**CUSTOMER CART ABANDONMENT**



**Shilpi Gupta A001**

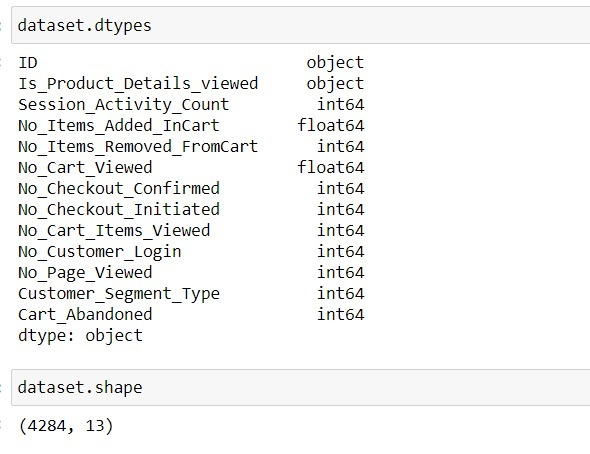
**Chelsi Fagniya A008**

**INTRODUCTION:**

Customer cart Abandonment can be referred as an e-commerce term to describe a visitor who leaves the web page before completing the purchase. Shopping cart abandonment is one of the most crucial problems for online businesses to overcome. Examples of abandonment include shopping cart abandonment, referring to visitors who add items to their online shopping cart, but exit without completing the purchase.

**DATASET:**

We have used Python as our primary software for the analysis part; here our first step is to import the dataset. After importing the dataset we have explored features of our dataset.



The dataset used contains 4284 rows and 13 columns.

The dataset consists of following 13 attributes:

|  |  |  |
| --- | --- | --- |
| S.No. | Name | Description |
| 1 | ID | The session id of the customer |
| 2 | Is\_Product\_Details\_Viewed | Whether the customer is viewing the product details or not |
| 3 | Session\_Activity\_Count | How many times a customer is going to the different pages. |
| 4 | No\_Items\_Added\_InCart | Number of items in cart |
| 5 | No\_Items\_Removed\_FromCart | Number of items removed from the cart |
| 6 | No\_Cart\_Viewed | How many times the customer is going to the cart page. |
| 7 | No\_Checkout\_Confirmed | How many times the checkout has been confirmed successfully by the customer. |
| 8 | No\_Checkout\_Initiated | How many times the checkout(successful as well as unsuccess) is being done by the user |
| 9 | No\_Cart\_Items\_Viewed | How many times a user is viewing the product from cart |
| 10 | No\_Customer\_Login | Number of times the customer had did log in |
| 11 | No\_Page\_Viewed | Number of pages viewed by the customer |
| 12 | Customer\_Segment\_Type | The customer falls under which category,i.e, 0 for Target Customer, 1 for Loyal Customer, and 2 for Untargeted customer |
| 13 | Cart\_Abandoned | Whether the customer is doing cart abandonment or not. This is the target variable that we need to predict |

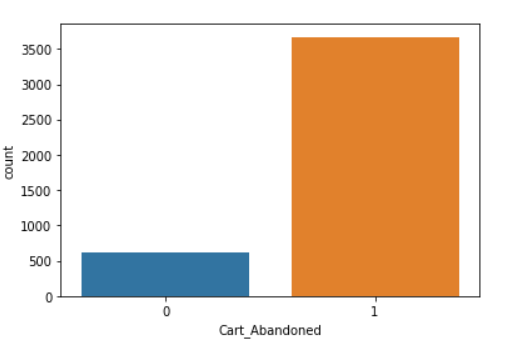
**EXPLORATORY DATA ANALYSIS**

Here we have performed both univariate as well as bivariate analysis to infer sights from our dataset. We have divided all numerical as well as categorical variables.



