University of Ibadan
Journal of Science and Logics in ICT Research

A Comparative Analysis of Ensemble Machine Learning Algorithms for Bank Customer Churn Prediction

Ogunsanwo Gbenga. O.
ogunsanwogo@tasued.edu.ng
Department of Computer Science, College of Science and Information Technology, Tai Solarin University of Education, Ogun State. Nigeria

Abstract
Customers churn became a serious issues to banks manager because customers have numerous options where to save their money. This justify why many researchers are attracted to this area. This study developed a bank customers churn predictive model. The study used dataset from kaggle.com repository. It consists of 10127 instances and 20 parameters. One Hot Encoder was used as data preprocessing on the dataset. The data was divided into 80% for training and 20% for testing. The predictive model was created using Long Short-Term Memory (LSTM), Ensemble LSTM, and Random Forest (RF). The results of the model revealed LSTM with F1 score of 0.94, accuracy of 0.9235, specificity of 0. 6635 sensitivity of 0.97, AUC of 0.95 and loss value of 0.1663. Ensemble LSTM with F1 score of 0.94, accuracy of 0.9057, specificity of 0.554, sensitivity of 0.98, AUC of 0.92 and loss value of 0.238. RF with F1 score of 0.97, accuracy of 0.95, specificity of 0. 774, sensitivity of 0.99, AUC of 0.99 and loss value of 0.15. The study concluded that RF outperformed both LSTM and Ensemble LSTM. Also pointed out that customer's gender, marital status, customer income category and age against attrition are determining factor for customer churn prediction. The model is recommended for banking sector to assist in decision making. Future work can be done using more ensembles techniques and perform more data expository

Keywords: **Customers, Churn, LSTM, Ensemble, Random Forest**

1. Introduction

In this technological age, the banking industry has seen witnessing tremendous competition with their customers having various options to choose from unlike before that they are confined in the box. Some IT companies have developed robust platforms to provide payment processing, lending, and investment services activities traditionally dominated by banks.

In Nigeria, commercial Banks are facing great competition from online banking system like Opay, Palmpay and so on due to their seamless operations and ability to open account with lesser stress. One of the major challenges for the management of any Bank is how to retain their customers and stop them from leaving their bank. Customers acquisition come with high cost rate and to lose this customers will definitely have negative impact on the bank's growth and revenue generation Hejazinia *et al* [7] pointed out the need of developing tactics of customer retention strategies for any company success. This justifies the need for any banks identify issues that lead to customer churn and to develop a means to combat this menace in order to retain their customers. The term Customer churn is usually referred as customer moving or leaving the company services. In today's world of business there is great need to respond to changes in customer behavior especially in banking industry.

In the past, there are numerous traditional methods of predicting churns like customer survey method, manual data analysis all these come with flaws such as not giving accurate information and sometime are time consuming. With the introduction of data mining techniques and exploring the power of machine learning (ML) methods enable the analysis of large amounts of customer data, leading to highly accurate predictions of customer churn The goal of this paper is to develop a predictive model for Bank customer attrition. This is achieved by analyzing Bank customer data, including demographic information, transaction records, customer Attrition, Customer Age, Gender,

Ogunsanwo Gbenga. O. (2025). A Comparative Analysis of Ensemble Machine Learning Algorithms for Bank Customer Churn Prediction, *University of Ibadan Journal of Science and Logics in ICT Research* (UIJSLICTR). Vol. 13, No.1, pp. 80 – 94.

Dependent count, Dependent count, Marital Status, Income Category, Card Category, Months on book, and other relevant factors.

This paper enhanced the current literature by addressing gaps and offering valuable insights into customer churn behaviour within the banking sector. The study applied an ensemble technique and build an ensemble LSTM model. The study analyzed effects of some parameter on the dependent variable Customers Attrition The findings would be useful to banks managers in creating proactive strategies to prevent customers churn and enhance customer retention, ultimately informing their decisions regarding churn behaviour.

2. Related Works

Many researchers have employed ML methods to forecast customers churn in the banking industry for instance, Aishwarya *et al* [1] analyzed customer churn in banking. The paper employed data mining method known as Gaussian mixture model clustering-based adaptive support vector machine (GMM-ASVM) to predict customer loss in the banking sector. The study analyzed consumer competence and loyalty using GMM in the banking industry.

Shadakshari *et al* [15] carried out a study on comparative analysis of ML algorithms for predicting Bank customers churn, leveraging a dataset that included customer transaction and demographic data. The study employed four ML techniques such as random forest, logistic regression, neural network, and decision tree. The model's validation was assessed using evaluation metrics such as accuracy, precision, recall, and F1-score.
Nikita *et al* [11] developed a ML model to forecast churned customers for Commercial Bank of Ethiopia (CBE) the study used 204,161 instances with eleven attributes the study employed supervised ML techniques such as LR, RF, SVM, KNN, and DNN to forecast customer churn. The standard metrics were used to evaluate the model. Hoang *et al* [8] carried a study on how customer segmentation influences the accuracy of customer churn prediction in banking, utilizing ML models. The study used various ML models such as k-means clustering to segment customers, LR, DT, KNN, SVM, and RF to predict customer churn.

Ogunsanwo [12] developed a machine learning model for employee Attrition predictive model. The study adopted IBR HR dataset obtained from kaggle.com repository; it consists of 1470 instances and 34 features. PCA was used for feature reduction. The predictive model was created with RF, BILSTM, SVM and LSTM. The study concluded that SVM perform better than BILSTM RF and LSTM in terms of the validating metric used. ML algorithms have the ability to detect patterns and relationships in e data that might not be easily recognized by humans, allowing for predictions based on historical data Mitchel *et al* [10]

3. Materials and Methods

This study proposes Long Short-Term Memory (LSTM) and Ensemble techniques that have been reported to have advantages over single technique. However this study intends to develop an ensemble with LSTM. For the Bank Churn prediction this study seeks to develop a predictive model with three techniques such as: LSTM, ensemble LSTM and Random Forest in order to determine which of the techniques perform better. The proposed framework for this study is shown in Fig 1.This provides a clear approach of this study.

