MoneyMouth: A Computational Analysis of Altruistic Crowdfunding Success on GlobalGiving.org

Stanford CS224N Custom Project

Danya Adib-Azpeitia
Symbolic Systems Program
Stanford University
dadib@stanford.edu

Abstract

NGOs often face knowledge barriers regarding how to write effective fundraising requests on crowdfunding platforms. This motivates the use of mixed computational methods to 1) identify the linguistic features associated with successful campaigns and 2) predict whether a campaign will be successful or not. Pearson correlation coefficients calculated on GlobalGiving.org projects suggest that features related to community-based thinking, storytelling, and readability are associated with campaigns that reach their target fundraising amount. The logistic regression model's most heavily-weighted linguistic features for the most part corroborate those results. The fine-tuned BERT model performed slightly better than the logistic regression at predicting a crowdfunding campaign's success.

1 Key Information to include

- TA Mentor: Manan Rai
- External Collaborators (if you have any): No
- Sharing project: No

2 Introduction

2.1 Motivating Problems

In charitable settings, the performance of persuasion faces unique challenges due to a potential lack of reciprocity. There is never a guarantee that what one chooses to fund (a non-profit's new program) will provide a return on the investment (demonstrated impact). Non-profits must often generate innovative methods to appeal to potential donors, especially on relatively nascent platforms such as crowdfunding websites. This "ask" is particularly challenging when it is mediated through a two-dimensional screen rather than a co-present setting, where the requester can rely on body language and other non-verbal communication elements to convince a potential donor.

Moreover, while entrepreneurial ventures frequently utilize crowdfunding strategies, NGOs may be overlooking the opportunity to compete to acquire funding and social pedigree in crowdfunding spaces. A recent report has shown that that donation-based crowdfunding should soon tally over $1 billion yearly (Massolution, 2021). Charities may also be looking to crowdfunding to address a younger audience interested in socially supporting projects and organizations. Despite the lucrative potential of crowdfunding, NGOs may not fully embrace this funding model as they may not know how to most effectively tailor their fundraising requests in crowdfunding spaces. Resultantly, the identification of attributes that can optimize a campaign's success is crucial in its applications.

2.2 Guiding Questions & Goals

Two critical questions guide this research: What linguistic features and styles does one employ to persuade good deeds, especially in the absence of a tangible reward? And which computational methods, whether top-down dictionary approaches or bottom-up automatic text analyses, best measure these linguistic features or predict a campaign's success?

Taken together, these two questions motivate explicit and implicit goals. Explicitly, these motivate the use of mixed computational methods to:

- **Goal 1 (G1):** identify the linguistic features associated with successful campaigns
- **Goal 2 (G2):** predict whether a campaign will be successful or not

Moreover, an implicit goal emerges from comparing computational methods: explore the tradeoffs between interpretability and predictive power. While traditional machine learning methods like logistic regression are highly interpretable (we can easily retrieve the features that were weighted most heavily), they are also much less sophisticated than larger language models empowered by deep learning, such as BERT. Meanwhile, these larger models are often considered "black boxes," employing automatic feature engineering which complicates interpretation of feature importance.

Our experiment incorporates both logistic regression and deep neural network methods to explore these tradeoffs. We use the Local Interpretable Model-Agnostic Explanations (LIME) package to assist with BERT interpretability, specifically to determine the relative importance of certain words in a campaign's success. Synthesizing these considerations, we hypothesize that logistic regression will better assist with interpretability **(G1)**, while BERT will perform better at prediction **(G2)**.

2.3 Potential contributions

The dominance of prose in online fundraisers and the reliance on individuals to choose which causes they will fund provides an environment for testing the impact of language use in enabling altruism, as donating to a crowdfunding campaign can be considered a successful behavior change compelled by persuasive stimuli. Thus, the answers to the guiding questions will illuminate psychological elements of altruism, such as fundraisers' and donors' motivations, and reveal successful communication strategies on emerging platforms, such as donation-based crowdfunding websites.

Hopefully, this project will democratize the knowledge needed to support the "underdogs" of the social impact world, equipping low- budget non-profits with high-impact skills. Beyond a pathway for service, this project also aims to unify ostensibly disparate academic disciplines such as communications, psychology, statistics, and machine learning. This cross-disciplinary approach can potentially contribute to some much-needed interpretability work in applying deep learning to language that is grounded in sociolinguistic theory. Additionally, the use of mixed computational methods can hopefully better inform social scientist researchers who are looking to expand their analytical toolkits.

