Understanding the Impact of In-Process Promotional Messages: Project Updates in Crowdfunding

Name: Eunjeong Lee
Student Number: 450105
Course Name: Master Thesis
Supervisor: Dr. Florian Deutzmann
Submission Date: 29-07-2017

Abstract

This research makes the first step towards understanding in-process promotions, specifically the message strategies used by creators in a crowdfunding setting. Drawing on a publicly available dataset from Kickstarter, consisting of 7,428 crowdfunding projects and an amount of 34,920 corresponding update posts, the models examine how update posts can be used as promotional messages in a crowdfunding setting and how they are positively associated with the total funding and the success of a campaign. The author shows that these effects differ by the type of message strategy used, namely: whether the project creator uses informative messages, which focus on campaign progress and rewards, or persuasive messages, which try to motivate sponsors to contribute to the campaign by emphasizing the campaign's goal and asking for help. To be specific, both informative and persuasive messages are significantly associated with both the total funding and the success of a campaign. Here, persuasive messages have a stronger association with the campaign performance than informative messages.

In addition, this paper finds that the effects of different types of messages also differ depending on the project category. Whilst there is no difference in the effects of both types of messages on total funding in different categories, the informative messages have a stronger association with the success of the campaigns in the Design and Technology category than those in the Art and Culture category. This paper also finds that distinguishing updates between informative and persuasive helps increase the model's explanatory power.

Preface

This master thesis is written in fulfilment of the requirements for the degree of Master of Science in Economics and Business at the Erasmus School of Economics, Erasmus University. I carried out research and wrote this work from November 2016 to July 2017.

I would like to take this opportunity to thank the two people that enabled me to see this project through until completion. First and foremost I would like to thank my thesis advisor Dr. Florian Deutzmann for his valuable feedback, especially during my stay in South Korea. Your detailed feedback and quick responses have helped immeasurably to improve the quality and the readability of my thesis. A special word of thanks also goes to Roderik Kelder. Without your support, I would not have been able to finish this piece. I appreciate the selfless effort you put into giving me feedback.

- Eunjeong Lee

Rotterdam, 2017

TABLE OF CONTENTS

I. Introduction

In the past years, increasingly sophisticated information technology has enabled new types of business transactions to be conducted over the internet. Examples of these are instances of co-creation (Kohler et al., 2011) such as crowdsourcing (Howe 2008, Leimeister et al. 2009), instances of open innovation (Lakhani and Von Hippel, 2003) such as the design of open source software programs and instances of crowdfunding activities (Ryu, 2016) ranging from fundraising to commerce. This paper focuses on the latter of these new types of business transactions: crowdfunding. Its increased and successful use over global online platforms has as of late attracted much public attention.

As new ventures face difficulties in attracting capital from angel investors, banks and venture capital firms, these opt for crowdfunding as an alternative, directly appealing to the public over online platforms such as Kickstarter or IndieGoGo to obtain funding for the production and launch of their innovative ideas. As the economic potential of online crowdfunding platforms has become apparent to the general public, these platforms have flourished. Each year millions of individuals contribute to the campaigns hosted on these platforms. In 2015 alone these platforms facilitated the exchange of U.S. $34 billion (Massolution, 2015).

This growth resulted in significant attention to media, U.S. legislators (JOB Act 2011) and the academia. A significant body of research (Belleflamme et al., 2014; Mollick, 2014; Agrawal et al., 2015) has investigated the factors that make crowdfunding projects a success and the factors that influence sponsors' motivation to fund a project. However, this body of scholarship reveals its shortcomings in its lopsided focus on the effects of the control factors in the *pre-launch phase* of the campaign on the campaign's result, such as the campaign's description, reward count and reward limit. As a result, we know only little about how creators interact with a project's backers[1] and market their projects *during* the campaign.

This study tries to fill this gap by investigating the in-process marketing during a campaign. It investigates the use of in-process promotional messages that campaign creators send to their project's backers in the form of project updates. In order to do this, this study uses the

[1] In this study, the terms "sponsors" and "backers" are interchangeable, unless otherwise specified.

framework proposed by Ducarroz et al. (2016), which identifies different strategies behind in-process promotional messages and analyzes their respective impacts on the final bidding price in an auction setting. This thesis applies this framework to a crowdfunding setting in order to analyze the impact of different types of in-process promotional messages on a campaign's performance. This thesis distinguishes between informative messages, which provide information on product attributes, and persuasive messages, which appeal to backers' emotions to elicit aid in achieving the funding goal. This thesis also looks at the varying effects of the two types of messages on two different categories of projects, namely 'Design and Technology' and 'Art and Culture'. To infer these messages' impact, this thesis carries out two tasks. The first task is to find out how these messages affect projects' total funding levels, up to and beyond their original funding goals. The second task is to find out how these messages affect crowdfunding projects' chances of success.

The empirical setting for this study's analysis is one of the oldest and largest crowdfunding platforms on the internet: Kickstarter. The platform is dedicated to bringing creative projects to life. Since its founding in 2009, the platform has raised over $3 billion dollars and thus enabled the successful funding for over 100,000 projects in 15 categories such as art, dance, fashion, technology and so on. This thesis examines a publicly available data set of 7,428 funded and unfunded Kickstarter projects to test its hypotheses. Specifically, it uses a regression analysis of 34,920 in-process promotional messages, or 'project updates', to analyze the differing impacts of two types of messages on projects' final funding amounts and the funding success.

The author observes that the funding levels and the success of the campaigns are correlated to both types of messages. More precisely, this paper finds that whereas both informative and persuasive messages have a positive impact on a campaign's performance, persuasive messages are generally found to be more effective than informative messages for campaigns in both 'Art and Culture' and 'Design and Technology' categories. While informative messages do not differ in their effects on the total funding level for either of the two different categories, they do differ in their effects on the success of a campaign. When looking at the relative effectiveness of the two types of messages per category, informative messages are found to show a stronger correlation to the success of 'Design and Technology' projects than to that of 'Art and Culture' projects. The effect of persuasive messages, however, does not differ by category. Thus, in order to effectively raise funds for a Kickstarter project, it is generally important to stimulate backers with more persuasive messages than informative messages.

This study contributes to a better understanding of the impact of the interaction between backers and creators on projects' total funding amounts as well as funding success. Firstly, this research provides insight into the impact of in-process promotional messages' strategies on backers' behavior and the overall campaign performance for those who aim to achieve funding success and for those who aim to maximize total funding levels. Secondly, this research identifies the optimal message strategy per project category by investigating the impact of different message strategies on different categories of projects. In doing so, this analysis enables project creators to make qualified decisions when deciding whether to implement a specific message strategy in any given project.

The remainder of this paper is organized as follows. In the next section, this paper presents the background theories and empirical findings on the factors underpinning successful crowdfunding campaigns. In section 3, this paper develops a conceptual framework based on the research from Ducarroz et al (2016). In section 4, this paper presents its research methodology, including information on the text mining process and the regression analysis. In section 5, this paper discusses the empirical context of this study, examines its primary data and defines its variables. In section 6, this paper conducts its empirical analysis and presents its findings. The last section finishes off with a discussion that includes managerial implications and limitations of this current study.

II. Literature Review

2.1 Discussion on Crowdfunding

The concept of crowdfunding started from a broader concept called crowdsourcing, which refers to "using the crowd to obtain ideas, feedback and solutions for corporate activities" (Bellflamme, Lambert, and Schwienbacher, 2014: p. 586). As these radically new ways of acquiring funding have gained in popularity, the academia has tried to capture their characteristics in several definitions of crowdfunding business models. Specifically, Bellflamme, Lambert, and Schwienbacher (2014: p. 588) defined crowdfunding as "an open call, essentially through the Internet, for the provision of financial resources either in form of a donation or in exchange for some form of reward and voting rights in order to support initiatives for specific purposes." Alternatively, De Buysere, Gajda, Kleverlaan, Marom, and Klaes (2012: p. 9) defined crowdfunding as "a collective effort of many individuals who networked and pooled their resources to support efforts initiated by other people or organizations." However, noticing that these definitions are either too broad and elusive, Mollick (2014: p. 2) came up with a new definition of crowdfunding. Specifically, he refers to crowdfunding 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 financing intermediaries."

Crowdfunding typically involves three players (Ordanini et al. 2011). A 'creator' offers a new project and attempts to get financed by sponsors, or 'project backers' (backers). These backers decide whether to support the campaign considering the compensation they receive for their contribution, which can include both tangible rewards such as rewards or cash, or intangible rewards such as the satisfaction of altruism. The last player is the 'intermediary platform' which connects creators and backers through its operator's screen and showcases creators' stories. Crowdfunding platforms have four common characteristics: platforms provide a standardized format through which the creators can pitch their ideas; a payment system that allows backing; funding related information such as the total goal amount reached, the duration of the campaign, the rewards offered and, lastly, tools that facilitate communication between backers and creators (Agrawal et al. 2011).

Crowdfunding is different from other traditional ventures in the sense that the relationship between creators and backers varies depending on the context and the nature of the funding effort. For example, backers could be donators to, investors in or even consumers of prototype products. Based on the context and the nature of the funding effort, scholars classify crowd-funding efforts differently. For example, Hemer (2011) posits that crowdfunding is categorized based on what backers receive for their contribution, the legal complexity and the degree of information asymmetry between creators and backers. Mollick (2014) identifies different categories in accordance with backers' goals: donation-based (e.g. Ammando), loan-based (e.g. Kiva), equity-based (e.g. Crowdcube) and reward-based (e.g. Kickstarter) crowdfunding platforms. This last model of crowdfunding, reward-based crowdfunding, is the most common form of crowdfunding and is considered one of its largest recent innovations (Fleming 2016). It is on reward-based crowdfunding that this paper will subsequently focus.

Reward-based crowdfunding has become common in all creative industries ranging from the funding of the production of music, games, inventions, and art (Agrawal et al. 2015; Burtch et al., 2013; Zheng et al., 2014). In reward-based crowdfunding, backers receive something non-financial, either material or immaterial, as a reward in return for their investment. An example of a material reward is a pre-purchase of a certain video game, while an example of an immaterial reward could come in the form of stretch goals[2] adding additional content to the game. This model of funding follows either of two basic principles: the "all-or-nothing" approach or the "keep-it-all" approach. In the all-or-nothing approach, projects are only funded if the funding goal is reached. Otherwise, any financial investments made are returned to the backers. Kickstarter is an example of a crowd-funding platform which uses the "all-or-nothing" approach. In the keep-it-all approach, which is less risky for creators, any financial investments made are collected by a project's creator regardless of the funding result of the campaign (Cumming at al., 2015; Haas et al., 2014; Bellflamme et al., 2014). IndieGoGo is an example of a crowdfunding platform which allows creators to choose between either of the "all-or-nothing" or "keep-it-all" systems. As Kickstarter only provides project creators with the "all-

[2] Stretch goals are not part of the official Kickstarter platform. Instead, it is a term set by the Kickstarter community as a way to set an additional target beyond the initial and official goal of the Kickstarter project and raise more money (Kickstarter, 2017).

or-nothing" approach, this thesis will focus its analysis on Kickstarter projects in order to guarantee consistency in its data.

2.2 Success Factors in Crowdfunding

Since its founding in 2009, Kickstarter has raised more than $2 billion for over 100,000 projects. Perhaps unsurprisingly, the top 10 most crowdfunded projects in Kickstarter from 2009 to July 2017 are all from the Design and Technology categories (refer to category details in Appendix A), which involve rewards in the form of materialistic goods. The most funded 'Design and Technology' project in the platform's history is the Pebble Time smartwatch which raised over $20,338,986 or about 4,067% of their original goal of $500,000 (Zipkin, 2015). However, Kickstarter has also been effective in raising funds for the 'Art and Culture' categories. When taking all art forms, such as art, theater, fine-art photography, film & video, music and dance together (refer to category details in Appendix A), this platform raises more money (323.6 million) for the arts than the total amount of funding provided through the U.S. governmental-run National Endowment for the Arts[3] (Boyle, 2013). For example, musician Amanda Palmer successfully raised U.S. $1.2m in order to make a new album and art book in June 2012, which led to an album release and concert tour end of that year (Dredge, 2014). In this way, argue Mollick and Nanda (2016), crowdfunding efforts contribute to the innovativeness of the art industry as artists with unconventional ideas, or artists that create art which is not commercially viable, have more chance of being able to start their art projects.

Several empirical studies have explored the factors that determine fundraising success in reward-based crowdfunding in two prominent areas: the pre-launch characteristics of a campaign that determine a crowdfunding project's total funding amount, and the factors that affect backer's motivation and behavior during a campaign. Building on the categorization of Yin et al. (2017) and Kunz et al. (2016), the author offers the following categorization of six success factors: The campaign's size, its presentation, the campaign's creator, its rewards, the campaign's communication and external influence. Firstly, regarding the campaign size, it was found that a campaign's duration and its funding goal have an impact on its total funding result

[3] U.S. federal government spending on arts and culture. Approximately 71 million are awards were made directly to organizations and individuals that apply through the NEA's funding categories in 2016.

(Mollick, 2014). Secondly, it was found that backers consider various elements of a campaign's description as a signal of its quality. According to Mollick (2014), campaign quality is shown among other things by preparedness, characterized by the absence of spelling errors, the inclusion of a video pitch and regular updates on the product's status. In line with this argument, the depth and quality of a project's description, as well as videos and images are found to have a positive influence on a campaign's successful funding (Koch et al., 2014; Wang et al., 2011; Zhou et al., 2015, 2016). Thirdly, a creator's existing social capital such as his number of connections on social media, and backing history such as his number of successful projects launched in the past, are also found to be an important marker of funding success (Zheng et al., 2014; Mollick 2014; Zvilichovsky et al., 2015, Lu et al., 2014). In line with this, geography is found to be linked to the success rate and nature of a campaign as local backers are more likely to contribute in the early phase of a campaign (Mollick, 2014; Agrawal et al., 2014). Kuppuswamy (2015) explained this phenomenon as the potential support of family and friends of the creators. Fourthly, the choice in the number of rewards, the scarcity of different kinds of rewards and the price of these different rewards also impact the final total funding amount (Xiao et al., 2014; Kunz et al., 2016). A fifth of all, competition, or the existence of similar projects competing for donations, has a negative effect on the final funding result (Meer 2014). Finally, being featured on the platform has been empirically proven to have a positive impact on a campaign's funding success (Mollick 2014).

The success of a crowdfunding campaign also depends on the behavior and motivation of its backers. Research on these different types of behavior and motivation can be divided into three categories: social influence, choice bias, and backers' personal motivation. Firstly, Kuppuswamy and Bayus (2013) found that all crowdfunding projects are characterized by a U-shaped pattern of funding support, meaning that there is a spur of activity in the early and the last stage of the funding cycle. They explain this phenomenon as the goal-gradient behavior of backers, or behavior in which backers donate when they perceive their contribution to have a maximum relative impact on a campaign's success. In practice, this means that new backers donate, or existing backers increase their existing donations either at the start of a campaign or when a campaign is close to reaching its funding goal. This is evident from a drop in support once a funding goal is reached. This finding is corroborated by the study of Burtch et al. (2013), which analyzed donation-based crowdfunding and found that crowdfunding efforts are characterized by substitution effects or, in other words, characterized by the fact that prior contributions have a negative impact on later backers' decisions to contribute to a project. Secondly, according to Simons et al. (2017), the reward types and the price point of rewards also influence the behavior of potential backers, as backers exhibit a middle option bias in their choice of reward. Backers who value the product higher are found to prefer higher priced option in the rewards (Hu et al., 2015).

Lastly, relatively few studies have focused on backers' motivations behind funding a project. Using a qualitative method, Gerber et al. (2012) and Ryu and Kim (2016) listed various motivations such as the desire to collect rewards, help others, support certain causes or be part of a community. For example, Ryu and Kim (2016) identified "angelic backers" as sponsors with a philanthropic motivation to support projects, who are not motivated by rewards and who tend to support film, play and charity projects. By contrast, "reward hunters" are backers which are highly motivated by rewards, are not motivated by philanthropic considerations and who mostly support the design and game related projects. This is also supported by other studies which find that backers show characteristics of both donors and investors (Belleflamme et al. 2014). The major findings of previous studies are summarized in Table 1 below.

Table 1. Summary of the Literature Review

Context	Author (s)	Major Findings
Communication	Kuppuswamy et al. (2013)	The number of updates is positively related to a campaign's success.
	Mollick (2014)	The regularity of updates is positively related to the total number of a campaign's backers
External influence	Meer (2014)	Competition, or the existence of similar campaigns at the time of a certain campaign's creation, has a negative relation to a campaign's success
	Mollick (2014)	Exposure, or the fact that a project gets presented to the public on Kickstarter's popularity indices, is positively related to a campaign's funding success
Geography	Agrawal et al. (2014); Kuppuswamy (2015); Mollick (2014)	Geography, or the location of a project creator in an area with a high population density, is positively related to a campaign's success
Motivation	Gerber et al. (2012); Ryu and Kim (2016)	A campaign's backers exhibit various motivations for funding a specific campaign ranging from the desire to collect rewards, help others or support certain causes to the desire to be part of a community.

Context	Author (s)	Major Findings
	Bayus and Kuppuswamy (2013)	A campaign's backers decide to fund a project when their contribution has the largest perceived impact on a campaign's success, i.e. at the start or at the end of a crowdfunding campaign.
	Burtch et al. (2013	Crowdfunding efforts are characterized by substitution effects, i.e. prior contributions have a negative impact on later backers' decisions to contribute to a project.
	Belleflamme et al. (2014)	A campaign's backers show both characteristics of donors and investors.
Project size	Mollick (2014)	The length of a campaign's duration and the size of its funding are positively related to the total funding result
Presentation	Mollick (2014)	A campaign's preparedness (the absence of spelling errors, the inclusion of a video pitch, regular updates) is positively related to the total funding result
	Koch et al. (2014); Wang et al. (2011); Zhou et al. (2015, 2016)	A project description, related images, and videos as well as the question of whether the founder has previously backed other projects influence funding success.
Project creator	Lu et al. (2014)	Crowdfunding is not only driven by a creator's own social network; it is also driven by third party information propagation.
	Mollick (2014)	A project creator's personal (social) networks, or the number of people that a creator has contact with, is positively related to a campaign's success.
	Zheng et al. (2014)	An entrepreneur's social network ties, obligations to fund other entrepreneurs, and the shared meaning of a crowdfunding project had significant effects on crowdfunding project performance both in China and the U.S.
	Zvilichovsky et al. (2015)	A project creator's backing history, or the number of projects that he has backed in the past, is positively related to a campaign's success.
Rewards	Kunz et al. (2016); Xiao et al. (2014)	The number of choices of rewards is positively related to a campaign's success The inclusion of relatively high-priced rewards among these choices is positively related to a campaign's success The scarcity of different kinds of rewards is positively related to a campaign's success.
	Hu et al. (2015)	A campaign's backers exhibit a clear preference in their reward choices for higher-priced reward options.
	Simons et al. (2017)	A campaign's backers exhibit a clear preference in their reward choices for the middle-priced reward option.

2.3 In-process Promotional Messages

As reviewed above, crowdfunding studies so far have focused on either creators' or backers' behavior and their impact on a given project's funding success. In this body of research, research on quality signaling to backers focused exclusively on those promotional actions which provide information on the campaign to potential backers *before* the campaign is launched. Conversely, relatively little research has been done on the development of the campaign *after* the campaign is launched and *during* its funding. Of the little research available, both Mollick (2014) and Kuppuswamy (2013) found that the number of updates issued is significant to the success of the project. The positive effect of social media exposure on the total amount a campaign raises is also supported by Lu et al. (2014), who argues that crowdfunding success is not simply driven by a creator's own social network. Instead, crowdfunding campaigns are more likely to rely on external media coverage for their funding success. They also mention that the existence of promotional activities on social media has a correlation with the number of backers but that in this, it is less important in determining the success of the campaign than its initial pre-launch design.

