

# Success Factors of Donation-Based Crowdfunding Campaigns: A Machine Learning Approach

Massara Alazazi  
University of Texas  
Rio Grande Valley  
massara.alazazi01@utrgv.edu

Bin Wang  
University of Texas  
Rio Grande Valley  
bin.wang@utrgv.edu

Tareq Allan  
Dakota State University  
tzallan@pluto.dsu.edu

## Abstract

*Crowdfunding has emerged as an alternative mechanism to traditional financing mechanisms in which individuals solicit financial capital or donation from the crowd. The success factors of crowdfunding are not well-understood, particularly for donation-based crowdfunding platforms. This study identifies key drivers of donation-based crowdfunding campaign success using a machine learning approach. Based on an analysis of crowdfunding campaigns from Gofundme.com, we show that our models were able to predict the average daily amount received at a high level of accuracy using variables available at the beginning of the campaign and the number of days it had been posted. In addition, Facebook and Twitter shares and the number of likes, improved the accuracy of the models. Among the six machine learning algorithms we used, support vector machine (SVM) performs the best in predicting campaign success.*

## 1. Introduction

Many individuals turn to the social media to solicit financial help from the general public, which is called the “crowd”, rather than the traditional financial fund seeking including business angels or venture capital funds [1]. This activity is called crowdfunding, which is a form of crowdsourcing that is facilitated by the Web 2.0 technologies. It not only opens the doors for people from all over the world to reach others and communicate, but also enables them to support others financially [2]. Crowdfunding is an open call funding mechanism that depends on small portions of funds from a relatively large number of people through online platforms for the purpose of financing a venture or project investment, without standard financial intermediaries [3]. Crowdfunding enables fundraisers to get financial support for their future ventures at a considerably low cost and risk

when compared with traditional forms of funding [1]. While crowdfunding is a relatively new phenomena, it has gained popularity, since lack of financial resources and limited access to them are identified as key problems for the operation of small and medium-sized enterprises [4].

Crowdfunding platforms support millions of crowdfunding campaigns from various categories and for different purposes. Although some campaigns would have similar goals or projects, their success rate may vary. Successful campaigns attract the crowd and persuade people about their goals and motives for the campaign. Furthermore, the ability to propagate those campaigns to other people through social media increases the social media buzz for the fundraising and thus increases its ability to succeed [5].

Despite the growing popularity of crowdfunding, there is a need to better understand this relatively new social phenomena [6]. Although many people turn to crowdfunding to support their projects financially, not all projects get the funds they aim for. For instance, only 44% of all projects on Kickstarter reached their goal [7]. Hence, there is a need to understand this funding variation. There is also a lack of understanding of the dynamics of successful crowdfunding [3]. Crowdfunding platforms provide categories of campaigns according to their purpose and description. These categories include emergency, medical, art, sports, nonprofit, and others. Variations of crowdfunding activities between different categories are not well studied.

In addition, most existing studies focus on reward-based crowdfunding platforms such as Kickstarter [3, 8]. In this study, we employ data from a donation-based crowdfunding platform, since there are few studies that examine this type of crowdfunding business model and funding activities. Our study applies a machine learning approach, which is rarely in crowdfunding research. Thus, our research purpose is to:

1. Identify key drivers of donation-based crowdfunding campaign success.
2. Compare the performance of different machine learning algorithms and the regression model approach in predicting donation-based crowdfunding campaign success.

In the remainder of the paper, we first discuss the related literature on crowdfunding as a social financial phenomenon. Next, we discuss the machine learning approach and compare different algorithms, followed by the empirical study including the data, analysis, and results. Finally, the paper discusses the results and future research directions.

## 2. Literature Review

### 2.1 Crowdsourcing and Crowdfunding Phenomena

The crowdsourcing phenomena flourished with the diffusion of information and telecommunications technologies, particularly the social media [9]. Crowdsourcing has been applied in many areas including crowdsourcing to obtain product specifications or improvement, crowdsourcing for answering academic problems, crowdsourcing for driving accident reporting, and crowdsourcing for innovative business ideas [10]. Accordingly, many applications and online platforms are founded including Amazon's Mechanical Turk and Dell IdeaStorm [10]. Incentives (e.g., rewards, feedback, and rivalry) in crowdsourcing have been studied by some researchers [11-14].

They found that rewards, positive feedback, and rivalry motivate individuals to provide their input and participate. The wisdom of the crowd, in which a large number of solvers contribute to a successful solution, promotes the application of crowdsourcing by many organizations and individuals [15].

