# PREDICTING FUNDRAISING SUCCESS IN REWARD-BASED CROWDFUNDING

by

Liqian Bao

A Dissertation Submitted in

Partial Fulfilment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

University of Wisconsin-Milwaukee

May 2023

ABSTRACT

PREDICTING FUNDRAISING SUCCESS IN REWARD-BASED CROWDFUNDING

by

Liqian Bao

The University of Wisconsin-Milwaukee, 2023
Under the Supervision of Professor Huimin Zhao

While there has been a rapid growth of the reward-based crowdfunding market, successfully achieving the target amount of a project is a big challenge. Entrepreneurs are eager to know the prospects of their campaigns. Investors also want to fund promising campaigns with high quality and low uncertainty. Therefore, effectively predicting the success of a project throughout the fundraising period is a crucial task for both entrepreneurs and investors. On the one hand, it provides guidance to entrepreneurs about project progress and potential, helping them adjust their campaigns in time. On the other hand, it helps investors manage their funding risks and reduce opportunity costs. However, most of the existing research has been aimed to explore the determinants of fundraising success, but much less attention has been paid to the success prediction problem. We explore the crowdfunding success prediction problem from two perspectives in the following two essays.

In Essay 1, we mine semantic features from comments to improve fundraising success prediction. More and more participants share and discuss facts and opinions about projects by posting comments, which can influence investors' funding decisions. Previous studies have mainly focused on quantity, sentiment, and linguistic features of comments, largely overlooking the value of semantic features, in predicting fundraising success. Rooted in information asymmetry and

herding behavior theories, we posit that discovering semantic signals from comments and distinguishing actor roles will benefit fundraising success prediction. We propose a framework with novel latent semantic features of comments. Empirical evaluation using data from a prominent platform demonstrates the utility of the framework and reveals interesting patterns in the dynamic predictive effects of semantic features for different actor roles.

In Essay 2, we apply features from multimodal data (texts, images, and videos) to improve fundraising success prediction. With the development in artificial intelligence and big data, multimodality has become one of the popular research areas of IS since multiple modalities can provide complementary information and improve the performance of the overall decision-making process. However, there is a lack of research providing a comprehensive investigation of multimodal data in crowdfunding. To gain a comprehensive review of linguistic and visual features of multimodality in crowdfunding, we propose a framework built on theories of Halliday's metafunctions framework of languages (1985), Kress and Van Leeuwen's functional visual design (1996), and Royce's intersemiotic complementarity of languages and visual images (1998) to explore relevant features representing the ideational, interpersonal, and textual metafunctions of multimodal data in crowdfunding. We have conducted several experiments to study the effectiveness of each metafunction, each modality, and their interactions in predicting crowdfunding success.

To

my husband,

my lovely son,

and especially my parents and parents-in-law

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

I want to express my sincere gratitude to my advisor and committee chair, Dr. Huimin Zhao, for his guidance, support, and encouragement throughout my Ph.D. program. I will always cherish the memory of how Dr. Zhao encouraged and motivated me to apply for the Ph.D. program. And throughout my Ph.D. program, he provided invaluable advice and guidance, whether it was related to coursework, academic research, or life in Milwaukee. His insightful suggestions in the developing of my dissertation idea, methodology, and framework are invaluable in the completion of my Ph.D. degree.

But what I value most about Dr. Zhao is his dedication to his students. I always remember that he said, "Even though I have a lot of other responsibilities to commit to, students' needs always come first". I was able to reach him within minutes, no matter what time of day or night it was. He respected every decision I made and offered the largest support he could provide. When I had to suspend my Ph.D. program for a year due to family issues, Dr. Zhao continued to advise me on my research projects, and he even reminded me of scholarship opportunities. His kindness and support during that time were greatly appreciated. Sometimes, pursuing a Ph.D. can be a tough, lonely, and stressful journey, but with the lead and guidance of Dr. Zhao, it is more like an adventure. He is a model of a good professor, and I hope I can follow his footsteps in pursuing my career.

I would also like to thank Mrs. Zhao for her support and help during my stay in Milwaukee, especially when I came across unexpected family issues. I cannot imagine how fortunate I am to meet such a kind couple, who enlightened my entire Ph.D. journey.

In addition, I would like to express my sincere appreciation to my other committee members: Dr. Sanjoy Ghose, Dr. Cheng Chen, Dr. Zhao Wang, and Dr. Gang Chen. Dr. Sanjoy Ghose, my minor advisor, has been a great inspiration to me and has led me to explore the connections between Marketing and Information Systems in some of my research projects. Dr. Cheng Chen, who shares similar research interests with me, provided me with valuable advice in developing my dissertation. Dr. Zhao Wang, the co-author of my first essay, offered concrete help on the framework, methodology, and technology aspects of this essay. Dr. Gang Chen, the co-author of my second essay, is an expert in multimodal deep learning, and I consulted and discussed with him during the entire formation stage of my second essay. Their help has been instrumental in developing my dissertation, and it has been an enjoyable journey under the guidance of my committee members.

Furthermore, I would like to take this opportunity to express my gratitude to all the faculty and staff at the Lubar College of Business. Studying and working in such a friendly and supportive family has been an enjoyable experience. I am grateful to Dr. Atish Sinha, who offered me PA work as a coordinator at CTI and SAP workshops, and provided me with an opportunity to teach the Business Intelligence course independently. These experiences have been invaluable in terms of administration and teaching skills. I would also like to thank Dr. Derek Nazareth and Dr. Mark Srite, who offered me seminars and advice on my research papers. Dr. Yang Wang gave me valuable advice in job hunting, and Dr. Odai Khasawneh usually stopped by to chat with me about my academic and job search progress. I am deeply appreciative of their guidance and support, which have played a precious role in my academic and professional journey.

I would also like to extend my gratitude to my classmates, Eun Jung Lee, Cong Zhang, Youngeui Kim, Zongxi Liu, Xiao Xiao, and Xiaoxia Zhu. It was a pleasure to have fellow panthers to study with and learn from. I would also like to thank my best friends in Milwaukee, Jing Chu, Shuxin Dong, and Lu Shi, who brought joy and color to my doctoral life.

Last but not least, I would like to express my sincere gratitude to my warm family. My husband Xin provided me with the strongest positive and stable emotional support during the stressful and challenging periods of my Ph.D. journey. I am also deeply grateful to my dear son Zhilin, who had to live without the company of his parents when he was just 13 months old. I am so happy to witness his healthy and joyful growth. I would like to extend a special thanks to my parents and parents-in-law, who devoted themselves to taking good care of my son, allowing me to focus on my dissertation and research projects. Their constant love and support have provided me with the strength and motivation to keep moving forward.

# CHAPTER 1

## Introduction

Crowdfunding is a form of finance that allows entrepreneurs to reach relatively small contributions from a large number of individuals through the Internet instead of traditional financial intermediaries (Cumming and Zhang, 2016; Mollick, 2014). It has received growing attention from entrepreneurs as a promising bridge for raising capital in recent years. It attracted over $11,580 million worldwide in 2019, and the global market size is expected to reach $21,370 million by the end of 2024, with a 16.5% CAGR in terms of revenue1. Harnessing the power of the crowd to fund small ventures, which are unlikely to get funded by traditional forms of funding (e.g., banks, venture capital, and angel investment), crowdfunding enables startup ventures and smaller entrepreneurs to participate and reach the funds since the transaction costs are dramatically reduced (Agrawal et al., 2014; Ahlers et al., 2015).

Reward-based crowdfunding is one type of crowdfunding that represents a relatively new way of acquiring capital. Investors in the reward-based crowdfunding fund a campaign (project) because they want early access to a new product rather than obtaining monetary benefits. The investors of successful campaigns are playing the role of early consumers, allowing them access to a new product earlier, at a better price, or with other special benefits (Mollick, 2014). Reward-based crowdfunding enables entrepreneurs to fulfill a specific purpose, which often focuses on the creation or distribution of a new product (Davis et al., 2017).

---

1 LP Information INC. Global Crowdfunding Market Growth (Status and Outlook) 2019-2024. ReportsnReports. 2019. https://www.reportsnreports.com/reports/2268612-global-crowdfunding-market-growth-status-and-outlook-2019-2024.html.

While there has been rapid growth of the reward-based crowdfunding market, successfully achieving the target amount of a project is a big challenge. Entrepreneurs are eager to know the prospects of their campaigns. Investors also want to fund promising campaigns with high quality and low uncertainty. Therefore, effectively predicting success throughout the fundraising period is a crucial task for both entrepreneurs and investors. On the one hand, it provides guidance to entrepreneurs about project progress and potential, helping them adjust their campaigns in time. On the other hand, it helps investors manage their funding risks and reduce opportunity costs. However, most of the current research aims to explore the determinants of fundraising success, but much less attention has been paid to the *success prediction* problem, which is the subject of this dissertation.

We study the success prediction problem of reward-based crowdfunding from two perspectives. The first essay mines valuable semantic features from comments to improve fundraising success prediction. The second essay explores effective features representing the ideational, interpersonal, and textual metafunctions of multimodal data in predicting fundraising success.

**Essay 1: Mining Semantic Features for Success Prediction in Reward-Based Crowdfunding**

Comments are the most convenient channel through which entrepreneurs and investors can communicate with each other on the crowdfunding platform. On the one hand, entrepreneurs and other investors can contribute their ideas by posting instantly progress and feedback on the comment page. On the other hand, investors can leave their opinions, suggestions, criticisms, or questions in the comment section. However, existing studies so far have mainly focused on the quantity, sentiment, or linguistic characteristics of comments, ignoring the predictive utility of the semantic of comments. To fill this gap, we propose a novel framework for mining semantic features from comments to improve fundraising success prediction. Based on information asymmetry and herding effect theories, we extract semantic intensities and keywords to predict fundraising success with a supervised learning approach. Since crowdfunding is a time-sensitive financial activity, we build several predictive models at different stages of the fundraising period to dynamically predict fundraising success. We demonstrate the predictive effectiveness of the proposed semantic features through empirical evaluation.

**Essay 2: Reward-based Crowdfunding Success Prediction with Multimodal Data**

Multimodality integrates multiple data modalities, such as linguistic, visual, gesture, color, design, and sound signals, to express the idea (Norris and Maier, 2014). With the development of artificial intelligence and big data, multimodality has become one of the popular research areas of IS since multiple modalities can provide complementary information and improve the performance of the overall decision-making process.

Crowdfunding campaign entrepreneurs rely on multimodal data, including the appropriate use of texts, images, and videos, to communicate the novelty and value of their ideas to backers (Yang et al., 2020). However, there is a lack of research providing a comprehensive investigation of multimodal data in crowdfunding. Based on linguistic and visual image metafunction theory, we discover and extract ideational, interpersonal, and textual metafunction features from multimodal data in crowdfunding platforms using deep learning models and Google Vision API. We demonstrate the predictive utility of three metafunctions and their interactions of single data modality and multimodality, in predicting crowdfunding success through empirical evaluation.

# CHAPTER 2

## Essay 1: Mining Semantic Features for Success Prediction in Reward-Based Crowdfunding

### 2.1 Introduction

No doubt entrepreneurs know better about their projects than investors. To alleviate such information asymmetry (Akerlof, 1970), entrepreneurs are encouraged to clearly describe their projects on the crowdfunding platform. Although entrepreneurs tend to provide valuable information in their project descriptions to help potential investors make decisions, they are prone to amplify the strengths of their projects to attract investors and sometimes may overlook their shortcomings or topics of interest to inventors, such as their shipping policies and product compatibility. Moreover, most investors are adventure hurters, it is difficult for them to build relational ties to entrepreneurs to reduce risks, and they are unlikely to get consultations from financial institutions who are experts in evaluating, monitoring, and managing risks (Huang and Knight 2017; Kim and Viswanathan, 2019). Potential investors are more likely to be influenced by herding behavior (Banerjee, 1992), that is, an investor may view preceding investors' opinions and mimic their decisions. Subsequent comments of projects may play an important role in predicting fundraising outcomes (success or failure), as they reveal opinions, attitudes, and decisions of existing investors, and potential investors who lack sufficient information can follow the lead investors with more complete and accurate information. In addition, typical concerns about a project from the investors' perspective are visible to potential investors through comments. Also, comments are updated more easily and frequently, thus providing a good channel for potential investors to receive instant progress and feedback from entrepreneurs. Therefore, we posit that the comments regarding a project may carry valuable information for predicting its likelihood to succeed.

However, there has been limited research on comment text mining to improve crowdfunding success prediction. The few studies (Chen et al., 2015; Lai et al., 2017; Lee et al., 2019; Li et al., 2016; Ryoba et al., 2020) using comments considered the comment quantity, sentiment, and linguistic characteristics, but largely overlooked the semantics of comments that reflects potential and valuable information influencing investors' funding decisions.

Consider the following real example of a thread consisting of a comment from an investor and a reply from the entrepreneur:

[Investor]:

*Love the concept and can't wait to try it out! Question: is there a way to make the sides of dividers higher (or just some of them)? I'm into baking and I'm worried, that it'll not be tall enough for it.*

[Entrepreneur]:

*Welcome to the project, we are working on higher sided versions now that will be available to add within the campaign or just after in the survey. Thanks so much for backing us! We are excited to show you the developments.*

The first comment delivers two types of semantics. One is that the investor supported and liked this product, potentially triggering herding behavior and attracting more potential investors in the future. The other is that the investor raised a suggestion and asked whether it can be possible, arousing a response from the entrepreneur (the second comment), which conveyed effective information to the potential investors who have similar concerns. If we treat all the comments equally by the quantity or linguistic features, it may limit our understanding about such semantics, which reduces information asymmetry or evokes herding behavior. To bridge this gap in the current literature, we strive to develop effective *semantic features* for fundraising success prediction in this design science research (Hevner et al., 2004).

Furthermore, entrepreneurs' comments and investors' comments play different roles in attracting funding. Investors' comments are regarded as important herd signals and impact other potential investors' decision making (Song et al., 2019). Entrepreneurs' comments provide important signals for investors to estimate their projects' quality, and their replies show their willingness to interact with investors, giving investors confidence to fund their projects. Such interaction has been investigated in online commerce (Chen et al., 2011) and crowdfunding (Song et al., 2019). Merging investors' comments and entrepreneurs' replies, as done in previous studies on fundraising success prediction (Chen et al., 2015; Lai et al., 2017; Lee et al., 2018; Li et al., 2016; Ryoba et al., 2020), may miss such interaction. Therefore, we propose *distinguishing the comments from different actor roles* (investors vs. entrepreneurs) in fundraising success prediction.

Additionally, as reward-based crowdfunding is a time-sensitive and time-variant financial activity, it is also important to consider the dynamics when predicting fundraising success. The behaviors of investors may vary as the fundraising proceeds. For example, Kuppuswamy and Bayus (2018) suggested that investors are more likely to contribute to a project in the first and last weeks as compared to the middle fundraising period. Moreover, the instantly updated information from comments in different stages of the fundraising period may evoke different effects on investors' funding decisions, and the effects of comments from different actor roles may evolve differently over the fundraising period. Therefore, we also investigate *how the predictive utilities of the comments from different actor roles evolve* over the fundraising period.

Specifically, we propose a novel framework for mining semantic features from comments to improve fundraising success prediction. Rooted in theoretical foundations, i.e., *information asymmetry* (Akerlof, 1970) and *herding behavior* (Banerjee, 1992), we address two types of comments: *informational* comments and *inclinational* comments. The informational comments regarding a project pertain to the project's characteristics, such as appearance, functions, features, and delivery, which may help to alleviate the information asymmetry. The inclinational comments reflect the investors' opinions or attitudes, which may trigger herding behavior and influence more potential investors. We define semantic features (semantic *intensities* and *keywords*) according to the two types of comments. We separate the comments by actor roles (entrepreneurs and investors) to further improve the predictive performance. We build prediction models at different stages of the fundraising period to examine the dynamic predictive utilities of the proposed features. We have evaluated our proposed framework using data from Kickstarter, the most prominent reward-based crowdfunding platform, and standard classification methods. The results demonstrate the predictive utilities of our proposed semantic features and distinguishing actor roles. Our analysis also reveals different dynamic patterns between the two actor roles.

Our study makes several contributions to the literature. While existing studies on the utilities of crowdfunding comments have only considered quantity, sentiment, and linguistic features, we propose novel semantic features effective for fundraising success prediction. We show the usefulness of distinguishing actor roles. We also demonstrate the dynamics of the predictive utilities of the proposed semantic features and distinguishing actor roles. Our study has managerial implications for the stakeholders in the crowdfunding market. Entrepreneurs can gain insights into the progress and prospect of their projects and accordingly adjust their projects promptly, investors can find and fund more promising projects, and crowdfunding platforms can make better recommendations to potential investors, based on predictions using our framework.

## 2.2 Related Work

### *2.2.1 Prediction of Crowdfunding Success*

Crowdfunding has received growing attention in recent years. Much research has been devoted to investigating factors influencing crowdfunding success. Most recently, scholars have started to look into the prediction of crowdfunding success. These studies have applied various machine learning methods, such as logistic regression (Liang et al., 2020; Yu et al., 2018), k-nearest neighbors (Etter et al., 2013; Ryoba et al., 2020), decision tree (Kaur and Gera., 2017; Yu et al., 2018), support vector machine (Chen and Shen, 2019; Li et al., 2018; Yu et al., 2018), random forest (Ahmad et al., 2017; Chen et al., 2015), and deep learning (Li et al., 2018; Wang et al., 2020; Yu et al., 2018).

Some project-related features, such as duration, category, goal amount, and numbers of images, videos, FAQs, and updates, have been frequently used in success prediction models. Besides such basic features, some studies have explored textual features of the description of a project. Chen and Shen (2019) used keywords from descriptions to improve prediction accuracy. Kaminski and Hopp (2019) found that linguistic styles of descriptions and the words the creators speak help to predict crowdfunding success. For example, the linguistic styles that aim to trigger excitement or social are better predictors of fundraising success while the words related to monetary depictions of venture reduce the success probability. Wang et al. (2017) studied textual descriptions from the sentimental aspect and showed that positive sentiment promotes campaign success and improves predictive accuracy. Yuan et al. (2016) extracted topical features from project descriptions to predict fundraising success. Some studies have explored features derived from social media, such as social network connections, tweets, number of Facebook shares for a project, and number of Facebook friends of creators. Li et al. (2016) found that social network-based features, especially features obtained at the beginning stage of a project, can help in improving predictive performance. Etter et al. (2013) derived features from tweets mentioning a campaign. Kaur and Gera (2017) further found that the numbers of tweets of distinct actor roles (e.g., backers, promoters) are useful for predicting the success of a campaign.

