FLIP ROBO

E-retail Factors for Customer Activation and Retention: A case study from Indian E-Commerce Customers

Submitted by:

Shaik Mahaboob Basha

ACKNOWLEDGMENT

These below references, research papers, data sources, professionals and other resources that helped me and guided me in completion of the project:

(PDF) E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers (researchgate.net)

https://www.datacamp.com/community/tutorials/exploratory-data-analysis-python

https://stackoverflow.com/questions/43962735/creating-barplots-using-for-loop-using-pandas-matplotlib

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

https://www.geeksforgeeks.org/graph-plotting-in-python-set-1/

https://www.machinelearningplus.com/plots/matplotlib-tutorial-complete-guide-python-plot-examples/

https://www.researchgate.net/publication/346412647_E-retail_factors_for_customer_activation_and_retention_An_empirical_study_from_Indian_e-commerce_customers

I would also like to thank My SME Mr. Shubham Yadav and FlipRobo Team for helping in completion of project.

INTRODUCTION

- Business Problem Framing

  E – Retailing Factors and Preferences that helps in customer retention and customer activation…

- Conceptual Background of the Domain Problem

  This project is based on the preference of people while doing online shopping, in terms of the website, payment mode, age-group preferences, city wise, discount wise etc. The end goal of the project to conclude a website is likely to offer a discount/offer for shopping to male / female basis their preference and an estimated number of male and female customers are more likely to retained and which are not.

- Review of Literature

  To run a successful online retail sore, emphasis must be given to the factors necessary for user satisfaction. However, these factors may differ among the customers as per their preferences. From the proposed and tested models: information, system and service qualities prove to be the well-established factors necessary for online business users' satisfaction.

  Nevertheless, e-tailers are required to integrate other benefits as well to create value to the customer. Empirically tested models have shown the importance of the quality of system, information, accessibility, security, and ease of use are needed to appeal to the consumer's utilitarian values. On the other hand, to appeal to the hedonistic values of the consumer, the e-store should incorporate elements that would offer value to the pleasure-seeking behaviour of customers; for example: adventure, gratification, role shopping etc. As to what motivates a customer's repurchase intention, the outcome of the study indicates that a combined approach designed to increase customer satisfaction, trust, and the overall net benefits of shopping on their e-store must be adopted to achieve their goals. Online retailers must focus on the information quality and system quality of the online retail websites to enhance customer satisfaction, which in turn will lead to repurchase decision. Both the utilitarian and hedonistic values led to customer satisfaction and considerably stimulate their repurchase intention, leading to loyalty. The result further suggests that customers' perception of utilitarian and hedonic values will inform their preference for a particular online store.

- Motivation for the Problem Undertaken.

Indian online retail industry has been experiencing good times since the last six years; as a result of the constantly growing internet penetration, deployment of modern infrastructures, and a robust ecosystem for e-retail start-ups. (Deloitte, 2018). Several e-commerce start-ups have commenced operation with innovative strategies, which differs from what was pioneered by first generation e-commerce companies. "More than 1200 start-ups came up in 2018, including eight unicorns, taking the total number to 7200 in India" (NASSCOM, 2019). India's B2C e-commerce revenue grew from €20 billion in 2017 to reach €25 billion in 2018 at a growth rate of 20% (Asendia, 2019). The number of internet users in India has been growing at a CAGR of 35% since the year 2007 according to a report by IAMAI-IMRB (2017). Having grown from just over a 100 million internet users in 2010, India has since touched 500 million internet users in 2018. Online retail businesses are positioning themselves to be able to take advantage of the massive internet user base by turning them into online shoppers. Indian e-commerce market is poised to surpass that of the United States, making it the second largest in the world in less than twenty years from now (PwC India, 2018).

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem
    a. Yeo-Johnson is used to scaling of the data.
    b. Logical Regression is used as statistical model.
    c. SVM, Random Forest, Decision Tree, Gradient Classifier are used to find out the accuracy.
    d. Confusion Metrix's, Classification Report are carried out while working on the model.