Crowdfunding is a form of crowdsourcing, and it has been defined by Mollick [3] as *"the efforts by entrepreneurial individuals and groups – cultural, social, and for-profit – to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the Internet, without standard financial intermediaries."* Crowdfunding is divided into four types depending on the return to the funders: reward-based, equity-based, loan-based, and donation-based [16].

Crowdfunding is used as an alternative source of small and medium-sized enterprise financing through the Internet to leverage large audience contributions [1]. In addition to the financial benefits,

crowdfunding allows more efficient investment decisions. It facilitates interactions between project creators and funders, who could be future consumers, eliminating geographic barriers [17].

Previous research studied the motivation factors to fund crowdfunding campaigns include connecting with others, learning, collecting funds for different projects, and distributing awareness regarding different issues [6]. The study applies a grounded theory method through semi-structured interviews. Another study explored the dynamics of crowdfunding [3]. It collected data from Kickstarter and found that Personal networks, project quality, and geography are associated with the success of crowdfunding efforts. Furthermore, a previous study focused on the drivers of crowdfunding success from IndieGoGo funding campaigns [18]. The study identified eight campaign success drivers including image, cause of need, picture appeal, perspective advocated, social comparisons, decisional control, labeling and request sizes. In addition, funding goal and the number of comments affect campaign success. Other research discussed more specific factors such as the emotional delivery impact on pro-social crowdfunding success. The study uses image classification software to analyze facial expressions in photos attached to the campaign. The study concluded that fundraisers use emotions to solicit money and attract funders, since there are usually no financial incentives for funders to support these campaigns. Emotions and visual expressions are key factors in affecting the success of the campaigns [19]. In addition, a previous study explored how physical distance would impact the fundraising effort through analyzing data extracted from Sellaband platform campaigns on artist-entrepreneurs with related geographic information on backers from Google maps. The study found that funding is not geographically constrained. However, geographic distance played a role in financing musical projects. This impact is apparent for investors who have personal connections with the artist-entrepreneur [8].

Other research efforts focused on crowdfunding from different perspectives. For instance, a previous study investigates crowdfunding from the economic perspective. Particularly, it looks at transaction costs, reputation, and market design and their impacts on the rise of non-equity crowdfunding. It collected data from Kickstarter to identify crowdfunding platform rules to maximize transaction volume [16]. In addition, focusing on the medical crowdfunding campaigns for organ transplantation behavior, a study by Durand et al. [20] found, after applying bivariate and multivariate analyses, that more positive

sentiment, lengthier campaign description, higher goal amount, and third person description positively affect the amount of fund raised by the campaign. As can be seen from these studies, the results are inconsistent and the factors affecting the success of crowdfunding campaigns are not very clear.

## 2.2 Crowdfunding and Social Media Outlets

Due to the social nature of crowdfunding, social media could play a prominent role in campaign fundraising through amplifying the social influence. Many studies focused on the determinants of crowdfunding success, but little attention is paid to the different impacts of campaign shares through social media and likes received on different categories of crowdfunding campaigns.

Social media interactions are a prominent part of many individuals' daily activities. Shares and exchanges of crowdfunding information through different social media outlets allow for faster expansion of crowdfunding campaign to a larger number of people. Crowdfunding is based on small donation from a large number of parties and thus social media could help in contributing more funds to the campaign [7].

Social media channels are used in crowdfunding to enable fundraisers and backers to share campaign information with peers and request their support [21]. The fundraiser's online social network is found to influence the success of the funding because it enhances the social capital of the fundraiser and motivates more contribution to the campaign [5]. Using social media platforms, fundraisers and the funders spread campaigns information among their friends and followers and increase the awareness about it.

Crowdfunding platforms realize the important role of social media. As a result, most of them create an easy way to share campaign links or information through providing buttons or shortcuts to share those campaigns in some popular social network sites such as Facebook and Twitter. This sharing might not only increase the possibility for fundraising success but also increase the popularity of the campaign and play a marketing role of its outcome product [7].

## 3. Machine Learning

Machine learning is an inductive process that employs algorithms and statistical models to predict patterns in big data [22]. Supervised machine learning builds a mathematical model based on a training data set that has both the input variables and

the output variable in order to learn the mapping function from the input to the output variable. Because machine learning does not require "rigid" statistical assumptions, it allows more effective prediction models [22].

