Computing (2021) 103:1939–1958
https://doi.org/10.1007/s00607-021-00926-w

SPECIAL ISSUE ARTICLE

Popularity versus quality: analyzing and predicting the success of highly rated crowdfunded projects on Amazon

Vishal Sharma[1] · **Kyumin Lee**[2] · **Curtis Dyreson**[1]

Received: 15 August 2020 / Accepted: 15 February 2021 / Published online: 28 May 2021

Abstract

Crowdfunding is a process of raising money (*funding*) for a project through a venture of large number of people (*crowd*). The popular online crowdfunding platforms Kickstarter and Indiegogo provide a stage for innovators worldwide to bring ideas to reality. Despite the popularity and success of many projects on the platforms, it is yet to be determined *whether successful projects always produce high quality products.* Previously, the quality of crowdfunded products (successfully funded projects from crowdfunding website that are available on Amazon) in the market (e.g., Amazon) has not been statistically and scientifically evaluated. There has been no previous study to understand whether a successful project will receive high/low ratings from customers in e-commerce sites like Amazon. To address this problem, we (i) compare crowdfunded products with traditional products in terms of their ratings on Amazon; (ii) analyze negative reviews of crowdfunded products; (iii) analyze characteristics of the successful projects (received ≥ 4 Amazon rating) and unsuccessful projects (received < 4 Amazon rating); and (iv) build machine learning models at three different stages, to predict high or low star ratings for a crowdfunded product. Our experimental results show that, on average, crowdfunded products received lower ratings than traditional products. Our ensemble model effectively identifies which product will receive high star-ratings from customers on Amazon. The dataset and code used in this manuscript are available at https://github.com/vishalshar/popularity_vs_quality_data-code.

Keywords Crowdfunded projects · Kickstarter · Amazon rating prediction

An early version of this manuscript appeared in the 2018 IEEE/ACM Proceedings of the ASONAM conference.

✉ Vishal Sharma
vishal.sharma@usu.edu

Kyumin Lee
kmlee@wpi.edu

Curtis Dyreson
curtis.dyreson@usu.edu

1 Department of Computer Science, Utah State University, Logan, UT, USA

2 Department of Computer Science, Worcester Polytechnic Institute, Worcester, USA

Mathematics Subject Classification 68T09

1 Introduction

Crowdfunding provides creators with opportunities to find investors, refine their ideas based on other users' feedback, have early adopters and potential brand advocates. The investors (backers) can fund interesting ideas and receive the products at an early stage at a lower price. Kickstarter is the most popular reward-based crowdfunding platform. Some of the very successful projects on Kickstarter are Pebble Time ($20.3M pledged), Coolest Cooler ($13.3M pledged), and Kingdom Death ($12.4M pledged). On one hand, crowdfunding provides creators an opportunity to get their ideas funded. On the other hand, there are reports of drawbacks of crowdfunding, for example, 9% of creators failed to deliver rewards/products that they promised to their backers [17] and 35% backers did not receive rewards on time, passing estimated delivery dates [39].

A reward-based crowdfunding project has three phases as shown in Fig. 1: (i) the *fundraising phase*; (ii) the *reward delivery phase*; and (iii) the *product sale phase*. In the past, researchers have studied only the first two phases, *fundraising* and *reward delivery*. Previous research has not considered the *product sales phase*. According to our study, successful projects that raised more money than their goal did not guarantee to produce high-quality products. This observation makes the phase *product sales phase* an important aspect. For example, Pebblebee Finder[1] in Fig. 2, a Kickstarter project, raised $214K (21 times more than its goal), but its product received low ratings (only 3.1) on Amazon. Due to the limited amount of available data, it is hard to predict which project will produce high/low quality product in the market. However, if we can predict with reasonably high accuracy, creators can further improve their projects, backers can support many promising projects (which will produce high-quality products), and buyers can purchase high-quality products after they are launched into the market. In this paper, we analyze the quality of crowdfunded and start-up products and predict which products will receive a high or low rating. Since it is not possible to measure the quality of a product objectively, we use customers' explicit feedback (e.g., star ratings) to measure a product's quality.

Recently, Amazon launched a marketplace named Launchpad[2] for listing crowdfunded and start-up products. By collecting and analyzing over two years of data from Launchpad, we formulated the following research objectives: (**RO1**) compare Launchpad products with traditional products (i.e., non-Launchpad products) on Amazon in terms of customer ratings; (**RO2**) analyze negative reviews of a Launchpad product to understand why customers did not like the product and learn about the product's drawbacks; (**RO3**) analyze characteristics of successful products (i.e., highly rated products) and unsuccessful products (i.e., lowly rated products); and (**RO4**) build machine learning models to predict which crowdfunded products will be rated high or low at the time of launch into the market.

1 https://kck.st/29yRJjZ.

2 https://www.amazon.com/Amazon-Launchpad/b?node=12034488011.

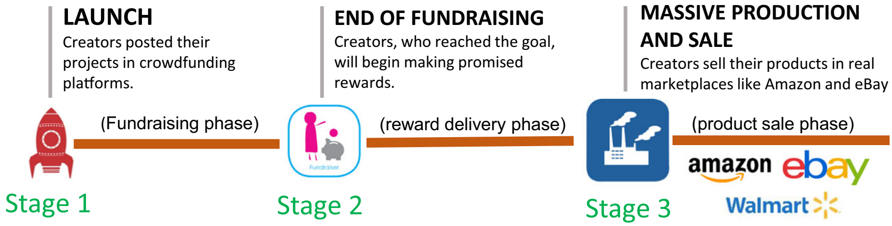

Fig. 1 Three phases of a crowdfunding project [33]

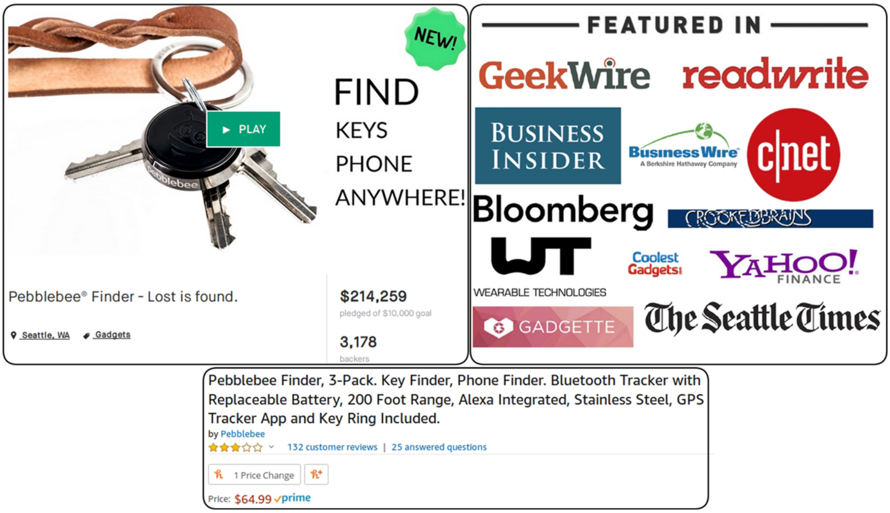

Fig. 2 PeebleBee project on Kickstarter (top) and low rating (3.1) on Amazon (bottom)

