Customer Churn: Predicting whether a customer

will change telecommunications provider

A report by Team Mindbenders

* **Understanding the Problem Statement:**

Customer churning is a problem being faced by the telecom industry due to various reasons. Hence in order to retain the overall market share we have created a machine learning model which can predict the chances of a customer getting churned. Using the results telecom giants can target these set of customers in order to retain them.

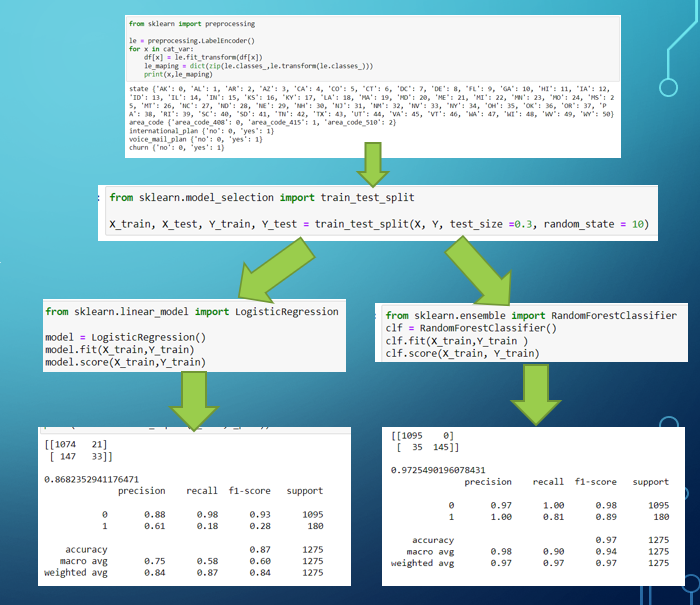
* **Detailed working:**

The task is to predict whether a customer will be churning by analyzing historic data.

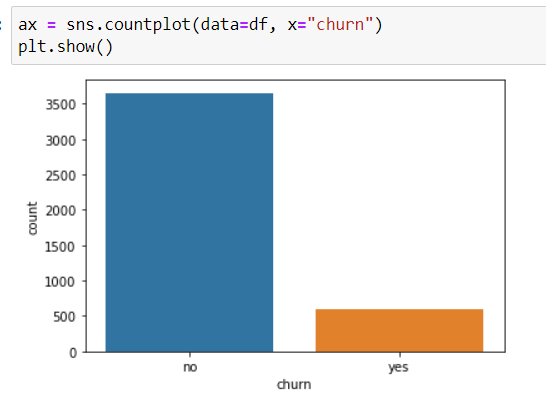
We took the data from Kaggle which considered of training and testing data but the test data was unlabeled hence we performed a train test split on the training data itself. Data consisted of 4250 rows and 20 columns. Data was clean with no missing values.

Our dependent variable is “churn” which is a categorical Boolean variable with yes and no as values.

1. Exploratory data analysis –
2. We ensured that there is no missing value in the dataset.
3. Deleted columns number\_vmail\_messages as most of the rows were 0.
4. Removed columns total\_minutes, total\_intl\_minutes to remove multicollinearity.
5. All other numeric variables had a normal distribution.
6. Feature engineering –
7. Made new columns total\_calls, total\_charge and total\_minutes by combining day, evening and night variables of each columns.
8. Created a variable account\_length\_cat which helped in visualizations. This variable is a categorical version of the account\_length.
9. Model Creation –
10. Label encoding – We have done label encoding to convert categorical variables to numeric variables.
11. Train test split – Since the test data was unlabelled we have splitted the training data in 7:3 ratio.
12. Logistic regression - Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.
13. Accuracy achieved – 86.8%
14. Random forest search - Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.
15. Accuracy achieved – 97.17%

