DATA 621: Final Project Proposal

Calvin Wong, Juanelle Marks, Kevin Benson, Ravi Itwaru, Sudhan Maharjan

11/20/2019

Introduction

We will be using the Telco Customer Churn dataset to study the customer behavior in order to develop focused customer retention programs. Customer churn occurs when customers or subscribers stop doing business with a company or service, also known as customer attrition. It is also referred as loss of clients or customers. One industry in which churn rates are particularly useful is the telecommunications industry, because most customers have multiple options from which to choose within a geographic location. There are times when a customer churns without actually intending to stop using the service. This is termed as Involuntary Churn. We will attempt to diagnose customer attrition and identify if voluntary or involuntary factors are in play.

Research Question

In this project we will attempt to answer the following research questions:

• Develop a regression model to predict customer churn by assessing their propensity of risk to churn

• Identify the factors of churn and identify if they are in fact voluntary or involuntary

Data Source

The dataset is sourced from Kaggle (<https://www.kaggle.com/blastchar/telco-customer-churn/>), and has been downloaded and saved as a raw CSV file at:

https://raw.githubusercontent.com/cwong79/DATA621/Calvin/Final%20Project/WA\_Fn-UseC\_-Telco-Customer-Churn.csv

This dataset includes 7,043 cases of 21 variables, which describes: -

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

Observations: 7043

Variables: 21

$ customerID : chr "7590-VHVEG" "5575-GNVDE" "3668-QPYBK" "7795-CFOCW" ...

$ gender : chr "Female" "Male" "Male" "Male" ...

$ SeniorCitizen : num 0 0 0 0 0 0 0 0 0 0 ...

$ Partner : chr "Yes" "No" "No" "No" ...

$ Dependents : chr "No" "No" "No" "No" ...

$ tenure : num 1 34 2 45 2 8 22 10 28 62 ...

$ PhoneService : chr "No" "Yes" "Yes" "No" ...

$ MultipleLines : chr "No phone service" "No" "No" "No phone service" ...

$ InternetService : chr "DSL" "DSL" "DSL" "DSL" ...

$ OnlineSecurity : chr "No" "Yes" "Yes" "Yes" ...

$ OnlineBackup : chr "Yes" "No" "Yes" "No" ...

$ DeviceProtection: chr "No" "Yes" "No" "Yes" ...

$ TechSupport : chr "No" "No" "No" "Yes" ...

$ StreamingTV : chr "No" "No" "No" "No" ...

$ StreamingMovies : chr "No" "No" "No" "No" ...

$ Contract : chr "Month-to-month" "One year" "Month-to-month" "One year" ...

$ PaperlessBilling: chr "Yes" "No" "Yes" "No" ...

$ PaymentMethod : chr "Electronic check" "Mailed check" "Mailed check" "Bank transfer (automatic)" ...

$ MonthlyCharges : num 29.9 57 53.9 42.3 70.7 ...

$ TotalCharges : num 29.9 1889.5 108.2 1840.8 151.7 ...

$ Churn : chr "No" "No" "Yes" "No" ...

For purpose of this project, our target variable will be derived from “Churn”, which identifies if that particular customer left Telco.

Methodology

In this project we will develop a regression model to predict whether a customer will leave Telco based on services utilized, account information and demographic. The regression model will most likely be a logistic regression, in which the binary response variable is:

• 1: Churn == “Yes”

• 0: otherwise

Potential predictor variables include various information about client utilization and demographic, including age and gender.

Once we develop the logistic model, we can use this to predict the probability of an individual discontinuing Telco’s services. By comparing the predicted probabilities to the actual values of the target variable, we can identify cases where the probability is high (>0.5) but where that individual is still a Telco user. The idea is these users have a high flight risk and methods of customer retention should be applied to keep them.