Predicting Potential Shoppers Based on Intent to Buy

Springboard Capstone #1 Project Proposal

Problem:

Online shopping is a huge and growing form of purchasing and represents a huge portion of B2C (Business to Customer) revenue. 69% of Americans have shopped online at some point (1), with an average revenue of $1804 per online shopper (2). 36% of Americans shop at least once per month (2)! Learning how and when shoppers will research and purchase goods online is important to businesses as they can use customer behavior insights to target advertising, marketing, and deals to potential customers to further increase their sales and revenue.

Data: Link (http://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset#)
The data set is a set of 18 features: 10 numerical and 8 categorical. This dataset has 12330 entries, split into 10,422 entries where the shoppers did not purchase and 1908 entries where the shoppers did purchase. Each entry is based on unique users in a 1-year period to avoid any trends specific to a specific campaign.

Approach:

The problem will be answered by creating a classifier model to predict whether the shopper will purchase an item, or if they will not purchase an item. Additionally, we can calculate an approximate likelihood that a shopper is bound to make a purchase as well as classify our shoppers into bins representing each approximate interest level.

As all entries in this dataset are pre-labeled, we will split the dataset into train/test subsamples and train our model on the training data. We can use our test data to validate our model, and give us an approximate accuracy level of the predictive model.

I propose to use an ensemble method such as a Random Trees Classifier to train a model to predict a shopper's intent. We can use several different ensemble methods and finalize our model by choosing the model with the best results.

References:

1. https://optinmonster.com/online-shopping-statistics/
2. https://www.invespcro.com/blog/global-online-retail-spending-statistics-and-trends/
3. http://maristpoll.marist.edu/wp-content/misc/usapolls/us180423_NPR/NPR_Marist%20Poll_Tables%20of%20Questions_May%202018.pdf#page=2
4. https://www.statista.com/topics/2477/online-shopping-behavior/