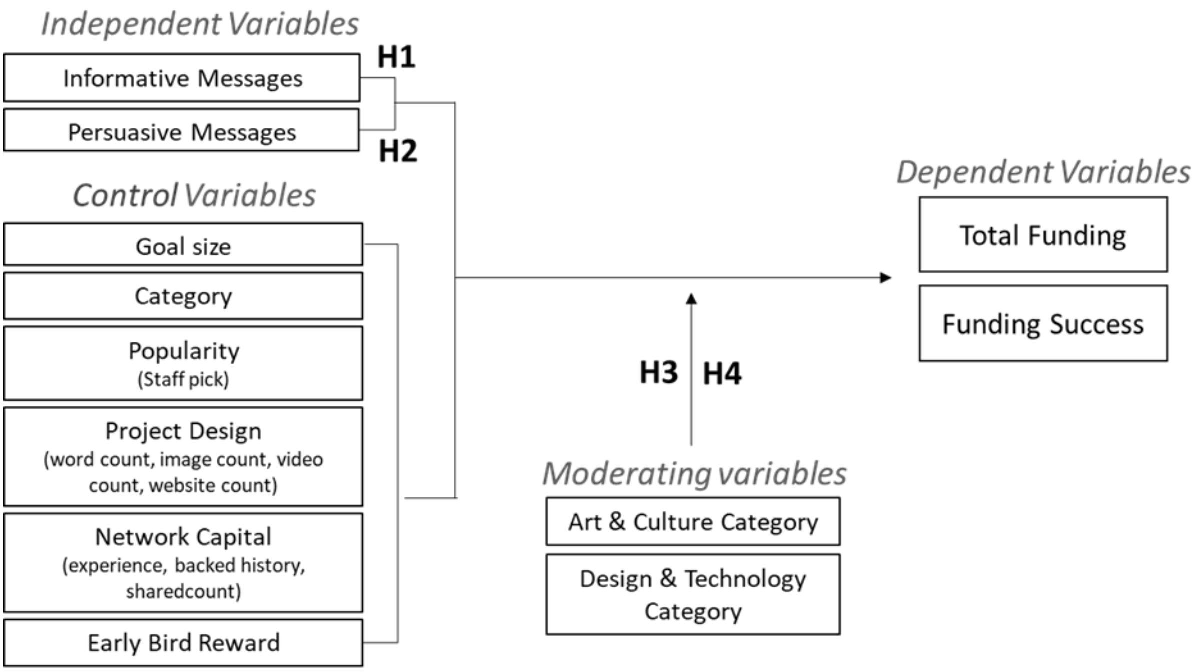
Other research, however, has stressed the importance of promotional messages in obtaining a maximum price in an auction setting. For example, Ducarroz et al. (2016) have examined how consumers respond to one particular type of intangible in-process promotions – messages – in auctions. These messages do not include promotional actions that are set before auction starts. Regarding message strategies, Ducarroz et al. draw a distinction between informative messages which focus on product attributes and persuasive messages which focus on motivating consumers. Their empirical results show that on an aggregate level, informative messages have an impact on the final auction result whilst persuasive messages do not. At the micro level, however, the two types of messages differed in their relative impact based on the timing and dynamics of the auction. For instance, persuasive messages proved to have a greater positive impact on new-bidder entry than informative messages. Conversely, for jump-bidders, which are bidders that bid repeatedly during an auction, these two types of messages both had significant negative impact. Finally, Ducarroz. et al. found that the bid rate during an auction had an impact on auctioneers' decisions to send out messages during bid intervals. Namely, auctioneers were found to be less likely to send out informative messages, and even less likely to send out persuasive messages, as the bid rate increased. Hence, these results suggest that if messages are not separated into distinct categories on the micro-level, the evaluation of their impacts can be misleading.

Since crowdfunding platforms can be seen as fulfilling an intermediary role comparable to those of auction houses, and in-process promotional messages can be seen as comparable to crowdfunding projects' updates, it can be argued that Ducarroz et al's (2016) framework is also applicable to crowdfunding context. For instance, project updates can motivate backers to spread the word about a specific campaign, either in person or in social media. In doing so, they exert an effect similar to that of in-process promotional messages during auctions. This research, therefore, adapts Ducarroz et al.'s (2016) framework on different types of message strategies, and its research methodology of auction-level analysis in order to investigate the impact of in-process promotional messages on a project's total funding amount and funding success.

III. Conceptual Framework

This section briefly introduces the conceptual framework for in-process promotional messages and discusses the theoretical background for the hypotheses that guide this research. The conceptual framework has two features. Firstly, this framework examines the factors that influence the total funding to be raised and the likelihood of funding success, which is the collective result of campaign characteristics as well as promotional activities. Secondly, this framework identifies different types of in-process promotional messages which may have different impacts on the campaign funding. The framework is described graphically in Figure 1 below.

Figure1. The Conceptual Framework



3.1 The Crowd-funding Setting

This research considers reward-based crowdfunding on the crowdfunding platform Kickstarter, in which a creator launches a campaign and subsequently directs promotional messages or so-called 'project updates' to backers during that campaign. This paper uses Kickstarter as it is not only the biggest reward-based platform but also uses only one system of funding (all-or-nothing) for all campaigns unlike other platforms such as IndieGoGo. On Kickstarter, updates are listed in a separate tab on the campaign page, which any member of the public can access. Updates are also sent out to a campaign's existing backers directly in the form of a personal message or email to backers. These messages help the creators to continue to be in touch with a campaign's backers in order to present gratitude and to communicate further information. The latter can include the progress of the current campaign or even offering backers new features that will be unlocked when a specific funding threshold is reached to entice these existing backers to contribute an even larger amount of money.

3.2 Message Types

This paper follows the model of message classification chosen by Ducarroz et al. (2016) in the auction setting. For their research, Ducarroz et al. (2016) opted for a dichotomy of generalized message types widely used in marketing: "informative" and "persuasive" messages (Farris and Albion 1979; Hunt 1976; Leffler 1981). This same distinction is applicable to the crowdfunding setting insofar that creators also issue different kinds of in-process promotional messages to backers in the form of project updates and insofar as backers have different motives that make them respond differently to these messages. In the following section, we try to leverage the theories behind these different message types and backer motivations, specifically prosocial motivation and self-determination theory, to examine the relationship between the two.

In their research on the informative role of advertisements, Vaughn (1980) and Aaeker and Norris (1982) proposed that informational messages carry a rational and cognitive appeal. These appeals relate information on the attributes of the product rather than trying to arouse a specific feeling in the consumer. Informative advertising aims to inform the consumer about the product, explain how it works and provide pricing and product information. Hence, we define informative updates as those messages that convey information about the attributes of existing campaigns, such as reminding backers of the campaign's existing specifics or providing information on the process of the campaign and changes in new rewards or goods (e.g., "This is the first of potentially 3 t-shirts backers may choose from!").

Generally speaking, informative messages incite a prosocial motivation; that is a motivation to help others. Prosocial motivations are directly linked to the perceived magnitude of the impact of the help. For example, Cryder et al. (2013) report that donations are higher when detailed information about the use of funds is given, and the perceived impact of a donation is thus higher than when this detailed information would not be given. This would mean that information on the process of the campaign would increase the perceived impact of each backer's investment. We can also link this to the concept of extrinsic motivation from self-determination theory (Deci and Ryan 1985, Ryan and Deci 2000). People are extrinsically motivated when they perform to achieve significant outcomes; in other words, they are influenced by external motivation such as rewards and relationship recognition. Hence, informative messages could influence backers with extrinsic motivation.

As for the persuasive role of advertisement, Koh and Leung (1992) found that persuasive advertising motivates people using emotional arguments rather than information and arouses a specific feeling in order to get them to purchase a certain product. The aim is to increase the demand for an existing good or service. Promotional activities could also include monetary promotion of free gifts to arouse excitement among existing customers. Hence, we define persuasive updates as those messages that aim to affect a goal rather than giving information on the campaigns. Examples include explicitly appealing for help with spreading information on the campaign and emphasizing the current funding percentage or the number of backers to elicit even more donations (e.g., “We have reached 50% of our total funding goal! Just a bit more guys! We can do it!”).

These persuasive messages also motivate prosocial behavior. Several experiment-based studies show that a feeling of prosocial impact leads to helping behavior. For instance, participants in a lab study were more willing to spend time on editing job applications when they received expressions of gratitude (Grant & Gino, 2010). Linking this to self-determination theory, we can thus infer that people are intrinsically motivated when they perform in order to gain inner satisfaction such as joy. Due to this motivation, such persuasive messages might incite the backers to spread the word or donate more money with the aim of helping the creators.

The author does not claim that the chosen two types of messages are the only appropriate form of classification as update posts in a crowdfunding setting can contain both informative and persuasive messages and also vary in length between different messages. However, the author felt that a dichotomous treatment was more appropriate than a continuous approach as the purpose of the updates are to a great extent one or the other. Moreover, as this paper shows later, distinguishing between these two types of messages increases the fit of our model.

3.3 Hypothesis Development

3.3.1 Impact of Messages on the Campaign Performance

Industry advisors insist that project initiators need to develop a campaign that communicates with traditional media, social media, bloggers and backers. In line with this argument, Steinburg (2012) notes that posting campaign updates is one way to generate online visibility and excitement around crowdfunding projects. In fact, argue Kunz et al. (2016), the total number of updates has a positive correlation with a campaign's success.

The cause behind such a correlation is not identified yet. However, previous research has made several assumptions. Firstly, Gefen and Straub (2004) argue that a creator's perceived social presence, the degree of awareness of the other person in a communicative interaction, increases customers' trust in sellers and reduces customers' perception of risks in social commerce settings. This leads to a higher willingness to purchase a creator's product. This also applies to the crowdfunding setting as updates are a way for a campaign's creator to interact with the crowd, hence increasing the willingness to back a project. Secondly, updates may motivate backers to talk about a specific project. This argument is corroborated by Lu (2014), who shows that "patrons", backers that back the project and promote it on social media channels, make up of 27% of all backers in the sample data. These arguments show that updates function as an effective medium for a project's creator to interact with the crowd, hence resulting in a positive impact on gaining additional backers through existing backers' information propagation.

With only the campaign page, the creator is uncertain about the likelihood of attracting more sponsors and sponsors are also uncertain of a project's quality or the certainty of its delivery. The creators can use messages to reduce these uncertainties by interacting with the sponsors through updates. Uncertainty reduction theory states that an interaction reduces the uncertainties by using verbal and nonverbal communication strategies (Claude & Warren, 1949). At the start of the funding phase of a campaign, backers are more uncertain as there have been fewer interactions between the creator and the crowd to establish trust and act as an indicator of the campaign's quality. This uncertainty can be alleviated by using updates as a medium for establishing trust. Hence, this paper posits the following hypothesis:

H1.1 The total funding increases as the number of informative updates issued in the campaign increases.

H1.2 The likelihood of success increases as the number of informative updates issued in the campaign increases

According to the expectancy theory of Vroom (1964), motivation is based on an individual's expectancy that a certain effort will lead to achieving a certain result or a goal. Backers act as a medium of information propagation because they are motivated by a campaign's goal and the desire to make sure that they receive a campaign's rewards in an all-or-nothing system such as Kickstarter (Gerber and Hui, 2013). Hence, persuasive messages that put emphasis on a campaign's funding level will remind backers of their common objective of reaching the funding goal. Therefore, this paper posits the following hypothesis:

H2.1 The total funding increases as the number of persuasive updates issued in the campaign increases.

H2.2 The likelihood of success increases as the number of persuasive updates issued in the campaign increases.

3.3.2 Impact of Messages on the Funding Success Modified by Category

The papers from Sheng Bi (2017) and Ryu (2016) examine crowdfunding platforms in China and Korea respectively. They found that the motivation and behavior of backers differ depending on specific project categories. For instance, in Chinese crowdfunding platform, backers in the art and entertainment category are more influenced by peripheral routes of information propagation (word of mouth) compared to those who backed projects in the science and technology category. Ryu also points that different categories appeal to different types of backers, characterized by vast differences in backer motivation. Angelic backers, defined as older backers which are motivated by philanthropic considerations, tended to support films, plays and charity projects. Conversely, reward driven backers preferred the design and game projects.

Kickstarter lets creators categorize their projects into one of 15 categories and 55 subcategories. This paper divides these categories into 'Art & Culture' and 'Design & Technology'. The difference between these two categories is that the campaigns in 'Design & Technology' mostly focus on providing materialistic rewards and commercial goods to backers while that is not the case for the campaigns in 'Art & Culture' category. Campaigns in the art and culture category consist of art, journalism, music, theater, dance, film and photography projects. Campaigns in the design and technology category consist of technology, design, fashion, games, comics, publishing, crafts and food projects. Appendix A shows specific examples within these sub-categories. Because of this difference in the nature of campaigns, the author expects that the former will attract angelic backers while the latter will attract reward driven backers. As different project categories appeal to different types of backers with different motivations, these different project categories must also have different types of optimal message strategies appealing to these motivations. For example, reward driven backers may find informative messages appealing while angelic backers find persuasive messages that directly ask for help or show gratitude appealing. Hence, this paper posits the following hypotheses:

H3.1 The effect of informative messages on total funding is stronger for the design & technology campaigns in comparison to art & culture campaigns.

H3.2 The effect of informative messages on the likelihood of success is stronger for the design & technology campaigns in comparison to art & culture campaigns.

H4.1 The effect of persuasive messages on total funding is stronger for the art & culture campaigns in comparison to design & technology campaigns.

H4.2 The effect of persuasive messages on the likelihood of success is stronger for the art & culture campaigns in comparison to design & technology campaigns.

IV. Research Methodology

This paper carries out a quantitative data analysis using two models. First, this paper uses multiple linear regression models of campaigns' total funding. Total funding raised by the *i*th campaign is the continuous dependent variable for this paper. Below is the regression model used:

(1) $\ln(\text{Total funding}) = \beta_0 + \beta_1 \ln(\text{NInfo}_i + 1) + \beta_2 \ln(\text{NPers}_i + 1) + \sum_{n=1}^{N} \delta_n Z_{in} + \mu_i$

(2) $\ln(\text{Total funding}) = \beta_0 + \beta_1 \ln(\text{NInfo}_i + 1) + \beta_2 \ln(\text{NPers}_i + 1) + \beta_3 (\text{Cat}_i * \ln(\text{NInfo}_i + 1)) + \beta_4 (\text{Cat}_i * \ln(\text{NPers}_i + 1)) + \sum_{n=1}^{N} \delta_n Z_{in} + \mu_i$

The second dependent variable is the probability of a campaign's success as measured on a dichotomous scale – success or failure. For this, this paper uses a binary logistic regression model to understand whether campaign success can be predicted based on the two types of messages. Below are the equations:

(3) $\Pr(\text{Funding Success}) = \beta_0 + \beta_1 \ln(\text{NInfo}_i + 1) + \beta_2 \ln(\text{NPers}_i + 1) + \sum_{n=1}^{N} \delta_n Z_{in} + \mu_i$

(4) $\Pr(\text{Funding Success}) = \beta_0 + \beta_1 \ln(\text{NInfo}_i + 1) + \beta_2 \ln(\text{NPers}_i + 1) + \beta_3 (\text{Cat}_i * \ln(\text{NInfo}_i + 1)) + \beta_4 (\text{Cat}_i * \ln(\text{NPers}_i + 1)) + \sum_{n=1}^{N} \delta_n Z_{in} + \mu_i$

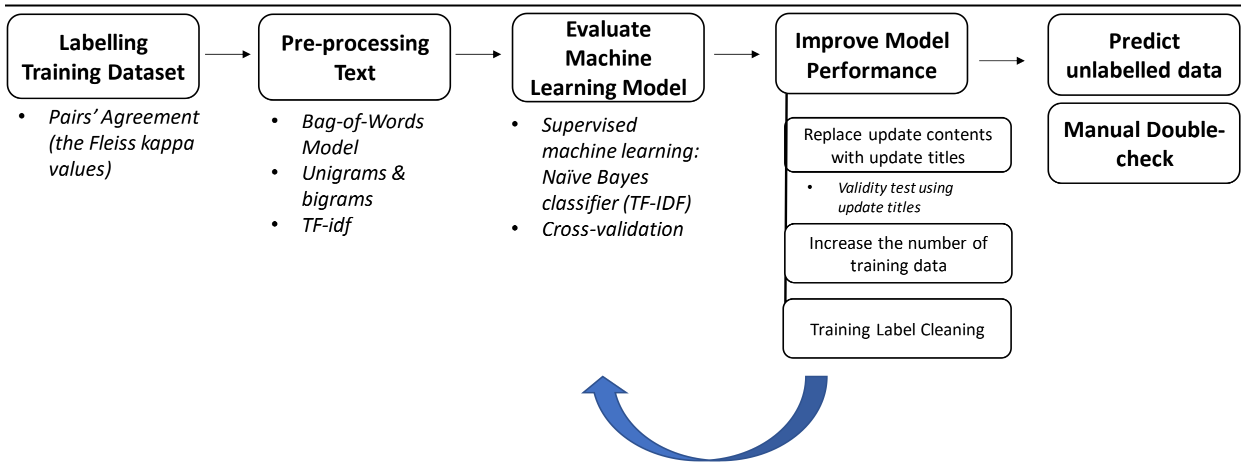
NInfo and NPers are the numbers of informative and persuasive messages sent during the *i*th campaign. Z_{in} are the control variables that will be further explained in section 4.2. The equation 2 and 4 shows the extended equation of 1 and 2 with interaction terms added to each independent variable to find out the moderating effect of different categories.

4.1 Independent Variable

Labelling each of the vast numbers of 34,920 update posts collected for this paper is difficult and very time-consuming (Ko & Seo, 2008). To deal with this problem, this paper uses "text mining" to assign labels from predefined 'informative' and 'persuasive' categories to update posts. Text mining is defined by Heyer (2009, p. 2) as "a set of computer based methods for a semantic analysis of the text that help to automatically, or semi-automatically, structure text, particularly very large amount of text." The methods of coding text operate either inductively, deductively or as a mix of both approaches sometimes referred to as an abductive approach (Wiedemann, 2015). Abductive research develops its categories from samples of the data and afterwards utilizes category systems for subsuming new data and testing hypotheses (Kelle, 1997). The research method employed in this paper which utilizes machine learning to classify texts is called “supervised learning”. The supervised learning algorithm uses the labelled dataset, which is a set of update posts that has been classified manually by the researcher, as a representative dataset for classification, and applies this classification to the rest of the unlabeled dataset to allow for automated classification (Ko & Seo, 2008). This automated classification method has been widely used in many research papers using different supervised learning algorithms such as naive Bayes (McCallum & Nigam, 1998; Ko & Seo, 2000), Support Vector Machine (SVM) (Joachims, 2001), and k-NN Nearest Neighbor (Yang, Slattery & Ghani, 2002).

The classification process for this paper is described in Figure 2 below. The process is comprised of five steps: Labelling the training set, pre-processing text, evaluating the supervised machine learning model, improving the model’s performance and finally predicting unlabeled data as well as manually checking ambiguous terms. This section aims to explain the theories and statistical modeling behind each process.

Figure 2. Text Mining Process



4.1.1 *Labelling the Training Dataset*

The major bottleneck of supervised machine learning is to label a large enough number of updates for the training dataset to accurately predict the rest of the unlabelled dataset. Once a certain amount of updates is manually labelled, the machine learning algorithm extracts characteristic patterns from documents of each category in a training phase (Wiedmann, 2015). To create a training set for a text classifier, the author faces two problems: how one can allocate updates to a certain category, and how one can make the process of manual classification more efficient to increase the number of the training dataset.

To solve these problems, the first task the author faces is to define each category. Laskey et al. (1989) in their paper find the task of classifying television commercials cognitively complex. Similarly, this paper too faced a problem in identifying creative strategies as exclusively emotional or cognitive, thus making exclusive categorization difficult. For example, update posts that asked backers for suggestions on possible new rewards did not fit perfectly in either of those classifications. In order to make this clearer, this paper opted to classify such posts as informative messages as they explicitly mention rewards and campaign progress. Hence, the author put in efforts to classify updates more clearly and to make a “code book” which tries to describe a category as accurately as possible (Krippendorff, 2013). As mentioned in section 3.2, the author chose to allocate one category per update post because the purpose of the updates is to a great extent one or the other. Following this code book, the author randomly selected 1000 updates, approximately 5% of the total number of updates, and codified these either as ‘informative’ or ‘persuasive’. The label was attached to each update if a certain entity fits into the definition of a category. The chosen subcategories are shown below in Table 2.

Table 2. Subcategories of Message Types

Strategies	Sub Categories
Informative messages	Campaign progress (product design, samples, artworks, shipping)
	Press and media share (sharing of press article/interview links)
	Stretch goals and new rewards (new reward-tiers)
Persuasive messages	Showing of gratitude (“Thank you!”)
	Appeals for exposure (“Please follow us on Twitter!”)
	Emphasis on goals (“We are halfway there!”)

To assess the quality of the author’s manual labelling, this paper followed the validation process used by Herzenstein et al. (2011). This validation process entails the separate labeling of a dataset by multiple people along certain criteria and then measuring the agreement of these labeling efforts by finding their Cohen kappa values (Cohen, 1960). For this paper, a marketing student placed the same training set of 1000 updates in one of the two dichotomous categories according to the code book. We then came together to discuss the unified type of message category for each update post. Cohen's κ was run to determine if there was an agreement between our judgements on whether 1000 update posts were either informative or persuasive. There was almost a perfect agreement between the judgments as $\kappa = .961$, $p < .005$. Considering the high level of agreement of 0.961 of the Cohen kappa values[4], labelling of updates this way is considered reliable.

The second task is to make the manual classification process more efficient. For instance, McCallum, Nigam, Rennie, and Seymore (1999) found that only 100 documents could be hand-labelled in 90 minutes and could only achieve just 30% accuracy. For this paper, as the goal is to use the classified data as an independent variable in the analysis, it is important to optimize the classification process so that each individual post is classified as accurately as possible, rather than merely estimating proportions of categories in populations correctly (Hopkins and King, 2010). Hence, the author must find a way to efficiently increase the training dataset in order to increase the fit of the model (Wiedmann, 2015). For this, the author chooses to replace 'update contents' with 'update titles'[5]. Examples of update titles and update contents can be found in Appendix B. 'Update contents' ranged from a minimum of 3 words to a maximum of 6023 words while update titles ranged from 1 word to 48 words. To check whether the classification of update titles can accurately represent the classification of update contents, this paper selected 100 randomly sampled update titles independent of contents from the previously labelled 1000 updates, manually coded them and compared them with the coding from their updates' contents. This process revealed a 95% match between the updates' titles and their contents.

