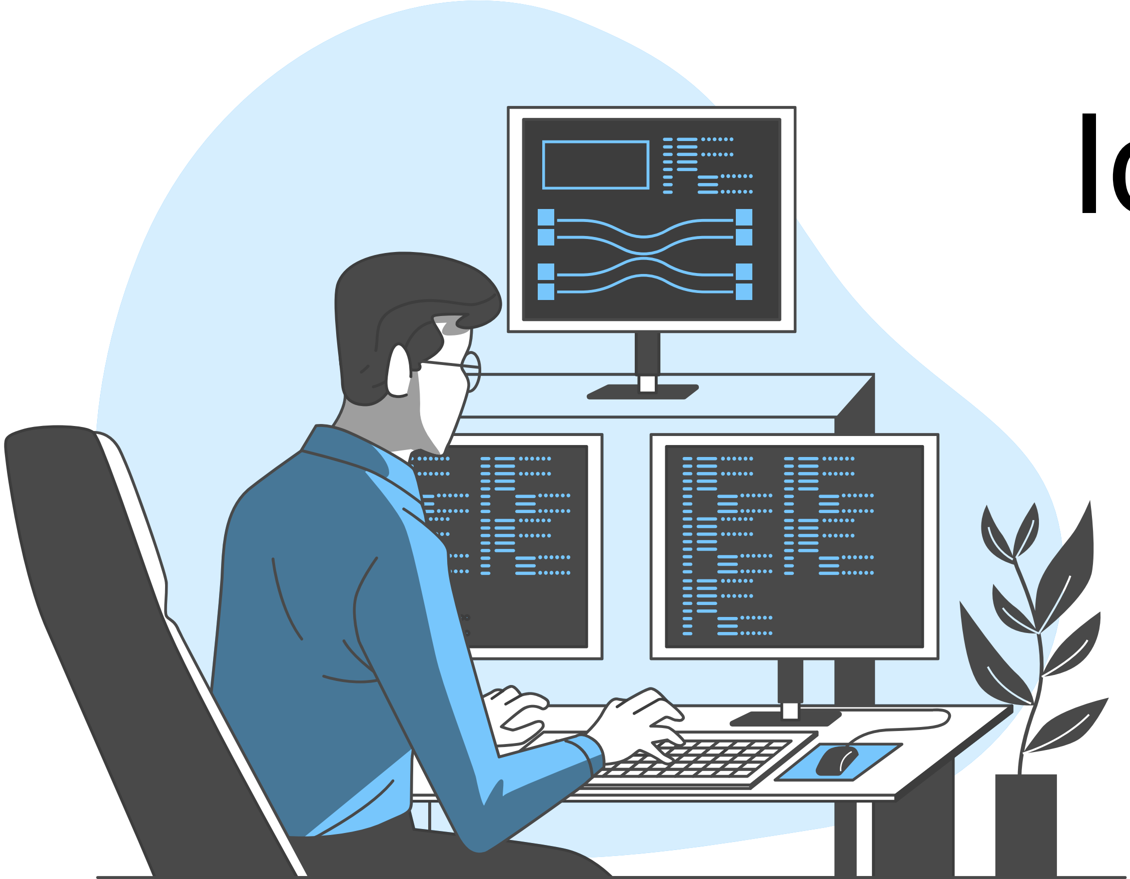


Identifying Telecom Churn

Capstone Project 3

Tom Shields

01

What is Customer Churn?

The number of paying customers who fail to become repeat customers.

02

Why does it matter?

Acquiring a new customer is more expensive than retaining an existing one.

