



A Comparative Analysis of Stacked Ensemble and Neural Network Models in Enhancing Customer Retention in Banking

¹Ismail Ahmad Ibrahim, ²Ameenu Abdulhameed Rabiu, ³Buhari Bashir and ⁴Salihu Abdullahi Audu

¹No. 13 Dutsinma Street Tudun Wada

²No. 70 Kundilar Gandu, Zoo Road Housing Estate

³No. 8 Funtua Crescent K/Kaura New Layout, Katsina.

⁴Department of Computer Science, Nasarawa State University

Corresponding Author's Email: abdulgiz@nsuk.edu.ng

Received on: April, 2025

Revised and Accepted on: May, 2025

Published on: July, 2025

ABSTRACT

Customer churn is a major problem for the banking sector, as it affects the profitability and sustainability of the business. Therefore, it is important to identify the customers who are likely to leave to take appropriate actions towards retaining them. The study aims to apply machine learning models to predict churn and identify the most effective model based on performance metrics such as log loss, ROC AUC, and accuracy using a dataset of 10,000 customers with 14 features. The study also utilizes domain feature engineering to get better predictive signals. We performed data analysis, and tested models including a neural network and stacked ensemble models comprising Random Forest, XGBoost, LightGBM, and Gradient Boosting with logistic regression as the meta-learner. The study performed a comparative analysis of the performance of a neural network with stacked ensemble model of all four tree models. The stacked ensemble model outperformed others, achieving a log loss of 0.2497, ROC AUC of 0.9609, and accuracy of 89.11% on the test set. This indicates that the ensemble model, by combining the strengths of individual learners, effectively handles complex patterns in the data, thereby providing a tool for reducing churn and improving customer engagement strategies in the banking sector.

Keywords: Machine learning, stacked ensemble, neural network, Customer Retention, Random Forest

1.0 INTRODUCTION

1.1 Introduce the Problem

Traditional methods for addressing churn, like surveys, have their limitations. However, with the expansion in various sectors, there has been a rise in the use of data mining and machine learning (ML) techniques, due to their capability to manage and analyse large volumes of data (Osman and Ghaffari, 2021). Customer churn (CC) refers to the process of identifying customers who are likely to cease their relationship with a business or service provider. It involves analysing historical data, customer behaviour to predict which customers are at risk of leaving, by leveraging ML models that identify potential churners before they leave. Customer Churn Prediction (CCP) enable companies to predict which customers are likely to leave, allowing them to take proactive steps to prevent it. These steps can vary depending on why a customer might leave and may include actions like offering discounts or additional services to retain them.

1.2 Purpose

Existing research, such as the work by Rahman and Kumar (2020) on machine learning based customer churn prediction, has laid a foundation in the use of individual classifiers such as K-Nearest Neighbors (KNN), support vector machine (SVM), Decision Tree,

and Random Forest (RF) within the banking sector, this paper seeks to further practices. Our research builds upon this foundation with the introduction of ML techniques, such as ensemble of tree-based models, executed through a stacking method that integrates the predictions of four individual tree classifiers an idea based on research by Oladimeji et al. (2023) and a neural network (NN) will then be applied as an individual model to be compared with the tree based models overall performance. The core of the stacking method relies on the outputs of individual tree-based models such as, XGBoost (XGB), LightGBM (LGBM), and Gradient Boosting classifiers (GBC) to form a composite prediction. This combined prediction is then used as input for a logistic regression classifier, serving as a meta-learner, to make the final churn prediction (Awasthi, 2023).

The focus of this paper is to compare the effectiveness of the above-mentioned stacked ensemble machine learning (SE-ML) models against NN in terms of predictive accuracy and reliability to identify churners. Additionally, domain-specific feature engineering will be employed to improve the accuracy of churn predictions. Meanwhile to evaluate the performance of these models, we will implement log-loss, AUC-ROC, Accuracy, precision, recall and F1 score as the major



metrics. Log-loss will allow us to assess the precision of the probability predictions, ensuring that the predicted likelihood of churn reflects actual outcomes. Also, the AUC-ROC metric will assist us in understanding the models' ability to distinguish between customers who will churn and those who will remain. The incorporation of these evaluation metrics will allow analysis of the models' predictive power. The project will elaborate on the dataset preprocessing, Exploratory data analysis (EDA), Feature importance and the splitting of the data into training and testing, the modelling techniques, evaluation process, comparative analysis and expected results.