- Data Sources and their formats

	1Gender of respondent	2 How old are you?	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? \t\t\t\t\t	10 What is the operating system (OS) of your device? \t\t\t	11 What browser do you run on your device to access the website? \t\t\t	12 Which channel did you follow to arrive at your favorite online store for the first time?	...	38 User satisfaction cannot exist without trust	39 Offering a wide variety of listed product in several category
count	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	...	269.000000	269.000000
mean	0.669145	2.959108	3.524164	2.672862	3.260223	1.676580	4.282528	1.776952	1.275093	1.360595	...	4.182156	4.148699
std	0.471398	1.066012	1.436586	1.651788	1.135887	0.843904	0.923426	0.797892	0.645429	0.897805	...	1.072162	0.842110
min	0.000000	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000	1.000000	1.000000	1.000000	...	1.000000	2.000000
25%	0.000000	2.000000	3.000000	1.000000	2.000000	1.000000	4.000000	1.000000	1.000000	1.000000	...	4.000000	4.000000
50%	1.000000	3.000000	4.000000	2.000000	3.000000	1.000000	4.000000	2.000000	1.000000	1.000000	...	4.000000	4.000000
75%	1.000000	4.000000	5.000000	4.000000	5.000000	2.000000	5.000000	2.000000	1.000000	1.000000	...	5.000000	5.000000
max	1.000000	5.000000	5.000000	5.000000	5.000000	4.000000	5.000000	3.000000	4.000000	4.000000	...	5.000000	5.000000