Using 'update titles' instead of 'update contents' was not only effective in increasing the accuracy of the classifier but also made the entire process of increasing the training set much less time-consuming. When using 'update contents', the model maintained its low accuracy of 82.3 ± 4.2% even with the doubling of the number of training data set from 1000 to 2000. However, when replaced with 'update titles,' the model's accuracy increased to 87.3 ± 4.2%.

---

[4] Cohen (1960) suggested the interpretation of the Kappa result as follows: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

[5] In order to view update's contents, the users must click on that respective update's title.

A possible explanation of the lower accuracy arising from the first method could be an overfitting of the data. Overfitting is defined by Hawkins (2004: p. 1) as "the use of models or procedures that violate parsimony – that is, that they include more terms than are necessary or use more complicated approaches than are necessary." The segmented words for machine learning even without sparse words amounted to 16384 attributes for 'update contents'. In other words, with so many variables and without clear distinctions in the types of messages, a model learned the detail and noise in the training data to the extent that it negatively impacted the performance of the model on new data. Therefore, it is appropriate to classify updates as informative or persuasive based on their 'update titles' rather than on the 'update contents'.

4.1.2 Pre-processing of Updates

When classifying bulks of text, data scientists commonly use the "bag-of-words model" to train a classifier. What the bag-of-words model does is to represent each update as an orderless multiset of words while labeling the frequency of occurrence of each word, which is also called a *feature* (Salton & MacGill, 1983). This has several levels of granularity – word level, sentence level and feature level (Kumar & Sebastian, 2012). As creators choose their update titles to be short pieces of information (1 to 48 words), the word level granularity aptly suits this setting.

As updates are varied in their contents, all of them hold a common problem – the presence of textual noise. It is required to pre-process the text to eliminate textual noise before applying text classification techniques. Agarwal, Godbole, Roy, and Punjani (2007: p. 3) define noise as "any kind of difference in the surface form of an electronic text from the intended, correct or original text." People are less careful about the lexical accuracy of written content in informal modes of communication or online documents. The noise from such text normally contains text disfluencies such as spelling errors, nonstandard words, missing punctuations, missing letter case, abbreviations, repetition (Agarwal et al. 2007). One example of textual noise from the update data is "SATTT-UUUURRRR-DDAAAAYYYY!!!" (update id: 16471). Thus, using Rapidminer Studio 7.5.001, the author pre-processed all the update titles as follows:

a) Tokenization: This process converts each update (document) into a 'bag-of-words', In other words, it splits the text into a sequence of tokens, which consists of one single word. Tokens can be used to build word vectors by generating n-grams in a later process, which will be explained below.
b) Replace contraction: words such as "You're" should be split into "You" and "Are".

c) Transformation to lower case: As character strings are capital sensitive, all capitalized letters are transformed to lower case.

d) Filter Stopwords: This method is based on the idea that removing non-discriminative words – that is "Stopwords" - reduces the feature space of the classifiers and helps them to produce more accurate results (Silva and Ribeiro, 2003). Stopwords are the most common words in text documents such as articles, which do not add any meaning to the documents. Examples of stop words are the, in, a, an, and with.

e) Stemming: According to Jivani (2011: p. 1930), the goal of stemming is "to reduce inflectional forms and sometimes derivationally related forms of a word to a common base in order to save time and memory space". For example, "agrees" and "agreed" will all be reduced to "agree". Various stemming algorithms for different languages can be chosen. Amongst them, the author chose "Snowball" developed by Porter which is a detailed framework of stemming developed from Porters Stemmer algorithm proposed in 1980 and also one of the most popular stemming algorithms proposed thus far (Porter, 2001). Porter developed the algortihm based on the idea that there are groups of smaller or smaller suffixes that make up all the 12,000 suffixes in the English language (Porter, 1980).

f) Remove Punctuation and numbers: All the punctuations are removed except for "!", "%". A paper from Nasukawa et al.(2007) shows that the extraction of syntactic entities such as noun phrases or parts of speech is not accurate in the absence of puncutations. Also, in sentiment analysis, the number of exclamation marks were taken as emotion intensifiers (Kumar & Sebastian 2012). In line with this, the author intuitively also noticed that "!", "%" and numbers are important markers to distinguish persuasive updates such as "we are 80%! Let's keep on sharing!". Hence, these three were transformed into "exclaim", "percent" and "number" respectively.

Rather than to allocate every individually stemmed word in the training corpus to a distinct feature column and to a value of frequency 1, *feature selection* can improve the accuracy and efficiency of the model as well as reduce overfitting (Ikonomakis, 2005). The author chose two ways to perform feature selection to be more discriminating about which words to provide as input to the learning algorithm: generating feature terms and choosing feature values (Forman, 2007). Regarding the former, the process involves an algorithm which creates a term of any consecutive sequence of tokens (n-Grams) in a document as feature terms rather than creating a term with just an individual word. For instance, having a single feature representing the phrase ‘stretch goal’ can be far more predictive than just having one feature for the word ‘stretch’ and another for the word ‘goal.’ In this paper, the author chose a maximum of 3 terms for each token (e.g., “a new reward”).

Once a decision has been made about generating feature terms, each term has to be weighed to quantify the relative importance of different terms in a selected set. For this, this paper chose TF-idf weighing (Term Frequency - Inverse Document Frequency) instead of a simple term frequency/occurrence count. In TF-idf, while all terms are weighed equally in normal ‘Term Frequency’ processing, the importance of each term increases proportionally to the number of times a word appears in each update but is offset by the frequency of the word in the corpus. Examples of the terms weighed down are “of” and “that” (Debole and Sebastiani, 2003; Ikonomakis et al., 2005). This can be done by multiplying the term frequency for each term by the Inverse document frequency as shown in equation 5 below:

$$(5) \qquad tf(t, d_i)\ x\ idf(t, D) = \frac{f_{t,d}}{\max f_{t',d}} \times log\frac{|D|}{|\#(f_t)|}$$

$\boldsymbol{f_{t,d}}$ = Raw count of a term t in a document d

$\mathbf{max}\,\boldsymbol{f_{t',d}}$ = Number of terms t’ in document d

$|\boldsymbol{D}|$ = Total number of documents in the corpus

$|\#(\mathrm{f}_{\boldsymbol{t}})|$ = Number of documents where term t appears

4.1.3 Evaluation of the Machine Learning Model

An ideal situation would be to use a machine learning classifier that is perfectly suitable for any given circumstance. However, the research literature does not feature any single dominant ‘one-size-fits-all’ algorithm. Furthermore, as the model is based on training data that is inevitably incomplete and is thus to a certain extent subjective when measured against all existing update data, prediction cannot be 100% accurate. For such reasons, the quality of a model can be evaluated by quality measures such as accuracy, ROC, precision, recall, and F1 (Baeza-Yates and Ribeiro-Neto, 2001; Asch, 2013) which will be utilized in choosing a classifier in this paper.

Using the agreed upon values as a training dataset that consists of 2000 manually labelled updates, the author of this study built a text mining, machine learning document classification system using the Multinomial Naive Bayes classifier in the Weka package (Hall et al., 2009) of Rapidminer Studio 7.5.001, a popular machine learning software. This paper based a machine learning method on the results from multiple 10-fold cross-validations on different classifiers to find the best one. In 10-fold cross-validation, the original sample of 2000 labelled updates is randomly partitioned into 10 equal sub size samples where a single subsample is retained as validation data to test the model, and the remaining 9 subsamples are used as training data. This is then repeated 10 times (folds) so that each subsample can be used once as a training data. This allows to optimally use all the labelled data to be used both for training and validation (Williams et al., 2015). The author opted for Naive Bayes amongst others such as SVM (support vector machine) and KNN as it delivered superior analytical results in terms of accuracy, AUC, precision, recall, and f-measures[6]. This can partially be explained by the fact that a Naive Bayes classifier does not need large training dataset to perform well (Domingos & Pazzani, 1997). Amongst Naive Bayes models, the Multinomial Naive Bayes classifier, which uses a multinomial distribution for each of the features, proved to be the most effective for carrying this research.

Multinomial Naive Bayes (MNB) captures word frequency information in documents. In other words, it estimates the conditional probability of a particular tokenized word in a class as the relative frequency of term *t* in documents belonging to class *c* (McCallum & Nigam, 1998). The individual word occurrences are considered to be the "events" and the document to be the collection of word events. Here, the length of the document or the order of the words is independent of the probability of each word event. Each document *d* is drawn from a multinomial distribution of words, which is similar to the "bag of words" model for documents. MNB classifier has been used by numerous people due to its simplicity, efficiency, and efficacy (Lewis et al. 1994; McCullum et al. 1998; Nigam et al. 2000; Russell et al. 2003). Equation 6 below shows the probability equation behind this classifier:

[6] 'Precision' is the number of positive predictions divided by the total number of positive class values predicted. Precision can be thought of as a measure of a classifier's exactness. 'Recall' is sensitivity on false positives that measures a classifier's completeness. 'F-measures' show the balance between precision and recall. 'ROC' analysis plots the sensitivity against the rate fallout (Powers, 2011). The AUC (area under a ROC curve) quantifies the overall ability of the test to discriminate between two categories (Hanely & McNeil, 1982).

(6) $$\mathrm{P}\left(d_i \middle| c_j;\theta\right) = P(|d_i|)|d_i|!\prod_{t=1}^{|V|}\frac{P(\omega_t|c_j;\theta)^{N_{it}}}{N_{it}!}$$

Θ: A parameter for a mixture model.

c_j : mixture components of a mixture model consist of C = {c1, ..., c|C|} whereby each component is parameterized by a disjoint subset of θ.

d_i: a document,

ω_t: a word

4.1.4 Improvement on Machine Learning Performance

Once the training set is classified using the MNB classification, its accuracy is then evaluated to decide whether to further improve the performance. One way to improve the accuracy is to increase the size of the training dataset (Zhuang et al., 1994; Foody & Mathur, 2004). Due to the replacement of 'update contents' to 'update titles', the author could easily increase the training dataset until there was no change in accuracy from 90.98± 2.3% up to a total of 3000 update titles. This includes the 1000 updates that were classified with a second reader for validation purposes (refer to section 4.1.1), and 2000 updates that were manually coded solely by the author.

Because 2000 updates were solely coded by the author, it is necessary to inspect the training data for any erroneous labelling, through (computer-assisted) *training label cleaning* (TLC) (Esuli & Sebastiani, 2013). For this, Esuli and Sebatiani (2013) introduce a technique called the Confidence-based Technique. This technique is used in this paper as MNB from the Weka package used in Rapidminer Studio 7.5.001 return a score of confidence in its own prediction. Using this confidence score, the mislabelled updates are top-ranked in decreasing order of confidence level. The rationale behind this technique is to investigate the mislabelled update titles with the highest confidence first and working down the rank list to start with the ones most likely to be misclassified. Using improved training data after correcting 11 labels as a classifier resulted in a final accuracy of 91.16 ± 1.16%. Appendix C shows the confusion matrix from the training dataset of 2932[7] updates and the AUROC (Area Under the Receiver Operating Characteristic curve). The model resulted in 91.16% ± 2.26% accuracy, with a

[7] When exporting the update messages from R, many update messages were erroneously duplicated. Hence, this paper removed these duplicates and cleansed the training dataset of 3000 updates to 2932 update messages.

precision of 90.91%, a recall of 94.25% and an f measure of 92.54%. The AUC for the model used in this paper lies in between 0.9 and 1.0 (0.961 ± 0.006). AUC evaluation differs in context. For example, 0.70 is considered strong effects in the field of applied psychology (Rice & Harris, 2005) and over 0.90 is considered excellent in a clinical context (Tape, 2005).

4.1.5 Prediction of Unlabelled Data

The investigation above also allowed the author to deal with the problem that terms coexisted both as part of the informative and persuasive categories, thus confusing the machine and misclassifying the updates. The terms that exist in both categories are “plan”, “midway”, “donors”, “stretch” and “free”. To give an example, because posts that mentioned the words “stretch goals” were classified as informative, the words “final stretch” or “home stretch” which referred to the final few days of a campaign track leading up to the campaign’s end were also wrongly classified as informative. To solve this problem, this paper filtered the updates with these terms after running the classifier on the unlabeled updates, which amounted to a total of 3,658 updates. Then the author double-checked each of them manually with the purpose of correcting possible mislabels.

4.2 Covariates

The regression model for this paper includes a total of 12 control variables in the vector Z (N = 12), described in Table 3. These variables are time-invariant quality signals of campaigns, mostly set *before* the start of the campaign. This paper needs to control for these factors in its analysis of the impact of in-process promotional messages on the outcome of a campaign, as all factors represent the *inherent* quality of the campaign except for the social media counts. These control variables will be explained in detail in data section 5.3.

4.3 Controlling for endogeneity

The problem of endogeneity may occur when the unobserved part μ_i in equation 1 of campaigns is also correlated with the number of informative and persuasive messages besides the total funding in the above equation, leading to biased and inconsistent estimators. For example, backers may observe a factor that is not observable to the researcher such as reading news articles on the campaign that affects their pledge, while this factor also affects the creator’s decision to post an update, such as about the news article. One can control for endogeneity using instrumental variables, a method also widely used in marketing literature

(Villas-Boas and Winer 1999; Yang, Chen, and Allenby 2003; Ducarroz et al 2016). What this achieves is that it decomposes independent variables into parts that are related with the error term μ_i. This separation hinders biased result from endogenous variables. An instrumental variable has two properties: changes in instrumental variables are (1) associated with changes in explanatory variables, (2) but do not lead to changes in the dependent variables (Bouwden, 1984).

In choosing instrumental variables that are highly correlated with the $\mathrm{NInfo_i}$ and $\mathrm{NPers_i}$ variables, this paper follows the method from Ducarroz et al. (2016). The instrumental variables chosen by Ducarroz et al. reflect the average number of messages issued from previous auctions of tickets with the same departure and destination cities. Similarly, an average number of informational messages in other similar campaigns can indicate that creators develop a specific message strategy to be employed in new campaigns they launch. Here similar campaigns are the past campaigns in the respective sub-categories such as dance, theater, and music. As these message strategies used in the past campaigns do not affect the result of the current campaign, they also meet the second property of the instrumental variable explained above. Hence, this paper uses the average number of informational messages and persuasive messages in the past campaigns in the respective sub-categories as instrumental variables (Refer to Table 3 below for descriptions). This paper later checks the validity of these instruments. Below are the functions of instrumental variables IV:

$$(7)\quad \mathrm{NInfo_i} = \beta_{\mathrm{NInfo},0} + \sum_{n=1}^{N} \delta_{\mathrm{NInfo},n} Z_{in} + \sum_{m=1}^{M} \tau_{\mathrm{NInfo},m} IV_{im} + \mu_{\mathrm{NInfo},i}$$

$$(8)\quad \mathrm{NPers_i} = \beta_{\mathrm{NPers},0} + \sum_{n=1}^{N} \delta_{\mathrm{NPers},n} Z_{in} + \sum_{m=1}^{M} \tau_{\mathrm{NPers},m} IV_{im} + \mu_{\mathrm{NPers},i}$$

V. Data

5.1 Empirical Setting

In this section, we discuss the empirical context of the study. Kickstarter, the representative reward-based crowdfunding platform, has raised close to 3 billion for over 100,000 projects since its founding in 2009. Campaign creators prepare a campaigns' contents, such as project descriptions, videos and images and the reward tiers, before the start of the campaign. They can choose any length of time up to 60 days for their campaign. Backers can decide to withdraw their pledge or to increase their pledge at any time of the campaign. An example of a campaign on the Kickstarter platform can be found in Appendix D.

5.2 Data Collection

Data was collected from finished projects on Kickstarter. The author of this paper used a customized computer script that automatically scraped data of the chosen sample of projects. However, the data of the specific financial contributions of each individual is not provided by Kickstarter. Consequently, the extracted financial data is limited to the contribution totals of both funding and the number of backers of individual projects. To ensure comparability between campaigns, first, the data set was limited to campaigns that were launched from January 2016 to April 2017, which also includes the most recent completed Kickstarter campaigns at the time of data gathering (April 15^{th,} 2017). This also means that this paper discarded data from campaigns that were still running at the time of extraction and were canceled or suspended. Thirdly, this paper controlled for the duration of a campaign by keeping only 'long campaigns' with a fixed duration of 30 days as the duration is found to influence the result of the campaign (Mollick, 2014). Fourthly, in line with the suggestion by Mollick (2014), this paper excluded projects with extreme funding goals of below US$100 or over $1 million to guarantee the representativeness of its analytical results. Fourthly, as this paper analyzes the contents of project updates, it is by necessity limited to the languages in which the author of the paper is versed. Consequently, this paper limits itself to English-language data stemming from four English-only speaking countries such as US, Great Britain, Australia and New Zealand. Countries such as Singapore or Canada with two or more official languages were excluded. Finally, this paper omitted campaigns with missing pages due to intellectual property disputes.

Following, this paper again reduced its sample depending on the collected update titles and contents. Firstly, it discarded those campaigns with private project updates only accessible to backers, as this paper was not legible to analyze these. Secondly, this paper ignored updates posted after the end of the campaign as these would distort analytical results for the updates posted within the fixed duration of thirty days. Thirdly, during the investigation of update posts, some update posts on the last day were found to be those posted after the campaign was over (the time the update is posted on Kickstarter is not specified on update posts). Hence, the update posts on the last day of the campaigns were discarded. The robustness is checked for this decision in the analysis part of section 6. Regarding covariates, this paper ignored campaigns that had written campaign descriptions in the form of images instead of texts. After cleaning the data for inaccuracies, this resulted in a dataset of 7,428 projects and 34,920 observed project updates.

5.3 Variable Definition and Measurement

As shortly described in Table 3 below, there are two dependent variables in the analysis: the total amount of funding that a campaign has raised and funding success, that is whether a campaign reached sufficient funds to reach its goal. Independent variables are the number of messages sent during the campaign for each type of messages. The control variables are the quality signals found in the above-mentioned literature regarding the success factors of crowdfunding in section 2. ‘Funding duration’, which is also one of the quality signals found in other literature, is not included as a control variable as the sample was only limited to 30-day campaigns.

- *Number of Backers:* The number of backers that funded the campaign. According to the paper from Kuppuswamy and Bayus (2013), backers are found to react to the actions of other backers which shows herding and bystander effects among backers.
- *Goal size*: Due to “all or nothing” threshold model that Kickstarter uses, the funds can only be collected once the goal is reached. This makes it important for project creators to set realistic goals as setting a too high goal may demotivate backers to fund the project while setting too low may result in project non-delivery. The amount of total funding that creators seek to raise for their crowdfunding campaigns has found to be negatively associated with success (Mollick, 2014).

Table 3. Variable Definitions and Measurement

Variable Name	Variable Definition	Measurement	Type
Dependent Variables			
Total Funding (DV1)	The amount of money that was pledged during the campaign (in US $)	Integer	Metric
Success (DV2)	Whether the campaign earned sufficient funds to reach its goal	Success = 1 Fail = 0	Dummy
Independent Variables			
NInfo (X1)	The total number of informative messages that a campaign creator posts during the campaign	Log scale	Metric
Npers (X2)	The total number of persuasive messages that a campaign creator posts during the campaign	Log scale	Metric
Control Variables			
NBackers (Z1)	The number of backers that backed the project	Integer	Metric
GoalSize (Z2)	The amount of money that a creator needs to complete their project	Integer	Metric
Category (Z3)	Whether the campaign is in one of the following categories: art, journalism, film, music, theater, dance, photography. The rest are campaigns in the following categories: design, fashion, games, comics, technology, crafts, food, publishing	Art and culture = 1 Design and tech = 0	Dummy
Staffpick (Z4)	Whether the campaign was presented in the staff pick section of the Kickstarter homepage	Staffpicked = 1 Not picked = 0	Dummy
Wordcount (Z5)	The number of words used in the campaign	Integer	Interval
Videocount (Z6)	The number of videos used in the campaign	Integer	Interval
Imagecount (Z7)	The number of images used in the campaign	Integer	Interval
Websitecount (Z8)	The number of websites of the creators, including homepage, blogs, Facebook page, Instagram account	Integer	Interval
Experience (Z9)	The number of campaigns launched by a creator before launching this specific campaign	Integer	Interval
Backedhis (Z10)	The number of other campaigns that the project creator backed before.	Integer	Interval
Sharedcount (Z11)	The total number of shares on social media sites such as Facebook, Linkedin, Google, Pinterest, Linkedin	Integer	Interval
Earlybcount (Z12)	Whether there was an 'Early Bird Reward's offered in the campaign	Earlybird = 1 No Eb = 0	Dummy
Instrumental Variables			
NInfoIns (iv1)	The average number of informative messages in past campaigns in the respective category	Integer	Interval
NPersIns (iv2)	The average number of persuasive messages in past campaigns in the respective category	Integer	Interval

- *Category:* This paper clustered these categories into two different categories: 'Art and Culture' and 'Technology and Design'. The former includes campaigns that are known to provide more intangible reward forms and consists of art, music, theater, dance, photography, and journalism. The latter focuses more on tangible rewards and consists of the categories technology, design, fashion, games, comics, publishing, crafts, food.
- *Popularity:* The popularity of a project can be detected through the possible mention of a project as a *Staff Pick*. Kickstarter has a dedicated team that spots and sorts projects that they find exceptional. These selections are featured as staff picks in a separate section in 'Projects We Love'. As their objective is to help the chosen campaigns to be exposed, they not only list them in a further search category but also reward the campaigns with a badge on their project description pages. Staff pick badges and this popularity list serve as a signal of a campaign's trustworthiness as it reduces the perceived risk of the backers (Flanagin et al., 2014).
- *Project Design*: Detailed project descriptions serve as an indicator of preparedness and increase a project's perceived quality for backers (Mollick 2014). A campaign's project descriptions are where creators can develop a project's story and convey who they are, what they have done, what product they are planning to deliver and why they need a campaign to be funded. The completeness of the information in a project's product description is found to be an important factor when obtaining financial resources from investors (Marom, 2012). In Kickstarter, you can develop a project's story in four different ways: through written descriptions, images, videos and extra information through a separate website. Therefore, this paper identifies four quality signals:
    - *Word count:* the product descriptions in e-commerce settings are found to be a decisive success factor (Palmer, 2002). As backers, who have yet to see the rewards in life, the more details help to reduce their risks.
    - *Image count / Video count*: stories in project descriptions can be effectively conveyed with the addition of images and videos. The visual appeals are found to be important in consumer decision-making process as it stimulates a variety of sensory channels at the same time (Lindgaard et al., 2006). Images and videos give a clear and vivid idea of what rewards may consist of, enticing backers to donate to the project based on the type of reward. They may also foster further

trustworthiness by featuring the project creator's appearance in the video or image, thus intimating a sense of familiarity between a project's creator and its backers.