While these studies considered exploring new features from texts, they derived features from project descriptions or social networks, largely ignoring the effects of dynamic interactions between entrepreneurs and investors on the crowdfunding platform. The profile or description of a project is largely static and is created by the entrepreneur only. If we only consider the antecedent texts, the important information from dynamic conversations is missing. It is hence necessary to track the comments dynamically throughout the fundraising period, which may be beneficial to predicting the success of the fundraising. Moreover, only a small portion of projects provide their connections to social networks, and the main channel to understand a project is the crowdfunding platform itself. The comment section is the only forum where investors and entrepreneurs can interact on the crowdfunding platform. On the one hand, potential investors are influenced by the information provided by entrepreneurs and investors to reduce information asymmetry. On the other hand, the opinions and attitudes of investors may trigger herding behavior and influence potential investors' decisions (Ahsan et al., 2018).

A few studies have used features derived from comments in fundraising success prediction. As our focus is on the predictive utilities of comments, we specifically discuss these studies, as well as studies suggesting comment-related factors influencing fundraising success, in the following subsection.

### *2.2.2 Crowdfunding Comments*

On the crowdfunding platform, entrepreneurs try to offer more information to the crowd to make the process more prepared, committed, and transparent (Kunz et al.,2017). While there are some channels, such as descriptions and updates, for entrepreneurs to deliver messages to investors, the main channel for investors to communicate with entrepreneurs or other investors is the comment section. Investors can leave their opinions, suggestions, criticisms, or questions in the comment section. Entrepreneurs and other investors can contribute their ideas by posting replies on the comment page.

Comment quantity, length, and sentiment (Clauss et al., 2018; Courtney et al., 2017; Kromidha and Robson, 2016; Petitjean, 2018; Wang et al., 2018) have been found to be important determinants for crowdfunding success. Kromidha and Robson (2016) and Petitjean (2018) found that the number of comments is positively related to crowdfunding success. Wang et al. (2018) found that comment quantity and reply length have positive impacts on crowdfunding success, and comment sentiment positively moderates the effect of comment quantity on crowdfunding success. Clauss et al. (2018) and Courtney et al. (2017) found that positive sentiments have positive effects on fundraising success. Moreover, several studies (Lagazio and Querci, 2018; Majumdar and Bose, 2018; Parhankangas and Renko, 2017) suggested that lexical complexity, content words (noun, verb), persuasion, authenticity/uncertainty, and emotionality (affection) impact fundraising success. For instance, Majumdar and Bose (2018) found that the presence of authentic words increases the likelihood of receiving funding.

Besides these studies on comment-related factors influencing fundraising success, recently, features derived from comments are getting attention in studies on crowdfunding success prediction. Existing studies so far have mainly focused on the quantity, sentiment, or linguistic characteristics of comments. It has been shown that the number of comments is an important feature for success prediction (Chen et al., 2015; Li et al., 2006; Ryoba et al., 2020). Lai et al.

(2017) found that readability and sentiment of comments contribute to increase in prediction accuracy. Lee et al. (2018) showed that linguistic features of comments contain valuable information for predicting fundraising success.

Beyond such quantity, sentiment, and linguistic analysis, we focus on semantic analysis of comments in this study. Comments play dual roles to leverage fundraising success. They can decrease information asymmetry or trigger herding behavior to attract more interests from potential investors. Therefore, we posit that these two effects of comments may be valuable to crowdfunding success prediction. Furthermore, we posit that the effectiveness of comments posted by different actors may also be different. For example, in terms of decreasing information asymmetry, entrepreneurs' comments may be more effective than investors' since entrepreneurs know much better about their own projects, while herding among potential investors may be amplified when comments are posted by their peers. However, these aspects of comments have not been adequately captured in past studies yet. We strive to bridge these research gaps in this study.

## 2.3 The Proposed Framework

We propose a framework for fundraising success prediction, mining semantic features from comments. Rooted in relevant theories, we identify two types of comments, train classifiers for the two types, and derive *semantic intensities* based on the classification results. We also extract representative keywords of each type and derive *keyword features* accordingly. Furthermore, we separate the comments by *actor roles* to boost the effectiveness of these features. Finally, based on the semantic intensities and keyword features of distinct actor roles, in addition to basic project features, we build machine learning models to predict the success of a project at different fundraising stages. Figure 2.1 outlines the proposed framework. Below, we describe the main novelties in the framework, starting with the underlying rationales.

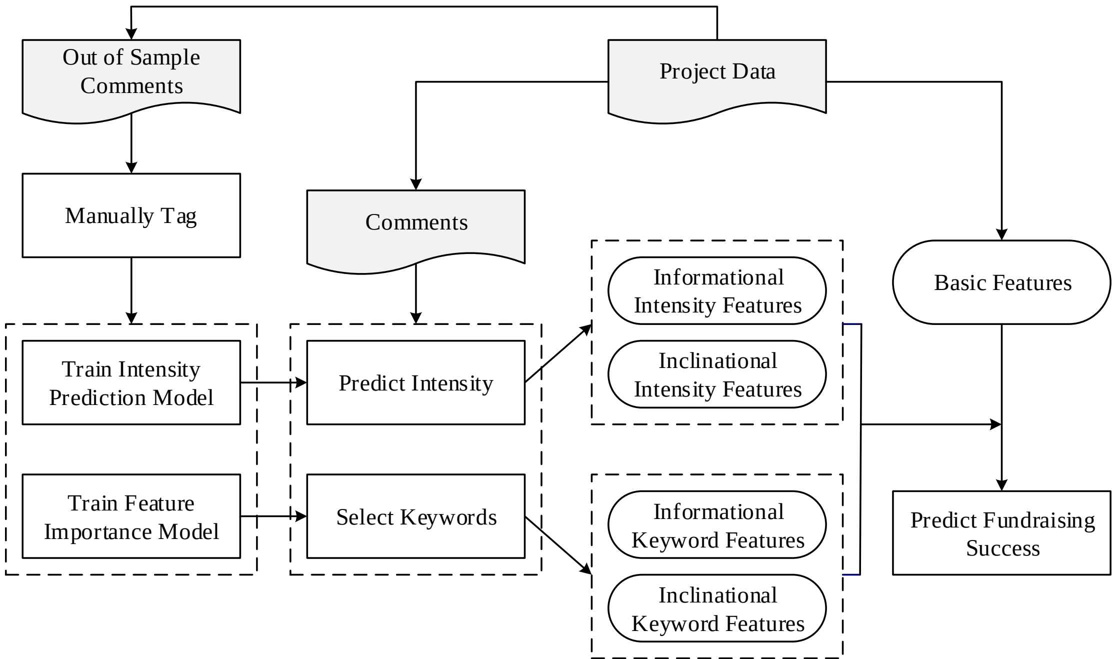



**Figure 2.1. The Proposed Framework**

### *2.3.1 Rationales*

By leaving comments, investors can ask questions on the functions or qualities of projects, while entrepreneurs can further explain their project progress and answer investors' questions. Comments contain a lot of signals on entrepreneurs' add-on information and investors' opinions or attitudes. Building on well-established theories, i.e., *information asymmetry* and *herding behavior*, we identify two aspects of comment semantics that may be useful for fundraising success prediction.

#### *2.3.1.1 Information Asymmetry in Crowdfunding*

The information asymmetry theory was first proposed by Akerlof (1970) to study asymmetric information in the car market, where sellers know better about car quality than buyers while buyers cannot effectively tell bad cars apart from good cars. Consequently, sellers with high-quality cars cannot sell better than average market prices. Spence (1973) further suggested that information asymmetry can be mitigated by signals, as people can transfer information to other parties to resolve the asymmetry. For example, “going to college” can be used as a credible signal of ability in learning when hiring new employees in the job market.

In crowdfunding, investors for a campaign can only get the product after the fundraising is over, and the delivery time may be as long as one year or more as it takes time for the idea to go from prototype to mass production (Wang et al., 2017). Moreover, almost anyone can initiate a campaign once passing the platform’s screening. The long delivery time and low entry barrier create difficulties for investors to assess product quality. Hence comes the most common problem in crowdfunding: information asymmetry. On the one hand, investors may not be able to assess the potential benefits and risks of crowdfunding projects accurately (Ahlers et al., 2015; Kuppuswamy and Bayus, 2018). On the other hand, it is also a challenge for entrepreneurs to transfer credible information on the potentials of their campaigns to investors. Therefore, decreasing the information asymmetry between entrepreneurs and investors is essential to achieve campaign success.

The main signals that can be judged are coming from project descriptions and communications between entrepreneurs and investors. Project descriptions are created solely by entrepreneurs, who may overstate the function and utility of their products to improve their pledges, inadvertently widening the information asymmetry. Conversely, as comments are created by both entrepreneurs and investors, the information is broader and more reliable. Hence, comments may contribute to decreasing the information asymmetry.

#### *2.3.1.2 Herding Behavior in Crowdfunding*

The herding behavior usually occurs in an environment with information asymmetry, where individuals may follow the actions of others when facing uncertainties in economic decision making (Croson and Shang, 2008; Liu et al., 2015; Simonsohn and Ariely, 2008). Herding is a prevalent phenomenon in crowdfunding since investors can observe other investors' behaviors online to guide their own decision making. Herding happens when a potential investor mimics others' decisions based on their descriptive social norms or follows well-funded campaigns (Burtch et al., 2015; Herzenstein, et al., 2011; Zhang and Liu, 2012). Two main types of herding behavior have been found in crowdfunding. First, potential investors may decide on their investment according to the popularity of a project, which may be measured by the number of current investors (Herzenstein et al., 2011; Lee and Lee, 2012; Liu et al., 2015). Second, investors may be more likely to fund projects that have more comments or social connections (i.e., electronic word-of-mouth) (Giudici et al., 2013; Kang et al., 2017; Kunz et al., 2017; Mollick, 2014). Moreover, herding is related to not only the number of investment actions or comments but also their shared attributes (Clauss et al., 2018). Therefore, we have reason to expect that the semantic signals conveyed by comments may also evoke herding behavior.

#### *2.3.1.3 Types of Comments*

Based on the theories discussed above, we distinguish two typical patterns of comments. One type of comments, which we call *informational comments*, describe or concern the characteristics, e.g., appearance, feature, function, reward, and delivery process, of a project. The other type of comments, which we call *inclinational comments*, convey attitudes or opinions toward a project or entrepreneur. Apparently, these two types of comments deliver different appeals to investors. Through Q&A between investors and the entrepreneur, informational comments provide more project-related signals which the entrepreneur may overlook in the project front page, updates, and FAQs. Through such comments, potential investors can get more accurate information and judge the concerns of other investors to make their decisions. Informational comments refine a project by reducing the information asymmetry. Inclinational comments may trigger herding behavior by revealing the inclination of investors. A potential investor may mimic the actions of current investors or be influenced by emotional appeals from the entrepreneur and current investors.

We provide two real examples to illustrate these two comment types. Figure 2.2 shows some of the informational comments on the electronic DIY product named “power Pi”, which is a power supply for Raspberry Pi. Potential backers (investors) asked questions on the model version, compatibility, utility, and shipping. The creator (entrepreneur) replied with answers. This information helps to reduce the information asymmetry effectively. The more comments with valuable information, the less uncertainty for backers about the project. The following backers can gain more confidence to make right decisions.

Figure 2.3 presents some of the inclinational comments on a project named “Cheat Sheets”, which designs cheat sheet sets for sheet pan cooking. Some comments express backers’ attitudes and opinions, e.g., “so existing”, “loved to see some colors”, “just backed you”, and “can’t wait to receive my set”.

A comment may contain both informational and inclinational signals. For example, the comment “I’d like purple honestly” delivered the backer’s inclination by the term “I’d like”, and meanwhile it suggested adding color purple to change the product’s appearance.

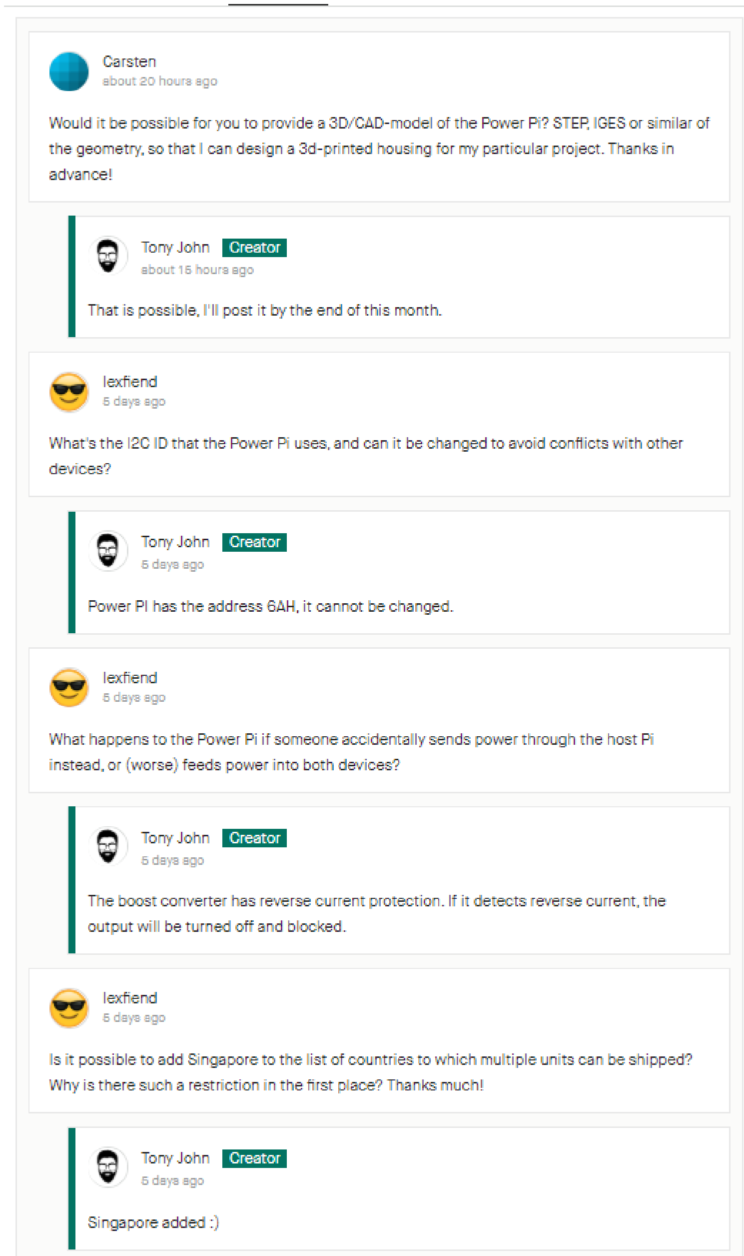



**Figure 2.2. An Example of Informational Comments**

These examples illustrate two different aspects of comment semantics, which may help potential investors make funding decisions. We therefore derive novel semantic features, including *intensities* and *keywords*, reflecting such two types of semantic signals from comments, for fundraising success prediction.

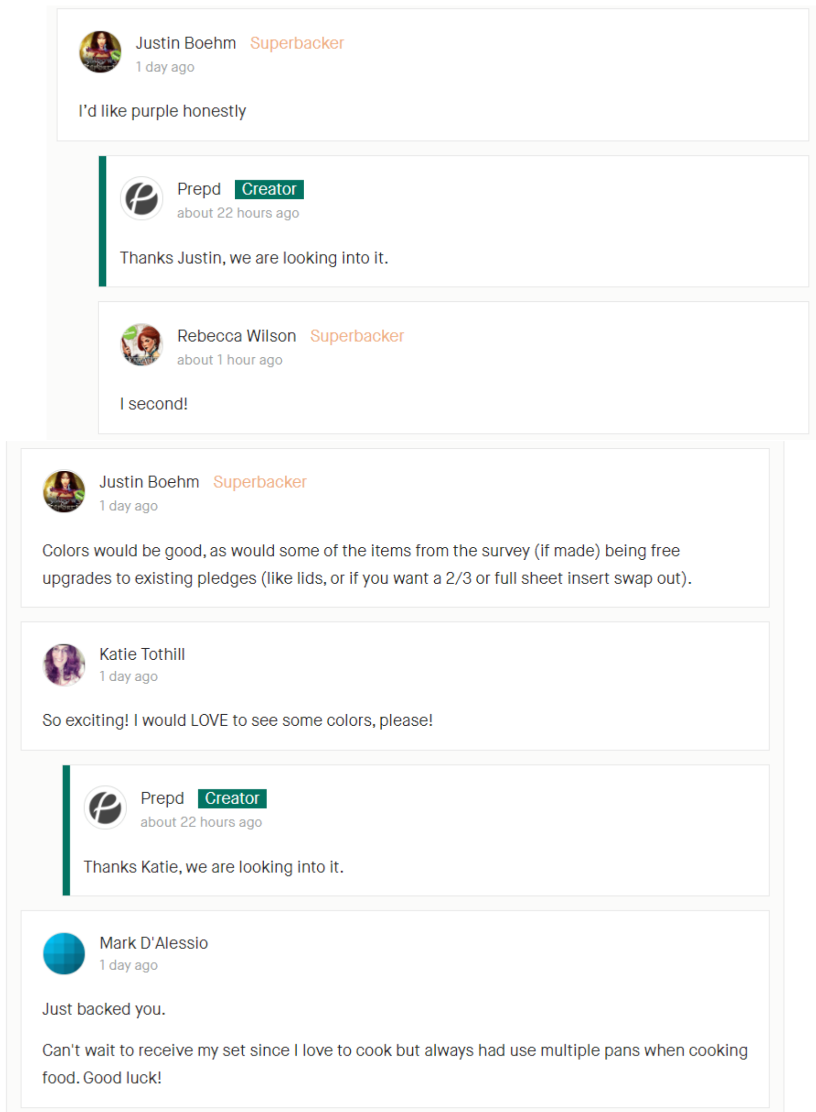



**Figure 2.3. An Example of Inclinational Comments**

### *2.3.2 Semantic Intensity Feature Extraction*

Semantic signals are hidden in comments, and crowdfunding platforms do not provide any direct information to classify the semantic types (i.e., informational and inclinational) of comments. Given the large number of comments on a crowdfunding platform, it is prohibitively costly to label all the comments manually. Thus, we propose a supervised learning approach for classifying the semantic types of comments.

