Customers Churn Prediction in Banks Using Artificial Neural Network.

Darshan B S
Department of Artificial Intelligence and Data Science
Global Academy of Technology
Bengaluru, Karnataka, India
darshanbsbs18@gmail.com

Darshan S
Department of Artificial Intelligence and Data Science
Global Academy of Technology
Bengaluru, Karnataka, India
darshandarsh5655@gmail.com

Ashwini Kodipalli
Department of Artificial Intelligence and Data Science
Global Academy of Technology
Bengaluru, Karnataka, India
dr.ashwini.k@gat.ac.in

Abstract - The loss of clients to rival banks, or customer churn, is a serious danger to financial organizations. Churn modeling provides a data-driven approach to address this problem. Churn models forecast which customers are most likely to leave by looking at account activity, service interactions, and customer demographics. Equipped with this understanding, financial institutions can proactively execute focused retention tactics such as tailored promotions, better customer support, and amplified correspondence. This proactive strategy reduces attrition, increases client loyalty, and eventually makes a bank more competitive. Churn prediction models play an important role in banking. These models examine consumer data using machine learning algorithms that consider credit score, age, duration, balance, and geography. Banks can identify customers who are likely to leave by assessing their turnover rate. Early detection enables tailored engagement via offers or services, which increases retention rates. Ultimately, lowering turnover leads to higher earnings and a better customer experience.

Keywords - Neural Network, Artificial Neural Network, Dropout, Classification, Customer Churn, Deep Learning,

I.INTRODUCTION

In today's highly competitive banking environment, keeping customers is crucial. Gaining new clients is an expensive undertaking that frequently calls for intensive advertising efforts and promotional deals. Conversely, current clients are a dependable and well-established source of income. But there's always the threat of consumer attrition. In the banking industry, the term "churn" describes the loss of a client who moves to another financial organization. The profitability of a bank may suffer significantly as a result. Research indicates that keeping current clients can be up to five times less expensive than finding new ones. Loyal clients also frequently use a greater range of services and produce more income, making them more profitable. Here's where churn modeling comes into play. Churn modeling is a potent statistical method that forecasts the chance of a customer quitting the bank based on past customer data. Through proactive implementation of targeted retention tactics, banks can identify clients who are at risk of leaving. The importance of churn modeling in banking is examined in this introduction, along with its advantages and insights. We'll also go over the different modeling approaches used and the kinds of consumer data that are used in churn models.

II. LITERATURE REVIEW

Customer Churn Prediction Model Using Artificial Neural Networks (ANN): A Case Study in Banking

In this, study utilized Churn Modelling dataset from Kaggle, comprising 10,000 samples with features like demographic data, transaction history, and customer behavior, to predict customer churn in a bank using an Artificial Neural Network. After thorough preprocessing involving cleaning, transformation, feature selection, and scaling, the ANN model, employing ReLU activation in hidden layers and by increasing depth, the model achieved an 86% accuracy rate. At last key findings indicated that higher churn among customers with credit scores of 600-700, around age group of 50, and earning salaries between 175,000 to 200,000. The ANN model outperformed Logistic Regression in performance metrics, suggesting actionable insights for banks to enhance retention strategies. And also stated that future research could focus on improving dataset quality with advanced feature selection methods to further enhance model effectiveness.

Predicting Customer Churn in Telecom Industry using Multilayer Perceptron Neural Networks: Modeling and Analysis:

This research focused on predicting customer churn in a telecommunication company using MLP (multilayer perceptron) neural network with real data from Umniah, it is a major telecom network in Jordan. The dataset included overall 11 attributes from 5000 randomly selected customers over a three-month period. Beyond churn prediction, the study also aimed to determine the relative importance of each input variable. Utilizing k-fold cross-validation with k=5, MLP model with 4 and 6 parameters in the hidden layer, it consistently demonstrated superior accuracy. Variable importance was assessed by training the dataset multiple times while excluding each of the variable, it made impact on accuracy and churn rate. And mentioned key findings that Total monthly fees, Total of international outgoing calls, and 3G service as the significant predictors of churn. At last concluded that these insights are crucial for optimizing Customer Relationship Management strategies in telecom companies.

