ATM, product type, and mobile banking are found to be significant, but not as much as the aforementioned features. Moreover, the remaining attributes are deemed less important, such as civil status, internet banking, and sex. From the dataset, it can be inferred that customers who use ATM cards and mobile banking are also less likely to churn from the bank. Moreover, customers with longer tenure are less likely to churn. Based on the outcome of the extra tree classifier and considering the restricted number of features, it can be observed that all attributes are of substantial importance in forecasting customer churn. As a result, all features have been chosen for the development of the model.

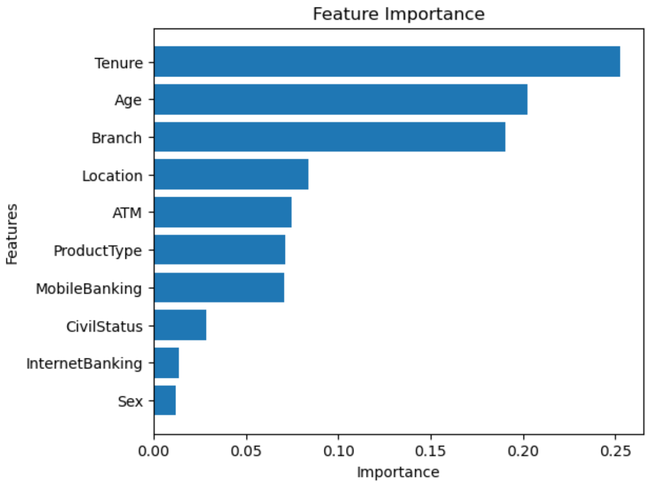

Figure 1: Feature importance testing results

4. Train and test split

There are various concerns that arise in relation to the selection of the train-test split. For instance, if an excessive amount of data is allocated for training purposes, there is the potential for bias, overfitting, and issues pertaining to generalizability. In such cases, the model tends to excessively memorize the training data, resulting in a lack of ability to generalize well to unseen data. Consequently, this can lead to suboptimal performance on new data or in situations that differ from the training data. Therefore, the selection of an optimal train-test split is a crucial step in the development of machine learning models. It serves to mitigate bias, enhance generalization, estimate uncertainty, facilitate fair model comparisons, identify overfitting, effectively select hyperparameters, and augment the trustworthiness of the models.

For the purpose of this research thesis, a variety of train test splits are used, including 90/10, 80/20, 70/30, and 60/40, in order to determine the optimal split. The optimal train and test split denote the most advantageous partitioning of a dataset into training and testing sets for the purpose of machine learning or data analysis endeavors. The ultimate aim is to establish a division that allows the model to be trained with maximum efficacy and evaluated with utmost precision.

To prevent any data splitting bias, stratified sampling is employed to maintain the same proportion of the target class in training and test sets, particularly for imbalanced datasets. Furthermore, k-fold cross-validation and group k-folds help provide a strong model assessment by utilizing all data points without leading to information leakage.

D. Proposed Model Development

1. Proposed Model Design

The proposed model is implemented for predicting customer churn, utilizing the data gathered from the Awash Bank Wolaita Sodo Region. As illustrated in Figure 2, the process of data preprocessing is performed in a sequential manner after the collection of data from the study area in order to remove any potential anomalies within the collected dataset. The dataset is partitioned into training and testing sets subsequent to the preparation phase. The model is trained by utilizing the training set, following which model performance is evaluated using the test set. The selection of the model for the study at hand is contingent upon several factors. As this research study is experimental, the optimal model is selected during the model selection stage, subsequent to an exhaustive experiments and performance evaluation. Finally, the results are documented and reported.

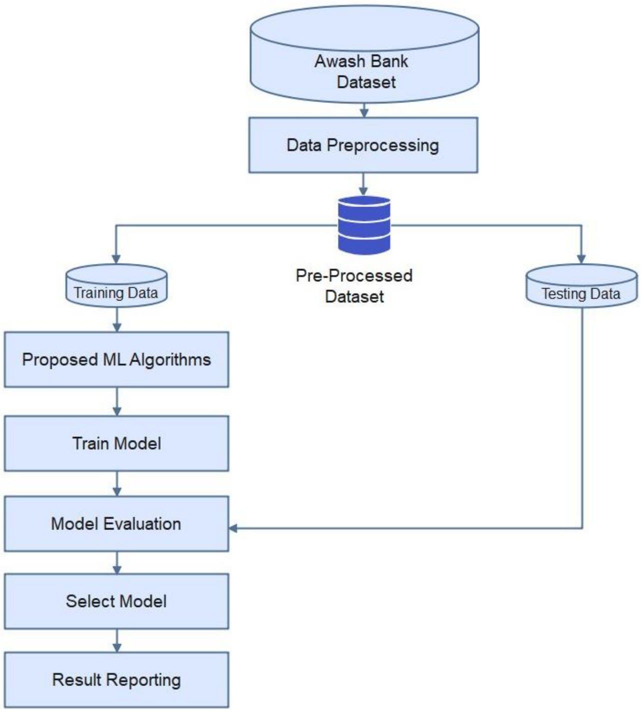

Figure 2: Proposed model design

IV. Experimental Results and Discussion

A. Experimental Setup

To enable the development of the proposed model, several experimental setups are imperative. These include the installation of Anaconda 3 version 2023.03-1, an open-source platform that enables the writing and execution of Python code. Within this platform, users have access to several code editors, such as Jupiter Notebook, PyCharm, Spyder, and more. In the present research endeavor, Jupiter Notebook version 6.5.2 was employed. Additionally, the creation of the proposed model demands the employment of diverse libraries and packages, including Pandas, Sklearn, TensorFlow, Matplotlib, Seaborn, Numpy, Imblearn, and Pickel, amongst others. These libraries provide a unique collection of packages and functions that can be utilized to execute a broad spectrum of responsibilities.

