MACHINE LEARNING ALGORITHMS TO PREDICT CUSTOMER CHURN IN A TELECOM COMPANY

Submitted by: Krishna Gopal Gowami

# Table of Contents

[Table of Contents 1](#_Toc527564310)

[PROJECT OBJECTIVE 2](#_Toc527564311)

[ABOUT DATA 2](#_Toc527564312)

[DATA PREPARATION 3](#_Toc527564313)

[MODELS 3](#_Toc527564314)

[Logistic Regression 3](#_Toc527564315)

[Naïve Bayes Classifier 3](#_Toc527564316)

[Random Forest Classifier 4](#_Toc527564317)

[Models Comparison 4](#_Toc527564318)

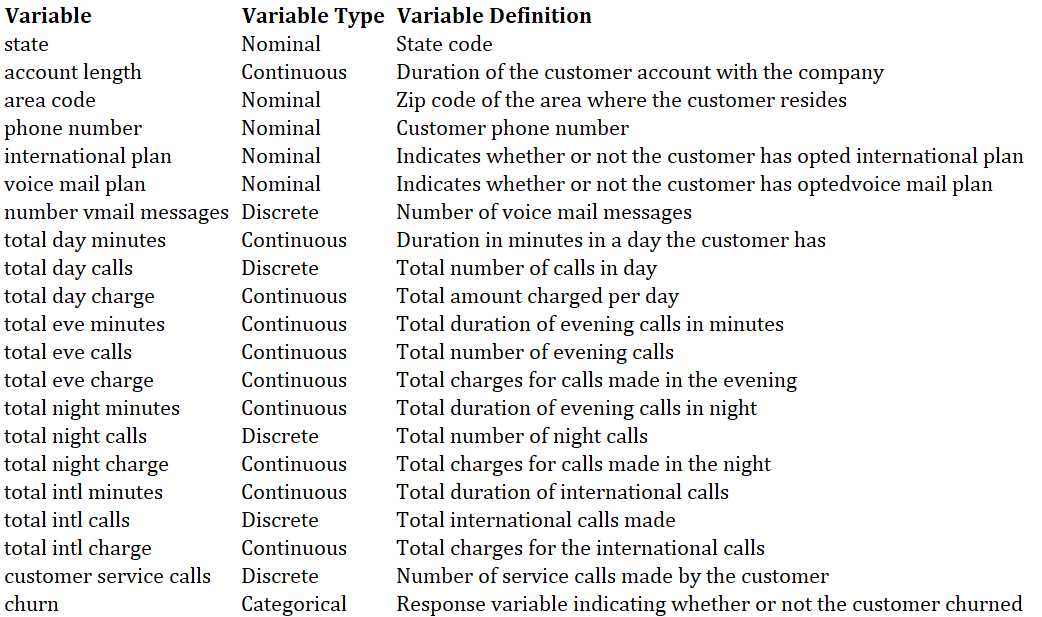
[GitHub Link 4](#_Toc527564319)

# PROJECT OBJECTIVE

The objective of the project is to fit machine learning algorithms to predict the customer churn. The model thus developed would enable the company to get early warning signs of customer churn and to take preventive actions for customer retention.

# ABOUT DATA

The data comprise of 21 variables with response variable being “Churn”. The data is a blend of demographic and customer life style depicting variables. The variable definition is as below:

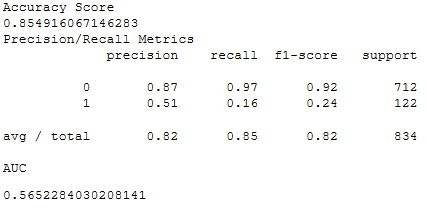


# DATA PREPARATION

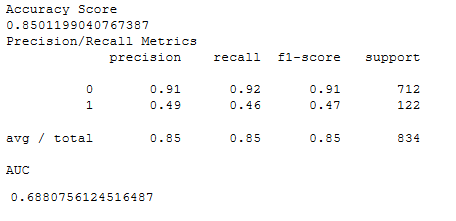
1. Missing value imputation
   1. No missing values found
2. Removal of unnecessary features
   1. The features “Area code” and “Phone number” do not add much value to the model, hence they can be omitted

# MODELS

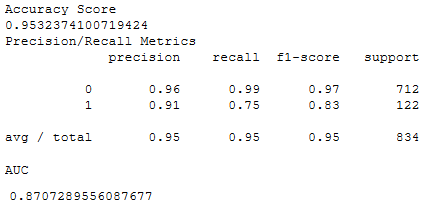
## Logistic Regression



## Naïve Bayes Classifier



## Random Forest Classifier



## Models Comparison

|  |  |  |  |
| --- | --- | --- | --- |
| # | Model | AUC | Model Fit |
| 1 | Logistic Regression | 0.57 | Average |
| 4 | Naïve Bayes Classifier | 0.69 | Descent |
| 6 | Random Forest Classier | 0.87 | Good |

## GitHub Link

<https://github.com/kgopal1982/Analytics/blob/master/CustomerChurnAnalyticsAngshu/CustomerChurnAnalysis.ipynb>