Understanding, predicting and preventing Churn

William RJ Cooper

August 20, 2024

1 Executive Summary

In this report we outline a model that can predict whether a customer will leave a business or not (churn) with around 86% accuracy, using a few pieces of information readily available to most banks. Preventing churn is vital in most businesses, and predicting it is the first step. This report outlines a model that:

- can predict customer churn can be predicted with 86% accuracy
- can output a percentage indicating how likely the customer is to leave or stay
- has a low false negative rate (11%) meaning it is unlikely for the model to predict that a churning customer will stay

This model for predicting customer churn and assigning a confidence value can be used to offer pre-made tiered packages to customers, with scaling value for money meaning that customers most likely to churn can be offered the most lucrative discounts or products to keep them from churning.

2 Introduction

Churn, customer attrition, customer retention, all refer to the loss of customers from a business. Henceforth referred to as simply churn, minimising the loss of customers from any business is beneficial to any business for several reasons:

- It can cost 5 times more to acquire and sign a new customer than retain an existing one
- According to a report by Harvard Business School, 'increasing customer retention rates by 5% increases profits by 25% to 95%'
- 80% of your future profits will come from just 20% of your existing customers (Leading on the Edge of Chaos: The 10 Critical Elements for Success in Volatile Times by Emmett C. Murphy, Mark A. Murphy, 978-0735203129)

Any scheme that seeks to understand, predict and prevent customer churn may be able to increase profits, reduce costs and increase customer satisfaction.

Businesses often keep large amounts of data about their customers, and whether or not a customer has churned would also be easy to calculate. It is for this reason that this problem is perfect for data-driven solution as if there is a link between customer data and churn, these methods can find it.

In this research we choose two datasets containing various information about customers including whether or not they have churned. Even though these datasets are from bank customers and I am not in banking myself, there is no reason why the approach we use here could not be used for other datasets from businesses in other areas. This is why I chose this project, because customer churn affects almost every business.

3 Methods

4 Results

5 Conclusion

6 Evaluation