Concretely, we make the following contributions:

– First, we compare crowdfunded products with over 300 million traditional products on Amazon and show that crowdfunded products are rated lower than traditional products.
– Second, we analyze negative reviews associated with crowdfunded products where customers expressed quality issues, poor customer service, and regret of the purchase.
– Third, we analyze distinguishing characteristics of successful (≥ 4 Amazon rating) and unsuccessful (< 4 Amazon rating) products.
– Fourth, we build machine learning models to predict which crowdfunded products will receive high/low star-ratings in the market.
– Finally, we make our data, code, and sentiment analyzer available for reproducibility of our work.[3]

[3] https://github.com/vishalshar/popularity_vs_quality_data-code.

2 Related work

In this section, we summarize some of the previous work related to crowdfunding websites. We discuss work related to predicting project success, factors affecting success, relationship of social network with project quality, and recommendations of projects to backers and creators.

Researchers have studied Amazon reviews for predicting product rating, using the text of reviews or annotations of the reviews. For example, Tang et al. [35] developed a neural network-based method that uses both reviews and author information. Gupta et al. [13] predicted ratings by using supervised learning. Qu et al. [30] proposed a bag-of-opinion technique and built a ridge regression model for a review rating prediction. Cheng at al. [5] extracted aspects of user reviews and proposed an aspect-aware rating prediction approach. Our method does not require reviews. Instead, it leverages information from crowdfunding website and predicts whether a product will be successful or not as soon as the product page is created on Amazon.

Crowdfunding platforms have been studied widely in recent years [3,6]. Greenberg et al. [12] compared different types of crowdfunding platforms. Solomon et al. [11] analyzed why and how crowdfunding works and why people create projects on these platforms. Kuppuswamy et al. [18] studied the behavior of Kickstarter backers. Dey et al. [8] studied the impact of a video campaign and the factors of a persuasive video campaign. There have been several studies related to the prediction of project success. Researchers [20] built classifiers based on directly extracted features from Kickstarter to predict the success of a project. Etter et al. [9] proposed a method to predict project success by using direct information and social features. Mitra et al. [26] studied a corpus of 45k projects and proposed novel text features for the prediction. Solomon et al. [34] discovered that early donation is the best strategy for an investor. Joenssen et al. [16] studied the scarcity management of 42k Indiegogo projects and concluded that it hinders achievement of a project's goals. Lee et al. [19] applied seq2seq deep neural network with sentence-level attention to process project-related text data for project success prediction.

Factors impacting a project success have also been studied. Joenssen et al. [16] showed that timing and communication were important factors affecting project success. Xu et al. [40] extensively studied how project updates are positively correlated to project success. Tran et al. [37] showed different factors associated with making a project successful. Geographical factors, project updates, and rewards have also been shown to impact a project's success on crowdfunding websites [10,34,40]. Other researchers [15,22] analyzed the role of social media and social communities in raising funds. Mollick et al. [27] studied the dynamics of success and failure of crowdfunding products and found that social and personal networks and project quality were associated with crowdfunding success. Tian et al. [36] studied critical factors related to project success in two Chinese crowdfunding platforms. Lin et al. [21] analyzed the competitiveness of projects on crowdfunding platforms as another factor. The recommendation of projects to backers and backers to creators has also been studied. An at al. [1] built a recommender system to recommend potential backers to creators. Rakesh et al. [31] built another recommender system to recommend potential projects to backers. Rakesh et al. [32] proposed a recommendation model to recommend projects to a group of investors.

Overall, researchers did not pay attention to the performance of crowdfunded products after they have been successfully funded. To complement the prior work, we collect data from Kickstarter and Amazon to study the performance of successful crowdfunded projects on Amazon. We compare and analyze characteristics of successful (i.e., rated high stars) and unsuccessful (i.e., rated low stars) products and then build machine learning models to predict a product's success. To our knowledge, we are the first to explore this topic and area of research.

3 Dataset description

Verifying whether a product in an e-commerce site (e.g., Amazon, eBay, and Walmart) was funded from crowdfunding platforms is not a trivial task because it is hidden information. In July 2015, Amazon launched a new marketplace called *Launchpad* to help crowdfunded startups bring products to the market. The Launchpad page lists products created via crowdfunding platforms like Kickstarter, Indiegogo, Hax, and CircleUp. We collected information on 3082 products from the Launchpad page, including associated reviews. Out of 3082 products, 2117 products had at least one review. Among 2117 products, 375 products were crowdfunded from Kickstarter. We also collected these products' Kickstarter project descriptions by searching and linking each product to Kickstarter's respective campaign page. In addition, we collected backers' location information obtained from the community page for Kickstarter projects as well as backers' comments. Table 1 presents statistics of the 375 Kickstarter projects, showing the mean number of FAQs, the number of collaborators, and pledged money. We observed that 68.5% of project creators' accounts were verified, 51.2% creators connected their Facebook pages, and 40.2% creators connected their Twitter accounts.