UNIVARIATE ANALYSIS

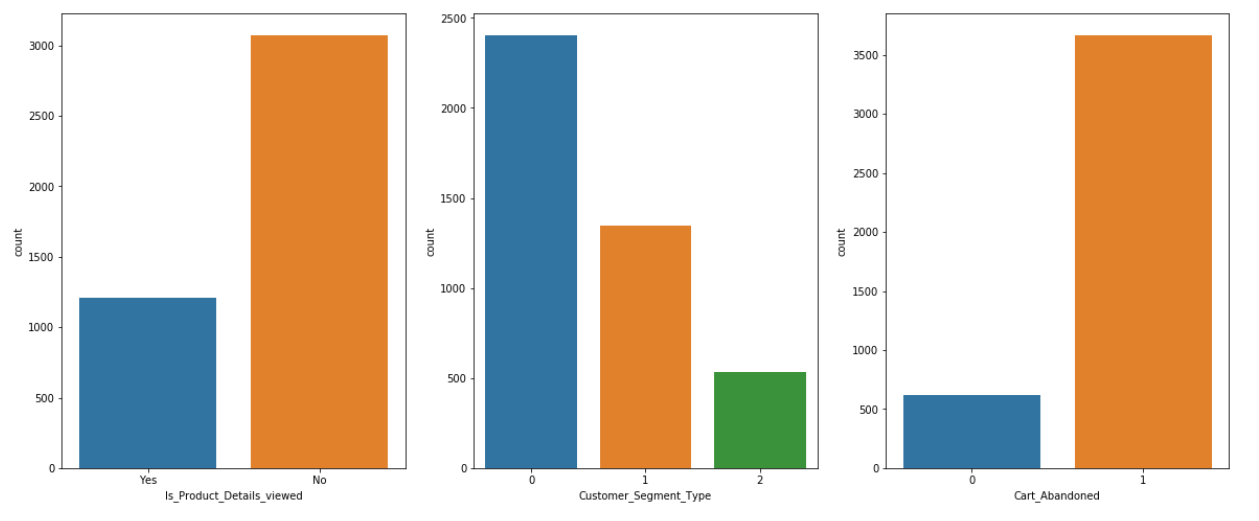
* Target variable – Cart Abandoned



It can be clearly seen that the data is an example of imbalanced dataset.

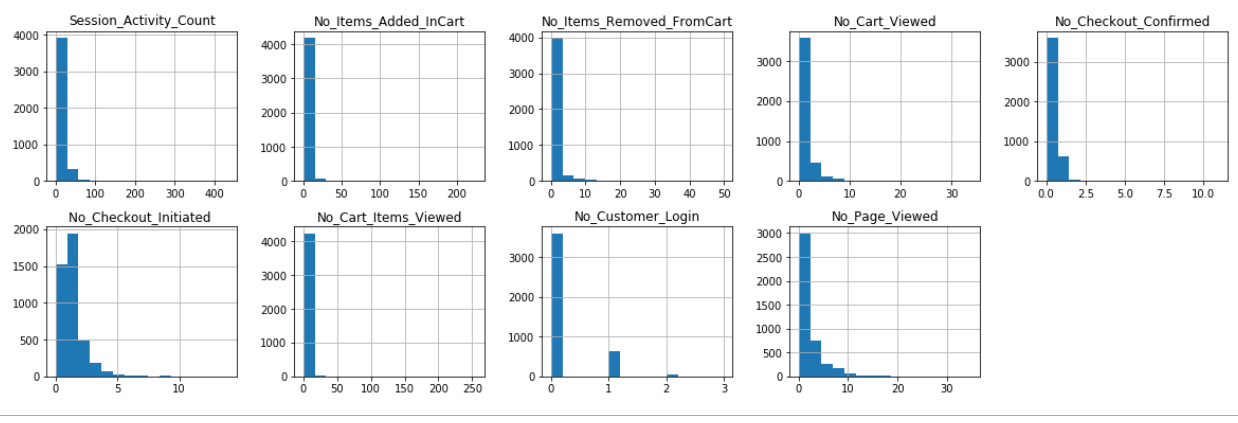
* Count plot for categorical attributes –

Is\_Product\_Details\_viewed & Customer\_Segment\_Type



Here we can observe that there are very few sessions in which the product details have been viewed by the user. And we can also observe that our targeted customers (0) are very high compared to untargeted (2) customers.

* Histogram for numerical attributes -



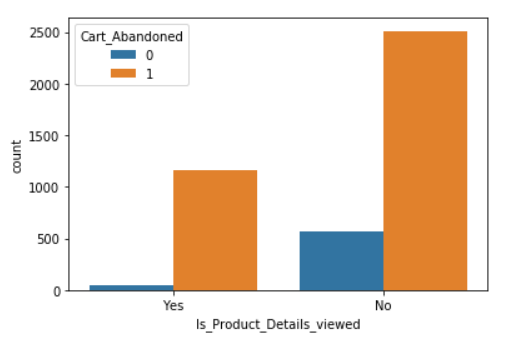
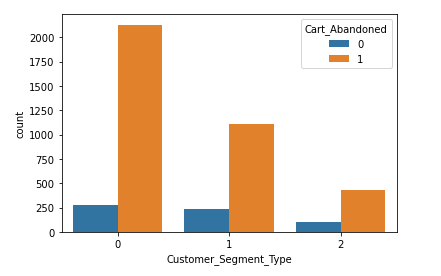
Here we can see that our data unevenly spread.