1.3 Research Question

To achieve the aims of this study, we have developed the following research questions:

- i. Among stacked ensemble models (SEM) and NN, which demonstrates superior performance in accurately predicting customer churn within the banking sector?
- ii. What are the challenges and limitations associated with using NN for CP within the banking sector, considering factors like imbalanced datasets, non-linear relationships between features, and scalability issues with large-scale banking data?

To address the research questions, we will compare the performance of SEM and NN in accurately predicting CC within the banking sector, analysing their effectiveness based on relevant metrics. Additionally, we will explore the challenges and limitations of using NN for CP in the banking sector, particularly focusing on handling imbalanced datasets, modelling non-linear relationships between features, and addressing scalability issues with large-scale banking data.

1.4 The Hypothesis

The SEM, which combines predictions from individual tree-based classifiers, will outperform the NN model in predicting CC.

2.0 MATERIALS AND METHOD(S)

2.1 Data Collection

This study utilizes the "churn modelling.csv" dataset downloaded from Kaggle, containing 10,000 records, with 14 attributes covering demographics, account behaviour. To safeguard customer privacy, some features and transactions in the data have been

anonymized. This dataset is well-recognized in the academic field and has been referenced in research published by Scopus-indexed journals. It's use in high-quality research is widespread, with articles appearing in well-known journals from IEEE, Springer, and Science Direct, in studies by Haddadi et al. (2024), Vu (2024), and (Huseyinov and Okocha, 2022). No additional data was gathered for this study, ensuring that the analysis is reproducible, relying on publicly available data of high quality.

2.2 Data Set Description

The features encompass a range of demographic, account, and transactional information designed to illustrate customer engagement and potential churn figure 3.1 below shows the first ten rows and the attributes of the data. An outline of the dataset is as follows: RowNumber: This is simply a row index. CustomerId: A unique identifier for each customer. Surname: The customer's surname, which has been anonymized. CreditScore: A value between 0 to 1000 indicating the customer's creditworthiness. Geography: The country where the customer's account is based. Gender: The customer's gender identified as Male or Female. Age: The age of the customer. Tenure: The length of time, in years, that the customer has been with the bank. Balance: The average balance maintained in the customer's account. NumOfProducts: The total number of banking products the customer is using. HasCrCard: A binary indicator of whether the customer has a credit card with the bank. IsActiveMember: A binary indicator of whether the customer is considered an active member. EstimatedSalary: The estimated salary of the customer.

2.3 Neural Network Model Implementation

The study aims to compare the predictive performance of NN and the SEC in this section we implemented a NN in pursuit of our research objectives. The NN model utilized in this study is a feedforward architecture, this network features a structure conducive to modelling complex, non-linear data relationships to align with the binary classification goals of the study, The initial network as depicted in Figure 1 composed of two dense hidden layers, each with six nodes, and utilizes the Rectified Linear Unit (ReLU) activation function. This function was selected due to its effectiveness in mitigating the vanishing gradient problem, as supported by early research.

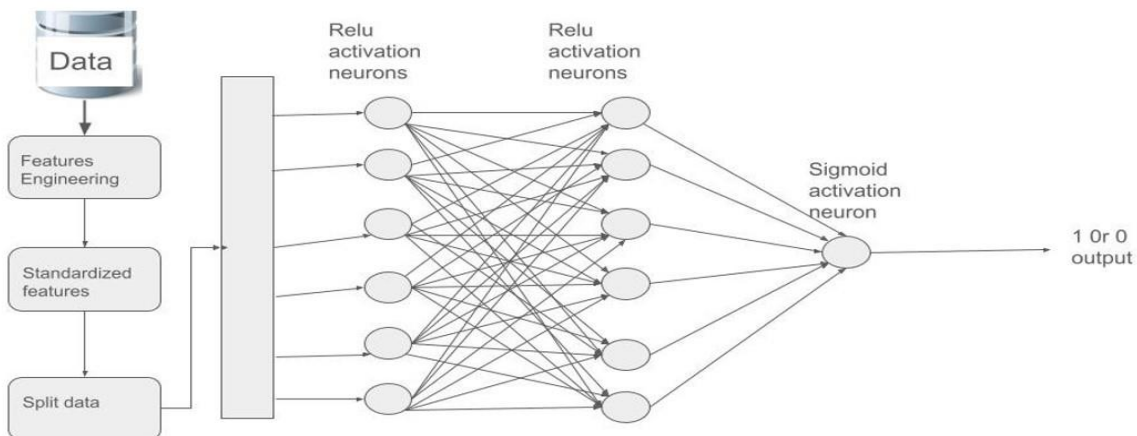


Figure 1: This illustrates the structure of the initial neural network used in the study, depicting the workflow from data collection, through feature engineering and standardization of features, to the final implementation with layers of ReLU activation neurons and a sigmoid output neuron for binary classification.