8 rows x 45 columns

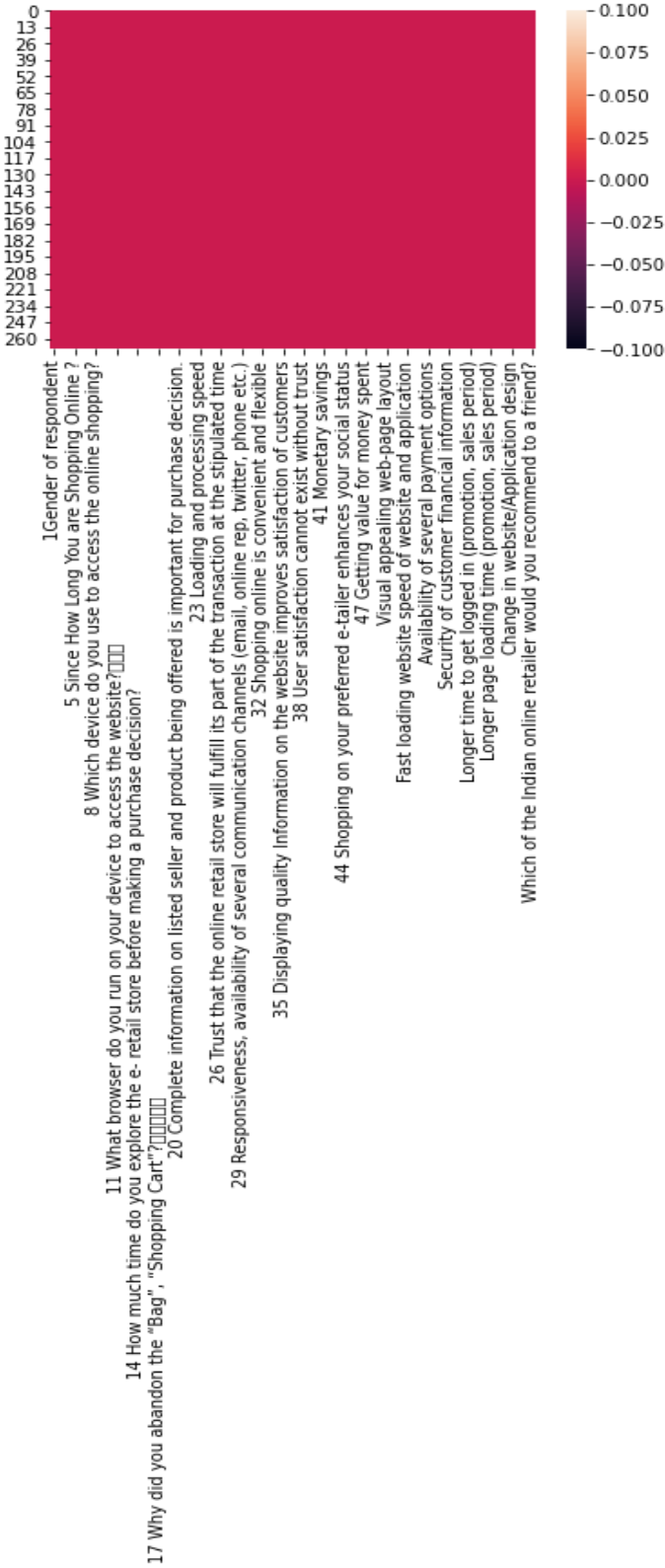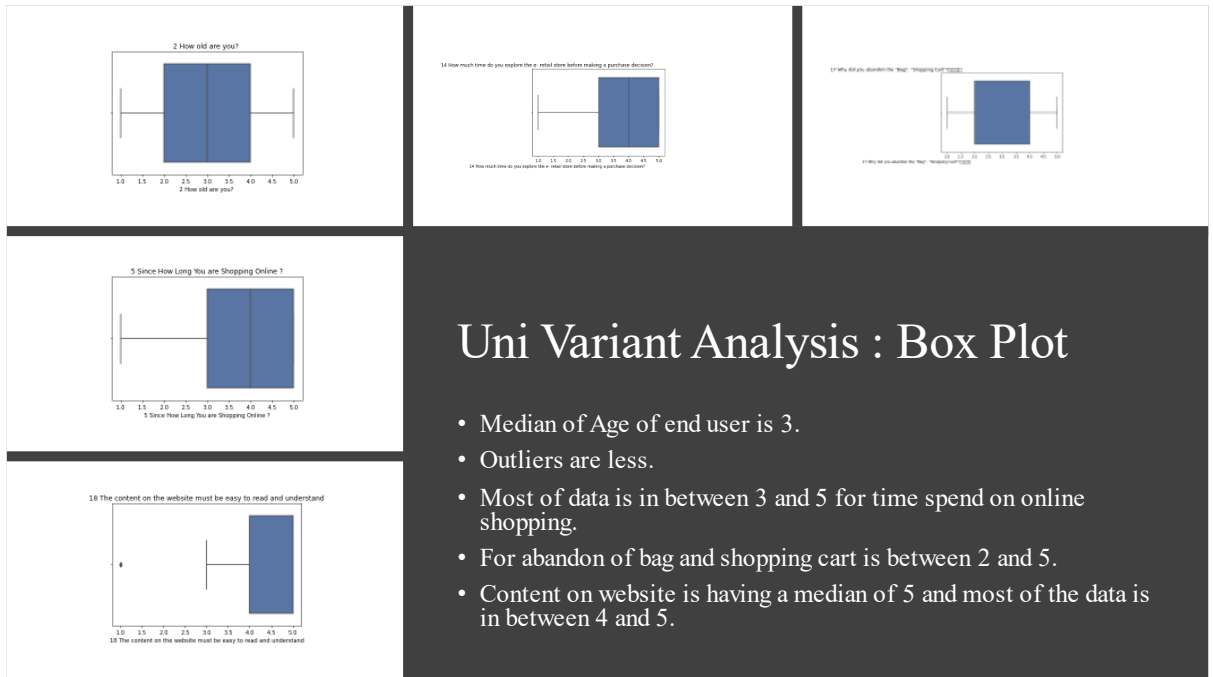| 40 Provision of complete and relevant product information	41 Monetary savings	42 The Convenience of patronizing the online retailer	43 Shopping on the website gives you the sense of adventure	44 Shopping on your preferred e-tailer enhances your social status	45 You feel gratification shopping on your favorite e-tailer	46 Shopping on the website helps you fulfill certain roles	47 Getting value for money spent
269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000	269.000000
4.349442	4.263941	3.914498	3.553903	3.223048	3.501859	3.282528	4.163569
0.755953	1.000485	0.693879	1.065869	1.219581	1.141564	1.178995	0.648773
2.000000	2.000000	3.000000	1.000000	1.000000	1.000000	1.000000	3.000000
4.000000	4.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000
5.000000	5.000000	4.000000	4.000000	3.000000	3.000000	3.000000	4.000000
5.000000	5.000000	4.000000	4.000000	4.000000	4.000000	4.000000	5.000000
5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

