Customer Retention Case Study

Customer retention refers to a company's ability to turn customers into repeat buyers and prevent them from switching to a competitor. It indicates whether your product and the quality of your service please your existing customers. It's also the lifeblood of most subscription-based companies and service providers.

Customer retention strategies are the processes and initiatives businesses put in place to build customer loyalty and improve customer lifetime value.

Keeping your current customers happy is generally more cost-effective than acquiring first-time customers. According to the Harvard Business Review, acquiring a new customer can be five to 25 times more expensive than holding on to an existing one. There are different views of different entities on this but according to me I think both the ways of creating & maintaining the trust are of challenging part.

Customer loyalty won't just give you repeat business. Loyal customers are more likely to give free recommendations to their colleagues, friends, and family. Creating that cycle of retained customers and buzz marketing is one way your company can cultivate customer loyalty for long-term success.

The benefits of retention are Cost savings, Positive word of mouth marketing, A better bottom line (Increasing retention rates) etc. & for that we use different strategies to improve customer retention like Deliver fast support, Personalize interactions, Invest in employees, Meet customers where they are, Gather customer feedback, Incentivize loyalty & more.

The problem statement contains the following description where we need to find out the customer loyalty or repeat purchase intentions of the customers. As we discussed previously about the importance of it we are going to find out the best possible result with different algorithms with machine learning. In the dataset there are different questions & answers given by customers about their experiences while purchase or after purchase. On that basis we will proceed with our finding on Customer Satisfaction.

"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."

We have done the following analysis of the dataset where we Imported necessary libraries so that we can work on datasets. After reading the dataset tried with conversion of String data into Integers for equalize the data type for process. We Checked for missing values & there were no missing values found in dataset. With the correlation among all the columns checked the correlation and found most of the data is positively correlated with each other.

In data visualization done the following visualizations :

First used Correlation Matrix for showing the correlation between all columns with Heatmap.

The result of the Correlation Matrix is on following GitHub link.

https://github.com/komalghatvilkar/Internship/blob/main/Customer_Retention_Case_Study%20-%20Assignment%204/Customer%20Retention-Correlation%20Matrix.png

Second, used Histogram for all dataset & plot Histogram for visualizing the data individually.

The result of the Histogram is on following GitHub link.

https://github.com/komalghatvilkar/Internship/blob/main/Customer_Retention_Case_Study%20-%20Assignment%204/Customer%20Retention-Histogram.png

According to the visualizations and understandings, we found that all the features or variables are important for analysis so haven't removed anything. There were no null values found which was good. There were some columns with strings so encoded them with label encoder for balance the data type for further process. From the output of correlation matrix, we can see that it is symmetrical i.e. the bottom left is same as the top right. It is also observed that most of the variables are positively correlated with each other. The Histogram shows that each variable distributed differently and as we can see the data has categorical values, histogram is better visualization graph to show the distribution. Firstly we split the data into x & y.

I have taken the output as the column 'Which of the Indian online retailer would you recommend to a friend?' because as per me we recommend the site we like or on which we trust or we always purchase from.

Then checked and removed the skewness of the data.

After that checked the outliers if any,found very less so didn't remove the outliersI used boxplot to check the outliers in dataset.

The result of the Boxplot is on following GitHub link.

https://github.com/komalghatvilkar/Internship/blob/main/Customer_Retention_Case_Study%20-%20Assignment%204/Customer%20Retention-%20Outliers%20In%20Data.png

After checking the outliers did the Train Test Split and find the best accuracy & random state :- Best accuracy is 1.0 on Randon_state 0. Then performed the model building and tried to use best possible model building techniques :

