Telecom Customer Churn Prediction

- Abhishek Kangale
- Akashnil Roy
- Ameya Kanawade
- Vivek John
- Moumita Ghosh
- Praveen Kumar
- Mano Aravind M

Agenda

- Business Problem
- Data Understanding
- Data pre-processing
- EDA Observations
- Modeling & selection of best model
- Classification Report
- Next Best Action

Business Problem:
Telecom Customer Attrition

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn.

Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn.

The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn.

Data Understanding

- The given dataset has 21 variables and 7043 records.
- Variables are customer ID, gender, monthly bill, total bill, payment method, subscriptions of the customer, partner, contract type, etc.
- All the above variables are used to predict whether a customer will churn or not.

Data pre-processing

One of the variables "Total Charges" was converted to "float" datatype as its datatype was "object" despite having numeric entries.

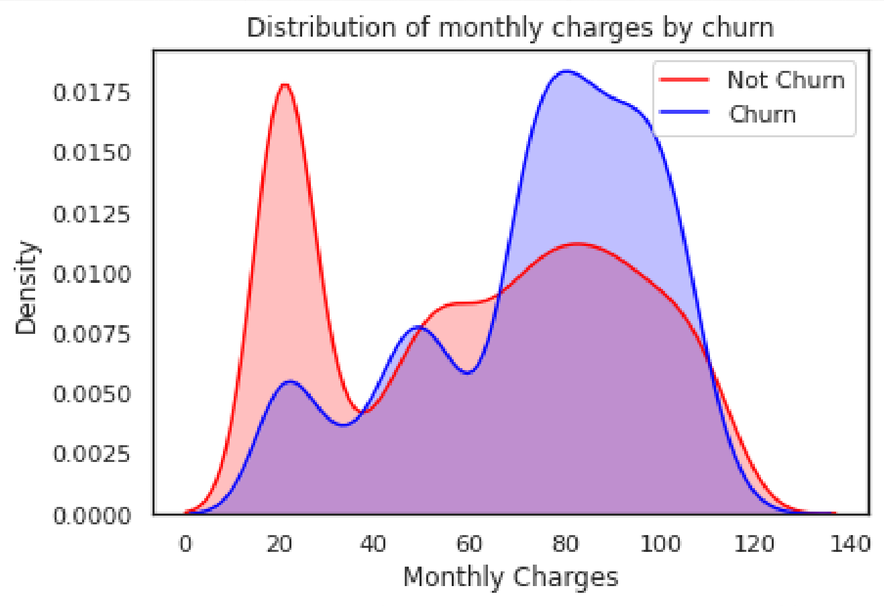
This same variable had missing values and its distribution was skewed.
Median Imputation was used to solve the problem.

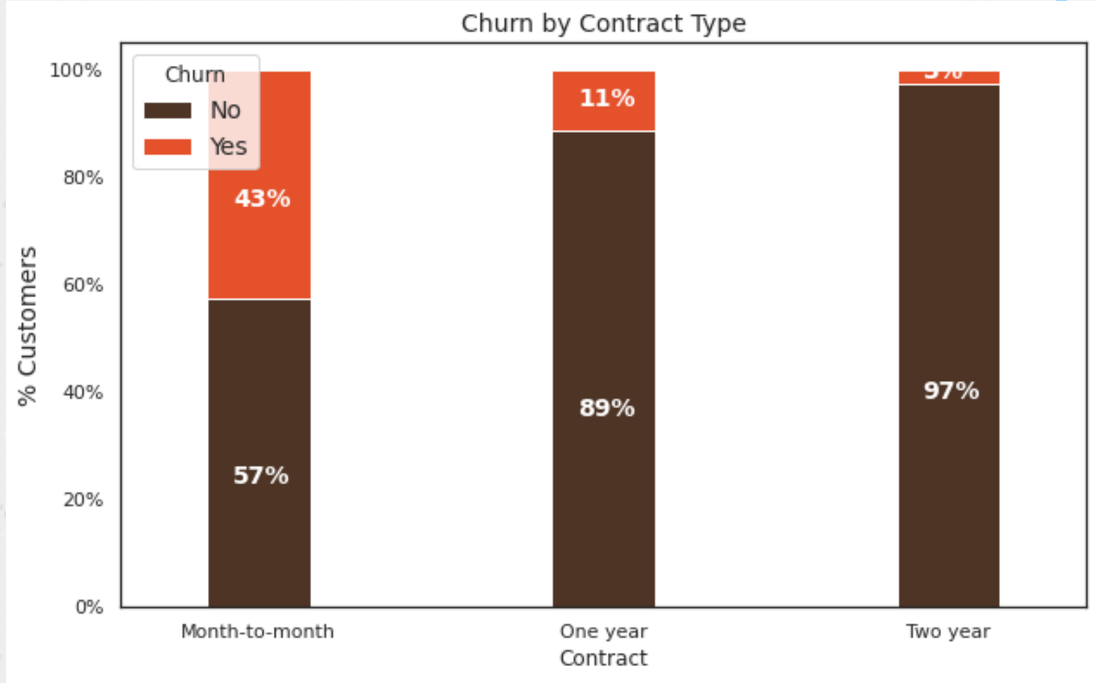
All the variables in the dataset were used in the model except the "customer ID".

