Predict Potential Usefulness and Positivity of Product Reviews

Romane Goldmuntz, Mansi Parikh, Jane Zhang, Hanh Nguyen, Joyce Yu

Business Problem

Both consumers and e-commerce businesses rely on the ubiquity of user-generated product reviews to make better-informed decisions. Reducing the costs of identifying relevant information from vast amounts of product reviews aids potential consumers with their purchase decisions and informs businesses of key product and/or service development areas to focus on based on their importance to their consumers. “Ratings and review content is having greater impact on consumer behavior in the COVID-19 era, providing the validation and social proof necessary to drive sales” (PowerReviews, 2020). Research suggests that consumers are seeking validation for their purchases in such uncertain times more than ever, yet the length of their review periods have shortened with increased stress and/or the greater need to multitask (PowerReviews, 2020). To positively influence consumers’ purchase intention in these accelerated and abbreviated consumer journeys with limited information, businesses may identify the most useful reviews in order to increase customer satisfaction via:

- Improving the product/service(s) offered and the overall customer experience based on feedback extracted from quality reviews, which would likely boost consumer confidence in purchases from the brand;
- Reducing friction in the review engagement process of the consumer journey by identifying and presenting high-quality user-generated information about the product/service.

While websites like Amazon allow consumers to vote for reviews found to be helpful, newly published ones that contain relevant, up-to-date information may go unnoticed when buried under top-voted reviews. Furthermore, users may focus on reviews with the highest and lowest ratings, neglecting useful information provided in average-rated reviews. To cope with the inundation of product reviews old and new, businesses may wish to assess and rank reviews by their usefulness and positivity. In particular, we are interested in understanding the properties of product reviews that signal their usefulness, and ultimately, creating predictive models that output the helpfulness and positivity of product reviews based on the review content.

Business Impact

Our product is a dashboard that provides the 5 positive reviews and 5 negative reviews that were found the most helpful by consumers to Amazon’s vendors. We are aiming at providing them consumer feedback and other strategic insights (industry trends, competitor information) that will help them improve the quality of their product or service.

We also believe that our deliverable will benefit Amazon as it will improve vendors' experience on the platform and increase their engagement with their consumers.

Data

We have chosen to focus on Amazon reviews datasets that are related to the beauty industry:

- All Beauty reviews (UCSD) - 370K reviews that represent 33K products
- Luxury Beauty reviews (UCSD) - 575K reviews that represent 12K products

We are planning to implement substantial manual filtering to identify the reviews that are suitable as model inputs. For example, we will ensure that all recent reviews published at the time that the data was pulled are to be eliminated from analysis given that they have had little time to accumulate adequate votes on the platform to be deemed useful. Additionally, since we are planning to display at least 5 positive and 5 negative reviews for each product, we will likely focus our analysis on the products that have a sufficient number of reviews. The dates on which reviews were published will be used to perform the first type of filtering, after which we will ensure that there are enough products for analysis.

Since Amazon reviews only use votes as an indication of helpfulness (i.e. there are no opposing indicators, e.g. for irrelevance) and the number of views per review is unknown, it is unclear whether or not a given review was found to be more helpful than others due to differences in views and votes accumulated over time. As a consequence, a review's helpfulness will be inferred from its relative vote count compared to that of other reviews.

Finally, the products are not labeled with the distributor or manufacturer (vendor), so we may have to manually assign these labels in order to utilize such information for related features on the vendor dashboard to be built.

Methods

Visualizations

As part of our exploratory data analysis, some of the visualizations that will be constructed to verify assumptions include:

- A strip plot of product reviews by rating and votes;
- A correlation heatmap of review parameters (e.g. message length, title length, number of votes);
- Vocabulary lists / word clouds of common opinion words used in positive and negative reviews;
- A histogram of all product reviews from available datasets by votes.

In order to provide high-quality visualizations of review-related insights to the end user (i.e. Amazon vendors), the following visualizations will be presented on a dashboard:

- A list of the top five most useful positive and negative reviews by product;
- Word clouds of keywords from both positive and negative reviews
  - for a given product sold by the vendor;
  - for similar products sold by competing Amazon vendors;

    - for all products in the given product category;
- A list of actionable recommendations for product/service improvements based on all product reviews analyzed.

