

Sentiment Analysis and Product Recommendation on Amazon's Electronics Dataset

Capstone Project 2: Proposal

1. What is the problem you want to solve?

Problem:

The Internet has revolutionized the way we buy products. In the retail e-commerce world of online marketplace, where experiencing products are not feasible. Also, in today's retail marketing world, there are so many new products are emerging every day. Therefore, customers need to rely largely on product reviews to make up their minds for better decision making on purchase. However, searching and comparing text reviews can be frustrating for users. Hence we need better numerical ratings system based on the reviews which will make customers purchase decision with ease.

Goal:

During their decision making process, consumers want to find useful reviews as quickly as possible by rating system. Therefore, models able to predict the user rating from the text review are critically important. Getting an overall sense of a textual review could in turn improve consumer experience. Also, it can help businesses to increase sales, and improve the product by understanding customer's needs.

In this project, the amazon review dataset for electronics products will be considered. The reviews and ratings given by the user to different products as well as reviews about user's experience with the product(s) will be considered.

The main goal for this project is to develop a model to predict user rating, usefulness of review and recommend most similar items to users based on collaborative filtering.

2. Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

Amazon is our client. The company wants to develop a software tool that will identify the positive and negative words which customers use when they write reviews for the home and kitchen products as their purchase inclination.

Getting an overall sense of a textual review could in turn improve better consumer decision making experience. Also, it can help businesses to increase sales, and improve the product by understanding customers' needs and problems.

3. What data are you using? How will you acquire the data?

Amazon Electronics Reviews and Product Data:

For this project, Amazon electronics product reviews and meta data will be downloaded from <http://jmcauley.ucsd.edu/data/amazon/>

4. Briefly outline how you'll solve this problem.

Outline:

1. Data collection

Tools: Python, sql, python Requests, json packages, jupyter notebook

2. Data wrangling

Tools: Python, jupyter notebook, Matplotlib, pandas, numpy

3. Exploratory analysis

Tools: Python, jupyter notebook, Matplotlib, pandas, numpy

4. Visualization

Tools: Python, matplotlib and seaborn

5. Sentiment Analysis and Recommendation system

Tools: scikit,NLTK

6. Summary and Report

Tools: Jupyter notebook

Outcomes:

In this project, following insights will be achieved:

- Predicting ratings based on reviews
- Usefulness on large volume of reviews
- Rating vs number of reviews
- Rating vs proportion of reviews
- Helpful proportion vs Number of reviews
- Rating vs helpfulness ratio
- Top 20 most reviewed products
- Bottom 20 reviewed products
- Positive and negative words

5. What are your deliverables?

Deliverables include code, detailed report and slide deck.