Predicting customer churn in banking industry using neural networks:

This article explores the application of neural networks in the banking industry as a dataset of 1866 clients. It focuses on factors such as sex, age, private status, average monthly income, internet banking usage, and the number of bank products used to predict client churn. The study highlights that clients using multiple bank products are less likely to leave compared to those with fewer products. The neural network model, optimized with three hidden layers (8, 4, and 2 neurons), enables the accurate prediction of client churn risk based on input parameters. In this the author identifies a potentially valuable group—such as students initially seen as unprofitable due to limited product usage—that banks could focus on them by providing student loans with favorable interest rates and says that it leads to long-term profitability and client retention strategies.

Customer churn is considered one of the significant issues facing the banking sector, and keeping clients is essential to the success of the company. Customer churn is the process through which clients stop doing business with one bank and switch to another [1]. Banks now depend heavily on churn prediction algorithms to predict and stop client churn. Customer Churn Prediction is a type of Client Relationship Management (CRM) where an organization develops a model that forecasts whether a customer is intending to leave or reduce its purchases from a company.

Managing customer churn and improving customer relationship management (CRM) is essential for banks so they are able to protect their loyal customers, ensure growth, and improve customer service. [5]. Banks face one of the most difficult challenges when it comes to retaining churn-prone customers.

The researcher built a predicted churn model, and 12 classification algorithms were applied in the study to a dataset of actual credit cardholders' behavior from a major Chinese commercial bank.

The research work [17] used a NN to predict the exhaustion of customer in cellular service. Data features such as age, gender, consumer status (retired, student, employed, unemployed), average monthly income, and whether the client uses 2 or more products were used as control variables for NN.

Customer churn is considered one of the significant issues facing the banking sector, and keeping clients is essential to the success of the company. Customer churn is the process through which clients stop doing business with one bank and switch to another [1]. Banks now depend heavily on churn prediction algorithms to predict and stop client churn. Customer Churn Prediction is a type of Client Relationship Management (CRM) where an organization develops a model that forecasts whether a customer is intending to leave or reduce its purchases from a company.

Managing customer churn and improving customer relationship management (CRM) is essential for banks so they are able to protect their loyal customers, ensure growth, and improve customer service. [5]. Banks face one of the most difficult challenges when it comes to retaining churn-prone customers.

Artificial neural networks

Artificial neural networks (ANNs) are inspired by the structure and function of the human brain. They consist of interconnected nodes, loosely mimicking biological neurons. These nodes process information and transmit signals to other nodes, with the strength of the connections influencing the overall output. Unlike traditional programming, ANNs learn through training on massive datasets. By adjusting the connections between nodes based on the accuracy of their outputs, the network progressively improves its ability to recognize patterns and make decisions.

This ability to learn and adapt makes ANNs powerful tools for a wide range of applications. They excel at tasks involving complex data analysis, like image recognition in facial scanning software or spam filtering in email. They can also be used for prediction, such as forecasting stock prices or weather patterns. Additionally, ANNs are at the core of natural language processing, allowing machines to understand and generate human language, which is crucial for virtual assistants and machine translation.

A standard feedforward multi-layer perceptron neural network consists of input, output, and hidden layers. Neural networks are classified into single-layer perception (SLR) and multilayer perception (MLP) networks. There might be one or more buried layers of neurons in between. These intermediate levels are referred to as hidden layers, and the nodes inside them are known as hidden nodes as they do not get direct input from outside sources. Each link has a corresponding weight. Figure shows the architecture of Artificial Neural Networks.