The above network further reconfigured and fine-tuned with a focus on achieving optimal generalization. We implemented Grid search to obtain the best combination of hyperparameters that delivers the best performance on our dataset Table 1 shows the hyperparameters tuned.

Table 1: Model Hyperparameters and Their Optimal Values

Model Hyperparameter	Tuning Range	Best Hyperparameter
Hidden Layer Sizes	(6), ((50, 100))	50
Activation	('relu', 'sigmoid')	Relu
Optimizer	('adam')	'adam'
Loss	('binary crossentropy')	'Binary crossentropy'
Epochs	(20, 50, 100)	20
Batch Size	(10, 32, 64)	10
Validation Split	(0.0001, 0.001)	0.001

These selected hyper-parameters are the results obtained through a grid search, with 10fold cross-validation to ensure model evaluation and minimize bias an idea based on the research of (Bent'ejac et al., 2021). The optimal neural network configuration featured a logistic activation function, an alpha value of 0.001, and 50 hidden layers aiming to enhance model accuracy while preventing overfitting, binary cross-entropy as the loss function due to its relevance for binary classification outcomes and conducted training over 20 epochs with a batch size of 10 as shown in the table above.

This configuration balances resource use with the model's ability to learn from the training data effectively. Throughout the training process, we monitored accuracy and loss metrics to assess the model's learning progression. The model's accuracy on both training and validation datasets offers insights into its generalization capabilities, while the visualizations of cross validation which will be discussed in subsequent sections helps in detecting overfitting Hinton et al. (2012).

The implementation of NN is tailored to address the challenges associated with CCP in the banking sector such as, managing imbalanced datasets, non-linear relationships, and ensuring model scalability. The design choices and training strategy was made to contribute to answering our research questions, focusing on the performance of NN against SEC in the domain of CP.

2.4 Stacked Ensemble Models Implementation

The SEC is utilized to enhance CP in the banking sector. This model combines the tree-based algorithms, such as XGBoost, RF, LGBM, and GBC as discussed in the previous sections, each for stability and efficiency in binary classification tasks. Figure 3.11 below provide visualized structure of the approach. The implementation of the stacked ensemble model involved fine-tuning individual classifiers and integrating their predictive capabilities through a meta-learner an idea based on the research of (Awang et al., 2021). Each base estimator XGBoost, RF, LGBM, and GBC was fine-tuned with

parameters.

Table 3.2 summarizing the parameters used for each model during the implementation step of the tree

ensemble models. The selection of parameters is similar to those utilized by Kalule et al. (2023) for optimizing the performance of each classifier.

Table 2: Model Hyperparameters Overview

Model	Hyperparameters
Random Forest	n estimators: 200, criterion: 'entropy', random state: 0
Gradient Boosting	n estimators: 100, criterion: 'squared error', random state: 42
LightGBM (LGBM)	boosting type: 'gbdt', num leaves: 31, max depth: 3, learning rate: 0.1, n estimators: 100, subsample for bin: 200000
XGBoost	objective: 'binary:logistic', max depth: 3, learning rate: 0.1, n estimators: 100, random state: 0

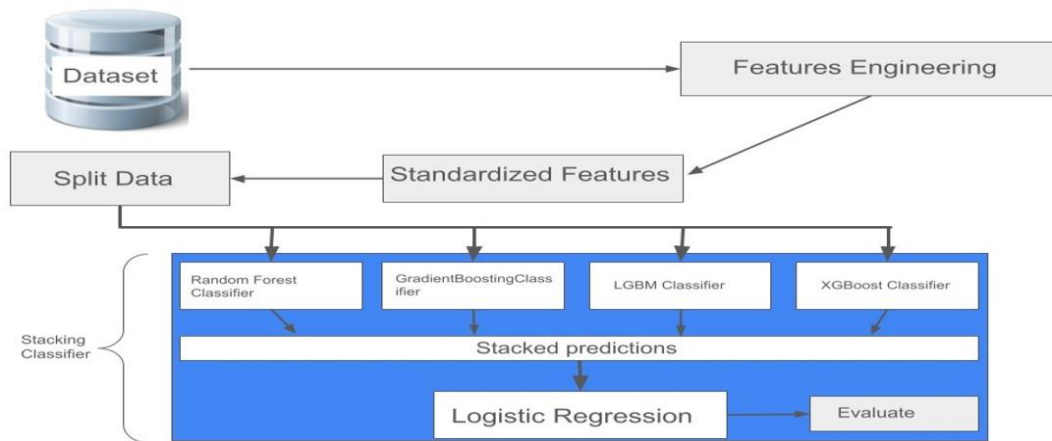


Figure 2: This diagram illustrates the structured workflow of the stacking ensemble model employed in the study

