

Business Understanding

David Ricardo Pedraza Silva

April 2021

Introduction

Our idea is to use purchase data on amazon, like reviews and purchase history, in order to device predictions which would allows us to make better suggestions/publicity in order to maximize the utility of said suggestion/publicity.

Determining Business Objectives

The objective is to create greater margins, as for any venture. However the way of doing this is what concerns us. We want to optimize the amount of information shown. We have two resources here: customer's attention and the money poured into banners/publicity, etc... We want to use as little as possible of this two resources in order to increase margin. The way of doing this is maximising the likelihood of showing publicity to customers of products that they will actually consider purchasing.

Tasks

- to draw a correlation between past reviews and future purchases using existing public data.
- To find a model which presents banners that feature a product with better than random chance of being purchased.
- increase customer satisfaction by offering a most efficient use of their time.
- To do so by not increasing operation costs, or at least not so much as to offset any gains made by this project.

Business Success Criteria

- For our model to work using a control group outside our initial data set. A standard deviation above what would be expected by mere randomness is a good goal.
- To be reliably used without offsetting the accuracy the previous model had procured. It must be at least as accurate.
- The project finishes on time.

Assessing The Situation

There are data compilations from Amazon reviews from the past couple of years. Legislation changes fast, however, and there is not complete certainty that any attempt at scraping/mining won't result in legal complications. Further investigation is required.

There is no inside person within the company working with me, however the wealth of data on this site and how far it has spread should suffice for our operation to be a successful one, at least in principle. Truth is we can only trust data from a few years ago to a certain extent – determining what it is is out of the scope of this project, hence why it would be ideal for our data to be as recent as possible.

Resources Required

For now our applications aren't that computationally demanding, a personal home computer should suffice. We also count with this data set ¹ which is nicely formatted in JSON format for our use.

Risk and Contingencies

We have two main concerns:

- For our data to be poor quality.
- For the project to take more than expected.
- For our model to be less than successful.

In the first case, we'll have to scrape. In that case the second concern becomes ever more present and a new concern appears:

- Legality regarding the scrapping of Amazon data.

Determining Data Mining Goals

To relate reviews made by users with the time in which they were made. After that we want to determine the likelihood of a certain item to be purchased given this historical data. Using the data we are given our task becomes organizing it so as to use it as a predictive tool.

Project Plan

1. To understand the business and our objectives (done)
2. To gather the data using local storage
3. To understand the structure of this data

¹<http://jmcauley.ucsd.edu/data/amazon/>

4. To model a solution using said data
5. To evaluate the model
6. To release our model