The catgorical variables in the dataset were converted to numeric type. By using get_dummies function.

Variables like monthly contract type, online security and tech support were positively correlated with churn variable. While, tenure, two year contracts were negatively correlated with churn variable.

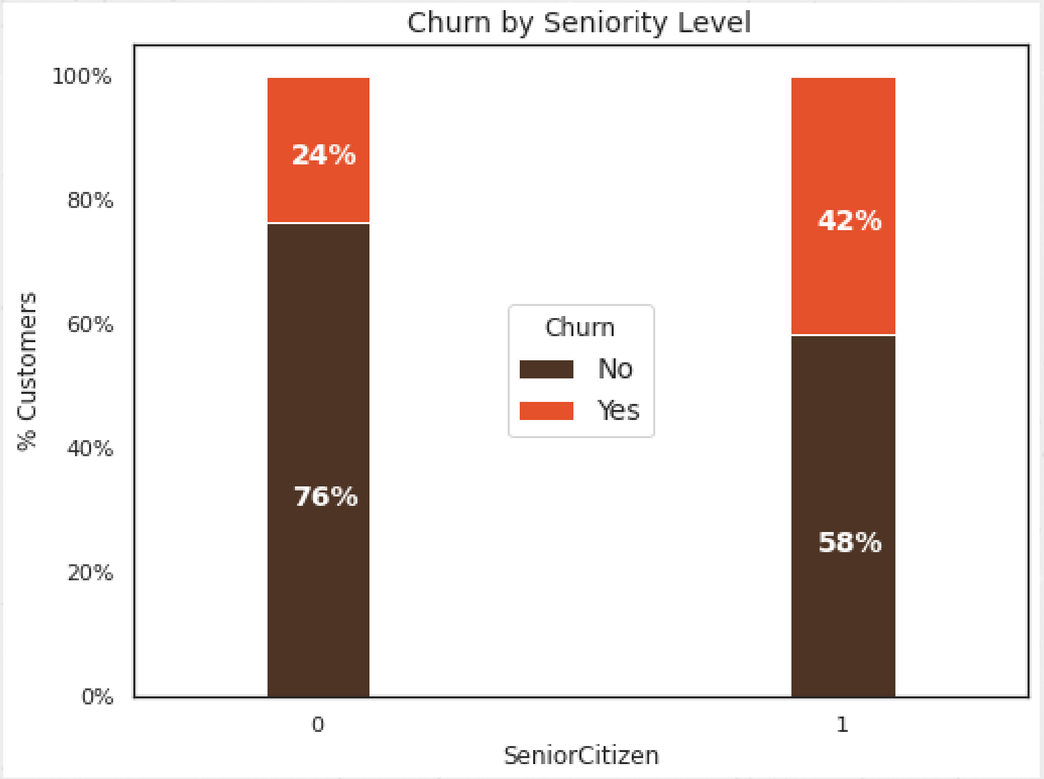
EDA Observations

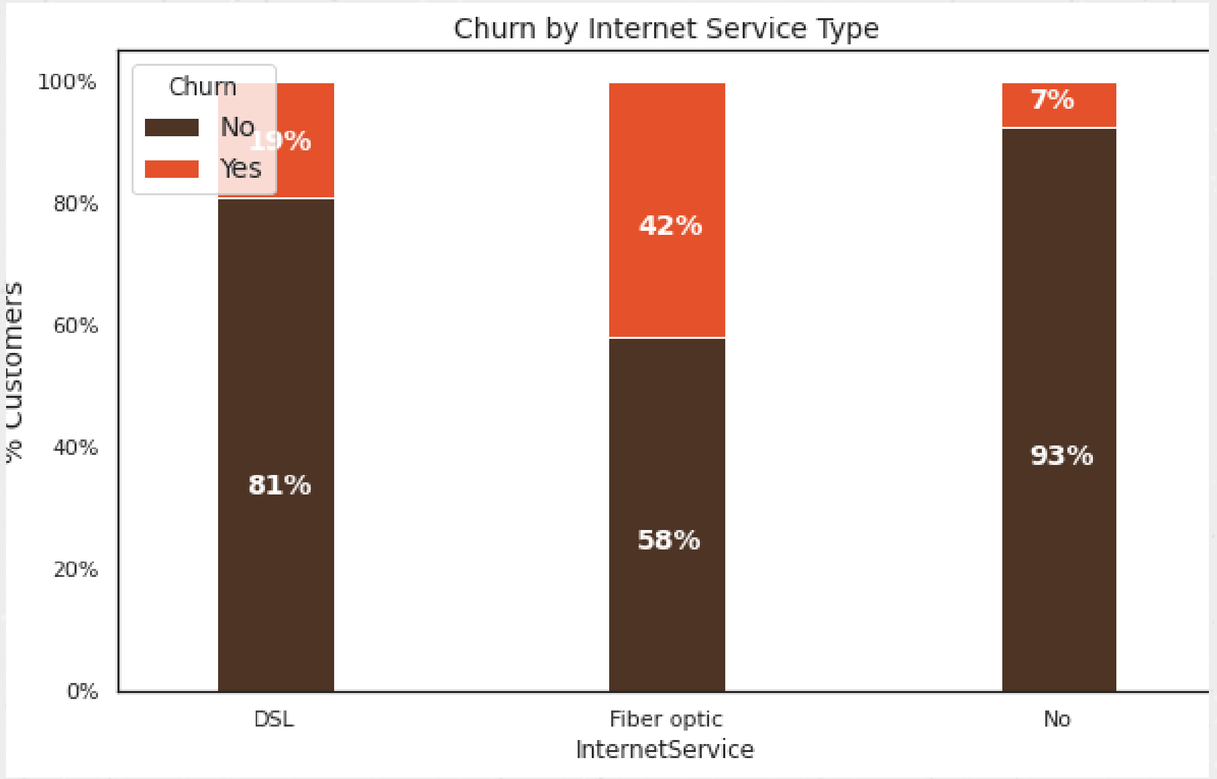




According to the graph, higher % of customers churned when the monthly charges were high.

The customers who were having a month to month contract had a very high churn rate.





Interestingly, senior citizens have almost double the churn rate than younger population.

Churn rate is highest for Internet service type Fiber Optic and lowest for customers with no internet connection.

Modeling

CLASS IMBALANCE

The data was imbalanced. The number of non-churn cases were 73%. The number of churn cases were 27%.
This was resolved through oversampling of minority class by SMOTE.

SCALING

The dataset was standardized using MinMax Scaler (Non-Gaussian) to prevent any biases creeping in due to the skewness in the values of different variables.

MODELS USED

- Logistic Regression
- Random Forest Classifier
- Support Vector Machine (SVM)
- Xgboost

Selection of Best Model

Logistic Regression (LR) performs well according to the classification report.

The main aim is to retain customers for the company . As newer customers are difficult to acquire without proper strategies.

The LR model satisfies the business requirements as per performance metrics in the classification report.

Classification Report

	precision	recall	f1-score	support
0	0.84	0.91	0.87	1546
1	0.68	0.53	0.60	567
accuracy			0.81	2113
macro avg	0.76	0.72	0.73	2113
weighted avg	0.80	0.81	0.80	2113

This report shows precision and recall is 68% and 53% respectively of minority class.

Area under the ROC curve is highest for Logistic Regression Model ie. **0.84**

In business terms, out of all the customers who have actually churned, the model is able to predict that 53% people have churned.

Out of all the predicted churn customers only 68% customers have actually churned.

Most important features obtained from the dataset are total charges of the customer, monthly contract type, internet service (Fiber optic) type.

Next Best Action

In this scenario, the important features can be looked into for those customers who are churning. So that marketing campaigns, advertisements can be designed for customer retention by focusing more on those important features.

According to EDA, Customer churning is less when customers stay for a longer tenure with the telecom company. Therefore affordable and convenient loyalty plans should be designed to attract customers.

Source code:

https://colab.research.google.com/drive/1t_QNeL5LfihfTOOTa07tTSCGfAGVkdoX

Thank You