3.0 RESULTS AND DISCUSSION

In this section, we present the results of the models used for CCP in the banking sector. To determine which models, deliver the most accurate and generalized results, we finetuned each model and measured their effectiveness across key performance metrics. Our aim is to optimize and compare the predictive power of the models, reflected in metrics such as accuracy, precision, recall, F1 score, ROC AUC. The Table 3 below encapsulates the results by presenting them in a tabulated format, we provide a comparative analysis, underlining the strengths and areas for improvement for this approach.

Table 3: Model Performance Metrics

Model	Log Loss	ROC AUC	Accuracy	Precision	Recall	F1 Score
Neural Network	0.2665	0.9546	0.8845	0.89	0.88	0.88
Random Forest	0.2719	0.9591	0.8939	0.89	0.91	0.90
Gradient Boosting	0.2708	0.9571	0.8889	0.89	0.89	0.89
LGBM	0.2680	0.9574	0.8908	0.91	0.92	0.90
XGBoost	0.2731	0.9559	0.8870	0.88	0.91	0.89
Stacking Ensemble	0.2497	0.9609	0.8911	0.89	0.90	0.89

This tabular representation is followed by visualizations that depict the stability of the SEM and NN performance over validation results. Variations in log loss and ROC AUC are important, as they provide insights into the consistency and

generalizability of each model. The evaluation of different machine learning models, as shown through metrics over multiple validation folds, has provided us with insights into each model's performance and stability. Figures offer visual

representations of these metrics, illustrating the models' behaviours during the training process and their validation outcomes.

3.1 Neural Network Training and Validation

Figure 3 illustrates the training and validation accuracy for the NN model. The graph shows a convergence of training and validation accuracy, indicating a consistent performance as the number of iterations increases. Figure 4 displays the training and validation loss, where a decrease in loss denotes model learning improvement. The ROC curve in the third graph figure 5 is a plot of the true positive rate against the false positive rate at various threshold settings. The AUC is an indicator of the model's ability to distinguish between the classes. The curve stays well above the diagonal line, which would represent a random guess, confirming the model's predictive performance.

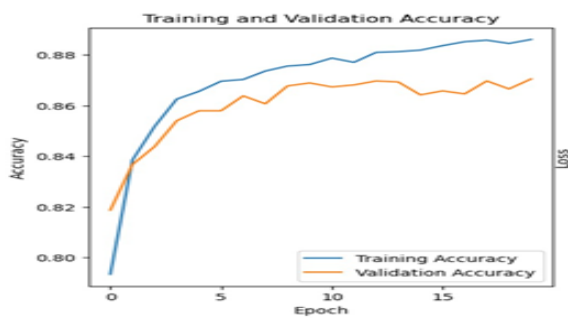


Figure 3: The chart shows the neural network training and validation accuracy over epochs, indicating learning progress.

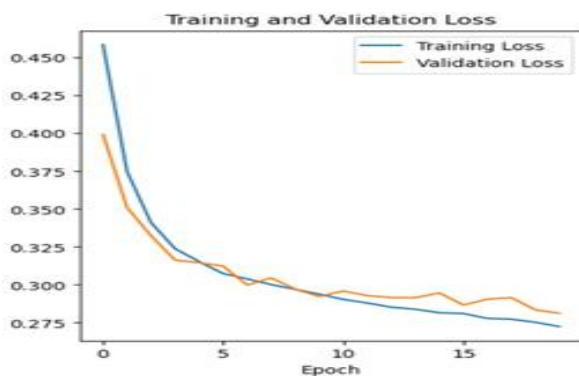


Figure 4: The chart illustrates the decline in training and validation loss over epochs for a machine learning

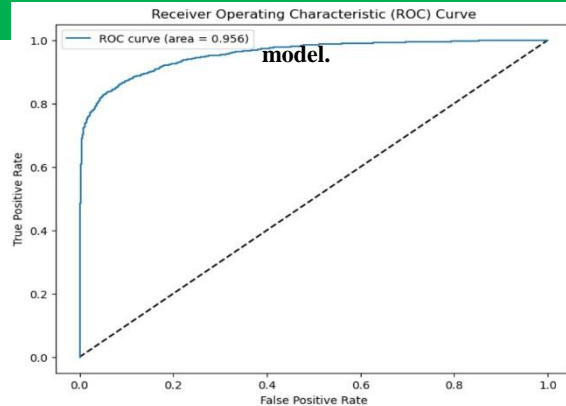


Figure 5: The graph displays an ROC curve demonstrating the model's classification performance with an AUC of 0.953.