03

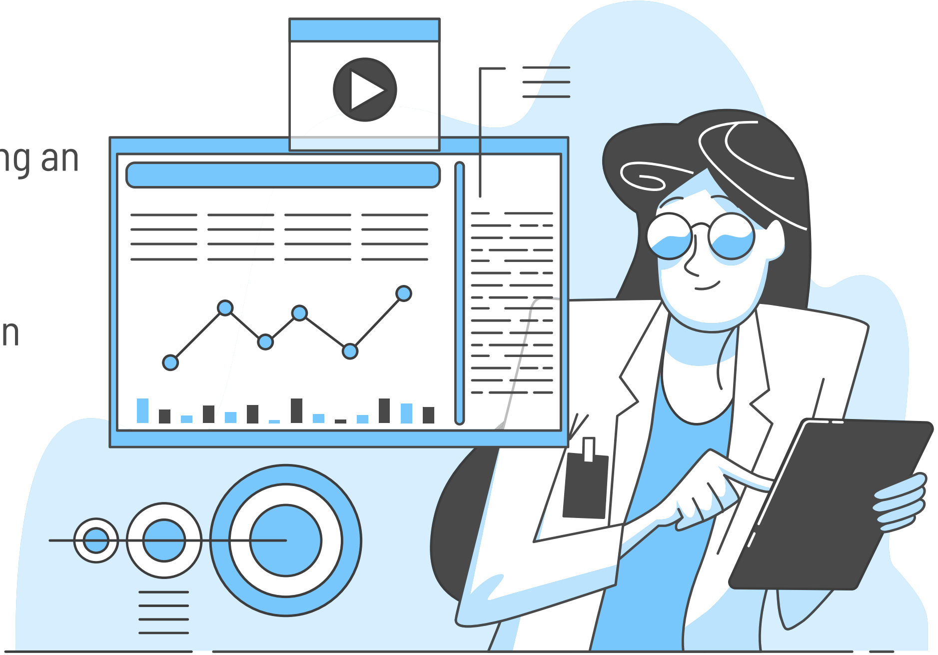
Problem Identification

Identify customers who are in danger of churning.

04

Solution

Offer these customers promotions to avoid churn



The Data

Data consists of information of 7043 customers' service contracts, the type of services offered, payment details etc. Data was extracted from IBM Cognos Analytics Data Collection.

Our Target

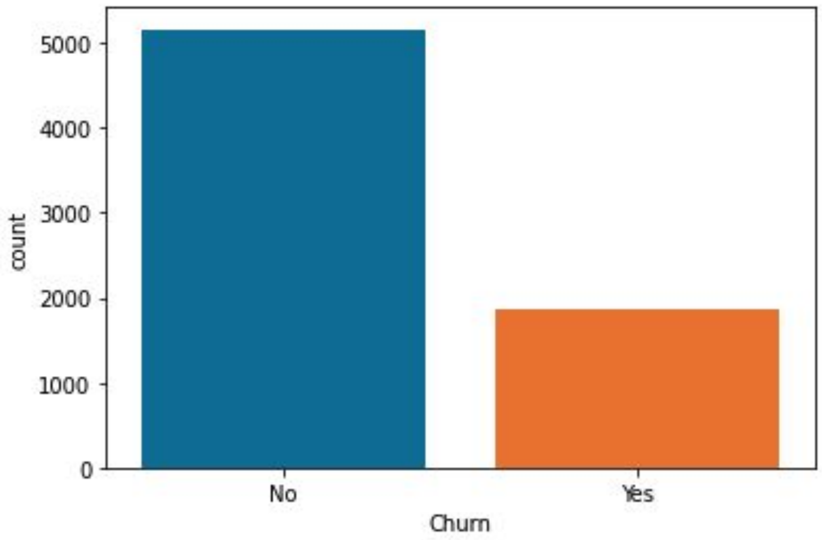
The dataset initially consisted of 21 columns and there were no missing values to have to impute. The target variable is labeled "Churn" and is a binary variable with Yes (if they have churned) and No (if they haven't churned) options.

Understanding the Data

Timespan: Manipulating the “tenure” variable we were able to deduce that the data was collected over 72 months, or 6 years.

Balance: This data set is neither balanced nor heavily imbalanced (Churn No:%73 vs. Yes: %27).

Data Type: The numeric data centers around what customers were charged and how long they were retained. The categorical data centers are demographics and services purchased.