BIVARIATE ANALYSIS

* FOR CATEGORICAL

1. Is\_Product\_Details\_viewed Vs. Cart\_Abandoned
2. Customer\_Segment\_Type Vs. Cart\_Abandoned

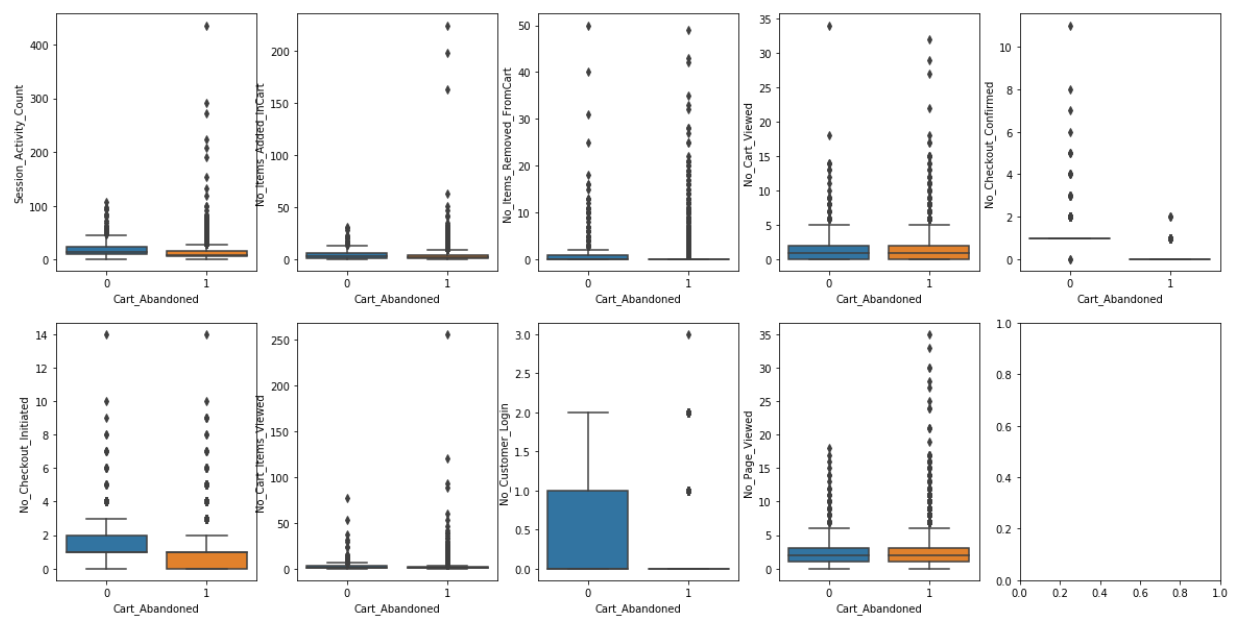
If a customer is viewing the product details then what is the chance that he is doing cart abandonment? How does customer segment type (0 for target customer, 1 for loyal customer, and 2 for untargeted customer) affect cart abandonment?

Here we can see that cart abandonment is seen more when the product details are not viewed.

