Project Proposal

October 12th, 2018

Regarding:

Predicting Customer Churn

Springboard Course Work
Capstone Project #1

Cliff Robbins
cliff@gearforgesoftware.com
M 612.701.2998

Foreword

The goal of the project is to demonstrate data science skills on a real customer problem leveraging the full suite of data science tools and skills.

As a data scientist, I will be working on various datasets that pre-exist in various forms. Part of this project will be a determination of the dataset, cleanup and experimentation to prove or disprove my hypothesis.

I will follow the scientific method during this project and use the scientific method as a guide. The steps I will be using are:

1. Formulation of a Question
2. Hypothesis
3. Prediction
4. Testing
5. Analysis

To stay true to this form, I will ensure that my finale outcome will include the capability for outside investigation, experimentation and validation.

Proposal

My project will focus on a problem that 28 million business face each day of operation, customer churn.

Definition

Customer churn, also known as customer attrition, customer turnover or customer defection is the loss of clients or customers. Many companies include customer churn rate as part of their monitoring metrics because the cost of retaining current customers compared to acquiring new customers is much less.
Within customer churn there is the concept of voluntary and involuntary churn with voluntary being a customer leaves on their own choice while involuntary could be attributed to customer relocation to a long term care facility, death or customer relocation in a different state/geography. In most analytical models, involuntary churn is excluded from the metric.

Formulation of a Question

When a company first starts up, the founding members can typically handle all of the various customer concerns. As the company continues to grow, the founders can no longer service all of the various clients with support handled by a customer service team. The customer service team focuses on current issues and a proactive approach is lost.
As the company grows, the company still cares about its clients; however, due to the large customer base they can no longer address each and every customer. This is a real problem for companies. How does a company proactively predict if a customer is happy or unhappy? How does a company know if a customer is so unhappy that they are willing to leave? If a company knew if a customer was getting ready to leave, could they reach out to the customer and mend the relationship?

Hypothesis

I believe past customer data can predict future customer churn.

Prediction

If I had past customer data that showed various features and whether they stayed or churned we could use that data to predict future outcomes of current customers.

Testing

To test my hypothesis, I will use a set of customer data with various features along with whether they churned or not.
The data has 7043 rows and can be found at:
https://www.kaggle.com/blastchar/telco-customer-churn
The dataset has the following features:

- customerID - Customer ID
- gender - Customer gender (female, male)
- SeniorCitizen - Whether the customer is a senior citizen or not (1, 0)
- Partner - Whether the customer has a partner or not (Yes, No)
- Dependents - Whether the customer has dependents or not (Yes, No)
- tenure - Number of months the customer has stayed with the company
- PhoneService - Whether the customer has a phone service or not (Yes, No)
- MultipleLines - Whether the customer has multiple lines or not (Yes, No, No phone service)
- InternetService - Customer's internet service provider (DSL, Fiber optic, No)

- OnlineSecurity - Whether the customer has online security or not (Yes, No, No internet service)
- OnlineBackup - Whether the customer has online backup or not (Yes, No, No internet service)
- DeviceProtection - Whether the customer has device protection or not (Yes, No, No internet service)
- TechSupport - Whether the customer has tech support or not (Yes, No, No internet service)
- StreamingTV - Whether the customer has streaming TV or not (Yes, No, No internet service)
- StreamingMovies - Whether the customer has streaming movies or not (Yes, No, No internet service)
- Contract - The contract term of the customer (Month-to-month, One year, Two year)
- PaperlessBilling - Whether the customer has paperless billing or not (Yes, No)
- PaymentMethod - The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- MonthlyCharges - The amount charged to the customer monthly
- TotalCharges - The total amount charged to the customer

The following target will be used to understand if the customer churned or not.

- Churn - Whether the customer churned or not (Yes or No)

Analysis

To determine if we can predict the churn rate, I will use a logistic regression algorithm which is a classification algorithm focused on a binary outcome. To ensure accuracy, I will conduct cross-validation and grid search on the model then tune the model based on the best outcome. This will ensure we minimize miss-classification.

I will split the data 80/20 into a training and test set of data to validate the model.

I will also test the various features to ensure they are all necessary to produce the best model.

Deliverables

To ensure repeatability, I will conduct my testing using a Jupyter Notebook which I'll check into a public repo for others to validate my findings. I will also provide a

presentation that outlines the findings of this problem along with further testing suggestions and future research.

Recap

The goal of this project is to attempt to predict customer churn rate based off past customer data. I will do this using a classification algorithm inside of a Jupyter Notebook. My final results will be published in a presentation with all work available via github.