- *Website count*: a creator can have multiple channels to feature additional information for the campaign. These channels include websites, Youtube channels, Facebook pages etc. For some projects, these additional channels hold external functionalities such as the ability to vote for changes in rewards or the addition of rewards, fostering tighter communities among backers. The possession of a website also conveys professionalism and commitment on the side of creators.

- *Network Capital:* Researchers have found out that social capital of the entrepreneurs and their connections influence the success of their financing efforts (Shane and Cable, 2002). Social network connections are found to have a significant positive effect on the number of backers (Mollick, 2014; Lu et al. 2014; Kunz et al. 2016). The aforementioned research used the number of Facebook friends of the creators as a measurement of network capital. However, this paper does not use this variable as it is impossible to identify accurately the number of Facebook friends at the launch of the campaign. In other words, it may be possible that the creators expanded their connections during and after the campaign, especially if the project succeeded. Instead, this paper uses three other variables to measure network capital:
    - *Experience*: The experience of a creator in launching a campaign means the creator had a chance to expose his or her brand before the launch of the campaign. Even better, some creators that continue creating campaigns with different products in the same product line have already garnered a fan base from the previous campaigns. It is likely that the creator would send an email out to the previous backers of the new project that they might again be interested in a given new project. Moreover, it also means that the creator is already aware of factors determining crowdfunding processes, such as update frequency, presentation, and promotional methods. In other words, as previous experience increases the expertise of the creator, it will also impact on the result of the campaign.
    - *Backed History*: Funding other campaigns help to build trust within the community, and thereby, influence on the willingness to back (Crosby et al., 1990). Creators were found to work together to promote each other's campaigns through updates to spread words, meaning the backers from the other campaigns could also become

part of a project creator's social network capital (Colombo et al., 2015; Zvilichovsky et al., 2015).

- *Sharedcount*: Sharedcount is the number of times the campaign links were shared on various social media channels. These include Facebook, Google, Pinterest and LinkedIn.

- *Early Bird Rewards:* Kickstarter's tools allow campaign creators to limit their rewards, specifying exact quantities. The funding graph of campaigns by Kuppuswamy and Bayus (2016) shows that the first and last week of the campaign drive the most pledges. In practice, pledges during the first week of the campaign are partially driven by "early bird" rewards, where only a handful number of backers can have a chance to get rewards for a lower price. This can be explained by the fact that quantitative limits imply scarcity which serves as a quality signal in marketing by creating a sense of uniqueness and distinctiveness (Stock & Balachander, 2005).

5.4 Descriptive Statistics

Table 4 contains the descriptive statistics of our total sample. The dependent variable, the total funding amount, ranges from $2 to approximately $3.5 million in our data set. The average of the total funding amount is $20455 (SD = 108028.05), the distribution of which is skewed to the right. The second dependent variable, funding success, shows that the overall success rate of the Kickstarter campaigns in our sample is 69.6%. Surprisingly, this value is twice as high as the average success rate of 35.38% published on Kickstarter statistics page. One possible explanation could be that the success rate is higher amongst campaigns with a fixed duration of 30 days. Although Mollick (2014) found that a campaign's duration is positively related to a campaign's success within a range from 1 to 60 days, the Kickstarter team recommends a duration of 30 days or less to generate a greater sense of urgency amongst potential backers. It may be possible that the increased rate of success of the Kickstarter projects analysed in this paper serves to prove this point, as this paper dealt exclusively with 30-day campaigns. The moderating variable, category, shows that campaigns belonging to the Art & Culture category consist of 33.6% (n = 2,494) of the whole sample, while the Design & Technology consists of the rest of 66.4% (n = 4,934). Figure 3 below shows that a higher proportion of campaigns in Art & Culture category have succeeded compared to that of Design & Technology category. The independent variable will be discussed further in the next section.

Table 4. Summary Statistics

Variables	N	Min	Max	Range	Sum	Mean	SD
Totalfund	7428	2	3560643	3560641	151941687	20455.26	108028.05
Success	7428	0	1	1	7428	0.70	0.46
NInfo	7428	0	86	86	21637	2.91	3.90
NPers	7428	0	18	18	13283	1.79	1.92
NBackers	7428	1	85581	85580	1788985	240.84	1424.21
Category	7428	0	1	1	7428	0.34	0.47
GoalSize	7428	101	1000000	999899	113439238	15272	39573
Staffpick	7428	0	1	1	1478	0.20	0.40
Wordcount	7428	26	4871	4845	6690290	900.69	652.61
Videocount	7428	0	12	12	7106	0.96	0.67
Imagecount	7428	0	95	95	56636	7.62	9.35
Websitecount	7428	0	8	8	10587	1.43	0.97
BackedHis	7428	0	890	890	75270	10.13	29.97
Experience	7428	1	79	78	14302	1.93	2.93
Sharedcount	7428	0	165275	165275	4167393	561.04	2747.01
Earlyb	7428	0	1	1	806	0.11	0.31
NPersinst	7428	1.38	2.17	0.79	13283	1.79	0.28
NInfoinst	7428	1.10	5.83	4.74	21637	2.91	1.3

Figure 3. Funding Success Per Category

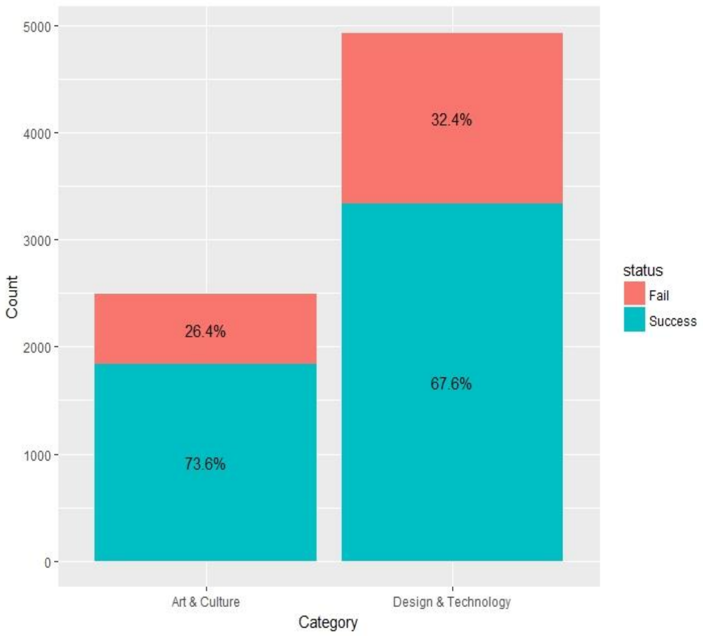
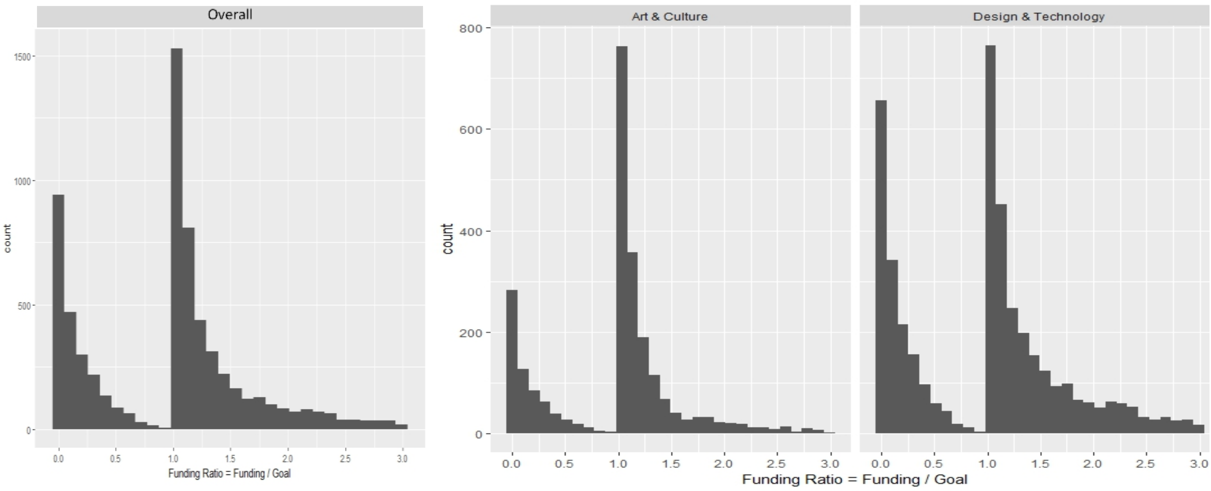


Figure 4. Distribution of the Funding Ratio



The distribution of the ratio of the total funding to the funding goal is illustrated in figure 4 above. If the funding ratio is below 1.0, that means that the campaign has failed. The overall distribution graph shows that most Kickstarter campaigns fail or succeed by small margins. This finding is in line with the finding from Mollick (2014). While 50% of failed projects could only reach 10% of their funding goal, 50% of successful projects reached to up to 125% of their funding goals. One possible explanation for this phenomenon is provided by Kuppuswamy and Bayus (2013) in their paper on goal-gradient effects. Once a funding goal is within reach, potential backers perceive their contributions as being relatively more important in attaining that goal. Hence, these backers are increasingly motivated to help a campaign reach its goal. Conversely, once a campaign has reached its funding goal, potential backers perceive their contributions as being relatively less important and are therefore less motivated to invest money in it.

Furthermore, projects in the Art and Culture category were generally funded closer to their funding goal. For example, among successful campaigns, 81.4% of all campaigns in the Art and Culture category were funded somewhere between 100-150%, compared to only 52% in the Design and Technology category. This means that the remaining 48% of campaigns in the Design and Technology category are overfunded by more than 50%. One possible explanation for this difference is that backers are more motivated to fund successful campaigns in the Design and Technology category, because regardless of a backer's perceived contribution or intrinsic motivation, the campaigns in the category always offers a material reward in the form of a gadget, design bag, and video games unlike those in Art and Culture categories.

5.5 Kickstarter Update data

This section will focus solely on the independent variables – informative and persuasive messages – of this paper. Kickstarter update data is posted in the 'Updates' tab of the campaign page. These messages are displayed online and are visible to anyone visiting the campaign page. The updates will not only show up in the backers' activity feed on the Kickstarter website and in its app but are also sent to backers' emails. The data contain the update URL, exact time, update title and update contents of each update message. Table 4 provides examples of some of the update titles for each type of messages. These examples are apparent on the term frequency plot in Figure 5, which shows that the top 3 frequently found terms for informative messages are "goal", "stretch" and "new", while it is "day", "percent", "funded" for persuasive messages. As the terms "stretch" and "goal" shows the highest frequency, these were coded separately for extended analysis in the result section. Among all informative updates (n = 21,637), stretch goal-related updates consist of 10.1% (n = 2,184).

The total number of updates analyzed amounts to 34,920 update posts, out of which 62% (n = 21,637) consists of informative messages whereas 38% (n = 13,283) consists of persuasive messages. When analyzing the number of informative and persuasive messages per category, the author found that updates sent out for projects in the 'Design and Technology' category contained a larger proportion of informative messages (63.9%) than for those updates sent out for projects in the 'Art and Culture' category (56.7%). Furthermore, figure 6 reveals the right-skewed distribution of each type of message update. Although the range of informative messages extends up to a total of 83 messages of a campaign, the boxplot reveals that the median of the number of messages is only 2 for a campaign. This means that only a few campaigns, in fact, contain more than 10 informative messages. Conversely, the range for persuasive messages extends up to 15, with a median of 1. This shows that only a few posts contain more than 5 persuasive messages.

Table 4. Examples of Updates

Message Type	Update titles
Informative Messages	• New Reward Level • Stretch Goal #3: Vintage Leather Strap with Quick Release! • New Designs • Game System Update #3 • Story Announcement and SNEAK PEEK ARTWORK!
Persuasive Messages	• 75% of the way there! • £1000 reached! Massive thanks to everyone who pledged! • Loved Project by Kickstarter Staff and You! • ***HELP!!! WE HAVE NOT PASSED OUR INITIAL GOAL!!!*** • DropArt has facebook page, twitter, and KickBooster

Figure 5. Term Frequency of Each Type of Message

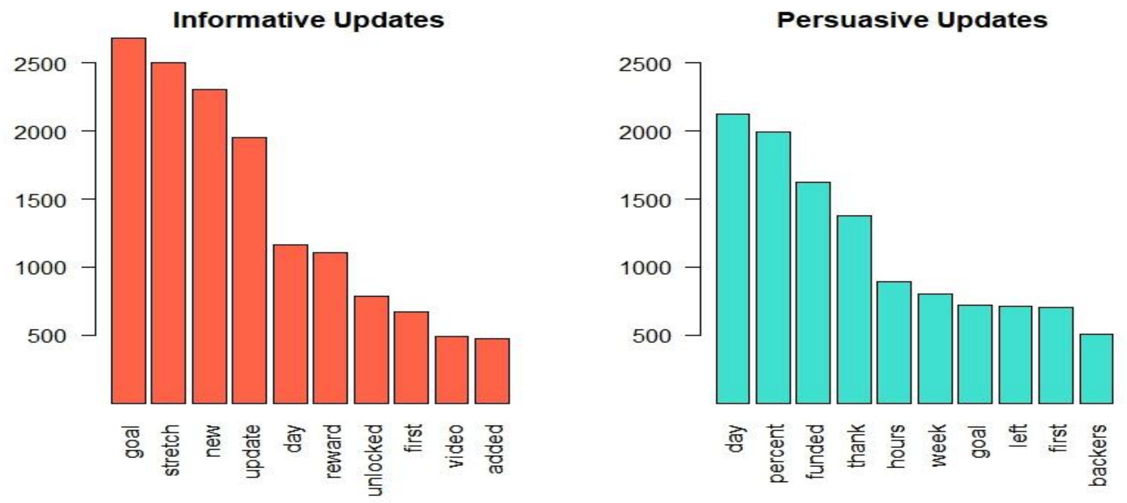
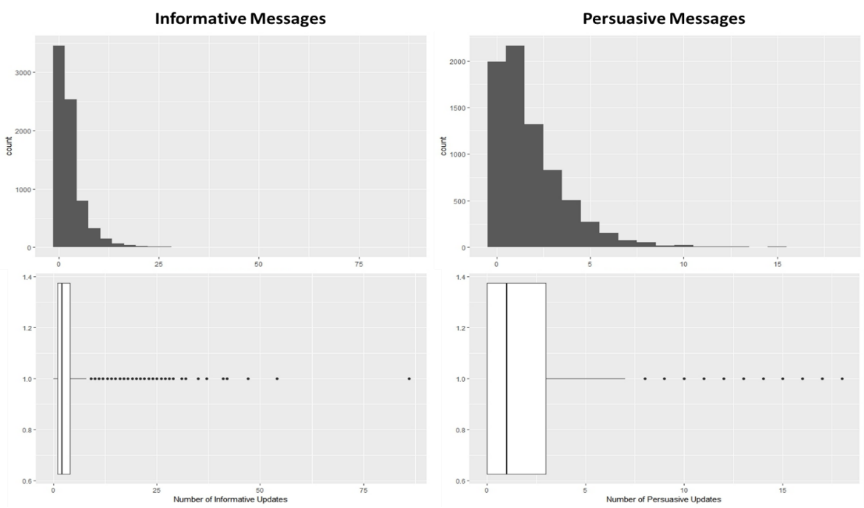


Figure 6. Distribution of Updates per Type



VI. Analysis and Result

6.1 Testing the Assumption

Before running the linear regression model, it is necessary to check whether this paper's assumptions regarding the performance of the linear regression analysis are met. Failure to meet these assumptions will result in biased or invalidated results. There are five key assumptions that the linear regression analysis needs to satisfy: 1) A linear relationship 2) Homoscedasticity 3) Multivariate normality 4) No or little multicollinearity 5) No auto-correlation (Janssens et al., 2008). This paper checks these assumptions below in accordance with the guide provided by Janssens, Wijnen, Pelsmacker and Kenhove (2008) in the order listed above.

As the relationship between advertising and demand is found to be non-linear in much of the literature (Cacioppo and Petty 1979; Rao and Miller 1975; Tellis 1988), the number of promotional update posts are also assumed to be non-linear. Hence, this paper used AIC among four forms (linear, diminishing return, inverse-U, and S-curve), to test which functional form best described the relationship between the total funding and the number of informative and persuasive messages. As diminishing returns (log form) provided the best fit, the author used this form for estimating the Equation 1 (refer to Appendix E.1).

Homoscedasticity means that residuals have the same variance for each value of the independent variable. This can be checked through scatter plots which showed the residuals getting larger as the prediction moves from small to large, indicating heteroscedasticity. In order to improve the model, this paper took a natural logarithm of the variables that showed asymmetrical distributions. The variables include the dependent variable 'Total Funding' and four other control variables, namely 'number of backers', 'number of shares', 'number of words' and 'goal size' (refer to Appendix E.2).

One requirement of multiple linear regression analysis is that the residuals of the regression (the error between observed and predicted values) should be normally distributed. The histogram of standardized residuals that satisfies the shape of the normal curve is reported in Appendix E.3. The fourth assumption is that there is no multicollinearity between variables. High correlations among predictor variables may lead to unreliable and unstable estimates of regression coefficients. Appendix E.4 displays the correlation matrix between all variables. If Pearson's r is close to 1, this means that there is a strong relationship between two variables.

All of the correlations of the variables differ significantly from zero but only one variable is greater than 0.6. This one variable that shows a high correlation with the total funding level represents the number of backers. However, because this variable is the control variable, any coefficients of the variables of interest are not affected. Therefore, multicollinearity is not a problem for the model.

A linear regression model requires each observation to be made independently of each other, meaning residuals of the independent variables should not be correlated of the residuals. The Durbin-Watson test on our model showed a value of 2.215. The value that lies in between 1.5 and 2.5 indicates a safe zone. Therefore, there is no autocorrelation in this model.

6.2 Dealing with Outliers

For model 1, where the author investigates the effect of message types on total funding process, this paper ignores the outliers for the reason explained in section 6.5.4. However, for the second model on funding success, calculating the model without the outliers was found to change the significance of the independent variables. Using the leverage statistic allows this paper to discover the campaigns which have a major influence on the predicted value (Janssens et al., 2008). If the outliers are found to have larger leverage value than the mean of leverage[8], which is 0.02 in this paper, the author concluded that the outlier is 'influential'. This led to the removal of 27 campaigns, leaving 7401 campaigns for the binary logistic regression analysis.

6.3 Impact of Messages on Campaign Performance

The coefficient estimates for the total funding model are reported in Table 5 and those for the hypothesis test in Table 7 below. In this paper, the author reports standardized coefficients[9] so as to be able to compare the effects of one variable to another. Coefficients were deemed significantly different from 0 if the significance (p-value) was lower than 0.05. Model 1.1

---

[8] Mean of leverage value equals to the number of parameters to be estimated divided by the sample size. For this paper, the mean 'leverage' value is therefore 0.02 (≅ 15/7428) (Janssens et al., 2008).

