Milestone 1: Customer Churn Prediction in banking

Customer Churn Prediction in banking

Aditya Sumbaraju

Bellevue University

DSC680: Applied Data Science

Dr. Brett Werner

https://github.com/adityasumbaraju/aditya_portfolio/tree/main/Customer_Churn_Prediction_in_banking

Milestone 1: Customer Churn Prediction in banking

Topic

The objective of this project, "Customer Churn Prediction in banking," is to reduce customer attrition by identifying potential churn candidates beforehand using a prediction model and taking proactive measures to make them stay. I would be considering the churn modeling dataset from Kaggle.

Business Problem

Customer churn exists across businesses in many sectors, especially since it significantly impacted banking. A company's growth depends on the high acquisition and low attrition rate. A high attrition rate represents a considerable investment loss, and both time and effort need to be channeled into replacing them. Predicting when a client is likely to leave and offering them incentives to stay can offer considerable savings to a business.

Predicting churn (attrition) is an essential factor for any current subscription-based business. A slight fluctuation in churn can significantly impact the bottom line of any business. Hence it is vital to know- "Is this customer going to leave us within X months?" Yes or No?

Research Questions:

- What is the percentage of customers churned and retained?
- Does the credit score impacts churn count?
- Is it possible to retain customers before they try to exit?

Milestone 1: Customer Churn Prediction in banking

Datasets

The dataset consists of 10000 observations and 12 variables.

- Independent variables contain information about customers.
- The dependent variable refers to customer abandonment status.

Variables:

- **RowNumber**: corresponds to the record (row) number and does not affect the output.
- **CustomerId**: Contains customer_ids of the bank.
- **Surname**: Surname is considered as PII (personally identifiable information) of the customer. Data needs to be Masked for ethical considerations or removed if there is no significance.
- **CreditScore**: This variable will significantly affect customer churn; the higher the credit score, the fewer chances to exit the bank.
- **Geography**: Location of Customer can be a potential churn factor.
- **Gender**: An interesting variable to explore and identify the churn factor.
- **Age**: Older customers are less likely to leave their bank than younger ones. It contains Age of the customer.
- **Tenure**: This variable signifies customer loyalty. Ideally, loyal customers are less likely to leave, and loyalty is gauged based on the term of stay.
- **Balance**: This variable provides a savings account balance. A vital indicator of customer churn is that customers with a lower balance in their accounts are more likely to exit the bank than those with higher balances.

Milestone 1: Customer Churn Prediction in banking

- **NumOfProducts**: Number of products that a customer has purchased through the bank during their tenure.
- **HasCrCard** : Boolean variable(0=No,1=Yes). It signifies whether or not a customer has a credit card. The hypothesis is people with a credit card are less likely to leave the bank.
- **IsActiveMember** : Boolean variable(0=No,1=Yes). Inactive customers are more likely to leave the bank. This is an important variable for our prediction use case.
- **EstimatedSalary**: Customers with lower salaries/balances are more likely to leave the bank than those with higher wages.
- **Exited**: Boolean variable(0=No,1=Yes) Does the customer leave the bank?

```
#information about the data
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   RowNumber        10000 non-null  int64  
 1   CustomerId       10000 non-null  int64  
 2   Surname          10000 non-null  object 
 3   CreditScore      10000 non-null  int64  
 4   Geography        10000 non-null  object 
 5   Gender           10000 non-null  object 
 6   Age              10000 non-null  int64  
 7   Tenure           10000 non-null  int64  
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64  
 10  HasCrCard        10000 non-null  int64  
 11  IsActiveMember   10000 non-null  int64  
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64  
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```

Dataset source:

https://www.kaggle.com/kmalit/bank-customer-churn-prediction/data

Milestone 1: Customer Churn Prediction in banking

Methods

I would use CRISP-DM to build a bank customer churn prediction model. Below are the phases I would be targeting to achieve a better model in this methodology.

- Data collection
- Data preparation and preprocessing
- Modeling and testing
- Model deployment and monitoring

Ethical Considerations

Data protection is the primary factor for banks to win customers' trust. The first is the need for "clear consent" to use personal data. I would emphasize the users being aware of how banks store customers' data, how they manage it, and their rights. Finally, proper data regulation needs to adhere while accessing customer PII. And companies must act with maximum diligence and responsibility to preserve security, privacy, and adequate use of data.

Challenges/Issues

1. Anticipating whitespaces in data and need to work on data alignment.
2. Incorrect variable types.
3. Python package:related issues.
4. I am relying on the Kaggle dataset as this is usually clean, and I anticipate no missing or insufficient data that needs to be substituted with dummy data.
5. Inaccurate or messy customer data

Milestone 1: Customer Churn Prediction in banking

References

Kaemingk, D. (2018, August 29). *Reducing customer churn for banks and financial institutions*. Qualtrics. Retrieved January 14, 2022, from https://www.qualtrics.com/blog/customer:churn:banking/

Tausend, F. (2018, September 19). *Eliminating churn is growth hacking 2.0*. Medium. Retrieved January 14, 2022, from https://blog.markgrowth.com/eliminating:churn:is:growth:hacking:2:0:47a380194a06

Tausend, F. (2019, January 15). *Hands:on: Predict customer churn*. Medium. Retrieved January 14, 2022, from https://towardsdatascience.com/hands:on:predict:customer:churn:5c2a42806266

Wuraolaifeoluwa, . (2021, May 21). *Prediction of customer churn in a bank using machine learning*. Medium. Retrieved January 14, 2022, from https://wuraolaifeoluwa.medium.com/prediction:of:customer:churn:in:a:bank:using:machine:learning:5a456c184ed1

Álvarez, C. (2018, September 25). *Big Data and privacy: New ethical challenges facing banks: BBVA*. NEWS BBVA. Retrieved January 14, 2022, from https://www.bbva.com/en/big:data:privacy:new:ethical:challenges:facing:banks/

Altexsoft, .. (2019, May). *Customer churn prediction using machine learning: Main approaches and models*. KDnuggets. Retrieved January 16, 2022, from https://www.kdnuggets.com/2019/05/churn-prediction-machine-learning.html