Analysing and Predicting Success of Crowdfunding Campaigns

Glenn Jin Wee Chia
glenn_chia@alumni.sutd.edu.sg
Singapore University of Technology and Design
Singapore

Kwan Hui Lim
kwanhui_lim@sutd.edu.sg
Singapore University of Technology and Design
Singapore

Abstract

Crowdfunding have gained popularity as a platform for individuals to harness the power of social solidarity to raise public funds for a range of objectives, such as for community projects, social good, medical expenses, etc. While there are many crowdfunding campaigns with altruistic goals, not all campaigns go viral and many more do not even meet their fundraising targets. This paper analyses the various factors that contribute to successful campaigns and develops five models for predicting fundraising success based on various combination of campaign features. We experiment on a dataset of 18,473 crowdfunding campaigns and discuss our main findings in terms of how different factors affect campaign success.

CCS Concepts

• **Information systems** → **Data mining**; • **Human-centered computing** → *Collaborative and social computing*; • **Computing methodologies** → *Machine learning*.

Keywords

Crowdfunding, Data Mining, Machine Learning, Prediction

ACM Reference Format:
Glenn Jin Wee Chia and Kwan Hui Lim. 2024. Analysing and Predicting Success of Crowdfunding Campaigns. In *The 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL '24), December 16–20, 2024, Hong Kong, China.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3677389.3702580

1 Introduction

Crowdfunding platforms, such as Kickstarter [5], GoFundMe [9] and Indiegogo [16], have emerged as a popular avenue for users to raise funds from the online public by harnessing the power of social solidarity. Organizers of crowdfunding campaigns can state their objectives, which ranges widely from supporting medical expenses and charitable causes to business ventures and sporting competitions. Crowdfunding has the potential for an extensive outreach like how a campaign to raise funds for an elderly lady who was attacked in San Francisco went viral and received significant coverage by major US news outlets. This campaign was highly successful and "raised over $897,000 of its $50,000 goal", with the recipient committed to donate these funds "back to the Asian American community to combat racism" [4]. However, only a minority of campaigns go viral and many do not even achieve their desired funding targets.

JCDL '24, December 16–20, 2024, Hong Kong, China

ACM ISBN 979-8-4007-1093-3/24/12
https://doi.org/10.1145/3677389.3702580

Figure 1: Wordcloud of Description of Campaigns

Every fundraising campaign ultimately aims to acquire donations from individuals to contribute towards their desired target amount. There are a multitude of elements that a potential donor interacts with on the crowdfunding platform that influences his/her decision to donate to a particular cause. This overall objective has spurred research into studying and understanding the underlying reasons behind successful crowdfunding campaigns, including from the perspectives of business theories [1, 11], social and psychology [2, 3], machine learning [7, 13] and others.

Contributions. In this paper, we contribute to the literature by analysing crowdfunding campaigns to investigate which factors of a campaign contributes to its success. This paper makes the following contributions: (i) We collected a dataset of 18,473 crowdfunding campaigns and extracted various important campaign features to use in this work; and (ii) We developed numerous machine learning models based on various combinations of campaign features, for predicting the fundraising success of a campaign.

2 Models and Baselines

There are multiple features of a campaign that could be utilised for the fundraising success prediction task. The following list shows the set of features and their representations as input to the various models we experiment with.

Campaign Features (Extracted): There are existing campaign features, such as Title (t), Subtitle (st), Description (d), Amount Targeted (at), Beneficiary Availability (b), Category (c), Duration (dur) that we directly extract from the crowdfunding campaign page. Textual features (t, st, d) first undergo white-space removal, lower-case conversion, stopwords removal, lemmatisation, before being represented as word-level TF-IDF. Numerical features (at, dur) are represented as integer values, while categorical features (c) are represented using one-hot encoding.