- Logistic Regression model :- The Accuracy is 100.0

- Decision Tree Classifier :- The Accuracy is 100.0

- Random Forest Classifier :- The Accuracy is 100.0

- SVC (Support Vector Classifier) :- The Accuracy is 100.0

Next did Cross Validation for all the models used & it shows the output -

- Cross Validation Score Of Logistic Regression Model : 1.0

- Cross Validation Score Of Decision Tree Model : 1.0

- Cross Validation Score Of Random Forest Model : 1.0

- Cross Validation Score Of SVC Model : 1.0

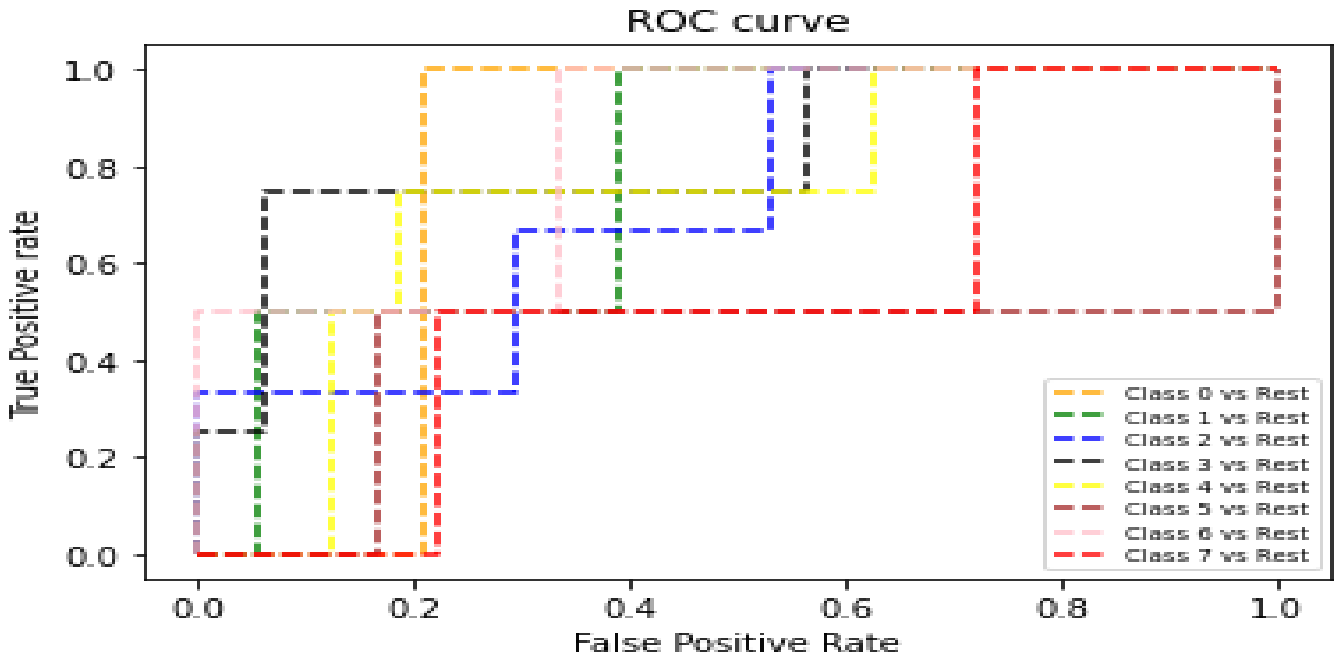
Then with the Hyperparameter Optimization Technique 'GridSearchCV' fit the data in model :- The result is Grid Search CV 1.0

Finally, plotted the ROC AUC Curve for the output & got the Conclusion as per the results found those are all the models are showing the exact result for the dataset with 100% accuracy. The final accuracy with GridSearchCV is 100%

Please find the GitHub links for ROC curve & Conclusion to refer.

https://github.com/komalghatvilkar/Internship/blob/main/Customer_Retention_Case_Study%20-%20Assignment%204/Customer%20Retention-ROC%20Curve%20of%20Dataset.png

https://github.com/komalghatvilkar/Internship/blob/main/Customer_Retention_Case_Study%20-%20Assignment%204/Conclusion%20Data.JPG



Index	Original_Data	Predicted_Data	Result
0	7	7	Flipkart.com, Paytm.com, Myntra.com, snapdeal.com'
1	6	6	Flipkart.com'
2	2	2	Amazon.in, Flipkart.com, Myntra.com'
3	0	0	Amazon.in'
4	0	0	Amazon.in'
5	7	7	Flipkart.com, Paytm.com, Myntra.com, snapdeal.com'
6	3	3	Amazon.in, Myntra.com'
7	0	0	Amazon.in'
8	1	1	Amazon.in, Flipkart.com'
9	0	0	Amazon.in'
10	0	0	Amazon.in'
11	2	2	Amazon.in, Flipkart.com, Myntra.com'
12	1	1	Amazon.in, Flipkart.com'
13	0	0	Amazon.in'
14	2	2	Amazon.in, Flipkart.com, Myntra.com'
15	6	6	Flipkart.com'
16	0	0	Amazon.in'
17	6	6	Flipkart.com'
18	7	7	Flipkart.com, Paytm.com, Myntra.com, snapdeal.com'
19	0	0	Amazon.in'
20	0	0	Amazon.in'
21	6	6	Flipkart.com'
22	6	6	Flipkart.com'
23	6	6	Flipkart.com'
24	0	0	Amazon.in'
25	0	0	Amazon.in'
26	5	5	Amazon.in, Paytm.com, Myntra.com',

Index	Original_Data	Predicted_Data	Result
27	1	1	Amazon.in, Flipkart.com'
28	0	0	Amazon.in'
29	1	1	Amazon.in, Flipkart.com'
30	0	0	Amazon.in'
31	0	0	Amazon.in'
32	0	0	Amazon.in'
33	1	1	Amazon.in, Flipkart.com'
34	3	3	Amazon.in, Myntra.com'
35	2	2	Amazon.in, Flipkart.com, Myntra.com'
36	1	1	Amazon.in, Flipkart.com'
37	0	0	Amazon.in'
38	1	1	Amazon.in, Flipkart.com'
39	0	0	Amazon.in'
40	0	0	Amazon.in'
41	0	0	Amazon.in'
42	7	7	Flipkart.com, Paytm.com, Myntra.com, snapdeal.com'
43	1	1	Amazon.in, Flipkart.com'
44	4	4	Amazon.in, Paytm.com'
45	3	3	Amazon.in, Myntra.com'
46	3	3	Amazon.in, Myntra.com'
47	0	0	Amazon.in'
48	7	7	Flipkart.com, Paytm.com, Myntra.com, snapdeal.com'
49	0	0	Amazon.in'
50	3	3	Amazon.in, Myntra.com'
51	4	4	Amazon.in, Paytm.com'
52	0	0	Amazon.in'
53	0	0	Amazon.in'

The results shows that the dataset is correct & we can proceed with this dataset further.

Thank you...!!!