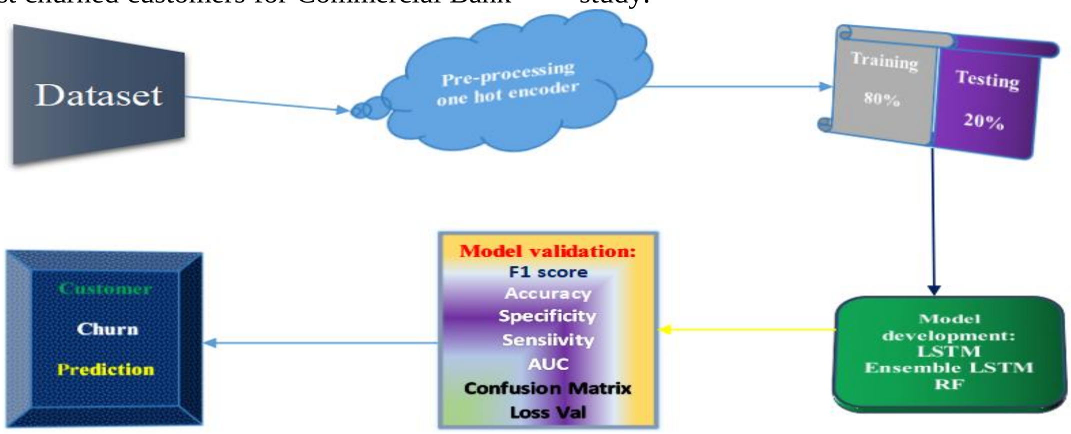
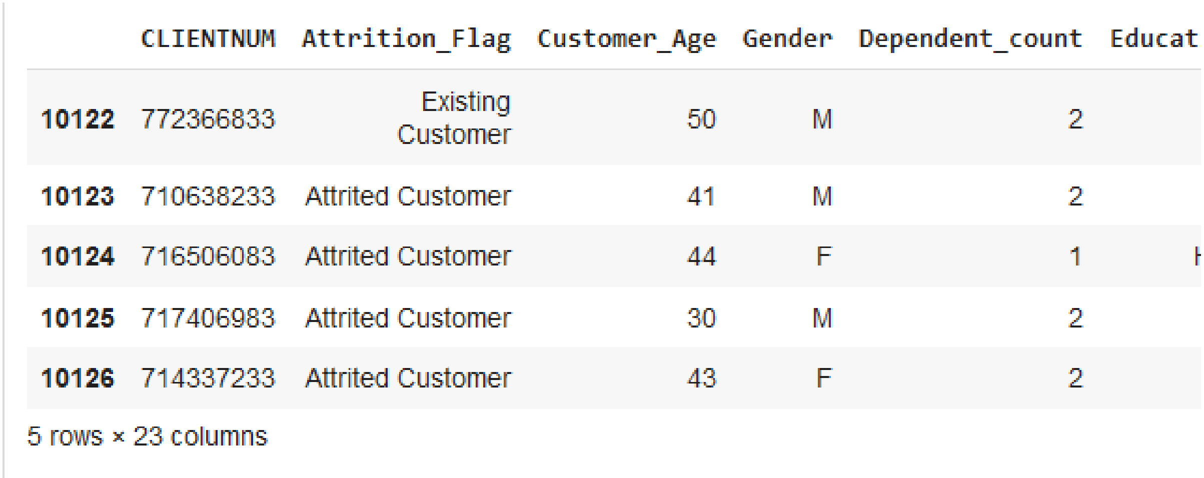


Figure 1 Flow diagram of the Model



	CLIENTNUM	Attrition_Flag	Customer_Age	Gender	Dependent_count	Educat
10122	772366833	Existing Customer	50	M	2	
10123	710638233	Attrited Customer	41	M	2	
10124	716506083	Attrited Customer	44	F	1	H
10125	717406983	Attrited Customer	30	M	2	
10126	714337233	Attrited Customer	43	F	2	

5 rows × 23 columns

Figure 2 sample of the dataset used

3.1 Data Acquisition

The dataset used for the Bank customer churn predictive model was downloaded from kaggle.com. The dataset consists of 10127 instances and twenty parameters. The Attrition-Flag was used as the dependent variable while the other variable consists Customer Age', 'Gender', 'Dependent count', 'Dependent count', 'Marital Status ', 'Income Category', Card_Category',' Months on book, Total Relationship Count, 'Months_Inactive_12_mon', 'Contacts_Count_12_mon','Credit_Limit etc as seen Figure 2.

3.1.1 Data Pre processing

Data preprocessing techniques was performed on the dataset in order to increase the accuracy of the predictive model. The first preprocessing done was numerical features transformation using Label Encoder in which the feature ranges are in 0.1. Thereafter categorical features were done using One Hot Encoder to handle unknown variable. Feature scaling: the data was scaled using MinMaxScaler, the perturbations are applied on the scaled feature values.

Dataset Used

The preprocessed dataset used for the study was divided into 80% training (8101.6) and 20% testing (2025.4) as seen in figure 3.

3.2 Classification Algorithms

3.2.1 LSTM

LSTM is an example of recurrent neural network (RNN) architecture build to tackle the vanishing gradient problem that affect the performance traditional RNN when processing long sequences. LSTM have memory cells used to store information over a long period. The cells are controlled by gates that control the flow of the information Hochreiter *et al* [9]. These gates are divided into three namely: Input gate which control the flow of new information into memory cell. Forget gate control the information that needs to be eliminated and Output gate which regulates the flow of information from the cell to the networks output Gers *et al* [6]

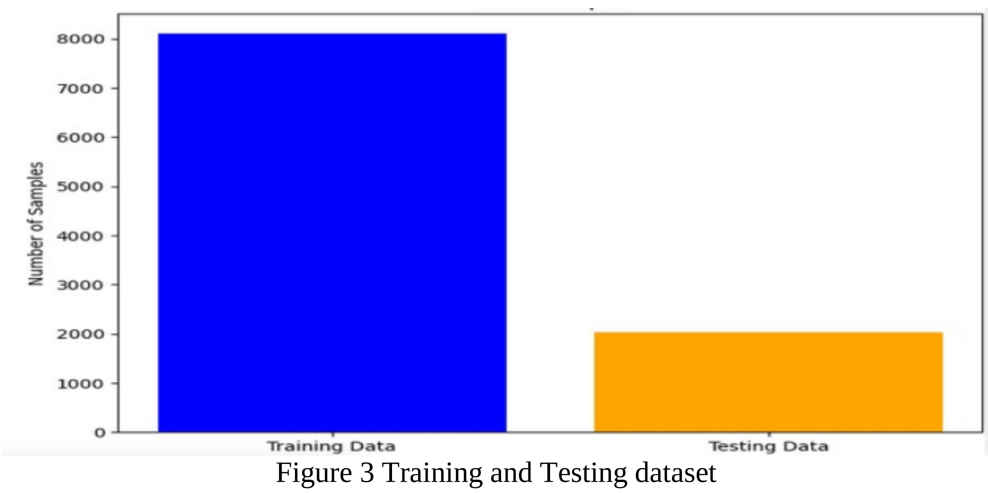


Figure 3 Training and Testing dataset

3.2.2 Ensemble LSTM

Ensemble learning is a method that merges several individual models to enhance prediction accuracy and robustness, demonstrating its effectiveness across different areas of ML. Applying this approach to LSTM networks, known as Ensemble LSTM, aims to leverage the strengths of multiple LSTM models to enhance overall performance in sequential data tasks. Ensemble LSTM reduces the variance by aggregating predictions from diverse models, leading to more stable and reliable results Brown et al [2]. The approach used for the ensemble LSTM is stacking in which multiple LSTM as seen in figure 4 and using a meta-learner to combine their prediction Wolpert [18]

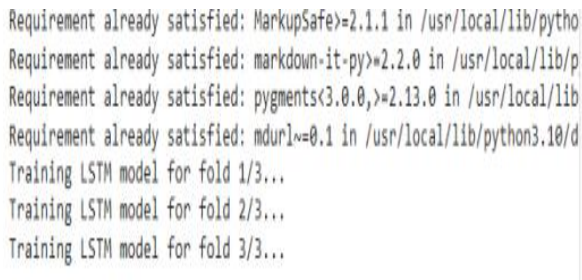


Figure 4 Ensemble LSTM model

3.2.3 Random Forest (RF)

RF is a strong and popular ensemble learning approach applied to both classification and regression tasks. It aggregates the predictions from numerous decision trees, leading to improved accuracy and robustness over those made by single trees. Random Forest consists of a collection of decision trees, each trained on various subsets of data and utilizing a random selection of features Breiman *et al* [4].

3.3 Validation Metrics

In this study, the following evaluation metric were used to validate the Bank Customers Churn predictive model developed namely: F1 score, Accuracy, Area under the curve (AUC), Confusion matrix, specificity, sensitivity.

