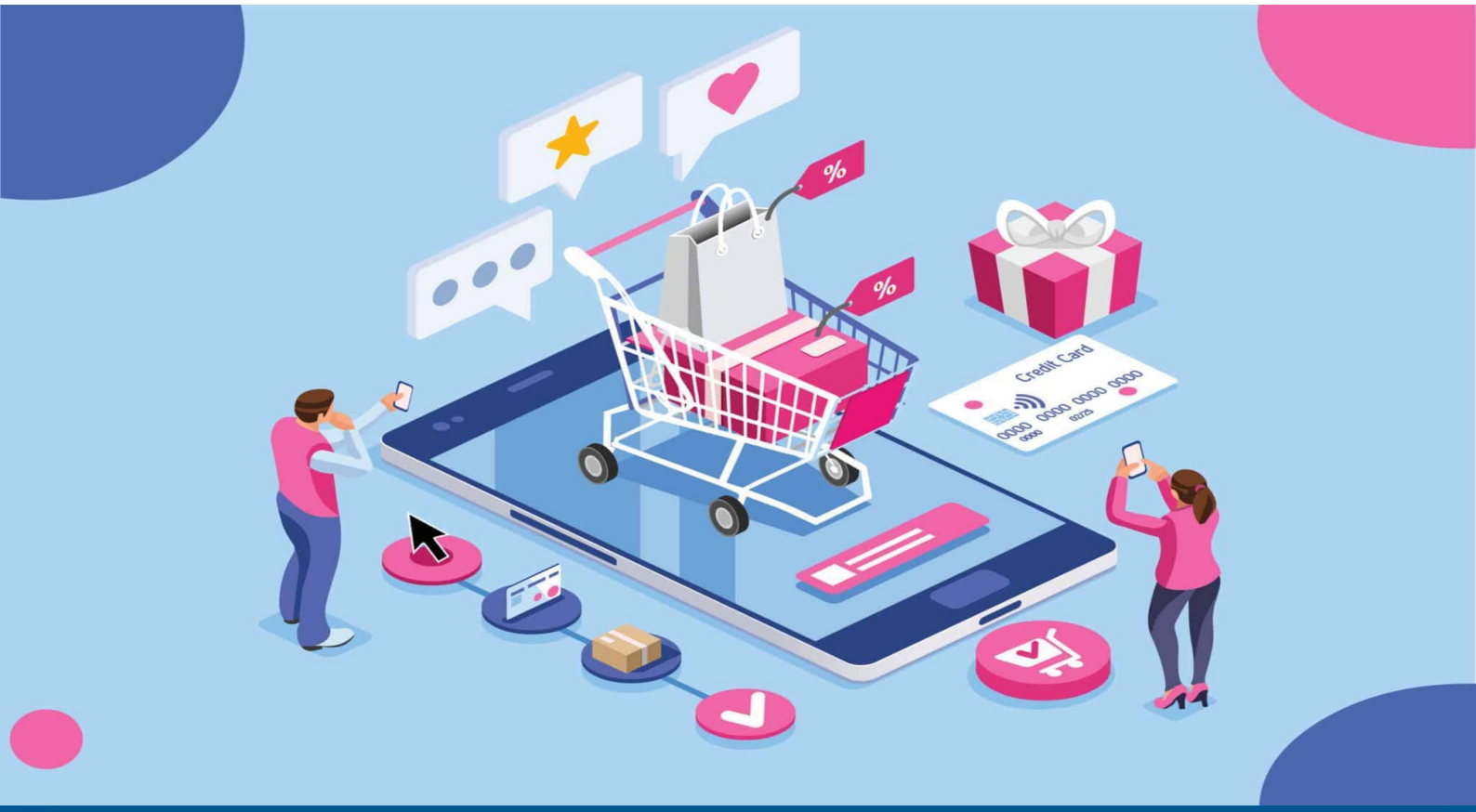
Flip Robo Technologies

Customer Retention – Project Report

Presented by Naveen Kumar Ranganathan



Introduction:

Globally customers are getting attracted towards online retailing; this is because e-stores usually offer them a variety of services and products according to their preferences. Convenience, round the clock availability, flexible pricing, discounts as well as free door step delivery are some of the major benefits of shopping online. Presently, more number of online retailers are beginning to experience increase in demand for products and service.