B. Data preparation for Model Development

After conducting the necessary preprocessing steps, the resultant dataset comprises eleven attributes, including, sex, age, civil status, location, branch, product type, tenure, ATM, Mobile banking, Internet Banking, and Churn. In order to select the most significant features, extra tree classifier feature selection was employed. ultimately culminating in a final dataset that consisted of 50,987 records. Furthermore, within this final dataset, 31,619 records were assigned to the "non-churn" class, while 19,368 were assigned to the "churn" class.

To tackle the problem of an imbalanced dataset, a method of SMOTE-ENN (Synthetic Minority Over Sampling-Edited Nearest Neighbors) sampling was employed. After employing this method to achieve balance within the dataset, the resulting dataset encompasses a total of 23,954 data entries for non-churn, while the remaining 22,739 data entries correspond to churn class. From the entirety of the dataset, it can be observed that 51.2% of the data corresponds to active accounts, while the remaining 48.8% represents closed account information, as depicted in Figure 3.

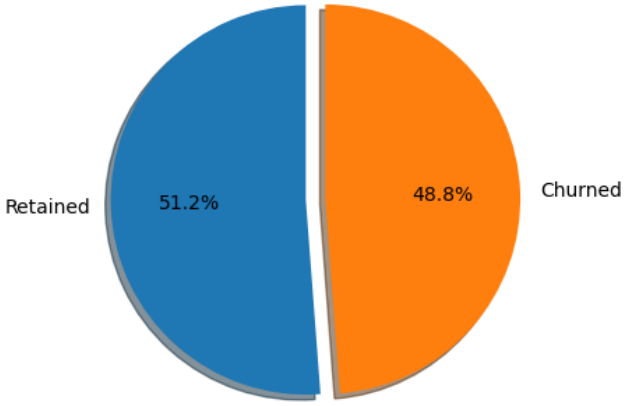

Figure 3: Visualization of the proposed research in percent

C. Result Discussion

A variety of performance measuring metrics including accuracy, recall, and precision were used to assess the performance of the model. Seven preeminent machine learning models were identified and tested for the study. The models include XGBoost, LigtGBM, DT, RF, DNN, GBM, MLP, and CNN. Other models including SVM, KNN, AdaBoosting, and logistic regression were excluded from this study as a result of their subpar performance. Furthermore, in the field of machine learning, hyperparameter tuning involves the careful examination and exploration of a set of optimal hyperparameter values for a specific model. In this study grid search hyperparameter tuning is implemented.

Various experiments were conducted on the aforementioned models for an optimal performance. When evaluating the performance of the models, it is evident that Random Forest outperforms the other models with an overall accuracy of 99.14% on the test set (unseen or holdout dataset). The GBM, XGBoost, LGBM, DT, LGBM, and MLP models achieved an overall accuracy of 99.12%, 98.69%, 98.34%, 98.23%, and 98.14%, respectively. On the other hand, the DNN and CNN achieved an overall accuracy of 96.33% and 93.19% respectively. Based upon the findings, it can be concluded that, within the context of the suggested investigation, machine learning models demonstrate a higher level of performance in comparison to deep learning models. This could potentially be attributed to the nature of the dataset; the proposed dataset is comprised of 50,987 observations. However, upon examining the number of columns, it becomes apparent that the suggested dataset solely encompasses 11 attributes. Consequently, this may be the underlying reason for the comparatively lower performance of the deep learning model. This is due to the fact that deep learning algorithms tend to be more effective in instances where the dataset exhibits a complex relationship, indicating a large number of rows and columns. Table 1 shows the train and the test accuracy of the proposed models.

Table 1: Train and test accuracy of proposed models

Models	Train Accuracy	Test Accuracy
RF	**99.88%**	**99.14%**
GBM	99.94%	99.12%
XGB	99.65%	98.69%
DT	99.06%	98.34%
LGBM	99.08%	98.23%
MLP	98.99%	98.14%
DNN	96.98%	96.33%
CNN	94.16%	93.19%

D. Machine Learning Models experimental result

1. Result for the Random Forest

The most optimal values for every parameter are meticulously chosen. These parameters include criterion, max_depth, min_samples_split, min_samples_leaf, random_state, and n_estimators. Table 2 displays the selected values.

Table 2: Selected parameter values for random forest model

Parameters	Values
criterion	entropy
max_depth	None
min_samples_leaf	1
min_samples_split	6
Random_state	1
n_estimators	200

As demonstrated in Table 3 of the classification report for the random forest, this model achieves a remarkable degree of accuracy at 99.14%. Moreover, when considering the precision, recall, and f1-score for classes 0 and 1, this model performs exceedingly well, achieving 99% for both classes. It is worth noting that the support for class 0 is 4548, whereas the support for class 1 is 4791.