3.3.1 Accuracy: It measures the percentage of instances that were classified correctly out of all the instances in the dataset. It's a modest and commonly used metric but normally have problem with imbalanced datasets.Sokolova *et al* [16]. It is calculated as seen in Equation (1)

Accuracy: (TP + TN) / (TP + TN + FP + FN) (1)
Where TP – true positive
TN – true negative
FP- false positive
FN- false Negative

3.3.2 Recall is also known as Sensitivity: This metric measures how many positive instances were predicted correctly out of all the actual positive instances. It shows how well we can find all positive cases. Powers [14]. It is calculated as seen in Equation (2)

Sensitivity: TP / (TP + FN) (2)
Where TP – true positive
FP- false positive
FN- false Negative

3.3.3 Specificity: This metric measures how many times negative instances were correctly predicted out of all the actual negative instances that actually exist. It shows how well negative cases can be identified Fawcett *et al* [5]. It is calculated as seen in Equation (3)

Specificity = TN / (TN + FP) (3)

3.3.4 F1-Score: The harmonic mean combines precision and recall to give a fair measure of both. It's useful when there are more instances of one class than the other Van [17]. It is calculated as seen in Equation (4)

F1-Score = 2 * (Precision * Recall) / (Precision + Recall) (4)

3.3.5 Receiver Operating Characteristic Area Under the Curve (ROC AUC): It differentiates positive and negative classes across different classification thresholds of the model's capability. It's good for evaluating models with probabilistic outputs **Bradley** [3].

3.3.6 A confusion matrix is a table commonly utilized to assess the performance of a classification model (or "classifier") on a test dataset with known true values. It serves as an effective means to evaluate the model's accuracy and gain insight into the kinds of error it commits. Usually, the confusion matrix is structured as a 2x2 table (for binary classification) with the following structure as seen in Table 1

Table 1: confusion matrix structure

Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

4. Model Developments and Result

4.1 The Performance Result of LSTM

The Loss indicates of how well the model is doing in terms of predicting the correct output. It quantifies the difference between the predicted values and the actual values. A lower loss value indicates that the model is making fewer errors and is better at predicting the target variable. In this study binary cross-entropy was used as loss function because it's a binary classification problem. The model is learning and improving when the loss is decreasing as seen in Figure 5. The **accuracy increasing** during training this indicates that LSTM model is learning to predict customer churn effectively. Table 2 shows the performance of LSTM model developed for banks customers churn predictive model.

Validation of Bank churn LSTM model Using Specificity: Specificity shows how the model's ability is correctly identified negative cases of customers who will not churn. The LSTM model got specificity of 0.6635 which reveals that the model correctly identifies 66.35% of customers who will not churn and so also that the model incorrectly identifies 33.65% of customers who will not churn as churners. The metric value increased as the threshold increased as seen in Figure 7.

Validation of Bank churn LSTM model Using Sensitivity: Sensitivity is also refers to as recall that quantifies the ability of a model to correctly identify the positive class that is customers who will churn. In this study the LSTM model have sensitivity of 0.9676 this means that the model correctly identifies 96.76% of the customers who are actually going to churn as seen in figure 8

Table 2 performance of LSTM Model

F1	Accuracy	Specificity	Sensitivity	AUC	Loss
0.94	**0.9235**	**0. 6635**	**0.97**	**0.95**	0.1663

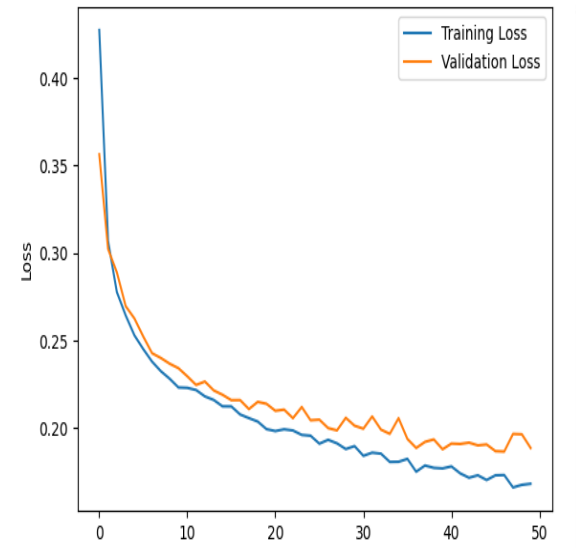


Figure 5 LSTM Loss Accuracy

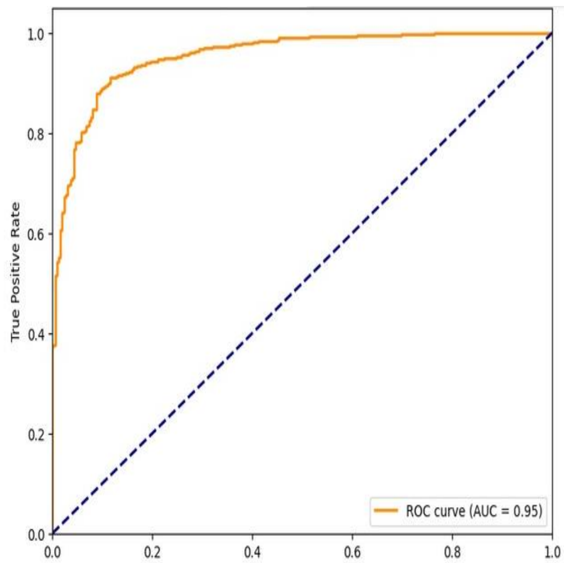
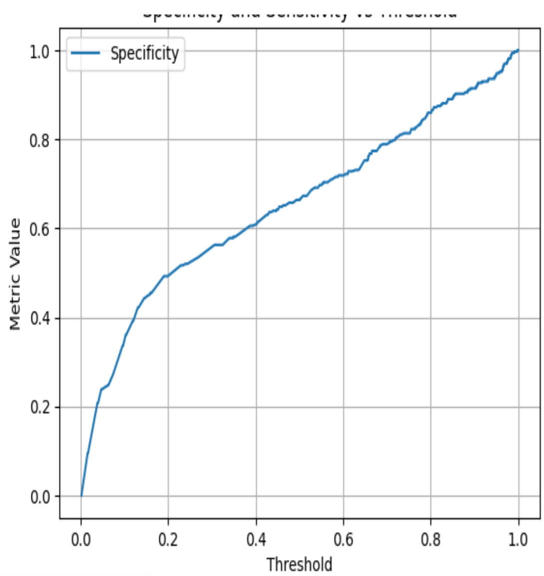


Figure 6 LSTM ROC curve



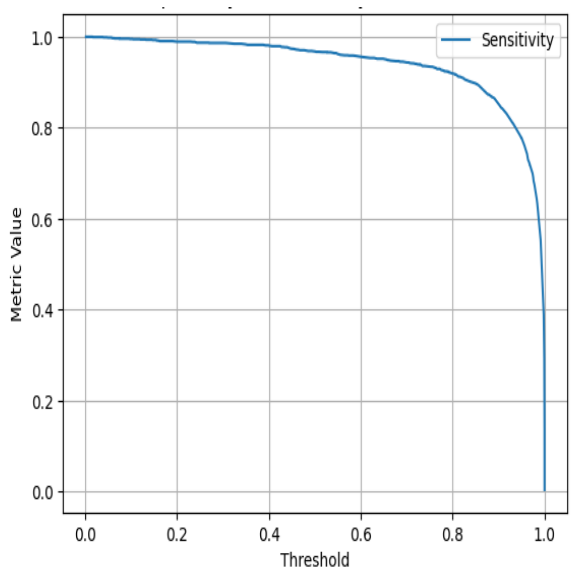


Figure 7 LSTM Specificity **Figure 8 LSTM Sensitivity**

Validation of LSTM model for bank customer churn prediction using ROC curve: The ROC curve for the LSTM model got the AUC 0.95 value. This would be considered a good performance, showing that the model has a good ability to distinguish between the positive and negative classes as seen in figure 6. A higher AUC value means the LSTM model is better at distinguishing between customers who will churn and those who will not.

Validation of Bank churn LSTM model Using Confusion Matrix

Bank churn LSTM model developed predicted that 217 customers correctly as not churning. Bank churn LSTM model predicted that **110** customers incorrectly as churning also predicted 55 customers incorrectly as not churning and model that predicted 1644 customers correctly predicted as churning as seen in Figure 9.