Solution Steps

01 Identify Churners

Predict potential churners

02 Create a Customer Profile

A list of customer characteristics

03 Design Possible Promotions

Create deals on services that will engage customers.

04 Initiate Outreach

Deliver promotions to customers in danger of churning.

Exploratory Data Analysis

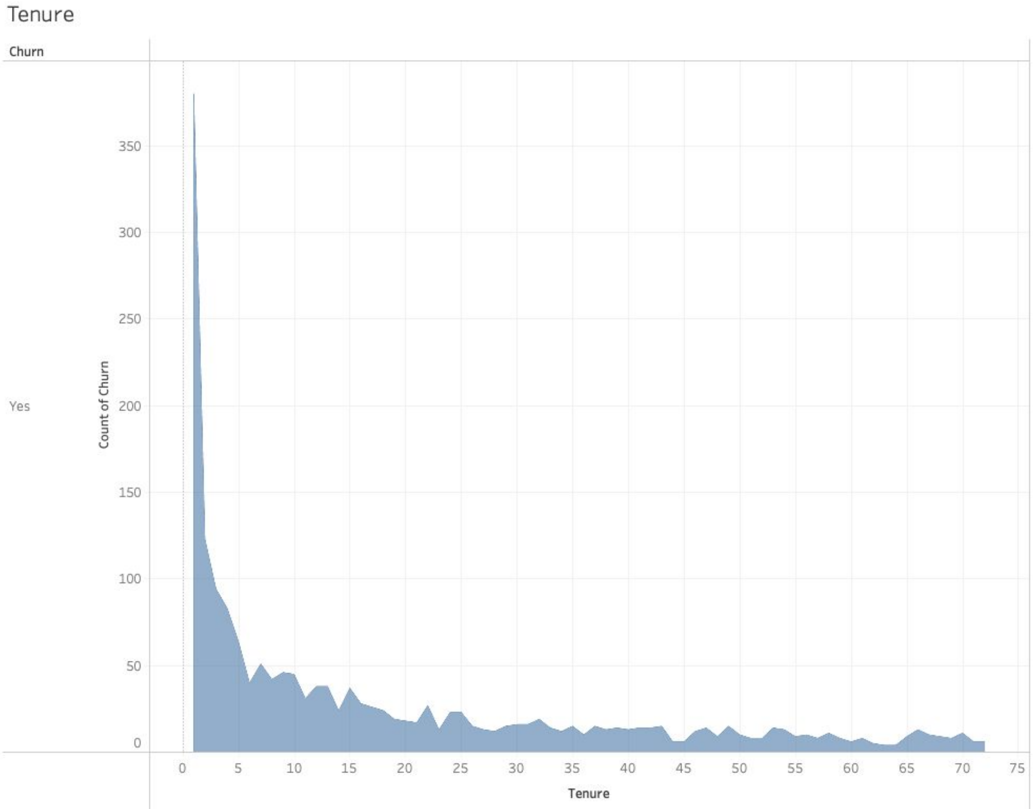
The data features can be divided into subcategories:

User/Demo Segment: Personal information centered around life circumstances

Services Segment: The regular services that customers subscribe to.

Account Segment: Numeric data involving monthly charges, total charges and tenure of customer.

Account Information



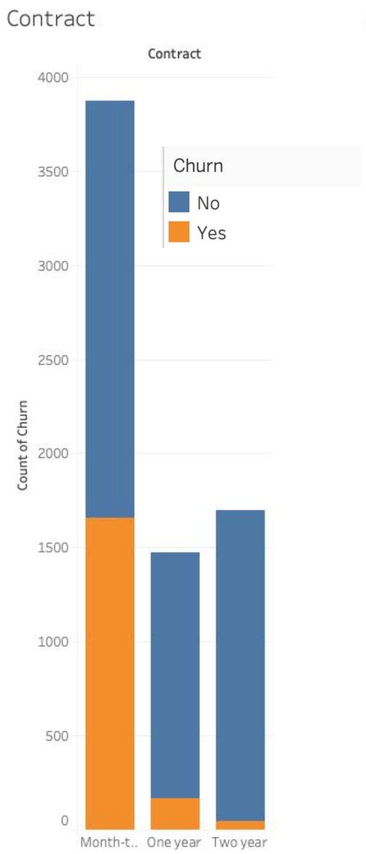
Takeaways:

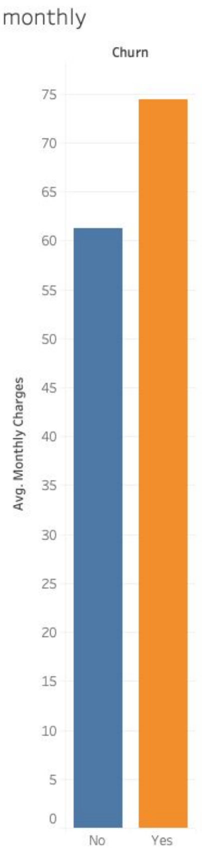
39.8% of all customer attrition happened with the first 5 months.

20.33% of all customer attrition happened after just one month.

This is clear from the large spike of customer churn over a brief time with the service

Account Information





Takeaways:

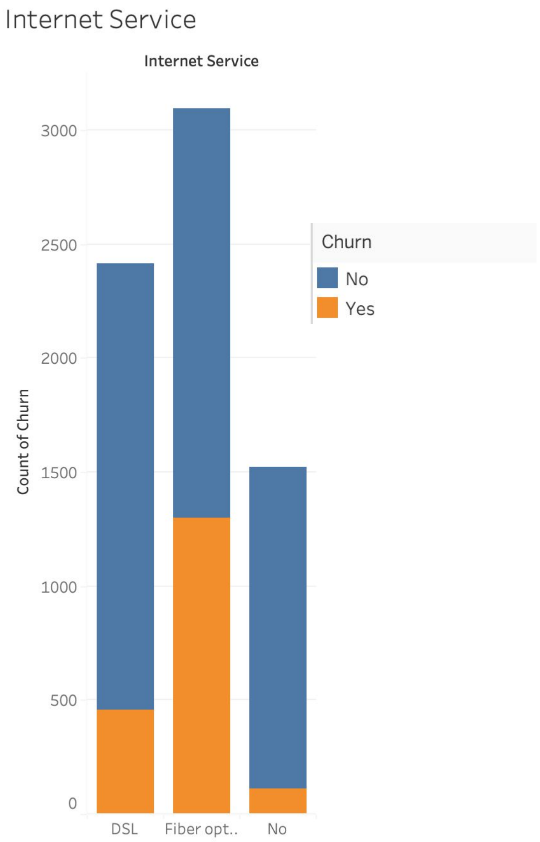
Customers with month-to-month contracts are more likely to churn.

Churn vs Retention

Month to month - 1655/2200 (42.71%)
One year contract - 166/1307 (11.27%)
Two year contract - 48/1647 (2.83%)

Customer with high monthly charges are more likely to churn. An monthly account balance over $61.27 is an indicator of potential attrition.

