2019

Project By:

Ritesh Natekar

10/24/2019



**Predicting Customer Churn**

**for**

**Telecom Company**

-PROJECT REPORT-

**Table of Contents**

1. Abtract 3

2. Business Strategy 4

3. Approach 5

4. Steps to prepare fit regression model 6

5. Project Results and it's Interpretion 7

7. Business Decisions after model analysis 10

Abstract

What is Customer Churn?

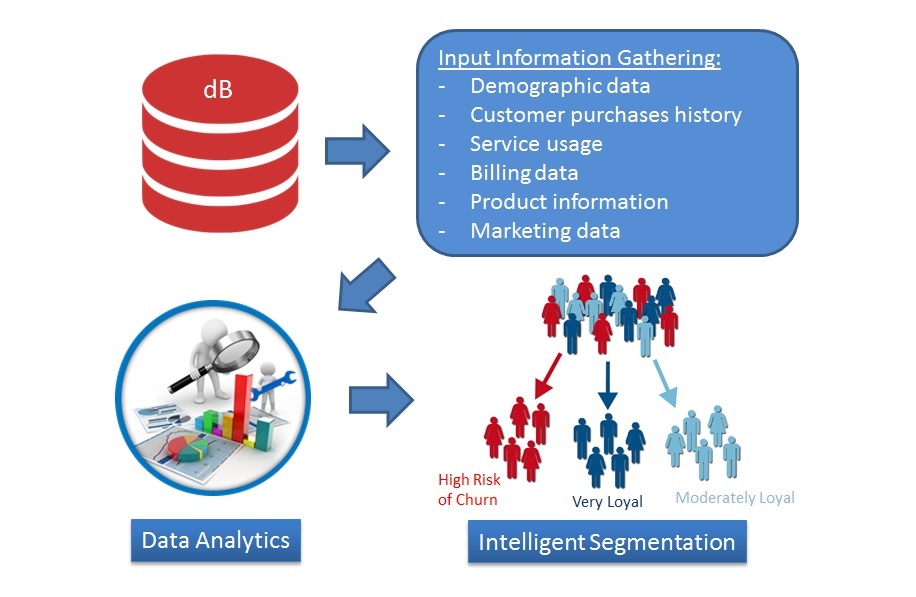
For a business in a stipulated period of time, customers can come under 3 major categories-  
a) Newly Acquired Customers  
b) Existing Customers  
c) Churned Customers

Churned Customers are those who have decided to end their relationship with their existing company. It can happen because of variety of reasons like-  
a) Bad customer Service  
b) Bad Onboarding  
c) Lack of Ongoing Customer Success

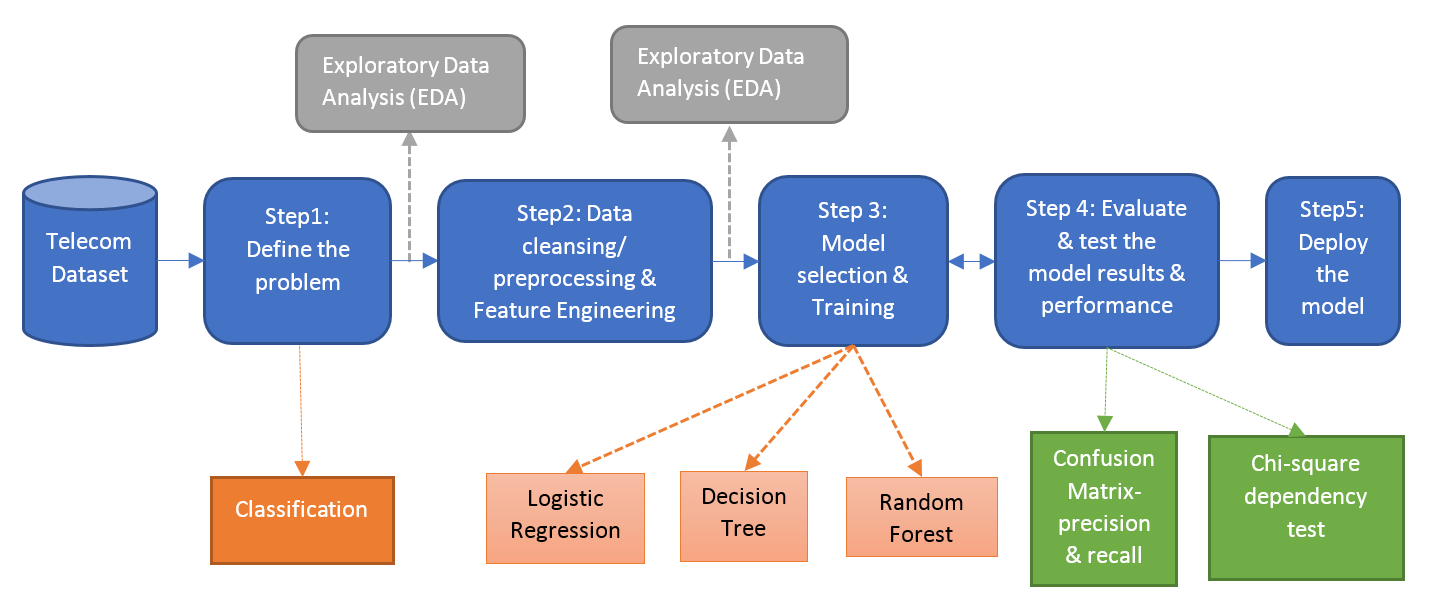
Why is it important?