4.2 The Performance result of Ensemble of three LSTMs

The Loss shows how well the model is doing in terms of predicting the correct output. It quantifies the difference between the predicted values and the actual values. A lower loss value indicates that the model is making fewer errors and is better at predicting the target variable. In this study binary cross-entropy was used as loss function because it's a binary classification problem. The model is learning and improving when the loss is decreasing as seen in Figure 10. The **accuracy increasing** during training this indicates that ensemble LSTM model is learning to predict customer churn effectively. Table 3 shows the performance of Ensemble LSTM model developed for banks customers churn predictive model.

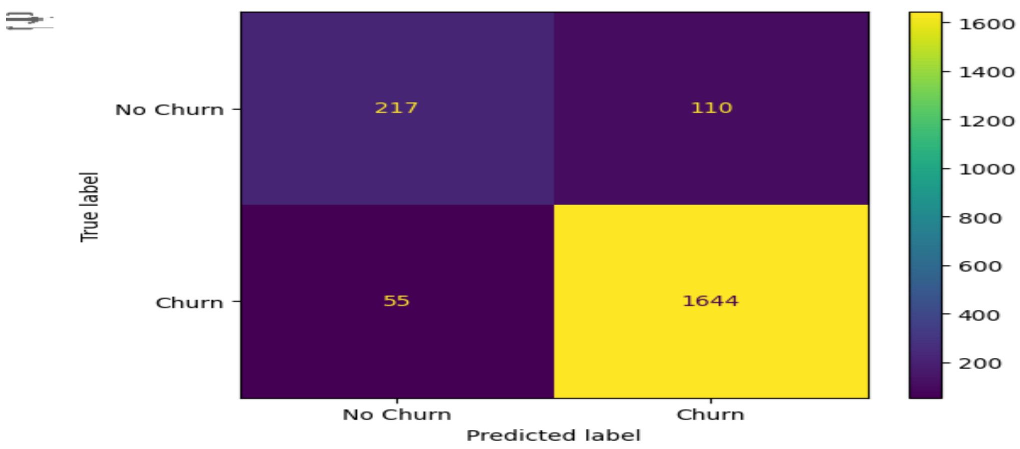
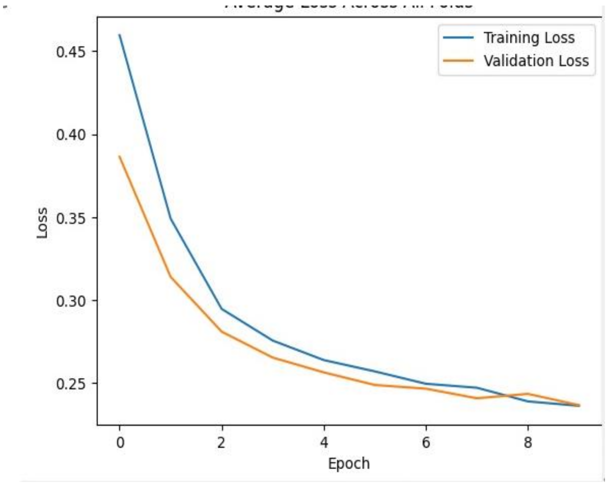


Figure 9 LSTM model Confusion Matrix

Table 3 Ensemble LSTM performance

F1	Accuracy	Specificity	Sensitivity	AUC	Loss
0.94	**0.9057**	**0. 554**	**0.98**	**0.92**	0.238



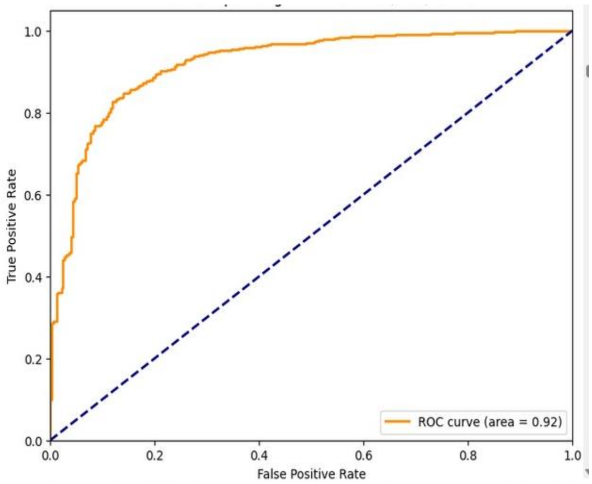


Figure 10 Ensemble LSTM Loss

Figure 11 ROC Ensemble LSTM

Validation of Ensemble LSTM model for bank customer churn prediction using Specificity:

Specificity quantifies the model's ability to correctly identify negative cases of customers who will not churn. The LSTM model got specificity of 0.55 which reveals that the model correctly identifies 55.4% of customers who will not churn and so also that the model incorrectly identifies 45.0 % of customers who will not churn as churners. The metric value increased as the threshold increased as seen in Figure 13a and 13b.

Validation of Ensemble LSTM model for bank customer churn prediction using Sensitivity:

Sensitivity is also referred to as recall that quantifies the ability of a model to correctly identify the positive class that is customers who will churn. In this study the LSTM model have sensitivity of 0.98 this means that the model correctly identifies 98.0 % of the customers who are actually going to churn as seen in figure 12

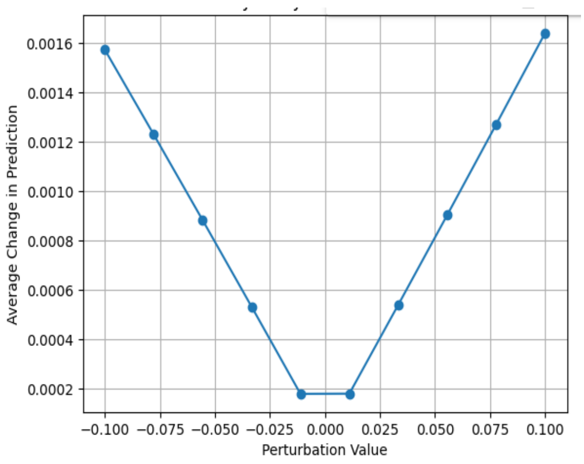
Validation of Ensemble LSTM model for bank customer churn prediction using ROC curve:

The ROC curve for the Ensemble LSTM model got the AUC 0.92 value. This would be considered a good performance, showing that the model has a good ability to distinguish between the positive and negative classes as seen in Figure 11. A higher AUC value means the Ensemble LSTM model is better at distinguishing between customers who will churn and those who will not.

Validation of Bank customers churn of Ensemble LSTM model Using Confusion Matrix:

Bank churn Ensemble LSTM model developed predicted that 180 customers correctly as not churning. Bank churn Ensemble LSTM model predicted that 147 customers incorrectly as churning also predicted 56 customers incorrectly as not churning and model that predicted 1643 customers correctly predicted as churning as seen in Figure 14

The Mean and Standard Deviation (SD)
The ensemble mean of 0.899 shows the churn prediction which is a binary classification problem predict an 89.9% chance of a customer churning. A lower standard deviation like the value generated by Ensemble LSTM model of 0.021 indicates that the predictions of the individual models are relatively close to each other as seen in figure 15



