SMU Data Science Review

Volume 2 | Number 1 Article 19

2019

Dare to Venture: Data Science Perspective on Crowdfunding

Ruhaab Markas
Southern Methodist University, rmarkas@gmail.com

Yisha Wang
Southern Methodist University, yishaw@mail.smu.edu

Dare to Venture: Data Science Perspective on Crowdfunding

Ruhaab Markas[1], Yisha Wang[1], John Tseng[2]
[1]Master of Science in Data Science, Southern Methodist University, Dallas, TX 75275 USA
[2]Independant Consultant
Dallas, TX 75275 USA
{Rmarkas, YishaW}@smu.edu, wc.john.tseng@gmail.com

Abstract. Crowdfunding is an emerging segment of the financial sectors. Entrepreneurs are now able to seek funds from the online community through the use of online crowdfunding platforms. Entrepreneurs seek to understand attributes that play into a successful crowdfunding project (commonly known as campaign). In this paper we seek so understand the field of crowdfunding and various factors that contribute to the success of a campaign. We aim to use traditional modeling techniques to predict successful campaigns for Kickstarter. We find emerging field of crowdfunding has many nuances due to various funding methods of online platforms. The importance of having relevant data is further highlighted as contributing factors identified from empirical studies were not published in the data set used for modeling. With the understanding of the field we were able to identify that the lack of features provided would hinder the creation of an accurate model. In conclusion we find that the need for complete and quality data is needed to create quality models.

1 Introduction

Crowdfunding has created an alternative space for entrepreneurs to finance their ventures. Entrepreneurs, often referred to as creators, seek out funds through crowdsourcing. For some creators, traditional financing may be an unobtainable source of funding due to the restrictive nature of risk adverse institutions. In addition to determining the amount of support they need to raise from a crowdfunding campaign, creators need to understand and utilize various factors that contribute to a successful campaign. Here we aim to understand the various aspect of a crowdfunding project and later apply this knowledge to our modeling. We focus on Kickstarter because it is the source of our modeling data.

Crowdfunding in its simplest form is the act of sourcing financing from multiple investors. The amount needed to finance the project is shared among many different individuals. These individuals often individually contribute a fraction of the total funding goal. Often creators use this financing method in lieu of traditional financiers who require set lending requirements. Traditional loan applications might request that entrepreneurs provide items such as collateral and a detailed business plan. Such requirements might automatically exclude creative endeavors in the arts like recording a new music album. In the case of modern crowdfunding, creators directly reach out to potential backers using online platforms. Notable platforms include Kickstarter, Indiegogo, and AngelList[1] in the United States. Backer review various aspects of the project on the Kickstarter platform to determine if a project is a worthy investment. In this study the focus is on campaigns from the Kickstarter platform.

1.1 Parts of a Crowdfunding Campaign

There are six key observed components in a crowdfunding campaign - the platform, the creator, the project, the pledge goal, the incentive, and the backers. Creators are the entrepreneurs who initiate the project. Often, creators seek to raise seed capital for their endeavor at an early stage of the project development. Creators set the financing need based upon the project. They also are responsible for setting up campaign parameters, communicating project aspects, creating investor interest, developing investor trust, providing project updates, and ultimately completing project & distributing deliverables.

Prior the launching a crowdfunding campaign, a creator must decide which crowdfunding platform is best suited their project needs. So far there have been 4 distinct types of crowdfunding platforms. These are reward-based, equity-based, charity/donation-based, and lending-based. In addition, not all platforms function within the same funding method. For example, Kickstarter and Indiegogo are both reward-based crowdfunding platforms. However, Kickstarter uses a “all or nothing” rule. In this case creators will only receive the funds raised if total funding goal is met. Kickstarter explains that this all-or-nothing method is key to reducing risk, motivate both funders & creators, and is a proven method. On Indiegogo, creators have the option of “flexible funding” in which creators will receive funding regardless of reaching funding goal2. Some creators might already have a portion of their seed capital and are only interested in gaining additional funding. Platforms can also be geared towards specific interest areas. Kickstarter’s introductory statement states that this platform is geared towards creative projects. Platforms will vet proposals prior to implementation to ensure that projects meet their definition of a legitimate project3.

The projects are the endeavor that a creator is seeking to finance. Projects can include starting a new business, building a product, recording an album, making a new game, and more. In the case of the Kickstarter platform, there are 8 distinct categories: Arts, Comics & illustration, Design &Tech, Film, Food & Craft, Games, Music, and Publishing. Depending on platform, the categories of projects might vary. Users who

---

[1] Additional notable crowdfunding websites include Crowdfunder, RocketHub, Crowdrise, Somolend, Appbackr, AngelList, Invested.in, and Quirky. The differentiating factors of these platforms can be seen in their funding method and type of crowdfunding campaigns they support.

