Churn Analysis

Using Survival Curve and Interpretable Machine Learning Model

Kyra Quan

Project Introduction

Key business objectives:

- Customers are the lifeblood of subscription business. Losing customers (churn) requires gaining new customers to replace — a 10X more expensive alternative than retaining existing.
- Solution: Use interpretable machine learning model to understand customer churn propensity

Project Introduction

Data Source:

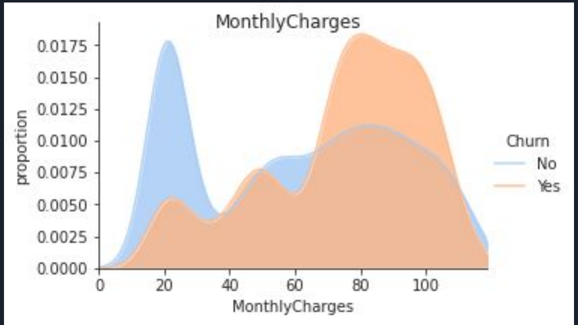
- ➢ IBM sample set for practicing data analysis on a real-world type of business problem
- ➢ 7043 observations and 22 variables that contain information about
    - o customer demographics
    - o services they signed up for
    - o account information
    - o churn (target variable)

Preliminary Analysis

- KEY POINTS:
  - tenure = Time, Churn = Target, Everything Else = Possible Predictors
  - change datatype of 'TotalCharges’
  - Inspected missing value, I found the missing values in 'TotalCharges' column all have 0 in 'tenure' column, which means they are all new customers. So I'll set every the missing value in 'TotalCharges' column equals to its 'MonthlyCharges'.

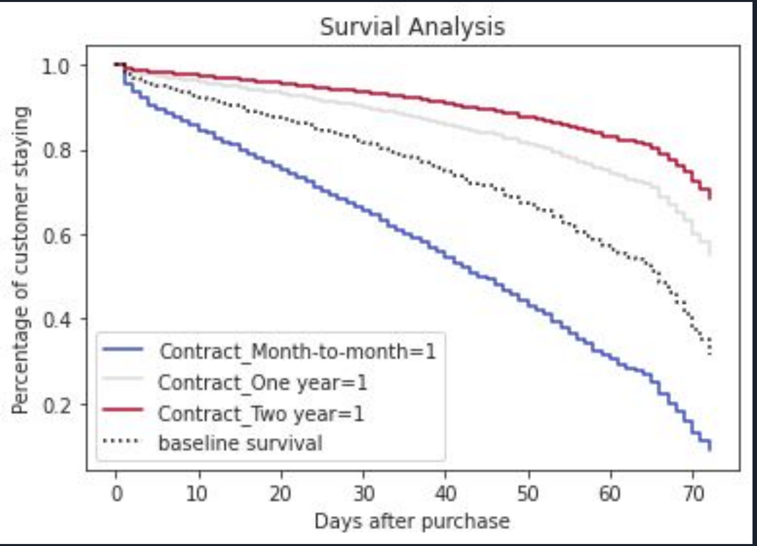
EXPLORATORY DATA ANALYSIS

- Key Points:
  - Explored categorical and numerical columns
  - Noticed the imbalanced distribution of target variable
  - Transformed categorical columns using one hot encoding
  - Identified customer churn problem using survival analysis



EXPLORATORY DATA ANALYSIS

- Why use survival analysis?
- Strength:
  - Survival Curves - communication tool that easy for business leaders to understand
  - Cox's proportional hazard model - used to incorporate multivariate analysis
- Weakness:
  - Not as high performance as Machine Learning
- Solution:
  - Build Machine Learning model to predict churn propensity

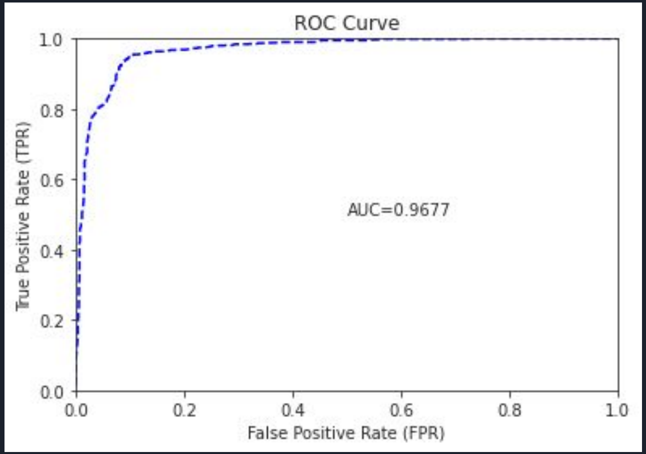


Insight: Chart shows that more than 50% of Month-to-month contract customers leave the company after 50 days of purchase

Model Building and Evaluation

- Built model using Random Forest:
    - Resampled imbalanced dataset
    - Tuned hyper parameters
    - Model achieved 0.93 AUC

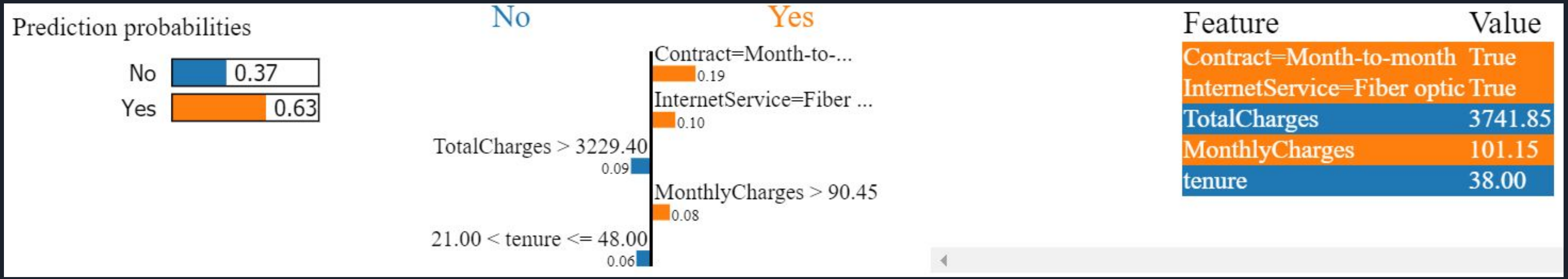
- Built second model using H2O autoML
    - Resampled imbalanced dataset
    - Best model achieved 0.96 AUC
    - Interpreted the model using LIME



Model Interpretation

- Now we can look at how the model has made that decision.

- We can see that it is attributed to his contract type, internet service type , total charges, monthly charges and tenure. The fact that he's paying 3741.85 of total charges and he's been a loyal tenure for 38 years has tried to pull down his likelihood of churn, but overall the model has decided that this person is 63% likely to churn.



Insights and Recommendations

Insights:

- Customers with high monthly charges and low tenure are more likely to churn.
- InternetService has a negative effect on Churn, and PhoneService has a null effect

Recommendations:

- Offer discounts to customers who are paying a high monthly charges and are very likely to churn. Further investigate on those customers to find out whether this promotion strategy is effective or not.
- Analyze the model at a more global level, and perform hypothesis tests to identify the non-effective attributes. Reduce those attributes to improve the company’s survey model and service plan.