Based on the theories discussed above, we first come up with a set of guidelines for classifying the comment types and have human experts manually label some comments (i.e., training data). Table 2.1 presents the guidelines that experts should follow when labeling the training comments. Afterward, we apply machine learning methods to train classifiers based on the labeled comments to predict the types of the remaining ones. Since some comments contain both types of semantic signals, we need to train two classifiers for predicting the two types, respectively. The prediction outcome of each classifier is a probability that a comment is of the target semantic type. Finally, we sum the predicted probabilities of each type of comments at the project level into an aggregate measure of *semantic intensity*. We used the sum, instead of average, of semantics of each project since the distribution of semantic features shows less predictive utility than the intensity of semantic features.

**Table 2.1. Guidelines for Labeling Comments**

| | |
|---|---|
| Informational | (1) Describing, commenting on, or asking about the appearance, quality, function, and features of the product; |
| | (2) Questions and replies on schedules, costs, rewards, shipping processes, and updates. |
| Inclinational | (1) Expressing interests, wishes, praises, appreciations, suspicions, and concerns toward the project or the entrepreneur; |
| | (2) Mentioning that the backer has already backed (pledged) the project or has backed previous projects by the same entrepreneur. |

### *2.3.3 Keyword Feature Extraction*

Besides the semantic intensity features, we also derive keyword features reflecting semantics in the two types of comments. For example, as informational comments are related to appearances, functionalities, etc., they are more likely to contain such words as “color” and “shipping”, whereas emotional words, such as “love” and “excited”, appear more frequently in inclinational comments.

We posit that such representative words also convey semantic signals.

We select a set of representative words (keywords) for each semantic type (informational comments or inclinational comments) based on the importance scores of the classifier. A keyword may be positively or negatively related to the semantic type. We count the occurrences of the positive (or negative) keywords in each comment and then sum the keyword occurrences in all the comments for a project into a positive (or negative) keyword feature, for each semantic type.

### *2.3.4 Prediction*

We apply the proposed semantic intensity and keyword features, in addition to some basic features, in building success prediction models. We further distinguish the actor roles (investors vs. entrepreneurs) in deriving the proposed features. Considering that the comments from investors and entrepreneurs play different roles in attracting funding, as discussed earlier, we posit that distinguishing the actor roles may help to improve predictive performance. Also, the predictive utilities of the semantic features based on different actor roles may evolve in different ways over time. Finally, considering the dynamic characteristics of comments, we predict fundraising success at different stages over the fundraising period.

## 2.4. Empirical Evaluation

### *2.4.1 Data*

We have evaluated our proposed framework at Kickstarter. Since Kickstarter was launched on April 28, 2009, “18 million people have backed a project, $5.2 billion has been pledged, and 185,417 projects have been successfully funded”, according to its website. Projects fall into 15 categories: arts, comics, crafts, dance, design, fashion, film & video, food, games, journalism, music, photography, publishing, technology, and theater. An entrepreneur (creator) can create a project page with a basic description and an optional video to explain the story behind the project, and people who are interested in the project (backers) can pledge the project. Throughout the fundraising period, the creator and backers can leave comments on the platform as further discussion. Kickstarter uses the all-or-nothing funding model, i.e., the creator will get funding only if the fundraising reaches its goal.

We developed web crawlers to collect data about all projects that started in 2017. The data include basic information (e.g., fundraising goal and duration) on the front page, project description, FAQ, updates, and comments for each project. We kept completed (successful or failed) projects in our evaluation and eliminated uncompleted (suspended or canceled) projects. There are 19,255 completed projects in total, 10,799 of which succeeded. Among the completed projects, 9,539 (7,083 succeeded) have comment postings. Since our research focuses on the utility of comments, we selected these 9,539 projects with a total of 548,092 comments for evaluation. Our proposed method is intended to improve predictive performance on projects with comments only. For projects without any comment, a reduced model without comment-related features would be used.

The data preprocessing mainly focused on comment text since our proposed semantic features are derived from comments. The distribution of comments may vary over time throughout the fundraising period. We divided the fundraising period of each project equally into four quarters and assigned each comment to a quarter based on its posting time. We predicted the fundraising outcomes at three time points, using comments posted during the first quarter (Q1), the first two quarters (Q1-2), and the first three quarters (Q1-3) of the fundraising period, respectively (we did not use the last quarter since it is meaningless for prediction). We also divided the comments based on the roles of their posters (i.e., creators or backers). Hence, we generated six datasets for comments according to these three fundraising stages and two actor roles. For each comment, we performed some standard preprocessing steps, e.g., removing punctuations, standardizing (upper or lower) case, and removing sparse terms (occurrence frequency below 0.5%). For each project, we generated a document-term matrix, in which each element $a_{ij}$ is the number of occurrences of word $j$ in comment $i$.

### *2.4.2 Feature Extraction*

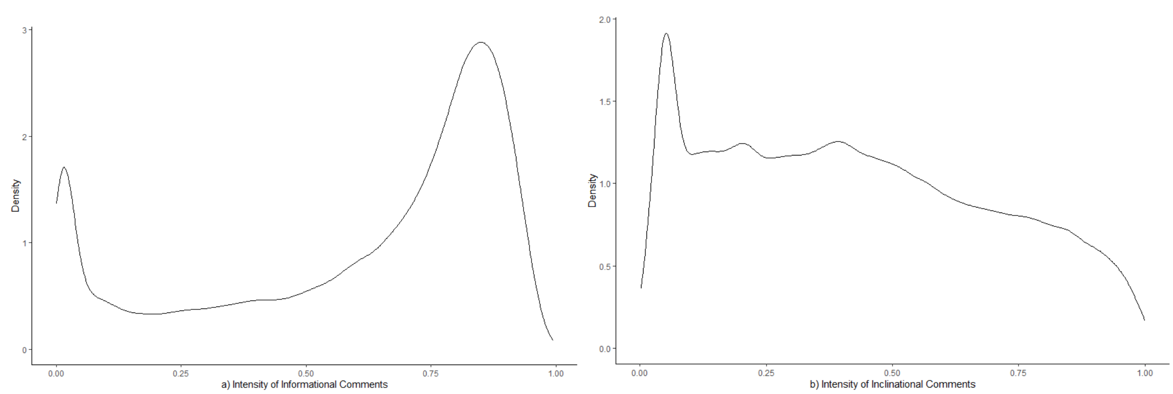
Our proposed semantic features consist of the intensities, as well as representative keywords, of informational and inclinational comments, respectively. To prepare training data, two PhD students in information systems manually tagged 5,000 randomly sampled comments following the aforementioned criteria (Table 1). Each comment may be tagged as being informational, inclinational, both informational and inclinational, or neither informational nor inclinational. The two taggers tagged the comments independently and then conducted a second-round discussion to reach an agreement on any conflict.

We applied four standard classification methods, i.e., logistic regression (LR), K-nearest neighbors (KNN), decision tree (DT), and random forest (RF), to train two classifiers (for informational comments and inclinational comments, respectively) based on the tagged comments. We evaluated the performance using ten independent rounds of ten-fold cross-validations. During each ten-fold cross validation, the dataset was divided into ten equal-sized subsets (folds). Each fold was used to estimate the performance of the classifier trained on the other nine folds. The splitting of folds was kept identical across classification methods. Overall, RF performed the best among the methods (Table 2.2), in terms of AUC, Kolmogorov–Smirnov (KS) statistic, and H measure (Hand, 2009). We therefore chose to use the RF classifiers to predict the types of the remaining comments. For the reliability of performance evaluation, we excluded the projects with the manually labeled comments in the subsequent experiments. For each of the remaining projects used for evaluation, we summed the predicted probabilities of its comments for the two types, respectively, as semantic intensity features (Figure 2.4 shows the density of comments over the intensity for each type).

**Table 2.2. Performance of Comment Type Classification**

| Classification Method | Informational | | | Inclinational | | |
|---|---|---|---|---|---|---|
| | AUC | KS | H | AUC | KS | H |
| LR | 0.899 (0.013) | 0.682 (0.029) | 0.555 (0.037) | 0.841 (0.018) | 0.571 (0.033) | 0.439 (0.039) |
| KNN | 0.785 (0.018) | 0.483 (0.033) | 0.309 (0.035) | 0.774 (0.016) | 0.433 (0.030) | 0.295 (0.028) |
| DT | 0.754 (0.022) | 0.488 (0.040) | 0.310 (0.046) | 0.769 (0.019) | 0.506 (0.038) | 0.335 (0.041) |
| RF | 0.912 (0.012) | 0.679 (0.030) | 0.572 (0.032) | 0.889 (0.015) | 0.655 (0.031) | 0.538 (0.038) |

Standard deviations are enclosed in parentheses.




a) Informational Comments

b) Inclinational Comments

**Figure 2.4. Densities of Comment Types**

We also selected the keywords for informational and inclinational comments as we found that the frequently used words in the two types of comments were different. The top 50 keywords (excluding the English stop words in the R package “tm”) for each type were selected based on the importance scores given by random forest (the importance score of a feature is the average loss of GINI entropy when the observed values of the feature are randomly permuted in out-of-bag samples). In addition, we determined the sign (positive or negative) of each keyword for each type based on the weight of evidence (Agterberg, 1992) and cross-checked the relationship (i.e., sign) with the coefficient of logistic regression (no conflict was found). We then counted the number of positive (negative) keywords appearing in the informational (inclinational) comments of a project as a semantic keyword feature, forming 2 (semantic types) * 2 (signs) = 4 keyword features in total.

To evaluate the effectiveness of our proposed semantic features, we compared them with typical features that have been studied before, i.e., sentiment features and linguistic features (Kaminski and Hopp, 2019; Lai et al., 2017; Lee et al., 2018; Wang et al., 2018). For sentiment features, we used SentiStrength to estimate the sentiment of each comment and then summed the positive (negative) sentiment strength scores (between 1 and 5) of the comments of each project. For linguistic features, we used stylistic features and lexical features (Abrahams et al., 2015). The stylistic features include readability (specifically, the Flesch Reading Ease Readability) and the numbers of characters, sentences, and words. The lexical features include the numbers of nouns, verbs, adjectives, and adverbs.

### *2.4.3 Classification Method Selection*

Before our main experiments, we evaluated four standard classification methods (i.e., LR, KNN, DT, and RF) using a set of basic features (summarized in Table 2.3), including ex-ante features and quantities of FAQs, updates, and comments. The ex-ante features are based on information provided by the creator before a project is launched, including how much pledges to fund (goal), how long the fundraising takes (duration), project description, category, the numbers of images and videos used to describe the project, and social media (Facebook, Twitter, and Instagram) connections. These characteristics have been shown to play a principal role in attracting the backers' interests and can be used to predict the fundraising outcome of a project. In particular, description length (Zhou et al., 2018), images/videos (Mollick, 2014), and social network connections (Kunz et al., 2017) have been shown to positively impact crowdfunding success, whereas duration and goal (Kunz et al., 2017) have been shown to negatively impact crowdfunding success. Besides, it has been shown that more communication leads to a higher probability for a project to get funded. In particular, the probability of project success increases as the numbers of comments (Kromidha and Robson, 2016), FAQs, and updates (Kunz et al., 2017) increase.

**Table 2.3. Basic Features**

| Feature | Description | Summary statistics | | | |
|---|---|---|---|---|---|
| | | Mean | SD | Max | Min |
| Duration | The fundraising duration (in days) of the project | 31.36 | 10.65 | 83.98 | 1.03 |
| Description length | The number of words in the project description | 712.00 | 591.30 | 4721 | 0 |
| Goal | The total amount of pledges (in dollars) the creator wishes to get | 34296.00 | 506709.80 | 33000000 | 1 |
| Image count | The number of images on the home page of the project | 13.93 | 14.52 | 153 | 0 |
| Video count | The number of videos on the home page of the project | 1.05 | 1.06 | 24 | 0 |
| FAQ count | The number of frequently asked questions | 1.34 | 3.78 | 121 | 0 |
| Update count (Q1) | The number of updates | 1.84 | 2.14 | 26 | 0 |
| Update count (Q1-2) | | 2.93 | 3.18 | 41 | 0 |
| Update count (Q1-3) | | 3.75 | 4.08 | 48 | 0 |
| Comment count (Q1) | The number of comments | 10.19 | 41.07 | 1297 | 1 |
| Comment count (Q1-2) | | 12.73 | 59.59 | 2131 | 1 |
| Comment count (Q1-3) | | 14.92 | 76.54 | 3062 | 1 |
| Category | The category of the project | Art: 9.03%; Comics: 10.87%; Crafts: 4.19%; Dance: 0.96%; Design: 2.52; Fashion: 10.54%; Film: 7.89%; Food: 7.31%; Games: 9.39%; Journalism: 0.96%; Music: 5.71%; Photography: 3.06%; Publishing: 12.07%; Technology: 13.43%; Theater: 2.07% | | | |
| Facebook connection | Whether the project connects to Facebook | Yes: 33.00%; No: 67.00% | | | |

| | | |
|---|---|---|
| Twitter connection | Whether the project connects to Twitter | Yes: 22.72%; No: 77.28% |
| Instagram connection | Whether the project connects to Instagram | Yes: 26.37%; No: 73.63% |

**Table 2.4. Predictive Performance**

| Method | Metric | Q1 | Q1-2 | Q1-3 |
|---|---|---|---|---|
| LR | AUC | 0.802 (0.024) | 0.799 (0.025) | 0.790 (0.022) |
| | KS | 0.481 (0.046) | 0.476 (0.050) | 0.465 (0.041) |
| | H | 0.291 (0.043) | 0.287 (0.045) | 0.262 (0.040) |
| KNN | AUC | 0.707 (0.029) | 0.704 (0.031) | 0.699 (0.025) |
| | KS | 0.344 (0.051) | 0.334 (0.050) | 0.325 (0.040) |
| | H | 0.154 (0.037) | 0.150 (0.039) | 0.141 (0.032) |
| DT | AUC | 0.679 (0.031) | 0.693 (0.031) | 0.711 (0.031) |
| | KS | 0.328 (0.053) | 0.343 (0.056) | 0.380 (0.058) |
| | H | 0.173 (0.045) | 0.170 (0.048) | 0.206 (0.047) |
| RF | AUC | 0.818 (0.021) | 0.820 (0.022) | 0.820 (0.020) |
| | KS | 0.499 (0.044) | 0.499 (0.041) | 0.499 (0.038) |
| | H | 0.333 (0.041) | 0.327 (0.042) | 0.325 (0.041) |

Standard deviations are enclosed in parentheses.

Table 2.4 summarizes the performance (in terms of AUC, KS, and H measure, estimated through ten independent rounds of ten-fold cross-validations) of each classification method based on the basic features. Random forest outperformed other methods in terms of all three performance metrics ($p < 0.01$). Thus, we selected random forest as the classification method in the subsequent experiments.

### *2.4.4 Experiments*

We conducted a series of experiments. As our goal is to predict, rather than to explain (Shmueli, 2010), we focus on predictive performance. First, we added our proposed semantic features, with/without distinguishing actor roles, to the basic feature set to assess the effects of adding the semantic features and distinguishing actor roles on predictive performance. Second, we also examined the effect of each actor role separately. Third, we compared our proposed approach with a benchmark using sentiment and linguistic features as a representative of existing approaches.

We ran all of the experiments at multiple time points (i.e., Q1, Q1-2, and Q1-3), so we can also analyze the dynamics of the effects of various factors over the fundraising period of a project. For every setting, we estimated the predictive performance in terms of AUC, KS, and H measure through ten independent rounds of ten-fold cross-validations.

## 2.5 Results

### *2.5.1 Effectiveness of the Proposed Semantic Features and Distinguishing Actor Roles*

To evaluate the effects of the proposed semantic features and distinguishing actor roles, we trained and tested three models, using the basic features (Table 2.3) only, basic features plus the semantic features without distinguishing actor roles, and basic features plus the semantic features while distinguishing actor roles, referred to as the "Baseline" model, the "Semantic" model, and the "Actor Roles" model (i.e., the proposed model), respectively. Table 2.5 summarizes the predictive performance results, and Table 2.6 summarizes the results of Friedman test and Dunn's pairwise post hoc test. The Actor Roles model outperformed the Semantic model, which outperformed the Baseline model, in terms of all performance metrics across all three quarters. This shows that our proposed semantic features contributed to predictive performance improvement and separating actor roles further amplified the effectiveness of the semantic features.

**Table 2.5. Predictive Performance of the Baseline, Semantic, and Actor Roles Models**

| Metric | Baseline | | | Semantic | | | Actor Roles | | |
|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q1-2 | Q1-3 | Q1 | Q1-2 | Q1-3 | Q1 | Q1-2 | Q1-3 |
| AUC | 0.818<br>(0.021) | 0.820<br>(0.022) | 0.820<br>(0.020) | 0.836<br>(0.021) | 0.843<br>(0.020) | 0.851<br>(0.019) | 0.839<br>(0.023) | 0.846<br>(0.021) | 0.855<br>(0.020) |
| KS | 0.499<br>(0.044) | 0.499<br>(0.041) | 0.499<br>(0.038) | 0.536<br>(0.043) | 0.549<br>(0.043) | 0.562<br>(0.040) | 0.540<br>(0.043) | 0.555<br>(0.045) | 0.574<br>(0.038) |
| H | 0.333<br>(0.041) | 0.327<br>(0.042) | 0.325<br>(0.041) | 0.353<br>(0.041) | 0.370<br>(0.041) | 0.384<br>(0.043) | 0.361<br>(0.044) | 0.379<br>(0.043) | 0.392<br>(0.043) |

Standard deviations are enclosed in parentheses.

**Table 2.6. Friedman Test and Post Hoc Dunn Test on the Baseline, Semantic, and Actor Roles Models**

| | | | *p*-value of pairwise comparison | |
|---|---|---|---|---|
| | | Average rank | Baseline | Semantic |
| Q1 | Baseline | 2.41 | | |
| | Semantic | 1.92 | <.001 | |
| | Actor Roles | 1.67 | <.001 | 0.007 |
| | Friedman $\chi^2$: 86.327 ($p < .001$) | | | |
| Q1-2 | Baseline | 2.86 | | |
| | Semantic | 1.69 | <.001 | |
| | Actor Roles | 1.45 | <.001 | <.001 |
| | Friedman $\chi^2$: 341.460 ($p < .001$) | | | |
| Q1-3 | Baseline | 2.95 | | |
| | Semantic | 1.74 | <.001 | |
| | Actor Roles | 1.31 | <.001 | <.001 |
| | Friedman $\chi^2$: 433.007 ($p < .001$) | | | |

### *2.5.2 The Effectiveness of Each Actor Role*

To further examine which actor role is more effective in predicting fundraising success and how the effects of different actor roles change over the fundraising period, we experimented with two additional models by applying semantic features based on creators' and backers' comments only, referred to as the "Creator" model and "Backer" model, respectively. Table 2.7 summarizes the predictive performance results, and Table 2.8 summarizes the results of Friedman test. In the first quarter (Q1) and second quarter (Q1-2), the Backer model significantly outperformed the Creator model. In the third quarter (Q1-3), the Creator model significantly outperformed the Backer model. For both models, the predictive performance improved over time, but in different degrees. In Q1, the Backer model performed notably better than the Creator model. In Q1-2, the Backer model still performed better than the Creator model, but the difference became smaller.