[2] Indiegogo also allows for a “fixed funding” which is similar to the “all or nothing” on Kickstarter.

[3] Kickstarter outlines these generalized requirements. Additional details are provided to creators.: “Projects must create something to share with others.”, “Projects must create something to share with others.”, “Projects can’t fundraise for charity.”, “Projects can't offer equity.”, and “Projects can't involve prohibited items.”.

are browsing the websites, can easily search and find projects within their interest category.

At the time of project creation, creators set a pledge goal which is the amount of capital they aim to raise. This amount depends on the project itself and is calculated by the creator. A pledge goal also includes an incentive and duration. On Kickstarter, pledge rewards are used as a tangible incentive. This is a key component of a reward-based crowdfunding model. Later in this paper we will touch upon other crowdfunding models. Investors are able to pledge funds within pre-determined thresholds set by the creator. Tied to each threshold are rewards. Forbes and Schaefer's study show that investor interest in the project is influenced by rewards and prefer rewards with meaningful connections to the actual project. For example, a comics & illustration project called "Naked Body: An Anthology of Underground Chinese Comics[4]" had a pledge of bracket of $10 + that will result in a digital copy of the book. A pledge of $22+ will result in a physical copy of the book. The reward is realized upon completion of actual project. Campaigning duration is set by creator and is mandatory per Kickstarter. Creators will have a limited amount of time to raise project funds.

Funders, or backers, are the individuals who invest in the project[5]. The financial burden needed to finance the project is shared among many different individuals. For example, Hexbot Robotics, raised $578,380 from 1,135 backers for their "Hexbot: The Modular All-In-1 Desktop Robot Arm for Everyone[6]" project. This is drastically different from the traditional financing path in which small number of (often singular) institution(s) will provide the total amount of capital (i.e. Business loan). Because of the riskiness of a new company a singular backer would require vigorous review of the business plan in order to reduce the risk of failure. There is very little requirement to be a backer[7] outside of financial ability and access to platforms. Backers do not need to come from an investment or finance background. Their motivations for supporting a project can be varied. Backers are not limited by reward thresholds since they can choose to pledge funds without reward. Because of these low limitations a wide scope of people is given access to investing in projects, startups, and new innovations that they previously did not have access to. Conversely creators have also been given more options of financing their ideas.

1.2 Traditional Financing vs. Crowdfunding

Crowdfunding gives creators access to the general public as alternative way to finance their new endeavor. Traditional financing requires creators to submit detailed business plans and sometimes collateral. Traditional investors or financiers required such detailed reviews because of the inherent riskiness of a startup. Startups can fail due to a number of reasons. In the case of failure, the institution would bear the brunt of the

[4] This project is a collection of cartoons drawn by independent cartoonists from China. Pledge goal was $8,000.00 but raised $13,612 with 352 backers. Campaign lasted from January 29,2019 and ended on February 27,2019.

[5] In this paper we will use the term backers interchangeably with investors.

[6] As of February 24, 2019, with 12 more days left in campaign duration.

[7] Backers of equity-based crowdfunding have more requirements than traditional reward based or charity-based models.

loss. This risk adverse behavior limits the type of projects that are funded. With crowdfunding, the loss is divided among many different backers. This is not to say that backers thoughtlessly fund interesting projects. Forbes and Schaefer show that backers correlate the quality of a campaign to the potential success of a project. Signalers such videos, grammar, active communication, and more are used by backers to decide on pledging their support.

In addition, Traditional financing such as a business loan requires creators to pay back funds with interest. In the case of crowdfunding platforms like Kickstarter, creators pay a certain percentage of total funds raise (given that the pledge goal is reached) and a payment processing fee per pledge. Each platform will provide guidelines of the accountability and liability of a creator. For example, for our given platform, liability is outlined to fall upon the creators and not on the platform. Creators are to be responsible for delivering rewards or refunding. Although, it seems unclear when a project has failed or still in progress despite missing deadlines.

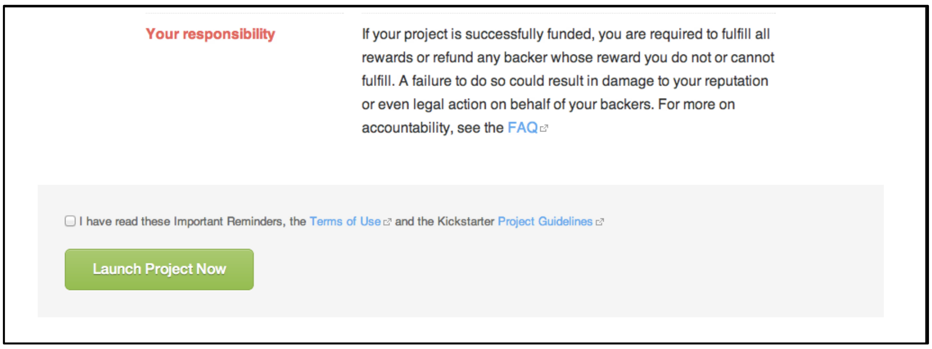

Fig. 1. Kickstarter disclosure outlining creator responsibility upon fulfilling rewards.

