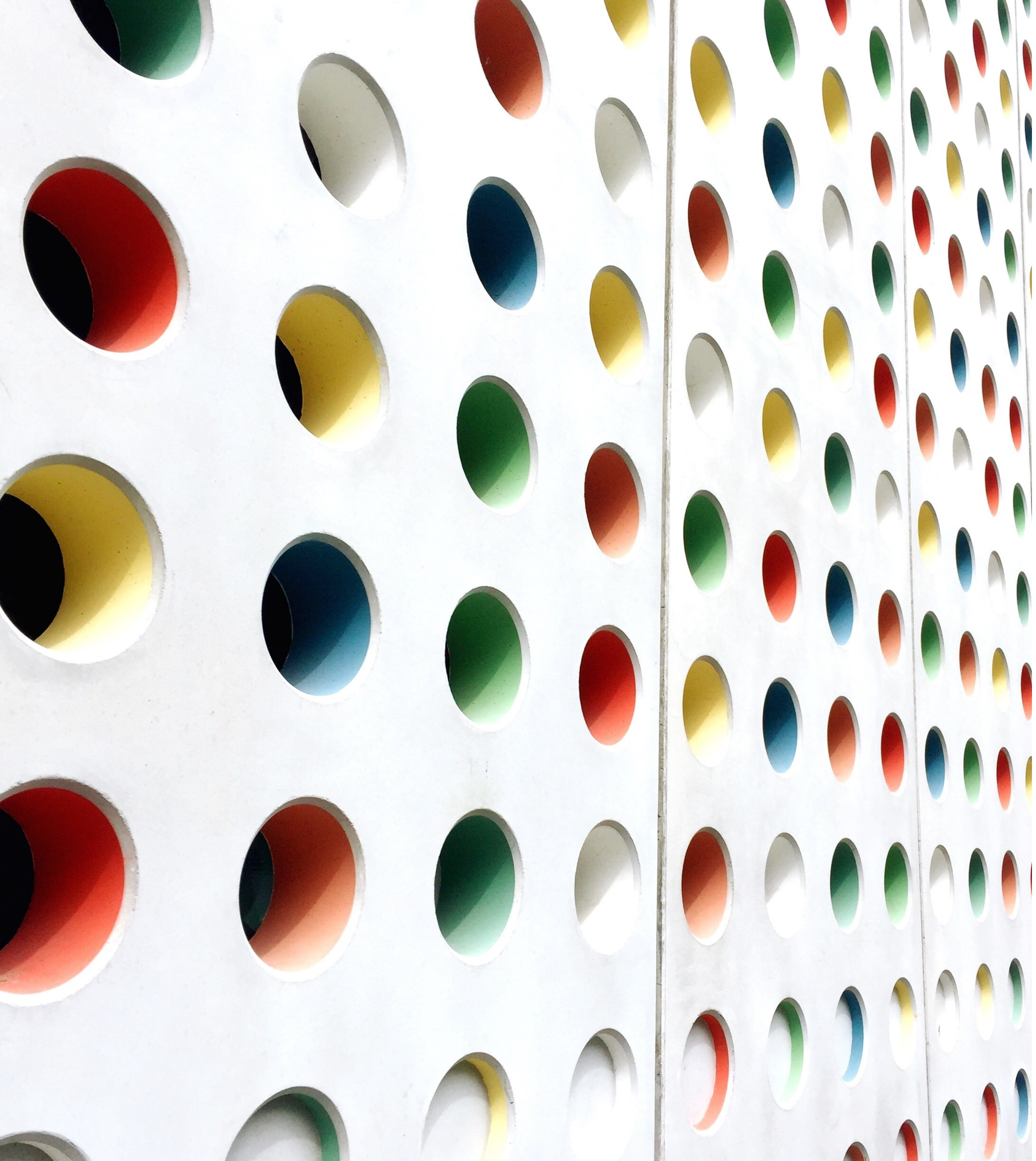


Text and Cluster Analysis of Facebook Reviews on Trustpilot

NICK KACHANYUK

12/11/2020

WILLAMETTE UNIVERSITY

Research Questions

Motivation

- Trustpilot is business review site where anyone with an email or Facebook account can register and leave a review about a business company.
- Trustpilot states that they combat fake reviews (spam, non-genuine experience) and reviews that do not follow community guidelines (hate speech, etc.)