Finally, in Q1-3, the Creator model even outperformed the Backer model. This shows that the predictive utility of the creator's comments increased faster, relative to the backers' comments, as the fundraising proceeded.

**Table 2.7. Predictive Performance of the Creator and Backer Models**

| Metric | Creator | | | Backer | | |
|---|---|---|---|---|---|---|
| | Q1 | Q1-2 | Q1-3 | Q1 | Q1-2 | Q1-3 |
| AUC | 0.833<br>(0.022) | 0.843<br>(0.021) | 0.855<br>(0.019) | 0.842<br>(0.022) | 0.849<br>(0.021) | 0.852<br>(0.019) |
| KS | 0.532<br>(0.040) | 0.557<br>(0.044) | 0.571<br>(0.037) | 0.541<br>(0.041) | 0.560<br>(0.043) | 0.564<br>(0.038) |
| H | 0.354<br>(0.043) | 0.376<br>(0.043) | 0.394<br>(0.043) | 0.363<br>(0.043) | 0.379<br>(0.043) | 0.384<br>(0.043) |

Standard deviations are enclosed in parentheses.

**Table 2.8. Friedman Test on the Creator and Backer Models**

| | Average rank | |
|---|---|---|
| | Creator | Backer |
| Q1 | 1.68 | 1.32 |
| | Friedman $\chi^2$: 37.705 ($p < .001$) | |
| Q1-2 | 1.63 | 1.37 |
| | Friedman $\chi^2$: 20.280 ($p < .001$) | |
| Q1-3 | 1.28 | 1.72 |
| | Friedman $\chi^2$: 60.255 ($p < .001$) | |

### *2.5.3 Comparison with the Benchmark Model*

Some previous studies (Kaminski and Hopp, 2019; Lai et al., 2017; Lee et al., 2018; Wang et al., 2018) have demonstrated the effects of text analysis on predicting crowdfunding success. They usually concentrated on linguistic features and sentiment features. To examine the effectiveness of our proposed semantic features relative to these linguistic and sentiment features used in previous studies, we compared the proposed (i.e., Actor Roles) model with another model (referred to as "Benchmark" model) using the linguistic and sentiment features instead of our proposed semantic features. For completeness, we also tested another model combining all the features in the Benchmark and Actor Roles models, referred to as the "Combined" model. Table

2.9 summarizes the predictive performance results, and Table 2.10 summarizes the results of Friedman test and Dunn's pairwise post hoc test. The Actor Roles model outperformed the Benchmark model in terms of all performance metrics across all three quarters, showing that our proposed semantic features are more effective than the linguistic and sentiment features. The Combined model did not perform better, and even performed worse sometimes (in Q2 and Q3), than the Actor Roles model, showing that linguistic and sentiment features may not contribute to further performance improvement over and above our proposed semantic features and may even cause some overfitting.

**Table 2.9. Predictive performance of the Actor Roles, Benchmark, and Combined models**

| Metric | Benchmark | | | Actor Roles | | | Combined | | |
|---|---|---|---|---|---|---|---|---|---|
| | Q1 | Q1-2 | Q1-3 | Q1 | Q1-2 | Q1-3 | Q1 | Q1-2 | Q1-3 |
| AUC | 0.818<br>(0.022) | 0.815<br>(0.024) | 0.821<br>(0.021) | 0.839<br>(0.023) | 0.846<br>(0.021) | 0.855<br>(0.020) | 0.831<br>(0.023) | 0.838<br>(0.022) | 0.848<br>(0.020) |
| KS | 0.508<br>(0.040) | 0.504<br>(0.049) | 0.511<br>(0.045) | 0.540<br>(0.043) | 0.555<br>(0.045) | 0.574<br>(0.038) | 0.523<br>(0.046) | 0.544<br>(0.042) | 0.565<br>(0.039) |
| H | 0.330<br>(0.039) | 0.323<br>(0.044) | 0.330<br>(0.043) | 0.361<br>(0.044) | 0.379<br>(0.043) | 0.392<br>(0.043) | 0.350<br>(0.045) | 0.363<br>(0.043) | 0.386<br>(0.045) |

Standard deviations are enclosed in parentheses.

**Table 2.10. Friedman Test and Post Hoc Dunn Test on the Actor Roles, Benchmark, and Combined Models**

| | | | *p*-value of pairwise comparison | |
|---|---|---|---|---|
| | | Average rank | Benchmark | Actor Roles |
| Q1 | Benchmark | 2.47 | | |
| | Actor Roles | 1.68 | <.001 | |
| | Combined | 1.85 | <.001 | 0.107 |
| | Friedman $\chi^2$: 106.040 ($p$ < .001) | | | |
| Q1-2 | Benchmark | 2.84 | | |
| | Actor Roles | 1.25 | <.001 | |
| | Combined | 1.91 | <.001 | <.001 |

| | | | | |
|---|---|---|---|---|
| | Friedman $\chi^2$: 384.772 ($p$ < .001) | | | |
| Q1-3 | Benchmark | 2.94 | | |
| | Actor Roles | 1.29 | <.001 | |
| | Combined | 1.77 | <.001 | <.001 |
| | Friedman $\chi^2$: 432.180 ($p$ < .001) | | | |

### *2.5.4 Dynamic Predictive Utility of Comments*

Entrepreneurs and investors would like to continuously track the evolution of projects throughout the fundraising periods and adjust their decisions/actions accordingly. Comments for a project are accumulating over the fundraising period, providing more and more information for predicting the fundraising outcome. We therefore examined the evolution of predictive performance over time (specifically, over the first three quarters). In this regard, we compared the basic features, sentiment and linguistic features used in previous studies, and our proposed semantic features (i.e., the Baseline, Benchmark, and Actor Roles models). Table 2.11 summarizes the results of Friedman test and Dunn's pairwise post hoc test.

Surprisingly, there were no significant improvements over the three quarters for the Baseline model, and even for the Benchmark model, there was no significant improvement from Q1 to Q1-2 and only marginally significant ($p$ > .01) improvement from Q1-2 to Q1-3. On the other hand, the predictive performance of the Actor Roles model continuously improved significantly over time (from Q1 to Q1-2 to Q1-3). Intuitively, as the fundraising proceeds, the performance of a prediction model should gradually improve due to the gaining of more information. However, we did not observe an apparent trend in the Baseline model and early stages of the Benchmark model. This implies that the improvement in predictive utility throughout the fundraising period may be minor for the quantity, sentiment, and linguistic features of comments but considerable for our proposed semantic features. Therefore, only counting the quantity or simply relying on sentiment and linguistic information may not be enough, and examining the semantics is important, to fully capture the dynamic predictive utility of comments.

**Table 2.11. Friedman Test and Post Hoc Dunn Test over Fundraising Period**

| | | | *p*-value of pairwise comparison | |
|---|---|---|---|---|
| | | Average rank | Q1 | Q1-2 |
| Baseline | Q1 | 1.94 | | |
| | Q1-2 | 1.98 | 0.672 | |
| | Q1-3 | 2.08 | 0.224 | 0.934 |
| | Friedman $\chi^2$: | 2.927 ($p$ = 0.231) | | |
| Benchmark | Q1 | 2.07 | | |
| | Q1-2 | 2.08 | 1.000 | |
| | Q1-3 | 1.85 | 0.024 | 0.013 |
| | Friedman $\chi^2$: | 10.167 ($p$ < .001) | | |
| Actor Roles | Q1 | 2.30 | | |
| | Q1-2 | 2.00 | 0.001 | |
| | Q1-3 | 1.70 | <.001 | 0.001 |
| | Friedman $\chi^2$: | 55.207 ($p$ < .001) | | |

## 2.6. Discussion

Our empirical evaluation yields some interesting findings. First and foremost, our results show that adding semantic features of comments helped to improve predictive performance, thus shedding light on the important role of comment semantics in prediction. This finding confirms that signals inducing herding behavior or reducing information asymmetry indeed have predictive value. Moreover, distinguishing actor roles of comments amplified the effectiveness of comment semantics in prediction, revealing that investors (backers) and entrepreneurs (creators) express different semantics through comments. It is therefore useful to study the predictive effects of different actor roles separately.

By tracking the dynamic effects of comments, our evaluation shows that the semantic signals led to significant improvement in predictive performance throughout the whole fundraising period, and this improvement kept growing over time. This finding implies that as fundraising proceeds, the semantic signals gradually accumulate and hence become increasingly more valuable for predicting fundraising success.

With regard to the predictive utilities of actor roles, it is interesting that comments from entrepreneurs and those from investors showed different trends over fundraising stages. Our results highlight the important role of investors' comment semantics in an early stage and the value of entrepreneurs' comment semantics in a later stage. This finding implies that comments of investors contain more semantic information that is valuable to predict crowdfunding success early on, while comments of entrepreneurs become more effective later on.

Our study has implications for both research and practice. For research, our study brings comment semantics to attention, contributes a new way to discover the semantics of comments, and demonstrates the usefulness of semantic features in fundraising success prediction. In addition, our study demonstrates the amplified effects of semantic features on separate actor roles, which is ignored in previous research. Finally, our study reveals the dynamic effects of different actor roles in different stages of fundraising periods.

For practice, our findings suggest to entrepreneurs that identifying the semantics of comments can be very beneficial for improving their success probabilities across the fundraising period. They should generate comments containing more information about their projects to attract potential investors' attention. As for investors, our study provides an exemplar for forecasting the success of projects based on comments, thereby saving their time and reducing opportunity costs. Potential investors should pay attention to the semantic signals of investors and entrepreneurs to make better funding decisions. Finally, our study also offers practical tips for crowdfunding platforms to improve their algorithms for recommending promising projects to potential investors.

## 2.7 Conclusion

As a growing number of participants are willing to communicate on crowdfunding platforms, entrepreneurs, investors, and platforms are hoping to find patterns of comments to predict fundraising outcomes. Previous studies have mainly focused on the quantity, sentiment, and linguistic characteristics of comments, largely overlooking the value of semantic features. Rooted in information asymmetry and herding behavior, we have proposed a framework to predict fundraising success by identifying novel latent semantic features of comments from different actors. Our empirical evaluation demonstrates the utility of the proposed framework. Our study makes several contributions to research and has managerial implications for various stakeholders.

Our study opens up several avenues for future research. First, while we developed and evaluated the semantic features specifically in fundraising success prediction, their usefulness in other related problems, e.g., funding amount prediction (Zhao et al., 2017) and fraudulent project detection (Siering et al., 2016), may be evaluated in the future. Second, while we only examined comments on the crowdfunding platform itself, future research may further extend to comments from other related sources, e.g., social media. Third, while we proposed some features reflecting signals that may induce herding behavior or reduce information asymmetry, other semantics, e.g., topics mentioned in comments, may be explored in future research. Fourth, while we distinguished comments by the actor roles, future research may further incorporate the interactions between entrepreneurs and investors, for example, whether investors' comments have been responded to promptly and appropriately by entrepreneurs.

# CHAPTER 3

# Essay 2: Reward-based Crowdfunding Success Prediction with Multimodal Data

## 3.1 Introduction

As we enter the information age, an increasing number of websites, including e-Commerce and share economy websites, are incorporating multimodal data to enhance data diversity and attract users' attentions. Crowdfunding platforms, as a form of online microfinance, also recommend entrepreneurs post multimodal data to attract investors. On the one hand, potential investors are unlikely to receive consultations from financial institutions on evaluating, monitoring, and managing risks (Huang and Knight, 2017; Kim and Viswanathan, 2019), and hence rely heavily on the information posted on crowdfunding websites. The incorporation of multimodal information can enhance the trustworthiness of fundraising projects. On the other hand, multiple modalities can convey multi-level characteristics of products or services, including colors, movements, feelings, and emotions, to potential investors more effectively than plain texts. This expansion of product or service can help potential investors make better informed decisions. Hence, it becomes necessary to leverage multimodal data in crowdfunding.

The impact of description texts and the presence and number of images and videos have been extensively studied in crowdfunding (Carradini and Fleischmann, 2022; Mollick, 2014; Yang et al., 2020; Zhou et al., 2018). However, previous research has largely overlooked the influence of visual features of images and videos on the fundraising outcome. Visual features attract more attention from the readers (Riegelsberger, 2002b) and are more easily processed than plain texts

(Chan and park, 2015). In an online environment, visual features can increase the text credibility and evoke emotional appeal in readers (Garrett, 2002), which may lead to favorable behaviors. For example, colors can influence the favorability of venture investment decisions (Chan and Park, 2015), and hedonic features of images can induce a pleasant attitude toward an online store (Fiore et al., 2005). Similarly, facial features of images can enhance the appeal and trustworthiness of e-commerce websites (Riegelsberger, 2002a; Riegelsberger, 2003; Cyr et al., 2009), which may in turn improve the outcome for e-Commerce. Reward-based crowdfunding shares many similar characteristics to e-Commerce, where stores or an entrepreneur would post description texts, images, or videos to introduce their products or services, and customers cannot receive the actual products or services before the payments. Therefore, we posit that visual features may also impact the fundraising outcome, and multimodal representations incorporating visual features may improve the predictive effectiveness of fundraising outcomes.

To provide a comprehensive analysis of linguistic and visual features of texts, images, and videos in crowdfunding, we propose a framework that is built on three earlier frameworks specifically designed for analyzing languages and images: Halliday's metafunctions framework of languages (1985), Kress and Van Leeuwen's functional visual design (1996), and Royce's intersemiotic complementarity of languages and visual images (1998). As the theory behind linguistics is functional rather than formal, Halliday first defined "three kinds of meaning that are embodied in human language as a whole, forming the basis of the semantic organization of all natural languages" (Halliday, 1985). These are metafunctions, namely, ideational, interpersonal, and textual metafunctions. Ideational metafunction focuses on the content of discourse and represents the activity sequences of processes, participants in the processes, and circumstances associated with the process (Halliday, 1985). Interpersonal is described as "the speaker or writer doing something to the listener or reader by means of language" and displays the social relationships and values between speakers and listeners or writers and readers (O'Halloran, 2008b; Royce, 2013). Textual metafunction shows the interweaving of ideational and interpersonal choices into cohesive and coherent units of meaning (Halliday, 2004).

Besides the application of metafunctions in linguistic communication, some works have been built up on visual modes, to explain what features make multimodal text visually-verbally coherent. Kress and Van Leeuwen (1996) suggested that visual images fulfill the metafunctions of representation of experiential world, the interaction between participants and viewers, and the compositional arrangement of visual resources. Royce (1998) identified the intersemiotic complementarity of language and visual images, where "visual and verbal modes semantically complement each other to produce a single textual phenomenon". He introduced three elements, represented participants, interactive participants, and visual compositional features (Royce, 2013), which correlate with Halliday's three metafunctions. The represented participants are all the elements or entities that are actually present in the visual, whether animate or inanimate. The interactive participants are the participants who are interacting with each other in the act of reading a visual, which represents the social relations between the viewer and the visual (Royce, 2013; Kress and van Leeuwen, 1996). Compositional features capture more fully the sense of two modes interacting with each other to provide coherent intersemiotic messages.

Building on these theories, we propose a framework to analyze the metafunctions of multimodal data in reward-based crowdfunding platforms, to investigate the following research questions: (1) whether each metafunction of each modality is valuable for predicting fundraising outcome; (2) whether multimodality improves the predictive performance over single modality in terms of each metafunction; (3) whether the combination of metafunctions improves the predictive performance over a single metafunction.

To address the first research question, we identify and extract representation features of ideational, interpersonal, and textual metafunctions from text, image, and video modalities in crowdfunding. We then build prediction models to assess the predictive capability of each metafunction separately. In response to the second research question, we investigate the interactions between multiple data modalities, in terms of each metafunction, by comparing the predictive performances of single modality and multimodality models. Coping with the third research question, we examine the interaction between three metafunctions by building predictive models on all possible combinations of metafunctions for multimodal data.

## 3.2 Literature review

### *3.2.1 Systemic Functional Linguistic and Metafunctions*

Systemic Functional Linguistic (SFL) was first developed by linguist Michael Halliday in the early 1960s as a tool for teaching Mandarin (Halliday, 1985) and later extended to the English language (Halliday, 1994). It serves as the central theoretical framework for systemic functional approaches to multimodality, which analyze the function and meaning of semiotic resources (O'Halloran, 2008a). Semiotic resources are theorized as realizing three different meaning functions, known as metafunctions. This meta-functional system is used to interpret how semiotic resources simultaneously construct experiences and logic (ideational meaning), enact social relations (interpersonal meaning), and organize a structured text (textual meaning).

Drawing on the insights of Halliday, researchers have gradually extended the systemic function theory to non-verbal semiotic resources and media, such as visual design (Kress and van Leeuwen, 1996), advertisement (O'Halloran, 2008b), website (Djonov, 2005; Bateman, 2008), and music (Van Leeuwen, 1999). Kress & van Leeuwen (1996) argued that the same metafunctions used in linguistic analysis can be applied to visual resources. The three metafunctions—ideational meaning, interpersonal meaning, and compositional (textual) meaning—can be identified in visual imagery. The term "compositional" has been used in place of Halliday's term "textual" because it more fully captures the sense of two modes within one page interacting with each other to provide a coherent intersemiotic message (Royce, 2013).

Systemic Functional Linguistics (SFL) and its metafunctions provide a comprehensive functional representation of language and visual design. As a result, they have become fundamental theories in the areas of linguistics, communication, education, and social sciences. However, they are rarely used as root theories in the information systems area. Although some literature has studied valuable feature capture referring to SFL theory, such as computer-mediated communication (Abbasi and Chen, 2008), social media (Dong et al., 2018), and tacit knowledge elicitation (Zappavigna and Patrick, 2010), prior studies extract text features only, overlooking their power in dealing with multimodal data. Our review of related studies indicates that no work has been done to analyze multimodality in information systems rooted in the metafunctions theory. Therefore, we aim to explore whether identifying and extracting metafunctions from multimodality can help to solve IS problems. Given that crowdfunding websites provide multiple semiotic modes, we analyze multimodal data of crowdfunding based on the metafunctions theory to predict fundraising success.