These visualizations provide an overview of the NN's learning dynamics and its effectiveness in classifying the given data. The increase in accuracy and the AUC value are promising, yet the gap between the training and validation accuracy suggests room for further optimization to achieve better generalization

3.2 Stacked Ensemble Training and Validation

The performance of the stacked ensemble model across 10-fold cross-validation depicted in figure 15 shows a degree of predictive ability. The top-left graph details the log loss metric for each fold, where a lower log loss signifies a confident and accurate model prediction. The plot shows some variation in performance across the folds, indicating the model's sensitivity to the data. In the top-right, the ROC AUC scores are displayed for each fold, scores remain above 0.955 for most folds, confirming a good level of separability between the classes predicted by the model. Accuracy in The bottom-left graph show fluctuation in the graph may point to particular folds where the model either performed very well or underperformed. Finally, the bottom-right graph shows the ROC curve, plotting the TP rate against the FP rate, with an AUC of 0.953. This value indicates a strong predictive ability, where the model does well in distinguishing between the classes across varying thresholds. The performance across these metrics, despite some variability, suggests that the model offers reliable predictions for the task at hand.

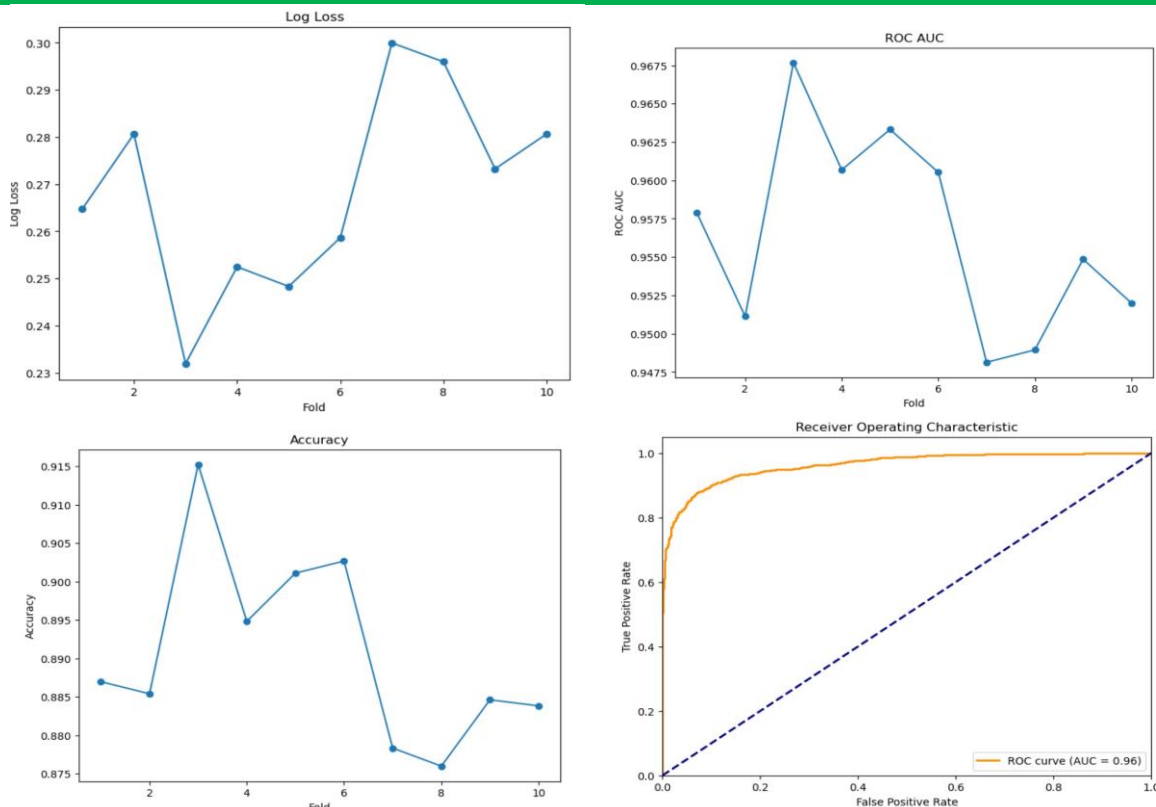


Figure 6: Overview of model performance metrics. From top left to bottom right: Log loss metric across 10 folds, ROC AUC scores across 10 folds, accuracy variation across 10 folds, and ROC curve with an AUC of 0.953.