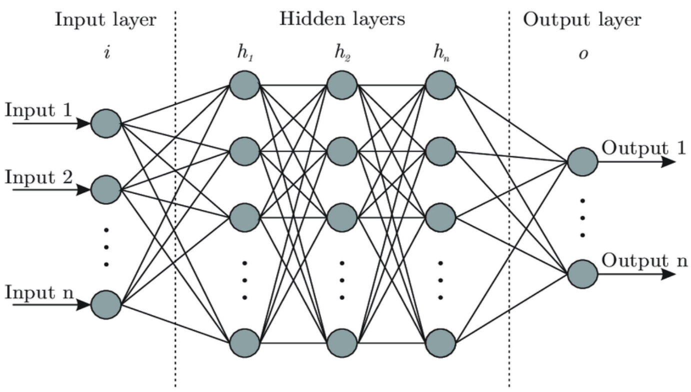


Fig. Artificial Neural Networks Architecture

Tools and Libraries

Pandas: is the most popular Python programming language package that offers powerful, expensive and flexible data structures that make data manipulation and analysis easy
Numpy: is the fundamental package for scientific computing in Python programming language that contains a

powerful N-dimensional array object and also useful in linear algebra.

Seaborn: is a Python data visualization library based on matplotlib that provides high-level interface for drawing attractive and informative statistical graphics.

TensorFlow: is an open source Python library used for machine learning applications such as neural network and used Keras as a backend.

Keras: is a neural network framework for Python programming language that provides a convenient way to define and train almost any kind of deep learning model.

Flowchart: illustrate the step by step procedure on how to predict customers churn with the multilayer perceptron of artificial neural network architecture.

Matplotlib: is an amazing visualization library in Python programming language for two-dimensional plots of arrays. One of the greatest benefits of visualization is that it allows high dimensionality data to visualize and easily understandable and it consists of several plots like line, bar, scatter, histogram and so on.

Sci-kit Learn*:* it is a free machine learning library for Python programming language that designed to interoperate with Python numerical Numpy and scientific libraries SciPy. Also, it can be used for classification, regression and clustering algorithms including support vector machine, linear and logistic regression, random forests, gradient boosting, decision tree, K-means and so on.

Dense A dense neural network, also known as a Fully Connected Neural Network (FCNN) or Multi-Layer Perceptron (MLP), is a type of neural network architecture where each neuron in one layer is connected to every neuron in the next layer. This design allows the network to learn complex patterns and relationships in data

Dropout: Dropout is a regularization technique used in deep learning to improve model performance and prevent overfitting in neural networks. It works by randomly setting a percentage of input and hidden units to zero during the training process

Loss: the Loss function is a method of evaluating how well your algorithm is modeling your dataset. It is a mathematical function of the parameters of the machine learning algorithm

Epoch: An epoch in machine learning means one complete pass of the training dataset through the algorithm. This epoch's number is an important hyper parameter for the algorithm. It specifies the number of epochs or complete passes of the entire training dataset passing through the training or learning process of the algorithm

III. METHODOLOGY

Churn Modelling, an open-source dataset obtained from the Kaggle website, was used for the study. Data cleaning, data transformation, feature selection, feature scaling, and data splitting into train and test sets were all done before the data was analyzed. The performance matrices will be used to assess the trained machine learning model. In the event that the suggested model performs poorly, certain model optimization approaches must be applied in order to get the desired outcomes. In the banking industry, the purpose of the ANN model is to forecast whether or not a client would churn. In this study, machine learning has been used to forecast client attrition in the banking sector.

Dataset

The dataset, which is openly accessible, contains a variety of characteristics and information on a bank's clientele, including transaction history, demographic data, and customer behavior.

Among the primary characteristics of the dataset are:

• Customer demographic data, such as age, gender, and location.

• Transaction History: length of service, credit score, and amount owed, quantity of goods, projected income, etc.

• Customer behavior, such as HasCrCard, IsActiveMember, etc.