Table 3: Classification report for random forest model

Classes	Precision	Recall	F1-score	Support
0	99%	99%	99%	4548
1	99%	99%	99%	4791
Accuracy			99%	9339
Macro avg	99%	99%	99%	9339
Weighted avg	99%	99%	99%	9339

From Figure 4, it can be observed that the random forest model, with a sample size of 4548 records, has accurately classified 4512 records and misclassified 36 records for class 0. Similarly, for class 1, this model from the same sample size of 4791 records correctly classified 4747 records, while the remaining 44 records were misclassified as class 0.

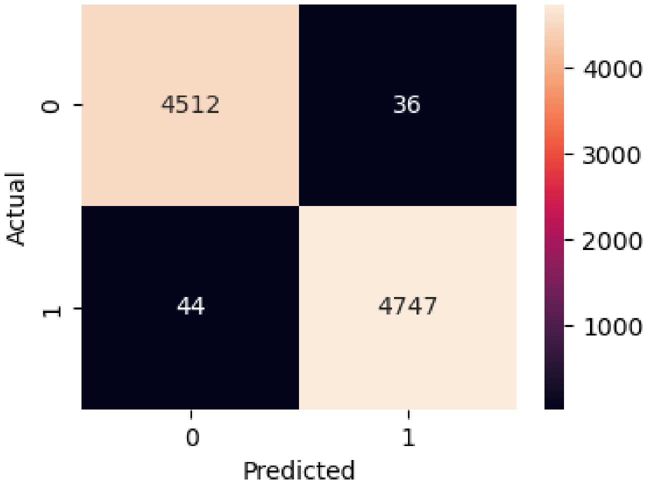

Figure 4: Confusion matrix for random forest model

2. Extreme Gradient Boosting (XGBoost) result

The grid search methodology is utilized to elect the optimal parameters, and the most suitable parameter values are determined, as presented in Table 4.

Table 4: Selected parameter values for XGBoost model

Parameters	**Values**
colsample_bytree	1
learning_rate	0.1
max_depth	7
n_estimators	300
subsample	0.9

As is evident from the classification report presented in Table 5, the XGBoost model demonstrates a remarkable accuracy of 98.69%. The precision, recall, and f1-score values for both classes 0 and 1 are 99%. It is noteworthy that class 0 has a support of 4548, whereas class 1 has a support of 4791.

Table 5: Classification report for random forest model

Classes	**Precision**	**Recall**	**F1-score**	**Support**
0	99%	99%	99%	4548
1	99%	99%	99%	4791
Accuracy			99%	9339
Macro avg	99%	99%	99%	9339
Weighted avg	99%	99%	99%	9339

As per the outcomes derived from the model, as indicated in Figure 5, for class 0, it can be observed that out of a total of 4548 records, 4487 records have been accurately classified as class 0. However, the remaining 61 records have been mistakenly classified as class 1. In the instance of class 1, out of a total of 4791 records, 4730 were accurately classified as class 1, while the remaining 61 records were erroneously classified as class 0 by the model.

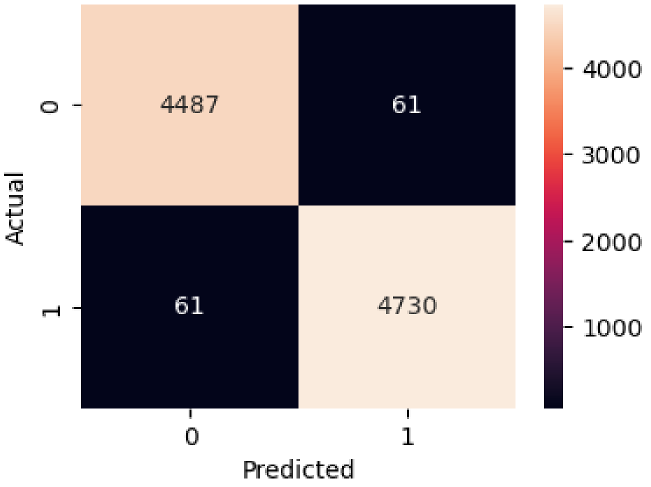

Figure 5: Confusion matrix for XGBoost model

3. Decision Tree Experimental Result

The most suitable parameter values for decision tree model are determined, as presented in Table 6.

Table 6: Selected parameter values for decision tree model

Parameters	**Values**
criterion	entropy
max_depth	None
min_samples_leaf	1
min_samples_split	12

As evident from the classification report, the decision tree model has demonstrated a remarkable accuracy of 98.34%. Furthermore, the precision, recall, f1-score, macro average, and weighted average of the decision tree model all stand at a commendable 98%. It is noteworthy that the total classification support for both classes amount to 9339. Table 7 displays the report pertaining to the decision tree model.

