# Customer retention report

Problem Statement

The dataset is related to customer satisfaction, that leads to the success of online stores. A comprehensive review has been carried out to purpose a model for customer satisfaction and customer retention. And also the dataset shows, which factors influences the customer repeat purchase intention i.e customer retention.

The source of this dataset are some Indian online shoppers, like Myntra, Amazon, Flipkart etc.

Results indicate the e-retail success factors, which are very much important for customer satisfaction and customer retention.

This is a multiclass classification problem.

First, we import all the necessary libraries. Then path of the dataset: r'C:\Users\JAYASHREE\Downloads\Customer\_retention\_dataset\customer\_retention\_dataset.xlsx' is loaded in the jupyter notebook.

There are 269 rows and 71 columns in the dataframe.

Columns: 1Gender of respondent

2 How old are you?

3 Which city do you shop online from?

4 What is the Pin Code of where you shop online from?

5 Since How Long You are Shopping Online ?

6 How many times you have made an online purchase in the past 1 year?

7 How do you access the internet while shopping on-line?

8 Which device do you use to access the online shopping?

9 What is the screen size of your mobile device?

10 What is the operating system (OS) of your device?

11 What browser do you run on your device to access the website?

12 Which channel did you follow to arrive at your favorite online store for the first time?

13 After first visit, how do you reach the online retail store?

14 How much time do you explore the e- retail store before making a purchase decision?

15 What is your preferred payment Option?

16 How frequently do you abandon (selecting an items and leaving without making payment) your shopping cart?

17 Why did you abandon the “Bag”, “Shopping Cart”?

18 The content on the website must be easy to read and understand

19 Information on similar product to the one highlighted is important for product comparison

20 Complete information on listed seller and product being offered is important for purchase decision.

21 All relevant information on listed products must be stated clearly

22 Ease of navigation in website

23 Loading and processing speed

24 User friendly Interface of the website

25 Convenient Payment methods

26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time

27 Empathy (readiness to assist with queries) towards the customers

28 Being able to guarantee the privacy of the customer

29 Responsiveness, availability of several communication channels (email, online rep, twitter, phone etc.)

30 Online shopping gives monetary benefit and discounts

31 Enjoyment is derived from shopping online

32 Shopping online is convenient and flexible

33 Return and replacement policy of the e-tailer is important for purchase decision

34 Gaining access to loyalty programs is a benefit of shopping online

35 Displaying quality Information on the website improves satisfaction of customers

36 User derive satisfaction while shopping on a good quality website or application

37 Net Benefit derived from shopping online can lead to users satisfaction

38 User satisfaction cannot exist without trust

39 Offering a wide variety of listed product in several category

40 Provision of complete and relevant product information

41 Monetary savings

42 The Convenience of patronizing the online retailer

43 Shopping on the website gives you the sense of adventure

44 Shopping on your preferred e-tailer enhances your social status

45 You feel gratification shopping on your favorite e-tailer

46 Shopping on the website helps you fulfill certain roles

47 Getting value for money spent

From the following, tick any (or all) of the online retailers you have shopped from;

Easy to use website or application

Visual appealing web-page layout

Wild variety of product on offer

Complete, relevant description information of products

Fast loading website speed of website and application

Reliability of the website or application

Quickness to complete purchase

Availability of several payment options

Speedy order delivery

Privacy of customers’ information

Security of customer financial information

Perceived Trustworthiness

Presence of online assistance through multi-channel

Longer time to get logged in (promotion, sales period)

Longer time in displaying graphics and photos (promotion, sales period)

Late declaration of price (promotion, sales period)

Longer page loading time (promotion, sales period)

Limited mode of payment on most products (promotion, sales period)

Longer delivery period

Change in website/Application design

Frequent disruption when moving from one page to another

Website is as efficient as before

Which of the Indian online retailer would you recommend to a friend?

Using info() method, we see that there are no null values. Hence, we are good to proceed further.

Next, we have to encode the data, as data are stored in string format. So, we need to encode data into int. after encoding data is stored in another excel file.

Data.describe() method gives the summary statistics of the dataset.

Then we check for the correlation between the columns of the dataset using data.corr() . It explains to what extent all the columns are related to each column.

We found that, the features like, clear and complete relevant information about the product, easy to navigate the application, a good web browser, user-friendly interface of the application, online shopping is convenient and flexible, net benefits, gaining loyalty program, the trust factor, privacy of the customer, the empathy, availability of assistance, availability of several payment options, complete and relevant information of products, speedy order delivery, return and replacement policy, monetary benefits and discounts etc gives much towards the customer retention and satisfaction, that further lead to success of online shopping.

Next step is the visualization. I have use countplot to analyse the categorial features graphically. From this analysis, we get which group has the most numbers, hence got a hint of classification.

Also, I think the dataset is sometimes imbalanced.

These are the findings. Next step includes train-test-split data for training and testing purpose. After that we can apply different algorithms for this multiclass classification problem.