Service Information

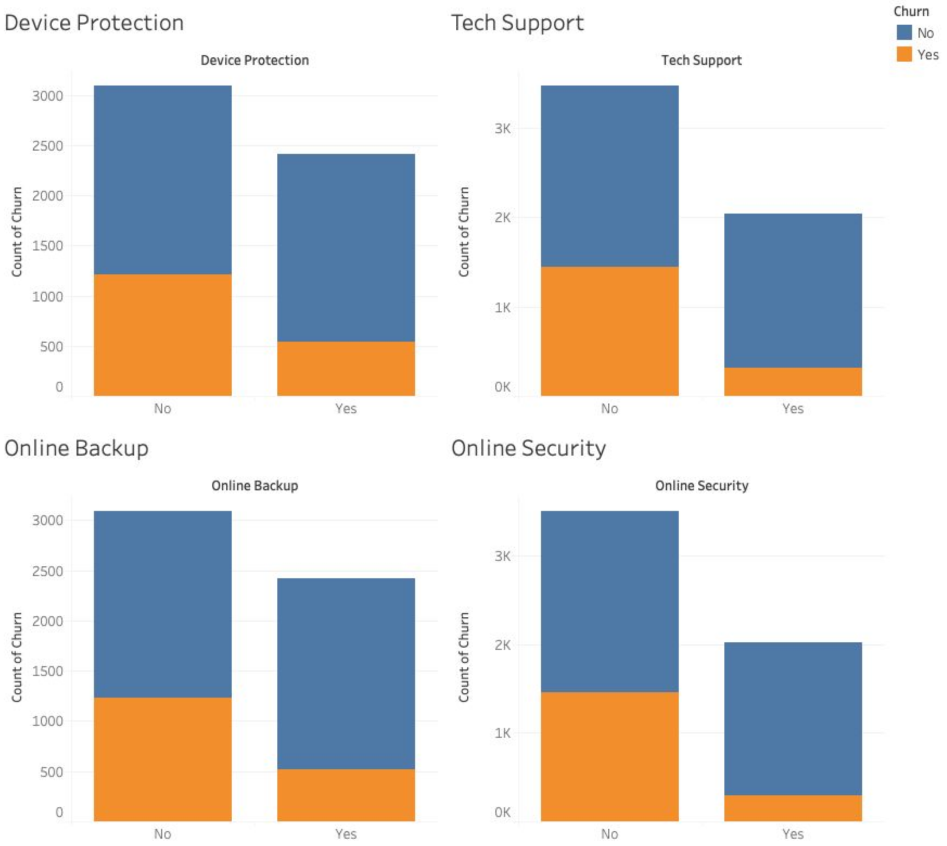


Takeaways:

Customers that subscribed to internet services using fiber optic cable were more likely to churn.

Fiber optic - 41.89% churn rate
DSL - 19% churn rate
No internet - 7.43% churn rate

Support Service Information



Takeaways:

Customers that do not subscribe to device protection, tech support, online backup or online security are more likely to churn.

Customer churn rate when not subscribed to:

Device protection - 39.14%
Tech support - 41.65%
Online backup - 39.94%
Online security - 41.78%

Modeling

Metric of Success

<u>The nature of Churn problems:</u>

- When measuring customer churn it is best to not use accuracy as a measure
- Churn data is imbalanced data
- The attrition rate (customers who churn) will not be the same as the retention rate (customers who stay).
- We can fix this problem through random sampling

The best metrics:

- A boxplot of F1 scores (balance of precision/recall) to help inform which default model works best
- The precision-recall curve is used for evaluating the performance of binary classification algorithms
- The confusion matrix helps us sort out how many churned customers our model missed

Comparing Models With Cross Validation Scores

Algorithm	F1 Score (Mean)	F1 Score (Standard Deviation)
Random Forest	**0.485**	**0.344**
Logistic Regression	**0.483**	**0.346**
Adaboost	**0.484**	**0.347**
Decision Tree	**0.472**	**0.334**
Gradient Boosting	**0.483**	**0.345**
SVC	**0.470**	**0.338**

Characteristics of Random Forest Model

- Consists of a large number of individual decision trees that operate as an ensemble.
- Random state ensures that the splits that you generate are reproducible
- A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models
- To classify a new object based on attributes, each tree gives a classification.
- The forest chooses the classification having the most votes

Prioritizing Metrics

Minimizing the number of missed churn customers. The smallest amount of false Negatives.

1

2

Maximize the number of identified customer who churn. Largest amount of true positives.

3

Maximizing the number of identified customers who didn’t churn. Largest number of true positives.