Machine learning approaches have been used in the information systems and social media disciplines. Previous research has used a machine learning algorithm to filter web pages and retrieve efficient and more precise search results by filtering out irrelevant documents from all sets of retrieved documents. The machine learning approach uses both content and structure analysis [23]. In addition, machine learning is used to predict individuals' behavior from online communities' content data mining. For instance, a recent research uses machine learning techniques to classify user-generated text from a smoking cessation community. The study identifies the smoking status of users with a high performance [24]. Furthermore, Twitter data is employed in a study for user classification using a rich feature set for the purpose of detecting political affiliation, identifying ethnicity and detecting affinity for a specific business [25].

Despite the increasing use of machine learning on other IS research topics, most previous research did not apply the machine learning approach to predict the performance of crowdfunding campaigns and used traditional statistical methods. The only exception we know is a study that applied the random forest method to predict the success of crowdfunding campaign at the time of campaign launch with an average accuracy of 65% [26]. However, no research has examined how social media shares can predict crowdfunding campaign success using the machine learning approach. In this study we utilize some of the techniques of supervised machine learning including linear regression, classification and regression tree (CART), support vector machine (SVM), k-nearest neighbors (KNN), neural network and random forest. The purpose of this process is to predict the amount of funds raised per day, which is a continuous variable, based on some of the variables related to the campaign.

Now we briefly define each of the applied machine learning techniques. First, the linear regression model is used when the output variable is continuous and it attempts to fit data with the best line (hyper-plane) that passes through the points [27]. It models a linear relationship between the predicted dependent variable and the (input) independent variables. The second approach is the (CART) [28]. It is a multistage classifier and an inductive rule-based learner in the form of a tree in which successive decision functions are used to classify an

unknown sample. Basically, a decision tree has “a root node, a number of interior nodes, and a number of terminal nodes” [29]. Each node “corresponds to a test  $X$  over a single attribute of the input data and has a number of branches, each of which handles an outcome of the test  $X$ . Each leaf node represents a class that is the result of decision for a case” [30]. The third techniques is SVM, which is a universal learner that is based on the structural risk minimization principle of the computational learning theory [31]. It is usually used for solving pattern recognition problems through mapping data to a higher dimensional input space and then constructing an optimal separating hyperplane [32]. It finds the hypothesis that minimizes the true error bounds [31]. The fourth algorithm is KNN. The input of the classifier is the  $k$  closest training examples in the feature space. It stores the complete training data. “New examples are classified by choosing the majority class among the  $k$  closest examples in the training data” [33]. The fourth algorithm is the random forest. It is an ensemble classifier that creates multiple decision trees, using a random subset of training dataset and variables [34]. The last technique is the neural network and was first developed to model the human brain. Neural network uses term weights to recognize patterns in data. It consists of three layers: input, hidden, and output layer. The hidden layer learns the mapping between the input and the output [35].

## 4. Empirical Study

### 4.1 Sample

A customized computer script was developed to automatically scrape data on every active project on GoFundMe.com. Data on 9,948 crowdfunding campaigns were collected in April 2019. GoFundMe is a donation-based crowdfunding platform that requires no financial fee for launching crowdfunding

projects. It allows for interactions between creators and backers through comments and replies. The study collected data on campaigns from different categories. Table 1 shows the categories and the number of campaigns in each category. The collected data include campaign title, text of story, date created, funds raised, goal amount of funds, number of funders, number of Facebook and Twitter shares, number of likes received, number of updates to the story, number of photos uploaded, and the fundraiser team. After campaigns with missing data were removed, we obtained data on 9,935 valid campaigns. Table 2 shows the descriptive statistics of the sample. The positive emotion and negative emotion were calculated based on the sentiment in the campaign text description using LIWC2015.

**Table 1. Campaign distribution over categories**

Category	Campaign Count
Animal	960
Business	973
Community	970
Competition	976
Creative	970
Education	978
Emergency	970
Faith	231
Medical	972
Memorial	974
Nonprofit	974

**Table 2. Campaign descriptive statistics (N=9,935).**

Variable	Min	Median	Mean	Max	Std. Dev.
\$ Raised/(# Days+1)	1.64	109.20	398.52	184168.44	2,292.32
\$ Goal	100	10,000	155,200	1,000,000,000	10,225,125
# Days	0	90	80.17	180	48.33
# Facebook/Twitter Shares	0	205.5	793.4	999000	10317.4
# Likes	1.0	65	216.8	351000	3554.71
Positive Emotion	0.00	4.45	4.74	100	2.49

Negative Emotion	0.00	1.00	1.29	27.27	1.32
------------------	------	------	------	-------	------

## 4.2 Variables and empirical models

Table 3 summarizes the definitions of the variables. Because GoFundMe does not require fundraisers to specify the time frame for their campaigns, we estimated the amount raised per day to take into consideration the number of days a campaign had been posted. We used ten dummy variables to represent eleven categories of fundraising campaigns with animals being the base category.