1.3 Types of Crowdfunding Business Models

There are three overarching types of crowdfunding models: charity-based, equity-based, and for-profit-based. For-profit types of crowdfunding can be broken down into smaller subsections as reward-based and donation-based. These would include peer-to-peer lending and micro-lending. Kiva, Kickstarter, and Indiegogo are examples of for-profit model.

Equity-based models are a much newer type of model which was made available since 2012 with the implementation of the JOBS Act8. Platforms for equity-based models include AngelList and EquityNet. For the scope of this paper, we focus on a reward-based model because of the clear transaction between creator and backer. In addition, dataset for equity-based models were difficult to locate.

Unlike reward-based for-profit platforms, charity-based platforms such as GoFundMe differ in the behavioral motivations of the backer. Backers are not

[8] Also known as Just Start Our Business Start up act.

specifically investing in a project but rather giving donations without a specified return. For example, GoFundMe has a lot of campaigns raising money for victims of tragic situations. The drivers for donations can more emotionally or sentimentally driven and can vary depending on backers. This makes influencing features difficult to quantify. In research done by Forbes and Schaefer, backers of reward-based projects are more discerning investors. These backers looked for sound business models and consequently, projects that will result in long term success. In this study we aim to leveraging Kickstarter crowdfunding data to predict successful crowdfunding projects.

Kickstarter is a reward-based "all or nothing" crowdfunding platform that focuses on creative innovations. With 8 overarching categories of projects. As of February 24, 2019, the website has raised $3.68 billion dollars in successful campaigns. The platform has successfully funded 158,411 unique projects with majority of funding goals raising with $1,000 to $9,999 (86,879 projects). The music category holds the largest count of successful project. This is followed closely by Film & Video. There have been 273,224 unsuccessful projects, 20% of which raised 0% of the funds. Largest bucket of unsuccessful projects raised between 1% to 20% of their funding goal (172,721). The category of Film & Video has the largest category of unsuccessful projects. The platform is widely used with 15.80 million backers with a 3rd being repeating backers. The statistics of the website are refreshed daily. Kickstarter self-published the data set used in this research project.

2 Subject Matter Research

2.1 Prior Research

A large majority of prior research completed on crowdfunding were focused in economic or behavioral studies. We hope to differentiate ourselves from those studies by apply data science approaching and producing a predictive model. Prior research has aided in allowing us to have a deeper knowledge of crowdfunding. We build upon this to make educated decisions in determining the correctness of outcomes and adjusting for outliers.

Research by Forbes and Schaefer, "Guidelines for Successful Crowdfunding", we see that crowdfunding is a worthy of academic research as it is an exponentially growing market. Few research studies have been found that incorporates predictive modeling to determine successful crowdfunding projects. The results of these few studies are not freely published. We will evaluate valid features selected for modeling by incorporating qualitative analysis of prior studies. In additional we would be able to leverage the prior research results to better tune our model. Also, we can reasonably assume more additional and detailed studies into the different business models of crowdfunding will be needed in the future as the field continues to expand. For example, the expansion in the volumes of campaigns can be seen with the growth shown by Forbes and Schaefer in the UK market. They point out that from 2011 to 2012, the UK volume of campaigns increased by 91%. Also, unknown expansion in the emergence of equity-based crowdfunding.

When determining to use data from various different crowdfunding platforms we selected the Kickstarter platform for research because of the popularity of use and data availability. There are other large platforms such as Indiegogo that are not incorporated into our modeling. According the Forbes and Schaefer study, users of varying platform show more trust in the Kickstarter platform. In addition, the differing funding methods of the two platforms can cause potential unknown variational effects on the results because of perceived risks by the backers. On Kickstarter, backers will not remit funds to creators unless the fund amount requested is backed. Therefore, a clearer distinction of success or fail.

The behavior of both the creators and funders are needed to determine the success of the projects. Forbes and Schaefer also argue that the business model of the platform also is a strong driver of outcome. For example, from their qualitative study, they found that funders are swayed to further funds projects that other backers have already funded. They perceive that the greater percentage funded are projects of better quality thus influencing overall success. This perception is blind to the reality of projects that require larger funding goals in which a pledge would have a smaller influence on percent increase. In short, backers are more likely to invest in a project when they see others funding the project. Kickstarter notes this behavior as an advantage and encourages both backer and creator to promote their project[9]. This can be both beneficial side effect and a deterring factor for projects that require larger seed capitol amounts. On one hand backers encourage or "influence" other backers to fund projects they deem worthy. On the other hand, larger projects might not get fully funded even though they raised larger actual amount. Here it is seen that backers are deterred by larger amount projects because of the perception of lower percent funded. However, despite of this Cumming, LeBeouf, and Schweinbacher argue that larger non-scalable projects are more likely to be successful when using the all or nothing projects due to the reduced risks for the backer.