[9] Standardized regression coefficients remove the unit of measurement of variables to allow for comparison between variables. Standardized variable is calculated by subtracting the mean from the variable and dividing by its standard deviation. This results in standardized variables having a mean of zero and a standard deviation of 1, meaning the linear regression goes through the origin. Hence the intercept also becomes zero. These values are regressed to standardized coefficient (Janssens et al., 2008).

depicts the linear regression model's results on total funding without the moderator, while model 1.2 is run while taking interaction effects into account. For both models, 74.3% of the variation in the total funding amount can be explained by a variation in the model's independent variables (R^2). This 'adjusted R square' value differs significantly from zero (p = 0.000).

The model 1.1 result shows that both the number of informative messages (β_1= 0.026, 95% CI [0.013, 0.039]) and the number of persuasive messages (β_2= 0.088, 95% CI [0.075, 0.101]) have a significant effect on the total funding level of a campaign, supporting H1.1 and H2.1. In order to test the hypothesis that these two coefficients are significantly different from each other, the author used the 50% overlap technique introduced by Cumming (2009). In the event that the confidence interval overlaps by less than 50%, the beta weights are considered significantly different from each other ($p < 0.05$). For this paper, corresponding 95% confidence intervals of the standardized coefficients are estimated. As it can be seen from above, these two coefficients show no overlap. Hence, persuasive messages are statistically considered significantly more effective than informative messages in increasing the total funding. Model 1.2 shows that adding interaction terms, and thus letting the model take account of the differences between categories with respect to the effects of each type of message on the total funding level, does not change the adjusted R-square value. Importantly, however, the interaction between both types of messages and categories are found insignificant (β_3= -0.015, $p > 0.05$, β_3= 0.002, $p > 0.05$), rejecting both H3.1 and H4.1.

The coefficient estimates for the funding success model are reported in Table 6 below. Model 2.1 shows the result of the binary logistic model on the funding success without the moderators and model 2.2 with the moderators. In this paper, the author reports the coefficient as well as the odd ratio E(β) to allow for an interpretation of the magnitude of the each effect. As suggested by Janssen et al. (2008), the model fit, adjusted R^2 count, is calculated from the 'overall percentage correct' value from the classification table as shown in Appendix F to avoid incorrect conclusions. The model fit shows that the full model with and without the interaction effect reduces the prediction error respectively by 64.2% and 64.0%, in comparison with the null model. The chi-squared model showed significance, meaning that the full model is better than the null model and that coefficients of the variables differ from zero.

Table 5. Regression Result on Total Funding

Models	Model 1.1 (main results: total funding)			Model 1.2 (with moderator)		
	Coefficients	t	Sig	Coefficients	t	Sig
Intercept	0	4.102	0.000	0	3.952	0.000
Independent Variable						
lnNInfo	0.026***	3.912	0.000	0.034***	4.059	0.000
lnNPers	0.088***	13.493	0.000	0.088***	10.99	0.000
Moderating Variable						
Info x Category				-0.015	-1.444	0.149
Pers x Category				0.002	0.222	0.825
Control Variable / IV						
lnNbacker	0.646***	83.26	0.000	0.646***	83.978	0.000
lnGoalSize	0.140***	20.802	0.000	0.141***	21.357	0.000
Category	0.050***	7.533	0.000	0.059**	4.669	0.000
Staffpick	0.026***	4.067	0.000	0.027***	4.244	0.000
lnWordcount	0.056***	7.649	0.000	0.056***	4.271	0.000
Videocount	0.063***	9.926	0.000	0.062***	10.493	0.000
Imagecount	0.060***	8.211	0.000	0.060***	8.628	0.000
Websitecount	0.024***	3.842	0.000	0.024***	4.111	0.000
Experience	0.014**	2.235	0.025	0.014**	2.090	0.037
BackedHis	-0.013**	-1.995	0.046	-0.013**	-1.973	0.049
lnSharedcount	0.088***	13.735	0.000	0.088***	13.857	0.000
Earlybird	0.022***	3.564	0.000	0.022***	3.620	0.000
NInfoinst	0.026**	3.106	0.002	0.026***	3.162	0.002
NPersinst	-0.064***	-7.241	0.000	-0.065***	-7.371	0.000
p (ANOVA)	0.000			0.000		
Adjusted R^2	0.743			0.743		

Dependent variable: ln(Total Funding)
Category (1 = Art and Culture, 0 = Design and Technology)
*** p < 0.01, ** p < 0.05, * p < 0.1

Model 2.1 shows that both persuasive messages (β_1= 0.218, 95% CI for Exp(β)[2.673, 3.561]) and informative messages (β_2 = 1.217, 95% CI for Exp(β) [1.103, 1.403]) are statistically significant. Because the independent variables were log transformed before running the analysis, one unit difference in messages equals to $\ln(1 + 1)$. Hence, an increase in informative messages by 1 unit corresponds to an increase in success likelihood by 2.48 times ($e^{(0.218+\ln(2))}$). Likewise, an increase in persuasive messages by 1 unit corresponds to an increase in success likelihood by 6.75 times ($e^{(1.217+\ln(2))}$). Similar to model 1.1, model 2.1 also shows that the confidence intervals of two types of messages do not overlap, hence, the effect of persuasive messages is significantly stronger than that of informative messages.

Table 6. Regression Result on Funding Success

	Model 2.1 (main result: funding success)			Model 2.2 (with moderator)		
	Coefficients	**Exp(β)**	**SE**	**Coefficients**	**Exp(β)**	**SE**
Intercept	1.742***	5.709		1.57***	4.804	0.000
Independent Variables						
lnNInfo	0.218***	1.244	0.061	0.307***	1.36	0.075
lnNPers	1.127***	3.085	0.073	1.219***	3.385	0.091
Moderating Variables						
Info x Category				-0.278**	0.757	0.124
Pers x Category				-0.286	0.751	0.148
Control Variables / IVs						
lnNbacker	1.326***	3.767	0.041	1.328***	3.775	0.041
lnGoalSize	-1.229***	0.293	0.043	-1.226***	0.294	0.043
Category	-0.714***	2.042	0.091	1.129***	3.092	0.172
Staffpick	0.571***	1.769	0.124	0.577***	1.781	0.124
lnWordcount	0.252***	1.287	0.071	0.254***	1.289	0.071
Videocount	0.352***	1.422	0.073	0.35***	1.419	0.073
Imagecount	0.002	1.002	0.006	0.001	1.001	0.006
Websitecount	0.027	1.027	0.046	0.027	1.028	0.046
Experience	-0.006	0.994	0.024	-0.005	0.995	0.024
BackedHis	0.003	1.003	0.003	0.003	1.003	0.003
lnSharedcount	0.129***	1.138	0.017	0.126***	1.134	0.017
Earlybird	0.272	1.312	0.141	0.268*	1.307	0.142
NInfoinst	0.923***	2.517	0.201	0.94***	2.559	0.203
NPersinst	-0.305***	0.737	0.047	-0.318***	0.728	0.048
p (ANOVA)	0.000			0.000		
Adjusted R^2 count	0.642			0.640		

Dependent variable: Funding success (1 = Success, 0 = Fail)
Category (1 = Art and Culture, 0 = Design and Technology)
*** p < 0.01, ** p < 0.05, * p < 0.1

Model 2.2 shows the interaction effect of the category on message types. For both categories, in general, goes that persuasive messages are found to be more effective than informative messages. For example, increasing the number of a persuasive message by 1 unit is found to be approximately three times more effective than informative messages on the likelihood of success of campaigns belonging to the ‘Art and Culture’ category. Conversely, persuasive messages are found to be twice more effective than informative messages for campaigns in the ‘Design and Technology’ category. The result shows that emphasizing the funding state and asking for help while promoting social network pages do stimulate excitement among backers and stimulate them to fund more or help in spreading the word about the campaign.

While there is no change in the effect of persuasive messages on funding success for either category (β_4 = -0.286, p > 0.05), the effect of informative messages on the success of a campaign is found to vary for different categories (β_3 = -0.278, p < 0.05). If the number of informative messages increases by 1 unit, it is 2.059 times ($\cong e^{(0.307-0.278\ +\ln(2))}$) more likely for a campaign to succeed for the 'Art and Culture' category, while it is 2.719 times ($\cong e^{(0.307\ +\ln(2))}$) more likely for the 'Design and Technology' category. This means that informative messages are more effective in the 'Design and Technology' category (β = 0.307) than in the 'Art and Culture' category in succeeding (β = 0.029). This supports H3.2. On the other hand, the effect of persuasive messages does not differ between categories as the interaction term was found to be insignificant, rejecting H4.2. The above two findings indicate that those who back categories in the 'Design and Technology' are more stimulated to reach a funding goal by messages related to rewards, stretch goals and campaign progress than in 'Art and Culture' category. The fact that there is no interaction effect of categories on the funding level while there is on the funding success requires an explanation. One reason could be that the effect of informative messages differs by category for those campaigns that have not yet reached their funding goals.

Table 7. Hypothesis Test

Hypothesis	
H1.1 The total funding increases as the number of informative updates issued in the campaign increases.	√
H1.2 The likelihood of success increases as the number of informative updates issued in the campaign increases	√
H2.1 The total funding increases as the number of persuasive updates issued in the campaign increases.	√
H2.2 The likelihood of success increases as the number of persuasive updates issued in the campaign increases.	√
H3.1 The effect of informative messages on total funding is stronger for the design & technology campaigns in comparison to art & culture campaigns.	X
H3.2 The effect of informative messages on the likelihood of success is stronger for the design & technology campaigns in comparison to art & culture campaigns.	√
H4.1 The effect of persuasive messages on total funding is stronger for the art & culture campaigns in comparison to design & technology campaigns.	X
H4.2 The effect of persuasive messages on the likelihood of success is stronger for the art & culture campaigns in comparison to design & technology campaigns.	X

6.4 Evaluation of Instrumental Variables

To avoid the problem of endogeneity, this paper included instrumental variables for the number of informational and persuasive messages as mentioned in equation (3) and (4). In order to evaluate relevant instrumental variables, the number of informative and persuasive messages were regressed on other control variables and instrumental variables. The report in Appendix G shows that the two instrumental variables – the average number of informative and persuasive messages – are significantly correlated with the independent variable. Therefore, the use of instrumental variables is relevant in this model.

6.5 Robustness check

6.5.1 Distinction between Informative and Persuasive Messages

The empirical results in the previous sections show the different impact of each message type on the total funding amount and the funding success. Persuasive messages have a much stronger relationship with both the total funding and the success compared to informative messages regardless of the category. Moreover, the moderating variable of category indicates that informative messages alone varied in their effects on funding success depending on which category of the project they provided updates for, as they showed a stronger effect in Design and Technology category than in Art and Culture category. If this distinction is reasonable, the statistical fit should increase for the model that classified the messages into two rather than the model that uses generic update numbers. This was tested using the Akaike Information Criterion. As expected, AIC was lower (24282.41) for model 1 used in this paper than that of the generic model (24541.93). Likewise, AIC was lower (4317.16) for model 2 than that of the generic model (4499.98). This shows that the statistical fit is better for this model than the generic model.

6.5.2 Analysis on Stretch Goals

In order to carry out a more in-depth analysis, this paper divided informative messages into two different types: those that are related to stretch goals, and those that are not. Descriptive statistics for the stretch goals and the rest of the informative messages are reported in Appendix H.1 and the regression results in Appendix H.2. This division reveals no change in fit (R^2), although the number of variables increased. Regarding the variable coefficients, all three types of messages are found to be positively correlated with the total funding amount with their respective effectiveness varying in the descending order of persuasive (β = 0.087, 95% CI [0.075, 0.100]), stretch-goal related (β = 0.027, 95% CI [0.014, 0.040]) and rest of informative messages (β = 0.016, 95% CI [0.003, 0.029]). To test whether there is a difference between coefficients, the author again used the 50% overlap rule by Cumming (2009). Persuasive messages do not overlap with other two coefficients, meaning it is significantly more effective on the total funding level than rest of the two. However, confidence intervals of stretch-goal related informative messages and the rest of informative messages overlap, hence calculation becomes necessary. The first half of the average of the overlapping confidence intervals is calculated (0.00975) and then added to the lower bound estimate of stretch-goal related messages (0.014), which yielded 0.02375. As the upper bound limit of the rest of informative messages (0.029) exceeds 0.02375, the difference between the standardized beta weights of two types of messages (Δβ = 0.011) is not considered significantly different. In other words, the effect of the messages related to stretch goals and the rest of informative messages on the total funding amount does not differ.

However, the effect of the message's contents on the likelihood of a campaign's funding success shows a different result. Whereas persuasive messages (β = 1.109, Exp(β) = 3.03, 95% CI [2.596, 3.466]) and stretch-goal-related informative messages (β = 3.426, Exp(β) = 3.755, 95% CI [15.698, 78.769]) have a significant effect on the success of a campaign, the rest of the informative messages do not. The confidence intervals between the beta coefficient of persuasive messages and stretch-goal-related messages do not overlap and are hence significantly different from each other. This is certainly because stretch goals are the goals set by those who aim to get overfunded, meaning successful campaigns will certainly mention stretch goals rather than those who have not. There exists no interaction between categories and messages for both models. However, it is surprising that the effect of stretch goal messages is found to be much more significant than that of persuasive messages when it comes to the likelihood of success. As stretch goal related messages only consist of 10% of the total informative messages and yet have a significant impact on the likelihood of success while the rest of 90% do not, implementing stretch goals for marketing purposes is meaningful.

6.5.3 Split-Half Analyses

To provide further insight into the results of the regression analyses, the author used a bootstrapping method on half of the randomly chosen sample to carry out a robustness check. If the conceptual framework and the model used in this paper are stable, a random draw of samples from the original dataset should be consistent with the results from the half. As reported in Appendix I.1, the result from the half of the dataset is consistent with the regression result from the entire dataset for both model 1 and 2.

This paper also used the parameter derived from half of the sample to predict the total funding of the other half of the dataset. The absolute mean difference of the predicted total funding from the real total funding was 2562.36. However, as reported in Appendix I.2, a paired-sample t-test shows no significant mean difference between the two ($p = 0.456$). Therefore, the model 1 is reasonable as a prediction model. However, for the model 2, the model based on half of the sample could only correctly predict 60.93% of the second half of the data. Hence, the model 2 is not reasonable as a prediction model.

6.5.4 Exclusion of Outliers

Here, we check the result of the model without the outliers, as suggested by 'Casewise diagnostics- outliers outside 2 standard deviations' using IBM SPSS Statistics 24. The diagnostics indicated the existence of 399 campaigns out of 7428 campaigns in which the difference between the actual and the predicted value of the total funding level does not lie in a range of two standard deviations of the mean residual. This left 7033 campaigns for analysis. The result from the hypothesis in the original model is highly consistent with the model without outliers, as shown in Appendix J. Therefore, it is not justifiable to remove the outliers from the original model.

6.5.5 Reduction in the Sample

Before the main analysis, the author discarded 1,849 updates that were posted on the last day of the campaign as many of them were found to be posted after the campaign had finished. In order to carry out the robustness check, the two models are run with 37,021 update posts including the discarded data. The result is reported in Appendix K. While the effect of informative messages and persuasive messages are both significant on both of the main models, the effects on total funding are no longer significant when the discarded data is included. Likewise, the effects of informative messages on funding success became insignificant when the discarded data is added. Hence, the influence of the last day data is considerable, validating the removal of the updates posted on the last day of the campaign.

6.6 Impact of Other Control Variables

This section will report the results of the model's control variables in order to gain a more in-depth understanding of the significance of the above findings. The author has recorded standardized coefficients in Table 5 so as to compare the effect of control variables. In their effects on the total funding amount, all twelve control variables are found to be highly significant. First, the number of backers was found to have the strong positive effects on the total funding levels (δ_1 = 0.651, 95% CI [0.631, 0.661]). This is not surprising as the total funding consists of the accumulation of individual backers' pledges[10] and the number of backers is found to be quality signals for future backers. The second highest significant control variable was the goal size (δ_2= 0.144, 95% CI [0.127, 0.153]), followed by the number of shares online (δ_{11}= 0.089, 95% CI [0.075, 0.100]). In comparison to shares online, staff-pick was found to have a lower positive impact on the total funding (δ_4= 0.028, 95% CI [0.014, 0.039]), suggesting that backers access the campaigns directly through the link rather than through the Kickstarter main website. Furthermore, the campaigns in Art and Culture category is found to be more positively correlated with the total funding level compared to the ones in Design and Technology (δ_3= 0.050, 95% CI [0.037, 0.063]) even though the number of campaigns and the total funding levels in the sample was higher for the latter. The project design was also found to be positively correlated with the total funding. Because the confidence levels overlap, the number of videos and images, as well as the number of words on project descriptions, are equally effective on total funding (δ_6= 0.066, δ_7= 0.065, δ_5= 0.056; 95% CIs [0.050, 0.075], [0.045, 0.075], [0.042, 0.070]). This suggests that visual materials on the campaign page are an as effective way of communicating with the backers as explaining in words. As expected, the number of campaigns that the creator launched previously is also positively related to the total funding (δ_9= 0.013, 95% CI [0.002, 0.027]) with the help of prior brand exposure, but much less significance compared to other factors. Surprisingly, the result of a project creator's backing history has been found to be negatively related to the total funding (δ_{10}= - 0.013, 95% CI [-0.025, 0.000]) in contrast to the result from Zvilichovsky et al. (2015). Our finding signifies that there may not be any reciprocity on crowdfunding platforms. Lastly,

[10] In Kickstarter, you can only pledge to one reward.

the existence of an early bird reward is positively related to total funding (δ_{12}= 0.022, 95% CI [0.010, 0.033]).

In the covariate effects on the funding success, only seven out of twelve are found to be significant. Amongst seven, five show a positive effect on the success of the campaign; these are the number of backers ($\delta_1 =$ 0.651), whether the campaign was staff-picked or not ($\delta_4 =$ 0.571), the number of words ($\delta_5 =$ 0.252), the number of videos ($\delta_6 =$ 0.352), and the number of shares ($\delta_{11} =$ 0.129). However, the goal size ($\delta_2 =$ -1.229) reduces the likelihood of success, indicating that unrealistic goal size should be avoided. Also, in contrast to the first model, the campaigns in Art and Culture category are negatively correlated with the likelihood of success compared to the ones in Design and Technology (δ_3 = -0.714) although the proportion of success is higher for the former category.

VII. Conclusion

7.1. General Discussion

Crowdfunding is a novel way to solicit funds for a wide variety of projects from an unknown large group of people – the crowd. Given the relative dearth of material on crowdfunding in the academia, this paper offers new insights into those success factors that influence the crowdfunding campaigns' total funding amounts *during* the time that the campaign is running. It does this specifically by investigating the effects of in-process promotional messages in a form of update posts on campaigns launched on Kickstarter. In this paper, the author makes a first step towards quantitatively examining the marketing strategies *during* the campaign that creators can implement to increase the number of pledges from backers. In distinguishing the types of promotional messages, the author divided these into informative and persuasive messages following the framework by Ducarroz et al. (2016).

For this, this paper analyzed 7,428 projects on the crowdfunding platform Kickstarter and collected 34,920 update posts. It controlled twelve different variables by data extracted from finished crowdfunding projects with a fixed duration of 30 days. This paper introduced a text mining process that enabled classification of update messages into two different types. This paper also posits that distinguishing between two different types of strategies help increase the statistical fit of the regression model used for analysis.

The findings generated useful insights for campaign creators. Both informative and persuasive messages have positive effects on the total funding levels and the success of a campaign. Generally speaking, persuasive messages are found to be more effective than informative messages in either category. In terms of relative effect, however, informative messages are more effective to elicit support from backers that support 'Design and Technology' projects than from those that support 'Art and Culture' projects in reaching the funding goal. The author also demonstrated the predictive power of the model 1 by using a split-half analysis. Next, this paper suggests how these findings should be of interest for crowdfunding creators and intermediaries. Then, this paper compares the results with the result of the auction setting introduced in the paper of Ducarroz et al. (2016). Lastly, this paper discusses its limitations and potential directions for future research.

7.2 Managerial Implications

For creators who seek to launch a campaign in Kickstarter, this paper introduces two message strategies for update posts, subdivided into six types of messages depending on the topic of the message. This classification provides a basis for creators which updates to use for which type of campaigns. There are some tips that creators can take. Firstly, posting updates is important for raising funds and reaching the funding goal. Secondly, the findings in this paper indicate that persuasive messages are more effective than informative messages. Creators should, therefore, should choose their update posts wisely by making more use of the persuasive types of messages than the informative messages.

This paper shows that the interaction between creators and backers are related to the funding amount and the success of the campaign. Currently, there are only two ways of doing so on Kickstarter platform: Updates and comments. Backers can react to updates by posting comments and liking comments, while creators can answer questions from backers through comments. However, because of such simplicity, creators have to use external websites for more complex interaction such as voting for new rewards or stretch goals. The intermediaries can, therefore, develop tools for more interaction such as voting, feedback or rating systems that help to solicit more funding for creators and to provide clearer quality signals for backers. Furthermore, stretch goals related update posts were found significantly related to total funding and the funding success. This means that many creators add "stretch goals" which is an unofficial system in Kickstarter. The intermediary such as Kickstarter can, thereby, facilitate this by building a tool that shows these stretch goals on the front page of the campaign to make it more noticeable for potential backers.