3.3 Evaluation and Comparative Analysis

This section compares SE and a NN models to find the best predictive performance for CC in banking. Key metrics like Log Loss, ROC AUC, accuracy, precision, and F1-Score evaluated. The analysis highlights differences in their accuracy in predicting customer behaviour.

3.3.1 Neural Network Evaluation

The NN model, optimized through grid search with cross-validation, achieved a mean accuracy of 88%. The best parameters found for the NN were 'activation': 'logistic', 'alpha': 0.001, 'hidden layer sizes' (50,). The classification report shows a balanced performance in terms of precision, recall, and f1-score for both classes (0.89 across the board for both classes). ROC AUC of 0.9546 and a Log Loss of 0.2665 indicate a good model performance, especially considering the complexity and variability in customer behaviour data.

3.3.2 Stacked Ensemble Evaluation

The stacked ensemble model, combining predictions from XGBoost, RF, LGBM, and GBC with a LR meta-learner, shows a performance with an accuracy of 89.11%, an ROC AUC of

0.9609, and a lower Log Loss of 0.2542. This indicates that the ensemble method, which leverages the strengths of multiple models, offers a generalized approach for CCP. The classification report for the SEM also reflects a better balance in precision and recall, leading to a higher f1-score compared to the SEM model.

3.4 Evaluation Metrics

The study will evaluate churn prediction models using the following metrics:

- i. **ROC AUC:** ROC AUC shows how well the model distinguishes between customers who will leave and those who will stay. A higher ROC AUC means better performance. It measures the model's ability to correctly identify churners and non-churners without the influence of a threshold. The best possible score is 1, indicating perfect prediction, while a score of 0.5 suggests no better than random guessing.
- ii. **Log Loss:** Log loss is a method that measures the accuracy of a model's prediction, especially when there are only two outcomes. It calculates how much the model's predicted probabilities differ from the actual class labels, with lower values indicating accurate predictions.



- iii. **Accuracy:** Accuracy calculates the proportion of correct predictions made by the model out of all samples.
- iv. **Precision:** Precision evaluates the positive predictive value—the level of true positives among all cases predicted as positive. It informs selectivity but ignores false negatives undetected. High precision minimizes wasted retention efforts from false churn alerts (Umayaparvathi and Iyakutti, 2016).
- v. **Recall:** In contrast, recall calculates sensitivity or the true positive rate—the share of actual positives correctly identified by the model.
- vi. Recall is crucial for discovering the majority of customers likely to churn for proactive interventions. Models maximizing recall promote complete identification of high-risk cases (Duchemin and Matheus, 2021).
- vii. **F1 Score:** The F1 Score combines precision and recall through their harmonic mean to measure overall accuracy, accounting for both false positives and false negatives.
- viii. The F1 score is a key metric that balances precision and recall to accurately evaluate a model's prediction of customer churn. This, along with other metrics, offers a comprehensive view of the model's accuracy, thoroughness, and reliability.

3.5 Comparative Analysis: Stacked Ensemble vs. Neural Network

The SEM shows a higher accuracy (89.11%) compared to the neural network model (88%). This suggests that the ensemble method show slightly better at generalizing from the training data to unseen data. The ensemble model's higher ROC AUC score (0.9609 vs. 0.9573) further supports its superior ability to distinguish between classes. Both models demonstrate comparable performance in terms of precision, recall, and F1-scores, with variations across classes. However, the ensemble model shows a more balanced performance between the two classes, which can be important in applications like churn prediction where identifying churners (recall) and minimizing false alarms (precision) are important. The Stacked ensemble model also outperforms the neural network in terms of log loss (0.2542 vs. 0.2665), indicating better probability estimates by the ensemble model. This is pivotal in applications where the confidence of predictions is as important as the predictions themselves. However, Despite the NN's slightly lower performance metrics in this study, its capacity for handling larger, more complex datasets suggests that with additional data, NNs may offer advantages.