Research Questions

- What are the different topics (technical issues, censoring, social media experience, etc.) that the Facebook reviews are written on Trustpilot?
- How these reviews can be explored via text and sentiment analysis to detect reviews that are not genuine?

Existing Theory & Evidence

- Fang, Xing, and Justin Zhan. "Sentiment Analysis Using Product Review Data." *Journal of Big Data*, vol. 2, no. 1, 16 June 2015, 10.1186/s40537-015-0015-2. Accessed 23 Apr. 2019.
- Key Takeaways:
  - Sentiment analysis is hindered due to fake, spam, and non-genuine reviews that are submitted
  - Neutral sentiments make it difficult to categorize reviews into distinct categories
  - They found that categorizing reviews becomes difficult when classifying reviews to their specific star-scaled ratings which caused them to report F1 scores less than 0.5
    - In my EDA, I found a similar problem occurring

Existing Theory & Evidence (continued)

- Hiremath, Prakash & Algur, Siddu & Shivashankar, S.. (2010). Cluster Analysis of Customer Reviews Extracted from Web Pages. Journal of Applied Computer Science & Mathematics. 4.
- Key Takeaways:
  - Features/IV can be used to segment reviews into categories (most significant, more significant, significant, and insignificant)
  - Found that a significant number of features belong to the insignificant review category

Existing Theory & Evidence (continued)

- Ng, James. "Natural Language Processing (NLP) to Analyse Product Reviews by Online Shoppers." *Medium*, 16 Apr. 2020, towardsdatascience.com/natural-language-processing-nlp-analysis-of-product-reviews-by-online-shoppers-7a5966f5e615. Accessed 10 Dec. 2020.
- Key takeaways:
  - Topic modeling has three uses (descriptive, predictive, and prescriptive analytics)
  - Descriptive is about understanding what the reviews are written on (my focus in the project)
  - Predictive is about understanding how the product will sell, attract new customers, and so on given a certain rating
  - Prescriptive is about looking at reviews left by the consumer and recommending them other products based on their previous reviews using an algorithm

Dataset and Variables

Data

- Obtained from Kaggle
- A collection of reviews written on Trustpilot about Facebook
- 3261 total reviews; only 58 were used in further analysis

Variables

- Time (when the review was written on Trustpilot)
- Num_Reviews (total number of reviews on Trustpilot associated with a particular account)
- Rating (a score given by the Trustpilot reviewer)
- id (identification number for a review)
- review (all the written text including both the title and the body of the review)
- afinn_sentiment (total sentiment score of a review generated after sentiment analysis)
- negative (binary variable indicating overall negative sentiment)
- neutral (binary variable indicating overall neutral sentiment)
- positive (binary variable indicating overall positive sentiment)

Methodology

Sentiment analysis

- AFINN lexicon (word sentiments w/ -5 to +5)
- New variables (afinn_sentiment, positive, neutral, negative)

Cluster Analysis

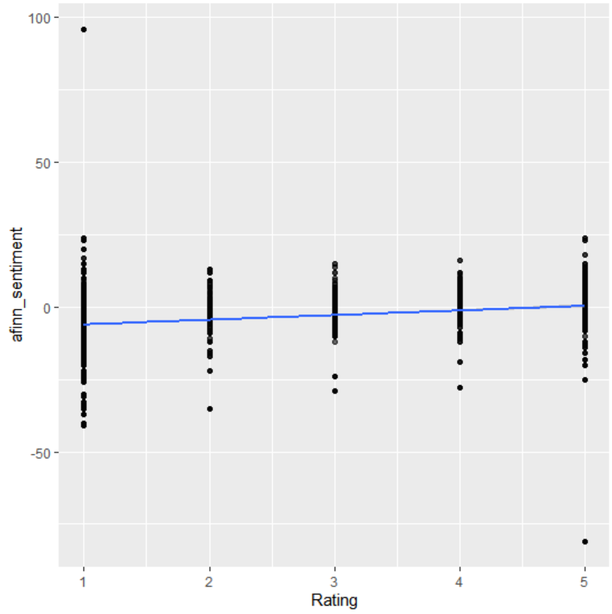
- Elbow, Silhouette, and Gap Statistic method