Campaign Features (Emotions): Using the text description of the campaign, we generate additional features based on the emotions and sentiments portrayed by the campaign. Similar to other works [12, 14], we utilize the NRC lexicon of emotions to compute

Table 1: Summary results of five models based on various combination of features, including Title (t), Subtitle (st), Description (d), Amount Targeted (at), Beneficiary Availability (b), Category (c), Emotion (e), Duration (dur)

Model	Metric	d	d+at	d+at+b	d+at+b+c	d+at+b+c+e	d+at+b+c+e+dur	d+at+b+c+e+t	d+at+b+c+e+t+st
Naive Bayes	Accuracy	0.79	0.8	0.81	0.81	0.81	0.65	0.80	0.79
Naive Bayes	Precision	0.51	0.68	0.69	0.69	0.69	0.59	0.66	0.66
Naive Bayes	Recall	0.50	0.59	0.60	0.61	0.61	0.63	0.61	0.61
Naive Bayes	F1	0.46	0.60	0.61	0.62	0.62	0.58	0.62	0.63
Logistic Regression	Accuracy	0.66	0.66	0.67	0.67	0.68	0.72	0.69	0.69
Logistic Regression	Precision	0.50	0.60	0.61	0.62	0.62	0.62	0.61	0.60
Logistic Regression	Recall	0.50	0.65	0.66	0.67	0.67	0.65	0.66	0.64
Logistic Regression	F1	0.50	0.59	0.60	0.61	0.61	0.63	0.61	0.60
SVM	Accuracy	0.61	0.70	0.70	0.70	0.70	0.70	0.70	0.70
SVM	Precision	0.51	0.60	0.60	0.60	0.60	0.60	0.60	0.60
SVM	Recall	0.51	0.63	0.63	0.63	0.63	0.63	0.63	0.63
SVM	F1	0.50	0.60	0.60	0.60	0.60	0.60	0.60	0.60
Random Forest	Accuracy	0.46	0.74	0.78	0.75	0.78	0.78	0.75	0.75
Random Forest	Precision	0.50	0.57	0.60	0.59	0.61	0.61	0.57	0.60
Random Forest	Recall	0.51	0.55	0.56	0.58	0.56	0.56	0.56	0.58
Random Forest	F1	0.43	0.56	0.56	0.58	0.57	0.56	0.56	0.59
XGBoost	Accuracy	0.74	0.78	0.75	0.79	0.77	0.79	0.78	0.77
XGBoost	Precision	0.51	0.65	0.62	0.66	0.64	0.67	0.65	0.64
XGBoost	Recall	0.51	0.63	0.63	0.64	0.64	0.64	0.63	0.65
XGBoost	F1	0.50	0.64	0.62	0.65	0.64	0.65	0.63	0.64

this Emotion (e) feature of 10 floating point values, with eight representing the emotions of joy, anticipation, sadness, surprise, fear, trust, anger and disgust, and two representing the sentiments of positive and negative.

Based on these eight features (t, st, d, at, b, c, e, dur), we came up with eight different combinations of feature sets. These eight feature sets are then provided as input into various machine learning models, which includes Naive Bayes [17], Logistic Regression, Support Vector Machines (SVM) [8], Random Forest [10] and XGBoost [6]. For training these models, we also require the ground truth label of whether a campaign was successful funded. We generated these labels based on the features of amount targeted and amount raised, i.e., a campaign is successful if the amount raised is at least the amount targeted, and unsuccessful otherwise.

3 Experiments and Results

Dataset Description. We collected a total of 18,473 crowdfunding campaigns from a major crowdfunding platform. Each collected campaign belongs to one of 19 categories, e.g., charity, community, education. faith, environment, etc, and also includes features such as its title, subtitle, description, amount raised/targeted, beneficiary, organizer, created date, etc. Figure 1 shows a wordcloud of the text descriptions of the campaigns in our dataset.

Data Pre-processing. We further pre-processed the dataset by filtering out campaigns with funding amounts set to be excessively low or high. For this purpose, we set a lower bound of $100 and upper bound of $100,000, which respectively removed 0.44% and 3.74% of campaigns from our dataset. We also filtered out campaigns that were not complete, e.g., with a very short or no text description, which removed 0.22% of the dataset.

Training and Evaluation. The processed dataset is then split into 75% for training our models and the remaining 25% for evaluation. We utilized the metrics of Accuracy, Precision, Recall and F1 score for this task of campaign funding success prediction. Table 1 shows the summary results of these experiments.