The reward incentive is another aspect of backer behavior. Backers responded better to rewards that were of actual value rather than gimmicky items (such as a branded sticker or t-shirt). An example of this was discussed earlier in the paper.

Backers also indicated that marketing strategy is also very important in determining if they would fund a project. This includes the use of an informative videos. Forbes and Schaefer offer suggestions based off of their research to aid in the creation of their project video. Kickstarter also strongly suggests to creators to create video as well. The videos purpose is the inform backer about the project, connect with the backers, and to display the potential of the project. Those who create a poorly written or conveyed projects page and video are perceived be a lower quality project thus risky to invest in. These communication methods are markers of quality.

Duration of campaign is also an influencer of the successful funding of a project. This is a required feature of the platform. The timeline is between 0 to 60 days. Kickstarter FAQ suggests for campaigns to be between 0 to 30 days because "Campaigns with shorter durations have higher success rates, and create a helpful sense of urgency around your project". Solomon, Ma, and Wash evidence that most backers will pledge funds within the first 20 days of a campaign. The next largest volume of

[9] Additional details on this topic can be found on Kickstarter 101 page titled "Why is funding all – or – nothing?"

backers happens within the last days of the project. They conclude that having project timeline cause inefficiencies for projects. Backers behavior show that they withhold funding projects till closer to the deadline. One major adverse effect is that by withholding funds, backer delay signaling to other potential backers to also fund the project. Solomon, Ma, and Wash also recommend backers to donate early on in the project rather than withholding support in the end. Backs ultimate desire the overall success of the project they support. Those who care more about the project and donate at the end of the project often donate more than the average backer. The only benefit to those who pledge more than other is through the reward. Often times higher money amounts will be tied into larger rewards.

Because of these behavioral factors, those seeking to fund their own projects should consider the type of platform, funding goals, reward options, and marketing strategy (i.e. video) according to Forbes and Schaefer.

2.2 Benefits and Drawbacks

Benefits for utilization of crowdfunding for seed capital can be varied for different types of creators. For creators who are not experienced in entrepreneurial endeavors, the crowdfunding platform are away to receive seed capital without having to seek out banks or venture capitalists. Backers themselves will have vetting standards but might not be a vigorous or rigid as traditional investors. They are also more likely to invest in those they are more passionate about even when they do not meet set standards. Overall has created an alternative method for financing, especially for project outside of the traditional startup/entrepreneurial scope.

For creators who are more established, such as a small company, benefit by creating exposure to potential customers. These creators might also utilize crowdfunding to fund riskier or experimental projects. For example, Cotopaxi[10] used Kickstarter to fund their creation of the Libre sweater. By funding this project through crowdfunding, Cotopaxi was able to receive the seed money they needed to make the sweater[11]. The reward served as an early bird special for a sweater. In addition, Cotopaxi was able to gauge and pique consumer interest online[12].

Creators can also utilize crowdfunding by evidencing proof – of – concept in order to receive formal investments from traditional investors. In this way creators prove that

[10] Cotopaxi is a small athletic wear company based in Salt Lake Utah. As of January 2019, they had successfully funded 3 projects on Kickstarter. 2 realized projects and 1 cancelled project.

[11] The Libre sweater had an initial campaign goal of $20,000. At the end of the campaign the project had successfully raised $389,890.00 (approximately 20 times the requested amount).

[12] After the Kickstarter campaign, had ended consumers were able continue funding the project on their Indiegogo campaign. The consumer could still fund for a reward of a sweater. Today, the Libre sweater is sold on the Cotopaxi store page as a regular merchandise. Later project for a travel backpack called “Allpa” used Indiegogo. Indiegogo is not an all- or – nothing platform like Kickstarter. This shows that there is a learning curve for creators alike to know which platform and crowdfunding model works best for their projects. Cotopaxi has funded 4 projects on Indiegogo.

there is significant interest in their project from consumers. These cases creators do not need to receive all seed capital from their campaign but rather fill in a financial gap.

There are drawbacks identified for using crowdfunding. The first notable drawback is that creators must understand and determine the optimal platform for their project needs. Each platform draws in different types of backers. In the case of using crowdfunding as a proof-of-concept, some traditional financiers view crowdfunding campaigns to signal lower quality endeavors. The riskiest draw back for crowdfunding is the potential for legal repercussions for failing to deliver rewards in reward-based projects. The onus of the project reward is upon the creator not platform.