Indian online retail industry has been experiencing good times, as a result of the constantly growing internet penetration, deployment of modern infrastructures, and a robust ecosystem for e-retail start-ups. Several e-commerce start-ups have commenced operation with innovative strategies, which differs from what was pioneered by first generation e-commerce companies.

A number of home-grown small e-retailers are struggling to make a good market share, while the big players are also facing a rough competition. Correspondingly, the decision making by these e-retailers becomes very much important in terms of customer activation and retention

Hence, there is stringent requirement of research on the dynamics of the Indian online retail industry and coming out with the models for successful development of this industry.

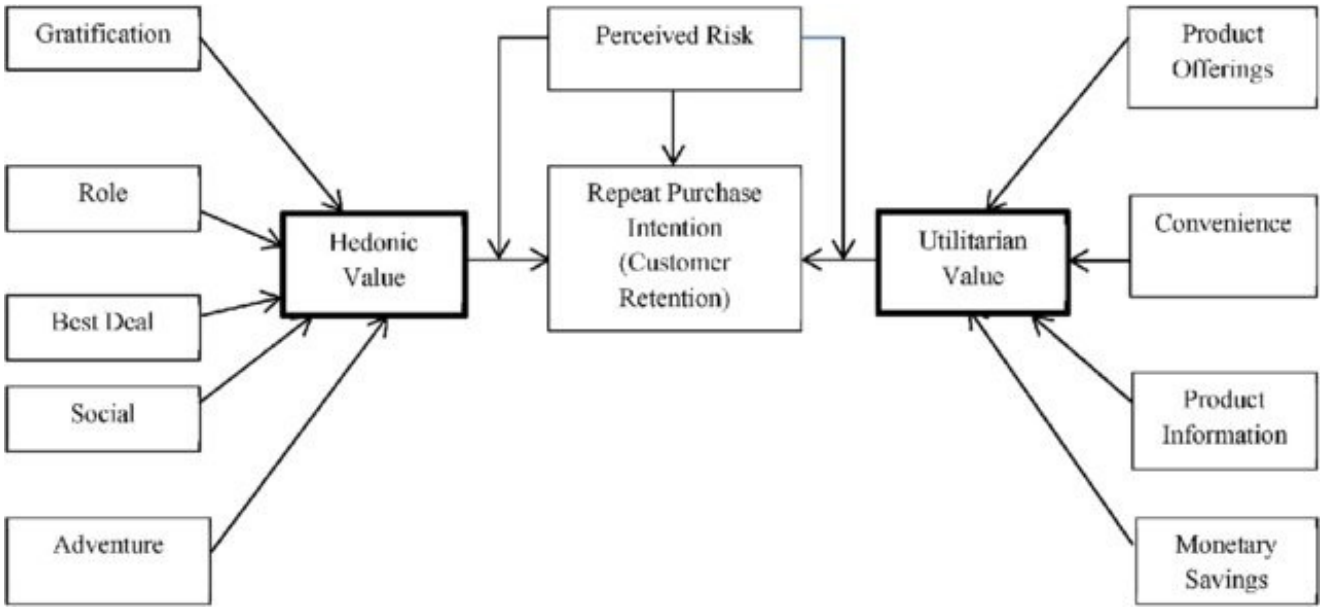
Case Study:

E-retail factors for customer activation and retention: A case study from Indian e-commerce customers

Customer satisfaction has emerged as one of the most important factors that guarantee the success of online store; it has been posited as a key stimulant of purchase, repurchase intentions and customer loyalty. A comprehensive review of the literature, theories and models have been carried out to propose the models for customer activation and customer retention. Five major factors that contributed to the success of an e-commerce store have been identified as: service quality, system quality, information quality, trust and net benefit. The research furthermore investigated the factors that influence the online customers repeat purchase intention. The combination of both utilitarian value and hedonistic values are needed to affect the repeat purchase intention (loyalty) positively. The data is collected from the Indian online shoppers. Results indicate the e-retail success factors, which are very much critical for customer satisfaction.