- Data Pre-processing Done

```
In [13]:   1  #Checking Presence of NaN Values
           2
           3  df.isnull().values.any()S

Out[13]:  False
```

Uni Variant Analysis : Box Plot

- Median of Age of end user is 3.
- Outliers are less.
- Most of data is in between 3 and 5 for time spend on online shopping.
- For abandon of bag and shopping cart is between 2 and 5.
- Content on website is having a median of 5 and most of the data is in between 4 and 5.

- Data Inputs- Logic- Output Relationships

  The Customer Retention Analysis dataset contains a target column and the target column is the gender column and the rest columns are categorical in nature.

- Hardware and Software Requirements and Tools Used

```
In [66]: #Importing Libraries

import numpy as np
import pandas as pd
import sklearn
import seaborn as sb
import matplotlib.pyplot as plt

from sklearn.linear_model import LogisticRegression as lg
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix,classification_report
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB

from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier


import warnings
warnings.filterwarnings('ignore')
```

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)
    a. Yeo-Johnson is used to scaling of the data.
    b. Logical Regression is used as statistical model.
    c. SVM, Random Forest, Decision Tree, Gradient Classifier are used to find out the accuracy.
    d. Confusion Metrix's, Classification Report are carried out while working on the model.

- Testing of Identified Approaches (Algorithms)
    a. SVM
    b. Decision Tree Classifier
    c. Random Forest Classifier
    d. Gradient Classifier