The target variable is a binary value that depends on whether the client has closed his account (left the bank) or is still a customer. This may be ascertained by looking at the customer's activity level or if they have completed any transactions in a specific time frame. Customers who remain with the bank are depicted as 0, while those who go are represented as 1. This dataset consists of 10,000 rows and 14 different columns. Table I provides a summary of the attributes shown in the dataset:

Row number	Row Numbers from one to ten thousand
Customer ID	It is a unique ID for the customer identification
Surname	It shows the surname of the customer
Geography	Location of the customers
Gender	Male or female
Age	The age of the customer
Tenure	Number of years the customer joined the bank
Balance	Customer balance
No of products	Number of products the customer is using
Has Credit Card	Binary flag, if the customer holds a credit card or not
Is Active Member	Binary flag, the customer is an active member with the bank or not
Estimated Salary	Estimated salary of the customer salary in dollars
Exited	Binary flag 1 for closing the account and 0 if the customer is retained

Data processing

An essential phase in the machine learning (ML) process is data processing. It entails getting the data ready for use in creating and honing machine learning models. Data processing makes sure your machine learning models get clear, insightful data for precise forecasts and insights. Similar to laying the foundation for a home, data processing is an essential phase in machine learning. A machine learning model requires clean, processed data to work successfully, much like a home needs a strong foundation to be sturdy.

Data Cleaning

Because it guarantees the quality of the data used to train models, data cleaning is an essential step in the machine learning process. Errors, inconsistencies, or missing numbers in dirty data can produce unreliable and erroneous findings. Finding and resolving these problems, such as eliminating duplicates, rectifying formatting mistakes, and managing outliers, are all part of the cleaning process. Data scientists can guarantee that their models are trained on a solid basis and produce more accurate forecasts and trustworthy insights by cleaning the data.

Data Transformation

An essential first step in any machine learning effort is data transformation. It entails taking unprocessed, sometimes erratic, raw data and transforming it into a format appropriate for analysis and machine learning model training. This may entail modifying the data by scaling features or encoding categorical variables, as well as cleaning the data by managing missing values and outliers. Data transformation aids in the more efficient learning and correct output of machine learning models by making the data more consistent and comprehensible.

Feature Selection

A critical phase in machine learning is feature selection, which is selecting the most useful characteristics from your data. It aids in helping you concentrate on the important details, much like organizing your closet. There are several ways in which you may enhance your machine learning model's performance by selecting the appropriate features. First, it simplifies the data, which helps the model learn more quickly and easily. It also aids in preventing overfitting, a condition in which a model becomes overly adept in memorization of training data and is ill-suited to adapt to new data. Lastly, by emphasizing pertinent characteristics, you may simplify the interpretation of your model and gain a better grasp of the variables that are most crucial to your predictions.

Feature Scaling

In machine learning, feature scaling is an essential step in the preparation of data. In essence, it unifies the variety of characteristics in your dataset. Assume you have the following characteristics: age (0–100 years) and income (in millions). Without scaling, a machine learning model may give more weight to reducing the income characteristic than age, which would result in predictions that are off. This is addressed by feature scaling strategies like normalization and standardization, which raise each feature to a comparable range and guarantee that each feature contributes equally to the model's learning process. This increases many machine learning algorithms' efficiency and accuracy.

Data Splitting

A crucial stage in machine learning is data splitting, which divides a dataset into distinct subgroups. The machine learning model is trained, validated, and tested using these subsets. The majority of the data is called the training set, which is used to train the model so that it can identify patterns in the data. By assessing the model's performance and avoiding overfitting, the validation set (optional) assists in fine-tuning the model during training. Finally, the model's generalizability and real performance on unseen data are evaluated using the test set, which the model has never seen before. Machine learning practitioners may efficiently train, optimize, and assess their models by dividing the data.