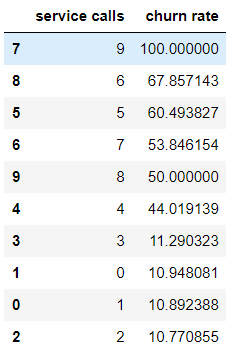


1. Uniqueness of the solution –

Table

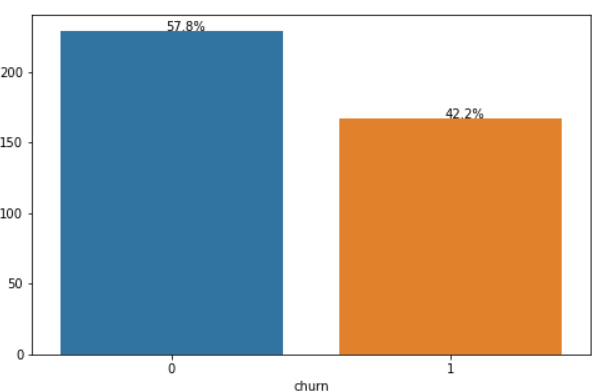
Description automatically generated

1. Label encoding – We have done label encoding to convert categorical variables to numeric variables.
2. Train test split – Since the test data was unlabelled we have splitted the training data in 7:3 ratio.
3. Logistic regression - Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.
4. Accuracy achieved – 86.8%
5. Random forest search - Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.
6. Accuracy achieved – 97.17%
7. How to retain churning customers (The way ahead):
8. Using our model telecom companies can first identify the set of churning customers to target for retention campaign.
9. Make sure that the customer is not facing a bad service issue. Bad services include less internet connectivity, rising costs, unstable network etc. This can also be focused in the top states that have the maximum churning and other important factors that are mentioned above.
10. Once the issue is identified the company should ensure that it is resolved.
11. Another strategy can be to offer incentives to customers from time to time and throw up long term offers. Some examples of long term offers include:
    * + Subscribing at an annual plan and getting 2 months of free subscription.
      + Cashbacks on monthly recharges.
      + Offering competitive pricing based on other service providers.
12. Visualizations and insights –
13. Once a users calls 4 times the churning is more than 44%.

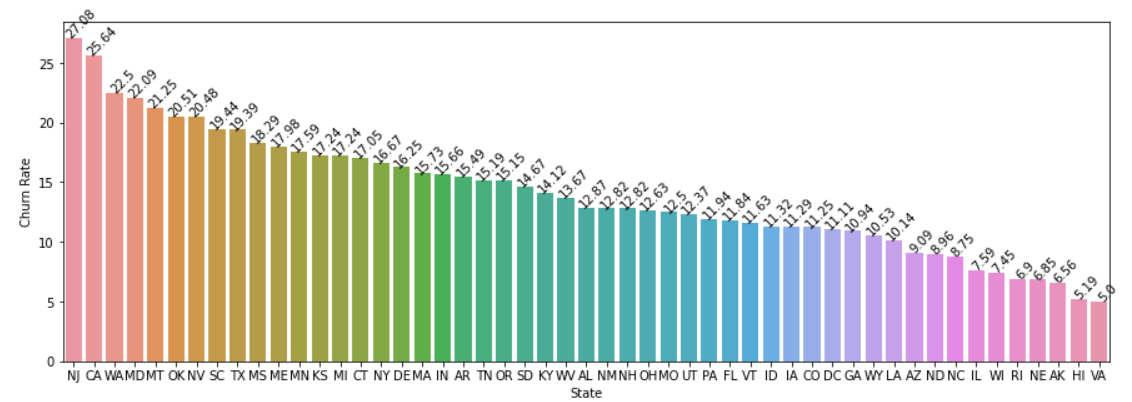


1. %people churning with an international plan

Findings - If a person has an international plan then he has a probablity of 42.2% of churning.



1. State wise churning rate



1. Number of churned customers vs total charges. Most of the customers churn when they cross a spend between 70 to 80 dollars

Chart, histogram

Description automatically generated