

Predict Credit Card Customer Attrition

Problem Statement: Attrition of credit cards has a negative influence on the card's firm P&L, resulting in a loss of sales and revenue as well as asset degradation. Account attrition can be caused by a variety of causes, such as unhappiness with the card's value proposition not fulfilling the customer's expectations, and so on. Leveraging the more significant of these factors. This helps in building strategies for countering account attrition

The task: The objective of this task is to determine the probability of attrition of an account holder within the next six months.

Objective: In recent months, the business manager of a bank's consumer credit card portfolio has faced a serious client attrition problem. This has a negative influence on the business. The company manager wants to use data analytics to figure out what the main causes of attrition are. She also wants to have an ability to understand the customers who are likely to close their accounts with the bank in near future, so that she can focus her efforts, well in advance to retain those customers.

Dataset: The dataset is shown below

```
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CLIENTNUM                             10127 non-null  int64
1   Attrition_Flag                         10127 non-null  int64
2   Customer_Age                           10113 non-null  float64
3   Gender                                 10127 non-null  object
4   Dependent_count                         10127 non-null  int64
5   Education_Level                         10127 non-null  object
6   Marital_Status                         10127 non-null  object
7   Income_Category                         10127 non-null  object
8   Card_Category                           10127 non-null  object
9   Months_on_book                         10127 non-null  int64
10  Total_Relationship_Count                10127 non-null  int64
11  Months_Inactive_12_mon                  10127 non-null  int64
12  Contacts_Count_12_mon                   10127 non-null  int64
13  Credit_Limit                           10123 non-null  float64
14  Total_Revolving_Bal                     10127 non-null  int64
15  Trans_Amt_Oct12                         10127 non-null  int64
16  Trans_Amt_Nov12                         10127 non-null  int64
17  Trans_Amt_Dec12                         10127 non-null  int64
18  Trans_Amt_Jan13                         10127 non-null  int64
19  Trans_Amt_Feb13                         10127 non-null  int64
20  Trans_Amt_Mar13                         10127 non-null  int64
21  Trans_Count_Oct12                       10127 non-null  int64
22  Trans_Count_Nov12                       10127 non-null  int64
23  Trans_Count_Dec12                       10127 non-null  int64
24  Trans_Count_Jan13                       10127 non-null  int64
25  Trans_Count_Feb13                       10127 non-null  int64
26  Trans_Count_Mar13                       10127 non-null  int64
dtypes: float64(2), int64(20), object(5)
memory usage: 2.1+ MB
```

The data is a sample of credit card customer accounts as of Mar-2013 and the attrition has been defined as customers closing their credit card account within the next 6 months (Apr-2013 - Oct-2013). The sample has a mix of snapshot data as of Mar-2013 and historical data covering the past 12 months prior to Mar-2013, including their demographic profile and transaction information. The Dataset is shown below, it has 27 columns and the total size is 2.1 KB, It contains some missing values in the columns which are Customer_Age and Credit_Limit.

Model: Logistic Regression classifier is used to build the model. It is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression is estimating the parameters of a logistic model (a form of binary regression).

Params:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=100,
                    multi_class='warn', n_jobs=None, penalty='l2',
                    random_state=None, solver='warn', tol=0.0001, verbose=0,
                    warm_start=False)
```

Results: 91% Precision score and f1score is achieved on test data.