- Run and Evaluate selected models.

```python
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier()
rfc.fit(x_train,y_train)
rfc.score(x_train,y_train)
predrfc=rfc.predict(x_test)
print('Accuracy Score:',round(accuracy_score(y_test,predrfc),2),'\t')
print('\n','Confusion Matrix:','\n','\n',confusion_matrix(y_test,predrfc))
print('\n','Classification Report:','\n','\n',classification_report(y_test,predrfc))
```

```
Accuracy Score: 0.93

 Confusion Matrix:

 [[11  2]
 [ 1 29]]

 Classification Report:

               precision    recall  f1-score   support

            0       0.92      0.85      0.88        13
            1       0.94      0.97      0.95        30

     accuracy                           0.93        43
    macro avg       0.93      0.91      0.92        43
 weighted avg       0.93      0.93      0.93        43
```
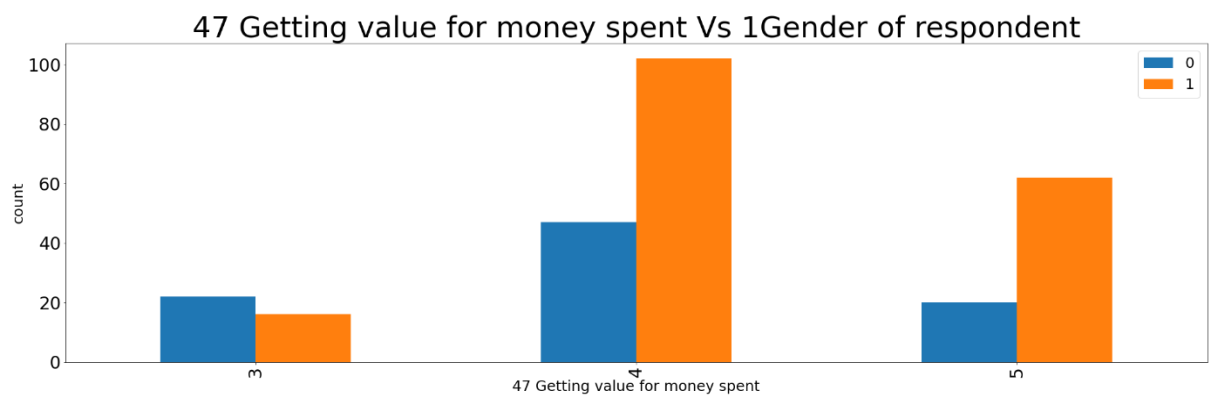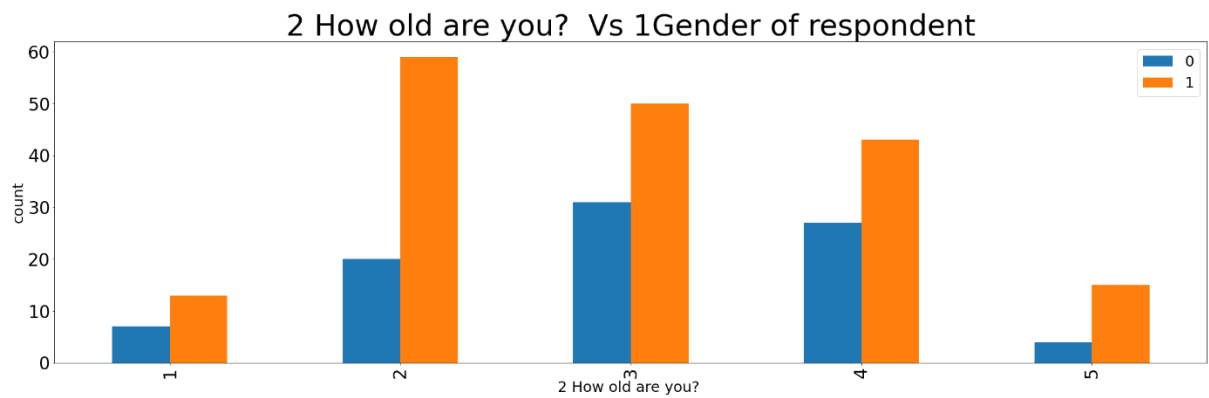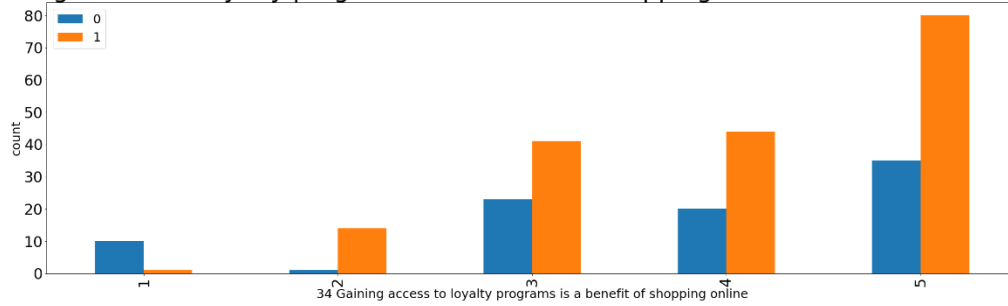
- Visualizations

2 How old are you? Vs 1Gender of respondent

8 Which device do you use to access the online shopping? Vs 1Gender of respondent

47 Getting value for money spent Vs 1Gender of respondent

31 Enjoyment is derived from shopping online Vs 1Gender of respondent

34 Gaining access to loyalty programs is a benefit of shopping online Vs 1Gender of respondent

9 What is the screen size of your mobile device? Vs 1Gender of respondent
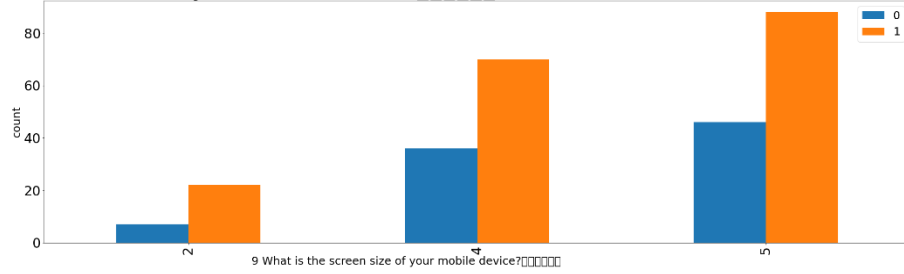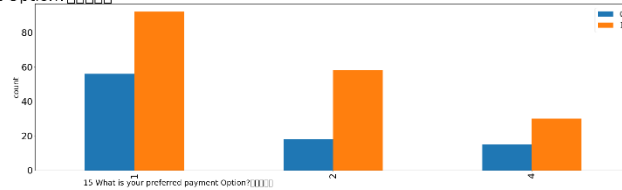

