Allyssa Utecht, Jimmy Horvath, and Andrew Plate

BUS306 - Text Analytics

Section 1 Group 6

**Kickstarter Final Project: Written Report**

The Kickstarter Final Project presented our group with the challenge of creating a function that would consistently determine the future success or failure of a Kickstarter campaign through various language analysis tools. Our group sought to find the factors that would best indicate whether or not backers would be willing to support a certain project, solely based off the marketing pitch for said Kickstarter campaign. Through research of various sources ,and trial and error with testing our function on sample Kickstarter campaigns, our group believes we can accurately predict whether a Kickstarter campaign will be successful in being backed. There is a strong, casual connection between the language used by project creators and the subsequent reaction of potential backers. As we see in multiple articles from Mitra and Gilbert, Kaminski and Hopp, and Zhou, Zhang, Du and Qiao, there is quantitative data illustrating that the words used, and how many words are used, directly influence whether a potential backer chooses to fund the campaign or not.

Our initial approach to the challenge of predicting a Kickstarter campaign success or failure, via a text analysis of the campaign description, was fairly simple. We initially started with a similar model to the one we used for the earnings report project. We created and defined the fog function, as well as the sentiment analysis. Then, we defined the kick\_predict function, separated into story and risk sections. We originally ran a fog of the entire text and added this to the score, as well as a fog of risks section and subtracted this to the score; we did this because we thought if the risks section was more complicated, it meant the campaigners were hiding something. Next, we put the entire text into lowercase, followed by tokenizing the entire text. Then, we ran a Lexicon analysis on the text, using a list of positive and negative words from the Kaminski and Hopp article. We also included some financial-related words in the ‘negative’ category, again due to insights derived from Kaminski and Hopp. Their study concluded that “all terms related to monetary depictions of the venture reduce the chances to reach the campaign goal successfully.” This was contrary to our initial beliefs and something we adjusted to quickly in accordance with our research. Furthermore, we originally focused on the readings “Language that Gets People to Give: Phrases that Predict Success on Kickstarter” and “Money Talks, A Predictive Model on Crowdfunding”. The former of the two outlines specific phrases that can influence potential backers through the ideas of social proof and liking. In this, people are more likely to comply with a person or product if they like them. Positive remarks about another person’s attitudes and performance increases liking. The “Money Talks” reading viewed project descriptions as marketing tools for crowdfunding campaigns and valued 3 aspects of the “amount of data, ease of understanding, and objectivity”. Based on our prior knowledge with the earnings report project, we felt these articles would initially lead us in the right direction.

Following our research through various articles, we began to create our function. We originally had many issues with our fog function, because you cannot run fog() of something that is an integer. After much troubleshooting, we realized that our sentence tokenize function was accidentally returning the length of the sentence - an integer - and not a string of text, so we had to adjust that portion of the fog function, and then the errors were fixed. We also ran into many issues with the phrases that separated the story section from the risks section; some of the sample texts didn’t have a risks title, so we had to make an if/else statement there. Additionally, we had to repeatedly adjust weights on the various score factors, to improve the accuracy of the returned scores. We decided to eliminate the negative fog of the risks section, because we realized that a higher fog might actually be better - it is better to be more transparent about the risks of a technology or design product, and that it might just take more words and more jargon to explain. For the testing, and text samples gathered, we began with 30 randomly selected Kickstarter campaigns, 16 positive and 14 negative, from the excel sheet given by Professor Davis. After some testing, we reverted back to research and looked into the “How Much is Too Much? The Effects of Information Quantity on Crowdfunding Performance” article. This reading influenced us to add a length function to both, the entire text, and risks section. The article found that there is an “upside down U-shaped curve” that correlates the relationship between length and success of Kickstarter campaigns. In this, campaigns that are on the extreme ends of either too few or too many words do poorly in garnering backers, while those in the middle have found better success. The article determined that the optimal word count, where the funding increase peaks, is roughly 1681 words. Our group chose to set our optimal length to 1700 words, because the graph for technology and design campaigns peaks at a slightly higher word count than other categories.