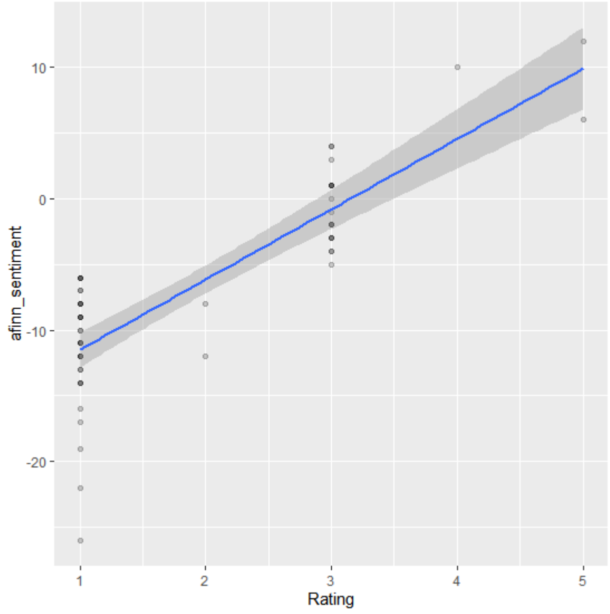
Topic Modeling

- LDA with Gibbs random sampling
- Perplexity per number of topics
- k =2 and k =7 were used; where k is the number of topics

Results (sentiment analysis)

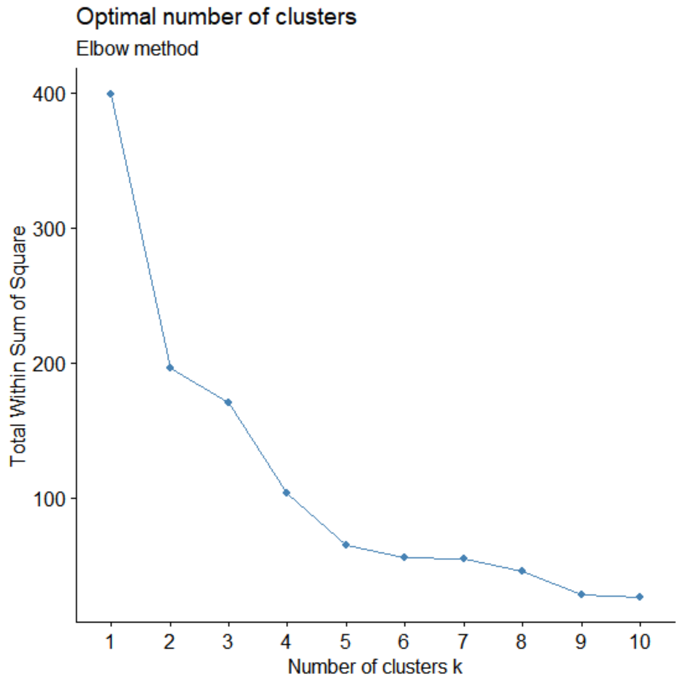
- Reduced the number of reviews from 3261 to 58
- Found that many reviews were given inappropriate ratings; for example, reviews with high sentiment scores but with a rating of 1, or reviews with a low sentiment scores but with a rating of 5
- One of these reviews was written in English and Danish; there were more Danish words and they expressed negative sentiment but there were randomly placed English words that expressed positive sentiment and were not stop words. Coincidence?