3 Methods

Here we need to define what is means to have a successful campaign. We find that there are 2 distinct kinds of success. The first success of a creator is to raise total or more amount of funds requested on the platform. This first success is binary – if funding goal is reached, then success. If funding goal is not reached, failure.

The second type of success is the aim of all creators - to have their project be completed with ultimate goal to be reached whether it be a product, charity, song, game, etc. This kind of total success in dependent upon many factors outside of funding. For backers, this would mean to receive the promised reward. These pieces of the puzzle include quality of communication, timeline, actual demand for product, quality of leadership from creators, and etc.

For the purpose of this study, we focus on success type 1. The outcome of our predictive mode is to predict if a project will reach the funding goal or not. Success type 2 is beyond the scope of the data. The data does not indicate if reward deliverables were completed.

4 Data

4.1 Data Collection

The data used for this project was published on Kaggle by Kickstarter. Two CSV files were available holding data from 2016 and 2018. Files varied in column size and had to be merged. Duplicates were removed. Missing values were addressed as well. Only 1% of the data provided were missing values. Because of this we chose to delete those rows rather than imputing due to time restrictions. Most of the missing values were in the column "USD pledged". After cleaning data and removing redundant features, we were left with 13 variables and 319K rows. Column "state" was the variable that indicated the success/failure of the project.

There was some key limitation in features in the provided data set. Influencing variables that were previously identified by prior search such a marker of quality (video, communication, shares, etc.) were not present. Duration of campaign was able to be extracted via date difference between start and ending date. In future attempts in modeling campaign successes, we suggest that data be available that includes features such as number of shares, video, number of prior successful campaigns given creator, frequency of communication, and other identified qualitative features.

4.2 Data Analysis

Exploratory Data Analysis (EDA) is one of the first steps in a data science workflow once the data has been acquired. As the name suggests, EDA refers to the process of learning about the data through visual analysis. Domain expertise is critical to explore specific aspects and questions of the data, but more general approaches can also be applied if there is less domain familiarity. Here we will complete an exploratory analysis of the data that was mined and wrangled. It's important to explore the data to further understand of the measures of center and variance. These basic descriptive statistics can provide simple but powerful statistics on the variables of interest. However, combining categorical data can also provide necessary insights to understand the context of the data with which one works with.

The fundamental nature of crowdfunding data is financial. This data set contains data on the amount of funding received for various projects. One of the preliminary bits of analysis to explore is funding over time and by category. Looking at funding data will give insights into the success and adoption of the platform as a mechanism of raising funds.

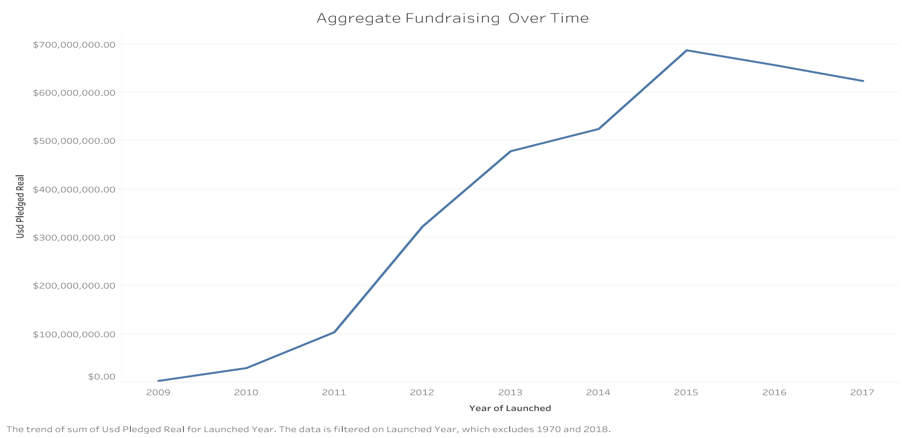

Fig. 2. Amount of funds raised by year excluding 2018 (incomplete year).

Looking at the funding over time gives us some simple but powerful insights. The platform has seen year-over-year growth in the amount of funds raised on an annual basis. The total amounts of funds raised exceeds $3.8B with 2015 accounting for the biggest year. However, there seems to be a downward trend with a gradual decay in 2016 and 2017. We will not explore the causal reasons for this decline, but this could be an important successive analysis to pursue.

Figure 3 illustrates the top categories sorted by the amount of funding they receive from backers. Games, design, technology, film & video, and music comprise the top five categories. There is little statistical separation after the top five. A pareto distribution indicates that the top five categories account for ~80% of overall funds raised.

