Machine Learning Engineering Capstone Proposal

Lakshya Khandelwal
July 15' 2020

Online Shopper's Conversions Prediction

Domain Background

Everytime a customer browses some product online, the search engine and visited website can store relevant session information. This data is heavily used to predict what kind of sessions will actually convert to purchases (revenue).

This can highly enhance the profit making ability of a brand/ ecommerce platform by targeting such potential sessions / individuals.

Due to improvement in machine learning algorithms as well as data storing and processing capabilities, companies can store huge amounts of data (even at individual customer's session level). This makes usage of sophisticated machine learning algorithms very effective.

Problem Statement

Using session level user data for online shopping interaction, we will try to develop intuition into how various features / attributes lead to click conversion in terms of revenue.

Data Set Information:

https://www.kaggle.com/roshansharma/online-shoppers-intention

The dataset consists of feature vectors belonging to 12,330 sessions.The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period.

Attribute Information:

The dataset consists of 10 numerical and 8 categorical attributes.

The 'Revenue' attribute can be used as the class label.

"Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration" represent the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL information of the pages visited by the user and updated in real time when a user takes an action, e.g. moving from one page to another. The "Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site. The value of "Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and

then leave ("bounce") without triggering any other requests to the analytics server during that session. The value of "Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that were the last in the session. The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction. The "Special Day" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentina's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8. The dataset also includes operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

Solution Statement

Supervised learning approach is taken in order to predict revenue conversion using given features. I try to identify important features and use models like SVM and Decision Tree.

Evaluation Matrix

I plan to use classification accuracy as a measure and hope I get accuracy in north of 80 %