Deep Learning Approach to Predicting Success of Medical Crowdfunding Campaigns

Stanford CS224N Custom Project

Advait Patil
Department of Computer Science
Stanford University
advaitp@stanford.edu

Abstract

Crowdfunding campaigns are often used to raise funds for medical and surgical procedures when patients are unable to afford hospital bills. Therefore, they represent a patient population that is at incredible risk and often underinsured. It is important for these patients to know early on whether their crowdfunding campaign is likely to succeed and to provide earlier warnings if there is a low likelihood of their campaign to work. Here, we attempt to solve these problems by developing a deep learning algorithm that can predict whether a medical crowdfunding campaign is likely to succeed early on so timely feedback can be provided to those that need funds for medical procedures. Specifically, we implemented the Hugging Face DistilBERT model with frozen weights and a binary classification head for prediction of whether a campaign would succeed. We compared this with fine-tuning DistilBERT by unfreezing and training all weights. We used tSNE for visualization of embeddings to try to find meaningful clusters. Then, we decided to conduct more granular analyses by using a pre-trained scispaCy model for medical disease classification to categorize surgical subtypes and identify specific areas of unmet need. This work presents a path forward towards better understanding the surgical crowdfunding landscape in a vulnerable patient population.

1 Key Information to include

This project idea was generated through a conversation with surgeons from Stanford's Department of General Surgery.

All data acquisition and subsequent ideas were and will be conducted independently, and there are no external mentors. All work conducted on this project will be for this class only. This work builds on previously-conducted research; all methods described in this paper were written for this course.

2 Introduction

2.1 Medical Crowdfunding

Online crowdfunding platforms have emerged as an increasingly popular funding modality to cover healthcare costs, especially for vulnerable populations, and are directly linked to health disparities and gaps in social safety-net systems. However, the nationwide prevalence and determinants of success of crowdfunding campaigns for medical care remain poorly understood.

Not only would tools to predict success of campaigns assist with increasing funds that are donated, but they would help highlight important gaps in care that could be addressed through non-crowdfunding means.

2.2 Surgical campaigns: an understudied space

Within the medical world, surgical crowdfunding costs remain completely unstudied. However, it has been well documented that costs of surgery represent a significant burden on patients, especially in the U.S. healthcare ecosystem. Therefore, a tool that would allow patients creating crowdfunding campaigns to could predict success early on would help improve success rates. More broadly, NLP approaches to tackle challenges and costs associated with surgical care remain understudied compared to other areas of healthcare.

2.3 Our goal

Our approach to help tackle these problems is to develop a classifier that will determine whether a campaign is successful or not based on solely the campaign story. Here, we define "success" as whether a campaign was able to raise the total number of funds that it had requested. The story is typically a 20-200 word paragraph detailing the circumstances of the patient seeking funds, often describing the specific condition or procedure for which the funds are requested as well as life circumstances or challenges that may affect an ability to pay for said conditions or procedures. Because effects of social networks and social media are difficult to account for or control, such a tool could be used at the point of creation of a campaign to inform whether the campaign could succeed.

Towards this, we used the Python Beautiful Soup webscraper to scrape all all GoFundMe campaigns from 2010-2021. We then filtered these down to 66,514 surgery-related campaigns for subsequent analysis. We then trained a baseline Hugging Face DistilBERT model with frozen weights and a binary classification head, well as a fine-tuned DistilBERT model with all weights trained. Then, we decided to conduct more granular analyses by using a pre-trained scispaCy model for medical disease classification to categorize surgical subtypes and identify specific areas of unmet need. This work presents a path forward towards better understanding the surgical crowdfunding landscape in a vulnerable patient population.

3 Related Work

3.1 Medical Crowdfunding

Several studies have undescored the importance of understanding factors influencing medical crowdfunding campaign outcomes. Most notably, Hou et al. conducted the most recent systematic review on this topic in 2022. [1] They found that text-based analysis was the most promising area for forther study of crowdfunding campaigns.

However, to our knowledge, there have been no efforts using NLP approaches to predict success of medical crowdfunding campaigns in the English language in the published literature. The work of Wang et al. is the main piece in the literature examining the task of prediction of crowdfunding campaign success for medical conditions [3]. While very important, there are several limitations to this work. First, the dataset the model was trained on involved only crowdfunding campaigns from China. Given the large challenge of dataset shift, especially when considering both medical conditions and communication patterns across cultures and countries, this becomes a larger issue for translation.

