Shopping Cart Analysis Tool

Team Members: Gorantla Varun Kumar Chowdary, Singh Nikhil, Tandel Sonia
Mentor: Dawood Moazzem
Professor: Gary D. Boetticher

Introduction

Today, the websites are facilitating the users to buy products of their choice. There has been an exponential growth in the number of internet retailers and at the same time the amount of data generated by these retailers per second has increased. The challenges faced by the organizations include extracting insights from the overwhelming amount of customer data using a predictive analysis model and while doing so, maintaining the model performance The purpose of this project is to predict whether a customer is going to abandon a shopping cart before it happens.

Project Scope

One of the concerns of online retail is when users select products to purchase, add those items to their shopping cart, and then end up exiting without making a purchase. This phenomenon is known as an abandoned cart. The number of potentially missed sales can result in a considerable missed opportunity in terms of revenue. If addressing the issue of abandoned carts can increase sales by even 1%, that means a lift of millions of USD per year.

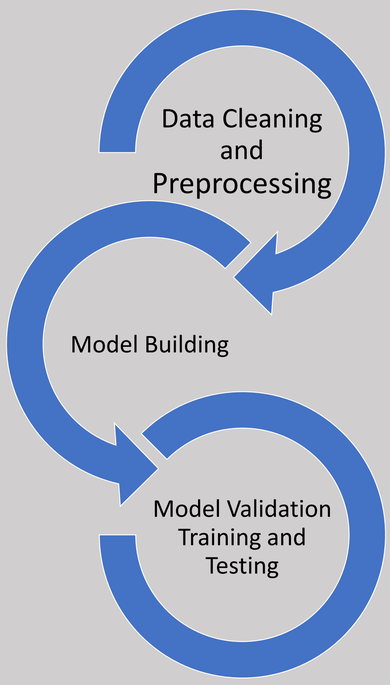
Methodology

Data cleaning and Pre-processing: Cleaning and preprocessing the data of size includes removing the duplicate values, columns with low variance, high correlation, and null values. Also, initial visualization is performed on the data to visualize the inconsistencies and eliminate outliers in the data.

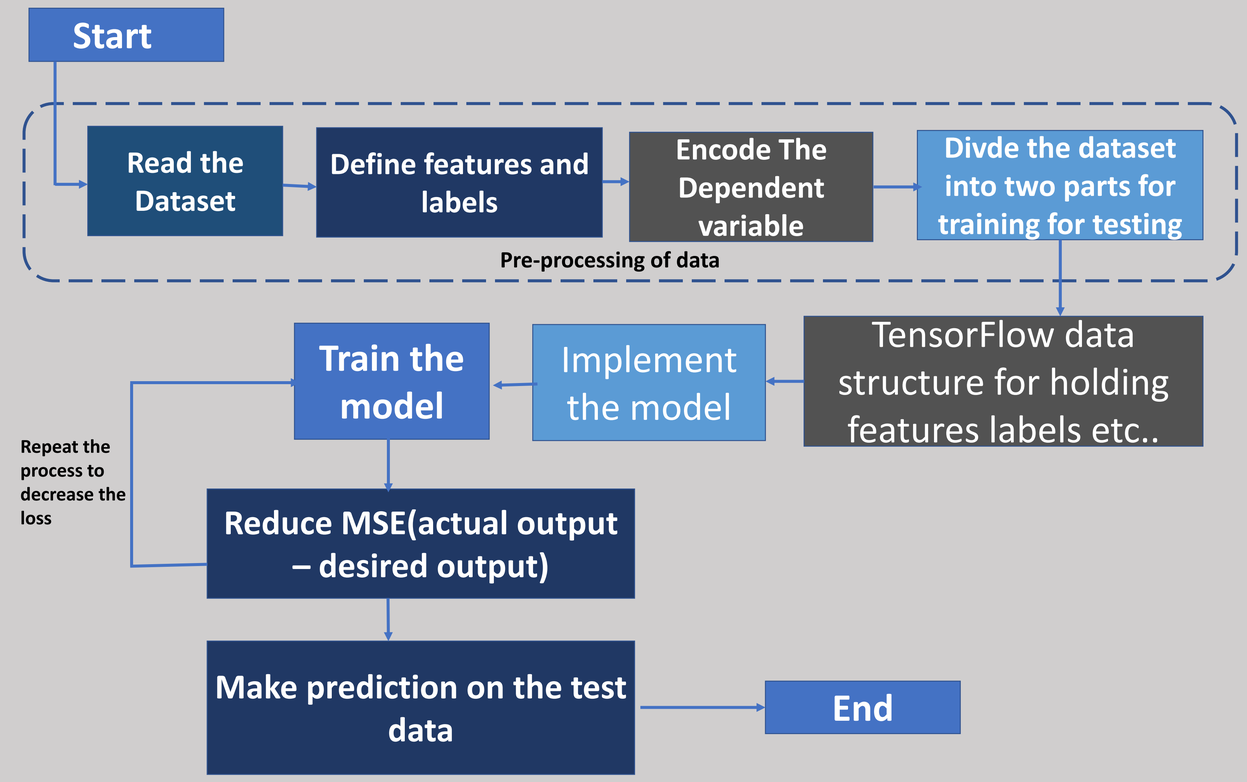
Model Building and Training phase: This phase focuses on building and training a predictive analysis model to identify the users who have abandoned the cart.

Validation phase and evaluation phase: This phase validates the outcome (prediction) of the resultant model of the previous phase against the known set of results.

Testing phase and Deployment phase: This phase evaluates the resultant model derived from the validation phase by applying the model to unlabeled data.



As, the problem is whether the customer abandoned the cart or not, it is a regression type problem.
The solution found was to use machine learning algorithms like ‘feedforward neural network to train the model. Python libraries like pandas, numpy, tensor flow and Spark libraries have been used for handling the large amount of customer data.



Results

This model is used to predict if a user will make a purchase or abandon the cart. 1009 predictions are made using this model, among which 826 users are predicted abandoning the cart and 183 users are predicted making a purchase. According to original data, 829 users abandoned the cart and 180 users made a purchase. The accuracy of the classifier is around 96%.

	Predicted:Abandoned	Predicted: Purchased
Actual: Abandoned	TN = 811	FP = 18
Actual: Purchased	FN = 15	TP = 165

Accuracy:
(TN + TP)/Total = (811+165)/1009 = 0.96
Error Rate:
equivalent to 1 – Accuracy
(FN + FP)/Total = 0.04

Conclusion

This model intends to help the organizations retain their customers thus, increasing the sales and improving the revenue.

References

[1] Adobe. (n.d.). Clickstream (Legacy) Data Feeds Help [Blog post]. Retrieved from https://marketing.adobe.com/resources/help/en_US/sc/clickstream/

[2] Adobe. (n.d). Traffic Variables sProps Inside Omniture SiteCatalyst [Blog post]. Retrieved from https://theblog.adobe.com/traffic-variables-sprops/

[3] Adobe. (n.d). Conversion Variables – Part I Inside Omniture SiteCatalyst [Blog post]. Retrieved from https://theblog.adobe.com/conversion-variables-part-i/