To address the first research objective (RO1), we obtained Amazon product dataset which consists of 82 million products listed on Amazon until July 2014 [24] a well as another Amazon dataset consisting of 230 million products listed on Amazon until

Table 1 Statistics of kickstarter dataset

Properties	Mean	Range
|FAQs|	5.5	[0–62]
|Collaborators|	4.04	[0–10]
Pledged money	$387,050	[$1,016–$20,338,986]
|Facebook friends|	508.5	[0–4998]
Product price on Amazon	$91.66	[$0.99–$1898.25]
Boolean properties	**True**	**False**
Account verified	275 (68.5%)	118 (31.5%)
Facebook connected	192 (51.2%)	183 (48.8%)
Twitter connected	151 (40.2%)	224 (59.7%)

Table 2 Rating distributions of traditional products in *Amazon 1* and *Amazon 2* datasets and the Launchpad products

Rating	|prod. in Amazon 1| [24]	|prod. in Amazon 2| [29]	|prod. in Launchpad|
1.0	4,265,230 (5.2%)	17,921,671 (7.8%)	27 (1.2%)
2.0	6,712,117 (8.1%)	10,817,447 (4.7%)	108 (5.1%)
3.0	7,049,301 (8.5%)	17,584,299 (7.6%)	685 (32.4%)
4.0	15,480,820 (18.7%)	36,949,729 (16.1%)	961 (45.4%)
5.0	49,169,663 (59.5%)	146,866,618 (63.8%)	336 (15.9%)
Avg.	4.19	4.23	3.69

Oct 2018 [29]. In the following sections, we call the 82 million product dataset as *Amazon 1 dataset*, and the 230 million product dataset as *Amazon 2 dataset*, 2117 Launchpad product dataset as *Launchpad dataset*, and the 375 Kickstarter product dataset as *Kickstarter dataset*.

4 RO1: comparing launchpad products with traditional products

In the first research objective, we compared rating distributions of *Launchpad* dataset with the *Amazon 1* dataset and *Amazon 2* dataset (i.e., traditional products) at macro and micro levels. As shown in Table 2, the average ratings in *Amazon 1* dataset, *Amazon 2* dataset and *Launchpad* dataset at the macro level were 4.19, 4.23 and 3.69, respectively. 59.5% of products in *Amazon 1 dataset* and 63.8% products in *Amazon 2* dataset received 5 stars, whereas only 15.9% products in *Launchpad* dataset received 5 stars. Note that each product's rating in Tables 2 and 3 was rounded to the nearest whole number.

To statistically analyze whether these datasets have different rating distributions, we performed Chi-squared test for independence. Tests between *Amazon 1* dataset and *Launchpad* dataset outputs X-squared = 2976.4, df = 4, and p value < 0.0001, and between *Amazon 2* dataset and *Launchpad* dataset outputs X-squared = 3706.07, df = 4, and p value < 0.0001. Since, Chi-squared distribution table value for df = 4 is 14.860, the result rejects our null hypothesis: rating is independent or not associated with a dataset. This means that the rating distributions of Amazon and Launchpad datasets are not similar. The result are aligned to our data as ratings are highly distributed between 5 and 4 star in *Amazon 1* and *Amazon 2* datasets, whereas 4 and 3 stars in *Launchpad* dataset.

To analyze average ratings at the micro level, we chose top 5 categories according to product counts in the *Launchpad* dataset. We compared average ratings in each category of *Amazon 1* and *Amazon 2* datasets with *Launchpad* dataset. From Table 3, we observed that product ratings in each category of *Launchpad* dataset were lower than ones in *Amazon 1* and *Amazon 2* datasets by the average of − 9.42% and − 10.5% respectively, with electronics being lowest of all by − 14.96% and − 22.8%. *Based on both macro and micro level analyses, we conclude that there are some gaps between*

Table 3 Average ratings in top 5 categories of Amazon 1, Amazon 2 and Launchpad datasests with their differences (%)

Rating	Amazon 1	Amazon 2	Launchpad	A. 1 vs. L.	A. 2 vs. L.
Electronics	4.01	4.07	3.41	– 14.96%	– 22.8%
Toys and games	4.15	4.23	3.97	– 4.34%	– 6.15%
Home and kitchen	4.19	4.19	3.76	– 10.26%	– 10.26%
Beauty and personal	4.15	4.22	3.77	– 9.16%	– 10.66%
Sports and outdoor	4.18	4.24	3.85	– 7.89%	– 9.2%
AVG rating	4.14	4.19	3.75	– 9.42%	– 10.5%

traditional products and Launchpad products on Amazon regarding quality (i.e., ratings).

5 RO2: analyzing negative reviews of the launchpad products

Previous studies [2,7,41] have shown that product ratings and reviews could be subjective, but they play a vital role in defining the quality of a product and are heavily used in product recommendation as well. It will be beneficial for the Launchpad (i.e., crowd-funded and start-up) product creators to know why their products received relatively lower customer ratings and how to improve their products based on negative reviews. By analyzing associated negative reviews, we are interested in understanding what are the negative responses of customers for the Launchpad products. Our approach is to develop a sentiment analyzer, which automatically extracts negative reviews and then perform topic modelling from the negative reviews to understand what topics the customers often mentioned. Such analysis can help understand products' drawbacks and steps required to improve.

Since a domain-specific sentiment analyzer performs better than a pre-built sentiment analyzer [14], we implemented a sentiment analyzer based on Naïve Bayes algorithm [28] using Amazon data. To create the training set for building the sentiment analyzer, firstly, we sampled 2000 products from each of 24 Amazon categories on the Amazon dataset. Given 48,000 products, we extracted the top positive and top critical reviews from each product. Note that Amazon internally chooses and displays top positive and top negative reviews of a product. A total of 96K reviews were extracted and used to build our sentiment analyzer.

We split our dataset into training (85%) and validation sets (15%) where we have 81,600 reviews for training and 14,400 for the validation. We used stratified splits of the dataset to maintain equal reviews in both classes. There could be a scenario where the classifier encounters a term which has not been seen. To handle such a scenario we used the Laplacian smoothing and gave the term equal probability for both classes. The Laplace smoothing for a $term_i$ in class c_j can be shown as below:

$$p(term_i|c_j) = \frac{num\ of\ term_i\ in\ class\ c_j + k}{(k+1)\star(Total\ number\ of\ terms\ in\ class\ c_j)},\ where\ k = 1$$