Model selection

To pick the best model from a pool of candidates in machine learning, model selection is an essential step. The method entails assessing many algorithms together with their setups, or hyper parameters, in order to ascertain which one works best for a given job and dataset. Not only does the optimal model perform well on training data, but it also generalizes well to new situations. This is ensured by model selection approaches such as cross-validation, which divide data into training and testing sets in order to evaluate the model's performance on unseen data. Finding a model that balances accuracy, complexity, and adaptability to new data is the ultimate objective.

Model Evaluation

An essential phase in machine learning is model assessment. In essence, it's evaluating the model's learning performance by looking at its homework. This entails evaluating the model's performance on hypothetical data using metrics. Different metrics are used for different tasks; for example, mean squared error is used in regression while accuracy is used in classification. We can evaluate the model's generalizability to fresh data and pinpoint its advantages and disadvantages by examining these measures. This aids in determining if the model is suitable for usage in practical applications.

Model Optimization

The technique of optimizing a machine learning model for optimal performance is known as model optimization. In doing so, accuracy and efficiency must be balanced. The model's capacity to learn from data may be enhanced by methods like pre-processing the input and adjusting the hyper parameters. Techniques like model compression and architecture selection can lower the amount of computing power needed for training and executing complicated models. All things considered, model optimization guarantees that you obtain the most accurate outcomes while maintaining the model's practicality.

IMPLEMENTATION

Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the biological neural networks found in human brains. Among ANNs, multilayer perceptron (MLP) neural networks are widely used for various tasks such as classification and regression due to their ability to learn complex patterns from data.

The methodology for building an MLP typically begins with problem formulation and data collection**.** It's crucial to clearly define the problem you want to solve, whether it's predicting outcomes, classifying data into categories, or another task.

Designing the MLP architecture is a critical step where the neural network's structure is defined. This includes determining the number of layers (input, hidden, and output layers), the number of neurons in each layer, and the activation functions used in each neuron. The first and the last layers represent the inputs and outputs of the system respectively. And also in MLP, each hidden layer node actually consists of two parts; the first one contains the summation function which calculates the sum of each input value multiplied by the corresponding weight.

Summation function can be represented by following equation:

$$S = \sum_{i=0}^{n} WiXi$$

Where n is the number of inputs to the neuron.

And the other part of the neuron is the *Activation Function*. Activation function determines whether a neuron should activate based on the input it receives. This function normally nonlinear, allowing neural networks to learn complex patterns and relationships in data. In this Sigmoid function is the most commonly used activation functions.

The use of the sigmoid function has many advantages in the modeling of complex systems based neural. Sigmoid function makes useful for binary classification and logistic problems, especially for feed forward neuron networks, it maps the input values to the value between 0 and 1. Its mathematical form is:

$$\delta(Z) = \frac{1}{1+e^{-Z}}$$

ANN finds ways to minimize some errors, so it uses learning algorithm. In this we depend upon back propagation algorithm as learning algorithm. Here, back propagation allows neural networks to learn from its mistakes, by comparing it to the actual outcomes and little by little it keeps on adjusting to get closer to the right value, it gradually improves the prediction. Forward propagation is straightforward and involves simply passing data through the network in the direction from input to output.

The adjustments are typically made using an optimization algorithm like gradient descent. The process of making predictions, calculating error, propagating it backward, and adjusting weights is repeated for multiple iterations (epochs) until the network learns to make predictions with minimal error on the training data

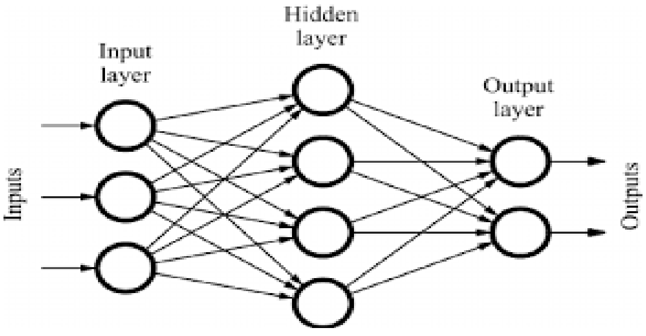


Fig. Feed forward MLP architecture.

