BLOG ON CUSTOMER CHURN ANALYSIS

1-Problem

Customer churn is when a company’s customers stop doing business with that company. Businesses are very keen on measuring churn because keeping an existing customer is far less expensive than acquiring a new customer. New business involves working leads through a sales funnel, using marketing and sales budgets to gain additional customers. Existing customers will often have a higher volume of service consumption and can generate additional customer referrals.

Customer retention can be achieved with good customer service and products. But the most effective way for a company to prevent attrition of customers is to truly know them. The vast volumes of data collected about customers can be used to build churn prediction models. Knowing who is most likely to defect means that a company can prioritise focused marketing efforts on that subset of their customer base.

Preventing customer churn is critically important to the telecommunications sector, as the barriers to entry for switching services are so low.You will examine customer data from IBM Sample Data Sets with the aim of building and comparing several customer churn prediction models.

2-Data Analysis

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

What are the 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.

**On boarding 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.

3-EDA

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)

4-Prediction Model

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 Logistic Regression to our new scaled data

Predicted the result and using the confusion matrix as the metric for our model.

5- Conclusion

Logistic Regression model was used for the prediction,because it had the highest accuracy score of more than 80%.Based on which the model could predict whether a employee would churn the job or not, while keeping all the other factors in point.