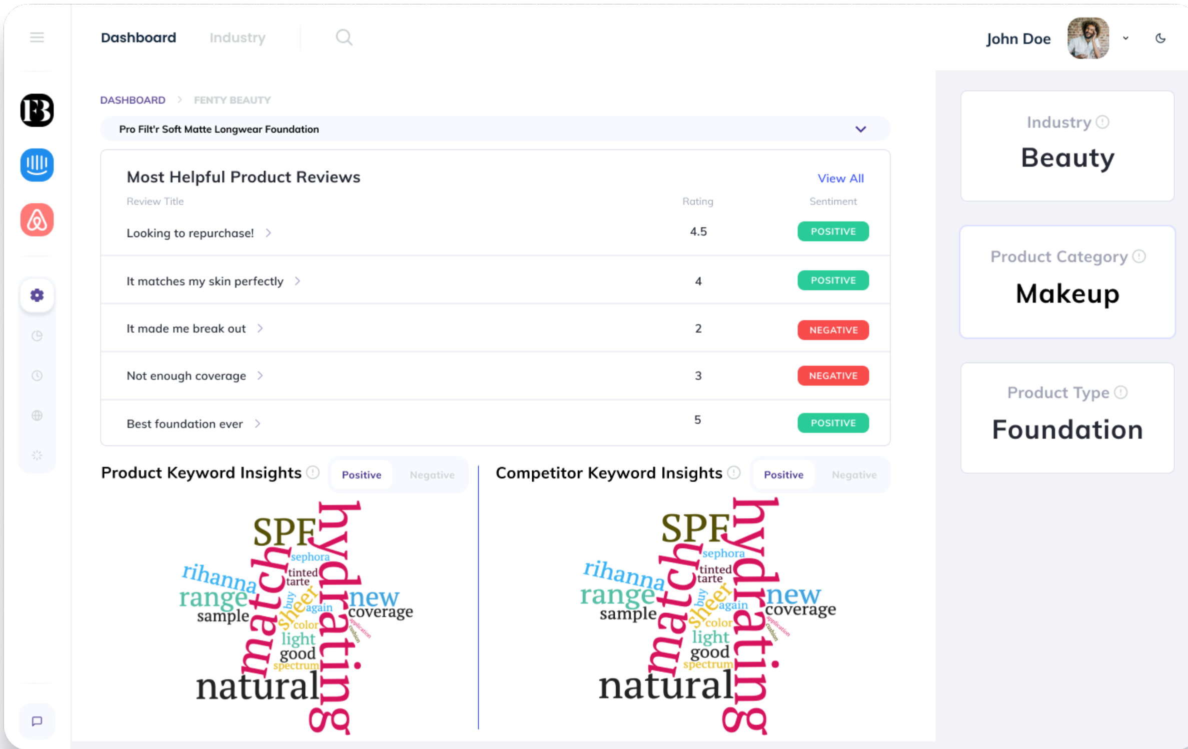
Models

- Predictive model to predict helpfulness against number of votes
    - Explanation: The most general traits of helpful online reviews are usually its level of detail (and usually that's a function of how long a post is), whether it touches on many different product attributes (and that thoroughness is usually achieved by a combination of several different word choices), and whether or not they provide sufficient feedback on a number of items that are important in a purchase decision, such as cost, durability, etc. Since these characteristics are mostly what influence people to cast a vote for a post and deem it "helpful," these will be used as the predictor variables. Unfortunately, this is still considered a rather subjective quality of a post, but if several people hold the same belief (as seen by the vote count), we can safely assume that the post is indeed helpful.
    - Details: Significant feature engineering is required to prepare the variables that will be inputs to the classification algorithm.
        - For the dependent variable (helpful or not), we will use some standardized or normalized version of vote count to set a threshold and classify a post into either of the two categories
        - We will have to parse the text reviews and assess whether the posts are inclusive of the traits mentioned above that deem a post helpful or not; to do this, we have established a vocabulary list to indicate when a post touches on these details
- Predictive model to predict if the review is positive or not (sentiment analysis)
    - Explanation: This model is aiming at predicting whether a review is positive or negative and at preventing any ambiguity with 2- and 3-star reviews. We can label 1-star reviews as negative and 4/5-star reviews as positive with a high certainty, but 2 and 3-star reviews are less obvious. We cannot rely only on the rating; we would have to read the review to understand whether the feedback is positive or negative. We are looking to automate this task by building a model that will identify the sentiment of these ambiguous reviews.
    - Details: We will be using several NLP techniques and algorithms whose performance for NLP tasks has been proven in the past (literature and applications) on the review text.