In addition, because the model was trained on data from only one crowdfunding website in China, the architecture lacks the ability to incorporate additional forms of information about campaigns that may be found on other websites. There are also important differences between mandarin chinese and english that may affect the ability of this model to perform on datasets derived from english-speakers.

Given the current state of the literature, there is significant opportunity and need to make progress on deep learning methods to predict success of medical crowdfunding campaigns.

4 Approach

4.1 DistilBERT

In this project, we trained a DistilBERT-based model with a classification head for prediction of whether a surgical crowdfunding campaign would succeed based on the story description. DistilBERT from Hugging Face has 40% less parameters than bert-base-uncased, runs 60% faster while preserving over 95% of BERT's performances. The inputs to this model are the tokenized versions of the stories (DistilBERT-fast tokenizer) with a maximum token length of 512. The output is a classification of whether the campaign reached its goal amount of funds or not.

4.2 tSNE

We used t-SNE, a popular method for visualization of high-dimensional data, to observe whether there were trends in clustering for the DistilBERT embeddings. We conducted this analyses to compare clustering across different levels of success in reaching the funding goal for each campaign.

4.3 scispaCy

Different surgical procedures can have significantly different costs incurred, meaning that there may be significant heterogeneity within the dataset. In addition, it is important to know which conditions incur disproportionate costs. To this end, we used the en_core_sci_md model from scispaCy, and applied it for Named Entity Recognition (NER) to classify crowdfunding campaigns into surgical subtypes. Given that cancer was disproportionately prevalent within the dataset, we decided to focus our analysis on surgical costs incurred across cancer subtypes.

5 Experiments

5.1 Data

We were unable to find any publicly available datasets online for English, US-based crowdfunding campaigns (or crowdfunding campaigns in general). We used the Python Beautiful Soup package to scrape all crowdfunding campaigns on the world's biggest crowdfunding platform, GoFundMe. This resulted in approximately 1.8 million URLs of scraped crowdfunding campaigns.

We excluded campaigns in non-US locations and those involving non-US currency, as well as checking for duplicates. We filtered these to surgery specific campaigns, curating a final dataset of 66,514 surgical campaigns from 2010-2021. All together, these campaigns represented nearly 1 billion dollars sought, with 354,849,732 dollars raised (35%). Crowdfunding campaigns had an approximately 13 percent success rate in raising the total funds they requested.

The specific input in our modeling task is the story text of each campaign, which gives a description of why the funds are requested and what they will be used for. We will be treating this as a classification problem to determine which campaigns will reach their requested funds based on solely the story description. An example of a campaign story follows:

" *on aug 7th my little sister angela (21) was airlifted from our hometown in pearsall, tx to san antonio after a ct scan revealed a massive tumor on her brain. she was admitted into icu due to hydrocephalus, and for further testing and observation. further testing showed that the tumor (schwannoma) was benign. on august 19th she underwent a 12 hour surgery and the neurosurgeon was able to remove the majority of the tumor. on sept 2nd after 3 weeks and 5 days in the hospital she was able to go home. a little over a month later on oct 5th she underwent a 2nd surgery due to the incision getting infected. she was released to go home on the 15th on medication. due to the remaining tumor being in the neck area her neurosurgeon and ent have been collaborating to determine the next steps best for angela. it has come down between radiation (which has a lot of side effects) and another surgery (risky because of the delicate nerves in the area). because of her being so young; doctors are leaning towards another surgery. angela has been healing from her surgeries but this is just the beginning to* *the long road ahead. medical bills have been coming in and the numbers are overwhelming. any contributions will go towards medical expenses for angela's journey. thank you all for your help and god bless!*"

5.2 Evaluation method

We decided to split our data into into train (80%), development (10%), and test (10%) sets. We are using AUROC and F1 scores to evaluate our model performance. We will then use tSNE as a dimensionality reduction technique to assess whether the DistilBERT embeddings cluster toegther based on funding success rate (by labeling by quartile for percent of funding goal reached). We then will use SHAP (SHapley Additive exPlanations) to assess relevant model features. Specifically, we plan to build on the work of Kokalj et al on extending SHAP explanations to transformer-based classifiers towards better model interpretability. Finally, we will examine a random sample of 100 successful and 100 non-successful campaign descriptions to identify any qualitative trends.

