



Northeastern University

Customer Retention Analysis

Project Proposal

INFO 7390: Advances in Data Sciences & Architecture
April 06th, 2018



Team 4

Eklavya Saxena 001850025
Ankur Jain 001206900
Amandeep Singh 001271649

Supervisor

Prof. Sri Krishnamurthy

Overview

Business can thrive and dive depending upon how happy your customers are, how loyal they are, and how willing they are to keep buying your services. Loyal customers are worth up to 10 times as much their first purchase and it is 6 to 7 times more expensive to land a new customer than to keep a current one.

According to a survey 91% of unhappy customers will not willingly do business with the company again.

Goals

A telecom company observes 'Customer Churns' also known as customer attrition, customer turnover, or customer defection, which is the loss of clients or customers. The estimated loss of income is approximately \$140K per month. These expenses are certainly resultant of attracting new customers, which is more expensive that retaining the current ones.

Given the historical data of 'loyal' and 'churn' customers of a telecom company, this project is aimed to:

- Understand the causes and/or scenarios causing attrition of customers
- Provide a predictive model that ranks the customers
- Learn the factors which are strongly correlated with churn rate
- Make recommendations to minimize the revenue loss



Data

Dataset Information and Description:

- Link: <https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/>
- The dataset consists of 7043 records/rows and 21 fields/columns
- Target Variable – Churn: Binary Classification {Yes or No}
- Following are the fields of the dataset:
 - 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

Process Layout

- **Data Wrangling** – data cleaning and handling data missingness
- **Exploratory Data Analysis** – draw out correlations and remove outliers
- **Machine Learning** – predictive modeling using supervised approaches and select the best model suited to the dataset
- **Pipeline** – create a pipeline to automate the feature selection and model training
- **Pickle** – serialize the model and store the model on AWS S3
- **Flask Web Application** – create a frontend web application to extract the pickled model and deserialize it to demonstrate the prediction and recommendation results
- **Docker** – dockerize an image for a platform independent implementation

Milestones

Timeframe	Deliverables
Day 1 – 2	Data Preprocessing & Exploratory Data Analysis
Day 3 – 6	Model Building, Training, Feature Selection, Pipelining
Day 7 – 9	Deployment of Pickled Models on Cloud, Build Web Application
Day 10 – 12	System Integration, Dockerize, Testing, Tuning, Documentation



References

- IBM Watson Analytics
- Customer Behavior Data – Improve Customer Retention
- <https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/>

