Customer Churn Analysis

A data-driven approach to proactively managing customer churn

# Executive Summary:

The ability to retain a customer in the service industry is a top priority for many organizations as it serves as a reoccurring/compounding revenue stream – the longer a customer is retained, the more profit is earned. Studies also show that the cost to acquire a new customer is anywhere from 5 to 25 times greater than the costs of retaining current customers (HBR, 2014). However, the traditional churn measures are lagging in nature – that is to say the customer has already made the decision to leave before the company knows they are severing ties. Most organizations opt to take a proactive approach toward reducing customer churn, and build teams of customer advocates whose main goal is to forge relationships with their customers and, hopefully, figure out customer’s needs and pain points before they close the door. However, this becomes challenging as the business begins to scale and each advocate’s book of business increases. Wading through the sea of customers in search of the ones who are in need of help can become downright impossible, and the need for effective customer management becomes clear.

In this study we take a data-driven approach to proactively managing customer churn. We use a real-life data set from the financial industry, customer data from a bank, and model the outcome of if the customer left the bank or stayed (in the specific period of time). The data set was downloaded from Kaggle.com (located [here](https://www.kaggle.com/shrutimechlearn/churn-modelling)), which is a popular data science platform. The outcome response is the column labeled “Exited”, and a value of 0 indicates the client stayed with the bank, where a value of 1 indicates they left.

When choosing the methods to apply to this classification problem, we make two arguments: first, we are ok with over identifying customers as “likely to churn” since more touch points with customers is never a bad thing (however, we hope to filter it down to something manageable by a customer advocate department). Secondly we don’t make “understanding the inner workings” of the model (interpretability) a top priority, which allows us a wider selection of methods to choose from - this study is strictly focused on finding an efficient way to identify customer churn from a large list of customers. Therefore, we opt for the powerful neural network as our leading candidate model since it works well with larger datasets and are more flexible to nonlinear problems. Secondly, we choose the logistic regression since there appears to be a lot of overlap between the outcome classes for many of the predictor variables (which is described in further detail within this project).

References:

Amy Gallo. Harvard Business Review. (2014, October). *The Value of Keeping the Right Customers.* <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>