Prediction:

We need to analyse the data to predict the E-retail factors for customer activation and retention



Getting insights about the dataset:

The data's are collected by questionnaire method. In which customer's had responded to 71 questions by ratings and preferred online retailer based on service quality, system quality, information quality, trust and net benefit.

Analysis:

- We have total of 269 rows and 71 columns in customer retention dataset.
- Out of which 70 columns datatypes are in object types (string value) and column pincode is in integer type.
- As our dataset is categorical, We can consider the dataset as supervised classification data.
- We can consider the column 'Which of the Indian online retailer would you recommend to a friend?' as our target column
- We dont find any null values in our dataset

Data Visualization:

For visualizing the dataset countplot has been used to find out the count of the values

- Based on the customer response we can identify Amazon was the highly recommended E-Retailer among customers followed by Flipkart
- Female customers are high when compared to male and both of them recommended Amazon and Flipkart as their online retailer
- 85% of the customers are from age category of 21-30 years, 31-40 years and 41-50 years
- Delhi tops the city list and Bulandshahr is in bottom of city list
- Customers who are shopping online for above 4 years uses amazon and flipkart
- 42% of customer purchases less than 10 times a year
- 52% of customer uses smartphone to access the internet for shopping on-line
- Google chrome was the highly used search engine
- 62% of the customers spend more than 11mins to explore the e- retail store before making a purchase decision
- 55% customers use credit/debit cards for payment and 63% of customers abandon frequently without making payment in that 49% customer finds better alternative offer

Feature Selection and Engineering:

Encoding:

Since dataset has a lot of string values. Encoding has been applied to convert the string data to numerical

Correlation:

In feature selection used correlation method to identify the positive and negative correlated values

Below are Negative correlated

- Enjoyment is derived from shopping online
- Gaining access to loyalty programs is a benefit of shopping online

Below are Positive correlated

- Presence of online assistance through multi-channel
- Easy to use website or application
- Reliability of the website or application
- Complete, relevant description information of products

Describe:

Applied describe method to check standard deviation, median, IQR percentile etc

- Mean is greater than median(50%) in most of the columns which shows data is skewed on right side
- Pincode 75% to max value difference is high which means outliers are present

Skew:

Applied skew method to check the range of skewness

- Keeping +/- 0.5 as the range for skewness, there are more number of columns which does not lie within the range

Since data are categorial not applying transformation methods

Zscore:

Z Score Method to detect and remove Outliers

- After applying Zscore, data loss percentage is 7.4%
- Our data is becoming biased as it is not considering the case of pincode which is the only column with integer value

Moreover our dataset is categorical, we will not remove any outliers and consider the original data

Variance Inflation Factor:

Variance Inflation Factor is used for Data cleansing, Feature Engineering.

- Our dataset is Multicolinearity which means multiple input factor which is affecting the Y Output.
- Using this method we found most of the columns VIF factor are infinity

Since our dataset is categorical we cannot remove columns based on VIF value.

Final Analysis:

We can use the original data for Machine learning Model, as our dataset is categorical and removing outliers and applying methods for skewness is not appropriate in prediciting or developing the model

For Machine learning Model we should use logistic regression, since our dataset is classification.

Conclusion:

Majority of the customers agree to the factor that the websites must be easy to read and understand, similar product to the one highlighted is important for product comparison, information on listed seller and product is important for purchase decision

When purchasing a product from online customers are majorly looking into these factors Ease of navigation in website, Loading and processing speed, Convenient Payment methods, Being able to guarantee the privacy of the customer Return and replacement policy.

87% of customers trust that the online retail store will fulfill its part of the transaction at the stipulated time and 97% of customers satisfaction improves when displaying quality Information on the website, while 80% User derive satisfaction when shopping on a good quality website or application