### *3.2.2 Multimodality in Crowdfunding*

A substantial body of research has demonstrated that various types of semiotic resources, such as texts, images, and videos, have an impact on fundraising outcomes. For instance, text length (Zhou et al., 2018), appearance and the number of images and videos (Bi et al., 2017; Hobbs et al., 2016; Liang et al., 2020; Mollick, 2014; Yang et al., 2020) positively impact the crowdfunding success. The sentiment (Wang et al., 2017) and topic (Yuan et al., 2016) of texts, color (Chan and Park, 2015), emotions (Hou et al., 2019), and human appearance of images (Anderson and Saxton, 2016), pitches, product complexity, owner appearance, and exemplification in the video (Dey et al., 2017; Kaminski et al., 2017; Korzynski et al., 2021) can help to predict the success of fundraising. However, only a few studies have analyzed the multimodality of crowdfunding projects. While one stream of literature has paid attention to explanatory study, it either only considers the statistics of texts, images, and videos or manually labels the visual contents on a small sample. For example, Carradini and Fleischmann (2022) investigated the effects of the numbers of images, links, gifs, videos, and galleries on crowdfunding success. They found that images, links, and video presences had a positive impact, while gifs and galleries did not. Grebelsky-Lichtman et al. (2018) analyzed verbal and nonverbal immediacy and nonimmediacy communication behaviors in crowdfunding videos. They found that verbal immediacy communication through videos has a positive effect on fundraising success, while verbal and nonverbal nonimmediacy has a negative effect. Yang el al. (2020) also conducted an explanatory study about the effects of the text length and image amount on pledges and backer amount. Even though Hou et al (2019) extracted emotional features from description texts, emotions, and objects from images, they only discussed the effects of title images on the crowdfunding platform.

Another stream of research focuses on predictive analysis, where researchers explore whether features from multimodality can improve predictive utility. Al-Qershi et al (2022) extracted linguistic features from texts, and content (objects and humans) and emotions from videos, discovering that humans contribute more than objects in predicting fundraising success, and words related to experience or perception are more important than cognitive words. Cheng et al (2019) applied Bag of Words (BoW) and word embedding (GloVe) to represent textual features and used a pre-trained VGG-16 model to extract visual features from crowdfunding images. Their extensive experimental results show that the image features could improve success prediction performance, particularly for project profiles with little text information. Kaminshi and Hopp (2019) built a predictive model using transformed texts, speech, and video contents, finding that the model with text, speech, and video features achieved the best performance in predicting campaign success or failure. Perez et al (2020) extracted sentiment, word importance, and named entities from description texts and proposed a pre-trained ResNet-152 model to extract emotion (eight types, e.g., sadness, fear, amusement), appearance (a color or an object) and semantic (the logit presence of predetermined objects in each image) features from images to predict fraud in crowdfunding projects. They found that textual data alone provides better performance in predicting fraudulent projects. Zhao et al (2022) examined how emotions are expressed in text and image modalities in a charity crowdfunding platform. Their experimental results show the superiority of the verbal when emotions are expressed via both text and image modalities in explaining crowdfunding campaign success, and the charity campaign appeals to persuasiveness if the same emotion appears in both text and image modalities.

In summary, the current literature on crowdfunding and its impact on fundraising outcomes has some limitations. Firstly, most studies have only analyzed the values of two data modalities in explaining or predicting fundraising outcomes. Secondly, the literature has only investigated a small set of features from multimodality, with most studies focusing on the emotional aspect of verbal and visual modalities, since there is no fundamental theory to support building a sound prediction framework that incorporates all relevant features. Thirdly, most studies focus on explanatory modeling, identifying salient influencing factors. However, in practice, a predictive model is more in need by stakeholders, providing insights into future trends and patterns, helping them make informed decisions in advance.

To the best of our knowledge, there is a gap in the research regarding a comprehensive investigation of multimodal data in crowdfunding. Prior studies have only considered one or two metafunctions and have neglected the interactions among the multiple modalities and among the three metafunctions (a comparison of studies is listed in Table 3.1). To address this gap, we propose a framework that explores features representing the ideational, interpersonal, and textual metafunctions of multimodal data in crowdfunding, respectively. We conduct several experiments to study the effectiveness of each metafunction, each modality, and their interactions in predicting fundraising outcomes.

**Table 3.1. Studies Related to Multimodal Data Analysis on Crowdfunding**

| Study | Target variable | Analysis type | Features | Modalities | Gap |
|---|---|---|---|---|---|
| Al-Qershi et al., 2022 | Success or not | Predictive | Linguistic features from texts, visual contents (number of humans, facial emotions) of videos | Texts and videos | Only investigated the objects and emotions from texts and videos, ignoring the experiential meaning, logical meaning, and compositional meaning of multimodality (ideational and textual metafunctions) |
| Carradini and Fleischmann, 2022 | Success or not | Explanatory | The numbers of images, links, gifs, videos, and galleries | Images and Videos | Only examined the numbers of images and videos, overlooked the textual and visual features of multimodality |
| Cheng et al, 2019 | Success or not | Predictive | Bag of words representation of texts; word embedding representation of texts; | Texts and Images | Overlooked the effects of social relations and interactions conveyed by texts and images (interpersonal metafunction) and the compositional features of texts and images (textual metafunction) |

| | | | | | |
|---|---|---|---|---|---|
| | | | CNN representation of images | | |
| Grebelsky-Lichtman et al., 2018 | Success or not | Explanatory | Verbal and nonverbal immediacy features | Videos （verbal and nonverbal） | Small sample size (120) with human coding; overlooked the experiential meaning, logical meaning, and compositional meaning of multimodality (ideational and textual metafunctions) |
| Hou et al., 2019 | Images' emotion; the number of backers; fundraising amount | Explanatory | Texts and title images' emotion; title images' composition; title images' color; body-ground relationship; the presence of text, human, and animal in the image | Texts and Images | Investigated the features of title images of the projects, overlooking the effects of other images; extracted only emotional features from texts |
| Kaminshi and Hopp, 2019 | Success or not | Predictive | Word vectors of texts and speech, objects from videos | Texts and videos | Overlooked the effects of social relations and emotions (interpersonal metafunction) and the compositional features of multimodality (textual metafunction) |
| Perez et al, 2020 | Fraud or not | Predictive | Emotions of texts; named entity and word importance of texts; color, texture, objects of images; face appearance of images | Texts and images | Overlooked the effects of experiential meaning and logical meaning of texts and images (ideational metafunction) |
| Yang et al., 2020 | Fundraising amount; the number of backers | Explanatory | Text length; number of images | Texts and images | Investigated the effects of text length, the number of images on fundraising amount, and the number of backers, ignoring the visual features of images |
| Zhao et al., 2022 | Success or not | Explanatory | Sentiments and emotions of texts and images | Texts and images | Only examined the emotions of image and text modalities, overlooking the experiential |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | meaning, logical meaning, and compositional meaning of texts and images (ideational and textual metafunctions) |
| Our research | Success or not | Predictive | Word embeddings; image embeddings; video embeddings; MFCC; social relation, polarity modality, and lexical density of texts; facial expression, objects, social distance, color, and composition of images and videos | Texts, images, and videos | Considers ideational, interpersonal, and textual metafunctions of each modality and multiple modalities |

## 3.3 Framework

Our framework builds on the theories of metafunctional framework on language proposed by Haliday (1985), functional visual design developed by Kress and Van Leeuwen (1996), and intersemiotic complementarity of language and visual image developed by Royce (1998). Halliday (1985) models the language as inter-related systems that are metafunctionally organized. Similarly, visual images also have metafunctions, including representing the experiential world (representational meaning), interacting with viewers (interactive meaning), and arranging the visual resources (compositional meaning) (Kress and Van Leeuwen,1996).

### *3.3.1 The Ideational Metafunction Representation*

The ideational metafunction of language demonstrates how we represent experience through the language (Halliday, 1985), including the experiential meaning and logical meaning between clauses. It encompasses the content and logico-semantic relations that reveal the organization of content (Djonov, 2005; Norris and Maier, 2014). This merafunction is concerned with analyzing the sequence of parts (i.e., words, word groups, clause, clause complexes, and paragraphs), which collectively develop the texts (O'Halloran, 2008a). Royce (1998) applied Halliday's ideational meaning in representing the visual structures. He identified the represented participants, which correlate to the ideational metafunction, as all the elements or entities that are present in the visual, whether animate or inanimate (Royce, 2013).

Drawing on their identifications of ideational metafunction for languages and visual images, we use deep representation features learned by neural networks to capture the ideational metafunction of modalities in crowdfunding.

### *3.3.2 The Interpersonal Metafunction Representation*

The interpersonal metafunction is realized by the clause as an exchange of information or an exchange of goods and services, and is basically concerned with enacting social relationships between the speaker or writer and the audience or viewer in a specific context of communication (Halliday, 2004). At the level of lexico-grammatical of texts, interpersonal meaning includes the forms of interaction and social interplay with others, polarity (positive and negative), and modality (degree of certainty and probability) (Halliday, 2004; Guijarro, 2010).

The interpersonal metafunction of the visual image involves features of contact, social distance, and modality between viewers and visual participants. Contact is constructed by any gaze or facial expression of the visual participants toward the viewers, and it conveys information in the form of a portrayal; social distance is determined by how close the visual participant appears to the viewer in an image, with close or long shots related to the degree of intimacy between visual participants and viewers (Kress and van Leeuwen, 2006). Modality is interpreted as the truth, credibility, and probability of what visual participants represent to viewers, and whether the information they offer is real or unreal (Royce, 2013). The pages that are relatively static, ordered, and less varied in color use present content as a higher modality and are more likely to convey factual information (Kress and Van Leeuwen, 2006).

Therefore, we extract the degrees of social interactions, emotions, and certainty of texts to represent text interpersonal metafunction. For interpersonal metafunction of images and videos, we extract the number of faces and facial expressions in the images/video frames to represent contact. We use the shot size of images/video frames (long, medium, or close) to represent social distance, and color and composition variation of images/video frames in each project to represent the modality. Moreover, we also include volume variation as a feature to represent the salience of audio.

### *3.3.3 The Textual Metafunction Representation*

The textual metafunction enables the function with ideational and interpersonal meanings. The textual component is focused on the analysis of lexical density and grammatical complexity. Halliday (1989) stated that written language becomes complex when it is lexically dense, as it involves packing a large number of lexical items into each clause, while spoken language is considered grammatically complex. Halliday (1996) further developed the measurement of lexical density as “the proportion of content (lexical) words - basically nouns, full verbs, adjectives, and adverbs derived from adjectives - over the total number of words in a text”. Thus, we use lexical density as a feature to represent textual metafunction of texts. We extract lexical words from the descriptions texts and calculate the proportion of lexical words in the texts to represent the textual meaning.

In visual design, achieving compositional cohesion requires attention to three factors: information value, framing, and salience of images (Kress and van Leeuwen, 2006). The information value refers to the placement of elements within the page, including their location in the center or margin, left or right. According to Kress and van Leeuwen (2006), elements on the left side of a visual image are considered as something already known, while the right side presents new information. The top and bottom of an image represent the realms of the ideal and the real, respectively. To capture this feature, we extract the location of the main body in each image.

Framing refers to the visual devices used to connect or separate the content on a page. Devices such as borders, spacing, and color can be used to frame content (Norris and Maier, 2014). Color is the key feature of image framing because it connects or separates important objects within simple pictures (Norris and Maier, 2014). Salience refers to the ability of a visual participant to capture the viewer's attention. Salience is related to size, movement, and contrast, and it is determined by a variety of features, such as the number and size of elements, tonal contrast, and color contrast (Norris and Maier, 2014). Therefore, we extract features of the image colors to represent framing, the number and size of main objects, and the color contrast to represent the salience of images. For salience, in addition to image-related features, we include audio variation to show the salience of audio in videos, as they are viewed as a sequence of images and audio. Table 3.2 summarizes the proposed metafunction representations of multimodality.

**Table 3.2. Metafunction Representations of Multimodality**

| Modality | Ideational Metafunction | Interpersonal Metafunction | Textual Metafunction |
|---|---|---|---|
| Text | Experiential meanings; logical meanings (textual elements) | Interaction and social relation; polarity; modality | Lexical density |

| | | | | |
|---|---|---|---|---|
| Image | | Visual elements | Contact;<br>social distance;<br>modality | Information value;<br>framing;<br>salience |
| Video | Image | Visual elements | Contact;<br>social distance;<br>modality | Information value;<br>framing;<br>salience (main object appearance time is included) |
| | Audio | MFCC representation | Modality (volume variation) | Salience (speech appearance time, music, speech, salience variation) |

The proposed framework is outlined in Figure 3.1. First, we collect multimodal data from the crowdfunding platform and then conduct a series of text preprocessing. Next, we extract the ideational, interpersonal, and textual representation features from multimodal data using deep learning models (BERT, CNN, 3D ResNets), Google Vision API, and Linguistic Inquiry and Word Count (LIWC). Finally, we develop machine learning models to predict fundraising success on the three metafunction representation feature sets of the multimodal data.

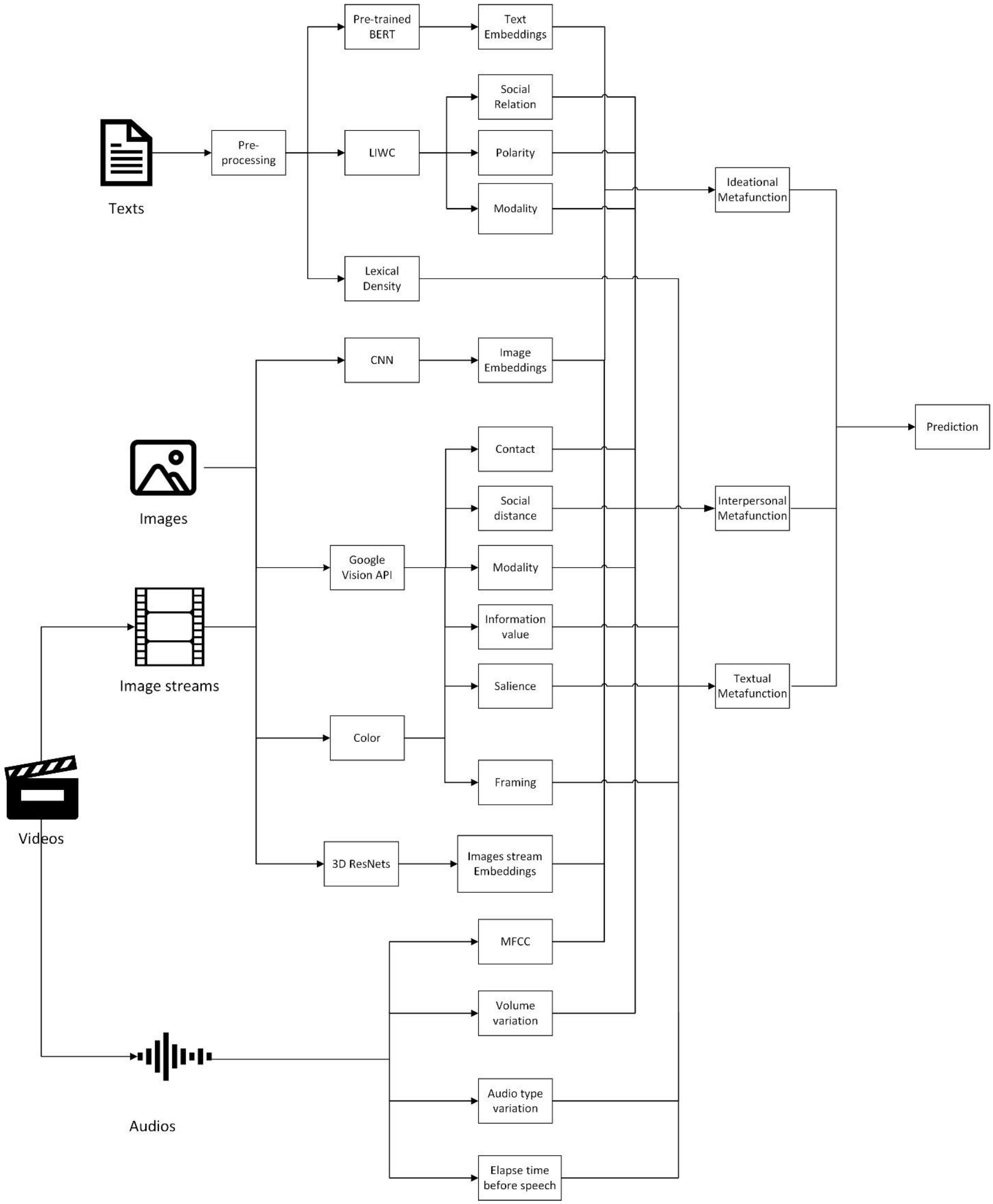



**Figure 3.1. Outline of the Proposed Framework**

## 3.4 Empirical Evaluation

### *3.4.1 Data*

We have evaluated our proposed framework on a dataset collected from Kickstarter.com, a well-known reward-based crowdfunding platform. We developed web crawlers to collect the projects that launched between 2014 and 2019, and then kept completed projects (successful or failed) for our evaluation and eliminated uncompleted (suspended or canceled) projects. As the object types of images can vary widely across different categories (e.g., there is more human images in the music category but less in the technology category), we select the music category, one of the top categories for launching projects, for empirical study. Since our research focuses on the multimodality of projects, we retained projects with multiple modalities for evaluation. In total, there were 8,559 completed projects, 6,164 of which were successful. Next, we extracted the texts, images, and videos from the description page of each project.

For text data, we did a series of text preprocessing. We removed stop words, numbers, punctuation, and HTML tags, converted characters to lowercase, and then lemmatized words to a base form. For videos, we selected the first video of each project for analysis, as it usually includes the main instruction of the project, which investors are most likely to watch. We then sliced each video to image sequences per second, and extracted the first 180 seconds (three minutes), which investors may pay the most attention to.

Besides the multimodality, we also extracted and standardized some meta features of each project, including fundraising goal, duration, description text length, number of images, and number of videos.

### *3.4.2 Feature Extraction*

#### *3.4.2.1 Ideational Metafunction*

Ideational metafunction reveals all the elements or entities that are present in the multimodal data including experiential and logical meanings. Deep neural networks are highly effective in extracting knowledge from a large and unstructured dataset. They have been used for implementing a universal learning approach across various application domains (speech, language, and vision understanding) (Alom et al., 2019).

