1. INTRODUCTION

1.1 Introduction & Background of Project

Customer satisfaction is considered as one of the most important characteristic that influences the customer loyalty, purchase, repurchase intentions and determines the success of any e-commerce website. Apart from customer satisfaction there five more characteristics such as service quality, system quality, information quality, trust and net benefit that contribute to the success of online store

A research was performed to investigate factors that directly influence customer's action, leading to repeated purchase from an e-commerce website. The data that was obtained from the Indian online customers showed that the utilitarian value and hedonistic values play crucial role in repeated purchase; further signifying the customer loyalty toward the website.

1.2 Problem Statement & Objective

This report aims to find the various parameters, that play crucial role in the success of the ecommerce business.

2. Research Design & Methodology

2.1 Data Collection

The dataset contains 71 columns with a unique set of parameters, each capturing the user's choice based on their experience. The following parameters were used to gather the information from 269 individuals.

1. Gender 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
48. From the following, tick any (or all) of the online retailers you have shopped from;
49. Easy to use website or application
50. Visual appealing web-page layout
51. Wild variety of product on offer
52. Complete, relevant description information of products
53. Fast loading website speed of website and application
54. Reliability of the website or application
55. Quickness to complete purchase
56. Availability of several payment options
57. Speedy order delivery
58. Privacy of customers' information
59. Security of customer financial information
60. Perceived Trustworthiness
61. Presence of online assistance through multi-channel
62. Longer time to get logged in (promotion, sales period)
63. Longer time in displaying graphics and photos (promotion, sales period)
64. Late declaration of price (promotion, sales period)
65. Longer page loading time (promotion, sales period)
66. Limited mode of payment on most products (promotion, sales period)
67. Longer delivery period
68. Change in website/Application design
69. Frequent disruption when moving from one page to another
70. Website is as efficient as before

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

All the data contained within this dataset are of object type.

2.2 Exploratory Data Analysis (EDA)

Exploratory data analysis is a process used to discover patterns and trends in the dataset by summarizing the information using statistical table and visual graphs.

2.2.1 Peeking inside dataset

The first step of EDA is to read and understand the dataset provided using pandas module; this involves getting insight of the raw data, determining its shape and analyzing the data type of each individual attribute.

2.2.2 Null Value Analysis

When the data provided in one or more items, or in entire unit is missing, it creates a void which has to be handled immediately before processing the data. This can be done by either removing the entire column if more than 30% of the data in the column is missing or by replacing the null values.

2.2.3 Data Visualization

Data Visualization is a technique used to analyze the data in the form of graphs or maps, so that it can be easy to understand and interpret the underlying data trends or patterns. It can be classified into three major categories which includes–

a) **Univariate analysis** - In this type of analysis, only a single variable is used to analyze the data. It is the easiest form of analysis, since it does test any relationship with other variables.

b) **Bi-Variate analysis** - In this type of analysis, two different variables that has relationship with each other are used to determine how they affect one another.

c) **Multi-Variate analysis** - When the analysis involves three or more variables, it is classified as multivariate.

2.2.4 Encoding Categorical Variables

This process involves converting the features of the dataset that are categorical (data that can be divided into group), into a set of binary variable that can be easily understood by the machine learning model. There are two distinct typs of categorical variables namely

a) **Ordinal data** - Categorical data that can be ranked.
b) **Nominal data** - Categorical data that cannot be ranked.

In this problem all the features in the dataset are categorical in nature and hence there is need to perform encoding based on the analysis.

2.3 Separating Input Data and Target Data

After the data is encoded, the dataset is divided into two halves; one half containing all the features of the dataset that will be used as an input for the machine learning model while the other half containing only the target/output variable.

Based on the objective of this project, 71st column ("Which of the Indian online retailer would you recommend to a friend?") of the dataset is selected as the target variable, since it is assumed that the customer will recommend the shopping website, only if they themselves are the frequent user of that platform , thus indicating the customer retention capability of the website. Remaining features are then used as an input for the model.

2.4 Class Distribution

Number of samples that belong to each class is referred to as class distribution. Whenever there is a bias or imbalance in this distribution, a sampling technique is used to balance the class; this can be done either using –

a. **Undersampling** - This technique deletes the rows of the majority class to balance the distribution and hence it is used only when there is more than sufficient amount of data.

b. **Oversampling** – This technique creates duplicate rows of minority class to balance the distribution and hence it is used there is less than sufficient amount of data.

2.5 Feature Scaling

Feature scaling is a technique frequent used in machine learning preprocessing to generalize the data points so that the distance between them is smaller. Two of the most widely used feature scaling techniques are –

a) **Min-Max Normalization** – It uses distribution value between 0 and 1 to re-scales all the feature values.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

b) **Standardization** - It makes the distribution mean=0 and variance=1 while re-scaling the feature values.

$$X_{new} = \frac{X_i - X_{mean}}{\text{Standard Deviation}}$$

2.7 SelectKBest

SelectKBest method with chi square as the function parameter is used to extract the features based on the K highest score, where K is the number of top features that are to be selected.

3. Conclusion

It was observed during the analysis, that the following top features extracted using SelectKBest method played a crucial role in customer retention.

index	Column Name	Score
16	"17 Why did you abandon the "Bag" "Shopping Cart"?	193.18328584995254
12	"13 After first visit how do you reach the online retail store?"	151.67201468427623
9	10 What is the operating system (OS) of your device?	146.55736456520276
42	43 Shopping on the website gives you the sense of adventure	142.86627466779575
10	11 What browser do you run on your device to access the website?	140.65821778865256
22	23 Loading and processing speed	139.50949533429676
69	Website is as efficient as before	137.87281579298389
18	19 Information on similar product to the one highlighted is important for product comparison	137.35805238246385
45	46 Shopping on the website helps you fulfill certain roles	131.51807380243082
28	"29 Responsiveness availability of several communication channels (email , online rep, twitter, phone etc.)"	120.37373824749625

46	47 Getting value for money spent	111.52399726142951
62	"Longer time in displaying graphics and photos (promotion, sales period)"	110.60301832921972
57	Privacy of customers' information	107.19978342147866