**Customer Churn Blog Submission**

Machine Learning is the word of the mouth for everyone involved in the analytics world. Gone are those days of the traditional manual approach of taking key business decisions. Machine Learning is the future and is here to stay.

However, the term Machine Learning is not a new one. It was there since the advent of computers but has grown tremendously in the last decade due to the massive amounts of data that’s getting generated, and the enormous computational power that modern-day system possesses.

Machine Learning is the art of Predictive Analytics where a system is trained on a set of data to learn patterns from it and then tested to make predictions on a new set of data. The more accurate the predictions are, the better the model performs. However, the metric for the accuracy of the model varies based on the domain one is working in.

Predictive Analytics has several usages in the modern world. It has been implemented in almost all sectors to make better business decisions and to stay ahead in the market. In this blog post, we would look into one of the key areas where Machine Learning has made its mark is the Customer Churn Prediction.

**What is Customer Churn?**

For any e-commerce business or businesses in which everything depends on the behavior of customers, retaining them is the number one priority for the organization. Customer churn is the process in which the customers stop using the products or services of a business.

Customer Churn or Customer Attrition is a better business strategy than acquiring the services of a new customer. Retaining the present customers is cost-effective, and a bit of effort could regain the trust that the customers might have lost on the services.

On the other hand, to get the service of the new customer, a business needs to spend a lot of time, and money on to the sales, and marketing department, more lucrative offers, and most importantly earning their trust. It would take more recourses to earn the trust of a new customer than to retain the existing one.

### ****Causes of Customer Churn?****

There is a multitude of reasons why a customer could decide to stop using the services of a company. However, a couple of such reasons overwhelms others in the market.

**Customer Service** – This is one of the most important aspects on which business the growth of a business depends. Any customer could leave the services of a company if it’s poor or doesn’t live up to the expectations. A study showed that nearly ninety percent of the customer leave due to poor experience as modern era deems exceptional services, and experiences.

When a customer doesn’t receive such eye-catching experience from a business, it tends to lean towards its competitors leaving behind negative reviews in various social media from their past experiences which also stops potential new customers from using the service. Another study showed that almost fifty-nine percent of the people aged between twenty-five, and thirty share negative client experiences online.

Thus, poor customer experience not only results in the loss of a single customer but multiple customers as well which hinders the growth of the business in the process.

**Onboarding Process** – Whenever the business is looking to attract a new customer to use their service, it is necessary that the on-boarding process which includes timely follow-ups, regular communications, updates about new products, and so on are being followed, and maintained consistently over a period of time.

### ****What Strategies could a Business Undertake to prevent Customer Churn?****

Customer Churn hinders or prevents the growth of an organization. Thus it is necessary that any business or organization has a flexible system in place to prevent the churn of customers and ensure its growth in the process. The companies need to find the metrics to identify the probability of a customer leaving, and chalk out strategies for improvement of its services, and products.

The calculation of the possibility of the customer churning varies from one business to another. There is no one fixed methodology that every organization uses to prevent churn. A churn rate could represent a variety of things such as – the total number of customers lost, the cost of the business loss, what percentage of the customers left in comparison to the total customer count of the organization, and so on.

Improving the customer experience should be the first strategy on the agenda of any business to prevent churn. Apart from that, marinating customer loyalty by providing better, personalized services is another important step one could undertake. Additionally, many organizations sent out customer survey time, and again to keep track of their customer experiences, and also seek reasons from them who have already churned.

A company should understand and learn about its customers beforehand. The amount of data that’s available all over the internet is sufficient to analyze a customer’s behavior, his likes, and dislikes, and improve the services based on their needs. All these measures, if taken with utmost care could help a business prevent its customers from churning

### ****Telecom Customer Churn Prediction****

Previously, we learned how Predictive Analytics is being employed by various businesses to prevent any event from occurring and reduce the chances of losing by putting the right system in place. As customer churn is a global issue, we would now see how Machine Learning could be used to predict the customer churn of a telecom company.

The data set could be downloaded from here – [Telco Customer Churn](https://www.kaggle.com/blastchar/telco-customer-churn)

The columns that the dataset consists of are –

* **Customer Id** – It is unique for every customer
* **Gender** – Determines whether the customer is a male or a female.
* **Senior Citizen** – A binary variable with values as 1 for senior citizen and 0 for not a senior citizen.
* **Partner** – Values as ‘yes’ or ‘no based on whether the customer has a partner.
* **Dependents** – Values as ‘yes’ or ‘no’ based on whether the customer has dependents.
* **Tenure** – A numerical feature which gives the total number of months the customer stayed with the company.
* **Phone Service** – Values as ‘yes’ or ‘no’ based on whether the customer has phone service.
* **Multiple Lines** – Values as ‘yes’ or ‘no’ based on whether the customer has multiple lines.
* **Internet Service** – The internet service providers the customer has. The value is ‘No’ if the customer doesn’t have internet service.
* **Online Security** – Values as ‘yes’ or ‘no’ based on whether the customer has online security.
* **Online Backup** – Values as ‘yes’ or ‘no’ based on whether the customer has online backup.
* **Device Protection** – Values as ‘yes’ or ‘no’ based on whether the customer has device protection.
* **Tech Support** – Values as ‘yes’ or ‘no’ based on whether the customer has tech support.
* **Streaming TV** – Values as ‘yes’ or ‘no’ based on whether the customer has a streaming TV.
* **Streaming Movies** – Values as ‘yes’ or ‘no’ based on whether the customer has streaming movies.
* **Contract** – This column gives the term of the contract for the customer which could be a year, two years or month-to-month.
* **Paperless Billing** – Values as ‘yes’ or ‘no’ based on whether the customer has a paperless billing.
* **Payment Method** – It gives the payment method used by the customer which could be a credit card, Bank Transfer, Mailed Check, or Electronic Check.
* **Monthly Charges** – This is the total charge incurred by the customer monthly.
* **Total Charges** – The value of the total amount charged.
* **Churn** – This is our target variable which needs to be predicted. Its values are either Yes (if the customer has churned), or No (if the customer is still with the company)

The following steps are the walkthrough of the code which we have written to predict the customer churn.

* First, we have imported all the necessary libraries we would need to proceed further in our code
* Just to get an idea of how our data looks likes, we have read the CSV file and printed out the first five rows of our data in the form of a data frame
* Once, the data is read, some pre-processing needed to be done to check for null, outliers, and so on
* Once the pre-processing is done, the next step is to get the relevant features to use in our model for the prediction. For that, we have done some data visualization to find out the relevancy of each predictor variables
* After the data has been plotted, it is observed that Gender doesn’t have much influence on churn, whereas senior citizens are more likely to leave the company. Also, Phone Service has more influence on Churn than Multiple Lines
* A model cannot take categorical data as input, hence those features are encoded into numbers to be used in our prediction
* Based on our observation, we have taken the features which have more influence on churn prediction
* The data is scaled, and split it into train and test set
* We have fitted the Random Forest classifier to our new scaled data
* Predicted the result and using the confusion matrix as the metric for our model
* The model gives us (1155 + 190 = 1345) correct predictions and (273 + 143 = 416) incorrect predictions

The entire code could be found in this GitHub [link](https://github.com/deysuman42/Data-Science-Projects/blob/Telco-Customer-Churn-Prediction/telco_customer_churn.ipynb)

### ****Conclusion****

We have built a basic Random Forest Classifier model to predict the Customer Churn for a telecom company. The model could be improved with further manipulation of the parameters of the classifier and also by applying different algorithms.