CUSTOMER CHURN PREDICTION MODEL USING DISCRIMINANT ANALYSIS

Team Members:

(1) Arjun Prabhu

(2) Farheen Shaik

(3) Harshil Chauhan

(4) Lakshmi Vadlamudi

(5) Mukesh Chowdary

Project Summary

Section 1: Problem Overview and Motivation

Abstract: -

The Problem is based on the domain of the Banking sector where the bank wants to predict the Churn of a customer depending upon the previous data of the customer. By churn it is meant that the bank wants to predict if a customer would retain or leave the bank next quarter depending upon their bank balance. It is important from a bank's perspective in order to maintain business and customer relationship Apart from that a bank can predict risk of losing the customer so then primitive measures can be taken to ensure that such conditions do not erupt.

Introduction

Today, every company wants to hear what their customers have to say. According to McKinsey's research, an unhappy customer will tell 9-15 people about their bad experience. Given the number of unhappy consumers you may have, that's already a lot of negative press. This will have a direct impact on your company's revenue and reputation. When your customers are happy, they trust the brand and become loyal to it. Customers that are loyal to a brand are more likely to buy from them again, and they account for a significant portion of the revenue. If this is not the case, no number of marketing efforts or promotions will be able to save your company if your customers are unhappy. Customers who are dissatisfied with a brand are more likely to abandon it in the future. Customer churn is the loss of a customer from a service or program as a result of this intern.

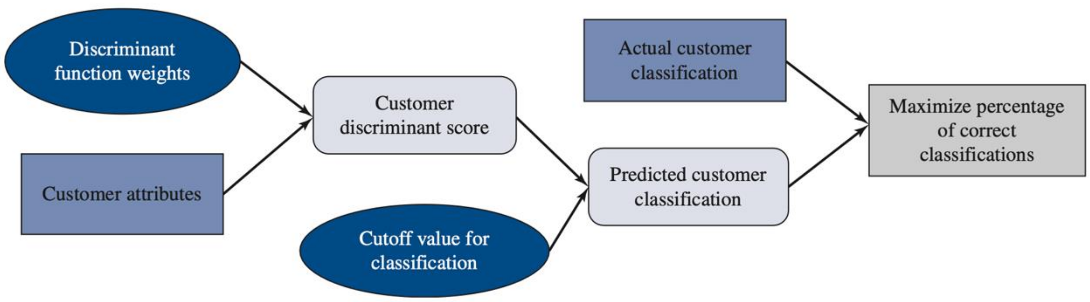
The expense of acquiring new customers is 6-7 times higher than the cost of keeping old ones. Rather of investing a large sum of money on obtaining new consumers, it is better to spend a little portion of it on enhancing your current processes and systems in order to keep existing customers. And it is here that we are constructing a predictive model for a bank to identify customers at risk of churn so that proactive reengagement programs may be implemented.

What is the problem?

The main problem is to predict if a customer would be leaving the bank or will be retained loyal to the bank.

Section 2: The Optimization Model

The customer churn prediction model is used to predict the customer retaining with bank or not using discriminant analysis model. Analysts in marketing and other disciplines of business utilize discriminant analysis as a statistical tool. Although it is comparable to cluster analysis, it is not the same. There are no specified clusters in cluster analysis.



- objective function: To develop Customer churn prediction model for loyalty programs and retention campaigns to retain as many customers as possible.
- decision variables: Weights used to form discriminant scores and the cutoff value for classification
- constraints:
  cutoff: To obtain the lower and upper limits of cutoff value we consider the maximum value (199992) and the lowest value of estimated salary (-199992)
  Weights: The weights are constrained to 1 and -1.
- model inputs: Credit score, country, gender, age, tenure, balance, products number, credit card, active member, estimated salary.

Section 3: Model Input Parameter Estimation and Data Requirement

Getting a new client is far more expensive than keeping an existing one. It is advantageous for banks to understand what causes a client to leave the organization. Companies can build loyalty programs and retention efforts to keep as many customers as possible by preventing churn.
In this case, we leverage customer data from a bank to build a predictive model for clients who are likely to churn.

The data set contains information for creating our model. We need to configure three things here:

1. Data source.
2. Variables.
3. Instances.

The data file contains 12 customer attributes for about 10,000 customers in a bank.

The variables are:

Customer_id, unused variable.

Credit_score, used as input.

Country, used as input.

Gender, used as input.

Age, used as input.

Tenure, used as input.

Balance, used as input.

Product number, used as input.