5.3 Experimental details

For our model, we chose to extend DistilBERT from Hugging Face, a faster and lighter version of the BERT base (www.huggingface.co/docs/transformers/model_doc/distilbert). As per the Hugging Face documentation, it has 40% less parameters than bert-base-uncased, runs 60% faster while preserving over 95% of BERT's performances as measured on the GLUE language understanding benchmark. We implemented this model with and without fine-tuning the weights. We used a classification head to go from DistilBERT embeddings to binary classification of campaign success. Specifically, we decided to go with a single dense output layer and then a sigmoid activation function with binary cross-entropy. Our adam optimizer rate was set to 3e^-5. Training on the full dataset took approximately 3 hours. The best model as determined by minimum loss on the development set was saved.

5.4 Results

After curating our dataset of surgical crowdfunding campaigns, we first sought to examine trends in the data and in funds requested. We were able to obtain the geographic distribution of campaigns by US state. These data are displayed in Figure 1.

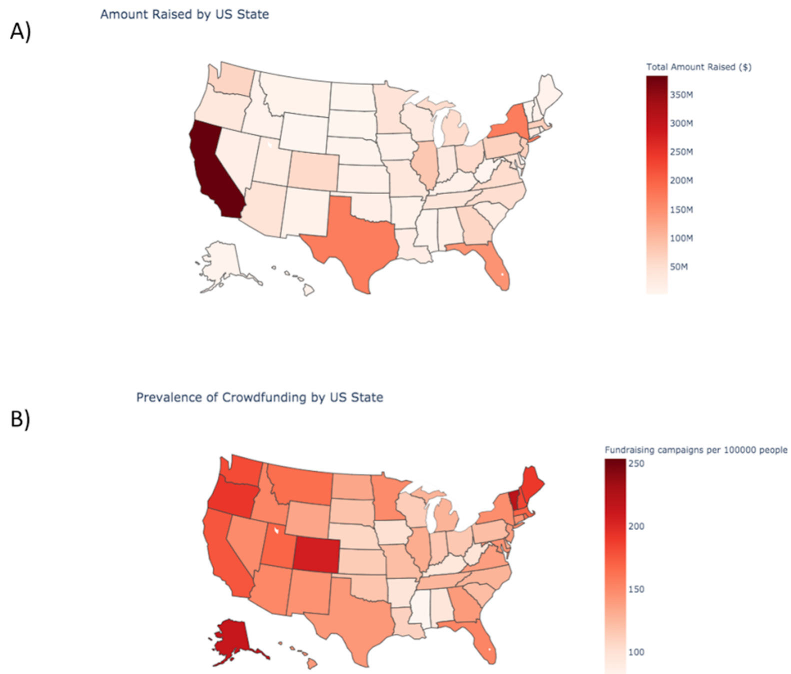

Figure 1: *National distribution of a) funds requested by crowdfunding campaigns by US state, and b) total number of crowdfunding campaigns by capita per US state.*

We find that California, New York, Texas, and Florida have crowdfunding campaigns requesting the greatest amount of funds. These are largely due to state size reflecting a large number of campaigns due to large populations. However, further investigation does reveal that even accounting for this effect, campaign funds requested are disproportionately high in California. This may be due to either a tendency for more expensive procedures not to be covered (e.g. insurance billing specifications) or a greater general awareness of online crowdfunding platforms. In addition, Alaska, Vermont, Colorado, and Maine had the most campaigns per capita (100,000 people). These represent states where patients may be underinsured and comprise a further area of study.

Next, we implemented the Hugging Face DistilBERT model with frozen weights and binary classification head as our baseline model. We compared this to a fine-tuned model of DistilBERT by unfreezing and training all of the weights. We trained these models for 20 epochs with the best model being retrieved and evaluated on the test set. AUROC and F1 scores are displayed in the Table below.

Model	AUROC (test)	F1 (test)
DistilBERT (frozen)	0.813	0.693
DistilBERT (finetune)	0.782	0.514

Overall, the accuracy of these models were higher than I expected. This is because predicting success of crowdfunding campaigns is a task that depends on many external variables, beyond just the campaign description (e.g. social network effects). Further work may allow us to improve further upon these efforts to achieve a higher-performing classifier.