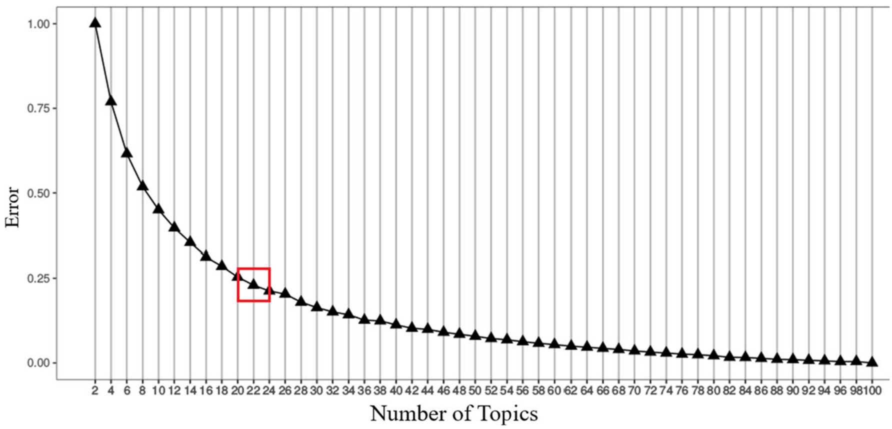

Fig. 3 LDA tuning using elbow method

Next, we performed negation handling by searching for the term 'not' followed by a word. We used a boolean sentence completion variable to store a state of a sentence if it has been negated or not. This state variable became false in the presence of a punctuation mark. As shown by authors [28], this step significantly improved the performance of the sentiment analyzer. We extracted bi-grams and tri-grams from each review as features. Our feature space using bi-grams and tri-grams was 4,107,869. To identify and rank distinguishing features, we performed the Mutual Information and chose top k values for the classification. To identify an optimal value of k, we performed a grid search on a value of k ranging from 100,000 to 3,000,000, and found the optimal value of k to be 2,415,000. In our experiments, we observed the accuracy of sentiment analyzer on the validation set was 81.1% with the optimal value of k.

After building the sentiment analyzer, we applied it to 32,446 reviews associated with the Launchpad products to extract negative reviews. The analyzer identified 3481 reviews as negative reviews. To understand what topics customers mentioned in the negative reviews, we performed topic modeling based on Latent Dirichlet Allocation (LDA) [4]. We observed that the error rate became stable around 22 topics as shown in Fig. 3, therefore we select 22 LDA topics using the elbow technique. Table 4 shows the most representative topics and corresponding reviews.

Customers expressed quality issues and poor customer service and regretted the purchase. Based on the analysis, we conjecture that even though creators of these products initially had exciting ideas and promising prototypes, when they moved to the mass production line, the quality of the products became poor. It could be because of a lack of experience and underestimated cost for production. These creators also provided less professional customer service. Therefore, these Launchpad products received lower ratings than traditional products from customers on Amazon. Figure 4 shows a word cloud, presenting frequent words used in the negative reviews. Customers often used terms like "product", "quality", "return", "money", "waste", and "service".

Next, we further analyzed category-specific terms in the negative reviews by measuring mutual information [23]. This study may reveal how to improve products in

Table 4 Top 5 LDA topics and reviews

Topic	Reviews
Disappointing products	Product arrived dried out or missing product. I'm disappointed...
Poor customer service	What I don't accept is that I emailed the company and have received no response. So now a faulty product and no customer service
Never worked	Piece of junk, never worked first time... Totally ticked off and Son is really upset
stopped working	This broke after 3 months of light usage, it just stopped working
Wasted money	Don't waste your money on this awful product. First Cube came and would not work; Another cube came...and did not work either

Fig. 4 Word cloud of negative reviews

a certain category. In particular, given terms $t_1, t_2, ..t_n$, mutual exclusion can help find terms which separate one class from the others. In this context, classes were 24 Amazon top categories, and terms were extracted from the negative reviews. Table 5 shows mutual information results of the top 4 categories selected by the number of negative reviews. The words in each category represent what people complained. In *Electronics* category, "speakers" was a common issue. When creators make products containing speakers, they should pay more attention to the quality of speakers. In *Cell*

Table 5 Mutual information results of top 4 categories

Electronics	Cell phone	Toys&Games	Home&Kitchen
Sound	Phone	Game	Cheese
Speakers	Case	Robot	Mason
Bluetooth	iphone	Kids	Phone
Camera	Screen	Grandson	Bags
Device	Mount	Play	Jars
Music	Charger	Christmas	Beer
Loud	Phones	Year	Pour
Laptop	Charging	Cards	Mattress
Computer	Cases	Loves	Coffee
Taste	Battery	Playing	Pillow

Phone category, "battery" was a common issue. Creators should pay more attention to batteries when they build cell phone related products.

6 RO3: characteristics of successful and unsuccessful products

So far, we have learned that the Launchpad products received lower ratings than the traditional products and analyzed the negative reviews of the Launchpad products. To protect backers in crowdfunding platforms and buyers in e-commerce sites against low-quality outcomes of products, we turn to study characteristics of successful products ($\geq$ 4 star) and unsuccessful products ($<$ 4 star).

In this section, we address two research questions: (i) Is there any positive correlation between raised money and Amazon ratings in the market?; and (ii) Are there any distinguishing characteristics between successful and unsuccessful products? To answer the research questions, we use *Kickstarter* dataset, which consists of 247 successful products (i.e., received $\geq$ 4 on Amazon), and 128 unsuccessful products (i.e., received $<$ 4 stars). Figure 5 shows the products as dots based on their pledged money (x-axis) and amazon rating (y-axis). We might assume that larger a product's pledged money is, the higher its rating might be. However, it was not the case in our analysis. There was no clear correlation pattern between these two properties. The Pearson Correlation between them was -0.08, showing they are not correlated. *In other words, being successful in raising funds on crowdfunding platforms does not mean that the creators will produce high-quality products and receive high ratings on Amazon.*

To answer the second research question, we computed the mean of various properties of the successful and unsuccessful Kickstarter products on Amazon. Table 6 shows the list of selected properties. We observed that successful products had a lesser number of FAQs than unsuccessful products in their Kickstarter project pages. It may indicate that backers/investors of unsuccessful products posted more concerns regarding the projects. For example, Kickstarter users asked more questions about products or

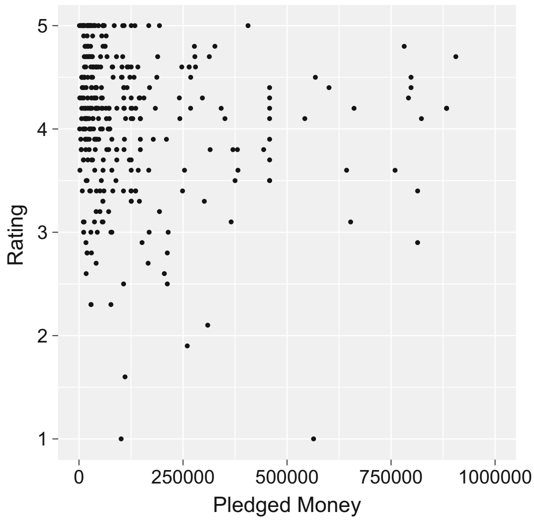

Fig. 5 Pledged money and rating of Kickstarter products whose pledged money $\leq$ \$1M. A dot indicates a product