Table 7: Classification report for decision tree model

Classes	**Precision**	**Recall**	**F1-score**	**Support**
0	98%	98%	98%	4548
1	98%	98%	98%	4791
Accuracy			98%	9339
Macro avg	98%	98%	98%	9339
Weighted avg	98%	98%	98%	9339

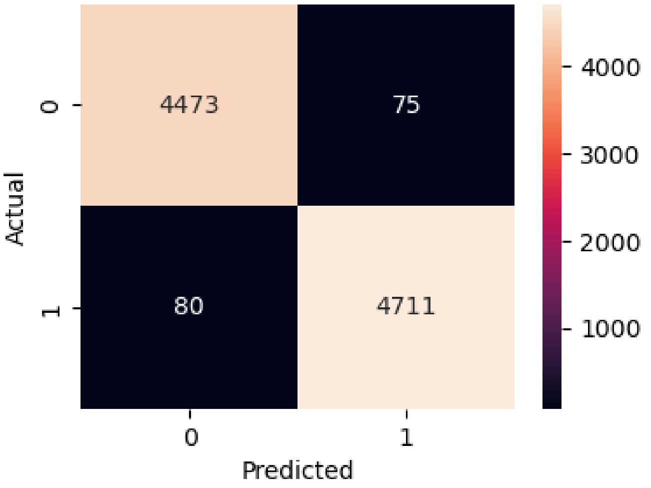

Figure 6: Confusion matrix for decision tree model

From the confusion matrix depicted in Figure 6, it is observed that out of the total of 4548 records, 4473 are aptly classified as class 0, while 75 exhibit misclassifications as class 1. For class 1, out of a total of 4791 records, 4711 have been accurately classified, while the remaining 80 have been incorrectly classified as belonging to class 0.

E. Results for the Deep Learning Models

1. Multi-Layer Perceptron (MLP) Experimental Result

There are various train-test splits utilized in the application of this model; however, the model boasts an impressive performance in the 80/20 split. Table 8 illustrates the most optimal values of the parameters in this particular model.

Table 8: Selected parameter values for MLP model

Parameters	**Values**
Learning_rate	0.001
Hidden_layer_size	(100,100)
Activation	Tanh
Solver	Adam

From the table presenting the classification report, it can be observed that the model exhibits an accuracy rate of 98.14%. Moreover, this model exhibits remarkable values for precision, recall, f1-score, macro average, and weighted average, all of which are at an impressive 98%. Table 9 presents the classification report for the MLP model.

Table 9: Classification report for MLP model

Classes	**Precision**	**Recall**	**F1-score**	**Support**
0	98%	98%	98%	4548
1	98%	98%	98%	4791
Accuracy			98%	9339
Macro avg	98%	98%	98%	9339
Weighted avg	98%	98%	98%	9339

As discernible from Figure 7 of the MLP confusion matrix, the present model adeptly categorizes 4471 of the 4548 total records as class 0. while the residual 77 are inaccurately allocated to the class. For class 1 of the 4791 records, this particular model accurately identifies 4694 records, while 97 records are erroneously categorized as class 0.

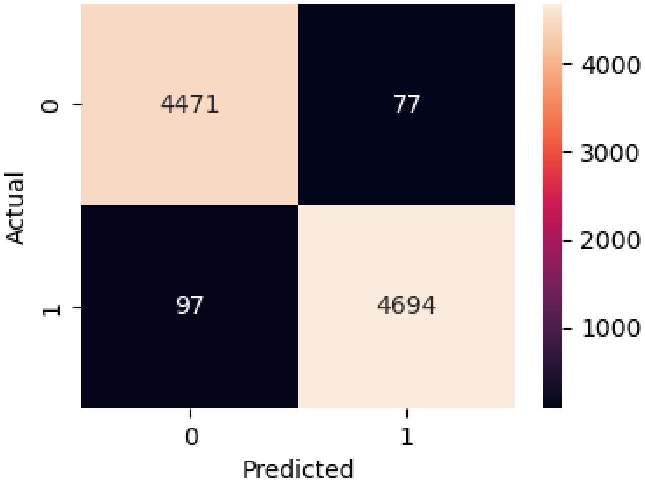

Figure 7: Confusion matrix for MLP model

F. One-Dimensional Convolutional Neural Network (1D-CNN) experimental result

Most often, the CNN model is deemed preferable for images; however, it exhibits commendable performance for the proposed tabular dataset. For the proposed model, a 1D-CNN is used because the dataset is tabular. The present model entails diverse parameters, namely the learning rate, optimizer, kernel size, and batch size, among others. Table 10 shows the exhibited optimal values for every parameter.

Table 10: Selected parameter values for 1D-CNN model

Parameters	**Values**
Learning_rate	0.001
optimizer	Adam
Kernel_size	3
Batch_size	128
Output layer activation	Sigmoid

For this particular model, various train and test splits have been executed, culminating in the 70/20/10 split attaining the most optimal outcome. This implies that 70% of the dataset is allocated for training purposes, 20% for testing, and the residual 10% for validation. With the implementation of a train-test split and 200 epochs, the aforementioned model attains a notable 93.19% level of accuracy. Furthermore, both classes 1 and 0 exhibit precision, recall, and f1-scores of 93%. Table 11 represents the classification report for the 1D-CNN model.

Table 11: Classification report for 1D-CNN model

Classes	**Precision**	**Recall**	**F1-score**	**Support**
0	93%	93%	93%	4548
1	93%	93%	93%	4791
Accuracy			93%	9339
Macro avg	93%	93%	93%	9339
Weighted avg	93%	93%	93%	9339

From Figure 8, the confusion matrix of the 1D-CNN model, it can be inferred that for class 0, out of a total dataset of 4548, this model accurately classified 4222 records as class 0, while misclassifying the remaining 326 records as class 1. For class 1, among a total of 4791 observations, the model accurately assigned 4481 observations to class 1, while the remaining 310 observations were erroneously classified as class 0.