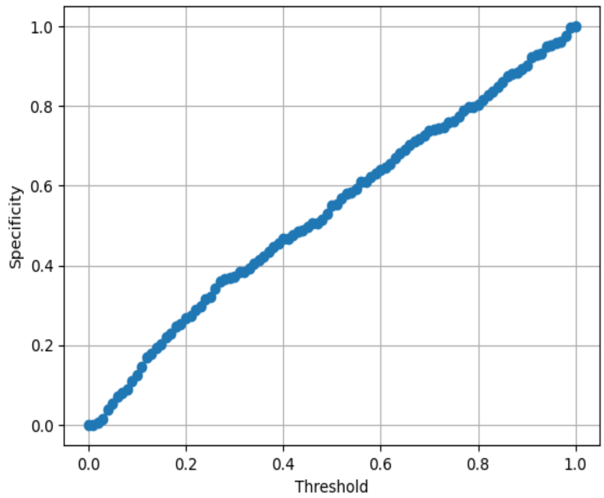
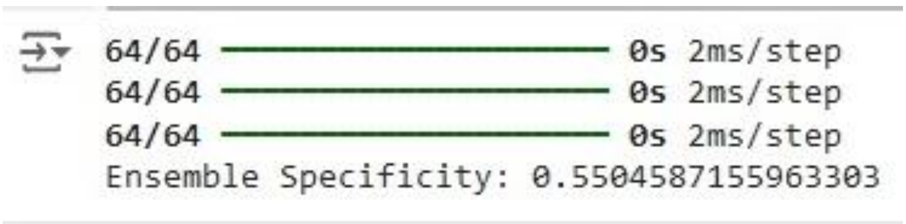
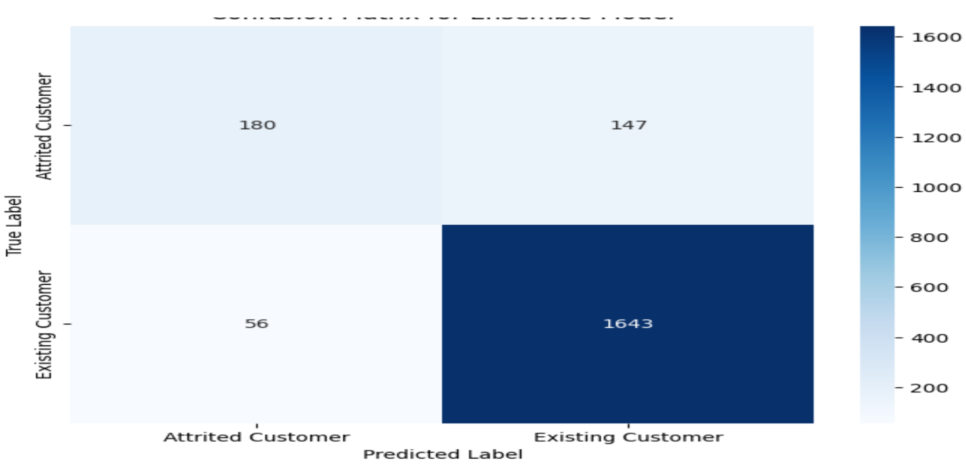


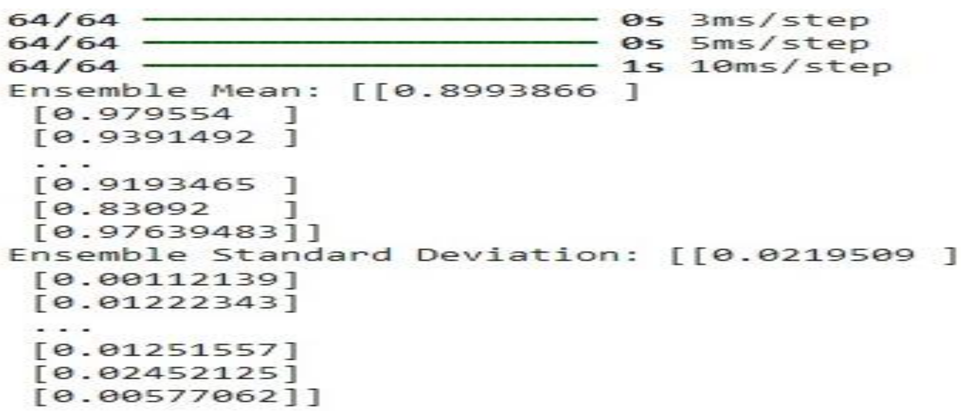
Figure 12 Ensemble LSTM Sensitivity **Figure 13a Ensemble LSTM Specificity**



```
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 2ms/step
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 2ms/step
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 2ms/step
Ensemble Specificity: 0.5504587155963303
```

Figure 13b





```
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step
64/64 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step
64/64 ━━━━━━━━━━━━━━━━━━━━ 1s 10ms/step
Ensemble Mean: [[0.8993866 ]
 [0.979554  ]
 [0.9391492 ]
 ...
 [0.9193465 ]
 [0.83092   ]
 [0.97639483]]
Ensemble Standard Deviation: [[0.0219509 ]
 [0.00112139]
 [0.01222343]
 ...
 [0.01251557]
 [0.02452125]
 [0.00577062]]
```

Figure 14 Ensemble LSTM Confusion Matrix Figure 15 Mean and S D of Ensemble LSTM

Performance Result of RF Model

The RF Model used for bank customers churn prediction as depicted in figure 16. The Lines connecting nodes represent branches, indicating the different paths taken based on the decision at each node. Each decision node shows the feature used for the decision (e.g., "Customer_Age <= 35.5") and the threshold value used to split the data. Table 4 shows the performance result of RF model with F1 score of 0.97and Accuracy, Specificity, Sensitivity, AUC and Loss of 0.95, 0. 774, 0.99, 0.99 and 0.15 respectively

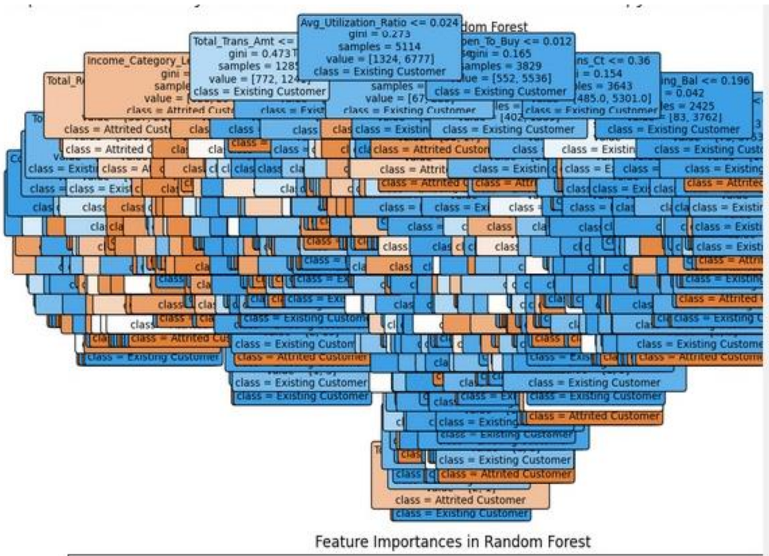


Figure 16 RF model

Table 4 the Performance of RF Model

F1	Accuracy	Specificity	Sensitivity	AUC	Loss
0.97	**0.95**	**0. 774**	**0.99**	**0.99**	0.15