3 Related Work

Many researchers have tried to identify features that contribute to fundraising success in specific contexts such as crowdfunding. Some posit that social factors such as geographical proximity and the fundraiser's social network greatly influence the success of a crowdfunding campaign (Agrawal, 2011; Belleflamme, 2014). Others assert that more consumer-based concerns impact success, such as the perceived quality of the product and rewards associated with contributing to the campaign (Mollick, 2014). However, most of this research has focused on for-profit crowdfunding, and these results may not necessarily hold in donation-based crowdfunding.

Literature investigating donation-based crowdfunding yields fewer results. Generally, scholarship focuses on the donor's intentions and identity, rather than the requester/recipient. A 2020 study titled "Mapping the Field of Donation-Based Crowdfunding for Charitable Causes: Systematic Review and Conceptual Framework" surveyed 92 publications and found that "a majority of publications from 2015 onward . . . using quantitative methodologies focuse[d] on antecedents related to individual donors, organizational promoters as main actors, and online channels and design-related features of campaigns as enablers" (Salido-Andres, 2020). This paper suggests few have used a computational linguistic approach to analyze donation-based campaigns. One study, conducted by researchers at the

University of Rochester did conduct text analysis, along with computer vision methods, to identify attributes of successful GoFundMe campaigns (currently the biggest donation-based crowdfunding platform). They found that third-person pronouns such as “she” and “he”, words regarding emotion, and social thinking are positively correlated with the success of a fundraiser in the Community & Neighbors category (Zhang, 2021).

4 Approach

We use five methods to address the goals listed in the abstract: Linguistic Inquiry and Word Count (LIWC), Pearson correlation coefficients, logistic regression, a fine-tuned BERT model, and Local Interpretable Model-Agnostic Explanations (LIME). At a high-level, LIWC assists with feature engineering, while Pearson correlation coefficient guides feature selection for regression. Logistic regression serves as a baseline for BERT, and LIME assists with BERT interpretability.

4.1 LIWC

LIWC is a popular text analysis tool used by computational social scientists. Developed by psychologist James Pennebaker and his team of researchers at the University of Texas at Austin, this program “counts words in psychologically meaningful categories” based on a rigorous assessment of word lists pertaining to those features. These lists contain words related to content (positive emotion, negative emotion, cognitive thinking, etc.) and function (pronouns, prepositions, articles, etc.), which reveal a speaker’s communication style. LIWC therefore assists with feature engineering, converting the textual data from the campaigns into numeric values. We decided on this top-down method because of its ease of use, its high validity, and its reproducibility across multiple data sets and studies.

(Note that we use LIWC-15," the version of LIWC developed in 2015; "LIWC-22" was released after our data had already been pre-processed.)

4.2 Pearson Correlation Coefficient

The PCC is a measure of the strength and direction of the linear association between two variables. This method helps address **G1**, identifying the linguistic features associated with a crowdfunding campaign’s success. We first calculate the "success ratio," the ratio of money raised to the funding goal (representing a campaign’s success). This then enables the calculation of the Pearson correlation coefficient between a project’s success and a LIWC category.

Baseline: We baseline with standard conventions in psychological research proposed by Cohen (1988), which determine the strength of a feature’s association. (See "Evaluation method" for more).

4.3 Logistic Regression

We take the LIWC features with the highest absolute correlation with our label as the feature inputs for our logistic regression model. (Features with correlation near zero will likely give little to no signal for our model to learn from.) Then, we train the logistic regression on these inputs to predict whether a campaign will be successful or not.

We needed to choose between treating the project as a binary classification problem (whether or not a project met its fundraising goal) or a regression problem (how much of its goal did a project meet). We originally decided that we wanted to do treat this as a binary classification problem because when looking at the ratio of funding to goal amount, we saw many outliers with extremely high ratios which would make it harder to train a regression model (After processing the completed projects, we noticed outliers where the ratio of money raised to the fundraising goal was greater than 1000. To avoid these outliers skewing the results of the correlation coefficient calculation, we assumed any project with a ratio greater than 0.8 is a success.

At first, we considered treating this as a regression problem, as a lot of information can be lost by labeling the output of the model as "successful" and "not successful." However, considering the intended use case and the outliers, we ultimately decided to go with binary classification.

Baseline The most positively and negatively-correlated features yielded from the Pearson correlation coefficient serve as a sort of "baseline" to compare against the features that the logistic regression model weighs most heavily. We use AUC to evaluate the logistic regression model's performance. These metrics will eventually serve as the baseline for our fine-tuned BERT model.