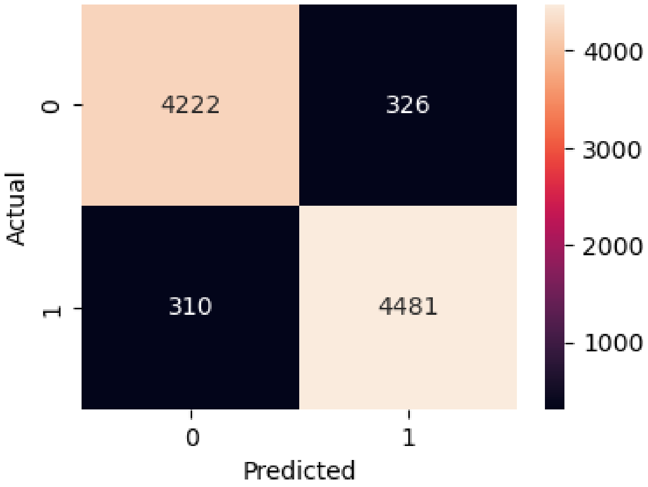

Figure 8: Confusion matrix for 1D-CNN model

As depicted by the training and testing accuracy graph of the 1D-CNN model in Figure 9, it is evident that the model is devoid of overfitting or underfitting predicaments. The model exhibits an optimal fit between its training and validation accuracy. The training and validation losses of the model are also fitting adequately without any concerns of underfitting or overfitting.

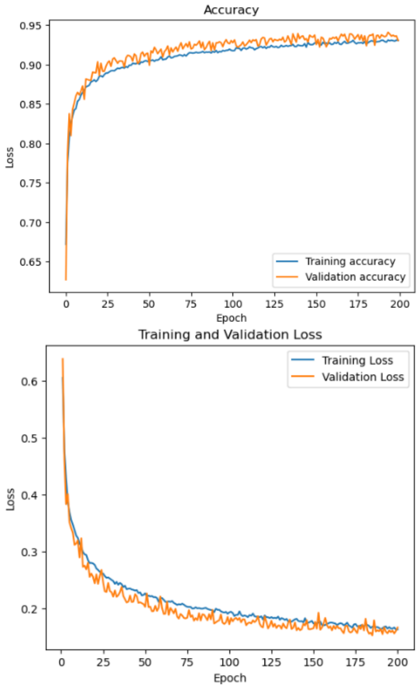

Figure 9: The training and validation loss and accuracy for the CNN model

Figure 10 illustrates the results of the ROC curve for the 1D-CNN model, achieving an impressive AUC of 98.39%.

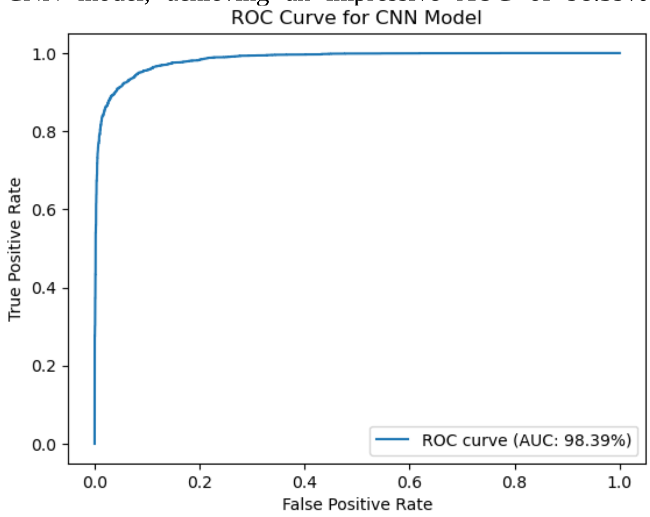

Figure 10: ROC Curve for 1D-CNN model

G. Deep Neural Network (DNN) experimental result

For the proposed investigation, one of the deep learning models is the DNN, which is employed. This model concerns parameters such as learning rate, optimizer, and batch size. The optimal value for each of these parameters has been delineated in Table 12. To construct this particular model, varying train-test splits were utilized, ultimately resulting in the superior performance of the 70% training, 20% testing, and 10% validation split.

Table 12: Selected parameter values for DNN model

Parameters	**Values**
Learning_rate	0.001
Optimizer	Adam
Batch_size	128
Output layer activation	Sigmoid

The deep neural network model is comprised of multiple layers, including an input layer, three hidden layers, three dropout layers, a batch normalization layer, and an output layer. At every stratum, a 0.25 dropout rate is implemented. Moreover, concerning the count of neurons present in each hidden layer, the first, second, and third layers, respectively, contain 200, 180, and 200.

This particular model employed the Adam optimizer in conjunction with binary cross-entropy loss and accuracy metrics. Additionally, it utilized a learning rate of 0.001 and underwent 200 epochs with 128 batch sizes. The DNN model demonstrates a remarkable level of accuracy, reaching 96.33%. Furthermore, it achieves precision, recall, and a f1-score of 96% for both classes 0 and 1. It is worth mentioning that the support for class 0 is 4588, while for class 1, it is 4813.

Table 13: Classification report for DNN model

Classes	**Precision**	**Recall**	**F1-score**	**Support**
0	96%	96%	96%	4588
1	96%	96%	96%	4813
Accuracy			96%	9041
Macro avg	96%	96%	96%	9041
Weighted avg	96%	96%	96%	9041

As depicted in Figure 11 confusion matrix outcome, it is evident that for class 0, out of a total of 4588 records, 4412 records are accurately classified as class 0, while 176 records are erroneously classified as class 1. Similarly, for class 1, 4644 records are correctly classified, while 169 records are mistakenly classified out of the same 4813 total records.

