Brooke Ann Coco
Project 4 Proposal
Topic Classification of Product Reviews
May 13, 2019

Overview:

The goal of this project is to improve product development and design by utilizing Natural Language Processing (NLP) to analyze consumer reviews and cluster opinions about certain aspects of a product, such as user interface, usability, intuitiveness, etc. Though the scope of this project would focus on a handful of products, this algorithm has the potential to scale to any type of product from microwaves to hospitals to corporations. The business case would be to enable producers and providers to identify what is and is not working from the eyes of their consumers. This algorithm could be used to supplement user research teams by not only confirming human findings, but also capturing insights that the team may have missed, overlooked, or deemed unimportant.

Data:

The data is being collected via scraping product review websites such as Trustpilot.com, aaareview.com, Amazon.com, Google.com, and customerreports.com. For my MVP, I would focus on the reviews of one type of product, for example microwaves, then expand out to a handful of other products.

Potential Flaws:

In a job setting, I would presumably have access to data such as focus group transcripts, interviews, etc. that was conducted by the company itself. Because I do not have access to this kind of information, I will have to rely on customer reviews that I am able to obtain online. Now this can be a legitimate source of information, however when retrieving information online there is always the potential for bias. For instance, I cannot discount the potential for fake reviews. One possible solution would be to limit my observations to reviews containing over 200 characters.

A second concern that would be more difficult to account for is the population bias. This study would only be representative of those who choose to write reviews online, not the general population. I would be impossible to suss out the motivation behind writing a review. Perhaps the reviewer was paid to leave a positive review. Perhaps a rival paid them to leave a negative review. Additionally, there is the potential that people who feel strongly one way or another about a product will leave a review, while those who feel neutral will not. Ideally, my sample would not be biased by these external factors, however, given the scope of this project and the resources currently at my disposal, I will not be able to account for these. The best I can do is be conscious of these biases throughout my analysis, as well as communicate these limitations to stakeholders.