Predicting Churn Customers

FINAL PROJECT

ISM 6136 - Data Mining

Lavanya Chilakalapudi

Meenaa Raviselvam

Nidhi Harishchandra Falak

Sarang Sunil Patil

Zhaoqi Jin

Table of Contents

Dataset

The dataset encompasses a comprehensive set of columns providing valuable insights into credit card customers' profiles and behaviors.
Link: **Predicting Churn Customers**

Columns

S.No	Column Name	Description
1	CLIENTNUM	Client number. Unique identifier for the customer holding the account.
2	Attrition_Flag	If the account is active 0 if closed, then 1.
3	Customer_Age	Demographic variable - Customer's Age in Years.
4	Gender	Demographic variable - M=Male, F=Female.
5	Dependent_count	Demographic variable - Number of dependents.
6	Education_Level	Demographic variable - Educational Qualification of the account holder (example: high school).
7	Marital_Status	Demographic variable - Married, Single, Divorced, Unknown.
8	Income_Category	Demographic variable - Annual Income Category of the account holder.
9	Card_Category	Product Variable - Type of Card (Blue, Silver, Gold, Platinum).
10	Months_on_book	Period of relationship with bank.
13	Total_Relationship_Count	Total no. of products held by the customer.
14	Months_Inactive_12_mon	No. of months inactive in the last 12 months.
15	Contacts_Count_12_mon	No. of Contacts in the last 12 months.
16	Credit_Limit	Credit Limit on the Credit Card.
17	Total_Revolving_Bal	Total Revolving Balance on the Credit Card.
18	Avg_Open_To_Buy	Open to Buy Credit Line (Average of last 12 months).
19	Total_Amt_Chng_Q4_Q1	Open to Buy Credit Line (Average of last 12 months).
20	Total_Trans_Amt	Total Transaction Amount (Last 12 months).
21	Total_Trans_Ct	Total Transaction Count (Last 12 months).
22	Total_Ct_Chng_Q4_Q1	Change in Transaction Count (Q4 over Q1).
23	Avg_Utilization_Ratio	Average Card Utilization Ratio.

Dataset Analysis

- It contains 21 features or variables per customer that describe their account attributes and activity.
- There is a mix of numerical and categorical features:
    - Numerical - Age, Credit Limit, Total Revolving Balance, Months on Book etc.
    - Categorical - Gender, Education Level, Income Category, Card Category etc.
- The target variable **Attrition_Flag** indicates whether the customer is churned or not. This is what we want to predict.
- It has 10127 rows implying it covers 10127 customers.
- The data seems complete with no missing values.
- Features like Age, Income Category, Utilization Ratio, etc. are very predictive of churn.

Business Questions

How can we develop a predictive model to identify customers at risk of churning from credit card services, enabling proactive measures to enhance customer retention and satisfaction?

1. Customer retention is crucial for credit card companies - Acquiring new customers costs more than retaining existing ones. If the bank can predict customers likely to churn, it can target retention efforts toward those high-risk individuals.
2. Churn leads to loss of revenue - By identifying potential churners early, the bank can take proactive actions through better offers, improved customer service, etc. This helps retain customers longer, leading to more interest payments, annual fees, interchange fees, and other revenue.
3. Competitive advantage - Accurate churn prediction models allow banks to keep churn rates lower than competitors. This directly protects and expands market share.

Planned Analysis

In our project, we will leverage a multifaceted approach, encompassing an array of essential analytical tools and methodologies. Visual plots and graphs will be instrumental in revealing data patterns and trends, helping us understand and communicate key insights effectively. We will employ summary statistics to gain a quick overview of data distributions and central tendencies, guiding our feature selection and analysis. Examining correlations between features will unveil intricate relationships within our data. Random Forest, a robust ensemble machine learning technique, will enable us to build predictive models, assess feature importance, and deliver accurate customer churn predictions. Moreover, we remain open to the incorporation of additional analytical tools and techniques, ensuring a comprehensive and data-driven approach throughout our project.