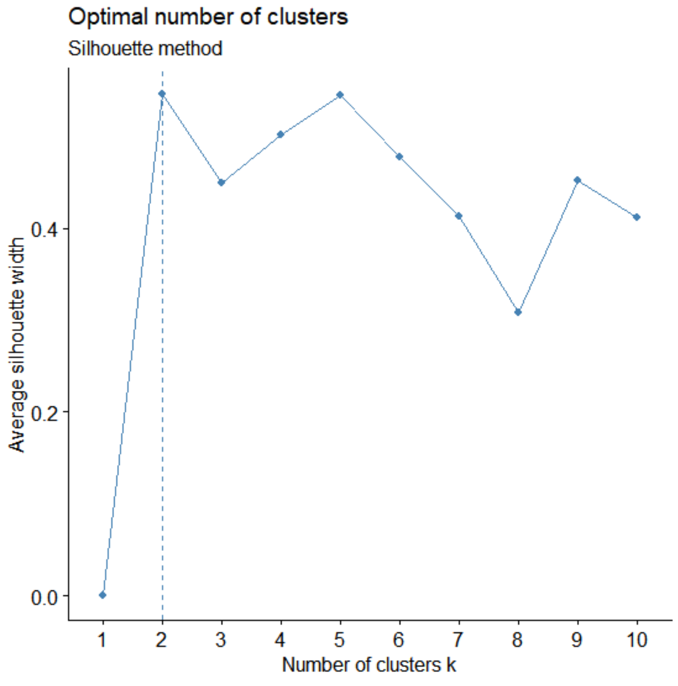


Results (cluster analysis)

