"Customer Retention"

Submitted by:

Sagnik Das

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to (Datatrained) also to my SME (Mr Sajid Choudhary). Who gave me the golden opportunity to do this wonderful project on the topic (Customer Retention), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

INTRODUCTION

- ## Business Problem Framing

  This problem helps to realise e-retail stores how they can better customer retention, by providing better shopping experience, better, better payment option, better comparison of products, etc. It also helps e-retail stores to know the likes and dislikes of the people of particular location and age group.

- ## Conceptual Background of the Domain Problem

  The problem has been divided by EDA steps, then PCA analysis, confusion matrix and logistic regression.

- ## Review of Literature

  This is a comprehensive summary of the research done on the topic of “Customer Retention”. The review helps e-retail stores to know the likes and dislikes of the people of particular location and age group, brand loyalty and whether they will recommend them to their friends.

- ## Motivation for the Problem Undertaken

  To learn about the effective process of customer retention policy that are being used by e-retail agencies.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  PCA Standardization

  $$z = \frac{value - mean}{standard\ deviation}$$

  Covariance Matrix computation

  $$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

  Covariance Matrix for 3-Dimensional Data

  Linear Regresion

  $$Y = a + bX$$

- Data Sources and their formats

  Data Provided by Flip Robo Technologies

- Data Preprocessing Done

- Data Preprocessing Done
- Remove duplicate or irrelevant observations
- Fix structural errors
- Filter unwanted outliers
- Handle missing data
- Changed strings into integers

- Validate and QA

- Data Inputs- Logic- Output Relationships

  Customer retention data sheet.

  EDA steps, then PCA analysis, confusion matrix and logistic regression to identify brand loyalty, potential new customers. Also, how they are different in different locations and vary with different age group.

- State the set of assumptions (if any) related to the problem under consideration

  Changed the name of locations into numerical

Delhi = 11
Greater Noida
=12
Karnal = 13
Bangalore = 14
Noida = 15
Solan = 16
Moradabad =
17
Gurgaon = 18
Merrut = 19
Ghaziabad = 20

  Changed the name of websites into numerical

Flipkart =
1
Amazon =
2
Paytm = 3
Myantra =
4
Snapdeal
= 5

- Hardware and Software Requirements and Tools Used

Hardware – Laptop

Software –

import pandas as pd, import os, import csv, import sklearn, import numpy as np, import matplotlib.pyplot as plt, import seaborn as sns, %matplotlib inline, from sklearn.preprocessing import MinMaxScaler,

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  EDA Method, PCA analysis, Covariance, Confusion Matrix, Linear Regression and prediction.

- Testing of Identified Approaches (Algorithms)
- EDA Method, PCA analysis, Covariance, Confusion Matrix,
- Run and Evaluate selected models
- Linear Regression and prediction.
- Key Metrics for success in solving problem under consideration

  Classification Matrix and Confusion Matrix
- Visualizations

  Contour, Heat map, pair plot, box plot.

- Interpretation of the Results

  Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

CONCLUSION

- Key Findings and Conclusions of the Study

  How to improve customer experiences in e-retail organizations. Also, to identify potential customers.

- Learning Outcomes of the Study in respect of Data Science

  EDA and PCA algorithm helped in cleaning excess data and helped it in putting it in visualised matter.

- Limitations of this work and Scope for Future Work

  To much data in string format which was needed to convert in integer format