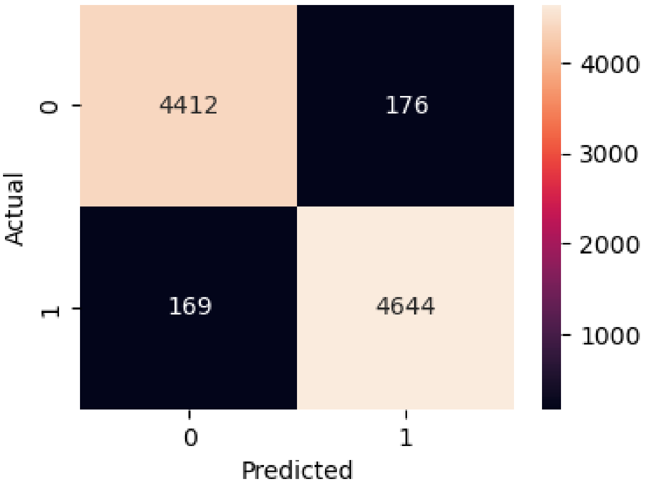

Figure 11: Confusion matrix for DNN model

The accuracy of both training and validation in the DNN model indicates that the model is devoid of any overfitting or underfitting issues. This observation is likewise evident in the training and validation losses. Figure 12 indicates the training and validation accuracy for the DNN model, also it shows the training and validation losses for the DNN model.

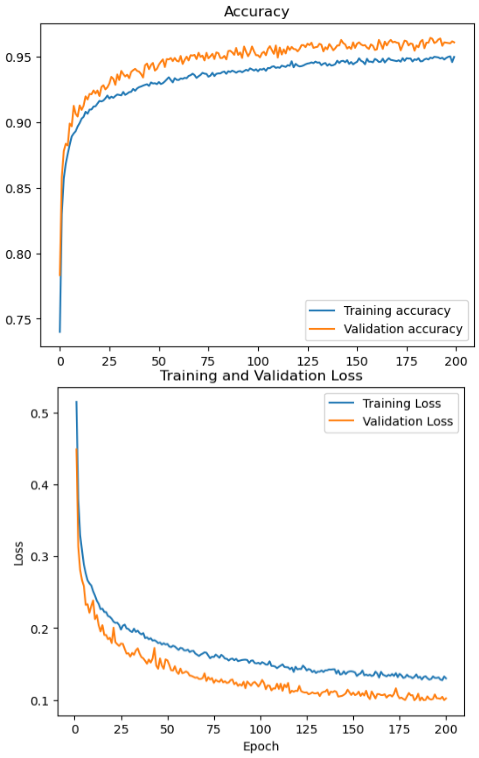

Figure 12: The training and validation loss and accuracy for the DNN model

Figure 13 depicts the ROC curve for the DNN model, exhibiting an impressive AUC value of 99.42%.

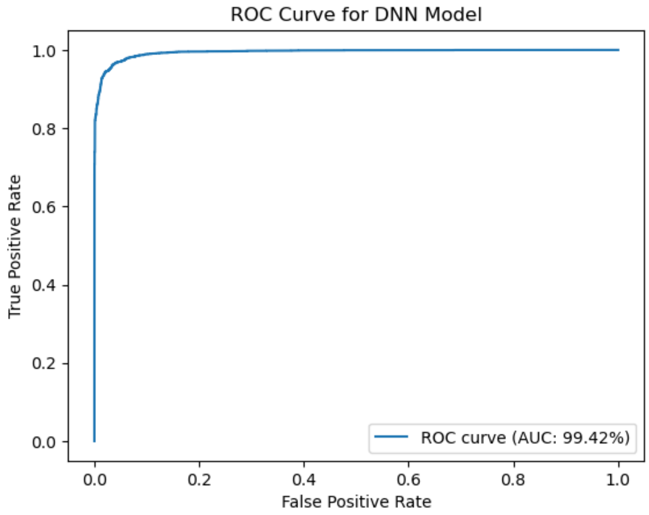

Figure 13: ROC Curve for DNN model

H. *Model Comparison Using Different Train-Test Splits*

For the purpose of electing the optimal train-test partition, various partitions are implemented on the proposed dataset, such as 90/10, 80/20, 70/30, and 60/40. As illustrated in Table 14, the models' performance across the four train-test splits exhibits a negligible discrepancy. However, it is noteworthy that the 90/10 and 80/20 splits predominantly showcase optimal performance in contrast to the 70/30 and 60/40 splits. For this study, the optimal split of 80/20 for the train test was chosen based on the obtained results. In the case of two deep learning algorithms, CNN and DNN, a validation dataset comprising 10% of the entire dataset was employed.