Interface

The final interface would be a dashboard with three main features:

1) View the top 10 reviews for a selected product (5 positive, 5 negative);
2) View insights extracted from sentiment analysis in the form of a word cloud;
3) View keywords (from both positive and negative reviews) from competitors' reviews.



The interface would be an end-user product marketed to vendors who sell their products on Amazon, allowing them to gain insights into the sentiments expressed in the reviews on their products. The vendor may select a specific product that they want to view the insights for. For each product, a summary of the most useful product reviews will be shown, along with buttons to be used to view further details of the reviews (e.g. date published, review message). Vendors may use the Positive/Negative toggle button to view a word cloud made of keywords from helpful positive/negative reviews.

Milestones

With consideration towards project time constraints and the skillsets of team members, the project work can be broken down to phases that result in the following four versions of the deliverable (in terms of data, analysis, and visualizations):

- **Version 1**: A dashboard that presents the top five positive and negative reviews ranked by helpfulness and positivity metrics output by the aforementioned predictive models for a given product.
  - *Estimated probability of completion*: 100%
- **Version 2**: In addition to functionality included in version 1, the dashboard would present the common keywords and phrases observed in the positive and negative reviews of a given product in the form of word clouds or similar text data visualizations.
  - *Estimated probability of completion*: 70%

- **Version 3**: In addition to functionality included in version 2, the dashboard would present the common keywords and phrases observed in the positive and negative reviews of competing products in the form of word clouds or similar text data visualizations.
  - *Estimated probability of completion*: 20%
- **Version 4**: In addition to functionality included in version 3, the dashboard would present a list actionable recommendations for product/service improvement based on the top-ranked negative product reviews.
  - *Estimated probability of completion*: 5%

Timeline

Date	**Deliverable**	**Details**
Week 1	Exploratory data analysis	● Data cleaning: ○ Remove non-voted and recently published product reviews, as well as duplicates in combined dataset ○ Standardize votes across reviews to select a threshold for helpfulness and positivity ● Labelling
Week 2	Model building	● Design and build the aforementioned two predictive models ● Build backend infrastructure (e.g. review ranking and fetching)
Week 3	Dashboard construction	Complete frontend infrastructure of web app (see ‘Milestones’)

Concerns

- No members have experience with building dashboards;
- Having 4 different versions may be ambitious given the anticipated time required to ramp up on dashboard construction;
- Filtering review stars and votes may not exclude certain edge cases that affect our analysis (e.g. reviews may include “fake” votes; buyers may have created “fake” reviews; reviews may have positive comments but negative stars; etc.)
- 10 reviews (5 positive and 5 negative) might be a small sample size to be representative of all customers’ opinions;

- Review content may be biased, as consumers might write their reviews according to their level of satisfaction on other aspects such as service or price, not just the products alone. For example, in the case of conflicts, customers might give out negative reviews regardless of the product's quality.
- Online reviews are subject to manipulation (e.g. "review reuse", "review hijacking", fake reviews from competitors).
- Review helpfulness is inherently a subjective quality since each post is open to interpretation by all of its readers, but there is some objectivity involved based on how many people found a review to be helpful (i.e. relatively high number of upvotes).
- 3-star reviews may have been intended to just be "neutral" (since they are in the middle of an odd-numbered Likert scale), though we are forcefully extracting even the slightest bit of polarity from them when they are not meant to be interpreted as either positive or negative.

References

PowerReviews. (2020, May 5). PowerReviews Market Trends Snapshot – April 2020. Retrieved from PowerReviews: https://www.powerreviews.com/insights/impact-consumer-ratings-covid-19/