To represent the text ideational metafunction, we applied transfer learning using a popular language representation model BERT (Bidirectional Encoder Representation from Transformers) (Devlin et al., 2018). A pre-trained BERT model is efficient to embed language structures, semantic patterns, and linguistic styles into deep representations to create state-of-the-art models for domain-specific text classification (Devlin et al. 2018).

Specifically, we processed texts into input embeddings (token embeddings, segment embeddings, and position embeddings) required by BERT and then initialized the BERT model with the pre-trained parameters[2]. We used the textual embedding outputs as input features of prediction models.

Krizhevsky et al. (2012) made a breakthrough in image classification using convolutional neural networks (CNNs), which significantly enhanced the description capability of image representation (Liu et al. 2018). Researchers have extracted different levels of image features using some CNN architectures, such as AlexNet (Krizhevsky et al., 2012), VGG-VD (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al. 2015), and CaffeNet (Jia et al. 2014). We applied

[2] https://github.com/google-research/bert/blob/master/multilingual.md

the VGG-16 architecture[3] (Simonyan and Zisserman, 2014), which has demonstrated superior recognition or classification accuracy in CNN (Alom et al., 2019), to map images into deep representations.

3D ResNets are popular 3D CNNs that are based on residual networks (ResNets) that aim to encode video data into a sequential embedding with spatio-temporal (3D) kernels (Hara et al., 2017. 3D ResNets have been designed to recognize actions in videos and have achieved better performance than relatively shallow 3D networks. Therefore, we adopted 3D ResNets[4] to extract sequential embeddings of image streams.

Regarding the audio modality, we used Mel Frequency Cepstral Coefficients (MFCC) to represent audio ideational features. MFCC was introduced by Davis and Mermelstein (1980) and has been widely used in automatic speech and speaker recognition.

Finally, we obtained text embeddings, image embeddings, and video embeddings from deep neural network models, as well as MFCC from audio to represent the ideational meanings of crowdfunding projects. As the number of embedding features is much larger than that of interpersonal or textual representative features, we conducted feature selection using long short-term memory (LSTM) attention. Some key parameter settings of the deep learning models are listed in Appendix A.

#### *3.4.2.2 Interpersonal Metafunction*

Interpersonal metafunction reveals interaction and social relations with viewers. For the text modality, we deployed the LIWC software (Pennebaker et al., 2015) to extract emotions, polarity, and modality from descriptions as interpersonal features. LIWC provides an effective and efficient method for studying the emotional, cognitive, social, and honest components present in verbal and

[3] https://github.com/minar09/VGG16-PyTorch
[4] https://github.com/kenshohara/3D-ResNets

written speech samples (Tausczik and Pennebaker, 2010). Its 80 language categories have been linked in hundreds of studies to interesting psychological processes. We obtained social relation (social, family, friends, females, males), polarity (positive and negative emotions), and modality (certainty, tentativeness, risk) of each project description text as interpersonal metafunction of the text.

To capture the interpersonal metafunctions revealed in images, we used Google Vision API to extract features of contact, social distance, and modality between viewers and visual participants. To represent contact, we extracted the face amounts and facial expressions (positive and negative) and calculated their statistics (i.e., maximum, minimum, mean, and standard deviation) among the images of each project. Since the size of the main human compared to the whole image reflects the distance between viewers and visual participants to some degree, we extracted the coordinates of the largest human in each image and then calculated his/her square ratio to the image size to present the long shot, medium shot, or close shot. We further calculate the relative statistics (i.e., maximum, minimum) of the ratios among the images of each project to represent social distance. Finally, since the standard deviation reflects the variation of a data series, we computed the standard deviation of color (i.e., warm/cool, saturation, brightness, and contrast) among the images of each project and the standard deviation of compositions: distances between main body mid-point and image horizontal/vertical central line in each project (visual balance), distances between main body mid-point and image equally-spaced horizontal/vertical lines in each project (rule of thirds) to represent the modality.

For videos, we sliced each video into image frames by each second. Similarly, we extracted representative features of contact, social distance, and modality from image frames. In addition, we also extracted some video-related features to represent interpersonal metafunction. Specifically, we calculated the ratio of human appearance duration to video duration, the maximum and average number of facial expressions (positive and negative), and audio volume variation to represent the contact.

### *3.4.2.3 Textual Metafunction*

Textual metafunction reveals the organization of the structure creating coherent texts or images. For the text modality, we extracted lexical words (i.e., noun, verb, adjective, adverbial, preposition, and conjunction) from the description texts and used the proportions of lexical words within the texts to represent the textual meaning.

To present information value of images, we extracted the location of the main body in each image and calculated the visual balance and rule of thirds of each image. We then used the means of these distances of each project to represent the information value.

For the framing of images, we calculated the average and maximum color ratio of each image to the whole project. Since salience shows the capability of a visual participant to attract the viewer's attention, we obtained the number of objects, the square ratio of the main body to each image in each project, and the color contrast of each project to represent salience.

For the video modality, we added the elapsed time before the human appearance and the elapsed time before the speech appearance to represent information value. We extracted the total/average appearance time and frequency ratios of music/speech/salience in each video to represent salience.

**Table 3.3. Descriptive Statistics of the Metafunction Features from Multimodal Data**

| Metafunction | Feature | Max | Min | Mean | Std |
|---|---|---|---|---|---|
| Meta | Goal ($) | 30,000,000 | 1 | 11，215.85 | 51，749.21 |
| | Duration (days) | 60 | 1 | 32.50 | 10.63 |

| | | | | | |
|---|---|---|---|---|---|
| | Description length (words) | 5518 | 1 | 466.34 | 430.63 |
| | Image count | 99 | 1 | 7.971756 | 8.08 |
| | Video count | 38 | 1 | 2.40 | 1.09 |
| Social relation (Text) | Female words within a text | 10.46 | 0 | 0.20 | 059 |
| | Male words within a text | 33.33 | 0 | 0.44 | 0.93 |
| | Friend words within a text | 16.67 | 0 | 0.29 | 0.44 |
| | Family words within a text | 20 | 0 | 0.21 | 0.47 |
| Polarity (Texts) | Positive emotion words within a text | 22.22 | 0 | 4.20 | 1.84 |
| | Negative emotion words within a text | 15.38 | 0 | 0.48 | 0.71 |
| Modality (Texts) | Certainty words within a text | 20 | 0 | 1.25 | 0.89 |
| | Tentative words within a text | 10 | 0 | 1.61 | 1.06 |
| | Risk words within a text | 14.29 | 0 | 0.16 | 0.33 |
| Lexical density (Texts) | Proportion of pronoun words within a text | 30.77 | 0 | 11.27 | 4.11 |
| | Proportion of prep words within a text | 33.33 | 0 | 13.17 | 2.61 |
| | Proportion of verb words within a text | 38.1 | 0 | 14.49 | 3.73 |
| | Proportion of adjective words within a text | 23.08 | 0 | 4.22 | 1.64 |
| | Proportion of conjunctive words within a text | 20 | 0 | 5.44 | 1.70 |
| | Proportion of adverbial words within a text | 20.00 | 0 | 2.91 | 1.45 |
| Contact (Images) | Maximum number of faces appearing in an image in each project | 46 | 0 | 2.60 | 3.08 |
| | Minimum number of faces appearing in an image in each project | 46 | 0 | 1.44 | 2.48 |
| | Average number of faces appearing in an image in each project | 46 | 0 | 1.86 | 1.50 |
| | Standard deviation of faces appearing in an image among each project | 5 | 0 | 0.44 | 0.89 |
| | Average probability of positive faces appearing in each project | 1 | 0 | 0.15 | 0.31 |
| | Maximum probability of positive faces appearing in each project | 1 | 0 | 0.28 | 0.45 |

| | | | | | |
|---|---|---|---|---|---|
| | Standard deviation of the probability of positive faces appearing in each project | 0.5 | 0 | 0.07 | 0.16 |
| | Average probability of negative faces appearing in each project | 1 | 0 | 0.27 | 0.39 |
| | Maximum probability of negative faces appearing in each project | 1 | 0 | 0.43 | 0.49 |
| | Standard deviation of the probability of negative faces appearing in each project | 0.5 | 0 | 0.10 | 0.19 |
| Social distance (Images) | Maximum ratio of square of main body to image square | 1 | 0 | 0.34 | 0.29 |
| | Minimum ratio of square of main body to image square | 1 | 0 | 0.19 | 0.25 |
| Modality (Images) | Standard deviation of distance between largest human body mid-point and image horizontal central line in each project | 0.21 | 0 | 0.03 | 0.04 |
| | Standard deviation of distance between largest human body mid-point and image vertical central line in each project | 0.22 | 0 | 0.02 | 0.04 |
| | Standard deviation of distance between largest human body mid-point and image equally spaced horizontal lines in each project | 0.40 | 0 | 0.04 | 0.08 |
| | Standard deviation of distance between largest human body mid-point and image equally spaced vertical lines in each project | 0.45 | 0 | 0.03 | 0.07 |
| | Standard deviation of warm color in each project (warm:1; cold:-1) | 1 | -1 | 0.07 | 0.87 |
| | Standard deviation of brightness in each project | 255 | 0 | 86.49 | 65.19 |
| | Standard deviation of saturation in each project | 252.94 | 0 | 65.58 | 55.13 |
| | Standard deviation of contrast in each project | 169.34 | 0 | 54.93 | 42.43 |
| Information value (Images) | Average distance between main body mid-point and image horizontal central line in each project | 0.46 | 2.76E-06 | 0.11 | 0.11 |
| | Average distance between main body mid-point and image vertical central line in each project | 0.47 | 1.34E-05 | 0.08 | 0.08 |

| | | | | | |
|---|---|---|---|---|---|
| | Average distance between main body mid-point and image equally spaced horizontal lines in each project | 0.93 | 0 | 0.30 | 0.19 |
| | Average distance between main body mid-point and image equally spaced vertical lines in each project | 0.94 | 0 | 0.26 | 0.16 |
| Framing | Average warm color among the images of each project | 1 | -1 | -0.26 | 0.86 |
| | Average brightness among the images of each project | 255 | 0 | 132.36 | 48.67 |
| | Average saturation among the images of each project | 252.94 | 0 | 90.04 | 51.18 |
| | Average contrast among the images of each project | 169.34 | 0 | 80.75 | 32.59 |
| | Maximum warm color among the images of each project | 1 | -1 | 0.03 | 0.99 |
| | Maximum brightness among the images of each project | 255 | 0 | 148.05 | 56.74 |
| | Maximum saturation among the images of each project | 255 | 0 | 108.04 | 60.58 |
| | Maximum contrast among the images of each project | 178.85 | 0 | 90.87 | 35.97 |
| Salience (Images) | Maximum ratio of largest human body square to image square in each project | 1 | 0 | 0.34 | 0.29 |
| | Average ratio of largest human body square to image square in each project | 1 | 0 | 0.25 | 0.24 |
| | Standard deviation of ratio of largest human body square to image square in each project | 0.50 | 0 | 0.05 | 0.10 |
| | Sum of contrast among the images of each project | 4726.61 | 0 | 217.26 | 311.42 |
| | Average contrast among the images of each project | 169.34 | 0 | 80.75 | 32.59 |
| | Maximum contrast among the images of each project | 178.85 | 0 | 90.87 | 35.97 |
| Contact (Video frame) | Total number of humans in the video | 43564 | 0 | 686.28 | 1939.23 |
| | Maximum number of faces appearing in a video frame in each video | 40 | 0 | 4.96 | 4.90 |
| | Minimum number of faces appearing in a video frame in each video | 15 | 0 | 1.07 | 0.64 |
| | Average number of faces appearing in an image in each project | 22.36 | 1 | 2.55 | 2.17 |

| | | | | | |
|---|---|---|---|---|---|
| | Standard deviation of faces appearing in an image among each project | 16.29 | 0 | 1.23 | 1.57 |
| | Total number of positive faces appearing in each video frame | 176 | 0 | 20.50 | 21.02 |
| | Average number of positive faces appearing in each video frame | 1 | 0 | 0.17 | 0.16 |
| | Standard deviation of positive faces appearing in each video frame | 0.5 | 0 | 0.31 | 0.14 |
| | Total number of negative faces appearing in each video frame | 176 | 0 | 20.50 | 21.02 |
| | Average number of negative faces appearing in each video frame | 1 | 0 | 0.36 | 0.20 |
| | Standard deviation of negative faces appearing in each video frame | 0.5 | 0 | 0.42 | 0.13 |
| Social distance (Video frame) | Maximum ratio of the square of main body to video frame square | 1.17 | 0 | 0.71 | 0.26 |
| | Minimum ratio of the square of main body to video frame square | 1.04 | 0 | 0.05 | 0.15 |
| Modality (video frame) | Standard deviation of distances between largest human body mid-point and video frame horizontal central line in each video | 124.33 | 0 | 43.95 | 24.51 |
| | Standard deviation of distances between largest human body mid-point and video frame vertical central line in each project | 184.29 | 0 | 16.78 | 11.73 |
| | Standard deviation of distances between largest human body mid-point and video frame equally spaced horizontal lines in each project | 520.06 | 0 | 239.37 | 59.70 |
| | Standard deviation of distances between largest human body mid-point and video frame equally spaced vertical lines in each project | 1068.11 | 0 | 135.26 | 55.79 |
| | Standard deviation of warm color in the video (warm:1; cold:-1) | 1 | 0 | 0.65 | 0.35 |
| | Standard deviation of brightness in the video | 103.89 | 0 | 30.33 | 16.73 |
| | Standard deviation of saturation in the video | 118.10 | 0 | 31.13 | 18.54 |
| | Standard deviation of contrast in the video | 72.31 | 0 | 21.27 | 12.96 |
| Modality (video audio) | Standard deviation of volume in the video | 51.94 | 0.17 | 41.24 | 14.00 |

| | | | | | |
|---|---|---|---|---|---|
| Information value (video frame) | The elapsed time before the first human appearance | 300 | 0 | 18.25 | 62.07 |
| | Average distance between main body mid-point and video frame horizontal central line | 260.03 | 0 | 75.12 | 44.22 |
| | Average distance between main body mid-point and video frame vertical central line | 534.05 | 0 | 40.90 | 25.75 |
| | Average distance between main body mid-point and video frame equally spaced horizontal lines in each project | 520.06 | 0 | 239.37 | 59.70 |
| | Average distance between main body mid-point and video frame equally spaced vertical lines in each project | 1068.11 | 0 | 135.26 | 55.79 |
| Information value (video audio) | The elapsed time before the speech appearance | 179.98 | 0.2 | 48.59 | 62.43 |
| Framing (video frame) | Average warm color among the video frames of each project | 1 | -1 | 0.03 | 0.68 |
| | Average brightness among the video frames of each project | 252.19 | 9.18 | 104.58 | 33.76 |
| | Average saturation among the video frames of each project | 253.27 | 0 | 97.42 | 36.95 |
| | Average contrast among the video frames of each project | 176.19 | 0 | 97.62 | 22.59 |
| | Maximum warm color among the video frames of each project | 1 | -1 | 0.86 | 0.50 |
| | Maximum brightness among the video frames of each project | 255 | 15.79 | 171.02 | 49.89 |
| | Maximum saturation among the video frames of each project | 255 | 0 | 175.29 | 55.28 |
| | Maximum contrast among the video frames of each project | 177.33 | 0 | 136.63 | 28.56 |
| Salience (video frame) | Maximum ratio of main body square to image square in the video | 1.18 | 0 | 0.71 | 0.26 |
| | Average ratio of main body square to image square in each project | 1.07 | 0 | 0.33 | 0.18 |
| | Standard deviation of ratio of main body square to image square in each project | 0.45 | 0 | 0.17 | 0.87 |
| | Average contrast among the images of each project | 176.19 | 0 | 97.62 | 22.59 |
| | Maximum contrast among the images of each project | 177.33 | 0 | 136.63 | 28.56 |

| | | | | | |
|---|---|---|---|---|---|
| | Standard deviation of contrast among the images of each project | | | | |
| Salience (video audio) | Ratio of music duration to video duration | 1.00 | 0 | 0.44 | 0.25 |
| | Ratio of speech duration to video duration | 1.00 | 0 | 0.42 | 0.28 |
| | Ratio of silence duration to video duration | 0.67 | 0 | 0.14 | 0.15 |
| | Average duration of music appearance in each video | 179.98 | 0 | 24.87 | 36.41 |
| | Average duration of speech appearance in each video | 179.98 | 0 | 15.59 | 24.93 |
| | Average duration of salience appearance in each video | 68.2 | 0 | 1.95 | 2.60 |
| | Maximum duration of music appearance in each video | 179.98 | 0 | 42.59 | 45.56 |
| | Maximum duration of speech appearance in each video | 179.98 | 0 | 15.59 | 24.93 |
| | Maximum duration of salience appearance in each video | 116.72 | 0 | 2.47 | 3.50 |
| | Minimum duration of music appearance in each video | 179.98 | 0 | 14.76 | 36.32 |
| | Minimum duration of speech appearance in each video | 179.98 | 0 | 7.81 | 23.75 |
| | Minimum duration of salience appearance in each video | 36.08 | 0 | 1.55 | 2.26 |
| | Standard deviation of duration of music appearance in each video | 89.19 | 0 | 11.22 | 16.11 |
| | Standard deviation of duration of speech appearance in each video | 86.83 | 0 | 8.07 | 10.55 |
| | Standard deviation of duration of salience appearance in each video | 48.52 | 0 | 0.41 | 1.19 |

### *3.4.3 Experiments*

To evaluate the predictive performance of the ideational, interpersonal, and textual metafunctions of multimodal data of crowdfunding, we conducted a series of experiments using five machine learning methods, i.e., logistic regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), random forest (RF), and XGBoost (XGB). In the first experiment, we included the meta features (summarized in Table 3.3) as the baseline model and then added our proposed ideational features to the meta feature set to form the ideational model. We used the ideational models to assess the predictive effects of ideational metafunction on each modality (text, image, and video) and multimodality. The first experiment involved 25 settings (5 combinations of modalities *5 classification methods).

Similarly, in the following two experiments, we examined the predictive utilities of interpersonal and textual representative features on each modality and multimodality. Both the second (interpersonal metafunction) and third experiments (textual metafunction) involved 20 settings (4 combinations of modalities * 5 classification methods).

