CSE 4/587

Data Intensive Computing

Project report Phase – III

**Team:** Lavan Theja Jada (50478770) & Likhitha Kothrepally (50471606) – CSE 587

**Project Title:** Bank Customer Churn Prediction

Problem Statement

Customer churn is one of the most important metrics for a growing business to evaluate. A churn problem is a general way of describing customers switching to another or cancelling out. Churn is a significant concern for banks because it can lead to a decline in revenue, profitability, and market share. In the context of a bank, customer churn means customers of a bank stop using their bank services and move to another bank for various reasons.

For instance, A bank has noticed an increase in customer churn, which is causing a decrease in revenue and a negative impact on customer satisfaction. The bank wants to understand why customers are leaving and what can be done to reduce churn. The bank has gathered data on customer demographics, account information like estimated salary, credit score and bank balance. By observing the key factors that contribute to customer churn in the banking industry and we develop a predictive model to help the bank retain customers and reduce churn rate. The main objective of our project is to develop a model that predicts churn to identify at-risk customers and intervene before they leave the bank.

Potential of the Project

The goal of our project is to analyse the data and develop a predictive model to identify customers who are at risk of leaving, as well as to identify the key factors that contribute to churn. Based on the insights gained, the bank will take action to retain customers and reduce churn.

Specifically, the project will address the following questions:

• What are the characteristics of customers who are most likely to churn?

• What are the key drivers of churn?

• Can we develop a predictive model to identify customers at risk of churning?

• What actions can the bank take to reduce churn and retain customers?

The outcome of this project will be a report summarizing the findings, a predictive model for identifying at-risk customers, and recommendations for reducing churn and improving customer retention.

Data Sources

The dataset contains 10,000 customer records with various features related to bank customers, such as their credit score, age, tenure, balance, the number of products they have with the bank, and their estimated salary. The target variable is "Exited," which indicates whether a customer has churned (1) or not (0).

In this project, we will be analysing the above features to find out whether a customer has churned or not and address the customers who are at risk of leaving.

We have collected dataset from Kaggle:

<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset>

Model Implementation:

Based on the outcomes of the evaluation metrics, we developed various Machine Learning models in phase 2 to estimate each customer churn. We then evaluated the model’s efficiency for the prediction.

|  |  |
| --- | --- |
| Model | Accuracy |
| Random Forest Classifier | 90% |
| Logistic Regression | 69.8% |
| Support Vector Machine | 80.6% |
| XG Boost | 91.3% |
| Gradient Boosting | 88.8% |
| MLP Classifier | 86.6% |

The XG Boost model is the most effective model for our data, providing the best outcomes, as can be seen from the models implemented above, which demonstrate their correctness. As a result, we created the product and chose the XG Boost model as the final model for customer churn prediction.

Model Selection:

XGBoost is known for its high performance and scalability and has been used successfully in a variety of applications, including predictive modeling and classification tasks.

In bank customer churn dataset, XGBoost can be a good choice for several reasons. XGBoost can handle high-dimensional data effectively and can automatically learn feature interactions that may be difficult to identify using other methods.

Validation Score –



Confusion Matrix –

Chart, treemap chart

Description automatically generated

Accuracy and Classification Report –

A screenshot of a computer

Description automatically generated with medium confidence

The output after performing XGBoost shows that the model has an overall accuracy of 91% on the test set. The precision and recall values for both classes (0 and 1) are also quite high, indicating that the model is performing well in predicting both customers who will not churn and those who will.

The f1-score, which is a measure of the model's balance between precision and recall, is also high at 0.91 for both classes. This suggests that the model is performing well in both identifying true positives and avoiding false positives.

The output suggests that the XGBoost model is a good fit for the bank customer churn dataset, and that it can accurately predict customer churn with a high degree of precision and recall.

Web Application

Tech Stack used to develop this Application are:

a. Frontend: HTML, CSS

b. Backend: Flask (Python Framework)

c. Plotting Graphs: Tableau

d. ML Model: XG Boost

Working Instructions

Step 1:

Below is the Working Directory of the Web Application. Navigate to the project directory

where all the files related to the project are present in the system as shown in figure

below.

Step 2:

Execute model.py file as **python model.py**

Run app.py file as **python app.py** in the local host to start flask API

Step 3:

Navigate to <http://127.0.0.1:5000/> URL to see web page where we can see two ways of predicting customer churn.

**1)** **To predict a single customer churn at a time:** To predict a single customer churn we designed a form to enter details of a customer a shown below.

A screenshot of a computer

Description automatically generated with medium confidence

**2)** **To predict multiple customers, churn at a time:** To predict multiple customers churn we added an upload option to upload a csv file containing customers data as shown below.

A close-up of a bank

Description automatically generated with low confidence

Step 4:

**1)** **To predict a single customer churn at a time:** After giving details of customer, click on predict to display the output as shown below.

A screenshot of a computer

Description automatically generated with medium confidence

A picture containing text, screenshot, font, line

Description automatically generated

**2)** **To predict multiple customers, churn at a time:** After uploading csv file, click on predict to display output as shown below.

A close-up of a bank customer

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with low confidence