RESULTS AND DISCUSSION

Confusion matrix

The resulting confusion matrix is illustrated in Figure

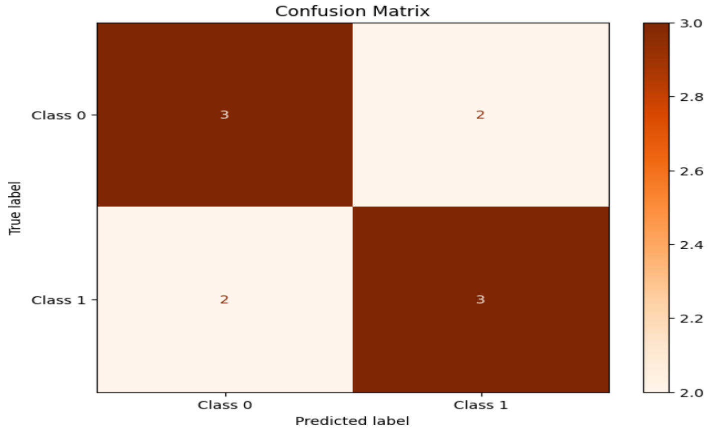


Fig. Confusion Matrix

Classification Report

```
79/79 [==============================] - 0s 1ms/step
Adagrad - Accuracy: 86.0400, AUC-ROC: 83.3899
Confusion Matrix:
[[1947   68]
 [ 281  204]]
Accuracy Score: 0.8604
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.97      0.92      2015
           1       0.75      0.42      0.54       485

    accuracy                           0.86      2500
   macro avg       0.81      0.69      0.73      2500
weighted avg       0.85      0.86      0.84      2500
```

Fig. Classification Report

Loss and Validation Loss curves

This shows how the training and validation errors have been reduced during 50 epochs

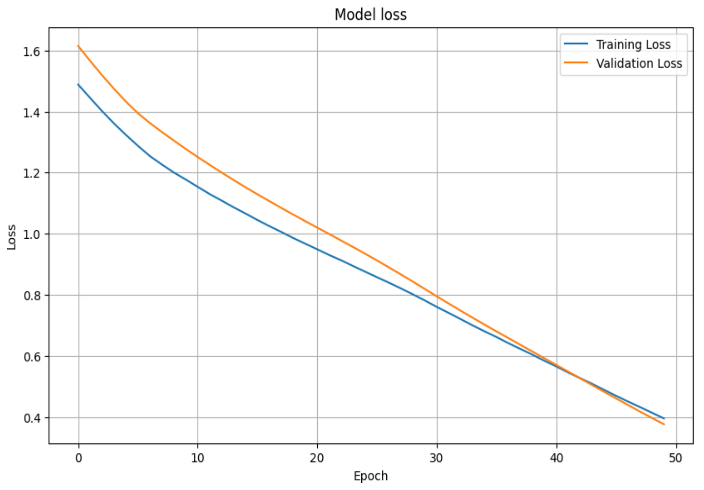


Fig. Loss and Validation Loss curves

Training and Validation Accuracy curves

This shows how the training and validation accuracy has been increased during 50 epochs.

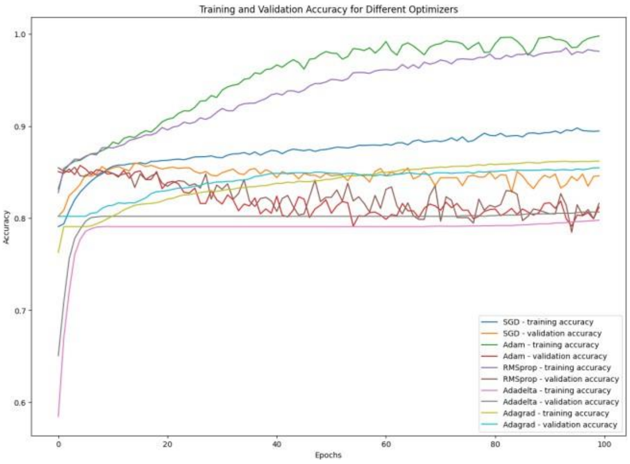


Fig.Training and Validation Accuracy curves

Receiver operating characteristic curve

(ROC) curve is a graphical visual that illustrates the predictive ability of a binary classifier system