To examine possible interactions among the metafunctions in influencing prediction performance, we compared the three metafunction feature sets, as well as the possible combinations of them, to examine possible interactions among the metafunctions in influencing the performance of prediction. This experiment involved 35 settings (7 combinations of metafunctions * 5 classification methods).

For every setting in the experiments, we estimated the prediction performance in terms of the area under the curve (AUC) and Kolmogorov–Smirnov (KS) statistic (Massey Jr, 1951), through ten independent rounds of ten-fold cross-validations, resulting in 100 estimates for each metric.

## 3.5 Results

### *3.5.1 Experiment 1: Ideational Metafunction of Multimodal Data*

To evaluate the effects of the proposed ideational features on single modality and multimodality, we have trained and tested five models using meta features and ideational representative features of four types of modalities, namely the "Baseline" model, "Text" model, "Image" model, "Video" model, and "Multimodal" model, respectively. Table 3.4 summarizes the predictive performance (AUC and KS) of the five classification methods of five models. Table 3.5 summarizes the results of the Friedman test and Dunn's pairwise post hoc test using the AUC metric (The test result using the KS metric is available in Appendix B).

Overall, the ideational metafunction of any modality and multimodality significantly performed better over the baseline model ($p$<.001), demonstrating the value of ideational metafunction in crowdfunding success prediction. Comparing the predictive effectiveness of different modalities, the metafunction of the multimodal data generated the best predictive performance ($p$<.001) over any single data modality. This implies that synthesizing data modalities helps improve the predictive utility. Furthermore, when comparing the predictive utility of a single modality, the results show that the ideational metafunction of text modality was significantly better than image and video modalities in predicting fundraising success. This suggests that text is an effective modality in conveying ideational metafunction to predict fundraising success.

**Table 3.4. Predictive Performance of Ideational Metafunction**

| Modalities | Metric | LR | LASSO | SVM | RF | XGB |
|---|---|---|---|---|---|---|
| Baseline | AUC | 0.690 (0.017) | 0.678 (0.018) | 0.669 (0.021) | 0.638 (0.016) | 0.710 (0.017) |
| | KS | 0.314 (0.030) | 0.303 (0.030) | 0.307 (0.035) | 0.235 (0.027) | 0.329 (0.029) |
| Text | AUC | 0.852 (0.015) | 0.834(0.016) | 0.837 (0.015) | 0.821 (0.016) | 0.838 (0.015) |
| | KS | 0.561 (0.032) | 0.528 (0.034) | 0.534 (0.031) | 0.500 (0.031) | 0.530 (0.030) |
| Image | AUC | 0.761 (0.020) | 0.734 (0.019) | 0.771 (0.021) | 0.761 (0.019) | 0.798 (0.018) |
| | KS | 0.411 (0.037) | 0.370 (0.036) | 0.433 (0.036) | 0.405 (0.034) | 0.461 (0.035) |
| Video | AUC | 0.816 (0.016) | 0.794 (0.017) | 0.795 (0.018) | 0.813 (0.016) | 0.828 (0.016) |
| | KS | 0.503 (0.032) | 0.472 (0.031) | 0.494 (0.031) | 0.495 (0.031) | 0.517 (0.031) |
| Multimodal | AUC | **0.882 (0.014)** | **0.862 (0.015)** | **0.869 (0.014)** | **0.852 (0.016)** | **0.878 (0.013)** |
| | KS | **0.618 (0.030)** | **0.572 (0.029)** | **0.608 (0.030)** | **0.559 (0.031)** | **0.604 (0.029)** |

Standard deviations are enclosed in parentheses.

**Table 3.5. Friedman Test and Post Hoc Dunn Test on Ideational Metafunction**

| Model | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | |
|---|---|---|---|---|---|
| | | Baseline | Texts | Images | Videos |
| Baseline | 4.99 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Texts | 2.11 | <.001 | | | |
| Images | 3.98 | <.001 | <.001 | | |
| Videos | 2.91 | <.001 | <.001 | <.001 | |
| Multimodal | 1.00 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1941.714 ($p$ < .001) | | | | | |

### *3.5.2 Experiment 2: Interpersonal Metafunction of Multimodal Data*

In the second experiment, we compared the predictive performance (AUC and KS) of interpersonal features of each modality and multimodality. As shown in Table 3.6 (results for the baseline model are the same as that in Experiment 1 and hence omitted), the interpersonal metafunction of any data modality showed better predictive capabilities over the baseline model ($p$ < .001). This highlights the value of social relationships in fundraising prediction. Comparing interpersonal metafunction of different data modalities, we found out that the multimodality resulted in the best performance over single modalities ($p$ < .001), indicating the superior predictive effectiveness of multimodality on interpersonal metafunction.

Interestingly, we observed that there is no statistically significant difference between the image model and video model, while either image or video model outperformed the text model ($p$ < .001). The results revealed that interpersonal metafunction conveyed by the video modality contributed almost equally as the image modality to the predictive effectiveness, and was more valuable than the text modality.

**Table 3.6. Predictive Performance of Interpersonal Metafunction**

| Modalities | Metric | LR | LASSO | SVM | RF | XGB |
|---|---|---|---|---|---|---|
| Text | AUC | 0.744 (0.020) | 0.723 (0.020) | 0.755 (0.020) | 0.762 (0.020) | 0.773 (0.018) |
| | KS | 0.397 (0.033) | 0.363 (0.033) | 0.404 (0.034) | 0.403 (0.034) | 0.427 (0.033) |
| Image | AUC | 0.778 (0.019) | 0.752 (0.018) | 0.773 (0.018) | 0.782 (0.017) | 0.795 (0.019) |
| | KS | 0.442 (0.034) | 0.400 (0.031) | 0.436 (0.031) | 0.442 (0.028) | 0.466 (0.034) |
| Video | AUC | 0.777 (0.019) | 0.752 (0.018) | 0.772 (0.018) | 0.783 (0.018) | 0.795 (0.019) |
| | KS | 0.442 (0.034) | 0.400 (0.031) | 0.435 (0.031) | 0.444 (0.031) | 0.465 (0.034) |
| Multimodal | AUC | **0.792 (0.018)** | **0.772 (0.017)** | **0.796 (0.018)** | **0.797 (0.018)** | **0.811 (0.017)** |
| | KS | **0.463 (0.033)** | **0.431 (0.031)** | **0.472 (0.033)** | **0.466 (0.035)** | **0.488 (0.032)** |

Standard deviations are enclosed in parentheses.

**Table 3.7. Friedman Test and Post Hoc Dunn Test on Interpersonal Metafunction**

| Model | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | |
|---|---|---|---|---|---|
| | | Baseline | Texts | Images | Videos |
| Baseline | 4.98 | | | | |
| Texts | 3.86 | <.001 | | | |
| Images | 2.54 | <.001 | 0.019 | | |
| Videos | 2.55 | <.001 | <.001 | 1.000 | |
| Multimodal | 1.07 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1781.042 ($p$ < .001) | | | | | |

### *3.5.3 Experiment 3: Textual Metafunction of Multimodal Data*

The predictive performance (Table 3.8) and Friedman test (Table 3.9) showed the predictive effectiveness of textual metafunctions in fundraising success prediction. The multimodality significantly outperformed any single modality, showing that increasing the data modality contributes to better predictive utility. Comparing the effectiveness of single modalities, the textual metafunction of videos overperformed that of images, which outperformed that of texts. This implies that ideational metafunctions conveyed by videos are more valuable in predicting fundraising success than those conveyed by text or image modality.

**Table 3.8. Predictive Performance of Textual Metafunction**

| Model | Metric | LR | LASSO | SVM | RF | XGB |
|---|---|---|---|---|---|---|
| Texts | AUC | 0.697 (0.020) | 0.684 (0.020) | 0.688 (0.021) | 0.713 (0.018) | 0.727 (0.016) |
| | KS | 0.313 (0.033) | 0.298 (0.036) | 0.315 (0.034) | 0.336 (0.029) | 0.360 (0.030) |
| Image | AUC | 0.736 (0.020) | 0.708 (0.019) | 0.729 (0.020) | 0.742 (0.021) | 0.762 (0.019) |
| | KS | 0.370 (0.035) | 0.338 (0.034) | 0.365 (0.034) | 0.370 (0.035) | 0.410 (0.034) |
| Video | AUC | 0.776 (0.019) | 0.750 (0.018) | 0.774 (0.018) | 0.788 (0.019) | 0.798 (0.018) |
| | KS | 0.439 (0.033) | 0.404 (0.031) | 0.434 (0.031) | 0.455 (0.031) | 0.473 (0.035) |
| Multimodal | AUC | **0.789 (0.018)** | **0.761 (0.018)** | **0.783 (0.019)** | **0.797 (0.017)** | **0.811 (0.017)** |
| | KS | **0.453 (0.033)** | **0.413 (0.033)** | **0.454 (0.035)** | **0.468 (0.033)** | **0.494 (0.032)** |

Standard deviations are enclosed in parentheses.

**Table 3.9. Friedman Test and Post Hoc Dunn Test on Textual Metafunction**

| Model | Average Rank | Adjusted $p$-value of Pairwise Comparison: Baseline | Texts | Images | Videos |
|---|---|---|---|---|---|
| Baseline | 4.69 | | | | |
| Texts | 4.22 | <.001 | | | |
| Images | 3.07 | <.001 | <.001 | | |
| Videos | 1.91 | <.001 | <.001 | <.001 | |
| Multimodal | 1.10 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1830.794 ($p$ < .001) | | | | | |

### *3.5.4 Experiment 4: Combination of Metafunctions of Multimodal Data*

Table 3.10 summarizes the predictive performance (AUC and KS) of combinations of metafunctions on multimodal data (results for the three metafunctions individually are the same as those in Experiments 1 to 3 and hence omitted). The results showed that combining any metafunction outperformed a single metafunction (i.e., ideational, interpersonal, or textual), highlighting the importance of considering multiple metafunctions in predicting fundraising success. However, the combination of all three metafunctions did not significantly outperform some combinations, namely ideational + interpersonal and ideational + textual. This implies potential conflicts or overfitting among metafunctions when combined together.

**Table 3.10. Predictive Performance of Metafunction Combinations**

| Model | Metric | LR | LASSO | SVM | RF | XGB |
|---|---|---|---|---|---|---|
| Idt+Ips | AUC | 0.887 (0.013) | 0.870 (0.014) | 0.877 (0.013) | 0.854 (0.015) | 0.884 (0.013) |
| | KS | 0.633 (0.030) | 0.586 (0.029) | 0.614 (0.029) | 0.560 (0.028) | 0.615 (0.027) |
| Idt+Txt | AUC | 0.886 (0.013) | 0.867 (0.014) | 0.875 (0.014) | 0.854 (0.015) | 0.883 (0.013) |
| | KS | 0.626 (0.029) | 0.583 (0.028) | 0.610 (0.029) | 0.560 (0.029) | 0.613 (0.028) |
| Ips+Txt | AUC | 0.803 (0.018) | 0.780 (0.016) | 0.802 (0.017) | 0.803 (0.017) | 0.822 (0.017) |
| | KS | 0.480 (0.032) | 0.441 (0.031) | 0.488 (0.033) | 0.477 (0.030) | 0.510 (0.032) |
| All | AUC | **0.888 (0.013)** | **0.871 (0.014)** | **0.877 (0.013)** | **0.853 (0.014)** | **0.884 (0.013)** |
| | KS | **0.633 (0.031)** | **0.490 (0.028)** | **0.609 (0.029)** | **0.558 (0.029)** | **0.616 (0.030)** |

**Table 3.11. Friedman Test and Post Hoc Dunn Test on Metafunction Combinations**

| | Average | Adjusted $p$-value of Pairwise Comparison |
|---|---|---|

| Model | Rank | Ideational (Idt) | Interpersonal (Ips) | Textual (Txt) | Idt+Ips | Idt+Txt | Ips+Txt |
|---|---|---|---|---|---|---|---|
| Ideational | 3.58 | | | | | | |
| Interpersonal | 6.20 | <.001 | | | | | |
| Textual | 6.61 | <.001 | 0.062 | | | | |
| Idt+Ips | 2.05 | <.001 | <.001 | <.001 | | | |
| Idt+Txt | 2.42 | <.001 | <.001 | <.001 | 0.136 | | |
| Ips+Txt | 5.18 | <.001 | <.001 | <.001 | <.001 | <.001 | |
| All | 1.95 | <.001 | <.001 | <.001 | 1.000 | 0.012 | <.001 |
| Friedman $\chi^2$: 2546.355 ($p$ < .001) | | | | | | | |

### 3.6 Discussion

Our experiments yield some intriguing findings. First and foremost, identifying ideational, interpersonal, and textual metafunctions of multimodal data has improved the predictive performance of fundraising success. This finding reveals that experiential and logical meaning, social interaction, and composition of the project description page have predictive values. Second, a comparison of the predictive utilities of the three metafunctions revealed that the ideational metafunction outperformed the interpersonal metafunction and the textual metafunction. This highlights the important role of the experiential and logical meaning of the project in fundraising success prediction. In addition, the metafunction combinations that include the ideational metafunction outperformed other combinations. This finding implies that effectively diffusing experiential and logical ideas to investors is a priority for predicting fundraising success.

In addition, our results show the superior predictive utility of multimodal data over any single modality on any metafunction, implying that increasing the diversity of data modalities is more valuable for predicting fundraising success. With regard to the predictive utilities of modalities, the interpersonal and textual metafunctions conveyed by videos and images were more valuable than those conveyed by texts, in predicting fundraising success. This implies that social connections or interactions and compositions are more effective to deliver more motive through engaged modalities, such as videos or images. Visual elements are more emotionally communicative than textual elements. In contrast, text modality demonstrated superior predictive value in conveying ideational metafunction. This reflects that texts are better at delivering logical and experiential meanings than image or video modalities when predicting crowdfunding success. Interestingly, we found that the image modality generated almost the same predictive utility as the video modality in conveying interpersonal metafunction. This implies that image modality has good enough prediction power in expressing social interactions and relations, and the additional movements in videos does not improve the predictive performance.

## 3.7 Contributions and Implications

Our study has implications for both research and practice. For research, our study is the first to adopt the metafunctions framework of languages (Halliday,1985), functional visual design (Kress and Van Leeuwen, 1996), and intersemiotic complementarity of languages and visual images (Royce, 1998) to discover metafunctions of multimodal data of crowdfunding projects. Our empirical study evaluated the effectiveness of ideational, interpersonal, and textual metafunctions in predicting fundraising success and demonstrated the predictive utility of any metafunctions and metafunction combinations. Moreover, our study revealed that adding modalities of data can significantly improve prediction performance. We discovered interesting patterns among metafunctions conveyed by different data modalities in crowdfunding success prediction. Specifically, our empirical results reveal a strong effect of interpersonal and textual metafunctions on image and video modalities, but a weak effect on text modality. In contrast, we discovered a superior effect of ideational metafunction conveyed by test modality than that conveyed by video or image modality.

For practice, our findings suggest that entrepreneurs should consider enriching their data with multiple modalities. Moreover, since different modalities are more effective at conveying different metafunctions, entrepreneurs should carefully select the appropriate modality to enhance predictive utilities of specific metafunctions. For example, if entrepreneurs want to deliver logical and experiential ideas, the text modality may provide more valuable information to predict success. On the other hand, if a project is immersive or perceptual in nature, a video or image modality may be more effective in improving success predictive utility.

Our proposed framework is based on the description pages of projects. It aims to help stakeholders to effectively predict campaigns' success once they are launched. Specifically, for entrepreneurs, predictive outcomes can assist in understanding the prospect of their campaigns, and in adjusting and optimizing their campaign strategy during the fundraising period. For investors, our proposed framework provides them with more information about the potential success of a campaign and helps them reduce the risk of investing in campaigns that are unlikely to meet the fundraising goals. Effective predictions help investors make informed investment decisions in advance, reducing their opportunity costs and maintaining their funding returns. For crowdfunding platforms, effective predictions can keep users engaged with a platform since the platform could provide them with valuable advice about campaign potentials, helping them make informed decisions, hence attracting more users and increasing the overall participation in the platform. Additionally, platforms can provide a more user-friendly website experience by recommending promising campaigns, which can help platforms develop a strong reputation for intelligence and reliability.

## 3.8 Conclusion

With the increasing popularity of multimodal data on crowdfunding platforms, entrepreneurs are increasingly using a variety of modalities to sell their ideas and make a good impression. However, there is a lack of systematic research exploring the effects of multimodal data functions on predicting fundraising success. Our study identified the metafunctions of multimodality in crowdfunding and demonstrated their predictive value. Our empirical evaluation also showed how different metafunctions affect data modalities.

Despite the valuable insights provided by our study, there are some limitations that may be addressed in future research. Firstly, while we have examined the predictive value of metafunctions in a human-engaged project category, further research is needed to assess their utilities in other categories, such as the technology category. Secondly, since we only examined the utilities of metafunctions on the Kickstarter platform, the generalizability of our findings needs to be validated on other reward-based crowdfunding platforms.

# REFERENCES


Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, *24*(6), 975–990.

Abbasi, A., & Chen, H. (2008). CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication. MIS Quarterly, 32(4), 811–837.

Agrawal, A., Catalini, C., & Goldfarb, A. (2014). Some simple economics of crowdfunding. *Innovation Policy and the Economy*, *14*(1), 63-97.

Agterberg, F. P. (1992). Combining indicator patterns in weights of evidence modeling for resource evaluation. *Nonrenewable Resources*, *1*(1), 39-50.

Ahlers, G. K. C., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice*, *39*(4), 955–980.

Ahmad, F. S., Tyagi, D., & Kaur, S. (2017). Predicting crowdfunding success with optimally weighted random forests. *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)*, 770–775.

Ahsan, M., Cornelis, E. F. I., & Baker, A. (2018). Understanding backers' interactions with crowdfunding campaigns: Co-innovators or consumers? *Journal of Research in Marketing and Entrepreneurship*, *20*(2), 252–272.

Al-Qershi, O. M., Kwon, J., Zhao, S., and Li, Z.（2022). Predicting Crowdfunding Success with Visuals and Speech in Video Ads and Text Ads. *European Journal of Marketing* (56:6), Emerald Publishing Limited, pp. 1610–1649.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., et al. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. Electronics, 8(3), 292.

Anderson, K. B., & Saxton, G. D. (2016). Smiles, Babies, and Status Symbols: The Persuasive Effects of Image Choices in Small-Entrepreneur Crowdfunding Requests. International Journal of Communication, 10(0), 22.

Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics, 84*(3), 488–500.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics, 107*(3), 797-817.