For our next analyses, we examined the DistilBERT embeddings using t-SNE. Clear clusters across the embeddings did not emerge (e.g. with respect to campaign types, or funds requested). Figure 2 shows a t-SNE plot of the embeddings colored by funding quartile (as a percent of requested funds obtained).

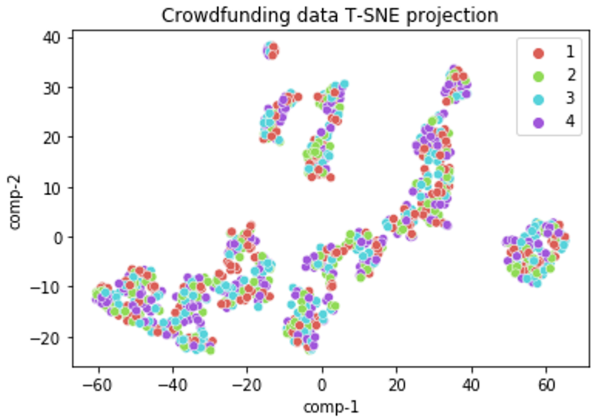

Figure 2: *t-SNE plot of DistilBERT embeddings colored by quartile of percent of funds obtained (quartile 1 = lowest success rate, quartile 4 = highest success rate).*

Finally, we used scispaCy for Named Entity Recognition to categorize crowdfunding campaigns by surgical subtype. Because it became apparent that most of the surgical funds were requested for cancer-related conditions, we decided to focus on these. The top 10 most prevalent conditions are shown in Figure 3.

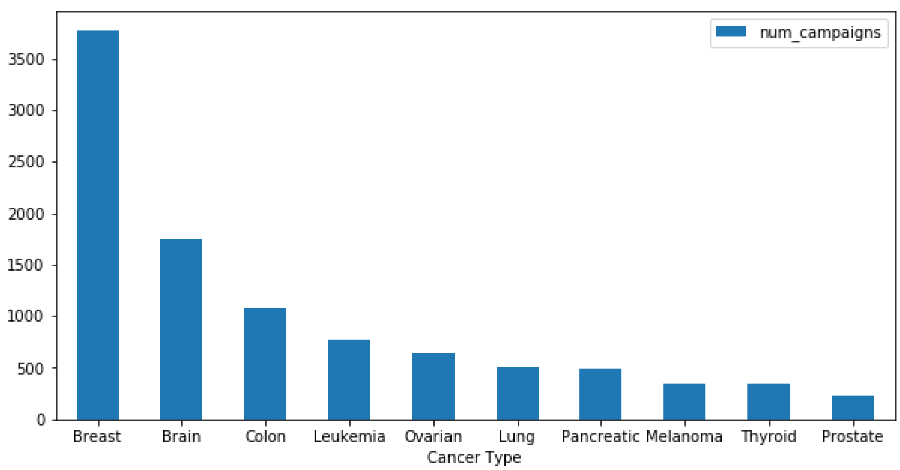

Figure 3: *Top 10 most common oncological conditions in the cancer crowdfunding data.*

We then decided to compare the prevalence of the most common cancer campaigns to the nationwide prevalence of the conditions to determine oncological types with disproportionate burden, as crowdfunding indicates difficulty with paying through alternative means (Figure 4).