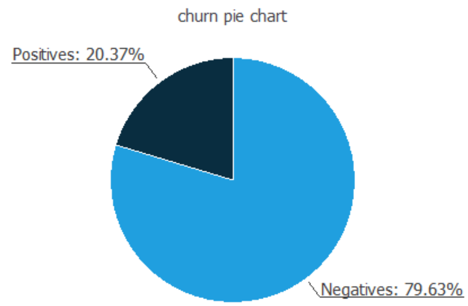
Credit card, used as input.

Active member, used as input.

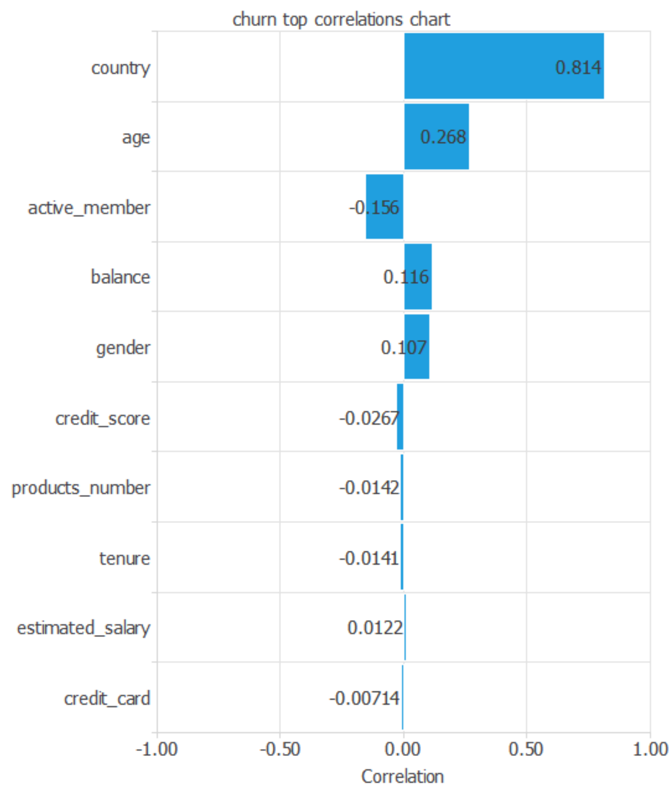
estimated salary, used as input.

churn, used as the target. 1 if the client has left the bank during some period or 0 if he/she has not.

The data distributions tell us the percentages of churn and loyal customers.



The inputs-targets correlations might indicate which variables might be causing attrition.



From the above chart, we can see that the country has a significant influence and that older customers have more probability of leaving the bank.

To solve the prediction model using Solver in Excel we used Evolutionary Solver.

Evolutionary Solver use genetic algorithms to develop good (near-optimal) solutions to more challenging situations, such as those in which the objective cell and/or restrictions are "non-smooth" functions of the changing cells.

Functions that we used to compute the data model are as follows:
Objective: To develop Customer churn prediction model for loyalty programs and retention campaigns to retain as many customers as possible.

Decision Variables: Weights, Cutoff

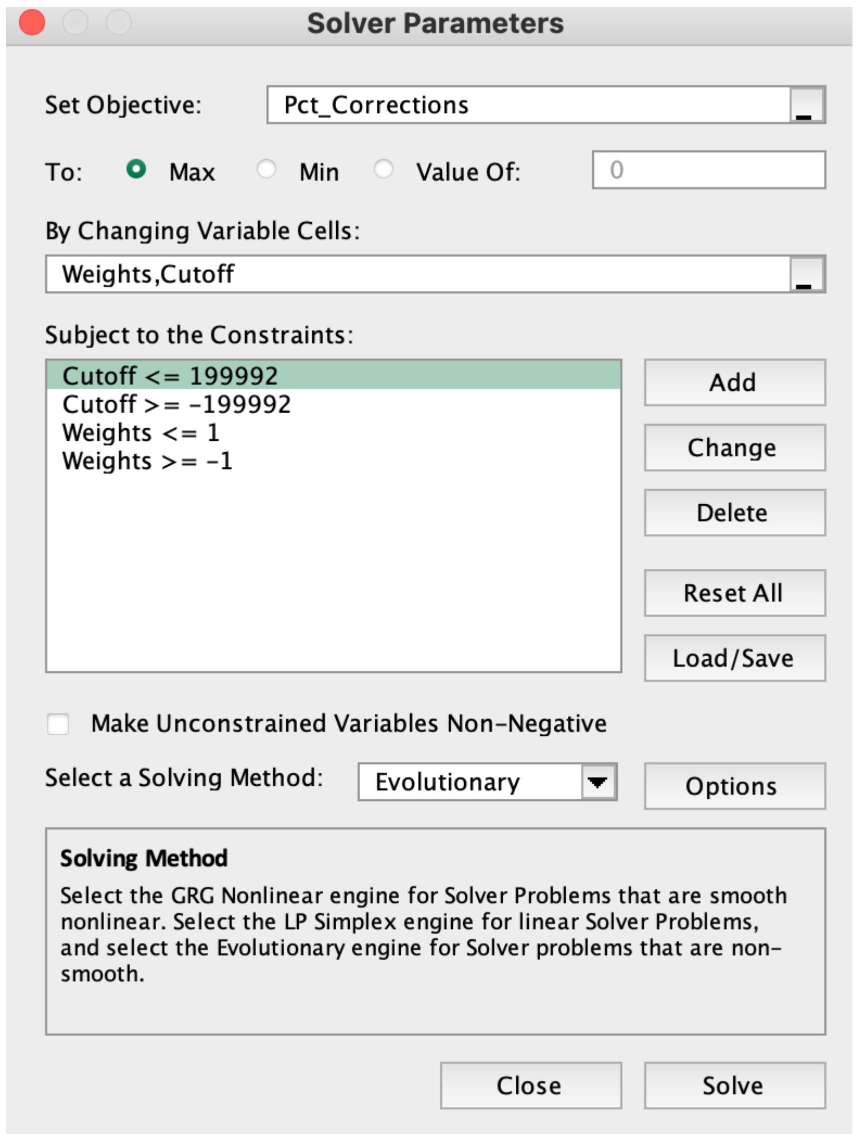
Constraints:

Cutoff <= 199992

Cutoff >= -199992

Weights <= 1

Weights >= -1



Model Inputs: Credit score, country, gender, age, tenure, balance, products number, credit card, active member, estimated salary.
We assumed an external weight factor to estimate the prediction model and the attributes used for the same are:

(1) Credit Score
(2) Country
(3) Gender
(4) Age
(5) Tenure
(6) Balance
(7) Products number
(8) Credit Card
(9) Active Member
(10) Estimated Salary

Objective parameters were set as Cutoff value for classification and Percentage Correct Classification.

An external estimation parameter for Salary was suggested as described to limit the data to an average of customers salary.

Min of Estimated Salary	11.58
Max of Estimated Salary	199992.48

An estimation of the data real/assumed is used to predict the customer churn model using classification matrix.

Section 4: The Optimization Solution Method

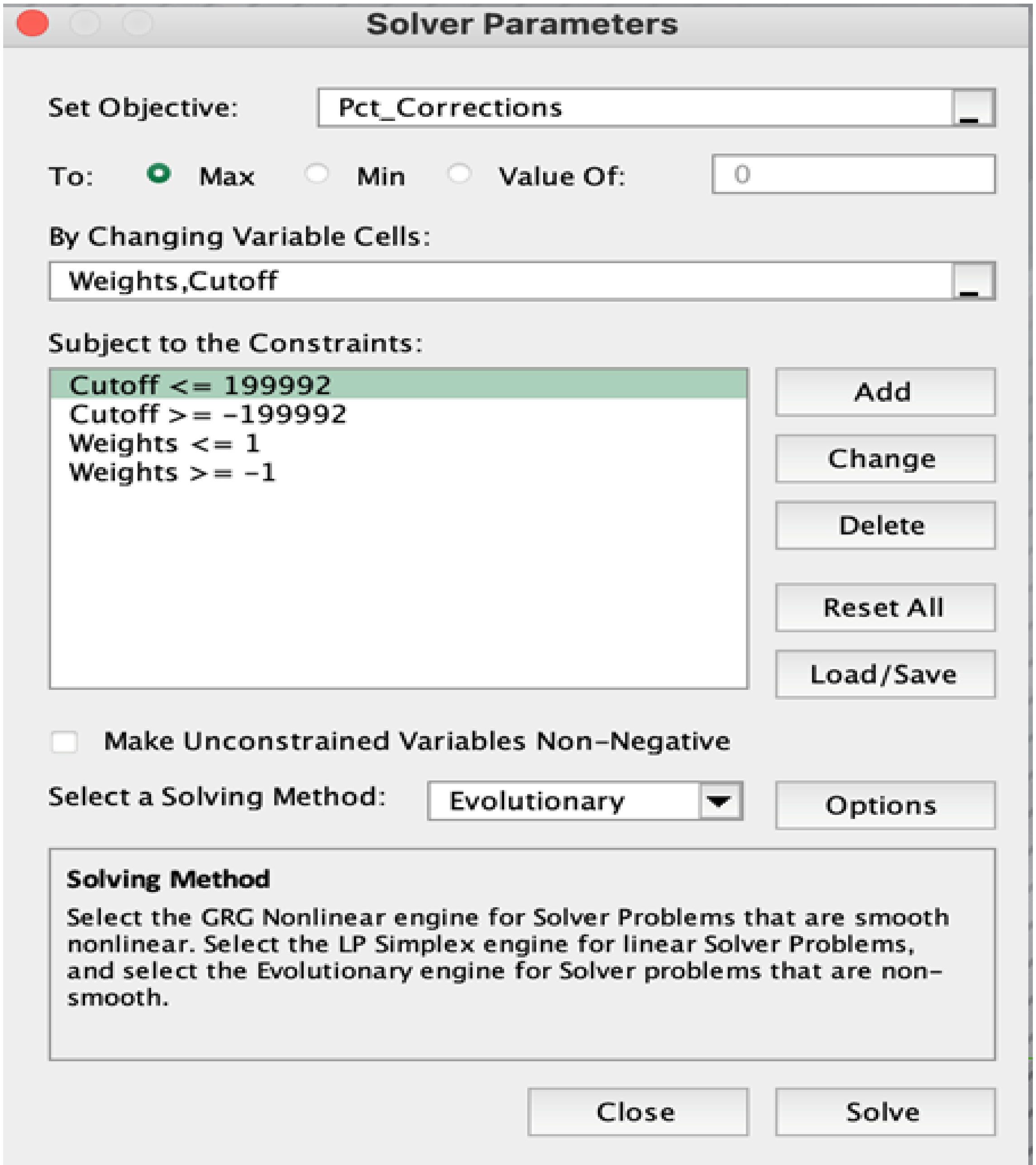
For this optimization model, we implemented the Evolutionary solver method. Typically, evolutionary algorithms are applied to provide acceptable approximation solutions to problems that are hard to address using other techniques. Obtaining a precise answer may be too computationally expensive, but sometimes a near-optimal solution can suffice. Evolutionary algorithms are never guaranteed to find an optimal solution for the problems because to their random nature, but they will usually find a good solution if one exists.