4.4 Fine-tuned BERT model

We downloaded a pre-trained BERT model and then fine tuned it: adding a shallow neural network on top of this BERT base, taking the embedding of the CLS token that is output by BERT and making the binary prediction. That is, we finetuned BERT following the format of single sentence classification tasks. we concatenated a campaign's title and description, feed the text through BERT, and take the output embedding of the CLS token as a dense representation of the campaign text. This embedding will be concatenated with the more traditional LIWC features, and finally fed through a shallow neural network to output the binary classification of whether or not this project will be successful.

Large transformer-based language models have shown state of the art performance on common NLP tasks. To try maximizing predictive performance, we finetuned a pretrained BERT model on the sequence classification task. For a project, the following project fields were hand-selected based on likely informativeness of the field in making the prediction: summary, need, activities, and long term impact. These four fields are found on the main project page of the GlobalGiving website and is the text which most readers will read when browsing projects. After concatenating these four fields, we take the output embedding of the CLS token as a dense representation of the campaign text and feed the embedding to the shallow neural network.

Notes on code attribution: The Pearson correlation coefficients is coded from scratch. LIWC was developed by James Pennebaker and UT Austin and is available for purchase (www.liwc.app). We downloaded a pre-trained BERT model from HuggingFace (https://huggingface.co) and then fine-tuned it by scratch. We used logistic regression from scikit-learn (https://scikit-learn.org).

Notes on method novelty: While other research has used a similar approach (Zhang et.al calculated PCC with LIWC features but employed a fusion classifier), we did not know this during development. In particular, we have not found any papers which specifically use the most correlated features yielded from the PCC as input for a LR, nor any which apply BERT to donation-based crowdfunding.

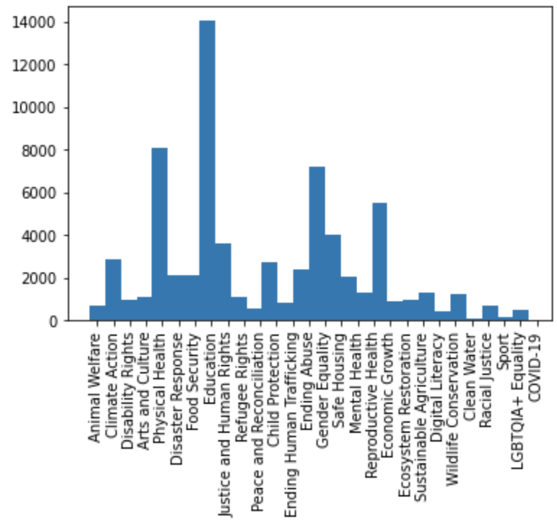
5 Experiments

5.1 Data

We analyzed linguistic determinants of success on GlobalGiving. Its position as the "first and largest global crowdfunding community that connects vetted nonprofits, donors, and companies globally" provides a specialized and vetted context that should reveal nuances in advanced persuasive techniques used for selective donors (rather than an average untrained eye). Non-profits must first undergo an intensive vetting process before they can publish their first crowdfunding campaign. These lengthy requirements imply that the projects on the website are from NGOs that should have experience communicating a pressing need, proposing a persuasive solution, and compelling donors. Yet, most campaigns do not reach its target fundraising goal. The data's robustness and organizations' varying fundraising success compelled me to use text analysis to study GlobalGiving and identify the features of successful persuasion in donation-based campaigns.

The dataset represents 27,000 campaign projects compiled from GlobalGiving.org using their public API. A "project" is one particular cause from an organization which someone can donate to. These pages are stripped of their pictures and years, and only contains texts about the campaign itself: as a project's "Title," "Summary," "Need," "Challenge," "Impact," and "Solution," as well as the fundraising target and resultant funding. The average word count of each project's associated text was 249 words.

Each of the 27,203 total projects yielded 93 LIWC numerical features (as described in "Approach"), which allowed for the calculation of the correlation coefficient between the success ratio and each of those 93 features. Future steps could include calculating LIWC features for projects grouped by a particular theme (such as COVID-19, Education, Racial Equality, etc.):



The plot above shows the breakdown of themes across the dataset. While not essential to training the logistic regression model, it is informative to see the frequency distribution of the different projects available on the Global Giving website. We see that the most common theme for projects is "Education", followed by "Physical Health" and "Gender Equality." This can inform further data segmentation and training future models in these thematic context to create better predictions.