Results and Discussion. Comparing the models, XGBoost has the overall best performance in terms of F1 score (up to 0.65), while Naive Bayes performs well in terms of Accuracy (up to 0.81). Examining the campaign features, we observe that the feature set d+at+b+c, comprising Description (d), Amount Targeted (at), Beneficiary Availability (b) and Category (c), offers good performance when used as inputs to the earlier two models. Across the models and features, we also note that adding Amount Targeted (at) to the Description (d) results in a good performance improvement but subsequent addition of more features only result in smaller improvements.

4 Conclusion and Future Work

We studied crowdfunding campaigns in terms of the factors that influence the success of fundraising. We analyzed 18,473 crowdfunding campaigns and developed five models using numerous campaign features for predicting fundraising success. Our main finding is that the campaign description plays a major role, with the amount targeted, beneficiary availability and category providing some minor improvements in terms of fundraising success. Given the role of campaign description, future work can explore the use of word or sentence embedding [15] to better predict the campaign success and explore the use of Large Language Models to improve campaign descriptions.

Acknowledgments

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (Award No. MOE-T2EP20123-0015). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the Ministry of Education, Singapore.

References

[1] Martin Angerer, Alexander Brem, Sascha Kraus, and Andreas Peter. 2017. Start-up funding via equity crowdfunding in Germany: A qualitative analysis of success factors. *The Journal of Entrepreneurial Finance (JEF)* 19, 1 (2017), 1–34.

[2] Aaron H Anglin, Jeremy C Short, Will Drover, Regan M Stevenson, Aaron F McKenny, and Thomas H Allison. 2018. The power of positivity? The influence of positive psychological capital language on crowdfunding performance. *Journal of Business Venturing* 33, 4 (2018), 470–492.

[3] Enrico Battisti, Elvira Anna Graziano, and Michael Christofi. 2022. Equity crowdfunding platforms and social media: a Twitter analysis. *International Journal of Entrepreneurial Behavior & Research* 28, 5 (2022), 1206–1221.

[4] Maina Chen. 2021. Chinese Grandma Who Fought Off Attacker to Donate Over $900K from GoFundMe to AAPI Community. *Yahoo News* (2021). https://news.yahoo.com/chinese-grandma-fought-off-attacker-193134140.html

[5] Perry Chen, Yancey Strickler, and Charles Adler. 2024. Kickstarter. https://www.kickstarter.com. Accessed: 2024-11-07.

[6] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[7] Chaoran Cheng, Fei Tan, Xiurui Hou, and Zhi Wei. 2019. Success Prediction on Crowdfunding with Multimodal Deep Learning.. In *IJCAI*. 2158–2164.

[8] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20 (1995), 273–297.

[9] Brad Damphousse and Andrew Ballester. 2024. GoFundMe. https://www.gofundme.com/. Accessed: 2024-11-07.

[10] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1. IEEE, 278–282.

[11] Jascha-Alexander Koch and Michael Siering. 2019. The recipe of successful crowdfunding campaigns: an analysis of crowdfunding success factors and their interrelations. *Electronic Markets* 29, 4 (2019), 661–679.

[12] Jolin Shaynn-Ly Kwan and Kwan Hui Lim. 2020. Understanding public sentiments, opinions and topics about covid-19 using twitter. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 623–626.

[13] Yan Li, Vineeth Rakesh, and Chandan K Reddy. 2016. Project success prediction in crowdfunding environments. In *Proceedings of the ninth ACM international conference on web search and data mining*. 247–256.

[14] Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. 321–327.

[15] Wenchuan Mu, Junhua Liu, and Kwan Hui Lim. 2024. Fast Bibliography Pre-Selection via Two-Vector Semantic Representations. In *Proceedings of the 2024 ACM/IEEE Joint Conference on Digital Libraries (JCDL'24)*.

[16] Danae Ringelmann, Slava Rubin, and Eric Schell. 2024. Indiegogo. https://www.indiegogo.com/. Accessed: 2024-11-07.

[17] Irina Rish. 2001. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. 3. 41–46.