7.3 Comparison with the Auction Setting

Although the aim of the paper is not to find similarities between the impact of in-process promotional messages in the auction and in crowdfunding settings, comparing these two settings provides additional insight into the generalizability of the proposed framework by Ducarroz et al. (2016). The result from the crowdfunding setting turned out to be opposite to the result in the auction setting. Ducarroz et al, in their paper, found that informative messages have an impact on the final auction price. However, persuasive messages were found insignificant. On the other hand, our paper shows a strong significant effect of persuasive messages on the total funding level and much less effect of informative messages. Although the two settings seem comparable in a sense that they are both intermediaries sourcing the funds from the unknown crowd in an online environment, there are factors that might explain the difference between the two research results. For instance, there is no helping behavior involved in bidders in an auction setting because auction participants are in a competition with each other with only one person ending up with a reward. Rather, crowdfunding requires backers to work together to reach the funding goal in order to be able to obtain their rewards. Furthermore, although there is a limit to certain rewards, everyone can get a reward as they are provided with a range of rewards to choose from.

7.4 Limitations and Directions for Future Research

In this paper, the author classified update posts into informative and persuasive messages. However, as discussed earlier, this proved to be difficult as not all messages were completely persuasive or informative. Some messages could equally be both informative and persuasive in their contents. This ambiguity could have been alleviated by using a continuous scale when carrying out the classification, but it would have remained nonetheless difficult to classify them in different weights of each type of messages using only term frequency. Furthermore, the online setting allowed in-process messages that are interactive, such as asking backers' advice on improving the product without the goal of selling the product to the audience. These kinds of promotional messages can even be seen on television nowadays (e.g., "use #hashtag on Facebook and win a free cup of coffee!"). Hence, researchers conducting future studies may wish to develop a new advertising typology updated to fit modern advertising messages.

This research only covers crowdfunding efforts on the Kickstarter platform, where they use an all-or-nothing system. Another reward-based crowdfunding platform, Indiegogo, offers a different keep-it-all system. These differences among crowdfunding platforms including donation-based and equity-based may alter the effects of in-process promotional messages. Therefore, researchers may wish to use this framework in different crowdfunding platforms. On top of that, Kickstarter is consistently improving its platform to include new features and sections. For instance, Kickstarter started a new section in June 2017 called 'Kickstarter Gold', which exclusively showcases the new projects by selected creators based on past success and innovation. Also, third parties such as the author are not able to capture the effect of the customized project recommendation system offered by Kickstarter for each individual backer. Such new features may result in different effects of identified success factors over time.

The aggregate effect of the messages is captured in this paper but fails to show a micro-view on their daily effects. In other words, it does not capture how messages affect or are affected during the funding phase on a daily basis. As campaigns have a specific duration, time-series data would provide a more thorough insight as the potential increase in backers is affected by the number of backers that have already funded the project[11]. To be specific, using daily panel data – indicating the daily incremental increase in funding for campaigns - would have allowed room to analyze the likelihood of gaining increased funding when posting specific types of messages. It would have also allowed room to analyze the impact of messages on social media on a given day depending on the type of messages. Also, this micro-level of analysis allows for the shift from backers to creators in analyzing the behavior of creators to find out the impact of funding rate on the posting behavior of the creators. Moreover, it may also allow the analysis of the impact of each type of message moderated by funding percentage. For instance, persuasive messages may be more effective near the deadline, or right before campaigns reach their goal. Lastly, whether lacking updates are related to the instances of withdrawal of backers during the funding can also be examined.

Although informative and persuasive messages are found to have a positive relationship with the total funding level and the funding success, it does necessarily infer a causal relationship between the two. The performance of the campaign could depend on the point of time the messages are sent. For example, persuasive messages may have a positive relationship with the total funding just because many are sent close to the deadline to stimulate backers to reach the goal. Sponsors are motivated by the target goal because later-stage contributions are generally believed to have a greater impact than early-stage contributions (Higgins and Brendl, 1995; Toure-Tillery and Fishbach 2011). In such cases, backers will be stimulated regardless of the types of messages, hence, invalidating the distinction between two types of messages.

This research shows that persuasive messages are positively related to a campaign's total funding amount. However, the effect is expected to diminish at a certain point as people do not like to be overtly confronted by the same type of messages and might perceive them as spam.

[11] The author does not provide time series analysis due to missing daily data. At the time of data collection, it was technically not possible to get time series data on the past campaigns as it is illegal to scrape the data from the private crowdfunding tracking websites.

Future research may wish to examine the optimal number of promotional messages for different types of promotions. The optimal number and the balance between the two types of message can then also be applied to other contexts such as email marketing, in-app push notifications, emails, Twitter and Facebook.

Reference

Aaker, A., & Norris, D. (1982) "Characteristics of TV commercials perceived as informative." Journal of Advertising Research, Vol 22(2), Apr-May 1982, pp. 61-70.

Agarwal, S., Godbole, S., Punjani, D., & Roy, S. (2007) How Much Noise Is Too Much: A Study in Automatic Text Classification, Proc. IEEE Seventh Int'l Conf. Data Mining, pp. 3-12.

Agrawal, A., Catalini, C., & Goldfarb, A. (2015). Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions. Journal Of Economics & Management Strategy, 24(2), 253-274. http://dx.doi.org/10.1111/jems.12093

Asch, V. (2013). Macro- and micro-averaged evaluation measure. http://www.cnts.ua.ac.be/~vincent/pdf/microaverage.pdf.

American Jobs Act: Read All the Details. (2017). whitehouse.gov. Retrieved 30 December 2016, from https://obamawhitehouse.archives.gov/blog/2011/09/09/american-jobs-act-read-all-details

Baeza-Yates, R., Ribeiro-Neto, B. (2011). Modern information retrieval: The concepts and technology behind search. Addison Wesley, New York, 2 edition.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. Journal Of Business Venturing, 29(5), 585-609. http://dx.doi.org/10.1016/j.jbusvent.2013.07.003

Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. Journal Of Business Research, 71, 10-18. http://dx.doi.org/10.1016/j.jbusres.2016.10.001

Bowden, R.J., & Turkington, D.A. (1984). Instrumental Variables. Cambridge, England: Cambridge University Press.

Boyle, K. (2013). Yes, Kickstarter raises more money for artists than the NEA. Here's why that's not really surprising. [online] Washington Post. Available at: https://www.washingtonpost.com/news/wonk/wp/2013/07/07/yes-kickstarter-raises-more-money-for-artists-than-the-nea-heres-why-thats-not-really-surprising/?utm_term=.c00fbf952ab7

Burtch, G., Ghose, A., & Wattal, S. (2011). An Empirical Examination of the Antecedents and Consequences of Investment Patterns in Crowd-Funded Markets. SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.1928168

Cacioppo, John T., & Richard E. Petty. (1979). “Effects of Message Repetition and Position on Cognitive Responses, Recall, and Persuasion,” Journal of Personality and Social Psychology, 37 (1), 97–109.

Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. Journal of Abnormal and Social Psychology

Claude, E.S., & Weaver, W. (1949). "The mathematical theory of communication (Urbana, IL." University of Illinois Press 19.7

Cohen J. (1960). A coefficient of agreement for nominal scales. Educ Psychol Meas. 20:37–46.

Cowles, D., & Crosby, L. (1990). Consumer Acceptance of Interactive Media in Service Marketing Encounters. The Service Industries Journal, 10(3), 521-540. http://dx.doi.org/10.1080/02642069000000056

Colombo, M.G., Franzoni, C. & Rossi-Lamastra, C., 2015. Internal Social Capital and the Attraction of Early Contributions in Crowdfunding. Entrepreneurship Theory and Practice, 39(1), pp.75–100.

Cryder, C., Loewenstein, G., & Seltman, H. (2013). Goal gradient in helping behavior. Journal Of Experimental Social Psychology, 49(6), 1078-1083. http://dx.doi.org/10.1016/j.jesp.2013.07.003

Cumming, G. (2009). Inference by eye: reading the overlap of independent confidence intervals. Statistics in Medicine, 28(2), 205-220.

Cumming, D., Leboeuf, G., & Schwienbacher, A. (2015). Crowdfunding Models: Keep-it-All vs. All-or-Nothing. SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.2447567

Lewis, D., & Gale., W (1994). A sequential algorithm for training text classifiers. In SIGIR-94, 1994.

De Buysere, K., Gajda, O., Kleverlaan, R., Marom, D., & Klaes, M. (2012). A framework for European crowdfunding. European Crowdfunding Network.

Debole, F., & Sebastiani, F. (2003). Supervised term weighting for automated text categorization. In Proceedings of SAC-03, 18th ACM Symposium on Applied Computing, pages 784–788, Melbourne, US. ACM. Press, New York, US

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. New York: Plenum

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29.

Ducarroz, C., Yang, S., & Greenleaf, E. (2016). Understanding the Impact of In-Process Promotional Messages: An Application to Online Auctions. Journal Of Marketing, 80(2), 80-100. http://dx.doi.org/10.1509/jm.13.0190

Dredge, S. (2014). *Kickstarter's biggest hits - why crowdfunding now sets the trends*. [online] the Guardian. Available at: https://www.theguardian.com/technology/2014/apr/17/kickstarter-crowdfunding-technology-film-games.

Esuli, A. and Sebastiani, F. (2013). Improving Text Classification Accuracy by Training Label Cleaning. ACM Transactions on Information Systems, 31(4), pp.1-28.

Farris, P.W., & Albion, M. (1981). Determinants of the Advertising-to-Sales Ratio. Journal of Advertising Research

Flanagin, A., Metzger, M., Pure, R., Markov, A., & Hartsell, E. (2014). Mitigating risk in e-commerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention. Electronic Commerce Research, 14(1),1-23.

Fleming, L., & Sorenson, O. (2016). Financing by and for the Masses: An Introduction to the Special Issue on Crowdfunding. California Management Review, 58(2), 5-19. http://dx.doi.org/10.1525/cmr.2016.58.2.5

Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing, 70, 627−633.

Forman, G. (2007). Feature selection for text classification. Computational Methods of Feature Selection, Chapman and Hall/CRC Press.

Gefen, D., & Straub, D. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. Omega, 32(6), 407-424. http://dx.doi.org/10.1016/j.omega.2004.01.006

Gerber, E., & Hui, J. (2013). Crowdfunding: Motivations and Deterrents for Participation. ACM Transactions On Computer-Human Interaction, 20(6), 1-32. http://dx.doi.org/10.1145/2530540

Grant, A., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. Journal Of Personality And Social Psychology, 98(6), 946-955. http://dx.doi.org/10.1037/a0017935

H. Akaike. (1973). Information theory as an extension of the maximum likelihood principle B.N. Petrov, F. Csaki (Eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest (1973), pp. 267-281

Haas, P., Blohm, I., & Leimeister, J. M. (2014). “An empirical taxonomy of crowdfunding intermediaries,” in Proceedings of the 35th International Conference on Information Systems (ICIS 2014), Auckland, New Zealand.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The Weka data mining software: An update. ACM SIGKDD Explorations Newsletter, 11, 10–18.

Hanley J.A., McNeil B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 143:29–36.

Hawkins, D. (2004). The problem of overfitting. Cheminform, 35(19). doi:10.1002/chin.200419274

Herzenstein, M., Sonenshein, S., & Dholakia, U. (2011). Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-Peer Lending Decisions. Journal Of Marketing Research, 48(SPL), S138-S149. http://dx.doi.org/10.1509/jmkr.48.spl.s138

Howe, J. (2008). Crowdsourcing: why the power of the crowd is driving the future of business.

Hu, M., Li, X., & Shi, M. (2015). Product and Pricing Decisions in Crowdfunding. Marketing Science, 34(3), 331-345. http://dx.doi.org/10.1287/mksc.2014.0900

Hunt, S.D. (1976) The Nature and Scope of Marketing. Journal of Marketing 40 (July): 17-28. Google Scholar CrossRef.

Heyer, G. (2009) Introduction to TMS. In Heyer, Gerhard, editor, Text mining services, volume 14 of Leipziger Beitr¨age zur Informatik, pages 1–14, Leipzig, 2009. LIV.

Ikonomakis, M., Kotsiantis, S., & Tampakas, V., (2005) Text classification using machine learning techniques. WSEAS transactions on computers 4.8 (2005): 966-974.

Janssens, W., Wijnen, K., De Pelsmacker, P. and Van Kenhove, P. (2008). Marketing research with SPSS. Harlow: Prentice Hall/Financial Times.

Jivani, A.G. (2011). A Comparative Study of Stemming Algorithms, International Journal of Computer, Technology and Application, Volume 2, ISSN:2229-6093.

Joachims, T. (2001). Learning to classify text using support vector machines. Kluwer Academic Publishers.

Kelle, Udo. (1997) Theory building in qualitative research and computer programs for the management of textual data. Sociological Research Online, 2(2), URL http://www.socresonline.org.uk/2/2/1.html.

Kickstarter.com. (2017). Kickstarter Stats — Kickstarter. [online] Available at: https://www.kickstarter.com/help/stats.

Klein, B., & Leffler, K. (1981). The Role of Market Forces in Assuring Contractual Performance. Journal Of Political Economy, 89(4), 615-641. http://dx.doi.org/10.1086/260996

Ko, Y., & Seo, J. (2000). Automatic text categorization by unsupervised learning. In Proceedings of the 18th international conference on computational linguistics (COLING'2000) (pp. 453–459).

Ko, Y., & Seo, J. (2009). Text classification from unlabeled documents with bootstrapping and feature projection techniques. Information Processing & Management, 45(1), 70-83. http://dx.doi.org/10.1016/j.ipm.2008.07.004

Kohler, T., Fueller, J., Stieger, D., & Matzler, K. (2011). Avatar-based innovation: Consequences of the virtual co-creation experience. Computers In Human Behavior, 27(1), 160-168. http://dx.doi.org/10.1016/j.chb.2010.07.019

Koch, J.-A., and M. Siering. (2015) "Crowdfunding Success Factors: The Characteristics of Successfully Funded Projects on Crowdfunding Platforms", in Proceedings of the 23rd European Conference on Information Systems, AIS, Münster, Germany.

Koh, W., & Leung, H. (1992). Persuasive Advertising and Market Competition. The American Economist, 36(2), 39-49. http://dx.doi.org/10.1177/056943459203600205

Krippendorff, K. (2013). Content analysis: An introduction to its methodology (3 ed. ed.). Los Angeles, Calif.: Sage.

Kumar, A. & Sebastian, T. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. International Journal of Intelligent Systems and Applications, 4(10), pp.1-14.

Kuppuswamy, V., & Bayus, B. (2013). Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.2234765

Kuppuswamy, V., & Bayus, B. (2017). Does my contribution to your crowdfunding project matter?. Journal Of Business Venturing, 32(1), 72-89. http://dx.doi.org/10.1016/j.jbusvent.2016.10.004

Kunz, M., Bretschneider, U., Erler, M., & Leimeister, J. (2016). An empirical investigation of signaling in reward-based crowdfunding. Electronic Commerce Research. http://dx.doi.org/10.1007/s10660-016-9249-0

Lakhani, K., & von Hippel, E. (2003). How open source software works: “free” user-to-user assistance. Research Policy, 32(6),923-943. http://dx.doi.org/10.1016/s0048-7333(02)00095-1

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

Laskey, H., Day, E., & Crask, M. (1989). Typology of Main Message Strategies for Television Commercials. Journal Of Advertising, 18(1), 36-41. http://dx.doi.org/10.1080/00913367.1989.10673141

Lawton, K., & Marom, D. (2012). The Crowdfunding Revolution: How to Raise Venture Capital Using Social Media. McGraw-Hill Professional.

Leimeister, J., Huber, M., Bretschneider, U., & Krcmar, H. (2009). Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition. Journal Of Management Information Systems, 26(1), 197-224. http://dx.doi.org/10.2753/mis0742-1222260108

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression!. Behavior and Information Technology 25 (2), 115–126.

Lu. S., Xie. X., Kong, & P.S. Yu. (2014). Inferring the impacts of social media on crowdfunding Proceedings of the 7th ACM international conference on Web search and data mining, pp. 573-582

Massolution. (2015). 2015CF: The Crowdfunding Industry Report available at http://www.crowdsourcing.org/

McCallum A. & Nigam K. (1998). A Comparison of Event Models for Naive Bayes Text Classification.https://www.researchgate.net/publication/2408220_A_Comparison_of_Event_Models_for_Naive_Bayes_Text_Classification

McCallum, A., Nigam, K., Rennie, J., & Seymore, K. (1999). A machine learning approach to building domain-specific search engines. In Proceedings of the sixteenth international joint conference on artificial intelligence (IJCAI-99).

McCallum A., Nigam K., Thrun S., & Mitchell T. (2000). Learning to classify text from labeled and unlabeled documents. In AAAI-98.

Meer, J. (2014). Effects of the price of charitable giving: Evidence from an online crowdfunding platform. Journal Of Economic Behavior & Organization, 103, 113-124. http://dx.doi.org/10.1016/j.jebo.2014.04.010

Miguel Villas-Boas & Russell S. Winer. (1999). "Endogeneity in Brand Choice Models," Management Science, INFORMS, vol. 45(10), pages 1324-1338, October.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal Of Business Venturing, 29(1), 1-16. http://dx.doi.org/10.1016/j.jbusvent.2013.06.005

Nasukawa, T., Punjani, D. Roy, S., Subramaniam, L.V., & Takeuchi, H. (2007). Adding sentence boundaries to conversational speech transcriptions using noisily labeled examples. In Proc. of AND07 Workshop in conjunction with IJCAI.

Ordanini, A., Miceli, L., Pizzetti, M., & Parasuraman, A. (2011). Crowd-funding: transforming customers into investors through innovative service platforms. Journal Of Service Management, 22(4), 443-470. http://dx.doi.org/10.1108/09564231111155079

Palmer, J. (2002). Web Site Usability, Design, and Performance Metrics. Information Systems Research, 13(2), 151-167. http://dx.doi.org/10.1287/isre.13.2.151.88

Porter M.F. (1980). An algorithm for suffix stripping, Program. 14, 130-137.

Porter M.F. (2001). Snowball: A language for stemming algorithms.

Powers, D.M.W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technology 2 (1), 37-63

Rajaraman, A.; Ullman, J. D. (2011). Data Mining. Mining of Massive Datasets (PDF). pp. 1–17. ISBN 978-1-139-05845-2.

Rao, A.G. & Miller, P.B. (1975). Advertising/Sales Response Functions. Journal of Advertising Research, 15 (April), 7–15.

Rice M.E., Harris GT (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum Behav 29: 615-620

Russell, S., & Norvig, P. (2003) [1995]. Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.

Ryan, R., & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology, 25(1), 54-67. http://dx.doi.org/10.1006/ceps.1999.1020

Ryu, S., & Kim, Y. (2016). A typology of crowdfunding sponsors: Birds of a feather flock together?. Electronic Commerce Research And Applications, 16, 43-54. http://dx.doi.org/10.1016/j.elerap.2016.01.006

Salton, G., & MacGill, M. (1983). Introduction to modern information retrieval (International ed. ed.). Auckland: McGraw-Hill.

Shane, S., & Cable, D. (2002). Network Ties, Reputation, and the Financing of New Ventures. Management Science, 48(3), 364-381. http://dx.doi.org/10.1287/mnsc.48.3.364.7731

Simons, A., Weinmann, M., Tietz, M., & vom Brocke, J. (2017) Which Reward Should I Choose? Preliminary Evidence for the Middle-Option Bias in Reward-Based Crowdfunding. In: Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS), Big Island, HI.

Steinberg, S., DeMaria, R., & Kimmich, J. (2012). The crowdfunding bible. [United States]: Read.me.

Stock, A., & Balachander, S. (2005). The Making of a “Hot Product”: A Signaling Explanation of Marketers’ Scarcity Strategy. Management Science, 51(8), 1181-1192. http://dx.doi.org/10.1287/mnsc.1050.0381

Silva, C. & Ribeiro, B. (2003). The importance of stop word removal on recall values in text categorization. In 816 Neural Networks, 2003. Proceedings of the International Joint Conference on, volume 3, pages 1661–1666. IEEE.

Tape Thomas G. The Area Under an AUC Curve. University of Nebraska Medical Center. http://gim.unmc.edu/dxtests/Default.htm

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of data clusters via the gap statistic. Journal of the Royal Statistical Society B, 63:411–423

Yang, S., Chen, Y., & Allenby, G. (2003). Bayesian Analysis of Simultaneous Demand and Supply. Quantitative Marketing And Economics, 1(3), 251-275. http://dx.doi.org/10.1023/b:qmec.0000003327.55605.26

Yang, Y., Slattery, S., & Ghani, R. (2002). A study of approaches to hypertext categorization. Journal of Intelligent Information Systems, 18(2).

Yin, Chunxiao, Mirkovski, Kristijan, Liu, Ivy Libo, Lim, Kai H., & von Briel, Frederik (2017) An empirical investigation on the impact of crowd participation on the degree of project success: The perspective of crowd capital. In Proceedings of the 50th Hawaii International Conference on System Sciences, IEEE, Waikoloa Village, Hawaii, pp. 14-23.

