September 2021

Telecom Customer Churn

Noah Scanlon

**Business Problem**

Customers are the lifeblood of businesses and retaining them over time is one of businesses’ greatest challenges. The negative of this is referred to as churn: The phenomenon when a customer becomes a previous customer (i.e. Mom cancels the Disney+ subscription because she didn’t like Raya and the Last Dragon).

Datapoints on customers, such as a customer’s service subscriptions, length of being a customer, payment method, demographic information, etc. help to predict the likelihood of a customer churning. Knowing whether and who is likely to churn gives the business a chance to strategize and act on retention, bolstering their bottom line.

Beyond just determining the likelihood of a given customer churning, exploring helps explain what forces are driving churn.

**Data Overview**

IBM Cognos Analytics creates sample datasets to use so to prove their products, this is one of them. To quote them:

*“The* ***Telco customer churn*** *data contains information about a fictional telco company that provided home phone and Internet services to 7043 customers in California in Q3. It indicates which customers have left, stayed, or signed up for their service. Multiple important demographics are included for each customer, as well as a Satisfaction Score, Churn Score, and Customer Lifetime Value (CLTV) index.”*

You can view IBM’ posting and full description of the data here: [Telco customer churn (11.1.3+)](https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113)

Though, the actual dataset I used was taken from here (same set, but usable): [Kaggle](https://www.kaggle.com/blastchar/telco-customer-churn)

**Anticipated Data Science Approach**

**Modeling Algorithm:** Logistic Regression (i.e. customer 1114 is 0.6% likely to churn this month)

**Evaluation Metric:** Area Under the Receiver Operating Curve (AUROC)

**Flow:**

* Import data
* Explore data and visualize relationships between features
* Feature-engineer any data that may be useful
* Use logistic regression as a model to predict churn likelihood
* Use K-means to segment customers based on churn and other value metrics

**Deliverables**

* Jupyter Notebook containing all code (.ipynb)
* Datasets used (.csv, .tsv, etc.)
* Final report describing the project (.docx and .pdf)
* Slide deck presentation describing the project (.ppt)
* All above published on GitHub (@ [noahscantron](https://github.com/noahscantron))