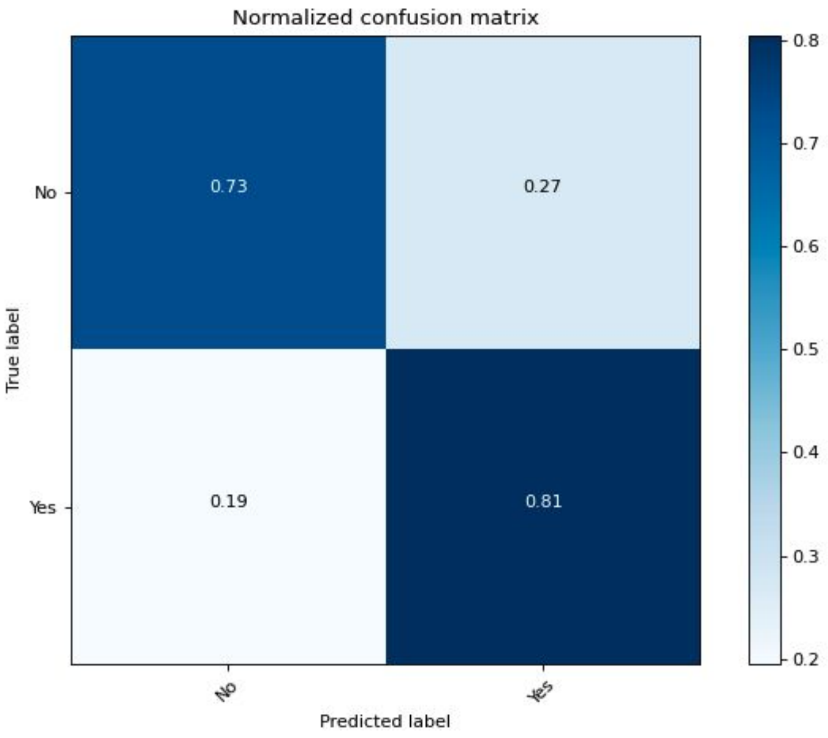
4

Minimize the number of missed customers who didn’t churn. The smallest amount of false Positives.

Confusion Matrix

True Positives

False Positives

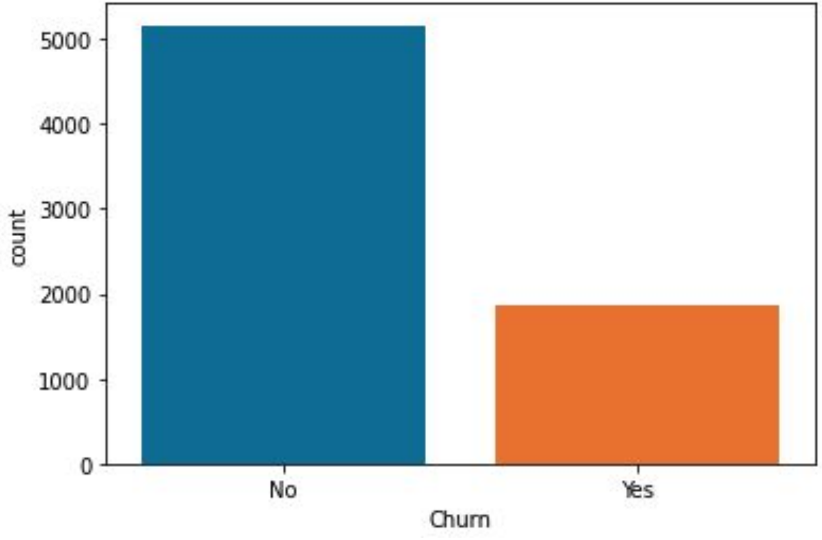


False Negatives

True Negatives

False Positives vs False Negatives

In regards to customer churn, false negatives may provide a falsely reassuring message to the company. This sometimes leads to inadequate efforts to retain a customer. So, it is more desirable to have too many false positives but a small number of false negatives.