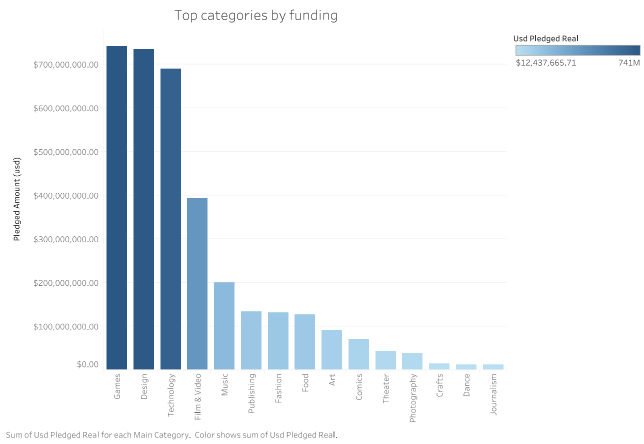

Fig. 3. Distribution of the top categories by the amount of funding received.

The amount of funding that a category receives could be driven by a variety of different factors. One factor could simply be the number of projects within each respective category. The underlying assumption here being that the higher number of projects a category contains, the more likely it is to raise higher funds. Figure 4 indicates that there is a positive linear correlation between these two variables. As the number of projects increases, so does the aggregate amount of funds raised overall. There are three outliers in this scatter plot all of which belong to the top three categories by overall funding amount. Generative research would be required to understand the causal reasons for this popularity, but it is evident that the top three categories are also among some of the most popular projects by the number of overall campaigns. Another interesting anecdote here is that even though the category film and video have the most number of projects, it is only the fourth largest category in terms of overall raised funds.

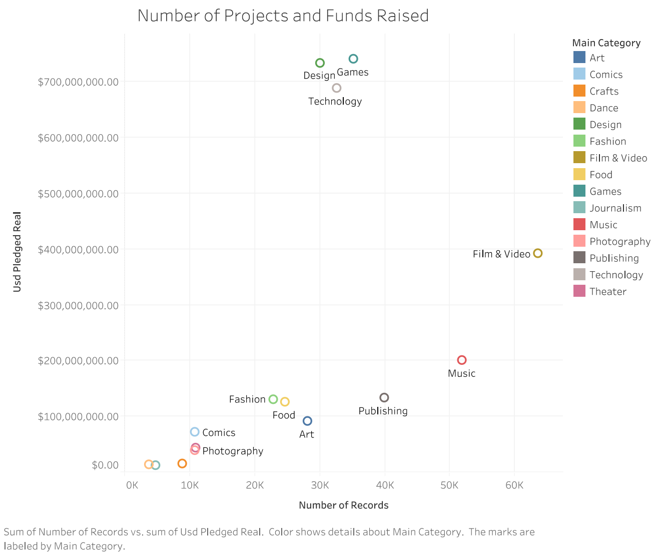

Fig. 4. Number of projects by the sum of number of projects compared to the amount of aggregate funds raised.

5 Statistical modeling

5.1 Model Building

The primary question we are searching to answer is how likely is it that a particular project will be funded. This can be a crucial insight as entrepreneurs raise capital in early stages of corporate development. Understanding this likelihood can help entrepreneurs optimize their strategy and approach to fundraising so that they can be successful in their efforts to fund their ventures. There are two main models that are used to predict this outcome and a stretch model depending on our learning and exposure: Logistic Regression, Random Forest, and Feed Forward Neural Networks. These first two models are considered shallow" models. Deep learning and neural networks involve multiple layers of non-linear processing which is not utilized in shallow models. A simplistic way to delineate these two approaches is that anything that is not classified as a deep learning approach can be considered shallow. Shallow models have their benefits. They tend to be more performant requiring less CPU power, require less data, and are usually more interpretable. All these factors are important considerations when evaluating model selection. Performance and availability of training data are self-explanatory, but why does a model output and decisioning need to be interpretable? Certain regulators have actually deemed this to be a critical component of machine learning[27]. The General Data Protection Regulation (GDPR) is a European Union data privacy law which has recently banned the use of modeling techniques that are not human interpretable.

Article 22. Automated individual decision making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision:
   (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;
   (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
   (c) is based on the data subject's explicit consent.
3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) apply and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

Fig. 5. Excerpt from Article 22 of GDPR**.**

This statute bans decisions based solely on automated processing that is not interpretable by humans. Financial institutions are most widely impacted from these laws as banks cannot use deep learning models to approve or decline decisions using black-box approaches like deep learning. As mentioned previously, this is where shallow algorithms are viable alternatives. We'll explore each of these models in more detail and results will be discussed in the subsequent section.