We tested two models that predict the dependent variable  $\ln \text{AmtRaisedPerDay}$ . Model 1 includes variables known at the beginning of the campaign ( $\ln \text{Goal}$ ,  $\text{NumPhotos}$ , dummy variables for campaign category, and sentiment including  $\text{PosEmotion}$  and  $\text{NegEmotion}$ ) plus  $\ln \text{NumDays}$  that takes into consideration the number of days the campaign had been on GoFundMe. Model 2 includes all variables

in Model 1 and three additional variables including  $\text{NumUpdatesPerDay}$ ,  $\ln \text{NumSocialMediaSharesPerDay}$  and  $\ln \text{NumLikesPerDay}$ .

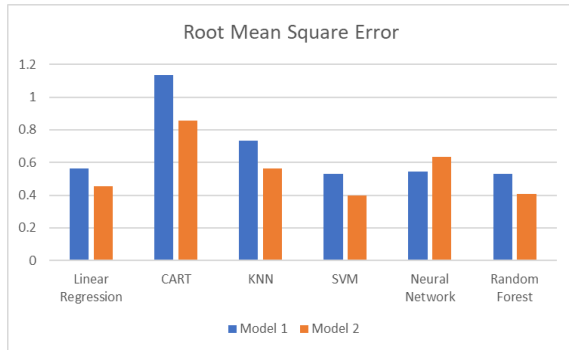
We randomly split the sample into a training dataset of 7,951 observations and a validation dataset of 1,984 observations. We used six different supervised learning algorithms including linear regression, CART, KNN, SVM, neural networks, and random forest on the training dataset and applied 10-fold cross validation. We then applied the fitted models to the validation dataset and calculated model performance metrics including the root mean square error (RMSE), R-squared, and adjusted R-squared. Table 4 summarizes the validation data testing results. Figures 1 to 3 compare the performances of Models 1 and 2 using six machine learning algorithms in predicting  $\ln \text{AmtRaisedPerDay}$  along three metrics – RMSE, R-squared, and adjusted R-squared.

**Table 3. Variables and definitions.**

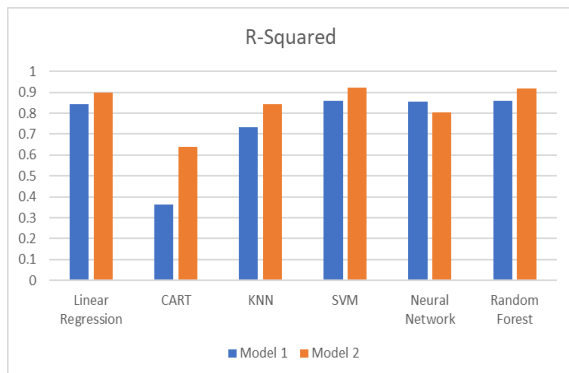
Variable	Definition
$\ln \text{AmtRaisedPerDay}$	The natural logarithm of the amount raised divided by one plus the number of days the campaign had been posted on GoFundMe.
$\ln \text{Goal}$	The natural logarithm of the campaign's fundraising goal.
$\ln \text{NumDays}$	The natural logarithm of one plus the number of days the campaign had been posted on GoFundMe.
$\text{NumPhotos}$	The number of photos the campaign posted.
<i>Business</i>	1 if a fundraising campaign is in the business category, 0 otherwise.
<i>Community</i>	1 if a fundraising campaign is in the community category, 0 otherwise.
<i>Competition</i>	1 if a fundraising campaign is in the competition category, 0 otherwise.
<i>Creative</i>	1 if a fundraising campaign is in the creative category, 0 otherwise.
<i>Education</i>	1 if a fundraising campaign is in the education category, 0 otherwise.
<i>Emergency</i>	1 if a fundraising campaign is in the emergency category, 0 otherwise.
<i>Faith</i>	1 if a fundraising campaign is in the faith category, 0 otherwise.
<i>Medical</i>	1 if a fundraising campaign is in the medical category, 0 otherwise.
<i>Memorial</i>	1 if a fundraising campaign is in the memorial category, 0 otherwise.
<i>Nonprofit</i>	1 if a fundraising campaign is in the nonprofit category, 0 otherwise.
$\text{PosEmotion}$	The positive emotion score of the campaign description as reported by LIWC2015.
$\text{NegEmotion}$	The negative emotion score of the campaign description as reported by LIWC2015.
$\text{NumUpdatesPerDay}$	The number of times updates had been posted to the campaign divided by one plus the number of days the campaign had been on GoFundMe.
$\ln \text{NumSocialMediaSharesPerDay}$	The natural logarithm of one plus the number of times a campaign has been shared through Facebook and Twitter divided by one plus the number of days the campaign had been on GoFundMe.
$\ln \text{NumLikesPerDay}$	The natural logarithm of one plus the number of likes the campaign received divided by one plus the number of days the campaign had been on GoFundMe.