While the SEM show higher accuracy and better performance metrics than the NNs, it's important to recognize the limitations in both approaches within the dataset's size and complexity.

3.6 Challenges and Limitations of Neural Networks and Stacked Ensemble

3.6.1 Neural Networks

Imbalanced Datasets: NNs require tuning and data preprocessing to manage imbalanced datasets effectively. Techniques like oversampling the minority class, under sampling the majority class, or using loss functions can mitigate this but introduce complexity.

Non-linear Relationships: While neural networks are capable of modelling non-linear relationships between features, identifying the right architecture and hyper-parameters to capture these dynamics without over-fitting can be challenging. This process requires experimentation and domain knowledge.

Scalability with Large-scale Data: Neural networks can be computationally intensive, especially with large-scale datasets. Training deep learning models needs resources and time, which can be a limitation for banks with limited resources or requiring quick updates.

3.6.2 Stacked Ensemble

Model Complexity: managing multiple models and their interactions can be a challenge, requiring tuning and validation to ensure that the ensemble's performance is reflective of the data. **Resource Requirements:** Like NNs, ensemble methods can be resource-intensive, although parallel training of models can mitigate this to some extent. **Overfitting:** Understanding which part of the data is causing inconsistencies made within a SE can be challenging than with a single model, making it harder to derive business insights directly from the model's predictions.

This comparative analysis shows that while stacked ensembles currently show better performance for the given dataset, neural networks hold considerable potential, especially in larger-scale data environments. Future work should not only aim to refine and compare these models but also explore approaches to harness the full capabilities of neural networks in churn prediction and beyond.

4.0 DISCUSSION

4.1 Interpretation of Findings

The NN model, optimized through grid search with 10-fold cross-validation, exhibited strong predictive



performance, achieving an ROC AUC score of 0.9546 and an accuracy of 88.45%. Despite these results, the exploration into SEM, incorporating base learners such as RF, XGB, LGBM, and GBC, presented an opportunity to surpass the NN's capabilities.

The analysis between the SEM and a NN model has reveals the capabilities of both machine learning strategies in addressing the issue of customer churn. Notably, while the SEM demonstrates higher accuracy (89.11%), superior ROC AUC (0.9609), and lower Log Loss (0.2497), indicating its more effective integration of predictions from multiple algorithms, the NN's performance suggests potential, particularly in scenarios involving larger or more complex datasets.

The finding that NNs with the right dataset size could match the predictive accuracy of SEM introduces an important consideration for future research. This insight underscores the importance of dataset characteristics and the scalability of models in achieving optimal performance.

The comparative analysis between the stacked ensemble model and the neural network shows the models ability to generalize from training data to unseen data. the ensemble model demonstrated slightly more balanced performance across classes, a factor in churn prediction where accurately identifying churners and minimizing false positives are important.

4.1 Practical Implications

The performance of the stacked ensemble model and the neural network has implications for both academic research and practical applications in the banking sector:

Our findings highlight the value of SEM and NN model in developing accurate churn prediction systems for the banking industry. The ability of these models to identify customers at risk of churn enables banks to implement retention strategies. However, the performance of neural networks, given their ability to leverage large datasets, suggests that banks with access to large amounts of customer data and the resources to support complex model training could benefit from integrating NN architectures.

These models not only enhance the accuracy of churn prediction but also offer banks avenue for targeted interventions aimed at addressing the needs of at-risk customers. For instance, personalized banking services, improved customer engagement programs, could be developed to enhance customer satisfaction and loyalty. This approach not only saves costs but also enhances

the customer experience by ensuring that interventions are relevant to those who receive them.

4.3 Challenges and Limitations

While both models present advantages, they also face challenges. NNs, for instance, require tuning and preprocessing to address imbalanced datasets, a common challenge in churn prediction. On the other hand, SEMs despite their superior performance on the dataset used, entail complexity in management and may demand computational resources. The absence of strong correlations among most features, poses a challenge in feature selection and engineering. This highlights the role of domain expertise and feature engineering in enhancing the predictive models.

