**Project: Telco Customer Churn**

DSC630\_Predective\_Analytics

Name: Bhargava Gaggainpali

Project Summary

**Introduction:**

Customer churn is a major problem and one of the most important concerns for every company. Retaining an existing customer is many times more effective thanks gaining new customers. Generally, people only switch to a different company only when the service is not to the level of expectation in one or the other factor when compared to competitors, more than getting attracted to the specials and offers which are offered by the other companies. As the number of peoples who use a phone or a telecom product does not increase in huge volume, any company looking to improve the profits is trying to attract other company customers. At the same time retaining existing customers is very important. As retaining a current customer is ten times more productive than gaining a new one.

**Problem Statement:**

As customer churn directly effects the revenues of the companies, 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 and retain the customers.

**Solution Approach:**

‘Big data’, ‘Predictive Analytics’ and ‘Machine Learning’ has gained popularity and with the latest emerging tools, complex mathematical algorithm can be applied on large datasets to discover data patterns which were not possible or would take huge amounts of resources in past days. Using Predictive analytics as a study by using the past customer data in the Telecom industry, we can section out the pattern of the customers, who has churned out to a different company.

By identifying such a section of people in the list of current active customers of the company, and addressing their issues, will help in reducing the number of customer churn to other companies.

**Scope:**

Scope of the project is to use the sample dataset from Kaggle, which shows several categories of customers and also some influencing factors causing the customer churn in the telecom industry to perform customer segmentation and identify a pattern which influence the customer churn. Implement supervised machine learning process to achieve the goal.

**Document Overview:**

Detail all the steps which are followed to find the factors which are influencing the customer churn in the telecom industry. Which includes the source of the data, tools used to access data, processing mechanism to achieve the goal. Include the validation process to make sure the process is not biased.

**Data sources or plan for data:**

For the analysis, Data is gathered from “Data source from Kaggle” which has 7043 observations with 21 parameters.

Telco Customer Churn | Kaggle [Telco Customer Churn | Kaggle](https://www.kaggle.com/blastchar/telco-customer-churn)

https://www.kaggle.com/blastchar/telco-customer-churn

Given are list of the parameters:

1. customerID

2. gender

3. SeniorCitizen

4. Partner

5. Dependents

6. tenure

7. PhoneService

8. MultipleLines

9. InternetService

10. OnlineSecurity

11. OnlineBackup

12. DeviceProtection

13. TechSupport

14. StreamingTV

15. StreamingMovies

16. Contract

17. PaperlessBilling

18. PaymentMethod

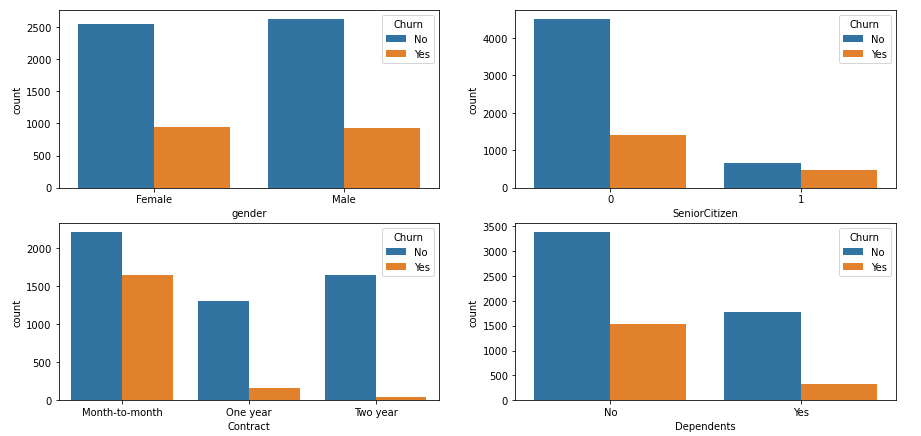
19. MonthlyCharges

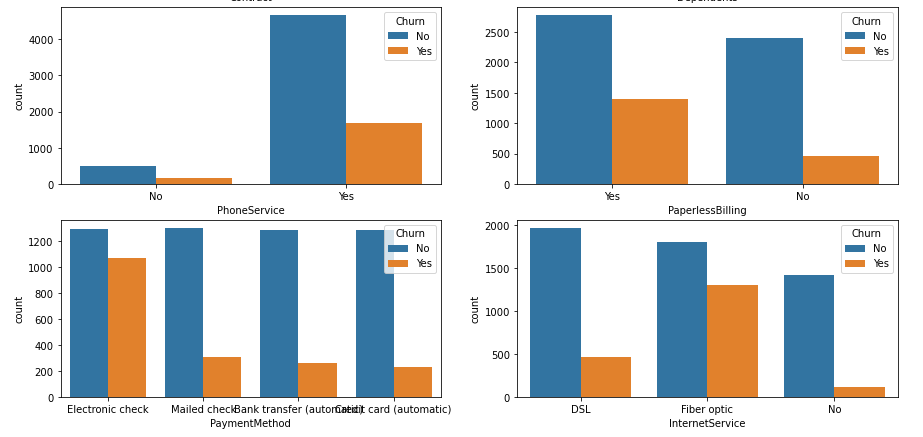
20. TotalCharges

21. Churn

**Exploratory Data Analysis:**

As the data is collected from internet, we have performed few steps to gather initial stats and check on the duplicate rows and null values to make sure the valid content of the data. Generate multiple charts on the variables to see the spread of the values in the variables.