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and other independent variables. Logistic regression does not try to predict the value of a continuous numeric variable and is not technically considered a regression model but does bear the name. It is a classification model with a probabilistic output that a given input belongs to a certain class. In our example this is a binary class of being funded or not being funded. The assumptions of linear regression require minimal collinearity, which assume independence of the input variables. Logistic regression is simply a non-linear transformation of the logistic regression model. The logistic distribution is an S-shaped function and contains the predicted probabilities between 0 and 1.

Random Forest is an ensembling method meaning multiple machine learning algorithms are used to obtain better predictive performance. The foundational building blocks of Random Forest are Decision Trees. Hence the name forest which is taken from the numerous trees used in the ensemble to arrive at the predictive output. Decision Trees use a tree-based model to consider possible consequences and outputs and the probability that these events occur. The model understands the critical features and which conditions the tree splits to ultimately arrive at the prediction. Ultimately, numerous decision trees are combined into a single model. By pooling predictions from numerous models, we can incorporate many more inputs based on different thresholds and conditions that the various trees were split on. This ultimately yields a more generalized prediction as opposed to evaluating the output of a single tree.

6 Results

Efficacy of these models will be evaluated using conventional machine learning error functions such as accuracy, precision, recall, and F1 which is the harmonic mean of these two scores. We'll explore what these metrics actually mean and also evaluate the actual outputs of the models across these dimensions as well.

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all the funded projects, how many were actually funded? High precision relates to the low false positive rate. Recall is also referred to as sensitivity which is the ratio of correctly predicted positive observations to the all observations in actual class. The question recall answers is: of all the passengers that were funded how many did we correctly predict as being funded? F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 can sometimes be more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at a more general metric like accuracy.

Accuracy results of the Logistic Regression and Random forest are .646 and .652 respectively. Given these models do not have particularly high results, we would like to explore more advanced technique like feed forward neural networks to achieve better results.

7 Ethical Analysis

Leveraging data science to determine successful crowdfunding campaigns will aid creators to reach their funding goals. There are ethical concerns in regards to relying upon these kinds of models. For example, creators might be deterred from posting projects in which there is low likelihood of success. This begs the question of how much importance users place upon imperfect algorithms verses human intuition & hope. This boils down to how much we as a society will let these types of result tell us how to aim our goals.

Users need also to keep in mind that the success of a project is not solely dependent upon reaching funding goals. Overall success of a startup depends upon the quality of the project and creator. Outside factors that are not predicted for can affect overall success. Things like market trends and consumers might shift away from the product they create.

Another ethical issue that arises is the potential for fraud or misuse. Creators might be tempted to start a campaign with the pursuit of monetary gain without intentions of ever delivering a product. Because there is no actual collateral or strictly define repercussions, fraudulent creators might not ever be persecuted for their dishonest actions. Credibility of a creator might not be transparent if they are new. Because of the nature of investing in startups, it is difficult to differentiate between fraudulent campaigns and those that have had significant delays in production. Backers might not be able to fully realize the situation until years down the line. On the other hand, harm to the credibility of honest creators and platform risk reputational damage due to dishonest creators.

Another ethical risk from reward-based crowdfunding is that backers assume that when funding projects for reward that the project will yield results. Vargras, Desari, and Vargras show an example of project failure past successful funding which led to eventual suing of creator for fraud when the deliverable was not met. Backers perceived that funding would ensure overall project success. In response to harmful losses to investors. The SEC has implemented regulation that limits on amount to raise, amount to invest, and other aspects of crowdfunding campaign. With the evolving and growing crowdfunding sector, regulator need to address changes in spaces such as lending-based or equity-based crowdfunding.

8 Conclusion

From this project we explored the emerging space of crowdfunding. Crowdfunding has created an alternative way for creators and backers to develop and invest. Prior to availability to online platforms creators seeking non traditional financing was limited to their network and geographical location. With the help of the internet, the public has been included in this entrepreneurial sphere.

This research was limited by the features provided in the data set. With the knowledge provided by prior research, we found that our data needed to include features that were deems important from qualitative research. This research was also limited to one reward-based platform and thus could not predict success outside of this platform.

Future works in this field is need for the different types of crowdfunding business models. Additional study is needed in the equity-based crowdfunding campaigns in both performance and impact. Behavioral studies in the motivation that drives backers for non-reward-based platforms is also needed. Overall more data is needed to make future research possible and meaningful. With the inclusive nature of the space, there are multiple features that need to be taken into consideration that could results in meaning findings.

References

1. Kickstarter 101, Kickstarter, PBC, 2018, www.help.kickstarter.com/hc/en-us/articles/115005047893-Why-is-funding-all-or-nothing-
2. Barnett, Chance. “Top 10 Crowdfunding Sites for Fundraising” Crowdfunding Systems Update. 2015
3. Gabison, Garry A. “Understanding Crowdfunding and its Regulations” JRC Science and Policy Report. European Union, 2015.