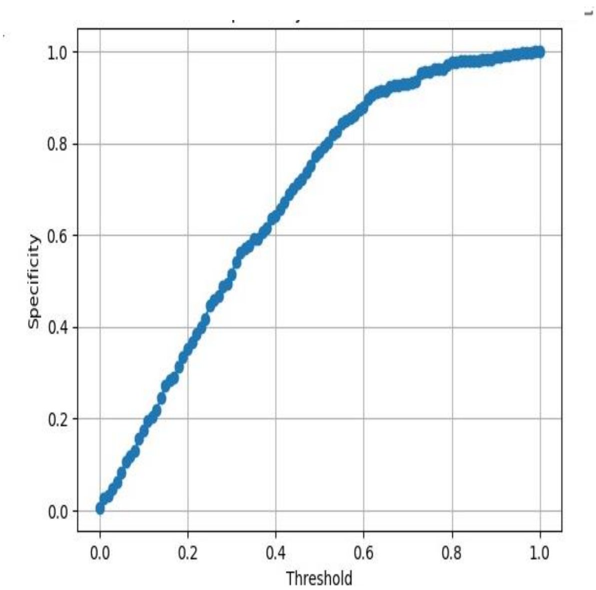


Figure 17 RF model Specificity

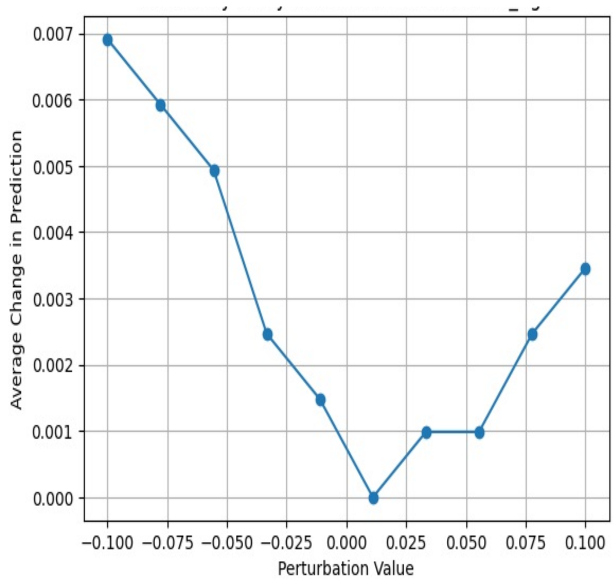


Figure 18 RF model Sensitivity

Validation of RF model for bank customer churn prediction using Specificity

Specificity quantifies the model's ability to correctly identify negative cases of customers who will not churn. The RF model got specificity of 0.77 which reveals that t**he** model correctly identifies 77.9 % of customers who will not churn and so also that the model incorrectly identifies 22.1 % of customers who will not churn as churners. The metric value increased as the threshold increased as seen in Figure 17

Validation of RF model for bank customer churn prediction using Sensitivity

Sensitivity is also referred to as recall that quantifies the ability of a model to correctly identify the positive class that is customers who will churn. In this study the RF model have sensitivity of 0.99, this means that the model correctly identifies 99.0 % of the customers who are actually going to churn as seen in figure 18

Validation of RF model for bank customer churn prediction using ROC curve:

The ROC curve for the RF model got the AUC 0.99 value. This would be considered a good performance, showing that the model has a good ability to distinguish between the positive and negative classes as seen in Figure 19. A higher AUC value means the RF model is better at distinguishing between customers who will churn and those who will not.

Validation of Bank customers churn RF model Using Confusion Matrix

Bank churn RF model developed predicted that 255 customers correctly as **not** churning. Bank churn RF model predicted that 72 customers incorrectly as churning also predicted 21 customers incorrectly as not churning and model that predicted 1678 customers correctly predicted as churning as seen in Figure 20.

4.3 Result of Exploratory Data Analysis

Attrition rate by Customers Gender

The Bank customers churn prediction generated the chi-squared statistic of 13.87: This value suggests a noticeable difference between the observed and expected frequencies of Gender and churn in the dataset. P-value of 0.000 is very small this provides a very strong evidence to reject the null hypothesis. Based on the Chi-squared test results, it can be concluded that there is a statistically significant association between gender and Attrition_Flag. This means that a customer's gender status is likely a factor that influences their likelihood of churning as seen in figure 21a and 21b.

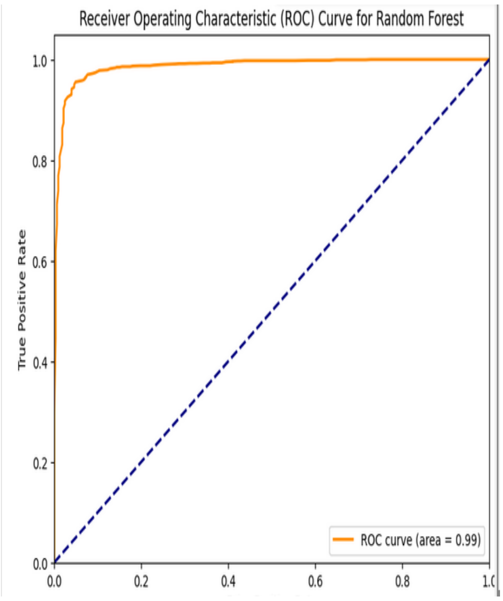


Figure 19 RF Model ROC

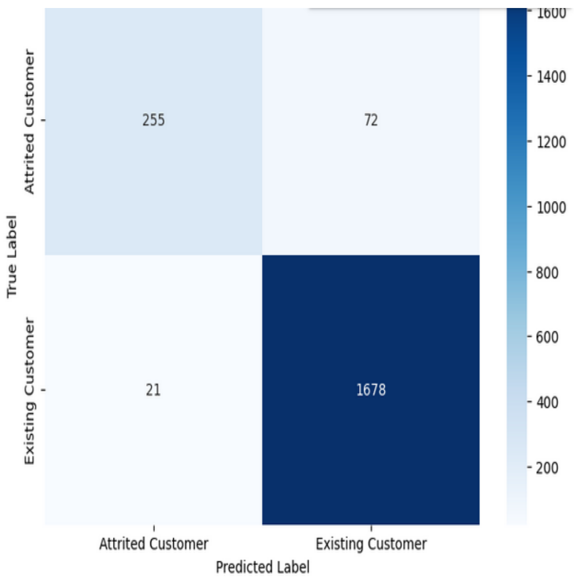


Fig 20 RF Model confusion matrix

Attrition rate by Income_Category

The Bank customers churn prediction generated the chi-squared statistic of 12.83: This value suggests a noticeable difference between the observed and expected frequencies of Income category of the customers and churn in the dataset. P-value of 0.025 is very small this provides a very strong evidence to reject the null hypothesis. Based on the Chi-squared test results, it can be concluded that there is a statistically significant association between Income category and Attrition_Flag.

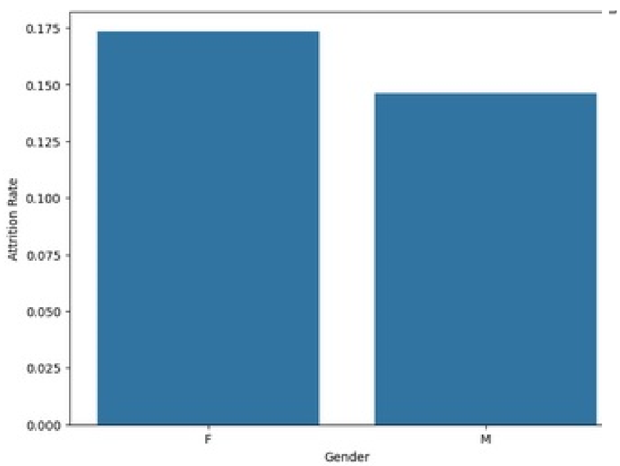
This means that a customer's Income category status is likely a factor that influences their likelihood of churning. The model results generated Income_Category of $120K with churn values of 0.173315, $40K - $60K with churn values of 0.151397, $60K - $80K with churn values of 0.134807, $80K - $120K with churn values of 0.157655 and Less than $40K with churn values of 0.171862. as seen in figure 22

Attrition rate by Marital Status

The Bank customers churn prediction generated the chi-squared statistic of 6.06. This value suggests a noticeable difference between the observed and expected frequencies of marital status of the customers and churn in the dataset. P-value of 0.109 is very small this provides a very strong evidence to reject the null hypothesis. Based on the Chi-squared test results, it can be concluded that there is a statistically significant association between marital status and Attrition_Flag. This means that marital status category status is likely a factor that influences their likelihood of churning. The model results generated Marital_Status for divorced with churn values of 0.161765 values, married with churn values of 0.151269 and single with churn values of 0.169414.as seen in figure 23