**Table 4. Validation dataset model performance metrics.**

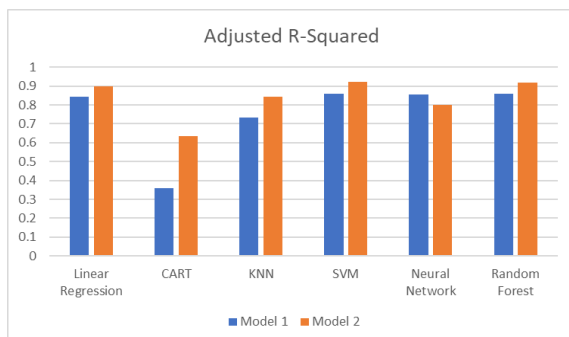
	Model 1			Model 2		
	RMSE	R-Squared	Adjusted R-Squared	RMSE	R-Squared	Adjusted R-Squared
<b>Linear Regression</b>	0.563	0.844	0.843	0.452	0.899	0.898
<b>CART</b>	1.137	0.363	0.358	0.858	0.637	0.634
<b>KNN</b>	0.734	0.735	0.733	0.562	0.845	0.843
<b>SVM</b>	0.531	0.861	0.860	0.398	0.922	0.921
<b>Neural Network</b>	0.544	0.855	0.853	0.634	0.802	0.800
<b>Random Forest</b>	0.532	0.861	0.860	0.409	0.918	0.917



**Figure 1. RMSE of prediction models using six machine learning algorithms.**



**Figure 2. R-Squared of prediction models using six machine learning algorithms.**



**Figure 3. Adjusted R-Squared of prediction models using six machine learning algorithms.**

As can be seen from the results, Model 1 with variables known at the beginning of each campaign and the number of days a campaign had been on GoFundMe had a RMSE ranging from .531 to 1.137, R-squared from .363 to .861, and adjusted R-squared from .358 to .860. The optimal hyperparameters as reported by the algorithms are as follows: sigma = 0.0477 and C = 1 for SVM, k = 5 for KNN, cp = 0.0846 for CART, size = 5 and decay = 0.1 for neural network, and mtry = 8 for random forest.

Among the six algorithms, SVM produced the best result across the three model performance metrics. These results suggest that Model 1 using the SVM algorithm can predict the amount raised per day with a high level of accuracy based on predictors known at the beginning of each campaign and the number of days the campaign had been on GoFundMe.

In Model 2, RMSE ranged from .398 to .858, R-squared ranged from .637 to .922, and the adjusted R-squared ranged from .634 to .921. The best values for the hyperparameters for Model 2 differ from those for Model 1. For SVM, the final values used for the model were sigma = 0.03936541 and C = 1. For KNN, the final value used for the model was k = 7. For CART, the final value used for the model was cp = 0.08414919. For the neural network, the final values used for the model were size = 5 and decay = 0.1. For random forest, 10-fold cross validation was used and the final value used for the model was mtry = 10. SVM again produced the best model performance across three metrics. Results from Figures 1 to 3 show that the model performance improved when NumUpdatesPerDay, lnNumSocialMediaSharesPerDay and lnNumLikesPerDay were added in Model 2 except for the neural network algorithm. Specifically, the RMSE decreased and both the R-squared and adjusted R-squared increased for five of the six algorithms. Even after taking into consideration the

numbers of variables in the two models, Model 2 still produced a higher adjusted R-squared compared with Model 1. Hence, the average daily numbers of updates, Facebook and Twitter shares, and likes contributed to the prediction of the amount raised per day.

Table 5 summarizes the linear regression results for both Models 1 and 2 based on the training dataset. Overall the two models produced consistent coefficient estimates. Specifically, the coefficient for

fundraising goal was positive and significant across both models. The coefficient for  $\ln$ NumDays was negative and significant in both models. Among the dummy variables for campaign categories, Community, Education, Emergency, Medical, Memorial, and Nonprofit had positive and significant coefficient estimates across both models, while Business and Competition had negative coefficient estimates across both models. The two sentiment variables were not significant in either model.