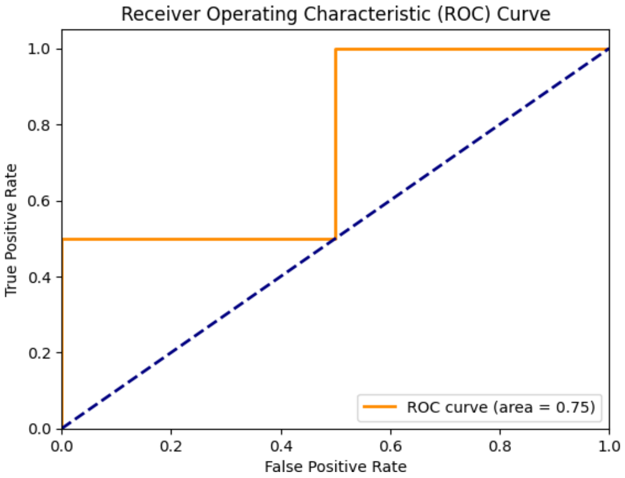
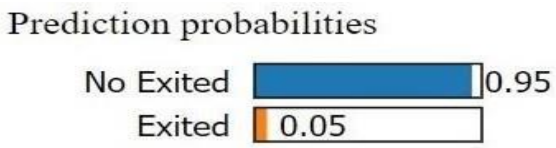
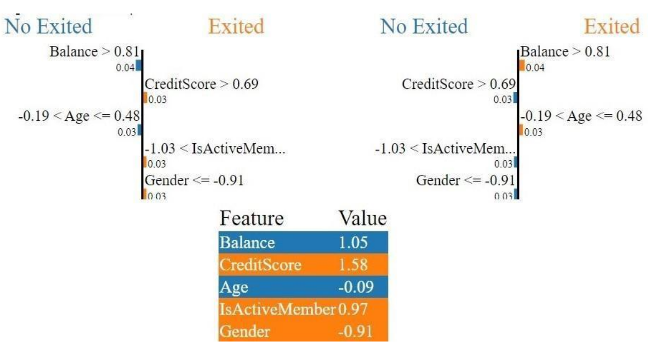


Fig.ROC Curve

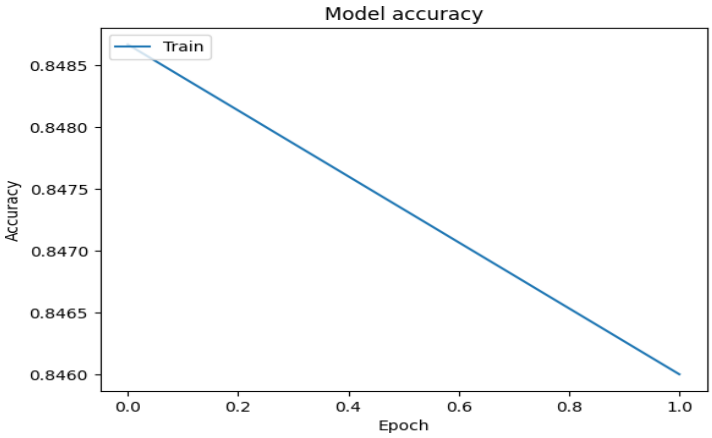




SUMMARY OF THE RESULTS

Ann optimizers	Class 0				Class 1			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
SGD	85%	0.88	0.94	0.91	85%	0.66	0.46	0.54
Adam	82.7%	0.88	0.90	0.89	82.7%	0.56	0.51	0.53
RMSprop	81%	0.88	0.88	0.88	81%	0.51	0.50	0.51
Adadelta	81%	0.81	0.99	0.89	81%	0.82	0.30	0.50
Adagrad	86%	0.87	0.97	0.92	86%	0.75	0.42	0.54

ACCURACY



CONCLUSION

The purpose of this study was to use artificial neural networks (Anns) to develop a robust model for forecasting customer attrition in the banking industry. The primary goal was to support banks in proactively identifying potential consumers and taking appropriate measures to prevent client attrition.

