Predicting Credit Union Customer Churn Behavior Using Decision Trees, Logistic Regression, and Random Forest Models.

by

Frederick Barr

A Capstone Project Submitted to the Faculty of

Utica College

May 2020

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in
Data Science

Abstract

Traditionally, U.S. Credit Unions served small and homogeneous populations. Managed by their members, Credit Unions have been granted tax exemption based on this cooperative structure. Today, Credit Union membership is expanding, increasingly heterogeneous and geographically dispersed.

Because commercial banks are corporate-owned, it is difficult for Credit Unions to compete due to the vast resources banks possess. The most crucial offering a Credit Union can provide is excellent customer service and the ability to provide financial assistance to its members.

Given the importance of customers as the most valuable assets of organizations, customer retention seems to be an essential requirement for any organization. The competitive atmosphere within which banks provide electronic banking services increases the necessity of customer retention.

The data used for this study include membership activity from January 2013 through July of 2019. It represents the counts of product purchases, such as loans, credit cards, and the number of banking transactions. The use of data from a prior year, rather than current information, better reflects the banking industry before the adverse effect of the COVID 19 pandemic.

According to the Chief Financial Officer of the Credit Union, providing the data, every member enables approximately $240.00 of average marginal income. This relatively small amount represents considerable income when multiplied by total membership. This paper examines three machine language models that can identify customer loss and provide the credit union's marketing team the ability to reach out and retain the member.

This paper intends to make use of performance comparisons to indicate the accuracy and percentage of probability that a customer wants to leave the Credit Union. A comparison of four statistical indicators, accuracy, classification error, root mean square error, and kappa value demonstrates that though all four models performed well, Random Forest provides the best overall performance. It provided the highest accuracy and the lowest error of all the models tested.

Keywords: Data Science-Business Analytics, Dr. Michael McCarthy, random forest, logistic regression, decision trees, churning, credit union, customer loyalty, customer retention.

Table of Contents

List of Illustrative Material

Introduction

Background of Study

Credit Unions are particularly sensitive to churning; a member stops doing business with the company, which results in a high amount of financial revenue spent by marketing to address this problem. Methods such as mass marketing and ad-hoc promotions attempt to maintain and strengthen customer retention.

Banking customers usually give their business to the same institution and tend to remain for a long time (Shevlin, 2019). A JD Power's U.S. Retail Banking Study indicated that only four percent of customers switched primary banks in 2018. (2019). Therefore, customer relationship management (CRM) concentrates on customer loyalty to attempt to retain as many as possible. Computer-based information systems provide CRM with both quantitative and qualitative aspects of membership behavior. Products and services used by the customer, provide the data that reflects customer behavior. This behavioral data acts as a potential source to predict churning (Chirica, 2013).

Marketing research literature indicates that "customer churn" is a term used in the banking service industry to show the customer movement from one provider to another.

Churn management is the term that describes an operator's process to retain profitable customers (Kumar & Kumar, 2019). Churn is also called "attrition" and often used to indicate a customer leaving the service of one company in favor of another company.

Selective and personalized marketing practices are starting to replace customer relationship management systems. Targeted and customized marketing methodology requires the identification of customers that are likely to stop using banking products or services and provide specific marketing or incentives for the customer to retain. The loss of banking customers generally results in a profit decline for the credit union in this study; the damage is about $240.00 per customer.

One method to control churning requires building useful and accurate prediction models. From a business perspective, two major analytical tasks present themselves. The first task is predicting those customers who are about to churn (i.e., leave the Credit Union).

The second task is assessing the most effective way that marketing can reach out to this customer and provide alternative methods to prevent customer loss.

In personal retail banking, a Credit Union must operate on a long-term customer strategy. As the customer relationships last a considerable duration, maybe decades, the company must address the value of a potential loss of a customer. The customer lifetime value analysis will help to overcome this challenge. This type of customer analysis requires the assessment of customer value, customer lifetime value, customer equity, and customer profitability. The problem in this concept is to define and measure the customer lifetime value during, or even before, the active stage of the customer relationship. The focus on customer churn determines the customers that are at risk of leaving.

The purpose of this research consists of examining several classification models that predict if a customer will churn by examining the model's accuracy,classification error, root mean square error and kappa value. The data consists of 30,000 observations and 26 attributes. The information belongs to this author's current employer, a mid-size Credit Union, established in 1954, and its current net worth is 1.8 billion dollars