Genetic algorithms are being used by Evolutionary Solver to develop good (near-optimal) solutions to particularly tough circumstances, such as when the objective cell and/or limitations are "non-smooth" functions of the changing cells. After Evolutionary Solver has obtained a good solution for smooth optimization models, GRG Nonlinear Solver can be used to try to find a slightly better solution.

<u>Steps to implement Evolutionary solver</u>

We first specify the Objective of this optimization problem, which is the percentage correction for the prediction or the accuracy of the classification matrix. The decision variables or the changing cells are selected; that is the weights and the cutoff score. Then the constraints, which are the range is entered for the cutoff score and the weights. The Evolutionary solver method is then selected and the solver is run for the customer churn prediction.

We have a few solving method options which are convergence and Mutation rate.
We often explicitly add upper bound and lower bound limitations on modifying cells while using Evolutionary Solver.


Solver Parameters
Set Objective:
Pct_Corrections
To:
Max
Min
Value Of:
0
By Changing Variable Cells:
Weights,Cutoff
Subject to the Constraints:
Cutoff <= 199992
Cutoff >= -199992
Weights <= 1
Weights >= -1
Add
Change
Delete
Reset All
Load/Save
Make Unconstrained Variables Non-Negative
Select a Solving Method:
Evolutionary
Options
Solving Method
Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.
Close
Solve

Section 5: Optimal Solution Structure and Insights

The Solution Shown in the above Pic is certainly not unique. Many other sets of weights and cutoff Values can obtain 79.71% correct classification rate, and we will probably obtain different solution from this solution. Also, you can see from the weights that the classification is based more heavily on Balance, with less weight placed on number of Products. Because of the positive weight on the balance, people with More Balance tend to be classified as existing Loyal Customers. Unfortunately, there is no reason to believe that these will work as well for another group of people.

Section 6: Conclusions, Model Limitations and Extensions

We have taken data consisting of 10,000 customers. Excel sheet contains the classification matrix where 1 is represented as customers who have left the bank during a certain period and 0 represents that customer have never left the bank. From the matrix, it is stated that 7 customers have left the bank at some point but have been classified as existing loyal customers. 2022 customers are loyal customers but have been classified as they have left the bank at a point. 15 customers have left the bank at a point and have been classified the same. Finally, 7956 customers have never left the bank and are classified the same. Hence, after running the solver, nearly 80% of the data was correctly classified.

References

[1] https://www.neuraldesigner.com/learning/examples/bank-marketing-campaign
[2] https://www.neuraldesigner.com/learning/examples/credit-risk-management
[3] https://www.neuraldesigner.com/solutions/churn-prevention
[4] https://www.neuraldesigner.com/solutions/customer-segmentation