

AI Proposal

Predicting Customer Churn in Banking Using Machine Learning

Submitted By:

Alaa Mohamed Tharwat

Amal Maher Abd El-Razek

Rahaf Fayez Abdallah

Rewan Mohamed El-Malah

Shrouk Meme El-Zoghby

Submitted TO:

Dr.Sara El-sayed El-Metwally

Introduction

customer retention is a critical component of sustained growth and profitability. Customer churn, where clients discontinue their relationship with a bank, directly impacts revenue and operational costs. This project leverages machine learning (ML) techniques to predict customer churn using demographic, behavioral, and transactional data. By identifying at-risk customers, banks can deploy targeted strategies to enhance customer satisfaction and reduce attrition.

Problem Statement

Customer churn is a major challenge for banks due to:

- 1. Low Switching Costs: Customers can easily switch to competing banks offering better services or rates.**
- 2. Complexity of Churn Factors: The factors influencing churn are diverse and deeply rooted in customer behavior, making accurate prediction difficult.**

This project aims to address the challenge by:

- 1. Developing an ML model to predict customer churn based on historical data.**

2. Analyzing the key drivers of churn to inform retention strategies.

Goals

The objectives of this project include:

- **Build a Predictive Model:** Develop a machine learning model to classify customers as likely to churn or remain.
- **Feature Analysis:** Identify and interpret the most significant factors influencing churn.
- **Improve Retention Strategies:** Provide actionable insights to design effective customer retention plans.
- **Evaluate Model Performance:** Use metrics like accuracy, precision, recall, and F1 score to ensure the model's reliability.

Approach and Methodology

Dataset

The dataset used is from Kaggle, ([DataSet](#)) containing key features for churn prediction:

- **Features:**
 - **CreditScore:** Customer's credit score.
- **Geography:** Country of residence (France, Spain, Germany).

- **Gender:** Male/Female.
- **Age:** Customer's age.
- **Tenure:** Years with the bank.
- **Balance:** Account balance.
- **NumOfProducts:** Number of bank products used.
- **HasCrCard:** Ownership of a credit card (Yes/No).
- **IsActiveMember:** Bank activity status (Yes/No).
- **EstimatedSalary:** Customer's salary estimate.
- **Target Label:**
- **Exited:** Whether the customer exited (1) or stayed (0).

Steps

1. Data Preprocessing:

- **Handle missing values, normalize numerical features, and encode categorical variables like Geography and Gender.**

2. Model Training:

- **Train an XGBoost classifier, known for its robustness in structured data tasks.**

3. Hyperparameter Tuning:

- **Optimize model parameters using GridSearchCV to improve accuracy and generalization.**

4. Evaluation:

- **Use metrics like AUC, precision, recall, and F1 score to assess performance.**

Related Work

Several studies have explored churn prediction in banking and related sectors:

- 1. Predicting Customer Churn in Banking Industry Using Random Forest Algorithm (2020): Focused on demographic and behavioral data to predict churn.**
- 2. Customer Churn Prediction in Telecom Using Machine Learning (2019): Demonstrated the adaptability of decision trees and gradient boosting for churn prediction.**
- 3. Comparative Study on Customer Churn Prediction Models in Retail Banking (2018): Compared algorithms like logistic regression, SVM, and neural networks for churn prediction.**
- 4. Application of Deep Learning to Predict Customer Churn in Banking (2021): Highlighted the effectiveness of neural networks and LSTM for analyzing transaction sequences.**
- 5. Explaining and Predicting Customer Churn Using SHAP Values (2022): Focused on interpretability of ML models for actionable insights.**