Table 6 Properties of successful and unsuccessful products

Properties	Unsuccessful	Successful
pledged money	\$528,400	\$313,800
|FAQs|	7.09	4.69
|comments|	934	1075
|images|	27.1	17.5
|negative comments by backers|	633	440
|projects backed by creators|	20.9	26.6
|Facebook friends|	359	773
|lists created by creators|	38	148.2
|posted tweets|	696	1889
|tweets liked by creators|	1397	1734
Product Price on Amazon	\$107	\$83

projects before backing the following projects: jamStik+[4] and Noke.[5] These products received low star ratings and were unsuccessful on Amazon. We also observed that the creators of unsuccessful products backed fewer projects than those with successful products, indicating that the creators of successful products are more experienced and active in the community. In the literature, researchers found that social network plays a vital role in a project's success [15]. We observed the same phenomena in our dataset. Creators of successful products had more Facebook friends and were more active on Twitter. They posted more tweets, created more lists, and liked more tweets. *This means creators who have richer and deeper social networks produce higher quality products.*

[4] jamstik+ The SmartGuitar: http://kck.st/2s0TLQ0.

[5] Noke: The World's Smartest Padlock: http://kck.st/1kU8ztT.

Table 7 Top 5 cities based on the number of backers in successful products and unsuccessful products

Backers city	% successful	Backers city	% unsuccessful
New York	8.64%	New York	3.60%
Los Angeles	8.52%	Singapore	3.58%
London	8.39%	London	2.92%
San Francisco	6.91%	Los Angeles	2.90%
Seattle	5.49%	San Francisco	2.81%

Another interesting factor is pledged money. The creators of unsuccessful products raised 59% more money than ones of successful products. However, the actual products were rated low by customers on Amazon. This means raising more money does not guarantee high-quality outcomes and may result in a complicated production situation. Other researchers also found that crowdfunded projects that received a large amount of pledged money were usually late in delivering the outcomes or products because these projects themselves were more sophisticated [38]. We applied a sentiment analyzer to backers' comments associated with crowdfunding projects. *Unsuccessful products had 69% more negative reviews.*

Next, we analyze the backers of the 375 Kickstarter projects to understand whether there is a different pattern between successful and unsuccessful products.[6] First, we split the backers into two groups: *new backers* and *returning backers*. A *new backer* is new to the Kickstarter platform and has never backed a project before, and a *returning backer* is an experienced backer who has backed at least one project before. An interesting question is whether more returning backers back successful projects than unsuccessful projects. In other words, do they have some sense to knowledge whether they can trust creators of Kickstarter projects and potentially give us a signal that their backing behavior could reveal whether a Kickstarter project will produce a highly rated product or not? We measured the accumulated number of returning backers of successful products and unsuccessful products separately to answer this question. Then, we normalized each accumulated backer count over the total number of backers in each group (i.e., a group of successful products and another group of unsuccessful products). We found that 19.83% of backers were returning backers in unsuccessful products, as compared to 40.35% in successful products. This indicates that the proportion of returning backers in successful products' projects was two times larger than in unsuccessful products' projects. *Successful products' projects were more attractive to the returning or experienced backers who potentially have knowledge whether the project creators would produce high-quality outcomes or not.*

The next interesting question is analyzing the location of backers. We obtained backers' location information from each Kickstarter project's community page, which shows top cities and countries with corresponding backers. Our analysis shows that 60% backers came from the USA and 40% backers from other countries. Table 7 shows the top 5 cities of backers associated with successful products' projects and unsuccessful products' projects. More backers in successful products' projects were

[6] The backer data was obtained from Kickstarter projects associated with 375 successful and unsuccessful Launchpad products.

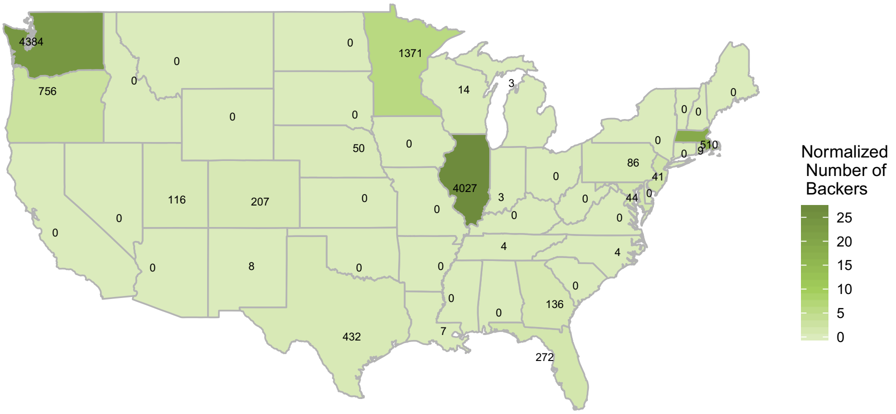

Fig. 6 US heat map associated with number of backers in successful products after removing states each of which has over 10K backers

from US cities than backers in unsuccessful products' projects. At a state level, since most backers, regardless of successful or unsuccessful products, were from major states like California, we removed states with more than 10k backers (to see a clear pattern) and normalized the number of backers by the respective state area. Figures 6 and 7 shows US heat maps associated with the number of backers in the remaining minor states of successful products and unsuccessful products, respectively. We observe that *Washington* and *Illinois* were popular states among backers of successful products, whereas *Oregon*, *New York*, and *Minnesota* were popular states among backers of unsuccessful products.

In summary, we found that successful and unsuccessful products have different characteristics in various properties. In the following section, we describe a list of features designed from the analysis and observation. Then, we build machine learning models to predict whether a product will be successful or not in terms of a star rating (i.e., product quality), and then evaluate their performance.

7 RO4: building predictive models

7.1 Feature engineering

In this section, we describe our proposed features that we use to build predictive models. The features are grouped by four categories: (i) Kickstarter project page features; (ii) Kickstarter creators' profile features; (iii) Kickstarter creators' Twitter profile features; and (iv) Amazon product page features.