4. Christopher Courtney Supradeep Dutta Yong Li. “RESOLVING INFORMATION ASYMMETRY: SIGNALING, ENDORSEMENT, AND CROWDFUNDING SUCCESS” School of Management State University of New York at Buffalo
5. Qianzhou Du, Weiguo Fan, Zhilei Qiao, Alan Gang Wang, Xuan Zhang, Mi Zhou. “Money Talks: A Predictive Model on Crowdfunding Success Using Project Description” AMCIS. 2015
6. Jacob Solomon, Wenjuan Ma, and Rick Wash. “Don’t Wait! How Timing Affects Coordination of Crowdfunding Donations” Computer Supported Cooperative Work (CSCW). Vancouver, BC. March 2015.
7. Anbang Xu1, Xiao Yang2, Huaming Rao3, Wai-Tat Fu1, Shih-Wen Huang4, Brian P. Bailey. “Show Me the Money! An Analysis of Project Updates during Crowdfunding Campaigns”
8. Garry Bruton Susanna Khavul Donald Siegel Mike Wright. “New Financial Alternatives in Seeding Entrepreneurship: Microfinance, Crowdfunding, and Peer-to-Peer Innovations” Baylor University January,2015
9. Elizabeth M. Gerber, Julie S. Hui, Pei-Yi Kuo. “Crowdfunding: Why People Are Motivated to Post and Fund Projects on Crowdfunding Platforms” Northwestern University
10. Hemer, Joachim (2011): A snapshot on crowdfunding, Arbeitspapiere Unternehmen und Region, No. R2/2011, Fraunhofer ISI, Karlsruhe
11. Gregory C Makris. “Crowdfunding: from startup businesses to startup science”: BMJ: British Medical Journal
12. , Vol. 350 (12 Jan 2015 - 18 Jan 2015)
13. Ajay Agrawal, Christian Catalini and Avi Goldfarb. “Simple Economics of Crowdfunding” The University of Chicago Press on behalf of the The National Bureau of Economic Research. Innovation Policy and the Economy
14. , Vol. 14, No. 1 (January 2014), pp. 63-97
15. Tim Grove “History Bytes: Experiments with Crowdfunding” American Association for State and Local History. History News, Vol. 70, No. 4 (AUTUMN 2015), pp. 5-6
16. Peter J. Loughran, Lee A. Schneider, Ebunoluwa A. Taiwo and Gabriel W. Lezra “The SEC Hands Out a Halloween Treat to Crowdfunding Supporters” Business Law Today, American Bar Association. 07 Dec 2015
17. Frank Vargas, Jennifer Dasari and Michael Vargas. “Understanding Crowdfunding: The SEC's New Crowdfunding Rules and the Universe of Public Fund-raising” Business Law Today, American Bar Association. 04 Dec 2015
18. Mohammadi, Al. Shafi, Kourosh. “Gender differences in the contribution patterns of equity-crowdfunding investors”. Small Bus Econ, 20 January 2017.
19. Cumming, Douglas J., Leboeuf, Gael, Schweinbacher, Armin. “Crowdfunding Models: Keep-It-All vs All-Or-Nothing” York University–Schulich School of Business. 31 May. 2015
20. Forbes, Hannah. Schaefer, Dirk. “Guidelines for Successful Crowdfunding”. Elsevier. Procedia CIRP, 9 May. 2017
21. Kickstarter Stats, Kickstarter, PBC, 2018, www.kickstarter.com/help/stats?ref=about_subnav.
22. Kickstarter Projects, Kaggle Inc., 2017, www.kaggle.com/kemical/kickstarter-projects.
23. Mollick, Ethan. “The Dynamics of Crowdfunding: An Exploratory Study.” Elsevier. Journal of Business Venturing, 13 Aug. 2013, www.sciencedirect.com/science/article/pii/S088390261300058X.
24. Marco Sahm, Paul Belleflamme, Thomas Lambert, Armin Schwienbacher Corrigendum to “Crowdfunding: Tapping the right crowd” Journal of Business Venturing, Volume 29, Issue 5, September 2014, Pages 610-611

25. Belleflamme, Paul, et al. “Crowdfunding: Tapping the Right Crowd.” Journal of Business Venturing, Elsevier, 29 Sept. 2013, www.sciencedirect.com/science/article/abs/pii/S0883902613000694.
26. Agrawal, et al. “The Geography of Crowdfunding.” NBER, The National Bureau of Economic Research, 25 Feb. 2011, www.nber.org/papers/w16820.
27. Han, et al, “Will New EU Regulations Starve Data-Hungry Deep Learning Models?” The Medium, Jan 31 2018, https://medium.com/syncedreview/will-new-eu-regulations-starve-data-hungry-deep-learning-models-25403795d26c