Analysis of Textual Features in Crowdfunding Campaign Success

Alexander J. Wang[1], Madison Pickering, MS[2], Christoph U. Lehmann, MD[2], Richard J. Medford, MD[2]
[1]St. Mark's School of Texas, Dallas, TX; [2]Clinical Informatics Center, University of Texas Southwestern Medical Center, Dallas, TX

Abstract

Crowdfunding is an important source of financial aid for health expenses. We examined whether certain text features were associated with higher funding success. Campaigns analyzed had three main themes: general health, cancer treatment, and long-term surgery. Of features tested, sentiment had the greatest association with campaign success, with positive sentiment increasing and negative sentiment decreasing funding. Links to CaringBridge stories also boosted funding. Overall, text features were moderate predictors of campaign success.

Introduction

In the last decade, crowdfunding has become an increasingly popular source of acquiring financial aid for health expenses[1, 2]. When faced with high medical costs, uninsured or underinsured families can appeal to the public to receive donations for treatment and maintain financial security. We sought to determine which textual features correlate with successful campaigns. Although funding may depend on other factors (e.g., race, gender)[3], campaigners have the most control over the written campaign portion, making optimization of the text crucial.

Methods

We scraped data from the GoFundMe website that included 120,308 medical campaigns created in the United States and its territories between September 2018 and August 2019. We used the *Text* field to derive inputs and the *Amount_Raised* field to measure success.

We created a logistic regression model to identify individuals who a) garnered funding greater than or equal to or b) less than the median of overall funds raised. Our final model included word count, average word length, average sentence length (excluding stop words), sentiment, readability, cluster number, and binary frequencies of 800 n-grams. We calculated sentiment using VADER and readability using Flesch reading-ease score. We clustered the data using a k-means model and TFIDF vectorization. We used eight clusters and manually identified three distinct topic categories for use in the final model.

Our list of n-grams included 300 unigrams, 300 bigrams, and 200 trigrams, selected to maximize frequency and correlation. For each text entry, we included the n-grams on a binary scale: 1 if the n-gram is present, 0 if not. Lastly, we classified campaigns by comparing the amount raised to the median.

Results

The logistic regression model achieved an AUC of 0.70. The features with the highest and lowest five coefficients are shown in Table 1 (with n-gram inputs excluding numbers, punctuation, and stop words). Campaigns with links to CaringBridge.org and positive sentiment were more likely to have funding success.

Table 1. Primary model coefficients.

Feature	Coefficient	Odds	Feature	Coefficient	Odds
caringbridge org visit	0.815603675	2.26	ask help like	-0.570431857	0.57
positive sentiment	0.787785959	2.20	ask help help	-0.57851383	0.56
asking anyone help	0.752966137	2.12	appreciated would	-0.589754187	0.55
cancer need help	0.732449282	2.08	anything appreciated	-0.637800216	0.53
active year old	0.688477255	1.99	*negative sentiment*	-1.45623786	0.23

The k-means model scored -72,427.67 over 92,089 data points. Using PCA, we visualized the data points and cluster centers (Figure 1). The clusters in the k-means model had significant overlap.

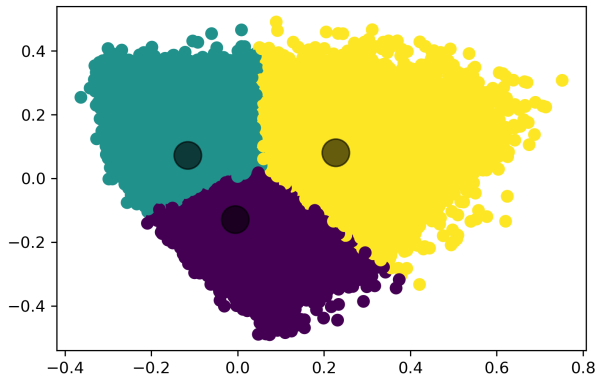

Figure 1. Visualization of k-means model using PCA.

Figure 2 shows the top 15 words in each cluster ranked by average TFIDF score. Data revealed three main themes: general appeals, cancer needs, and long-term surgeries.

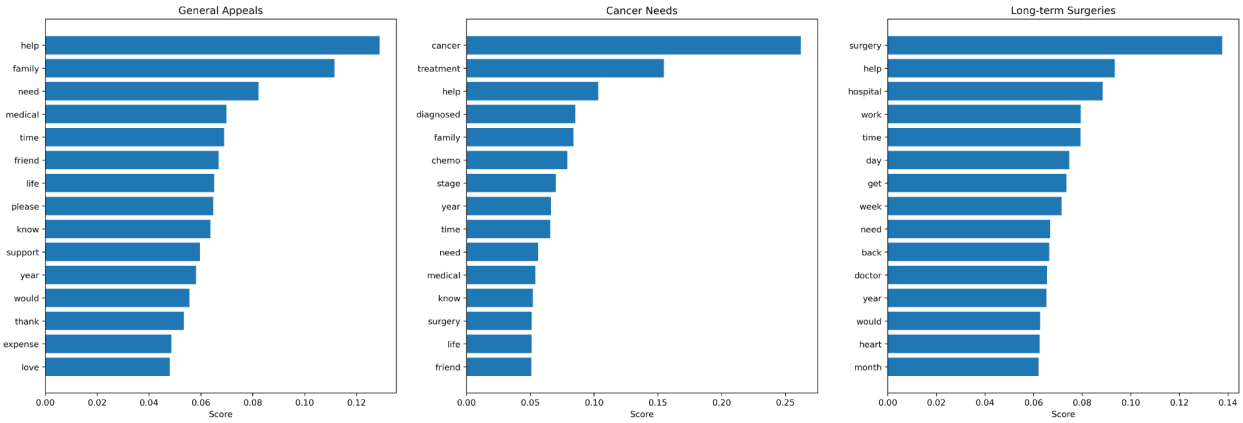

Figure 2. Top 15 words in each cluster, ranked by average TFIDF score.

Discussion

Text features are moderately associated with the success of crowdfunding campaigns. Of the features tested, sentiment played the greatest role in campaign success, suggesting that donors prefer optimistic campaigns and tend to avoid pessimistic ones. Links to caringbridge.org, a website for sharing medical updates, had the greatest positive impact, while desperate-appearing campaigns (e.g., pleas such as “anything [is] appreciated”) resulted in lower funding success.

References

1. Saleh SN, Lehmann CU, Medford RJ. Early crowdfunding response to the COVID-19 pandemic: cross-sectional study. J Med Internet Res. 2021 Feb 9;23(2):e25429. doi: 10.2196/25429. PMID: 33523826; PMCID: PMC7879716.
2. Saleh SN, Ajufo E, Lehmann CU, Medford RJ. A comparison of online medical crowdfunding in Canada, the UK, and the US. JAMA Netw Open. 2020 Oct 1;3(10):e2021684. doi: 10.1001/jamanetworkopen.2020.21684. PMID: 33104206; PMCID: PMC7588935.
3. Kenworthy N, Dong Z, Montgomery A, Fuller E, Berliner L. A cross-sectional study of social inequities in medical crowdfunding campaigns in the United States. PLoS One. 2020 Mar 5;15(3):e0229760. doi: 10.1371/journal.pone.0229760. PMID: 32134984; PMCID: PMC7058302.