Classification Report

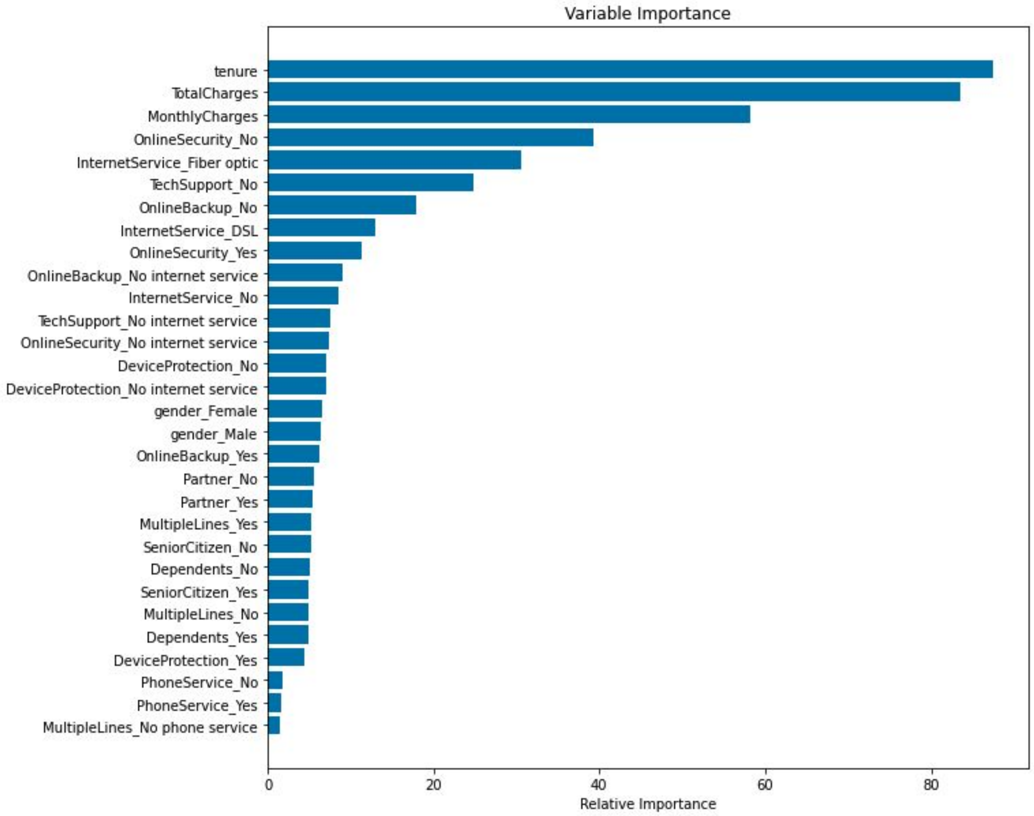
Labels	Precision	Recall	F1
Retained Customers	**0.90**	**0.73**	**0.81**
Churned Customers	**0.56**	**0.81**	**0.66**
Accuracy			**0.75**
Macro Average	**0.73**	**0.77**	**0.73**
Weighted Average	**0.80**	**0.75**	**0.76**

Nice!

Model identified 422 churn customers out of 524 actual churn customers with a recall of 81%.

We can reach out to these customers before they churn with incentives to stay.

Variable Importance in Prediction



Prioritizing Variables

1

Tenure: How long has a customer subscribed to the service?

2

Monthly Charges: How much money is the customer spending on a monthly basis?

3

Fiber Optic Internet Service: Are customers using the service?

4

Service Subscriptions: Are customers not subscribed to online security, tech support and online backup?

Recommendations

- Filter customer that have subscribed for less that two years, who have high monthly billing statements, that use fiber optic internet services, and that do not subscribe to online security, tech support, and online backup.
- Reach out to prospective churners through email with surveys regarding customer satisfaction, likelihood of churn and possible competitors, follow up with incentives
- Target these customers by building customer loyalty through discounts on longer contracts.
- Promotions should be centered around deals on longer contracts, partial credits for monthly billing statements, and discounts on bundled subscription services.
- Reevaluate price optimization through the use of mathematical tools to determine how customers will respond to different prices for its products and services through different channels
- Reevaluate the quality of fiber optic internet service, or customer service regarding fiber optic internet service, as it could be causing churn