**Table 5. Linear regression results based on training dataset (N=7,951).**

<b>Variable</b>	<b>Model 1</b>	<b>Model 2</b>
Intercept	6.354*** (0.061)	5.047*** (0.054)
<i>lnGoal</i>	0.223*** (0.005)	0.175*** (0.004)
<i>lnNumDays</i>	-0.981*** (0.008)	-0.569*** (0.009)
<i>NumPhotos</i>	0.000** (0.000)	0.000 (0.000)
<i>Business</i>	-0.849*** (0.028)	-0.414*** (0.024)
<i>Community</i>	0.180*** (0.028)	0.251*** (0.023)
<i>Competition</i>	-1.151*** (0.029)	-0.645*** (0.025)
<i>Creative</i>	-0.208*** (0.028)	-0.003 (0.023)
<i>Education</i>	0.336*** (0.028)	0.438*** (0.023)
<i>Emergency</i>	1.579*** (0.029)	1.064*** (0.025)
<i>Faith</i>	-0.381*** (0.046)	0.077* (0.037)
<i>Medical</i>	1.807*** (0.030)	1.245*** (0.026)
<i>Memorial</i>	1.615*** (0.029)	1.087*** (0.025)
<i>Nonprofit</i>	0.483*** (0.029)	0.558*** (0.023)
<i>PosEmotion</i>	0.003 (0.002)	0.003 (0.002)
<i>NegEmotion</i>	-0.002 (0.005)	0.000 (0.004)
<i>NumUpdatesPerDay</i>		-0.037 (0.056)
<i>lnNumSocialMediaSharesPerDay</i>		-0.021*** (0.003)
<i>lnNumLikesPerDay</i>		0.426*** (0.007)
R-Squared	0.857	0.907
Adjusted R-Squared	0.856	0.907

Notes: \*\*\*,  $p < 0.001$ ; \*\*,  $p < 0.01$ ; and \*,  $p < 0.05$ .

The number of photos in the campaign description had a significant but close to zero coefficient estimate in Model 1 and was nonsignificant in Model 2. The coefficient estimate for Creative campaigns was negative and significant in Model 1 but was nonsignificant in Model 2. The dummy variable for Faith campaigns had a negative and significant coefficient estimate in Model 1 but a positive and significant coefficient estimate in Model 2. Among the three additional variables in Model 2,  $\ln \text{NumSocialMediaSharesPerDay}$  had a negative and significant coefficient estimate,  $\ln \text{NumLikesPerDay}$  had a positive and significant coefficient estimate, and  $\text{NumUpdatesPerDay}$  was nonsignificant.

## 5. Discussion

### 5.1 Theoretical Contribution

We compare the performance of two models in predicting crowdfunding campaign success using six machine learning algorithms. Our research has the following contribution to theory.

First, our research is one of the first to predict donation-based crowdfunding campaign success using the machine learning approach. Using only variables available at the beginning of each campaign and the number of days it has been posted, our model using the SVM algorithm was able to predict the average daily amount received of the validation dataset with a high level of accuracy. In addition, adding average daily social media shares, likes and updates improved the model performance across all six algorithms used and the best performing model was again the SVM. Overall, the SVM performed better than the linear regression. Hence, our results highlight the importance of employing machine learning algorithms instead of traditional regression models in predicting crowdfunding campaign success.

Second, our results reveal the importance of fundraising goal and number of days on GoFundMe in affecting the average daily amount received. Specifically, a higher fundraising goal is associated with a higher average daily amount raised. Hence, prospective donors are more willing to donate a higher amount when the goal is set higher. In contrast, a longer listing period is associated with a lower average daily donation received. Previous research has revealed a substitution effect in crowdfunding where lenders are less willing to fund

campaigns that have received a large amount of contribution [36]. This leads to a higher donor enthusiasm at the beginning of each fundraising campaign and lower contribution as the donor enthusiasm wears off. As a result, the average daily amount received decreases as a campaign stays on GoFundMe for a longer period of time.

Third, our results show that campaigns in categories such as community, education, emergency, medical, memorial and nonprofit categories receive a higher average daily donation, while those in the business and competition categories receive a lower average daily contribution. These results reveal that donors on GoFundMe respond more favorably to life changing events and charitable causes than to business ventures. On donation-based crowdfunding platforms such as GoFundMe, prospective donors mainly contribute for altruistic reasons to help others in need.