Attrition rate by Customers Age

he model result of Age and Attrition_Flag in the dataset. The result pointed out that a customer's age is a factor that influences their likelihood of churning.as seen in Figure 24




```
Chi-squared test results:
Chi-squared statistic: 13.87
P-value: 0.000
```

Figure 21 a Attrition rate by gender

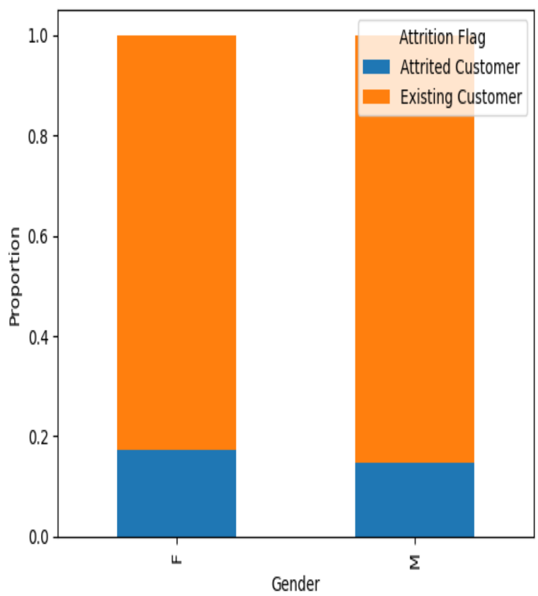
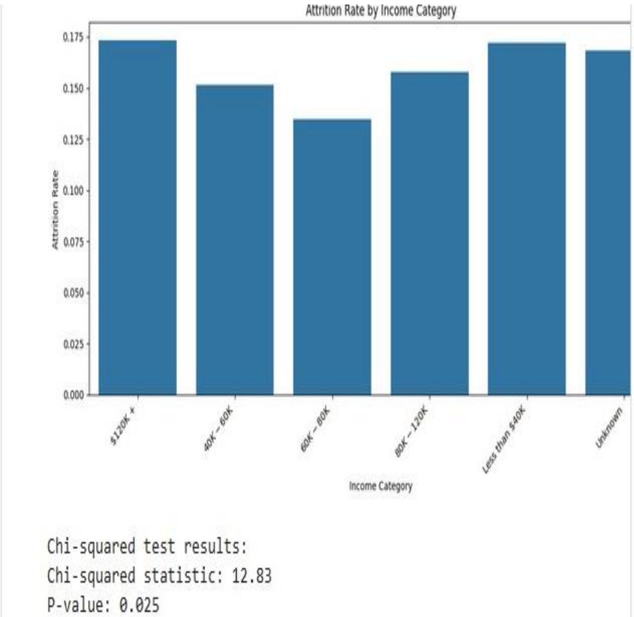


Figure 21 b Attrition rate by gender



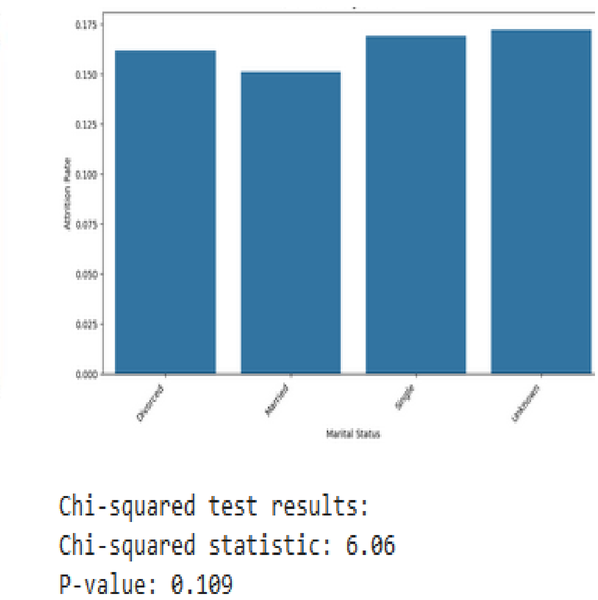


Figure 22 **Attrition rate by Income_Category** Figure 23 **Attrition rate by Marital Status**

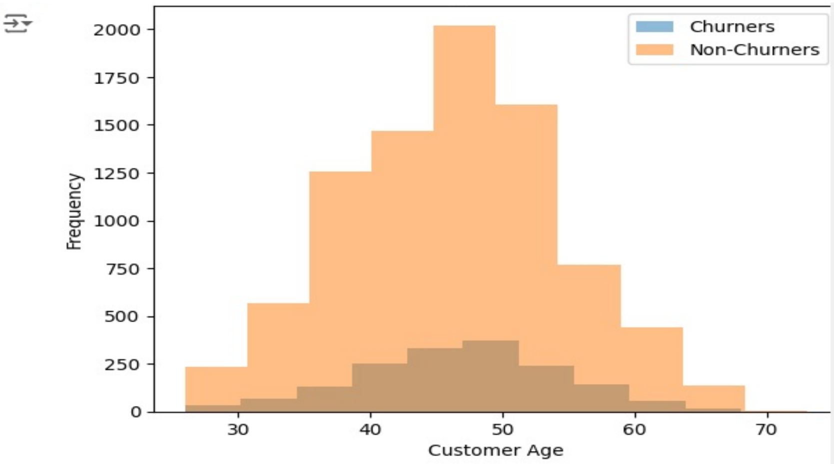


Figure 24 **Attrition rate by Customers Age**

4.2 Discussion of the findings

The comparison of the model used in the study revealed in Table 5 and figures 25, 26, 27 and 28. Based on the result it is revealed RF has the highest F1 score (0.97), followed by LSTM and Ensemble LSTM (both 0.94) since higher F1 score indicates better performance. In term of Accuracy: RF achieves the highest accuracy (0.95), followed by LSTM (0.9235) and Ensemble LSTM (0.9057). For Specificity RF demonstrates the highest specificity (0.774), followed by LSTM (0.6635) and Ensemble LSTM (0.554). Sensitivity (Recall): RF has the highest sensitivity (0.99), slightly Ensemble LSTM (0.98) and LSTM (0.97). AUC (Area Under the ROC Curve): RF exhibits the highest AUC (0.99), followed by LSTM (0.95) and Ensemble LSTM (0.92). Loss: Quantifies the model's error during training. RF has the lowest loss (0.15), followed by LSTM (0.1663) and Ensemble LSTM (0.238). The result of the study revealed that RF appears to be the best-performing model with the highest F1 score, accuracy, specificity, and AUC, and the lowest loss. The model reveals that female is like to churn with 0.173572 values more than male with 0.146153. The result of the model also revealed that customers with less than $40K are more likely to churn compare to other categories. The result of the model revealed that customers that are singles are more likely to churn followed by divorced compare to customers that are married. It also revealed that customers within the age of forty to fifty-five are likely to churn where the age between 50 are more likely to churn. The study is compare with the finding of Oladipo *et al* [13] with RF accuracy of 92%, Shadakshari *et al* [15] with 85 % accuracy for the RF model. The study outperformed the existing models in the literature in term of accuracy of 95% for RF model developed.