Tellis, Gerard J. (1988). Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice. Journal of Marketing Research, 25 (May), 134–44.

Tryon, Robert C. (1939). Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality. Edwards Brothers.

Vaughn, Richard. (1980). How Advertising Works: A Planning Model. Journal of Advertising Research, Vol. 20

Vichi. M. (1985). "On a flexible and computationally feasible divisive clustering technique" Rivista di Statistica Applicata, 18:199–208.

Vroom, V. (1964). Work and Motivation. Wiley, New York.

Wang, G., Liu, X., & Fan, W. (2011). A knowledge adoption model based framework for finding helpful user-generated contents in online communities. Proceedings of 2011 International conference on information systemems. Retrieved from http://aisel.aisnet.org/icis2011/proceedings/knowledge/15/.

Wiedemann, G. (2016). Text Mining for Qualitative Data Analysis in the Social Sciences (1st ed.)

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A. & Spasić, I. (2015). The role of idioms in sentiment analysis. Expert Systems with Applications, 42(21), pp.7375-7385.

Xiao, S., Tan, X., Dong, M. and J. Qi. (2014). How to Design Your Project in the Online Crowdfunding Market? Evidence from Kickstarter. In: Proceedings of the 35th International Conference on Information Systems. Auckland: Australia.

Zheng, H., Li, D., Wu, J., & Xu, Y. (2014). The role of multidimensional social capital in crowdfunding: A comparative study in China and US. Information & Management, 51(4), 488-496. http://dx.doi.org/10.1016/j.im.

Zhou, M., Du, Q., Zhang, X., Qiao, Z., Wang, G., & Fan, W. (2015). Money talks: A predictive model of crowdfunding success using project description. AMCIS 2015

Zhou, M.., Lu, B., & Fan, W. (2016). Project description and crowdfunding success: an exploratory study. Inf Syst Front. doi:10.1007/s10796-016-9723-1

Zhuang, X., Engel, B. A., Lozano-Garcia, D. F., Fernandez, R. N., & Johannsen, C. J. (1994). Optimisation of training data required for neuro-classification. International Journal of Remote Sensing, 15, 3271−3277.

Zipkin, N. and more, R. (2015). *The 10 Most Funded Kickstarter Campaigns Ever*. [online] Entrepreneur. Available at: https://www.entrepreneur.com/article/235313.

Zvilichovsky, D., Inbar, Y., & Barzilay, O. (2015). Playing Both Sides of the Market: Success and Reciprocity on Crowdfunding Platforms. SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.2304101

Appendix

Appendix A. Sub-categories

Category	Sub Categories	Examples
Art & Culture	Art	Ceramics, Conceptual Art, Digital Art, Illustration, Installations, Mixed Media, Painting, Performance Art, Public Art, Sculpture, Textiles, Video Art
	Theater	Comedy, Experimental, Festivals, Immersive, Musical, Plays, Spaces
	Dance	Performances, Residencies, Spaces, Workshops
	Music	Blues, Chiptune, Classical Music, Comedy, Country & Folk, Electronic Music, Faith, Hip-Hop, Indie Rock, Jazz, Kids, Latin, Metal, Pop, Punk, R&B, Rock, World Music
	Photography	Animals, Fine Art, Nature, People, Photobooks, Places
	Film & Video	Action, Animation, Comedy, Documentary, Drama, Experimental, Family, Fantasy, Festivals, Horror, Movie Theaters, Music Videos, Narrative, Film, Romance, Science Fiction, Shorts, Television, Thrillers, Webseries
	Journalism	Audio, Photo, Print, Video, Web
Design & Technology	Technology	3D Printing, Apps, Camera Equipment, DIY Electronics, Fabrication Tools, Flight, Gadgets, Hardware, Makerspaces, Robots, Software, Sound, Space Exploration, Wearables, Web
	Games	Gaming Hardware, Live Games, Mobile Games, Playing Cards, Puzzles, Tabletop Games, Video Games
	Comics	Anthologies, Comic books, Events, Graphic Novels, Webcomics
	Publishing	Academic, Anthologies, Art Books, Calendars, Children's Books, Comedy, Fiction, Letterpress, Literary Journals, Nonfiction, Periodicals, Poetry, Radio & Podcasts, Translations, Young Adult, Zines, Literary Spaces
	Fashion	Accessories, Apparel, Childrenwear, Couture, Footware, Jewelry, Pet Fashion, Ready-to-wear
	Design	Architecture, Civic Design, Graphic Design, Interactiv Design, Product Design, Typography
	Crafts	Candles, Crochet, DIY, Embroidery, Glass, Knitting, Pottery, Printing, Quilts, Stationery, Taxidermy, Weaving, Woodworking
	Food	Bacon, Community Gardens, Cookbooks, Drinks, Events, Farmer's Markets, Farms, Food Trucks, Restaurants, Small Batch, Spaces, Vegan

Appendix B. Update Titles and Contents

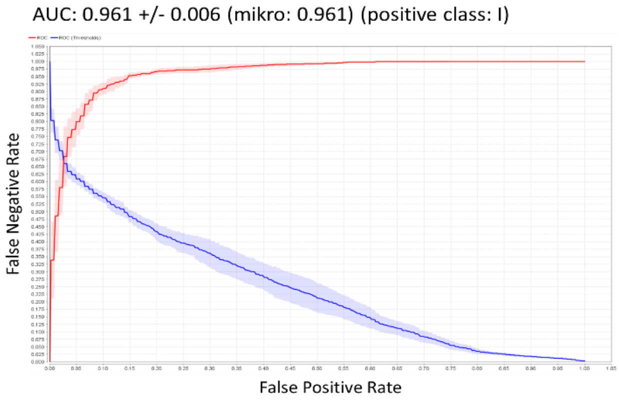
Update Title	Update Contents
Informative Messages	
Progress report	We are undeterred by the modest support to our project so far. If we do not meet our target we will still be going ahead with the magazine - it may just have to be digital-only to begin with.We are currently building a dedicated website for the magazine and will make announcement once it is ready. The website will have more information about what kinds of articles we are looking for, as well as advertising guidelines. In the meantime, keep an eye on our main website: www.nzcrafthub.co.nz for news. Regards, Skye
The Washington Post features The Story House!	Check out this great feature story in the Washington Post about the Story House. It was supposed to run last week, but got bumped by Brexit. It was a real pleasure to be interviewed by journalist Abha Bhattara; she spent a lot of time talking with me and getting to know the project. She writes a column about crowdfunding campaigns. It is very interesting to see the different projects that people create. The campaign is running slow and steady. We have 70 backers , $9,078 raised (which is 45%) with 16 days to go.
2 New Reward Tiers and Stretch Goals UNLOCKED!	You guys have made us so happy. Here are 2 new rewards tiers that you have unlocked with your awesomeness. $70 **STANDARD HC BOOK + 2 PRINT LEVEL** $130 **LIMITED EDITION HARDCOVER + 2 SIGNED PRINTS LEVEL** PLUS New cover art for ALL Limited Edition Hard Cover copies of KARIBA! More stretch goals coming soon..."
Persuasive Messages	
Day 1 - A Quick Thank You...	...to everyone who has backed so far!!!! If you're reading this, you're a Day One Backer and you have all of my gratitude. I'm humbled by such a quick response to this project and... well... stunned by all the shares, likes, and so on.So... Thank you!!!!!!! Much love for our Day One Backers!
Launched Facebook Advertising Campaign	Hi guys, as it's been incredibly difficult to get media sources interested in our campaign, we've launched a Facebook advertising campaign earlier this week. If you're viewing this as a result of one of those ads, welcome, we look forward to adding you to our community of backers! As always, if you have any questions about the project, feel free to let us know by sending us a message. Thanks!
FINAL STRETCH: PLEASE SHARE!	Hey all! To everyone who has supported me or this project in any way, either through boosting my statuses, donating money, or giving me hugs and kind words: I appreciate you so so so much! To those of you who have hit me up saying you'll donate, but have not yet gotten around to it: I will 100% be messaging you in the coming days! To any of you who have followed this project in any way and have not contributed in any way: PLEASE DO! We are entering the final stretch of the kickstarter, and I need your support more than ever! If you cannot personally contribute money, which I TOTALLY UNDERSTAND and do not fault you for [minimum wage - cost of living = negative money] I would ask you to please share this! Share this with your parents, your friends, your professors, your doctors, literally anyone you know who might be interested and might have the means to donate! Post it to any social media or internet forums you might be a part of! Your support literally means the world to me, because this comic is becoming my world. This is both the best comic I have ever done, and the most important to me personally in terms of implications and thematic content. I desperately want for this to become a reality. This is more than a "school project". This is the beginning of my career in narrative media. This is the beginning of my life. (Not to be dramatic, but I'm a writer- what did you expect?)All of you are so CUTE & FUN !!!!! Lots of love, Mark

Appendix C. Performance of Cross-Validation

Accuracy: 91.16% ±1.16%

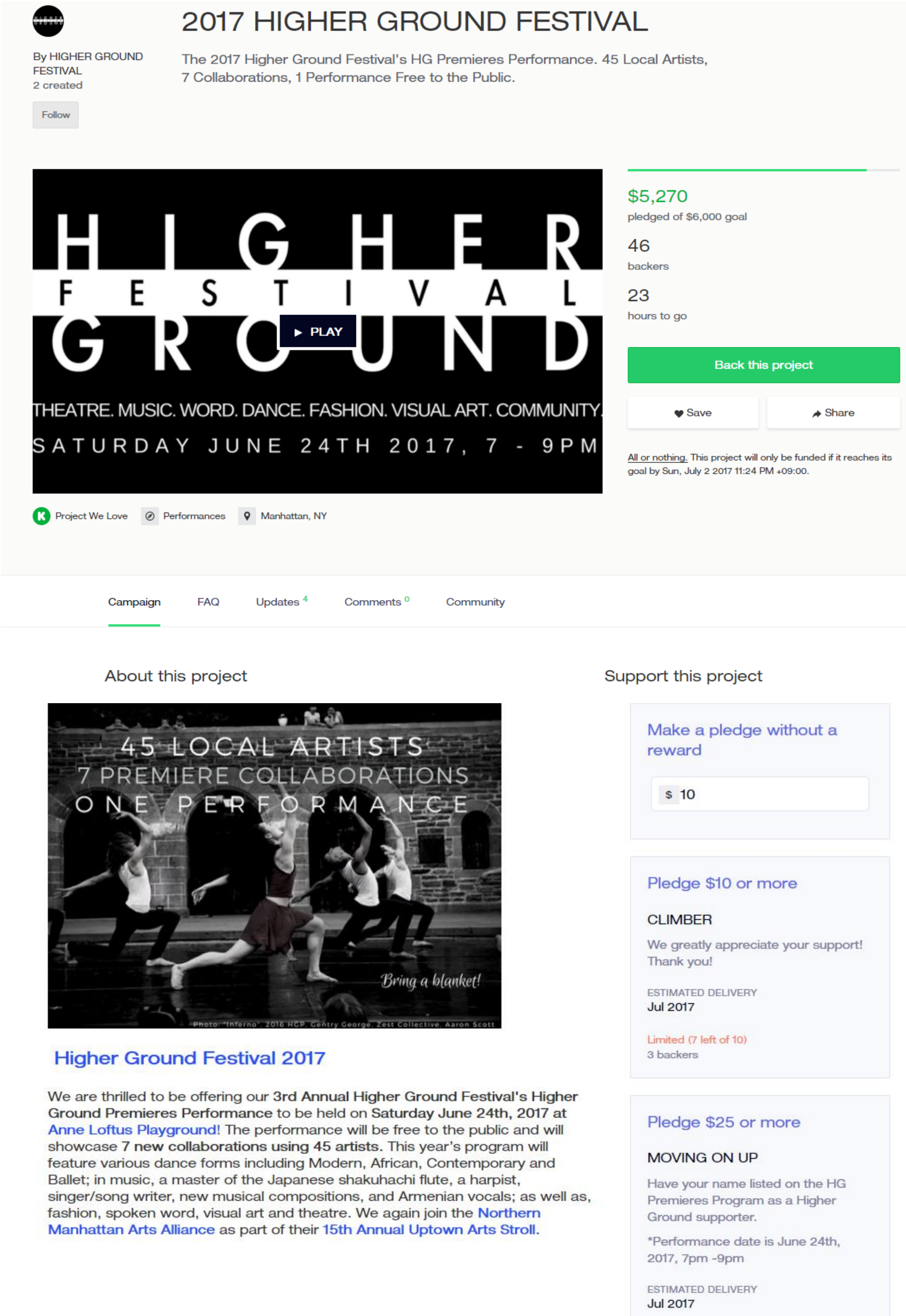
	true P	true I	class precision
pred. P	1066	98	91.58%
pred. I	161	1606	90.89%
class recall	86.88%	94.25%	

n = 2932



Note: In the case of binary classification task, the AUROC(Are under the Receiver Operating Characteristic) is commonly used as a summary statistics for the goodness of a predictor. The area indicates the probability that the predictor will predict and rank randomly chosen positive instances higher than a randomly chosen negative ones (Powers, 2011).

Appendix D. Example of Kickstarter Project Page



Appendix E.1 Model Fit Under Alternative Specifications of Independent Variables in the Regression equation

Relationship between TotalFunding and NInfo, NPers

	Linear x	Diminishing returns ln(x+1)	Inverted-U $x + x^2$	S-curve $x + x^2 + x^3$
AIC	129015.6	-141948.8	143147	130143.9

Note: AIC (Akaike Information Criteria) was used to find which functional performed best described the relationship between the number of each kind of message and the final auction price. (Akaike, 1973). These are:

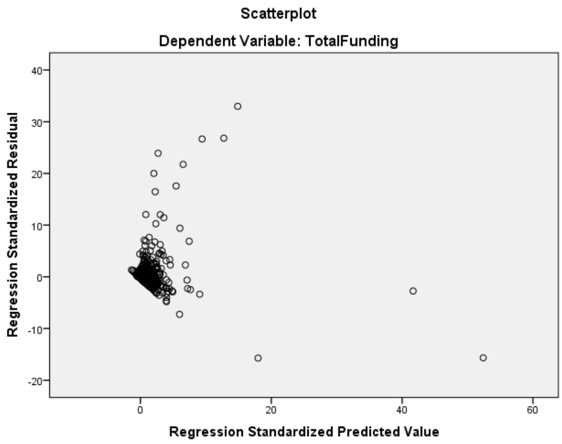
- Diminishing returns –transforming to log 10 of the number of each type of message;
- An inverted-U – transforming into linear and square terms.
- An S-curve – transforming into linear, square, and cube terms.

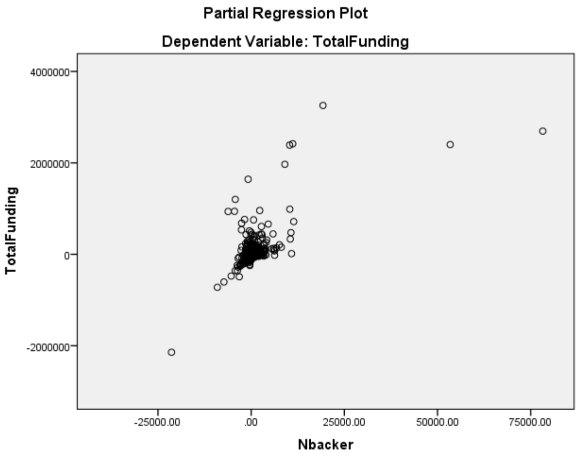
Below is the measure of model fit of AIC.

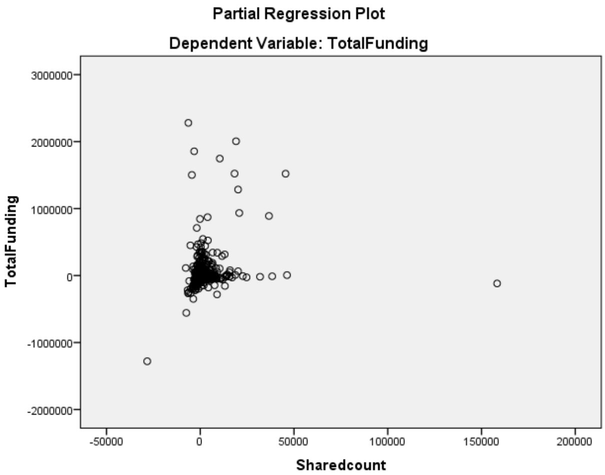
AIC = -2(log-likelihood) + 2K

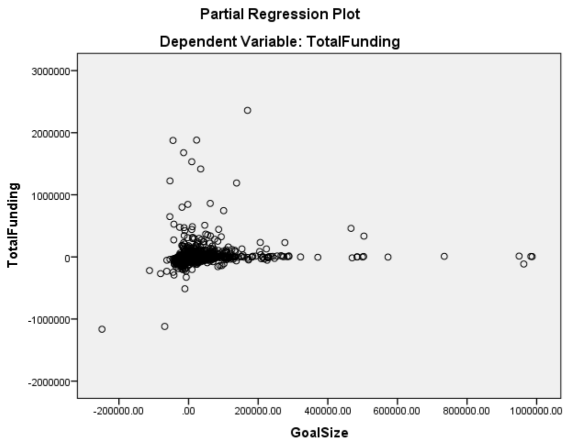
- K is the number of model parameters (the number of variables in the model plus the intercept).
- Smaller AIC indicates higher model fit

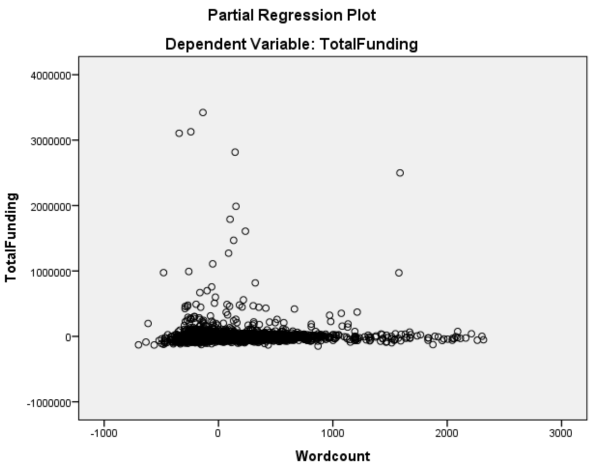
Appendix E.2 Heteroscedasticity (before transformation)



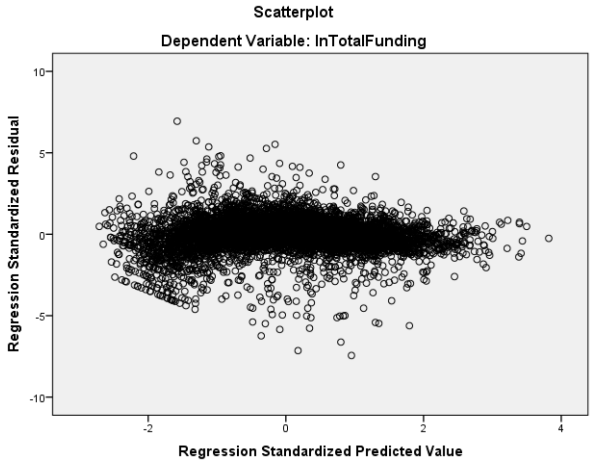


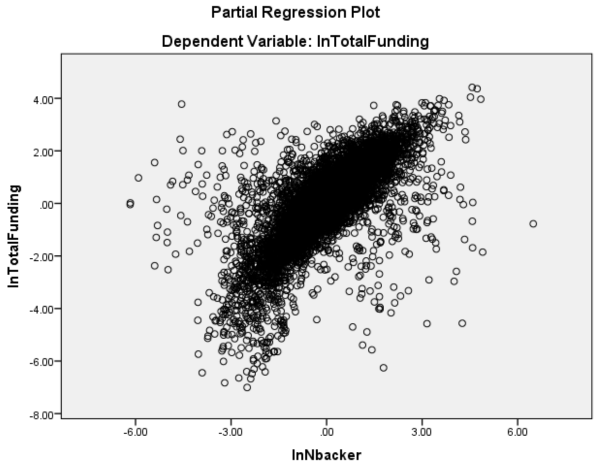


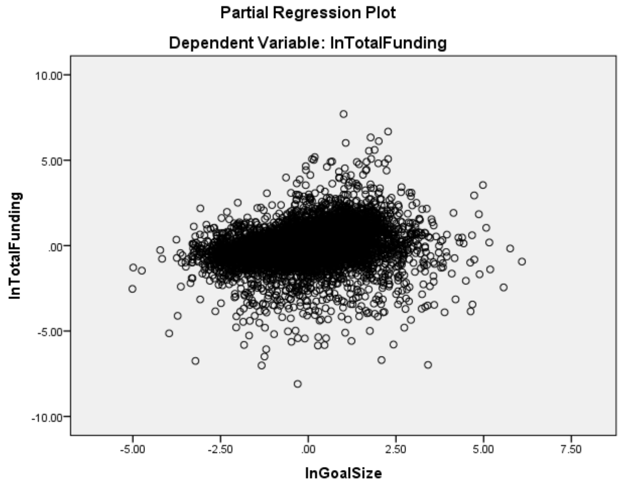


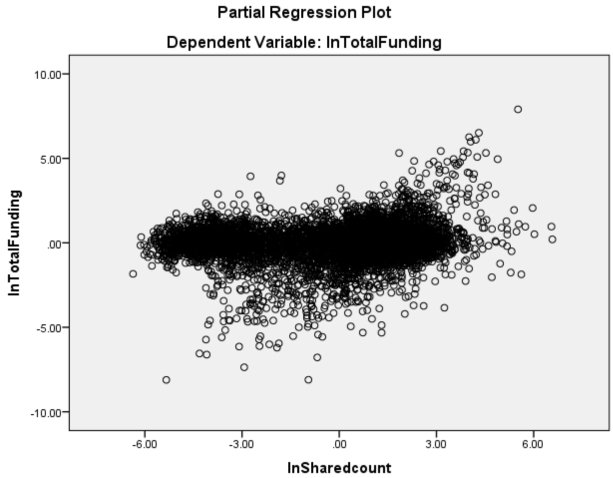


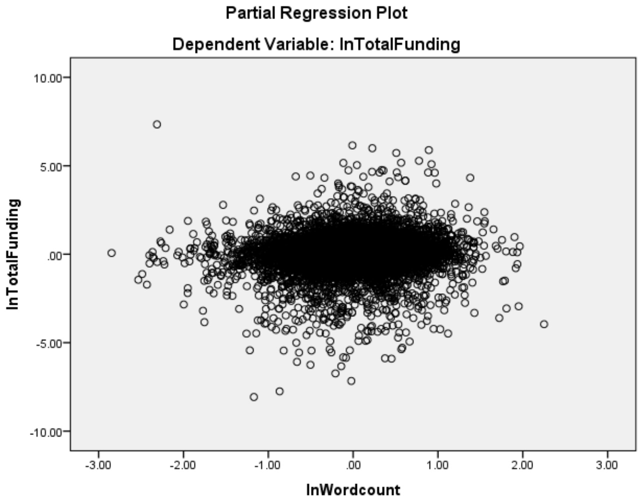
Heteroscedasticity (after transformation)