Following the implementation of this factor, our code was able to successfully predict all of the failures from the test sample, but only 5 of 16 successes. From this, we looked into changing the weights associated with each aspect of our code, especially within the lexicon analysis. From the Kaminski and Hopp article, we found that “linguistic styles that aim to trigger excitement or are aimed at inclusiveness are better predictors of campaign success than firm-level determinants.” The ability to make potential backers feel excited about a particular product or involved within said product are powerful determinants in predicting success of a Kickstarter campaign. From this, we added an additional lexicon of words and phrases to be analyzed in the story section, in addition to the original lexicon of positive and negative words and phrases. These words and phrases specifically pertain to reciprocity, scarcity, social proof and identity, and authority. As gathered from the Mitra and Gilbert article, the phrases ‘mention your’ and ‘pledged will’ relate to the reward backers will get if their projects are funded, and have very high β weights, illustrating the power of reciprocity. The phrases ‘given the chance’ and ‘option is’ represent scarcity and exclusivity and have similarly high β weights; people are more willing to fund the campaigns they think are unique or rare. Additionally, the authors explain that people act based on how others act, so we included the “social proof” phrases they mention, such as ‘has pledged’ and ‘pledged and’. However, possibly the most influential phrases have to do with authority: ‘project will be’ (β of 18.48) and ‘we can afford’ (β of 2.94). By expressing a sense of assurance that the project will be funded, and they will be successful, the campaigners illustrate expertise and confidence. Overall, testing and adjusting our model primarily relied on changing the weights given to each factor. The biggest issue we found was that any changes would increase the accuracy of the successful sample campaign predictions, but would decrease the accuracy of the failure sample campaigns - or vice versa.

While our original focus was on the story section, and this remained in the final model, we came to understand that the length of the risks section is important. We were not vying for a perfect model, but rather one that is just more accurate than not. Mitra and Gilber stress how important the language used in the project description is for predicting the success of the campaign. Our final model is a more intricate version of the original model, with a much more detailed function to separate the story and risks, length functions added, and an entire new set of phrases for the additional lexicon - based on phrases from Mitra and Gilbert, not just the positive and negative words from Kaminski and Hopp. Our final model puts heavy emphasis on the words of the story section, because this is where the campaigners have the opportunity to explain themselves, and where they include the most emotion and excitement. We made the optimal length of the campaign text slightly longer than what is typically advised, because the technology and design campaigns usually need more words to explain the intricacies of their product.

To entrepreneurs looking to create their own successful Kickstarter campaign there are multiple ways to garner support from backers. While not included in our own model, our group found through research that the inclusion of images and other media can positively affect the outcome Kickstarter campaigns. Media paired with the Kickstarter campaign allows potential backers to get a visual glimpse of what they would be investing in. From a literary standpoint, it is important for Kickstarter entrepreneurs to make potential backers feel included in the project. People feel a bias for things that they are involved with, and if the campaign can provide this aspect through the pitch, then a backer is more likely to commit their money. Furthermore, if a pitch conveys positive emotion and entrepreneurial excitement from the campaigner side, then potential backers will also be more likely to invest. Backers want to support campaigns that the creator is excited and motivated in, as it demonstrates a potential for higher success. Lastly, the length of the text must fit in the sweet-spot, of not rambling on but providing an adequate description of the product. Too little words leave the potential backer confused or unsure of the product, whereas too many words become overwhelmed or seem as rambling from the potential backers view. Entrepreneurs who display confidence, assurance, positivity, and emotion are typically the most successful, because they believe in their product the most and want it to succeed the most.

Bibliography

Zhou, Mi Jamie & Du, Qianzhou & Zhang, Xuan & Qiao, Zhilei & Wang, G. & Fan, Weiguo.

(2015). Money Talks: A Predictive Model on Crowdfunding and Success Using Project Description., August 2015., <https://www.researchgate.net/publication/315828262_Money_talks_A_predictive_model_on_crowdfunding_success_using_project_description?enrichId=rgreq-1b12b41ae5a483c17ae1e853b7f44890-XXX&enrichSource=Y292ZXJQYWdlOzMxNTgyODI2MjtBUzo0OTk3MjQ1Nzg5Nzk4NDBAMTQ5NjE1NTAzMjA1Mw%3D%3D&el=1_x_3&_esc=publicationCoverPdf>

Kaminski, Jermain C., and Christian Hopp. “Predicting Outcomes in Crowdfunding Campaigns with Textual, Visual, and Linguistic Signals.” *Small Business Economics*, vol. 55, no. 3, 9 July 2019, pp. 627–649, <https://doi.org/10.1007/s11187-019-00218-w>.

Moy, Naomi, et al. “How Much Is Too Much? The Effects of Information Quantity on Crowdfunding Performance.” *PLOS ONE*, vol. 13, no. 3, 14 Mar. 2018, p. e0192012, https://doi.org/10.1371/journal.pone.0192012.

Tanushree Mitra and Eric Gilbert. 2014. The Language that Gets People to Give: Phrases that Predict Success on Kickstarter. In Proceedings of Computer Supported Cooperative Work (CSCW’14), 49-61. http://dx.doi.org/10.1145/2531602.2531656