Churned Customers means a direct loss of Marketing Acquisition Cost and possible revenue which could be capitalized post sale. Hence, predicting possible customers who can churn beforehand can help us save this loss.

Business Strategy



Approach



* Getting Dataset

Dataset has been collected in a “.csv” format. Examined the dataset with all variables included. Understand the importance of each variable in the dataset. Checking with data type of variables and how much relevant information they are providing for model building.

* Choosing Regression Model

In the given dataset, having variables with categorical and continuous type of data. Dependent variable i.e. Customer Churn is of type categorical and other independent variable which correlate the dependent variable more accurately are of type categorical. We have explanatory variable as qualitative and Response variable as qualitative, so **Logistic Regression** model is best fit for predicting the response variable i.e. Customer Churn.

* Cleaning dataset

Once regression model is confirmed, to start the working on dataset, it is very important to clean it. Finding missing values/null values and replacing it with relevant values in order to avoid discrepancy in data.

* Development Model and Testing Model

Divide dataset into two different model:

1. Development Model
2. Testing Model

Development Model, It is used to develop regression model. Considering different explanatory variables to predict response variable and examined accuracy of model.

* In the predictive modeling, the data need to be partitioned into train and test sets. 70% of the data will be partitioned for training purpose and 30% of the data will be partitioned for testing purpose.
* In this project-

Train Data contains 4924 data points and Test Data contains 2108 data points

With 20 variables.

In this dataset, 4K+ customer records are used for training purpose and 2K+ records are used for testing purpose.

Testing Model, It is used to conduct testing on developed fit model.

* Developing Regression Model

Logistic Regression model, considering all relevant explanatory variables to predict response variable. Check the P-value, if variable having least p-values then it more accurate in predicting response variable.

Check AIC for every iterative model. The **Akaike information criterion** (**AIC**) is an [estimator](https://en.wikipedia.org/wiki/Estimator) of the relative quality of [statistical models](https://en.wikipedia.org/wiki/Statistical_model) for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for [model selection](https://en.wikipedia.org/wiki/Model_selection).

* Testing Model

Once model developed fully then perform testing on it. And validate it.

* Inferring data insights and business decisions.

Now model is ready to analyses the thing and make business decisions.

Steps to prepare fit Regression Model

* Loading Dataset into R environment.

Dataset with: Observations - 7043.

Variables - 20.

* Detecting and treating the missing values in dataset.

Here Total Charges contains missing values i.e. NA. But as only 11 missing values were there, so we preferred to remove those.

- Converting necessary variables into factors so that model can do well and predict the

the response variable very well.

* Splitting dataset into two model:

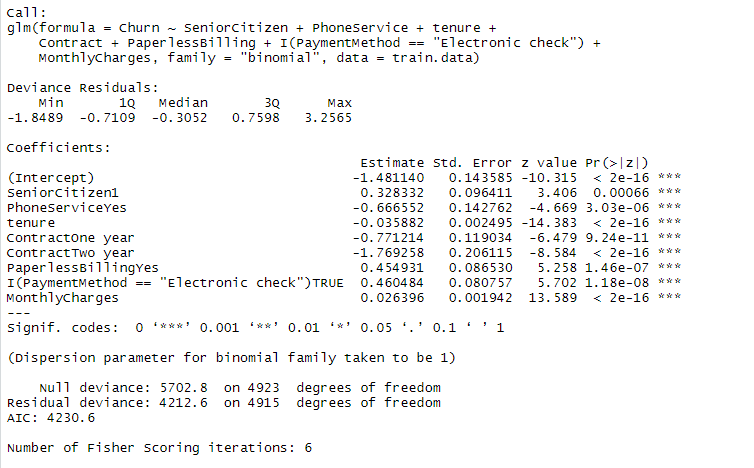
1. Development Model
2. Testing Model

* Preparing Final Regression Model

Initially provide all variables to model and observe the behavior. Check for least P-value and multicollinearity. Put variable like Senior Citizen, Phone Service, tenure, Contract, Paperless Billing, Electronic check Payment Method, Monthly Charge , these are contributing more to predict response variable i.e. Customer Churn or not.

Result and its Interpretation

Result for the best fit regression model:



Customer with contract of one year and two years are less likely to churn.

Company providing good phone service, that’s why customers with phone service are less likely to churn.

Customer with Paperless billing are likely to churn, so company need to improve their service in the area of paperless billing. Also the Payment method Electronic check, company needs to work on it, customers with payment method electronic check are more likely to churn.

Business Decisions after model analysis

* Customer with contract of one year and two years are less likely to churn.
* Company providing good phone service, that’s why customers with phone service are less likely to churn. But need to improve Internet Service Quality to retain more customer.
* Customer with Paperless billing are likely to churn, so company need to improve their service in the area of paperless billing. Also the Payment method Electronic check, company needs to work on it, customers with payment method electronic check are more likely to churn.
* Company need to consider different group of age while planning strategy. Senior citizens are more likely to churn.