Table 5 Comparison of the three Models

Model	F1	Accuracy	Specificity	Sensitivity	AUC	Loss
LSTM	0.94	0.9235	0. 6635	0.97	0.95	0.1663
Ensemble LSTM	0.94	0.9057	0. 554	0.98	0.92	0.238
RF	0.97	0.95	0. 774	0.99	0.99	0.15

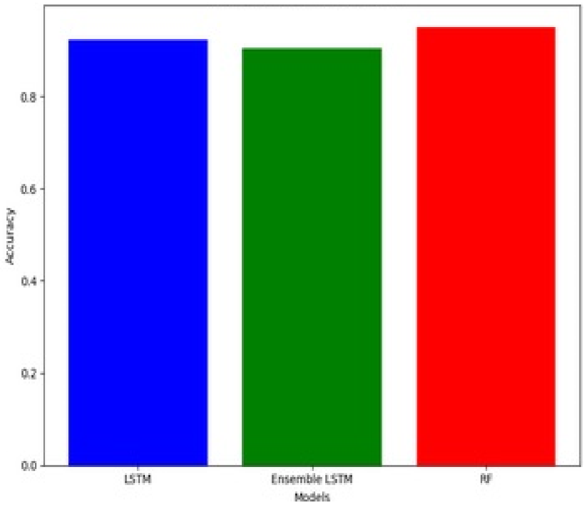


Figure 25 Models Accuracy comparison

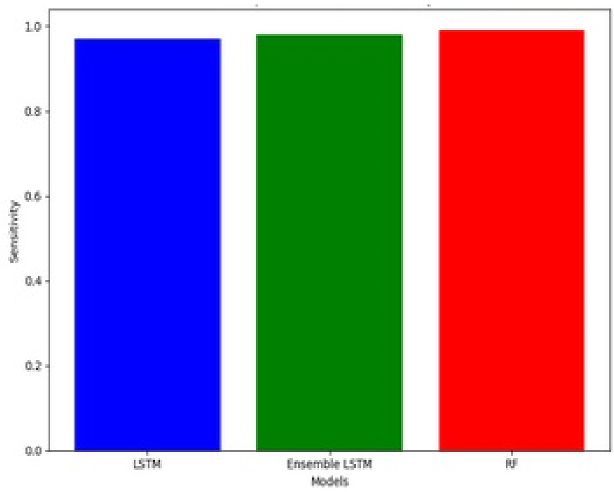


Figure 26 Models sensitivity comparison

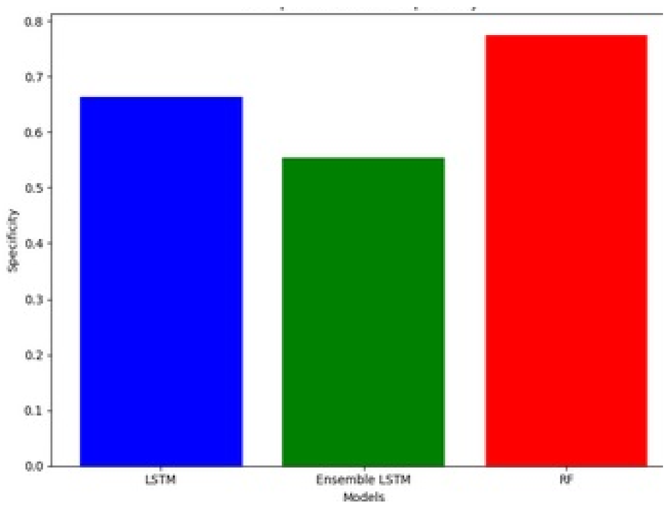


Figure 27 Models specificity comparison

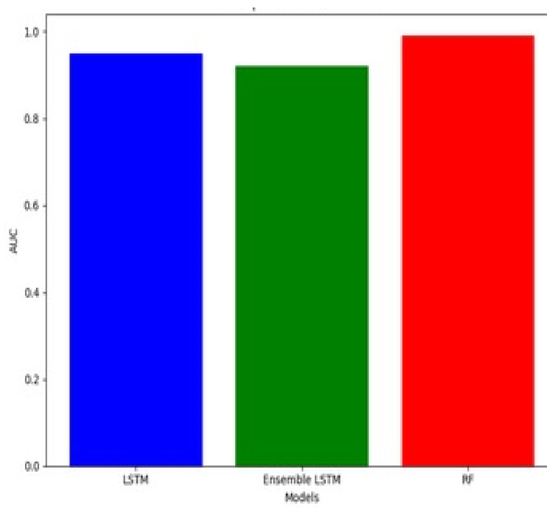


Figure 28 Models AUC comparison

5. Conclusions

Technology is aiding increase in the number of service providers in every sector in which banking industry is not left out of this development. We are leaving in the era as customers in banking sector have numerous options where to put their money. This makes customers churn prediction a hot topics for many researchers and banks managers. In this study, a bank customer's churn predictive model was developed using LSTM, Ensemble LSTM and RF.

The dataset used for this study was downloaded from kaggle.com repository in .csv format. Thereafter data preprocessing was done on the dataset In order to increase the performance accuracy of the model. The research conducted data mining on bank customer's dataset in order to explore the likelihood of churn by analyzing customers behaviour using ensemble techniques.

The experiment results showed that RF outperformed with f1 score of 97%, accuracy of 95%, specificity of 77% sensitivity of 99%, AUC 99% and loss of value 0.15 compared with LSTM and Ensemble LSTM. The exploratory data analysis carried also revealed that customers' gender, marital status, customer income category and age against attrition are determine factor for customer churn prediction. The model is recommended for banking sector to aid decision making to know likely customer that prone to churning in order to come up with retention strategies in order to bring and development to their banks. Future work can be done using more ensembles techniques and perform more data expository.

Declerations

Ethics approval : Not applicable
Consent for publication: Not applicable
Competing interests: No competing interests on the manuscripts
Funding: The study was self-funded

References

[1] Aishwarya Saxenaa , Anushi Singhb & Govindaraj M. (2023) Analyzing customer churn in banking Multidisciplinary Science Journal·DOI: 10.31893/multiscience.2023ss0310

[2] Brown, G., Wyatt, J. L., Harris, R., & Yao, X. (2005). Diversity creation methods: a survey and categorization. Information Fusion, 6(1), 5-20

[3] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159.

[4] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[5] Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.19

[6] Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. Neural computation, 12(10), 2451-2471.

[7] Hejazinia, R., & Kazemi, M. (2014). Prioritizing factors influencing customer churn. https://www.semanticscholar.org/paper/Prioritizing-factors-influencing-customer-churn-Hejazinia-Kazemi/991816bd37b7d23272f70683701ac7d726ef749e#citing-

[8] Hoang Tran, Ngoc Le & Van-Ho Nguyen (2023) Customer Churn Prediction In The Banking Sector Using Machine Learning-Based Classification Models Interdisciplinary Journal of Information, Knowledge, and Management, 18, 18,87-105. https://doi.org/10.28945/5086

[9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[10] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Education.

[11] Nikita Khandelal, Vikas Sakalle (*2023)* Customer Churn Prediction in Telecommunication, Medical Industry Using Machine Learning Classification Models", *6th International Conference on Contemporary Computing and Informatics (IC3I)*, vol.6, pp.1727-1734,.

[12] Ogunsanwo G.O (2024) Machine Learning Model For Employee Attrition Prediction FUW Trends in Science & Technology Journal,e-ISSN: 24085162; p-ISSN: 20485170; Vol. 9 No. 3 pp. 001 – 006

[13] Oladipo, I. D., Awotunde, J. B., AbdulRaheem, M., Taofeek-Ibrahim, F. A., Obaje, O. and Ndunagu, J. N. (2023). Customer Churn Prediction in Telecommunications Using Ensemble Technique, University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR), Vol. 9 No. 1, pp. 82 – 95.

[14] Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation.

[15] Shadakshari , Srijan Shashwat , Prashant Kumar Himanshu, Aniket Singh1 & Shivam Chaudhary (2024) A Comparative Study of Machine Learning Algorithms for Bank Customer Churn Prediction East African Scholars Journal of Engineering and Computer Sciences ISSN: 2617-4480 (Print) & ISSN: 2663-0346

[16] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information processing & management, 45(4), 427-437.

[17] Van Rijsbergen, C. J. (1979). Information Retrieval. Butterworth-Heinemann.

[18] Wolpert, D. H. (1992). Stacked generalization. Neural networks, 5(2), 241-259.