Kickstarter project page features These features were extracted from each product's associated Kickstarter page and its community page. They consist of a project goal, pledged money, number of images, number of videos, number of FAQs, number of comments, number of rewards by creators, number of backers in the least rewards, number of backers in maximum rewards, project description length, reward description

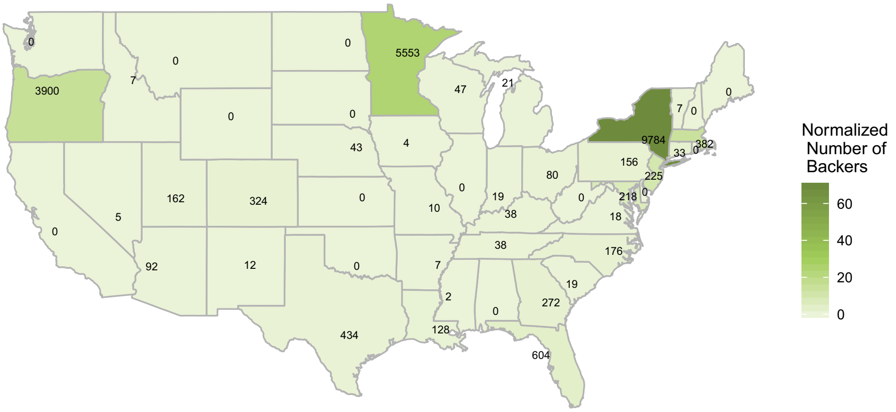

Fig. 7 US heat map associated with number of backers in unsuccessful products after removing states each of which has over 10K backers

length, a percentage of negative comments associated with the project, Coleman Liau readability scores [25] of the project page and reward descriptions, a ratio of pledged money to the goal of the project, number of comments from Superbackers,[7] the number of returning backers, and the number of new backers. The percentage of negative comments was calculated by a sentiment analyzer [28].

Kickstarter creators' profile features These features were extracted from the creators' profile. They consist of a number of backed projects, number of created projects, number of linked external websites, number of creators (e.g., a project may be created by multiple people), is the account verified?, is Facebook connected?, and number of Facebook friends.

Kickstarter creators' Twitter profile features 151 out of 375 Kickstarter project creators linked their Twitter profiles. We extracted their number of tweets, number of followers, number of followees, number of favorites, and number of lists. Missing values were treated by replacing them with the mean of the respective feature.

Amazon product page features Since we know which Kickstarter project is linked with which Amazon product, we further extracted features from an associated Amazon product page. These features consist of a category of the product, number of images, number of videos, product description length, and number of technical details. Besides, we measured a Levenshtein distance/title similarity between a product's Kickstarter title and its Amazon title. We only extracted the product page features available when it was newly created and listed to Amazon. We did not extract any feature from comments and reviews associated with the Amazon product page because we assume that our predictive model (which will be described shortly) will predict whether an Amazon product will be successful or not once its product page is just created.

[7] Superbackers are users who have supported more than 25 projects with pledges of at least $10 in the past year.

7.2 Experiments

Considering real scenarios where different types of information were available at different times, we built predictive models based on the available features in each of the following three stages (refer to Fig. 1):

– *Stage 1*: When a project is just launched on Kickstarter.
– *Stage 2*: At the end of the fundraising period.
– *Stage 3*: When the kickstarter project's respective product page is posted to Amazon.

Our objective of predicting at three stages is to understand how prediction accuracy changes as a crowdfunding project proceeds over time as well as what features affect the prediction. This analysis improves the planning of crowdfunding projects and their products before selling in the market. We extracted features from a Kickstarter page when a project is launched at *Stage 1*, extracted all features available on the Kickstarter page at the end of the fundraising period at *Stage 2*, and extracted features from both the Kickstarter page and a corresponding Amazon page at *Stage 3*. The Amazon page related features that were extracted when the Amazon page is created. We did not extract features related to reviews because the reviews are not available when an Amazon product page is newly launched.

In this experiment, we used the *Kickstarter dataset*, which consisted of 247 successful products (received $\geq$ 4 stars on Amazon) and 128 unsuccessful products (received $<$ 4 stars). To ensure that all of the features have distinguishing power between successful and unsuccessful products, we conducted feature selection by measuring mean decrease impurity from Random Forest. Table 8 shows the top five features at each stage. In particular, in stage 3, the five most important features were # of creators, # of images, product price on Amazon, # of comments from Superbackers, and # of FAQs. *Our analysis showed that successful products were originally initiated by a larger number of creators, less complicated (fewer images, fewer FAQs, and lower prices), and got more comments from Superbackers.* The interpretation makes sense because (1) a larger team usually has more human resources and experience, (2) less complicated projects would have a higher chance of success, and (3) more attention from experienced backers indicates a positive response to the project. Overall, all of our proposed features were important. Table 9 presents features extracted in each stage. We added previously available features to the following stage. For example, stage 3

Table 8 Top 5 features at each stage

Stage 1	Stage 2	Stage 3
# of images	# of creators	# of creators
Project description length	# of images	# of images
Reward description readability	# of creators' comments	Product price on Amazon
# of backed projects	Pledged money&goal ratio	# of Superbackers' comments
Reward description length	# of backed Projects	# of FAQs

Table 9 Features extracted at each stage

Stages	Features
Stage 1	Project goal, pledged money, # of images, # of videos, # of FAQs, # of rewards by creators, project description length, reward description length, Coleman Liau readability scores of the project page and reward descriptions, ratio of pledged money to the goal of the project, # of backed project, # of created project, # of linked external websites, # of creators, is the account verified? Is Facebook connected?, # of Facebook friends, # of tweets, # of followers, # of followees, # of favorites, # of lists
Stage 2	# of comments, # of backers in least rewards, # of backers in maximum rewards, percentage of negative comments, # of Superbacker's comments, # of creators' comments, # of new backers, # of returning backers
Stage 3	# of images on Amazon, # of videos on Amazon, Amazon product description length, Amazon Price, # of technical details, Similarity between a product's Kickstarter title and its Amazon's title based on Levenshtein distance

We added previously available features to the following stage

includes features available in stages 1, 2, and 3 and stage 2 includes features from stage 1 and 2.

To find the best classification algorithm in this domain, we chose four classification algorithms: Random Forest, SVM, AdaBoost, and Gradient Boosting Machines.[8] Then, we performed tenfold stratified cross-validation. To obtain each algorithm's optimal results, we tuned the predictive models using the trial and error method. For example, parameters *gamma* and *cost* in SVM were tuned to 0.1 and 1.0, respectively. Random Forest was tuned with *number of estimators* as 100. In Gradient Boosting Machines, we tuned parameters *interaction depth* as 3, *number of trees* as 100, and *shrinkage* as 0.1. The reported accuracies in Table 10 shows the average accuracy under tenfold stratified cross-validation in each stage. Gradient Boosting and Random Forest performed the best among the four base classifiers, achieving 0.730 and 0.746 accuracy at stage 3, respectively.

