The Archons

Customer Purchase Prediction

ML Project Final Review

BATCH - 17
M.Manjusha Devi
K.Divya
B.Nandini
K.Asha Deepthi

June 24, 2021

Overview

Title

Project Title : Customer Purchase Prediction

Problem Statement

To analyse the interest of the customers based on the activity he/she is performing on a E-commerce site. We predict the buying behavior of a site visitor, as this can have many implications such as E-commerce website will be able to suggest better target ads or figure out factors that may lead to increased sales.

Approach Followed

Approach followed:

- Firstly, we researched more about our idea and found suitable data.
- We initially explored the data and found imbalance in data.
- We cleaned the data and eliminated the collinearity problem.
- For the modeling and feature selection, we selected SGDClassifier, Random Forest Classifier, XGBoost Classifier.
- Calculated the Precision score, Recall score, test and train accuracy for all the models to find the most accurate model.

Technologies Used

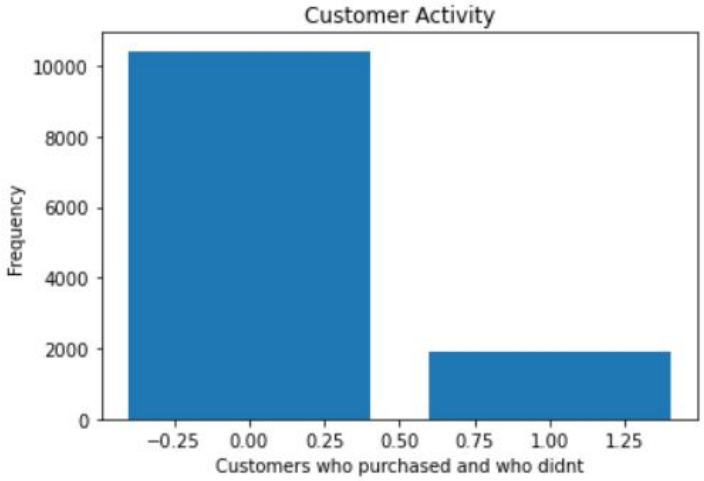
Tech Stack

- Python 3.8.1
  - Numpy, Pandas
  - sklearn
- LaTeX
- GitHub

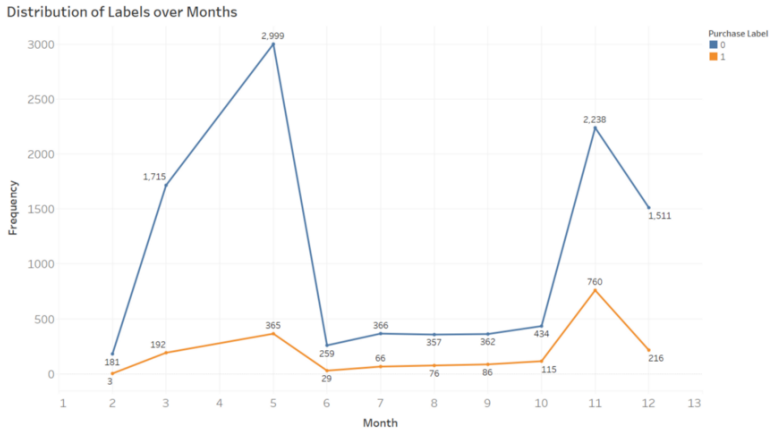
Data Set

Online Shoppers Purchasing Intention data set provided on the UC Irvine's Machine Learning Repository. *Click Here*

Learnings from Data Set



(a) Customers purchase activity



(b) Purchase Traffic every month

Figure: Learnings from data set

Learnings

- We got a clear idea about how to train the model and see the accuracy of Algorithm.
- We learnt thet the data is linear after visualizing the data, and implemented SGDClassifier, RandomForest Classifier, XGBClassifier.
- We were able to conclude that data generated from the customer clickstream is the most important for predicting purchase probabilities.
- We resolved the errors and applied the learnings which we understood by discussing and researching to improve the accuracy.

Results

Classifier	Train	Test	Precision	Recall
SGDClassifier	0.883313	0.881995	0.692308	0.509852
RandomForestClassifier	0.983982	0.890916	0.721683	0.549261
XGBClassifier	0.989964	0.893755	0.710526	0.598522

Table: Final Accuracy Scores of all Classifiers.

Project Demo

Problems Faced

Problems Faced

- 1.Finding the appropriate Data set and understanding the terms.
- To research on all the algorithms and check which one is the correct fit for our trained model.
- Checking the accuracy and again training the model for the same.

Conclusions

- Able to predict that e-commerce sales from online traffic would prove to be beneficial for any company.
- Predict that companies should focus on improving mobility between pages to encourage users to browse among different products
- Able to predict certain months where e-commerce companies should capitalize and provide additional sales, deals to encourage product sales.

References

ML Project Git Repository

Customer Purchase Prediction ML_Project 302 *Click Here*

Customer Purchase Prediction Through ML by Hannah Sophia Seippel

University of Twente.*Faculty of Electrical Engineering, Mathematics Computer Science*

Data Set

UCI Machine Learning Repository *Click Here*

Thank You