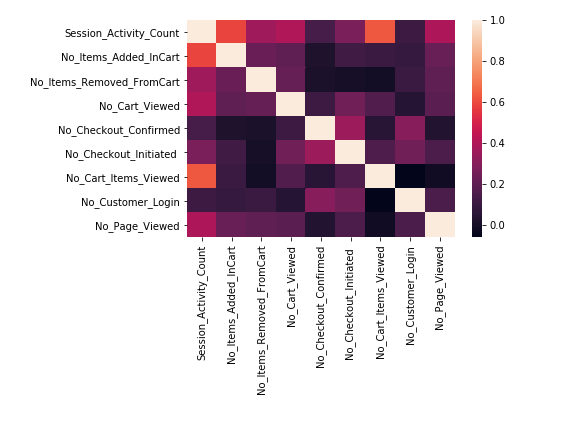
* FOR NUMERICAL

1. Numerical Attributes Vs. Cart\_Abandoned: By Boxplot



Here we can see that we have many outliers

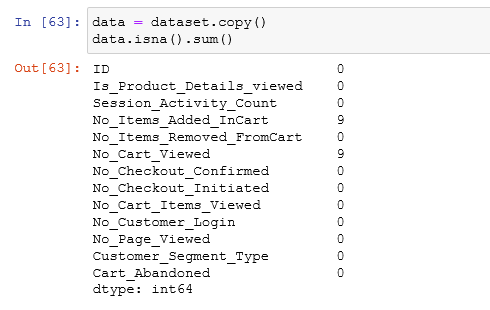
1. Correlation plot of independent attributes

**

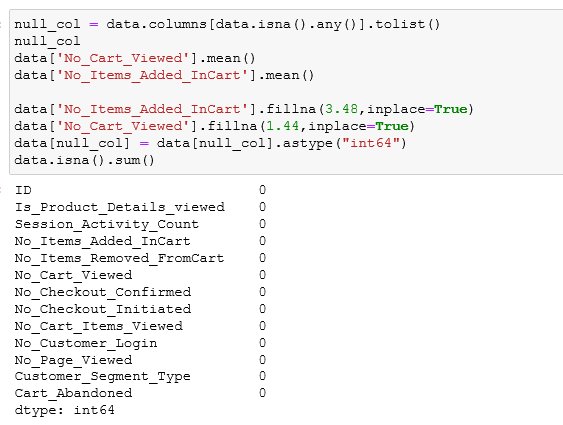
Here we can observe that our attributes are mostly independent of each other.

