Customer Analytics

A comprehensive and detailed approach using Machine Learning

Customer analytics is a process by which data from **customer** behavior is used to help make key business decisions via market segmentation and predictive **analytics**. This information is used by businesses for direct marketing, site selection, and **customer** relationship management.

Case study Background

An International E-Commerce company(Electronic goods) wants to use some of the most advanced machine learning techniques to analyse their customers with respect to their services and some important customer success matrix. They also have future expansion plans to India.

They have some specific key insights to be found out from their existing customer database.

Problem Statement - 1

- As a Data Scientist, they want you to build a model to predict if the shipments are or will reach on time or not.

For this, they want you to use various Logit/Probabilistic techniques with the most accurate model. The main models they want you to build and compare the accuracy are:

1. Logistic Regression
2. Support Vector Machines
3. Random Forest
4. XgBoost or any other boosting technique.

It is not only about the accuracy but also about the variables impacting the model and generating insights.

Problem Statement - 2

- They also want to know, if their shipments are reaching on time to their best customers who have a :

1. Good customer rating.
2. Good customer score.
3. Make recurring orders.
4. Highest payment buyers

To analyse these Problem statements you can use advance visualizations using ggplot2.

Problem Statement - 3

- They want you to create a customer segmentation using clustering algorithm of the customers to whom the shipments are not reaching on time.

Hint : Use only delayed customers data for clustering.

Problem Statement - 4

- Since the company also wants to start their operations in India, they want you to do a Sentiment analysis(Positive or Negative) of their competitors such as:

1. Amazon India
2. Flipkart
3. Snapdeal

You can use twitter sentiment Analysis to solve this problem.

Data Information : (Important)

ID	Id number of the customer
Warehouse_block	The company has a big warehouse which is divided in various blocks such as A,B,C,D and so on.
Mode_of_Shipment	The company ships the products by different modes of transport such as ship, air and road
Customer_care_calls	this variable indicates the number of calls made for enquiry of the shipment.(Sometimes customer make too many calls, hence the company wants to know that are these customers unknowingly favoured.
Customer_rating	The company has rated every customer on various parameters, 1 being the lowest (Worst), 5 being highest (Best)
Cost_of_the_Product	It is the cost of the product in USD
Prior_purchases	This variable indicates the number of prior purchases
Product_importance	The company has categorised the products in the range of high, medium and low based on various parameters
Gender	Male or female
Discount_offered	it is the percentage of discount offered on that specific product.
Weight_in_gms	It is the weight in grams
Reached.on.Time_Y.N	It is the Y variable, where 1 Indicates that the product has **NOT** reached on time and **0 indicates it has reached on time**

- The company operates in various states of USA, but the customer data is only specific to one state in the USA.
- The warehouse is located on the eastern part whereas, the state to which shipments are delivered is at the western part of USA.

Thank You.