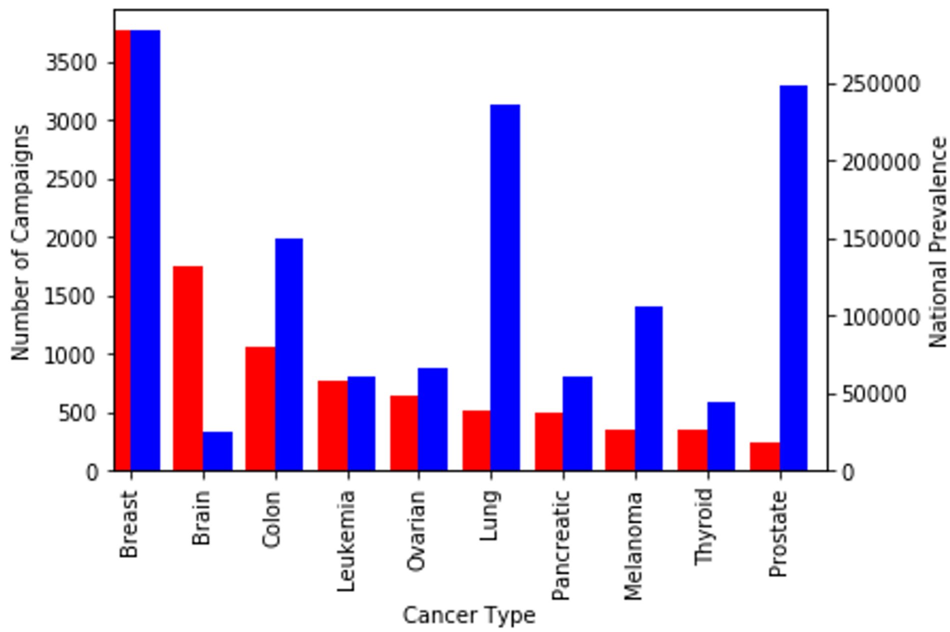

Figure 4: *Comparison of the 10 most common cancer subtypes for surgical crowdfunding campaigns to nation-wide prevalence.*

We see that brain and colon cancer appear to have a disproportionate burden compared to their nationwide prevalence. In contrast, prostate and lung cancer have a lower prevalence in crowdfunding campaigns compared to the nationwide prevalence, meaning these conditions incur a lower burden to patients seeking funds.

6 Analysis

We conducted qualitative analysis to better understand the behavior of our model.

After manually inspecting a random sample of 100 successful campaigns and 100 non-successful campaigns, we found that there were a number of successful campaigns meant to cover surgical costs for pets, often involving the word "vet". However, the word "vet" is also used in campaigns for military veterans, which comprise an often neglected patient population in terms of ability to pay for healthcare.

Therefore, approaches to filter out pet-related campaigns from our dataset while retaining those for medical conditions for military veterans should be investigated. We also found that campaign stories that were very long were misclassified as fully-funded. While there appears to generally be a correlation between story length of funding success up to a certain point, this behavior should be investigated further in depth.

7 Conclusion

In conclusion, we achieved several key goals through this work. First, we were able to develop a deep learning approach to predict crowdfunding campaign success from solely text-based story descriptions, and achieved reasonable accuracy given the difficulty of the task. Next, using scispaCy, a pre-trained NER model trained on a large biomedical/clinical corpus, we categorized campaigns into surgical oncology subtypes. From this analysis, we identified that brain tumor surgery involves a greater burden for funds requested in reference to nationwide prevalence compared to surgical procedures for other conditions.

There are several key limitations to this work. First, due to the inherent limitations of our webscraping data collection approach, we were unable to collect longitudinal data regarding online engagement with each campaign (e.g. number of facebook likes, retweets, shares, etc). Future approaches should consider 1) how to obtain this data at scale (not currently feasible with a webscraper), and 2) how to design a model that can account for these different forms of data.

In summary, we can define two broad classes of factors that can influence success of crowdfunding campaigns. The first are *intrinsic factors*, such as the story of a campaign for why funds are needed, which were the focus of this analysis. The other are *extrinsic factors* like the forms of longitudinal data described above. Future work should focus on incorporating such factors over the duration of a campaign to increase the possible predictive utility of a deep learning model in real-time.

8 References

[1] Hou X, Wu T, Chen Z, Zhou L. Success Factors of Medical Crowdfunding Campaigns: Systematic Review. J Med Internet Res. 2022 Mar 22;24(3):e30189. doi: 10.2196/30189. PMID: 35315779; PMCID: PMC8984822.

[2] Birkmeyer JD, Gust C, Dimick JB, Birkmeyer NJ, Skinner JS. Hospital quality and the cost of inpatient surgery in the United States. Ann Surg. 2012;255(1):1-5. doi:10.1097/SLA.0b013e3182402c17

[3] Wang, T., jin, F., Hu, Y., &; Chen, Y. (2019). Early Predictions for Medical Crowdfunding: A Deep Learning Approach Using Diverse Inputs. ArXiv Machine Learning. https://doi.org/10.48550/arXiv.1911.05702

[4] Sanh V. et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv Computation and Language. https://doi.org/10.48550/arXiv.1910.01108

[5] Devlin J., Ming-Wei C., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv Computation and Language. https://doi.org/10.48550/arXiv.1810.04805