Table 14: comparison of model's performance based on train-test split

Models	**90/10**	**80/20**	**70/30**	**60/40**
RF	99.31%	99.14%	98.97%	98.67%
GBM	99.25%	99.12%	99.12%	99.08%
XGB	99.04%	98.69%	98.74%	98.65%
DT	98.54%	98.34%	97.97%	97.56%
LGBM	98.52%	98.23%	99.29%	98.23%
MLP	98.31%	98.14%	96.93%	96.98%
	80/10/10	**70/20/10**	**60/30/10**	**50/40/10**
DNN	96.83%	96.33%	96.33%	96.05%
CNN	93.58%	93.19%	93.98%	93.28%

I. *Model comparison using different data balancing techniques*

Due to the unequal distribution of records for both classes in the dataset adopted for the intended research analysis, diverse methods for balancing the dataset are implemented to address the issue of imbalance. There exist various methodologies that can be employed in the context of imbalanced datasets, including under-sampling, over-sampling, the combination of both techniques, and utilizing the original dataset in its imbalanced form for model construction. The proposed model was formulated utilizing disparate datasets, including an imbalanced dataset, an under-sampled dataset, an over-sampled dataset, and a composite dataset.

Based on the outcomes of the model for every dataset, it can be observed that the combination balancing methodology surpasses other techniques. The utilization of the over-sampling technique to equalize the dataset exhibits the most exceptional performance, second only to the combination balancing method. The evaluation accuracy for each balancing method is presented in Table 15.

Table 15: Comparison of dataset balancing techniques

Models	**Unbalanced**	**Under-sampling**	**Over-sampling**	**Hybrid**
RF	87%	86%	88%	99%
GBM	87%	86%	88%	99%
XGB	91%	87%	89%	99%
DT	86%	85%	86%	99%
LGBM	86%	87%	87%	98%
MLP	88%	86%	87%	98%
DNN	88%	86%	86%	96%
CNN	86%	84%	86%	93%

J. *Comparison of proposed model performance with previous research models*

In the domain of customer retention, the precise anticipation of customer churn assumes an essential role in achieving business accomplishment. The advent of machine learning has presented itself as a formidable instrument for recognizing potential churners, empowering businesses to proactively execute strategies to retain valuable clientele. In this particular context, the evaluation of the effectiveness of suggested machine learning models for customer churn prediction is of paramount importance in order to appraise their efficacy and pinpoint areas where enhancements can be made. Numerous investigations have been carried out in this particular field. When we assess the efficacy of the models, the suggested model outshines in the majority of instances. This might be attributed to the meticulous selection of methodologies, such as data balancing techniques, normalization techniques, missing value handling techniques, encoding techniques for the conversion of categorical values into numerical values, feature selection techniques, model selection, and others. Based on extensive research, here we present a comparison of model performance with the most relevant preceding studies.

The primary research study [9], according to this study, the DNN exceeded the performance of the machine learning models with an accuracy of 79.32%. One plausible explanation for this could be the extensive dataset used, which amounts to 204,161 instances. Consequently, this abundance of data serves as a rationale for the superiority of the deep learning model over the machine learning model. As for data balancing, the researcher opted for the SMOTE method, yet due to the highly imbalanced nature of the dataset, it generates a significant number of fabricated data. This could account for the subpar performance of the model.

The second scholarly article [8], this particular research investigation was conducted using a dataset spanning two years, consisting of precisely 12,007 records that were gathered from the organization Lion Insurance. Using the aforementioned dataset, the optimal performance of the models yielded a DNN with an accuracy rate of 97.04%. Even if the performance of this particular model is commendable, the prediction made by the model is greatly contingent upon the premium attribute, as can be deduced from the outcome of the feature importance analysis. Furthermore, the researchers have not employed any technique to address this matter.

Table 16: Summery table for comparison of proposed model with previous models

Study	Model Type	Dataset Size	Accuracy	Key Findings
Proposed Model	RF	50,987	99.14%	Outperforms most local and global studies, emphasizing robust data handling and feature selection techniques.
[9]	DNN	204,161 instances	79.32%	Higher accuracy attributed to the extensive dataset; use of SMOTE led to high imbalances affecting performance.
[8]	DNN	12,007 records	97.04%	Strong performance influenced heavily by the premium attribute; no techniques applied to mitigate feature dependence.
[7]	KNN	54,623 records	99.91%	Close performance to proposed model due to careful technique selection in the study; also highlights effective modeling.

The third research [7] involved the utilization of a comprehensive set of 54,623 customer records sourced from the CBE. Using the aforementioned dataset, the KNN classification algorithm achieved an impressive accuracy rate of 99.91%. Furthermore, it is worth noting that the model performance exhibited in this particular research study is in close proximity to the proposed model performance. This is due to the researcher's careful selection of techniques for his study. In general, it can be stated that the proposed research model exhibited superior performance in comparison to the majority of local and global research studies that are related.

V. Conclusions

Customer churn denotes the deliberate discontinuation of the utilization of a product or service offered by a business entity. The primary objective of this research endeavor is to develop customer churn prediction model for the Awash Bank Wolita Sodo region. This research holds immense importance for the bank, as it enables the bank to safeguard its business and mitigate the decline in profit by minimizing and restricting customer churn. To construct the suggested model, data was obtained from the Awash Bank. The study selected 11 (eleven) attributed such as tenure, age, product type, location, branch, civil status, ATM status, internet banking status, mobile banking status, and churn (target class). After data preprocessing the dataset comprises of 31,619 active accounts and 19,368 closed accounts. In order to achieve balance within the dataset, the SMOTE-ENN technique is employed. Eight prominent machine learning models including XGBoost, LGBM, GBM, RF, DT, DNN, CNN, and MLP are tested. To determine the ideal values for each model, grid search hyperparameter tuning technique is utilized. Model performance has been observed to yield acceptable results, with Random Forest exhibiting a 99.14% accuracy rate. The GBM, XGBoost, DT, LGBM, and MLP models achieved accuracy rates of 99.12%, 98.69%, 98.34%, 98.23%, and 98.14%, respectively. The DNN has 96.33% accuracy, and lastly, the CNN has 93.19% accuracy. For the proposed model, Random Forest was chosen as the optimal model due to its impressive performance, exhibiting a remarkable 99.14% accuracy on the test dataset.