Additionally, we performed ensemble and stacking approaches of the best performing base models (i.e., Random Forest and Gradient boosting). The ensemble model uses multiple machine learning algorithms to produce a better performing model. In the ensemble approach, the output of each base model can be counted as a vote. There are two voting mechanisms: *Hard* and *Soft* voting. In *Hard* voting, the class receiving the majority of votes becomes the final output, whereas, in *Soft* voting, the output of probabilities for each class from each base model is averaged, followed by the output of the class with the highest averaged probability. As shown in Fig. 8, we used *Soft* voting since we only have two base models in the ensemble approach. Stacking is

[8] We also tried a neural network model which performed poorly, so we do not report its results.

Table 10 Prediction results at all stages (accuracy)

Algorithm	Stage 1	Stage 2	Stage 3
AdaBoost	0.679	0.679	0.698
SVM	0.698	0.711	0.712
Gradient Boosting Machines (GBM)	0.703	0.725	0.730
Random Forest (RF)	0.701	0.722	0.746
Stacking (base models: RF & GBM)	**0.711**	0.722	0.749
Ensemble (base models: RF & GBM)	0.701	**0.733**	**0.751**

Bold values represent the highest accuracy in respective column

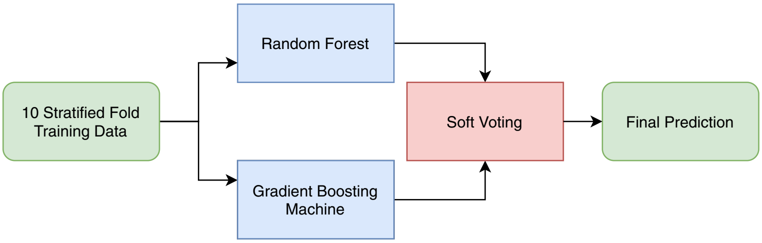

Fig. 8 Ensemble of random forest and gradient boosting

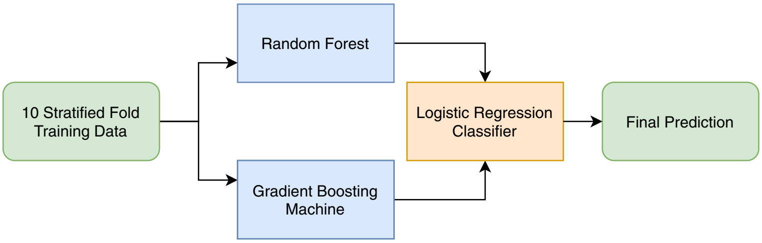

Fig. 9 Stacking of random forest and gradient boosting

another ensemble method, which uses the output of several base models as an input in the second layer of a machine learning algorithm (usually logistic regression classifier) as shown in Fig. 9. Our experiments used a logistic regression classifier at the second layer as the combiner of Random Forest and Gradient Boosting models in the first layer. The logistic regression model tries to learn the base models' optimum weight and produced the final prediction.

Table 10 presents our experimental results of the six models. Our stacked model outperformed the other models at *stage 1*, achieving **0.711** accuracy. Our ensemble model of Random Forest and Gradient Boosting machine outperformed the other modes at *stage 2* and *stage 3*, achieving **0.733** and **0.751** accuracy, respectively. Our ensemble model achieved the highest F1-Macro score at all stages as shown in Tables 11, 12 and 13. The accuracy has increased at the later stages in all models. Compared with the majority selection approach which always predicts a product's class as the majority

Table 11 Precision, recall, accuracy (Acc.), and F1 score of all models at Stage 1

Algorithm	Successful			Unsuccessful			Overall	Macro
	Precision	Recall	F1	Precision	Recall	F1	Acc.	F1
AdaBoost	0.737	0.821	0.770	0.539	0.406	0.435	0.679	0.603
SVM	0.716	0.906	0.798	0.616	0.296	0.381	0.698	0.590
GBM	0.734	0.865	0.792	0.640	0.391	0.467	0.703	0.629
RF	0.739	0.849	0.789	0.582	0.414	0.478	0.701	0.633
Stacking	0.714	**0.942**	**0.811**	**0.719**	0.266	0.369	**0.711**	0.590
Ensemble	**0.742**	0.841	0.786	0.597	**0.430**	**0.486**	0.701	**0.636**

Bold values represent the highest accuracy in respective column

Table 12 Precision, recall, accuracy (Acc.), and F1 score of all models at Stage 2

Algorithm	Successful			Unsuccessful			Overall	Macro
	Precision	Recall	F1	Precision	Recall	F1	Acc.	F1
AdaBoost	0.745	0.793	0.763	0.545	0.460	0.475	0.679	0.619
SVM	0.724	0.914	0.807	0.656	0.319	0.416	0.711	0.611
GBM	0.769	0.845	0.801	0.636	0.491	0.533	0.725	0.667
RF	0.753	0.866	0.804	0.632	0.446	0.516	0.722	0.660
Stacking	0.727	**0.934**	**0.816**	**0.704**	0.313	0.412	0.722	0.614
Ensemble	**0.779**	0.837	0.805	0.628	**0.530**	**0.565**	**0.733**	**0.685**

Bold values represent the highest accuracy in respective column

Table 13 Precision, recall, accuracy (Acc.), and F1 score of all models at Stage 3

Algorithm	Successful			Unsuccessful			Overall	Macro
	Precision	Recall	F1	Precision	Recall	F1	Acc.	F1
AdaBoost	**0.783**	0.768	0.770	0.556	**0.563**	0.546	0.698	0.658
SVM	0.725	0.911	0.806	0.689	0.328	0.427	0.712	0.616
GBM	0.769	0.857	0.806	0.663	0.485	0.537	0.730	0.671
RF	0.757	0.910	0.825	0.726	0.431	0.531	0.746	0.678
Stacking	0.740	**0.963**	**0.835**	**0.869**	0.337	0.461	0.749	0.648
Ensemble	0.781	0.877	0.823	0.690	0.509	**0.570**	**0.751**	**0.697**

Bold values represent the highest accuracy in respective column

instances' class (i.e., *successful* class), our best approach relatively improved accuracy by 14% at the stage 3.

Overall, the experimental results confirmed that our proposed features and framework successfully identified which crowdfunded product will be rated high or low by customers in the future even though we only used limited online data without actual reviews. Using this model, backers could know which project will likely produce high/low-quality products, and e-commerce customers could know which product will receive low or high star ratings even though there are no review available.

