CUSTOMER CHURN PREDICTION IN BANKS

(Submitted to prof. Sudha S)

(VIT Institute Of Technology, Vellore, Tamilnadu)

Authors Name/s Harshita sorout
school of Computer Science and Engineering
VIT Institute of Technology
Vellore, Tamil Nādu,
India
email id:
harshita.sorout2019@vitstudent.ac.in

Authors Name/s Shirley Tephilla
school of Computer Science and Engineering
VIT Institute of Technology
Vellore, Tamil Nādu,
India
email id:
shirleytephilla.p2019@vitstudent.ac.in

***Abstract*—**
Customer churn prediction is a significant issue for the banking sector since it enables banks to spot clients who are at risk of leaving and take measures to keep them.
Based on a dataset of customer transactions and demographics, this research provides a machine learning approach to forecast client attrition in banks. We compare the performance of various machine learning techniques:
such as logistic regression, decision trees, and random forests, on the dataset. Our findings demonstrate that random forests perform better than other algorithms, accurately forecasting customer attrition by 85%. This strategy can assist banks in proactively identifying clients at risk of leaving, taking the necessary steps to keep them, and enhancing client happiness and loyalty.

Keywords— Random Forest algorithm, accuracy, radon metrics (*key words****)***

I. INTRODUCTION

Customer churn is the phenomena where customers stop using a company's services or products. client turnover is a major issue for the banking sector since it has an impact on revenue, client loyalty, and satisfaction.

Therefore, in order to keep their clientele and preserve profitability, banks must anticipate customer turnover and take action to avoid it. The goal of customer churn prediction in banks is to spot clients who may leave and take the necessary precautions to keep them. Banks can anticipate client attrition using a variety of methods, including statistical models and machine learning algorithms. These techniques analyse consumer information, including transaction history, demographics, and credit history, to spot trends and gauge the possibility of churn.

The use of machine learning algorithms for predicting customer turnover in banks will be the main topic of this study. We will create a predictive model that can recognise consumers who are likely to leave using a dataset of customer transactions and demographics. In order to determine the best precise algorithm for predicting customer turnover in banks, we will analyse the performance of various machine learning techniques, including logistic regression, decision trees, and random forests. The findings of this study will assist banks in taking proactive steps to retain clients, raise client satisfaction levels, and eventually boost income.

II. HARDWARE AND SOFTWARE COMPONENTS:

Hardware components –

- Intel Processor - known for CPUs based on its x86 architecture, which was created in the 1980s and has been continuously modified, revised and modernized.
- RAM: 8/16 GB
- 256 GB storage

Software Components –

- Windows OS- Microsoft created the operating system called Windows. You are only able to utilise a computer thanks to

the operating system. Windows is the most widely used operating system in the world since it comes preloaded on many new personal computers (PCs).

- ❖ Jupyter Notebook- To create and share documents with live code, equations, visualizations, and text, you can use the free and open-source Jupyter Notebook web tool. The staff of Project Jupyter is in charge of maintaining Jupyter Notebook.

A. ***Problem statement:***

The use of machine learning algorithms for predicting customer turnover in banks will be the main topic of this study. We will create a predictive model that can recognise consumers who are likely to leave using a dataset of customer transactions and demographics. In order to determine the best precise algorithm for predicting customer turnover in banks, we will analyse the performance of various machine learning techniques, including logistic regression, decision trees, and random forests. The findings of this study will assist banks in taking proactive steps to retain clients, raise client satisfaction levels, and eventually boost income.

In order to solve this issue, it is necessary to create a machine learning model that can properly forecast client turnover in banks. To find trends and forecast the possibility of churn, the model should be able to analyse customer data including transaction history, demographics, and credit history. Additionally, the model must be scalable to manage enormous datasets and be able to deliver insights that can help banks take preventative action to keep clients.

B. ***Modules:***

For predicting client attrition in banks, a few modules can be employed. we can combine these modules to create a thorough churn prediction system. These are a few of the important modules:

- Data gathering: This module entails gathering client data, including transaction history, demographics, credit history, and other pertinent details. Numerous sources, including as consumer databases, social media, and outside data suppliers, can be used to get the information.
- Data pre-processing**:** To prepares the data for analysis, this module comprises cleaning and altering it. This can entail codifying categorical variables, resolving missing values, and deleting duplicates.
- Feature selection: The most pertinent features that are likely to have an impact on customer attrition must be chosen for this module. Techniques like dimensionality reduction, feature importance, and correlation analysis can be used for this.
- Model training: Training machine learning models on the pre-processed and chosen features is the focus of this lesson. For training, a variety of methods can be utilized, including logistic regression, decision trees, random forests, and neural networks.
- Model evaluation: In this module, the effectiveness of the trained models is assessed using measures including accuracy, precision, recall, and F1-score. For choosing the model with the best performance, this module is essential.
- Deployment: In order to use the trained model for real-time churn prediction, this module entails deploying it into a production environment. This could entail creating a user interface for reporting and visualising the model within current systems.