Fourth, a larger number of average daily social media share is associated with a smaller amount of average daily campaign donation, while the opposite is true for the average daily number of likes. These results suggest that it is not the number of social media shares but rather the number of positive support that leads to more funds raised. A higher social media share count is not helpful as the content can be either positive or negative. The unexpected social media, i.e. Facebook and Twitter, shares results could be due to the crowded content on these platforms, resulting in the shares having a minimal impact on the donation behavior. What really matters is how many people in the crowdfunding platform are interested in a particular campaign.

Fifth, the sentiment in the campaign description does not affect crowdfunding contribution. Combined with the previous result, this shows that it is how the contribution will be used but not how the sentiment is framed in the campaign narrative that affects donor decision.

### 5.2. Practical Implications

Our research has the following practical implications. First, crowdfunding platforms interested in predicting campaign performance should consider using machine learning algorithms to improve model performance. Using only variables known at the beginning of each campaign and the number of days it has been available, our model using SVM was able to predict average daily campaign donation with a higher level of accuracy than linear regression.



Second, the decreasing average daily contribution received as time goes by highlights the importance of promoting late-stage campaigns, especially on crowdfunding platforms where the fundraisers get nothing if the goal is not met. As the initial donor enthusiasm wears off, it is important to promote these fundraising campaigns by placing them in prominent places on the website to increase prospective donors' exposure and contribution so as to achieve the fundraising goal.

Third, donation-based crowdfunding platforms can educate fundraisers regarding the importance of appealing to prospective donors' desire to help others in need or facing tragic events to promote fundraising success. For campaigns in less popular categories such as business ventures and competitions, fundraisers can emphasize how their business ventures may help charitable causes and individuals or families going through life changing events.

Forth, the results of the study can guide the crowdfunding platform providers and campaign initiators to determine a feasible number of days that is the most appropriate to get the required financial contributions to the campaign.

## 6. Conclusion, Limitations, and Future Research Directions

In this research, we compare the performance of six different machine learning algorithms in predicting crowdfunding campaign success on GoFundMe and identify the drivers of the average daily amount received. Our results show that SVM performs the best and is able to predict the average daily amount received with a high degree of accuracy using variables available at the beginning of each campaign and the number of days the campaign has been posted. In addition, adding social media shares, likes, and the number of updates further improves the accuracy of the prediction.

Our research has the following limitations and direction for future research. First, we collected cross-sectional data at a single point in time. Longitudinal data may reveal richer temporal characteristics and show some patterns, which could be used as a prediction baseline. We plan on extending this study by collecting longitudinal data from GoFundMe on a weekly basis for several months for all active campaigns to analyze campaign success at different time points and discover how a campaign's activities progress with time. In addition, a longitudinal study would allow us to predict the future performance of a campaign using different machine learning algorithms and compare their predicting performance. Second, we only examined

crowdfunding success on GoFundMe. Future research could replicate our analysis on other crowdfunding platforms to examine the generalizability of our findings. Other machine learning algorithms such as the recurrent neural network algorithm could also be applied.

## 7. References

- [1] Schwienbacher, A. and B. Larralde, *Crowdfunding of small entrepreneurial ventures*. Handbook of entrepreneurial finance, Oxford University Press, Forthcoming, 2010.
- [2] Selwyn, N. *Web 2.0 applications as alternative environments for informal learning-a critical review*. in *Paper for CERI-KERIS international expert meeting on ICT and educational performance*. 2007.
- [3] Mollick, E., *The dynamics of crowdfunding: An exploratory study*. Journal of business venturing, 2014. **29**(1): p. 1-16.
- [4] Rakicević, Z., J. Omerbegović-Bijelović, and G. Lazić-Rasović, *SMEs Support Planning Improvement Based on its Suitable Structure*. Management (1820-0222), 2013(68).
- [5] Hong, Y., Y. Hu, and G. Burtch, *Embeddedness, pro-sociality, and social influence: Evidence from online crowdfunding*. MIS Quarterly, Forthcoming, 2018.
- [6] Gerber, E.M. and J. Hui, *Crowdfunding: Motivations and deterrents for participation*. ACM Transactions on Computer-Human Interaction (TOCHI), 2013. **20**(6): p. 34.
- [7] Moissejev, A., *Effect of social media on crowdfunding project results*. 2013.
- [8] Agrawal, A.K., C. Catalini, and A. Goldfarb, *The geography of crowdfunding*. 2011, National bureau of economic research.
- [9] Deng, X., K. Joshi, and R.D. Galliers, *The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design*. Mis Quarterly, 2016. **40**(2): p. 279-302.
- [10] Mo, J., S. Sarker, and S. Menon, *Know when to run: recommendations in crowdsourcing contests*. 2018.
- [11] Yang, D., et al., *Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones*. IEEE/ACM Transactions on Networking (TON), 2016. **24**(3): p. 1732-1744.
- [12] Singla, A. and A. Krause. *Truthful incentives in crowdsourcing tasks using regret minimization mechanisms*. in *Proceedings of the 22nd international conference on World Wide Web*. 2013. ACM.