8 Conclusion

In this manuscript, we compare Launchpad products with traditional products and find that Launchpad products, on average, receive lower ratings than traditional products on Amazon. We describe a domain specific sentiment analyzer to extract negative reviews of the Launchpad products. We apply topic modeling on extracted negative reviews and our results show that concerning frequent topics are product quality and poor customer service. We analyze the characteristics of successful and unsuccessful products. Based on this analysis, we show how to construct predictive models that automatically predict, with high accuracy, the rating for a product when launched in the market. Our ensemble model outperforms the other models, achieving 0.751 accuracy.

Acknowledgements This work was supported in part by the National Science Foundation under Award No. DBI-1759965, *Collaborative Research: ABI Development: Symbiota2: Enabling greater collaboration and flexibility for mobilizing biodiversity data*, and CNS-1755536, *CAREER*: Tracking, Revealing and Detecting Crowdsourced Manipulation. Opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

References

1. An J, Quercia D, Crowcroft J (2014) Recommending investors for crowdfunding projects. In: WWW
2. Bauman K, Liu B, Tuzhilin A (2017) Aspect based recommendations: rcommending items with the most valuable aspects based on user reviews. In: KDD
3. Belleflamme P, Lambert T, Schwienbacher A (2014) Crowdfunding: tapping the right crowd. J Bus Ventur 29(5):585–609
4. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3(Jan):993–1022
5. Cheng Z, Ding Y, He X, Zhu L, Song X, Kankanhalli M (2018) A^3ncf: an adaptive aspect attention model for rating prediction. In: IJCAI
6. Chung J, Lee K (2015) A long-term study of a crowdfunding platform: predicting project success and fundraising amount. In: HT
7. Cu T, Schneider H, Van Scotter J (2016) New product diffusion: the role of sentiment content. In: SIGMIS-CPR
8. Dey, S., Duff, B., Karahalios, K., Fu, W.T.: The art and science of persuasion: not all crowdfunding campaign videos are the same. In: CSCW (2017)
9. Etter V, Grossglauser M, Thiran P (2013) Launch hard or go home! Predicting the success of Kickstarter campaigns. In: COSN
10. Gerber EM, Hui J (2013) Crowdfunding: motivations and deterrents for participation. ACM Trans Comput Hum Interact 20(6):34:1–34:32
11. Gerber EM, Hui JS, Kuo PY (2012) Crowdfunding: why people are motivated to post and fund projects on crowdfunding platforms. In: CSCW
12. Greenberg MD, Pardo B, Hariharan K, Gerber E (2013) Crowdfunding support tools: predicting success & failure. In: CHI
13. Gupta N, Di Fabbrizio G, Haffner P (2010) Capturing the stars: predicting ratings for service and product reviews. In: NAACL HLT workshop on SS
14. Hamilton WL, Clark K, Leskovec J, Jurafsky D (2016) Inducing domain-specific sentiment lexicons from unlabeled corpora. In: EMNLP

15. Hui JS, Greenberg MD, Gerber EM (2014) Understanding the role of community in crowdfunding work. In: CSCW
16. Joenssen DW, Michaelis A, Müllerleile T (2014) A link to new product preannouncement: success factors in crowdfunding. In: SSRN
17. Kickstarter: Kickstarter fulfillment report. https://www.kickstarter.com/fulfillment (2017)
18. Kuppuswamy V, Bayus BL (2013) Crowdfunding creative ideas: the dynamics of project backers in Kickstarter. In: SSRN
19. Lee S, Lee K, Kim HC (2018) Content-based success prediction of crowdfunding campaigns: a deep learning approach. In: CSCW
20. Li Y, Rakesh V, Reddy CK (2016) Project success prediction in crowdfunding environments. In: WSDM
21. Lin Y, Yin P, Lee WC (2018) Modeling dynamic competition on crowdfunding markets. In: WWW
22. Lu CT, Xie S, Kong X, Yu PS (2014) Inferring the impacts of social media on crowdfunding. In: WSDM
23. Manning CD, Raghavan P, Schütze H et al (2008) Introduction to information retrieval. Cambridge University Press, Cambridge
24. McAuley J, Yang A (2016) Addressing complex and subjective product-related queries with customer reviews. In: WWW
25. McCallum DR, Peterson JL (1982) Computer-based readability indexes. In: ACM conference
26. Mitra T, Gilbert E (2014) The language that gets people to give: phrases that predict success on Kickstarter. In: CSCW
27. Mollick E (2014) The dynamics of crowdfunding: an exploratory study. J Bus Ventur 29(1):1–16
28. Narayanan V, Arora I, Bhatia A (2013) Fast and accurate sentiment classification using an enhanced Naive Bayes model. In: IDEAL
29. Ni J, Li J, McAuley J (2019) Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: EMNLP-IJCNLP. Association for Computational Linguistics, pp 188–197
30. Qu L, Ifrim G, Weikum G (2010) The bag-of-opinions method for review rating prediction from sparse text patterns. In: COLING
31. Rakesh V, Choo J, Reddy CK (2015) Project recommendation using heterogeneous traits in crowdfunding. In: ICWSM
32. Rakesh V, Lee WC, Reddy CK (2016) Probabilistic group recommendation model for crowdfunding domains. In: WSDM
33. Sharma V, Lee K (2018) Predicting highly rated crowdfunded products. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 357–362
34. Solomon J, Ma W, Wash R (2015) Don't wait! How timing affects coordination of crowdfunding donations. In: CSCW
35. Tang D, Qin B, Liu T, Yang Y (2015) User modeling with neural network for review rating prediction. In: IJCAI
36. Tian Z, Guan L, Shi M (2018) The key factors of successful internet crowdfunding projects-an empirical study based on different platforms. In: International conference on service systems and service management (ICSSSM)
37. Tran T, Dontham MR, Chung J, Lee K (2016) How to succeed in crowdfunding: a long-term study in Kickstarter. In: CoRR
38. Tran T, Lee K (2017) Characteristics of on-time and late reward delivery projects. In: ICWSM
39. Tran T, Lee K, Vo N, Choi H (2017) Identifying on-time reward delivery projects with estimating delivery duration on Kickstarter. In: ASONAM
40. Xu A, Yang X, Rao H, Fu WT, Huang SW, Bailey BP (2014) Show me the money! An analysis of project updates during crowdfunding campaigns. In: CHI
41. Zhang Y, Lai G, Zhang M, Zhang Y, Liu Y, Ma S (2014) Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: SIGIR