10 What is the operating system (OS) of your device? Vs 1Gender of respondent
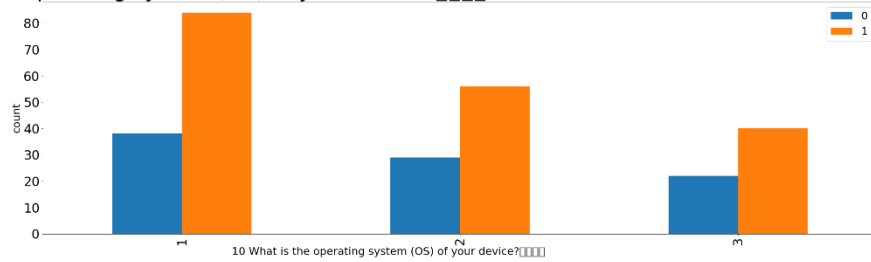
26 Trust that the online retail store will fulfill its part of the transaction at the stipulated time Vs 1Gender of respondent
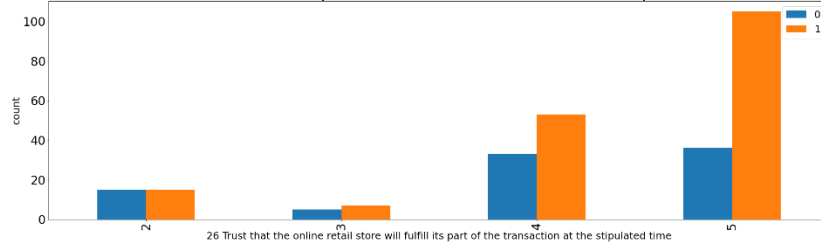
15 What is your preferred payment Option? Vs 1Gender of respondent

- Interpretation of the Results
    a. Credit / Debit Cards are widely used for online shopping by both male and female users.
    b. Both genders trust that online retail store will fulfil its parts of transaction at stipulated time
    c. Maximum prefer Windows / Windows Mobile for doing online shopping.
    d. Screen Size is different for Maximum Females and Males while do online shopping.
    e. Loyalty programs attract both the genders to a larger extend.
    f. Maximum Number of Males opted for indifference and most of the females feel that enjoyment is derived from Shopping Online.
    g. Both the genders agree that online shopping is Value for money
    h. Both Male and Females use Smartphones for accessing the online shopping
    i. Maximum males' users are between 31-40 Year old whereas maximum females users age is in between 21-30 Years

CONCLUSION

- Key Findings and Conclusions of the Study

  Online retailers must focus on the information quality and system quality of the online retail websites to enhance customer satisfaction, which in turn will lead to repurchase decision. Both the utilitarian and hedonistic values led to customer satisfaction and considerably stimulate their repurchase intention, leading to loyalty. The result further suggests that customers' perception of utilitarian and hedonic values will inform their preference for a particular online store. Hence, attention to the customer specific content is very much vital. The study recommends that the products and contents must be customers specific (personalised). The results of the study can help identify market segments for new e-retailers. Marketers can use the identified segments of online customers on the basis of their shopping values and decide accordingly on marketing strategies.

- Learning Outcomes of the Study in respect of Data Science

  Aim of the data analysis was to test and validate the proposed conceptual model and to identify the casual relationships that exist between variables. The research adopted several techniques to test, validated, check the fitness of the model as well as to test the hypotheses.

  Random Forest Classifier has given Accuracy of about 93% while data modelling. But for other model like SVM, Decision Tree etc accuracy is between 80% to 91%.

- Limitations of this work and Scope for Future Work

  The use of web-analytics in its classic form, which is to measure users' interaction with the website. However, web analytics is changing into digital analytics because of the multi-channel attribution like social media integration, social buying etc. Future studies may be conducted using the new digital analytics approach.