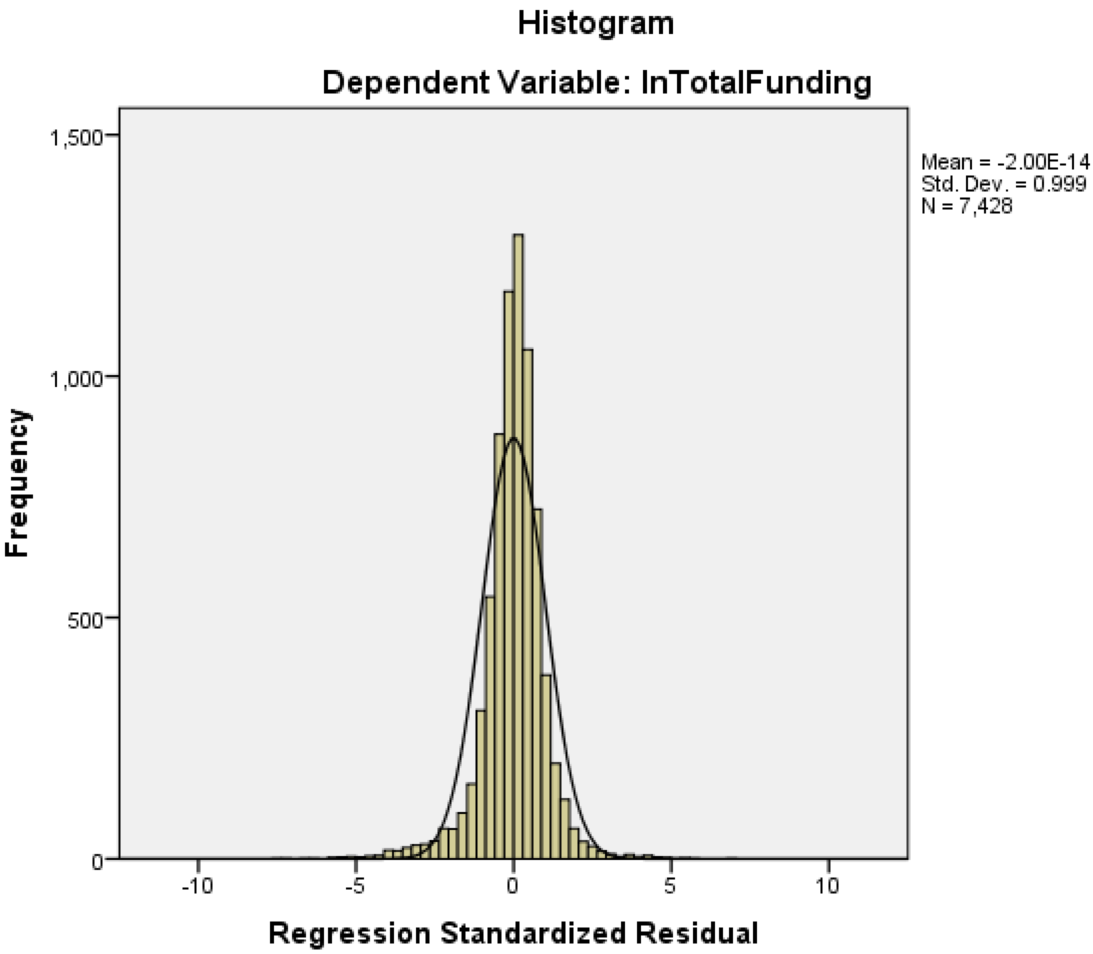








Appendix E.3 Histogram of Standardized Residual



Appendix E.4 Correlation Matrix

	Total Fund	Per	Info	NBacker	Goal	Word	Video	Image	Website	Experience	Backed His	Shared	NPers inst	NInfo inst	Category	Staff pick	Earlybird
Total Funding	1	.299**	.412**	.821**	.392**	.392**	.304**	.427**	.180**	.075**	.132**	.384**	.162**	.130**	-.060**	.375**	.192**
NPers	.299**	1	.140**	.320**	.123**	.318**	.120**	.300**	.157**	.126**	.187**	.083**	.239**	.331**	-.155**	.175**	.064**
NInfo	.412**	.140**	1	.396**	.074**	.208**	.121**	.227**	.109**	0.007	.101**	.192**	.159**	.109**	-.053**	.191**	.085**
NBacker	.821**	.320**	.396**	1	.261**	.354**	.233**	.397**	.189**	.119**	.190**	.332**	.234**	.218**	-.112**	.372**	.180**
GoalSize	.392**	.123**	.074**	.261**	1	.300**	.246**	.254**	-0.008	-.097**	-.059**	.246**	-.084**	-.023*	-.070**	.245**	.116**
Word count	.392**	.318**	.208**	.354**	.300**	1	.248**	.504**	.126**	.036**	.139**	.181**	.154**	.239**	-.135**	.244**	.124**
Video count	.304**	.120**	.121**	.233**	.246**	.248**	1	.244**	.067**	-.029*	.037**	.096**	.045**	.039**	-.037**	.131**	.123**
Image count	.427**	.300**	.227**	.397**	.254**	.504**	.244**	1	.146**	.092**	.159**	.167**	.267**	.287**	-.255**	.203**	.243**
Website count	.180**	.157**	.109**	.189**	-0.008	.126**	.067**	.146**	1	.263**	.162**	-0.013	.096**	.088**	-0.016	.096**	.043**
Experience	.075**	.126**	0.007	.119**	-.097**	.036**	-.029*	.092**	.263**	1	.264**	-0.017	.138**	.197**	-.104**	0.008	.024*
BackedHis	.132**	.187**	.101**	.190**	-.059**	.139**	.037**	.159**	.162**	.264**	1	.036**	.211**	.238**	-.126**	.106**	0.010
Shared count	.384**	.083**	.192**	.332**	.246**	.181**	.096**	.167**	-0.013	-0.017	.036**	1	.030**	.046**	-.070**	.198**	.075**
NPers inst	.162**	.239**	.159**	.234**	-.084**	.154**	.045**	.267**	.096**	.138**	.211**	.030**	1	.690**	-.315**	.088**	.070**
NInfo inst	.130**	.331**	.109**	.218**	-.023*	.239**	.039**	.287**	.088**	.197**	.238**	.046**	.690**	1	-.398**	.058**	.049**
Category	-.060**	-.155**	-.053**	-.112**	-.070**	-.135**	-.037**	-.255**	-0.016	-.104**	-.126**	-.070**	-.315**	-.398**	1	-0.012	-.191**
Staff pick	.375**	.175**	.191**	.372**	.245**	.244**	.131**	.203**	.096**	0.008	.106**	.198**	.088**	.058**	-0.012	1	.045**
Early bird	.192**	.064**	.085**	.180**	.116**	.124**	.123**	.243**	.043**	.024*	0.010	.075**	.070**	.049**	-.191**	.045**	1

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed).

Appendix F. Classification Tables

Null Model

		Predicted		
		Fundstatus		Percentage
Observed		Fail	Success	Correct
Fundstatus	Fail	0	2247	.0
	Success	0	5154	100.0
Overall Percentage				69.6

a. Constant is included in the model.

b. The cut value is .500

Full Model (Model 2.1 without interaction effect)

		Predicted		
		Fundstatus		Percentage
Observed		Fail	Success	Correct
Fundstatus	Fail	1692	555	75.3
	Success	249	4905	95.2
Overall Percentage				89.1

a. The cut value is .500

Adjusted R square count: {(1692 + 4905) – 5154} / {7401 – 5154} = 0.642

Full Model (Model 2.2 with interaction effect)

		Predicted		
		Fundstatus		Percentage
Observed		Fail	Success	Correct
Fundstatus	Fail	1695	552	75.4
	Success	256	4898	95
Overall Percentage				89.1

a. The cut value is .500

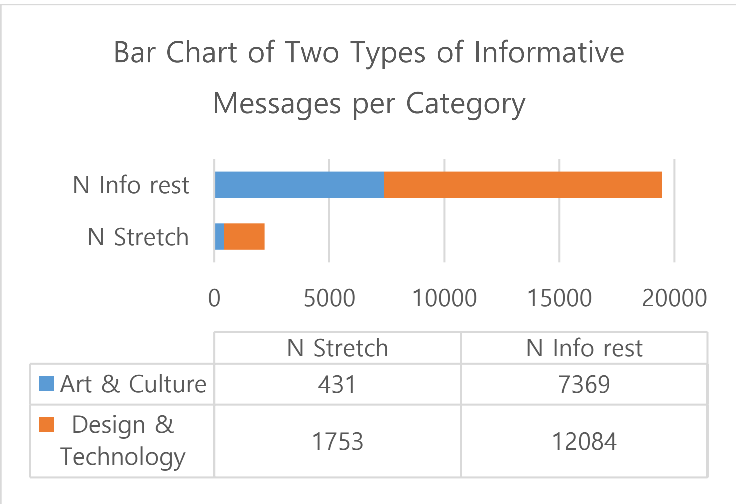
Adjusted R square count: {(1695 + 4898) – 5154} / {7401 – 5154} = 0.642

Appendix G. Regression on Instruments

Dependent Variable	**Number of Informative messages**			**Number of Persuasive Messages**		
	Coefficient	t	Sig.	Coefficient	t	Sig.
(Constant)		-0.211	0.833		1.528	0.127
lnGoalSize	0.030	2.554	0.011	-0.066	-5.530	0.000
Category	0.015	1.221	0.222	0.041	3.394	0.001
Staffpick	0.040	3.539	0.000	0.038	3.235	0.001
Wordcount	0.144	11.468	0.000	0.050	3.905	0.000
Videocount	0.010	0.929	0.353	0.021	1.839	0.066
Imagecount	0.078	5.972	0.000	0.059	4.380	0.000
Websitecount	0.059	5.448	0.000	0.037	3.329	0.001
Experience	0.025	2.216	0.027	-0.061	-5.407	0.000
BackedHis	0.055	4.991	0.000	0.014	1.202	0.230
Earlybird	-0.017	-1.539	0.124	0.009	0.793	0.428
lnSharedcount	-0.033	-2.999	0.003	0.073	6.425	0.000
lnNbacker	0.156	12.176	0.000	0.315	23.941	0.000
NInfoinst	**0.225**	**18.183**	**0.000**			
NPersinst				**0.065**	**5.429**	**0.000**

Appendix H.1 Descriptive statistics for informative messages that are related to stretch goals.

Variables	N	Min	Max	Range	Sum	Mean	SD
N stretch goals msg	7428	0	16	85	19453	2.62	3.56
N rest of Info msg	7428	0	85	16	2184	0.29	0.97



Appendix H.2 Regression Model with Stretch Goals

Models	Model 1.1 (with stretch goals)			Model 1.2 (with moderator)			Model 2.1 (with stretch goals)			Model 2.2 (with interaction)		
	Coefficients	t	Sig.	Coefficients	t	Sig.	Coefficients	Exp(β)	SE	Coefficients	Exp(β)	SE
Intercept		4.048	0.000		4.015	0.000	2.634***	13.927	0.46	2.547***	12.77	0.465
Independent Variable												
lnNPers	0.087***	13.418	0.000	0.085***	10.858	0	1.109***	3.03	0.073	1.2***	3.321	0.091
lnNstretch	0.027***	4.093	0.000	0.03***	4.353	0	3.426***	30.755	0.403	3.51***	33.448	0.432
lnNrestofInfo	0.016***	2.462	0.014	0.018**	2.281	0.023	0.015	1.015	0.061	0.014	1.014	0.075
Moderating Variable												
lnInfo x Cat												
lnPers x Cat				0.004	0.397	0.691				-0.259*	0.772	0.147
lnNstretch x Cat				-0.01	-1.598	0.11				-0.782	0.458	1.195
lnNrestofInfo x Cat				-0.004	-0.413	0.679				-0.015	0.985	0.124
Control Variable												
lnNbacker	0.643***	82.019	0.000	0.643***	82.004	0	1.333***	3.792	0.042	1.335***	3.8	0.042
lnGoalSize	0.143***	21.048	0.000	0.143***	20.97	0	-1.208***	0.299	0.043	-1.208***	0.299	0.043
Category	0.051***	7.71	0.000	0.053***	4.366	0	0.867***	2.381	0.087	1.047***	2.85	0.171
Staffpick	0.026***	4.064	0.000	0.026***	4.064	0	0.535***	1.707	0.124	0.543***	1.72	0.124
lnWordcount	0.057***	7.746	0.000	0.057***	7.775	0	0.232***	1.261	0.071	0.235***	1.264	0.071
Videocount	0.063***	9.971	0.000	0.063***	9.951	0	0.361***	1.434	0.073	0.362***	1.436	0.073
Imagecount	0.058***	7.874	0.000	0.058***	7.808	0	-0.007	0.993	0.006	-0.008	0.992	0.006
Websitecount	0.024***	3.811	0.000	0.024***	3.843	0	0.031	1.031	0.047	0.029	1.03	0.047
Experience	0.015**	2.277	0.023	0.014**	2.264	0.024	-0.022	0.978	0.023	-0.02	0.98	0.024
BackedHis	-0.014**	-2.227	0.026	-0.014**	-2.226	0.026	0.001	1.001	0.003	0.001	1.001	0.003
lnSharedcount	0.087***	13.556	0.000	0.087***	13.57	0	0.119***	1.127	0.017	0.118***	1.125	0.017
Earlybird	0.022***	3.502	0.000	0.022***	3.526	0	0.193	1.213	0.141	0.187	1.206	0.142
p (ANOVA)	0			0.000			0.000			0.000		
Adjusted R Square/count	0.744			0.744			0.641			0.646		

Dependent variable: ln(Total Funding) / Funding success (1 = Success, 0 = Fail)
Category (1 = Art and Culture, 0 = Design and Technology)
*** p < 0.01, ** p < 0.05, * p < 0.1

Appendix I.1 Split-half analysis (coefficient comparison)

Model 1: Linear Regression

	Original		Half-sampled	
	coefficient	significance	coefficient	significance
Independent Variables				
lninfo	**0.022**	**	**0.028**	***
lnpers	**0.080**	****	**0.095**	****
Control Variables				
BackedHis	**-0.015**	*	**-0.019**	**
Imagecount	**0.066**	****	**0.051**	****
Experience	**0.012**	*	**0.017**	*
lnShared	**0.092**	****	**0.083**	****
lnbacker	**0.642**	****	**0.651**	****
lngoalsize	**0.142**	****	**0.138**	****
lnWordcount	**0.039**	****	**0.045**	****
Websitecount	**0.027**	****	0.012	
Videocount	**0.067**	****	**0.054**	****
Category	**0.043**	****	**0.057**	****
Staffpick	**0.024**	***	**0.029**	***
Earlybird	**0.017**	*	**0.028**	***

Dependent Variable: ln(TotalFunding)
****p = 0.000, *** p < 0.01, ** p < 0.05, * p < 0.1

Model 2: Logistic Regression

	Original		Half-sampled	
	coefficient	significance	coefficient	significance
Intercept	**1.742**	***	**1.173**	
Independent Variables				
lninfo	**0.218**	****	**0.214**	***
lnpers	**1.127**	****	**1.113**	****
Control Variables				
BackedHis	**1.326**	****	**1.3**	****
Imagecount	**-1.229**	****	**-1.213**	****
Experience	**0.714**	****	**0.814**	****
lnShared	**0.571**	****	**0.444**	***
lnbacker	**0.252**	****	**0.244**	***
lngoalsize	**0.352**	****	**0.465**	****
lnWordcount	0.002		-0.001	
Websitecount	0.027		**0.14**	**
Videocount	-0.006		-0.009	
Category	0.003		0.007	
Staffpick	**0.129**	****	**0.131**	****
Earlybird	0.272	*	0.11	

Dependent Variable: Fund status
****p = 0.000, *** p < 0.01, ** p < 0.05, * p < 0.1

Appendix I.2 Split-half analysis (Prediction for Model 1)

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Real	8.0487	3714	2.11872	0.03477
Prediction	8.0354	3714	1.84032	0.0302

Paired Samples T-Test

				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Real - Prediction	0.01333	1.09003	0.01789	-0.02174	0.0484	0.745	3713	0.456

Appendix J. Comparison with and without outliers

	Original		Without outliers	
	coefficient	significance	coefficient	significance
Independent Variables				
lninfo	0.028	***	0.016	***
lnpers	0.090	***	0.052	***
Control Variables				
lnbacker	0.654	***	0.647	****
lngoalsize	0.142	***	0.151	****
Category	0.026	***	0.039	****
Staffpick	0.028	***	0.040	****
Wordcount	0.029	***	0.030	***
Videocount	0.067	***	0.053	***
Imagecount	0.066	***	0.030	***
Websitecount	0.027	***	0.007	
Experience	0.012	**	0.010	**
BackedHis	-0.012	*	-0.028	***
lnShared	0.087	***	0.032	****
Earlyb	0.020	**	0.007	

Dependent Variable: Fundstatus
****p = 0.000, *** p < 0.01, ** p < 0.05, * p < 0.1

Appendix J. Models including the discarded update posts

	Model 1 (Total Funding)				**Model 2 (Funding Success)**				
			95% CI for B					95% CI for EXP(B)	
	Standardized Beta	Sig.	Lower Bound	Upper Bound	B	Sig.	Exp(B)	Lower	Upper
(Constant)	0	0	1.109	2.278	0.1	0.478	1.105	0.839	1.454
lnnewNinfo	0.003	0.763	-0.044	0.06	0.151	0.47	1.163	0.772	1.753
lnnewNpers	-0.012	0.234	-0.123	0.03	**3.124**	0	22.736	15.3	33.787
lnNbacker	**0.841**	0	1.05	1.117	**-2.195**	0	0.111	0.082	0.151
lnGoalSize	**0.107**	0	0.12	0.186	**0.67**	0.009	1.954	1.183	3.229
Category	**0.041**	0	0.081	0.272	0.231	0.531	1.26	0.611	2.601
Staffpick	-0.02	0.069	-0.236	0.009	**0.41**	0.036	1.506	1.026	2.211
lnWordcount	**0.037**	0.002	0.042	0.186	**0.506**	0.013	1.659	1.113	2.473
Videocount	**0.051**	0	0.093	0.229	**-0.034**	0.022	0.966	0.938	0.995
Imagecount	0.021	0.094	-0.001	0.011	0.159	0.251	1.173	0.893	1.54
Websitecount	0.019	0.08	-0.005	0.095	-0.12	0.072	0.886	0.777	1.011
Experience	0.014	0.204	-0.008	0.038	**-0.017**	0.04	0.983	0.967	0.999
BackedHis	**-0.031**	0.006	-0.008	-0.001	0.084	0.094	1.087	0.986	1.199
lnSharedcount	**0.049**	0	0.023	0.059	**1.069**	0.027	2.913	1.126	7.535
Earlybird	0.014	0.194	-0.05	0.246	0.387	0.499	1.473	0.479	4.524
NPersinst	-0.015	0.301	-0.316	0.098	**-0.454**	0.001	0.635	0.486	0.831
NInfoinst	**-0.051**	0	-0.134	-0.038	**4.777**	0.003	118.759		