Ensuring the privacy and security of sensitive data is an issue for banks, who must navigate between protecting customer information and utilizing it to improve their predictive models. The deployment of technologies like NNs and SEMs demands a careful balance: adhering to regulatory and ethical standards, efficiently managing resources, and leveraging expert knowledge in data handling. Overcoming these challenges is essential for banks to use these technologies to minimize customer churn effectively

5.0 CONCLUSION

The study examined customer churn prediction in the banking sector using a dataset of 10000 customer records and 14 variables downloaded from kaggle database .the process began with data visualisation followed by feature engineering specifically domain-specific feature engineering was applied to find the most essential feature to improve the accuracy of churn predictions. the study used ensemble classifiers such as Random Forest, Gradient Boosting, LightGBM, XGBoost, as base learners and stacked them into logistic regression as a meta learner resulting in a more efficient model that would provide key insight to customer churn behaviour . We used a neural network and fine-tuned its hyperparameters to ensure the best performance. The paper compares the effectiveness of the stacked ensemble models (SEM) against Neural Networks (NN) in terms of predictive accuracy and reliability. The SEM emerged as the superior model, showcasing a Log Loss of 0.2497, ROC AUC of 0.9609, and an accuracy of 89.11%, outperforming the individual base models. This highlights the effectiveness of ensemble learning in handling complex patterns and interactions within the data, thus providing a reliable predictive framework for identifying potential customer churn



This study has added to the body of knowledge of data mining and machine learning by using stacked ensemble classifiers to examine churn prediction in the banking sector. The method used in this study has been able to combine the best classifiers used in previous researches to provide an effective model of churn prediction. The proposed method will be useful for more accurate prediction outcome.

Based on the findings and the challenges, The analysis was constrained by the size and the feature set of the dataset used. Future research could explore the implications of incorporating larger datasets with more features. Future research should also explore the integration of newer base learners and meta-learners to further refine the predictive capabilities of stacked ensemble models. As well as explore advanced neural network architectures. Additionally, developing hybrid models that combine the strengths of neural networks and ensemble methods could offer an avenue for improving churn prediction performance.

While the stacked ensemble model shows superior performance in this context, the choice between these models should be guided by the characteristics of the dataset, computational resources, and the need for model interpretability. This conclusion sets the stage for further exploration and innovation in predictive analytics to enhance customer retention strategies in the banking sector.

REFERENCES

- Awang, M. K., Makhtar, M., Udin, N., and Mansor, N. F. (2021). Improving customer churn classification with ensemble stacking method. *International Journal of Advanced Computer Science and Applications*, 12(11).
- Bent'ejac, C., Csörgő, A., and Martínez-Munoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967.
- Duchemin, R. and Matheus, R. (2021). Forecasting customer churn: Comparing the performance of statistical methods on more than just accuracy. *Journal of Supply Chain Management Science*, 2(3-4):115–137.
- Haddadi, S. J., Farshidvard, A., dos Santos Silva, F., dos Reis, J. C., and da Silva Reis, M. (2024). Customer churn prediction in imbalanced datasets with resampling methods: A comparative study. *Expert Systems with Applications*, 246:123086.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

- Huseyinov, I. and Okocha, O. (2022). A machine learning approach to the prediction of bank customer churn problem. In *2022 3rd International Informatics and Software Engineering Conference (IISEC)*, pages 1–5. IEEE.
- Kalule, R., Abderrahmane, H. A., Alameri, W., and Sassi, M. (2023). Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs. *Scientific Reports*, 13(1):9855.
- Oladimeji, O. M., Ajiboye, A. R., and Usman-Hamza, F. E. (2023). An optimized stacking ensemble technique for creating prediction model of customer retention pattern in the banking sector. *Gadua Journal of Pure and Allied Sciences*, 2(1):22–29.
- Osman, Y. and Ghaffari, B. (2021). Customer churn prediction using machine learning: A study in the b2b subscription based service context.
- Rahman, M. and Kumar, V. (2020). Machine learning based customer churn prediction in banking. In *2020 4th international conference on electronics, communication and aerospace technology (ICECA)*, pages 1196–1201. IEEE.
- Umayaparvathi, V. and Iyakutti, K. (2016). A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. *International Research Journal of Engineering and Technology (IRJET)*, 3(04).
- Vu, V.-H. (2024). Predict customer churn using combination deep learning networks model. *Neural Computing and Applications*, 36(9):4867–4883.
- Wang, C., Han, D., Fan, W., and Liu, Q. (2019). Customer churn prediction with feature embedded convolutional neural network: An empirical study in the internet funds industry. *International journal of computational intelligence and applications*, 18(01):1950003.
- Wu, H. (2022). A high-performance customer churn prediction system based on selfattention. *arXiv preprint arXiv:2206.01523*.
- Wu, X., Li, P., Zhao, M., Liu, Y., Crespo, R. G., and Herrera-Viedma, E. (2022). Customer churn prediction for web browsers. *Expert Systems with Applications*, 209:118177.