Pre-Processing These 27,000 projects are a subset of all the projects on the website that have been "completed" (i.e. no longer ongoing, having passed the temporal deadline of their campaign). After removing the ongoing projects and processing these outliers (e.g. where the ratio of money raised to the fundraising goal was greater than 1000; discussed further in project proposal), I determined the finished projects' target and resultant funding.

Task To summarize, we take as input a project's text (represented numerically by LIWC features) as a prediction of its success (whether it achieves at least 80% of its target fundraising goal or not).

5.2 Evaluation method

Pearson Correlation Coefficients: To interpret the effect size of correlation coefficients (CC) in psychological research, standard practice is to use Cohen's (1988) conventions: a CC of .10 signifies a weak association; a CC of .30 signifies a moderate association; a CC of .50 or larger signifies a strong or large association.

Logistic Regression: To evaluate predictive performance, we use area under the curve (AUC). We aim for an AUC greater than 0.5 as that means our model is performing better than chance. To evaluate interpretability, we examine the features weighted most heavily and compare it against the most correlated features yielded with the Pearson coefficient calculations.

BERT: We again use AUC to evaluate the fine-tuned BERT model's predictive performance, comparing it against the logistic regression's AUC. Evaluating the model's most heavily weighted features is more complicated with BERT, but we compare LIME's identified features with the features identified by logistic regression and the Pearson correlation coefficients.

5.3 Experimental details

Logistic regression: 30 features (with the highest absolute value for the correlation coefficient). The learning rate was 0.00001. There were 15000 steps using gradient ascent. The training time was 17.44 seconds.

BERT: We finetuned the bert-base-uncased model with a learning rate of 3^{-8} for 10 epochs. Training time took 9 hours, and we used a training batch size of 6 due to GPU memory constraints.

6 Results

6.1 Pearson Correlation Coefficient

Table 1: Most Positively Correlated LIWC Features	
LIWC Feature	Correlation Coefficient
Third-person singular pronouns ("shehe")	0.2601756233658381
Present focus	0.22870164294312398
Personal pronoun	0.20015600044489953
Pronoun	0.1852388990248392
Second person	0.1852388990248392

Four out of five of the LIWC features that most positively correlated with a campaign's success were pronouns, with third-person singular pronouns ("shehe") ranking highest (Table 1). Taking the baselines into account, third-person singular pronouns with a coefficient of 0.26 therefore had a moderate association with campaign success, and the rest of the features in Table 1 ($0.18 < x < 0.23$) have a moderate-to-weak effect.

Table 2: Most Negatively Correlated LIWC Features	
LIWC Feature	Correlation Coefficient
Words/sentence	-0.1623
Words > 6 letters	-0.1483
Conjunctions	-0.1092
Prepositions	-0.0984
Analytic	-0.0972

The most negatively correlated LIWC features support the efficacy of clear narrative thinking (Table 2). Campaigns with wordy sentences, longer words, conjunctions, prepositions, and analytic thinking were least likely to reach their target fundraising goal. These features reflect complexity, as Pennebaker wrote that, "Complex thinking generally involves bigger words, longer sentences, and more complicated sentences, often involving prepositions." While dense sentences can reflect verbal fluency, they are often harder to read. And while analytic thinking reflects cognitive complexity, it is not as compelling as narrative thinking. Taking the baselines into account, these features ($-0.16 < x < -0.09$) have a weak effect.

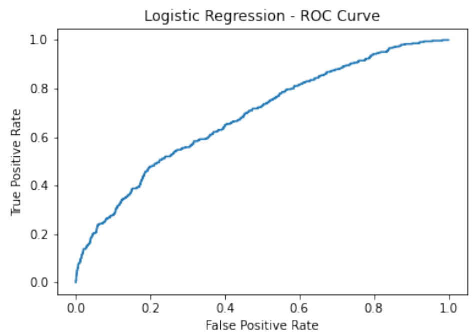
6.2 Logistic Regression

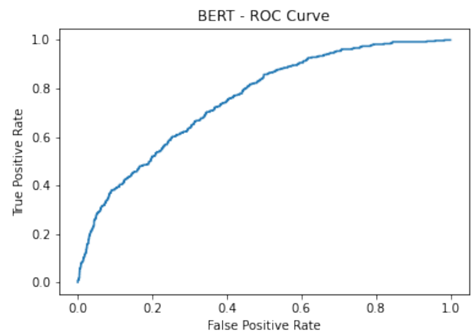
Table 3: Logistic Regression Feature Importance Weights	
LIWC Feature	Feature Weight
focuspresent	3.833559471295191
shehe	3.0135236714023805
Apostro	1.9073226304401008
time	1.6813563914155558
period	1.3779755645396496
Clout	0.8796438332378126
female	0.8110522227649374
focuspast	0.7494749093137564
Analytic	0.02002360059229464

Our model does somewhat confirm the most important linguistic features determined from our baseline of the correlation coefficients. See the plot below for performance evaluation, and Section 7 for interpretation.