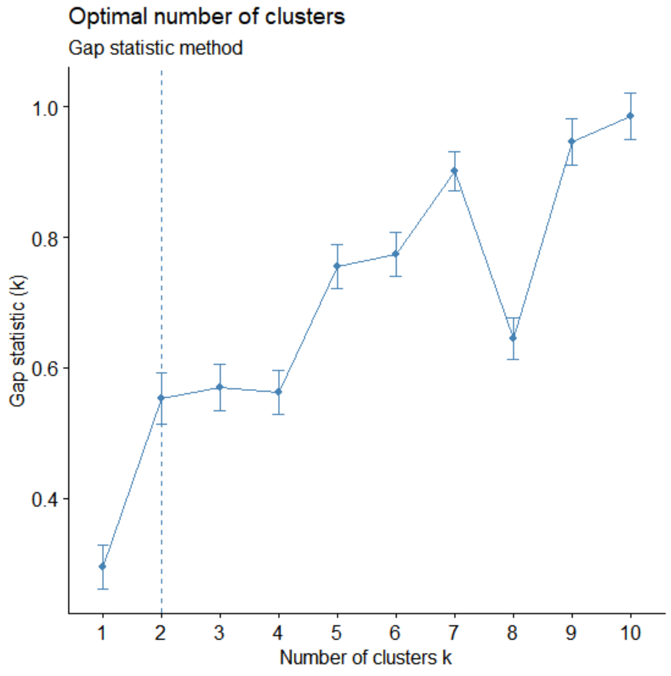
Elbow method



Silhouette method

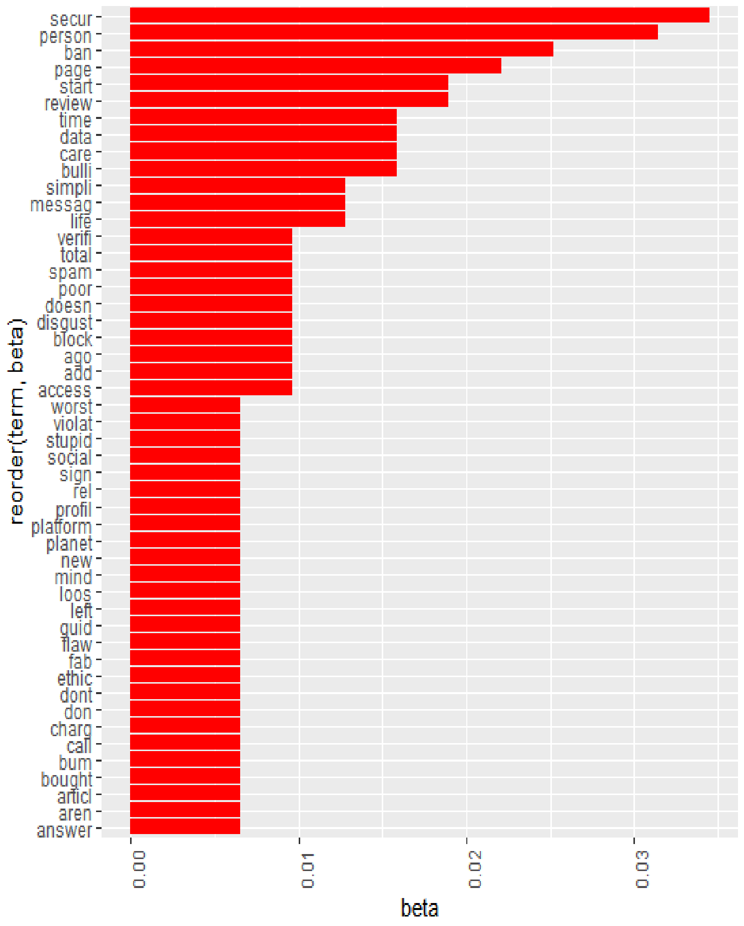


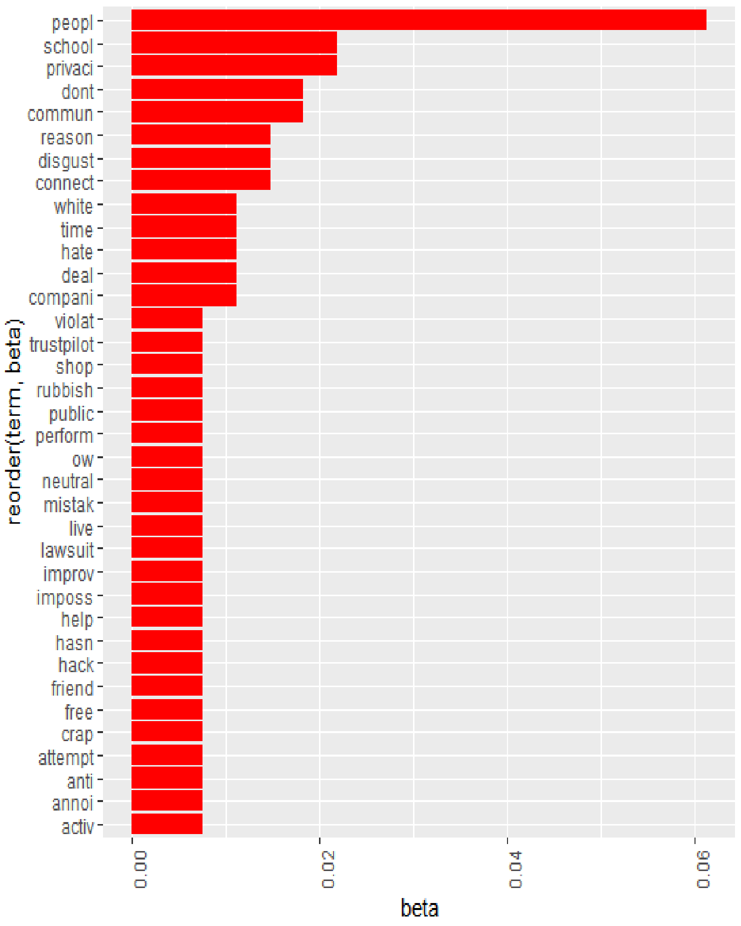
Gap Statistic method



Results (topic modeling)

- Both k =2 and k =7 were used in the LDA models; k = 7 was based on perplexity
- For each topic I examined top 10 beta values (top words per topic) and top 10 gamma values (top reviews per topic)
- Read 20 reviews obtained from the LDA results where k = 2, to find out what the topic themes were:
  - Topic 1: security issues, scammers, and privacy issues
  - Topic 2: censorship, poor customer service, poor social experience, poor FB community standards





Conclusions

- Sentiment analysis was useful in the EDA and creating new variables that would allow me to examine the quality of the reviews
- Cluster analysis was useful in determining the number of possible topics based on **numeric** variables
- Perplexity (how well a model fits new data) was useful in determining number of topics based on **text** data
- Topic modeling was useful in determining top words per topic and top reviews that represented each topic for a given k value.