VI. Recommendations

The future researcher is advised to consider the following recommendations: The initial suggestion pertains to the approach employed in managing an imbalanced dataset. It is not advisable to utilize the SMOTE oversampling technique if the dataset exhibits a high level of imbalance. This is because said technique generates a substantial number of falsified data points. Instead, it is recommended to employ the SMOTE-ENN technique to balance the dataset. This technique involves the use of SMOTE for oversampling the minority class and ENN for under sampling the majority class.

The second recommendation concerns the significance of features. In certain cases, the feature's importance may indicate a high level of importance for a particular attribute. In such instances, it would be more appropriate to either eliminate said attribute or explore alternative methods of handling it, unless the model heavily relies on said attribute due to its strong correlation with the class.

Thirdly, it is advised to employ label encoding for converting categorical values into numerical values. One of the drawbacks of using one-hot encoding lies in the issue of feature selection, as it leads to the merging of features with values, thus rendering the selection of features a more challenging task. Moreover, it is important to acknowledge that the current study did not make use of transactional specifics. Consequently, it is suggested that a researcher choose to integrate transactional specifics into their investigation by implementing a range of methodologies, such as long short-term memory networks (LSTM), recurrent neural networks (RNN), and other relevant techniques with meticulous data balancing techniques. Furthermore, it is recommended that other scholars carry out their research on various financial institutions, insurance corporations, telecommunication enterprises, and other relevant entities. Finally, to use in actual business applications of the churn prediction model, such as targeted marketing campaigns, proactive customer contact through CRM systems, and ongoing strategy development based on the model's findings.

AKNOWLEDGEMENNT

We would like to express our heartfelt gratitude to Zelalem Bekele, the director of Awash Bank in the Wolaita Sodo region, for allowing us to utilize their essential resources and facilities. These invaluable resources and facilities played a crucial role in the successful execution of our research project. Additionally, we wish to acknowledge the entire staff of Awash Bank in this region for their unwavering support and for providing critical insights and information.

DATA AVAILABILITY

The data supporting the findings of this study are available upon request from the authors.

FUNDING

No funding was received for conducting this study.

AUTHORS' CONTRIBUTIONS

All authors contributed equally to this work.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

ETHICAL STATEMENT

In this article, the principles of scientific research and publication ethics were followed. This study did not involve human or animal subjects and did not require additional ethics committee approval.

REFERENCES

[1] Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*, 1-13. https://doi.org/10.1186/s40854-016-0029-6

[2] Jamjoom, A. A. (2021). The use of knowledge extraction in predicting customer churn in B2B. *Journal of Big Data*, *8*(1), 110. https://doi.org/10.1186/s40537-021-00500-3

[3] Arnaldo, M. (2003). Origins and Early Development of Banking in Ethiopia. *UNIMI Economics Working Paper No. 04.2003*, http://dx.doi.org/10.2139/ssrn.667265

[4] Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer churning analysis using machine learning algorithms. *International Journal of Intelligent Networks*, *4*, 145-154. https://doi.org/10.1016/J.IJIN.2023.05.005

[5] Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, *14*, 100342. https://doi.org/10.1016/J.RICO.2023.100342

[6] Gebremeskel, K. (2013). *Application of data mining techniques to predict customers' churn at Commercial Bank of Ethiopia* (Master's thesis). Addis Ababa University, School of Graduate Studies, School of Information Science.

[7] Gebreegziabher, B. (2022). *Bank customer churn prediction model: The case of commercial bank of Ethiopia* (Doctoral dissertation, St. Mary's University).

[8] Kingawa, E. D., & Hailu, T. T. (2022). Customer Churn Prediction Using Machine Learning Techniques: the case of Lion Insurance. *Asian Journal of Basic Science & Research*, *4*(4), 60-73. https://doi.org/10.38177/ajbsr.2022.4407

[9] Seid, M. H., & Woldeyohannis, M. M. (2022, November). Customer churn prediction using machine learning: commercial bank of Ethiopia. In *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICT4DA56482.2022.9971224

[10] Rahman, M., & Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In *2020 4th international conference on electronics, communication and aerospace technology (ICECA)* (pp. 1196-1201). IEEE. https://doi.org/10.1109/ICECA49313.2020.9297529

[11] IEEE Communications Society. (2008). WICOM 2008: 2008 International Conference on Wireless Communications, Networking and Mobile Computing: October 12-1, 2008, Dalian, China. IEEE.

[12] Dalmia, H., Nikil, C. V., & Kumar, S. (2020). Churning of bank customers using supervised learning. In *Innovations in Electronics and Communication Engineering: Proceedings of the 8th ICIECE 2019* (pp. 681-691). Springer Singapore. https://doi.org/10.1007/978-981-15-3172-9_64

[13] He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. *Procedia computer science*, *31*, 423-430. https://doi.org/10.1016/j.procs.2014.05.286