**DATA PRE-PROCESSING**

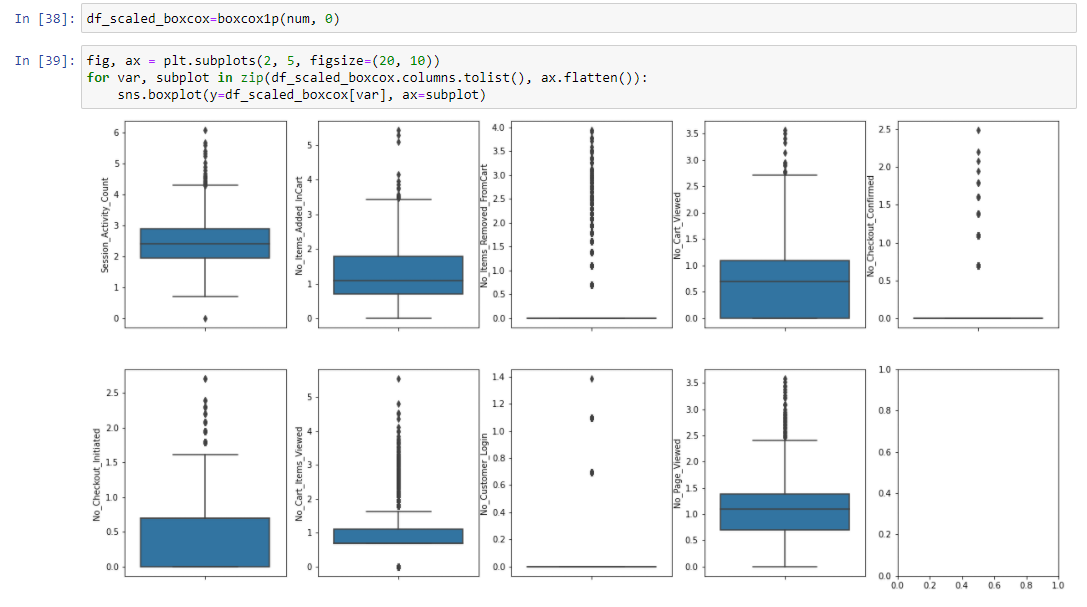
* Checking for missing values

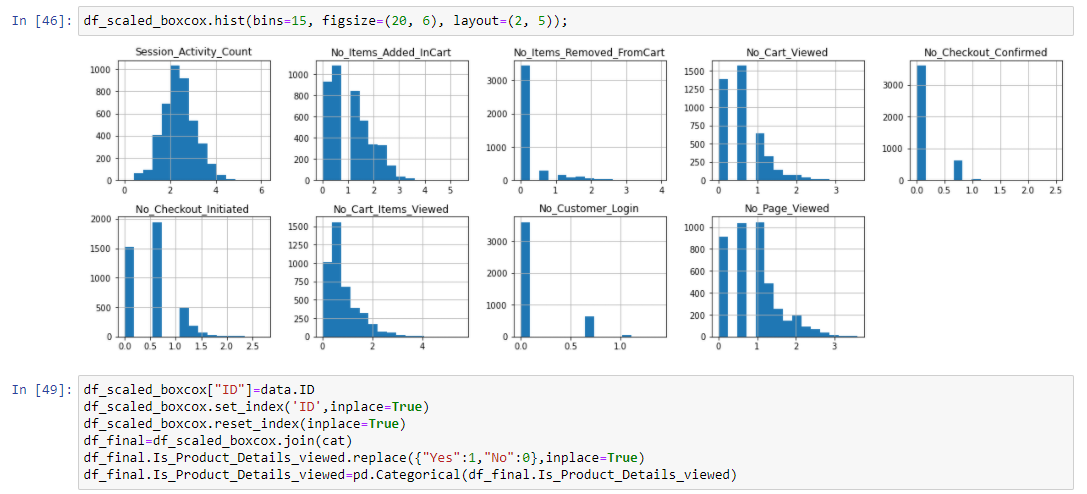


* Imputing missing values



* Taking care of outliers by normalizing data - By BoxCox Normalization



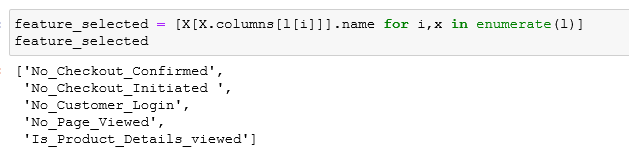


Here we can see that our data is almost normally distributed. Now we are ready to go to the next step i.e., feature selection.

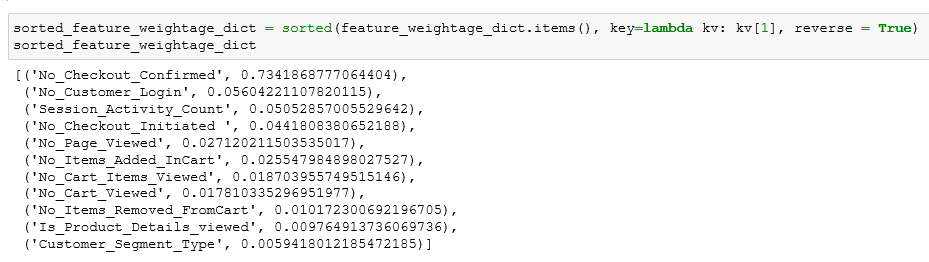
**FEATURE SELECTION**

Here we have taken two models through which we will get feature importance, by comparing the feature importance’s of both models we will select our best feature.

1. By RFE(Recursive Feature Elimination), the features selected are –



1. By Random forest classifier, the features selected are-

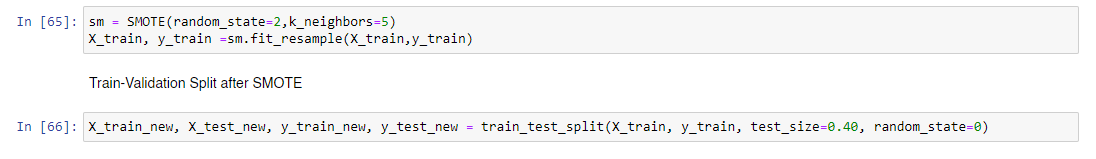


Hence on selecting common features from both models, we get the following features to build our model and predict the outcome.