6.3 BERT

Unfortunately, our interpretability method does not give global assessment of feature importance. That is, we must run LIME individually on each project, and its lengthy run time prevented us from being able to run LIME across all projects. We further examine interpretability in section 7.2.





The left figure shows the ROC curve yielding AUC of 0.686. Considering our baseline of 0.5, LR performs better than chance, and better than we expected.

The right figure shows the ROC curve yielding AUC of 0.754, showing that BERT finetuned on the sequence classification task outperforms LR trained with hand crafted features from LIWC.

7 Analysis

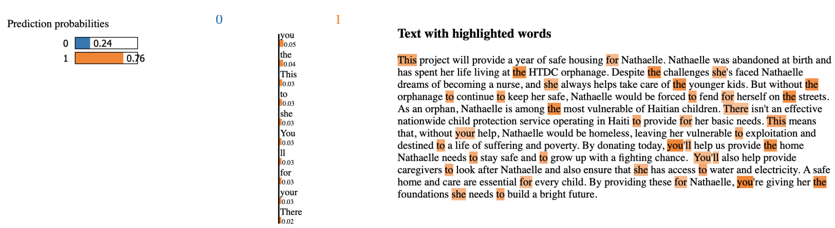
7.1 Pearson Correlation Coefficient vs Logistic Regression

Both the PCC and the LR highlight the importance of pronouns. LIWC researchers Tausczik and Pennebaker propose that, “Function words, such as personal pronouns, reflect attentional allocation,” suggesting that campaigns which met their target fundraising goals displayed community-based thinking. Third-person pronouns also reveal the presence of storytelling. Pennebaker defines this communication style as “narrative thinking,” and finds that “people who score high on [this] factor tend to have better social skills, more friends, and rate themselves as more outgoing.” By highlighting a specific beneficiary with third-person singular pronouns, fundraisers invite a donor’s intimate identification with their cause. Persuasion in charitable contexts circumvents the lack of direct reciprocity by funneling attention towards a person in deep need, maximizing the emotional reward of having helped a cause that feels real.

However, the most important features as determined by the model are not the same as the features which have the highest correlation coefficient. For example, "Analytic" still had a positive feature importance value, which contradicts the Pearson correlation coefficient calculations that found analytic to be negatively correlated with success. One explanation for this discrepancy is that the logistic regression model performance could have been better (see "Conclusion"), so the feature importance values of the LIWC features that it calculates could be unreliable. Another explanation is that the feature importance values of the logistic regression model could be influenced by correlations between features themselves. More work is needed to drop out highly correlated features before inputting them into the logistic regression.

7.2 Logistic Regression vs BERT

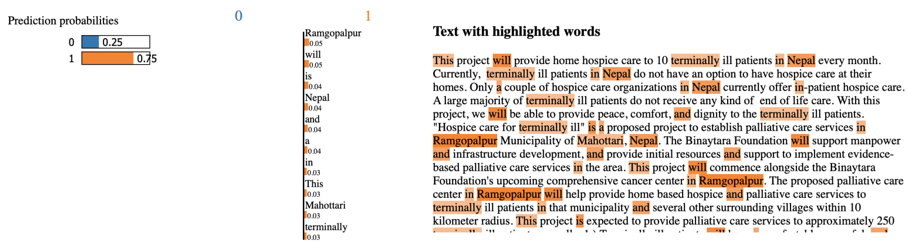
To more closely examine model interpretability and performance, we look at examples where BERT performed better than logistic regression and where BERT performed worse. (These mismatches should better reveal comparative (dis)advantages than particular examples where the models’ predictions agreed.) Specifically, we examine two examples where the BERT prediction was made with >=0.75 confidence, as these examples may be more informative in identifying through LIME which words BERT identifies as important.



The figure above shows a true positive for BERT (where BERT predicted the project would be mostly funded and it actually was funded) and a false negative for logistic regression (where logistic regression predicted the project wouldn't be mostly funded but it was indeed funded).