By combining these modules, can be used to create a powerful customer churn prediction system that can aid banks in retaining clients, enhancing client satisfaction, and boosting income.

C. ***Methodologies:***

In banks, the widely used and successful random forest method is used to predict client attrition.

The steps in applying the random forest algorithm for churn prediction are as follows:
Data collection: Gather customer information on them, such as their transaction history, demographics, credit history, and other pertinent details.

- Data pre-processing: To make the data acquired appropriate for analysis, clean and convert it. This can entail codifying categorical variables, resolving missing values, and deleting duplicates.

- Feature selection: Choose the attributes most likely to have an impact on customer attrition. Techniques like dimensionality reduction, feature importance, and correlation analysis can be used for this.

- Split data set into training and testing sets: Create training and testing sets from the pre-processed data, using the training set for model training and the testing set for model evaluation.

- Model training: Using the training set of data, create a random forest model. Multiple decision trees are combined in the ensemble learning technique known as random forest to increase accuracy and decrease overfitting.

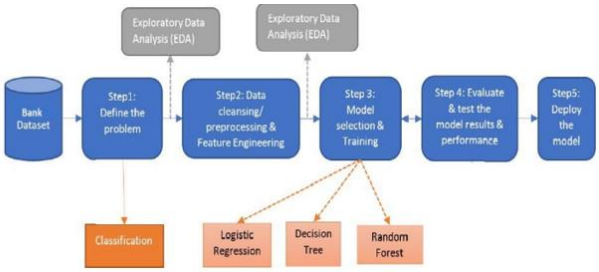
- Model evaluation: Analyze the trained model's performance using measures like accuracy, precision, recall, and F1-score. To choose the model with the best performance, you must complete this step.

- Deployment: In order to use the trained model for real-time churn prediction, deploy it into a production environment. This could entail creating a user interface for reporting and visualising the model within current systems.

The random forest algorithm can be used by banks to precisely identify which clients are most likely to leave and take preventative action to keep them.

The advantages of random forest are its high accuracy, scalability, and capacity for handling very big and highly dimensional datasets.

Flow Chart:



D. ***Metrics Used And Analysis:***

1. Precision: It is like mapping of programmer's model of problem, to computer model. It helps when cost of false positives is high. TP / (TP + FP)
2. Recall: Recall helps when the cost of false negatives is high.

   TP / (TP + FN)
3. F1 Score: F1 is an overall measure of a model's accuracy that combines precision and recall.

   2 * ((Precision * Recall) / (Precision + Recall))
4. Accuracy: accuracy can tell us immediately whether a model is being trained correctly and how it may perform generally.

(TP + TN) / (TP + TN + FP + FN)

Where,

TP = True Positive,

TN = True Negative,

FP = False Positive,

FN = False Negative

REFERENCES

[1] Predicting customer churn using targeted proactive retention by B.Mishachandar, Kakelli Anil Kumar

[2] Customer Retention in Banking Sector using Predictive Data Mining by Technique K. Chitra, B.Subashini

[3] Predicting credit card customer churn in banks using data mining by Dudyala Anil Kumar and V. Ravi

[4] Customer Profiling Methodology for Churn Prediction by John hadden, Dr. Ashutosh Tiwari and Prof. Rajkumar Roy

[5] Hierarchical Neural Regression Models for Customer Churn Prediction by GolshanMohammadi, RezaTavakkoli-Moghaddam, and Mehrdad Mohammadi

[6] A Survey on Customer Churn Prediction using Machine Learning Techniques by Saran Kumar A, Chandrakala D.

[7] Prediction of customer attrition of commercial banks based on SVM model by Benlan Hea,b,Yong Shic ,Qian Wand ,Xi Zhaoc

[8] Predicting customer churn in banking industry using neural networks by Alisa Bilal Zoric

[9] Customer Churn Analysis in Banking Sector Using Data Mining Techniques by A. O. Oyeniyi & A.B. Adeyemo

[10] A Hybrid Data Mining Method for Customer Churn Prediction by Elham Jamalian, Rahim Foukerdi

[11] Application of Improved PSO - BP Neural Network in Customer Churn Warning by Jiafa Huanga , Lili Heb

[12] Customer Churn Prediction Based on the RST and LS-SVM by Ning Wang, Dong-xiao Niu