- [13] Füller, J., K. Hutter, and M. Fries, *Crowdsourcing for goodness sake: impact of incentive preference on contribution behavior for social innovation*, in *Interdisciplinary Approaches to Product Design, Innovation, & Branding in International Marketing*. 2012, Emerald Group Publishing Limited. p. 137-159.
- [14] Jian, L., et al., *Managing the crowds: the effect of prize guarantees and in-process feedback on participation in crowdsourcing contests*. University of Connecticut School of Business Research Paper, 2018(18-11).
- [15] Brabham, D.C., *Crowdsourcing as a model for problem solving: An introduction and cases*. *Convergence*, 2008. **14**(1): p. 75-90.
- [16] Agrawal, A., C. Catalini, and A. Goldfarb, *Some simple economics of crowdfunding*. *Innovation Policy and the Economy*, 2014. **14**(1): p. 63-97.
- [17] Golić, Z., *Advantages of crowdfunding as an alternative source of financing of small and medium-sized enterprises*. *Zbornik radova Ekonomskog fakulteta u Istočnom Sarajevu*, 2014(8): p. 39-48.
- [18] Evers, M., C. Lourenço, and P. Beije, *Main drivers of crowdfunding success: a conceptual framework and empirical analysis*. Erasmus Universiteit, 2012.
- [19] Rhue, L. and L.P. Robert. *Emotional Delivery in Pro-social Crowdfunding Success*. in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018. ACM.
- [20] Durand, W.M., et al., *Medical crowdfunding for organ transplantation*. *Clinical transplantation*, 2018. **32**(6): p. e13267.
- [21] Hui, J.S., M.D. Greenberg, and E.M. Gerber. *Understanding the role of community in crowdfunding work*. in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 2014. ACM.
- [22] Sebastiani, F., *Machine learning in automated text categorization*. *ACM computing surveys (CSUR)*, 2002. **34**(1): p. 1-47.
- [23] Chau, M. and H. Chen, *A machine learning approach to web page filtering using content and structure analysis*. *Decision Support Systems*, 2008. **44**(2): p. 482-494.
- [24] Wang, X., et al., *Mining user-generated content in an online smoking cessation community to identify smoking status: A machine learning approach*. *Decision Support Systems*, 2019. **116**: p. 26-34.
- [25] Pennacchiotti, M. and A.-M. Popescu. *A machine learning approach to twitter user classification*. in *Fifth International AAAI Conference on Weblogs and Social Media*. 2011.
- [26] Greenberg, M.D., et al. *Crowdfunding support tools: predicting success & failure*. in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2013. ACM.
- [27] Ayodele, T.O., *Types of machine learning algorithms*, in *New advances in machine learning*. 2010, IntechOpen.
- [28] Lawrence, R.L. and A. Wright, *Rule-based classification systems using classification and regression tree (CART) analysis*. *Photogrammetric engineering and remote sensing*, 2001. **67**(10): p. 1137-1142.
- [29] Swain, P.H. and H. Hauska, *The decision tree classifier: Design and potential*. *IEEE Transactions on Geoscience Electronics*, 1977. **15**(3): p. 142-147.
- [30] Stein, G., et al. *Decision tree classifier for network intrusion detection with GA-based feature selection*. in *Proceedings of the 43rd annual Southeast regional conference-Volume 2*. 2005. ACM.
- [31] Joachims, T. *Text categorization with support vector machines: Learning with many relevant features*. in *European conference on machine learning*. 1998. Springer.
- [32] Suykens, J.A. and J. Vandewalle, *Least squares support vector machine classifiers*. *Neural processing letters*, 1999. **9**(3): p. 293-300.
- [33] Horton, P. and K. Nakai. *Better Prediction of Protein Cellular Localization Sites with the  $it\ k$  Nearest Neighbors Classifier*. in *Ismb*. 1997.
- [34] Belgiu, M. and L. Drăguț, *Random forest in remote sensing: A review of applications and future directions*. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016. **114**: p. 24-31.
- [35] Warner, B. and M. Misra, *Understanding neural networks as statistical tools*. *The american statistician*, 1996. **50**(4): p. 284-293.
- [36] Burtch, G., A. Ghose, and S. Wattal, *An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets*. *Information Systems Research*, 2013. **24**(3): p. 499-519.