Surprisingly, while models' predictions diverge, the features they identified as successful converge. We see that LIME identified 2nd-person pronouns, such as "you" and "your," and 3rd-person singular pronouns, such as "she" as most important. This result is corroborated by the findings from our baselines, with PCC finding 2nd and 3rd person pronouns among the top 5 features most positively correlated with campaign success, and LR weighting these features most heavily.

However, examining the subsetted LIWC scores for this particular project, we find that 'Analytic' feature had the higest value (0.91), followed by six-letter words (0.48). We see that this project rated highly on analytic thinking and other features which the logistic regression model weighted negatively, perhaps suggesting why logistic regression predicted this project would be unsuccessful. Moreover, 'you' and 'shehe' had LIWC scores of 0. This result is surprising; given that "you" and "she" appear readily in the text, we should expect this score to be >0. This error likely resulted from the data pre-processing phase, but more exploration is needed to understand exactly what went wrong and how large the ramifications of this error are on our models' predictive abilities.



The figure above shows a false positive for BERT (where BERT predicted the project would be mostly funded but it wasn't) and a true negative (where LR correctly predicted the project wouldn't be mostly funded). LIME identified proper nouns which refer to locations, such as "Ramgopalpur," "Nepal," and "Mahottari," future-focus verbs, such as "will," and other function words. Moreover, this project also rated highly on analytic thinking (0.88) and prepositions (0.57). These results somewhat diverge from those of the PCC and LR. Perhaps BERT failed because it placed too much importance on those locations, and in other projects, mentioning those locations was associated with a campaign's success.

8 Conclusion

In conclusion, storytelling, readability, and community-based language are the most pertinent textual features of a successful crowdfunding campaign. The Pearson correlation coefficients, logistic regression feature weights, and BERT interpretation for the most part converged on that analysis. And while BERT performed better than logistic regression, perhaps it is not optimized for the task of predicting campaign success, given the complexity of reasons why certain projects receive substantial funding and others do not, beyond the text alone.

Future work could combine both the BERT CLS embedding and the LIWC features, as well as including additional manual features which are not solely textual features for improved performance. For example, project funding goal amount, number of previous successful projects from the same organizer, images, NGO pedigree, amount of times shared, and location of the project could be useful features to include. Moreover, because we conducted this analysis on all 27,000 projects, these findings may not necessarily hold for particular thematic (e.g. "Education, Covid-19, Racial Equality") or structural categories (title, need, call to action). Further segmentation of the crowdfunding campaigns may reveal stronger correlations or other latent factors specific to a subset.

We recommend that individuals or non-profit organizations looking to raise money should focus on crafting a clear and cohesive narrative that humanizes a person in need, rather than creating a wordy request dominated by analytic thinking. We recommend to researchers the use of mixed computational methods to overcome the tradeoffs between interpretability and predictive power (and solutions like LIME to assist with inference). Hopefully, future improvements of interpretability methods for deep learning models may be able to give us the best of both worlds.

References

Ajay Agrawal, Christian Catalini, and Avi Goldfarb, "The Geography of Crowdfunding," February 2011, https://doi.org/10.3386/w16820 .

Belleflamme, Paul, Thomas Lambert, and Armin Schwienbacher. "Crowdfunding: Tapping the Right Crowd." Journal of Business Venturing 29, no. 5 (September 2014): 585–609. https://doi.org/10.1016/j.jbusvent.2013.07.003.

C.S. Richard Chan, Charuta Pethe, and Steven Skiena. Natural language processing versus rule-based text analysis: Comparing BERT score and readability indices to predict crowdfunding outcomes. In Journal of Business Venturing Insights, 2021.

Mollick, Ethan. "The Dynamics of Crowdfunding: An Exploratory Study." Journal of Business Venturing 29, no. 1 (January 2014): 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005.

James W. Pennebaker, The Secret Life of Pronouns: What Our Words Say about Us (New York: Bloomsbury Press, 2013).

Salido-Andres, Noelia, Marta Rey-Garcia, Luis Ignacio Alvarez-Gonzalez, and Rodolfo Vazquez-Casielles. "Mapping the Field of Donation-Based Crowdfunding for Charitable Causes: Systematic Review and Conceptual Framework." VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 2020. https://doi.org/10.1007/s11266-020-00213-w .

Zhang, Xupin, Hanjia Lyu, and Jiebo Luo. "What Contributes to a Crowdfunding Campaign's Success? Evidence and Analyses from GoFundMe Data." ArXiv:2001.05446 [Cs], January 15, 2020. http://arxiv.org/abs/2001.05446.