The most important variables for the prediction model were found using exploratory data analysis approaches. Eventually, an array of performance indicators, including the parameters like accuracy, precision, recall, and F1-score, were used to develop, train, and verify the ANN model. The results showed that the proposed ANN model could predict customer attrition with an accuracy rate of 86%.

Eventually, an array of performance indicators, including the parameters like accuracy, precision, recall, and F1-score, were used to develop, train, and verify the ANN model. The results showed that the proposed ANN model could predict customer attrition with an accuracy rate of 86%.

When compared to the LR model's performance, the ANN model has done better. The study pinpoints the precise areas that banks should focus on in order to aid in client retention. Although a thorough comparative analysis of several machine learning models, including Random Forest (RF), Decision Trees (DT), KNN (K-Nearest Neighbor), and Support Vector Machines (SVM), was not covered in the study, it does point the way in new avenues for further investigation.

Just 10,000 data samples were employed in the study's short and severely unbalanced dataset for the model-development procedure. Yet, in contrast to the selected dataset, bank data in the actual world will be far larger. In order to increase the churn prediction model's accuracy, future studies might look at a variety of machine learning approaches, including ensemble models, deep learning, and hybrid models. It is possible to compare the ANN model to other machine learning models.

Further investigation will provide more insights into the various Machine Learning techniques that may be used to train the dataset and will be able to determine which model is the best predictive.

Future studies can also concentrate on enhancing the dataset's quality by applying sophisticated feature selection techniques to improve the model's performance.

REFERENCES

1. Sharma, Anuj, and Dr Prabin Kumar Panigrahi. "A neural network based approach for predicting customer churn in cellular network services." *arXiv preprint arXiv: 1309.3945* (2013).

2. Amuda, Kamorudeen & Adeyemo, Adesesan. (2019). Customers Churn Prediction in Financial Institution Using Artificial Neural Network.

3. Tanishq Vyas , machine learning churn modelling using artificial neural network

4. Iranmanesh, Seyed & Hamid, Mahdi & Bastan, Mahdi & Shakouri G., Hamed & Nasiri, Mohammad Mahdi. (2019). Customer Churn Prediction Using Artificial Neural Network: An Analytical CRM Application.

5. Willi Akbar Satria , Iskandar Fitri , Sari Ningsih ,(2020). Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks

6. Bilal Zorić, Alisa. (2016). Predicting Customer Churn in Banking Industry using Neural Networks. Interdisciplinary Description of Complex Systems. 14. 116-124. 10.7906/indecs.14.2.1.

7. Adwan, Omar & Faris, Hossam & Jaradat, Khalid & Harfoushi, Osama & Ghatasheh, Nazeeh. (2014). Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks: Modeling and Analysis. Life Science Journal. 11. 75-81.

8. Ramesh, P., Jeba Emilyn, J. & Vijayakumar, V. Hybrid Artificial Neural Networks Using Customer Churn Prediction. Wireless Pers Commun 124, 1695–1709 (2022). https://doi.org/10.1007/s11277-021-09427-7

9. Satria, W. A., Fitri, I. and Ningsih, S. (2020) “Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks: Prediction of Customer Churn in the Banking Industry Using Artificial Neural Networks”, Jurnal Mantik, 4(1), pp. 936-943. Available at: https://iocscience.org/ejournal/index.php/mantik/article/view/871 (Accessed: 30June2024).