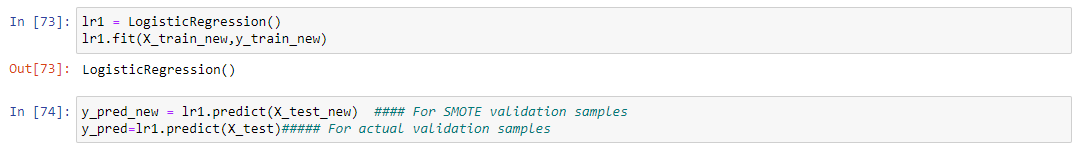


**OVER-SAMPLING USING SMOTE**

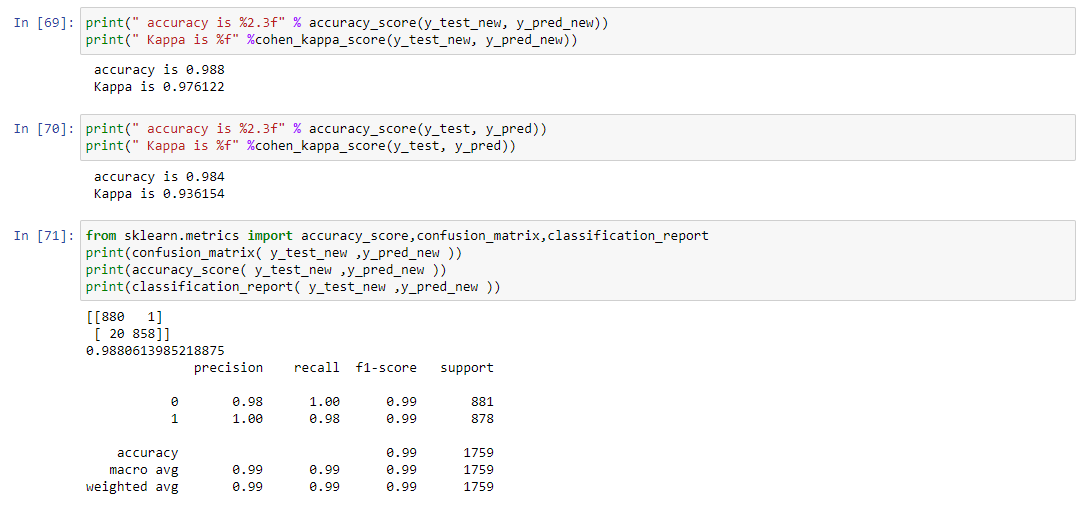
Since our dataset wasn’t balanced so we decided to use SMOTE technique (Amongst the 3 techniques namely under sampling, over Sampling and SMOTE) to balance it first.



**MODEL BUILDING AND PREDICTION**



## MODEL EVALUATION



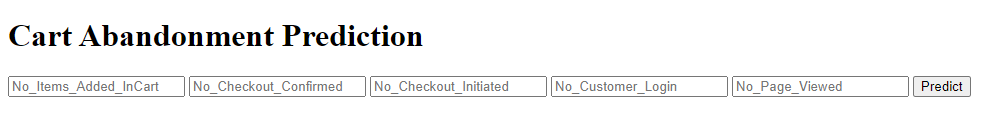
We can see clearly that our model has predicted the outcomes with an accuracy of 98.4%. As our kappa score is also high we can conclude that our model will give the same accuracy with other data.

## DEPLOYMENT

We have built an app for our model whose link is given below. Using this link one can predict whether the cart will be abandoned or not.

<https://cartabandonment.herokuapp.com/>

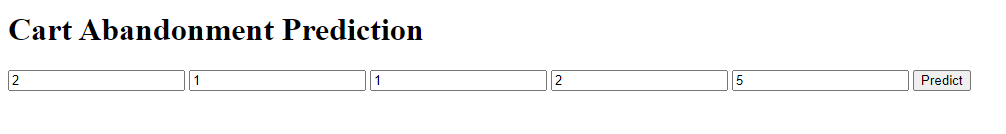
Step 1 – Open the link on browser.



Step 2 – Assign values of your choice to each feature. Refer to the table below.

The possible values of the features affecting cart abandonment can be –

|  |  |  |
| --- | --- | --- |
| **Features** | **Description** | **Possible Values** |
| No\_Items\_Added\_InCart | Number of items in cart | Any whole number |
| No\_Checkout\_Confirmed | How many times the checkout has been confirmed successfully by the customer. |
| No\_Checkout\_Initiated | How many times the checkout(successful as well as unsuccessful) is being done by the user |
| No\_Customer\_Login | Number of times the customer had did log in |
| No\_Page\_Viewed | Number of pages viewed by the customer |



Step 3 – Click the predict tab.

After assigning the values of choice to each feature the app will automatically predict whether cart abandonment has happened or not using the above model.

If the cart abandonment has happened then it will predict Cart\_Abandonment 1 otherwise Cart\_Abandonment 0.



**CONCLUSION**

By using this analysis we can predict whether a customer will do abandonment or not in the future. So that customers can be given some offers to complete the checkout successfully and avoid cart abandonment.