Springboard

Customer Churn Prediction in Telecom

Capstone Project 1

Alex Martkovich

Agenda

- **Introduction**
- **Historical data review**
- **Confusion matrix**

Introduction

A telecom company is looking to improve customer retention and decided develop focused customer retention programs based on customer churn prediction model.

The historical data includes information about:

- Customers who left within the last month
- 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 the company, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

In this document we've developed a few models, applied several techniques to improve prediction accuracy and built a business case.

NOTE: The analysis did not take into account probability of the customer retention after it has been correctly predicted.

Agenda

- Introduction
- Historical data review
- Confusion matrix

The Data contains 19 features and a target

#	Field	Description	Example
1	customerID	Customer ID	7590-VHVEG
2	gender	Whether the customer is a male or a female	Female
3	SeniorCitizen	Whether the customer is a senior citizen or not (1, 0)	0
4	Partner	Whether the customer has a partner or not (Yes, No)	Yes
5	Dependents	Whether the customer has dependents or not (Yes, No)	No
6	tenure	Number of months the customer has stayed with the company	1
7	PhoneService	Whether the customer has a phone service or not (Yes, No)	No
8	MultipleLines	Whether the customer has multiple lines or not (Yes, No, No phone service)	No phone service
9	InternetService	Customer's internet service provider (DSL, Fiber optic, No)	DSL
10	OnlineSecurity	Whether the customer has online security or not (Yes, No, No internet service)	No
11	OnlineBackup	Whether the customer has online backup or not (Yes, No, No internet service)	Yes
12	DeviceProtection	Whether the customer has device protection or not (Yes, No, No internet service)	No
13	TechSupport	Whether the customer has tech support or not (Yes, No, No internet service)	No
14	StreamingTV	Whether the customer has streaming TV or not (Yes, No, No internet service)	No
15	StreamingMovies	Whether the customer has streaming movies or not (Yes, No, No internet service)	No
16	Contract	The contract term of the customer (Month-to-month, One year, Two year)	Month-to-month
17	PaperlessBilling	Whether the customer has paperless billing or not (Yes, No)	Yes
18	PaymentMethod	Payment method - Electronic check, Mailed check, Bank transfer, Credit card	Electronic check
19	MonthlyCharges	The amount charged to the customer monthly	29.85
20	TotalCharges	The total amount charged to the customer	29.85
21	Churn (Target)	Whether the customer churned or not (Yes or No)	No

Dataset size and preprocessing

- The dataset has no missing values and requires no data cleaning

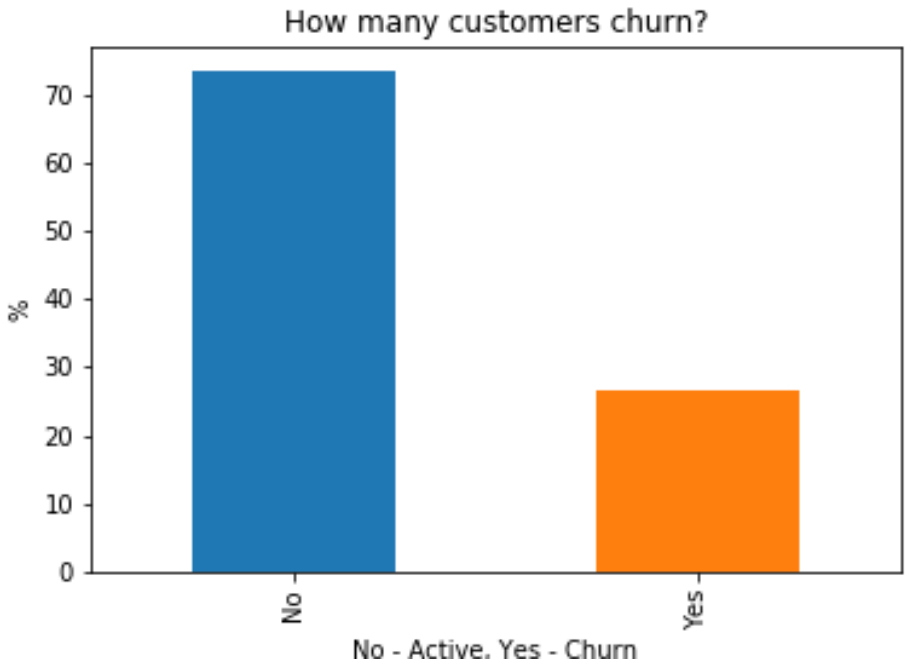
- Overall the file contains 7043 records and 20 columns

- In order to be able to perform modeling, the data requires preprocessing, including:
  - Converting string format to numerical
  - Converting categorical data to dummy values
  - Converting binary categorical to numerical

- After preprocessing the number of feature fields increased from 19 to 40
- Considering that the file contains significant amount of features, but no time series data or behavioral data, the feature engineering has not been done

Target variable

Approximately 26% of the customers churn



- Due to underrepresented data on Churning customers we can assume that the model will perform better at predicting non-churning customers

- We will test strategies to deal with this imbalance

Agenda

- **Introduction**
- **Historical data review**
- **Confusion matrix**

Confusion matrix – we are looking to improve precision and recall of churn prediction

		Predicted value	
		1 (Positive)	**0 (Negative)**
Actual value	**1 Positive**	**TP** **Customers who are predicted and will churn**	FN Customers who are predicted to stay, but will churn
	0 Negative	**FP** **Customers who are predicted to churn, but will stay**	TN Customers who are predicted and will stay

Analysis Focus

Net effect of predicted model parameters

Importance	Value	Event	Treatment cost	Retained Profit	Net effect	Objective
1	**TP**	Customers who are predicted and will churn	-$422*	$2,279**	$1,875	Maximize
2	**FP**	Customers who are predicted to churn, but will stay	-$422	0	-$422	Minimize
3	FN	Customers who are predicted to stay, but will churn	0	0	0	
4	TN	Customers who are predicted and will stay	0	0	0	

Analysis Focus

* Average monthly revenue per customer ~ $70.42. Offered promotion to retain a customer is 50% discount for 6 months ~ $422.55

** Average revenue per customer to date ~ $2,279

Model performance assessment

The model performance have been compared and optimized based on the following parameters:

1. **Recall** – number of accurately predicted churning customers out of total number of churning customers, to offer treatment and reduce forgone revenue (the higher Recall the better)

2. **Precision** – number of customers predicted to churn, but will actually stay, to reduce the total number and cost of offered treatment (the higher Precision the better)

3. **Revenue retained** – amount of revenue retained due to churn prediction model application out of the total potential forgone revenue

4. **Accuracy score** – overall model accuracy


Springboard