**Initial Observations:**

1. Variable “gender” has both Male and Female members and has equal customer churn.
2. For “Senior Citizen”, we see that churn is more when the customer is not Senior Citizen.
3. Churn is more when the “Contract” is ‘month-to-month’.
4. Churn is more when customers has no “Dependents”.
5. Customers having “Phone Service” has higher Customer churn.
6. Customers who opted for “Paperless Billing” has higher Customer churn.
7. Customers who generally pay with ‘Electronic Check’ has higher Customer churn.
8. Customers who are using ‘Fiber Optic’ as “Internet Service” has higher Customer churn.

**Technical Approach:**

As the Machine Learning model is a mathematical algorithm which can be applied on the numerical data sets, all the categorical values are converted to the numerical values by giving equal importance to each value.

The sample dataset is divided in to two sections. 70% for training the model and 30% to evaluate the model, before applying it to different Supervised Machine learning algorithms.

As the mathematical concept and formulae are different in different Machine Learning models, we applied the training data to three different models to check on the best model which suits the data.

Machine Learning Models:

Logistic Regression Model

Random Forest Classifier Model

Ada Boost Classifier Model

**Results:**

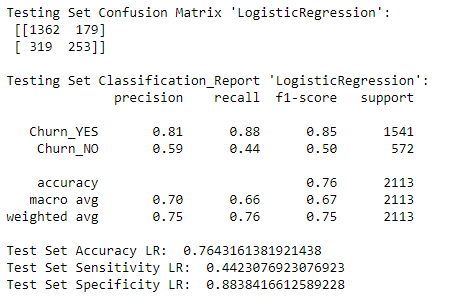
After the built of model using multiple Machine Learning algorithms, average accuracy of 77% was achieved.

Logistic Regression Model - Accuracy 76%

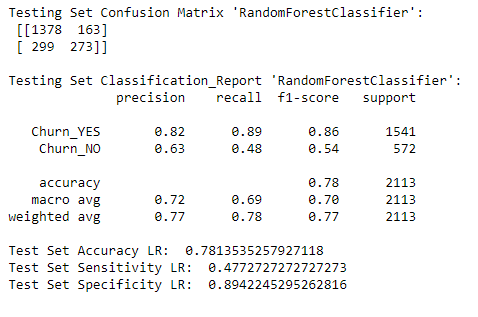
Random Forest Classifier Model - Accuracy 78%

Ada Boost Classifier Model - Accuracy 77%

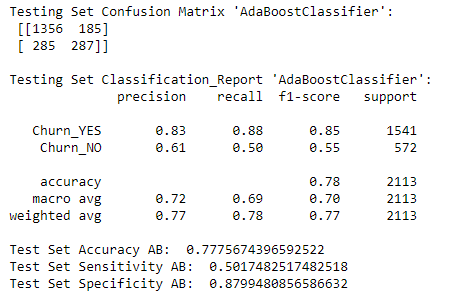












By identifying such a section of people in the list of current active customers of the company, and addressing their issues, will help in reducing the number of customer churn to other companies.

**Reference:**

1. “’Predictive Analytics’ The power to predict who will click, buy, lie or die” by Eric Siegel.

2. “‘Applied Predictive Analytics’ Principles and techniques for the professional data analyst” by Dean Abbot.

3. <https://www.kaggle.com/blastchar/telco-customer-churn>

4. [How Costly Is Customer Churn in the Telecom Industry? - The European Business Review](https://www.europeanbusinessreview.com/how-costly-is-customer-churn-in-the-telecom-industry/?__cf_chl_jschl_tk__=ec58e0210ef4250a8414742b7aaa8d8e28384753-1611086044-0-AXu6Mldw8Cpmi6L6s5SgV1PXJkIGPIbNtM2k3mk8dG4GwrfiMZ4fT3rVf0CPVbws4ipH6NuFN2fITmeBiQb8uSJYkXQKBcXB-e3XBH6IJWYli7zOM4eqnO9BCh0JE11v9IHHQZeKKMNRTUecYHq5Mahj6uNBBJ23n_LWFAdEF677ZQEQpPYruho0OLOLbdjBj-oKjPGHAZ52v2QMgPAHjyXRIDb-FEUu_sFhz8CW6q0KxFhOgoPWW_R8KZhMo4AOUlv2fbzwIi3oYhYXMSMOdAjpE9u7gRwD-0RQeFCEtTEzQqRmjC2QEbIGAHZKgbkn99TTuu5kBJLOuOoRLFhaD9d-oHKuQ27yx7WrU_mDPxYdp7qkByeE3o25LzRg460JyA)

<https://www.europeanbusinessreview.com/how-costly-is-customer-churn-in-the-telecom-industry/?__cf_chl_jschl_tk__=ec58e0210ef4250a8414742b7aaa8d8e28384753-1611086044-0-AXu6Mldw8Cpmi6L6s5SgV1PXJkIGPIbNtM2k3mk8dG4GwrfiMZ4fT3rVf0CPVbws4ipH6NuFN2fITmeBiQb8uSJYkXQKBcXB-e3XBH6IJWYli7zOM4eqnO9BCh0JE11v9IHHQZeKKMNRTUecYHq5Mahj6uNBBJ23n_LWFAdEF677ZQEQpPYruho0OLOLbdjBj-oKjPGHAZ52v2QMgPAHjyXRIDb-FEUu_sFhz8CW6q0KxFhOgoPWW_R8KZhMo4AOUlv2fbzwIi3oYhYXMSMOdAjpE9u7gRwD-0RQeFCEtTEzQqRmjC2QEbIGAHZKgbkn99TTuu5kBJLOuOoRLFhaD9d-oHKuQ27yx7WrU_mDPxYdp7qkByeE3o25LzRg460JyA>