Bateman, J. (2008). Multimodality and genre: A foundation for the systematic analysis of multimodal documents. Springer.

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.

Burtch, G., Ghose, A., & Wattal, S. (2015). The hidden cost of accommodating crowdfunder privacy preferences: A randomized field experiment. *Management Science*, *61*(5), 949–962.

Carradini, S., and Fleischmann, C. (2023). The Effects of Multimodal Elements on Success in Kickstarter Crowdfunding Campaigns. *Journal of Business and Technical Communication* (37:1), SAGE Publications Inc, pp. 1–27.

Chan, C. R., & Park, H. D. (2015). How images and color in business plans influence venture investment screening decisions. Journal of Business Venturing, 30(5), 732–748.

Chen, L.-S., & Shen, E.-L. (2019). Finding the keywords affecting the success of crowdfunding projects. *2019 IEEE 6th International Conference on Industrial Engineering and*

*Applications (ICIEA)*, 567–571.

Chen, S.-Y., Chen, C.-N., Chen, Y.-R., Yang, C.-W., Lin, W.-C., & Wei, C.-P. (2015). Will your project get the green light? Predicting the success of crowdfunding campaigns. *PACIS 2015 Proceedings*, 79.

Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, *48*(2), 238–254.

Cheng, C., Tan, F., Hou, X., & Wei, Z. (2019). Success Prediction on Crowdfunding with Multimodal Deep Learning. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2158–2164.

Clauss, T., Breitenecker, R. J., Kraus, S., Brem, A., & Richter, C. (2018). Directing the wisdom of the crowd: The importance of social interaction among founders and the crowd during crowdfunding campaigns. *Economics of Innovation and New Technology*, *27*(8), 709–729.

Courtney, C., Dutta, S., & Li, Y. (2017). Resolving information asymmetry: Signaling, endorsement, and crowdfunding success. *Entrepreneurship Theory and Practice*, *41*(2), 265-290.

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. MIS Quarterly, 33(3), 539–566.

Croson, R., & Shang, J. Y. (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, *11*(3), 221–233.

Cumming, D., & Zhang, Y. (2016). Alternative investments in emerging markets: A review and new trends. *Emerging Markets Review*, *29*, 1–23.

Davis, B. C., Hmieleski, K. M., Webb, J. W., & Coombs, J. E. (2017). Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: The influence of perceived product creativity and entrepreneurial passion. *Journal of Business Venturing*, *32*(1), 90–106.

Davis, S. V., and MERMELSTEIN, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing* (28), pp. 57–366.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

Dey, S., Duff, B., Karahalios, K., and Fu, W.-T.（2017). The Art and Science of Persuasion: Not All Crowdfunding Campaign Videos Are The Same. *in Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland Oregon USA: ACM, February 25, pp. 755–769. (https://doi.org/10.1145/2998181.2998229).

Djonov, E. (2005). Analysing the organisation of information in websites: From hypermedia design to systemic functional hypermedia discourse analysis. PhD Thesis: University of New South Wales.

Dong, W, Liao, S & Zhang, Z. (2018). Leveraging Financial Social Media Data for Corporate Fraud Detection. Journal of Management Information Systems. 35. 461-487.

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch hard or go home!: Predicting the success of Kickstarter campaigns. *Proceedings of the 1st ACM conference on Online Social Networks*, 177–182.

Fiore, A. M., Jin, H.-J., & Kim, J. (2005). For fun and profit: Hedonic value from image interactivity and responses toward an online store. Psychology & Marketing, 22(8), 669–694.

Garrett, J. J. (2002). The elements of user experience: User-centered design for the Web / Jesse James Garrett. (1st ed.). New Riders.

Giudici, G., Guerini, M., & Rossi Lamastra, C. (2013). Why crowdfunding projects can succeed: The role of proponents' individual and territorial social capital. *Available at SSRN 2255944*.

Grebelsky-Lichtman, T., and Avnimelech, G. (2018). Immediacy Communication and Success in Crowdfunding Campaigns: A Multimodal Communication Approach. *International Journal of Communication* (12:0), p. 27.

Guijarro, A. (2010). A Multimodal Analysis of The Tale of Peter Rabbit within the Interpersonal Metafunction. Atlantis, 32, 123–140.

Halliday, M. A. K. (1985). An introduction to functional grammar. E. Arnold.

Halliday, M. A. K. H., & Halliday, M. A. K. (1989). Spoken and written language. Oxford University Press, USA.

Halliday, M. A. K. (1994). An introduction to functional grammar. London: Edward Arnold.

Halliday, M.A.K. (2004). An Introduction to Functional Grammar. Third Edition. Revised by Christian M.I.M. Matthiessen. London: Edward Arnold.

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123.

Hara, K., Kataoka, H., and Satoh, Y. (2017). Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition. in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice: IEEE, October, pp. 3154–3160.

Hobbs, J., Grigore, G., & Molesworth, M. (2016). Success in the management of crowdfunding projects in the creative industries. Internet Research.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing*, *25*(1), 27–36.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.

Hou, J.-R., Zhang, J., & Zhang, K. (2019). Can Title Images Predict the Emotions and the Performance of Crowdfunding Projects? Proceedings of the 52nd Hawaii International Conference on System Sciences.

Huang, L., & Knight, A. P. (2017). Resources and relationships in entrepreneurship: An exchange theory of the development and effects of the entrepreneur-investor relationship. Academy of Management Review, 42(1), 80–102.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. Proceedings of the ACM International Conference on Multimedia - MM '14, 675–678.

Kaminski, J., Jiang, Y., Piller, F., and Hopp, C. (2017). Do User Entrepreneurs Speak Different? Applying Natural Language Processing to Crowdfunding Videos. *in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems,* CHI EA '17, New York, NY, USA: Association for Computing Machinery, May 6, pp. 2683–2689.

Kaminski, J. C., & Hopp, C. (2019). Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals. *Small Business Economics*, 1-23.

Korzynski, P., Haenlein, M., and Rautiainen, M. (2021). Impression Management Techniques in Crowdfunding: An Analysis of Kickstarter Videos Using Artificial Intelligence. *European Management Journal* (39:5), pp. 675–684.

Kang, L., Jiang, Q., & Tan, C.-H. (2017). Remarkable advocates: An investigation of geographic distance and social capital for crowdfunding. *Information & Management*, *54*(3), 336–348.

Kaur, H., & Gera, J. (2017). Effect of social media connectivity on success of crowdfunding campaigns. *Procedia Computer Science*, *122*, 767–774.

Kim, K., & Viswanathan, S. (2019). The experts in the crowd: The role of experienced investors in a crowdfunding market. MIS Quarterly, 43(2), 347-172.

Knox, J. (2007). Visual-verbal communication on online newspaper home pages. Visual Communication, 6(1), 19–53.

Kress, G., & van Leeuwen, T. (1996). Reading Images: The Grammar of Visual Design.London: Routledge.

Kress, G., & van Leeuwen, T. (2006). Reading Images: A Grammar of Visual Design(2nd ed.). London: Routledge.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, 1097–1105.

Kromidha, E., & Robson, P. (2016). Social identity and signalling success factors in online crowdfunding. *Entrepreneurship & Regional Development*, *28*(9–10), 605–629.

Kunz, M. M., Bretschneider, U., Erler, M., & Leimeister, J. M. (2017). An empirical investigation of signaling in reward-based crowdfunding. *Electronic Commerce Research*, *17*(3), 425–461.

Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. *The Economics of Crowdfunding*, 151–182.

Lagazio, C., & Querci, F. (2018). Exploring the multi-sided nature of crowdfunding campaign success. *Journal of Business Research*, *90*, 318–324.

Lai, C.-Y., Lo, P.-C., & Hwang, S.-Y. (2017). Incorporating comment text into success prediction of crowdfunding campaigns. *PACIS 2017 Proceedings*, 156.

Lee, C. H., Bian, Y., Karaouzene, R., & Ahmed, N. (2019). Roles of narrative in civic crowdfunded project: Linguistic style, message substance, and crowdfunding success. *Industrial Management & Data Systems*, (January 10).

Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications*, *11*(5), 495–503.

Li, H., Chen, X., Zhang, Y., & Hai, M. (2018). Prediction of financing goal of crowdfunding projects. *Procedia Computer Science*, *139*, 108–113.

Li, Y., Rakesh, V., & Reddy, C. K. (2016). Project success prediction in crowdfunding environments. *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, 247–256.

Liang, X., Hu, X., & Jiang, J. (2020). Research on the effects of information description on crowdfunding success within a sustainable economy—The perspective of information communication. *Sustainability*, *12*(2), 650.

Liu, D., Brass, D., Lu, Y., & Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. *MIS Quarterly*, *39*(3), 729-742.

Majumdar, A., & Bose, I. (2018). My words for your pizza: An analysis of persuasive narratives in online crowdfunding. *Information & Management*, *55*(6), 781–794.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16.

Norris, S., & Maier, C. D. (Eds.). (2014). Interactions, images and texts: A reader in multimodality (Vol. 11). Walter de Gruyter GmbH & Co KG.

O'Halloran, K. (2008a). Mathematical Discourse: Language, Symbolism and Visual Images. A&C Black.

O'Halloran, K. L. (2008b). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. Visual Communication, 7(4), 443–475.

Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, *32*(2), 215–236.

Perez, B., Machado, S. R., Andrews, J. T. A., & Kourtellis, N. (2020). I call BS: Fraud Detection in Crowdfunding Campaigns.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.

Petitjean, M. (2018). What explains the success of reward-based crowdfunding campaigns as they unfold? Evidence from the French crowdfunding platform KissKissBankBank. *Finance Research Letters*, *26*, 9-14.

Riegelsberger, J. (2002a). The effect of facial cues on trust in e-commerce systems. In Proceedings of HCI (Vol. 2).

Riegelsberger, J., Sasse, A., & McCarthy, J. (2002b). Eye-Catcher or Blind Spot? The Effect of Photographs of Faces on E-Commerce Sites, 383–398.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2003). Shiny happy people building trust? Photos on e-commerce websites and consumer trust. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 121–128.

Royce, T. D. (1998). Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. JASFL Occasional Papers, 1(1), 25-49.

Royce, T. D., & Bowcher, W. (Eds.). (2013). Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. In New Directions in the Analysis of Multimodal Discourse , 71–117.

Ryoba, M. J., Qu, S., & Zhou, Y. (2020). Feature subset selection for predicting the success of crowdfunding project campaigns. *Electronic Markets*, 1-14.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310.

Simonsohn, U., & Ariely, D. (2008). When rational sellers face nonrational buyers: Evidence from herding on eBay. *Management Science*, *54*(9), 1624–1637.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

Siering, M., Koch, J.-A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems, 33*(2), 421–455.

Song, Y., Wu, H., Ma, J., & Lu, N. (2019). Exploring the dynamic influences and interaction effects of signals on backers' investment in the crowdfunding market. *Information Technology & People*, *33*(2), 792–812.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going Deeper With Convolutions. 1–9.

Tausczik, Y., and Pennebaker, J. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* (29), pp. 24–54.

Van Leeuwen, T. (1999) Speech,Music, Sound. London:Macmillan.

Wang, N., Li, Q., Liang, H., Ye, T., & Ge, S. (2018). Understanding the importance of interaction between creators and backers in crowdfunding success. *Electronic Commerce Research and Applications*, *27*, 106-117.

Wang, W., Zhu, K., Wang, H., & Wu, Y.-C. J. (2017). The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. IET Software, 11(5), 229–238.

Wang, W., Zheng, H., & Wu, Y. J. (2020). Prediction of fundraising outcomes for crowdfunding projects based on deep learning: A multimodel comparative study. *Soft Computing*, *24*(11), 8323–8341.

Wang, W., Zhu, K., Wang, H., & Wu, Y.-C. J. (2017). The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. *IET Software*, *11*(5), 229–238.

Yu, P.-F., Huang, F.-M., Yang, C., Liu, Y.-H., Li, Z.-Y., & Tsai, C.-H. (2018). Prediction of crowdfunding project success with deep learning. *2018 IEEE 15th International Conference on E-Business Engineering (ICEBE)*, 1–8.

Yang, Jialiang, Li, Y., Calic, G., & Shevchenko, A. (2020). How multimedia shape crowdfunding outcomes: The overshadowing effect of images and videos on text in campaign information. Journal of Business Research, 117, 6–18.

Yuan, H., Lau, R. Y. K., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. Decision Support Systems, 91, 67–76.

Zappavigna, M., & Patrick, J. (2010). Eliciting tacit knowledge about requirement analysis with a Grammar-targeted Interview Method (GIM). European Journal of Information Systems, 19(1), 49–59.

Zhang, J., & Liu, P. (2012). Rational herding in microloan markets. *Management Science*, *58*(5), 892–912.

Zhao, K., Zhou, L., and Zhao, X. (2022). Multi-Modal Emotion Expression and Online Charity Crowdfunding Success, *Decision Support Systems* (163), p. 113842.

Zhao, H., Zhang, H., Ge, Y., Liu, Q., Chen, E., Li, H., & Wu, L. (2017). Tracking the dynamics in crowdfunding. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 625–634.

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.

Zhou, M., Lu, B., Fan, W., & Wang, G. A. (2018). Project description and crowdfunding success: An exploratory study. *Information Systems Frontiers*, 20(2), 259–274.

# APPENDICES

## Appendix A Parameter Settings

**Table A1. Parameter Setting of Pre-trained BERT**

| Parameter | Setting | Description |
|---|---|---|
| Textual embedding size | 768 | The output dimensionality of the CLS hidden cell of the pre-trained BERT |
| Token size | 256 | The padded dimensionality of tokens for textual corpus |
| batch_size | 64 | Number of samples per gradient update |
| epochs | 3 | Number of epochs to train the model |
| loss | binary_crossentropy | Loss function of the network |
| optimizer | adam | Optimization approach of the network |
| lr | 0.001 | Learning rate |
| beta_1 | 0.9 | The exponential decay rate for the 1st-moment estimates |
| beta_2 | 0.999 | The exponential decay rate for the 2nd-moment estimates. |
| decay | 0 | Learning rate decay over each update |

**Table A2. Parameter Setting of CNN**

| Parameter | Setting | Description |
|---|---|---|
| layers | 5 | Number of convolutional layers |
| Image dimension | 4096 | Image embedding dimensionality extracted from VGG16 |
| Output dimension | 64 | Output dimensionality of CNN |
| batch_size | 8 | Number of samples per gradient update |
| epochs | 200 | Number of epochs to train the model |
| loss | binary_crossentropy | Loss function of the network |
| optimizer | adam | Optimization approach of the network |
| lr | 0.00001 | Learning rate |

| | | |
|---|---|---|
| beta_1 | 0.95 | The exponential decay rate for the 1st-moment estimates |
| beta_2 | 0.999 | The exponential decay rate for the 2nd-moment estimates. |
| decay | 0 | Learning rate decay over each update |

**Table A3. Parameter Setting of 3D ResNet**

| **Parameter** | **Setting** | **Description** |
|---|---|---|
| num_classes | 400 | The number of output classes of the model |
| shortcut_type | B | The type of shortcut to use in the model |
| cardinality | 32 | The cardinality of the bottleneck blocks in the model |
| sample_size | 112 | The spatial size of the input frames to the model |
| sample_duration | 16 | The number of frames in each video clip |
| last_fc | false | Whether to include a fully-connected layer at the end of the model |

**Table A4. Parameter Setting of LSTM-Attention**

| **Parameter** | **Setting** | **Description** |
|---|---|---|
| units | 64 | The dimensionality of the output space |
| activation | sigmoid | Activation function |
| dropout | 0.0 | Fraction of the units to drop for the linear transformation of the inputs |
| return_sequences | False | Whether to return the last output |
| return_state | False | Whether to return the last state in addition to the output |
| batch_size | 64 | Number of samples per gradient update |
| epochs | 2 | Number of epochs to train the model |
| loss | binary_crossentropy | Loss function of the network |
| optimizer | adam | Optimization approach of the network |

## Appendix B Statistical Significance Test Results on KS

**Table B1. Friedman Test and Post Hoc Dunn Test on Ideational Metafunction**

| | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | |
|---|---|---|---|---|---|
| | | Baseline | Texts | Images | Videos |
| Baseline | 4.97 | | | | |
| Texts | 2.19 | <.001 | | | |
| Images | 4.00 | <.001 | <.001 | | |
| Videos | 2.82 | <.001 | <.001 | <.001 | |
| Multimodal | 1.02 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1900.346 ($p$ < .001) | | | | | |

**Table B2. Friedman Test and Post Hoc Dunn Test on Interpersonal Metafunction**

| | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | |
|---|---|---|---|---|---|
| | | Baseline | Texts | Images | Videos |
| Baseline | 4.96 | | | | |
| Texts | 3.73 | <.001 | | | |
| Images | 2.52 | <.001 | <.001 | | |
| Videos | 2.54 | <.001 | <.001 | 1.00 | |
| Multimodal | 1.30 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1590.961 ($p$ < .001) | | | | | |

**Table B3. Friedman Test and Post Hoc Dunn Test on Textual Metafunction**

| | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | |
|---|---|---|---|---|---|
| | | Baseline | Texts | Images | Videos |
| Baseline | 4.63 | | | | |
| Texts | 4.21 | 0.003 | | | |
| Images | 3.19 | <.001 | <.001 | | |
| Videos | 1.78 | <.001 | <.001 | <.001 | |
| Multimodal | 1.25 | <.001 | <.001 | <.001 | <.001 |
| Friedman $\chi^2$: 1698.296 ($p$ < .001) | | | | | |

**Table B4. Friedman Test and Post Hoc Dunn Test on Metafunction Combinations**

| | Average Rank | Adjusted $p$-value of Pairwise Comparison | | | | | |
|---|---|---|---|---|---|---|---|
| | | Idt | Ips | Txt | Idt+Ips | Idt+Txt | Ips+Txt |
| Ideational | 3.14 | | | | | | |
| Interpersonal | 6.16 | <.001 | | | | | |
| Textual | 6.44 | <.001 | 0.864 | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Idt+Ips | 2.16 | <.001 | <.001 | <.001 | | | |
| Idt+Txt | 2.49 | <.001 | <.001 | <.001 | <.001 | | |
| Ips+Txt | 5.39 | <.001 | <.001 | <.001 | 0.337 | <.001 | |
| All | 2.21 | <.001 | <.001 | <.001 | 1.000 | 0.849 | <.001 |
| Friedman $\chi^2$: 2370.354 ($p$ < .001) | | | | | | | |