ENVIRONMENTAL DATA SCIENCE SESSION

ENVIRONMENTAL DATA COLLECTION AND CLASSIFICATION IN CROWD-FUNDING PLATFORMS - EVIDENCE FROM KICKSTARTER

Isidora Ljumović*

Institute of Economic Sciences,
Belgrade, Serbia

Abstract:

Mass use of social networks and the increasing availability of internet technologies creates a series of possibilities for raising funds for entrepreneurs. Crowdfunding is one such option, in which individuals (project creators) can share their ideas with the general public (crowd) via the dedicated Internet platforms, resulting in getting project supporters (backers). The relevance of crowdfunding platforms is thoroughly described in the contemporary academic literature. Furthermore, crowdfunding has the potential to contribute significantly to the financing of environmental ideas and projects, hence accelerating sustainability.

The purpose of this paper is to highlight some of the techniques that can be used in the analysis of data collected from CF platforms, as well as to provide an insight into tests for differences between characteristics of the project with environmental concepts. The study uses a sample of 121,437 projects from the Kickstarter platform between 2011 and 2019. A t-test was employed to determine whether the differences among environmental and non-environmental campaigns. The results show that environmental campaigns are more successful, have a higher goal, attract more funds and investors, while the Kickstarter team favourites them. Analysis showed that quantitative field studies and big data analysis can offer a deeper analysis of the main characteristics of crowdfunding campaigns.

Keywords:

Crowdfunding, Kickstarter, sustainable, bigdata.

INTRODUCTION

A series of techniques (e.g., big data, data mining, deep learning, artificial intelligence, and so on) have become the driving force of enterprise transformation in the mobile internet and social Web era [1], while crowdfunding has become an increasingly important channel for entrepreneurs to raise funds for their start-up projects [2]. Mobile phones, Internet technologies, and social networking sites have all seen rapid growth and use in the last decade.

Many financial advances, including crowdsourcing, have been enabled by technological advancements such as the growth of internet platforms and the rising number of social networks (CF). Although it is a matter of raising a relatively small amount of money, it is possible to acquire significant funds [3], in a short amount of time.

Correspondence:

Isidora Ljumović

e-mail:
isidora.ljumovic@ien.bg.ac.rs

The three basic elements of CF are dedicated Internet platforms where individuals (project creators) can share their ideas with the wider public (crowd) and get project supporters (backers). Initial forms of CF involved donations or individuals who support projects (backers) and received rewards (perks) in exchange for their support. However, more recent modifications include receiving shares of the venture or interest from the loan [4]. Many crowdfunding websites have appeared in the previous ten years. Today's, most popular are Kickstarter, Indiegogo, Patreon, GoFundME, Crowdcube and others. They gather millions of ideas and numerous potential investors. In addition to CF, the range of alternative products scopes from, financing based on future income, online loans, peer to peer loans, cryptocurrencies (Bitcoin, Ethereum, XRP, Tether and others) [5-6].

Having in mind that researchers can acquire a large quantity of data from CF platforms, several methodologies for their analysis emerged. Detailed typology of methodological approaches used in research on CF is best explained in [4].

Almost all CF platforms enable browsing options for all campaigns: up-to-date, achieved and active, successful and failed. Search tools and the growing number of CF users enables a considerable amount of data suitable for the analysis of different project characteristics. Researchers have two options for data collection: to use search algorithms or more recently to employ scraper robots that crawl CF websites and collect data on campaigns. The amount of data collected can be more than a half-million per one CF platform, depending on its size. To illustrate, according to the statistic from the official Kickstarter webpage up to June 2021 only on Kickstarter, 525,851 campaigns are launched. The range of data available for collection, defers a lot, depending on the CF platform and the campaign itself. For example, the parameters related to the campaign can include the goal of the campaign as a desired amount of funds, the percentage of goal fulfilment, the amount of pledged money, the duration of the campaign, the number of investors. Factors related to the platform include special features, such as whether the staff of the platform marks the campaign as a favourite or socially responsible. There are also campaign creator parameters the number of friends on social networks, the number of updates and comments, the number and quality of posted images and videos related to the campaign.

The retrieved data from such platforms researchers can, among others, enable employing tools for big data analysis. Nevertheless, although this is an extremely convenient way to collect data, it requires cleaning, for instance eliminating ongoing campaigns, or campaigns with too low or too high set goals.

This research aimed at pointing out to the techniques that can be applied in the analysis of data collected from CF platforms. As many studies have shown, CF has a lot of potential for contributing to environmental concepts [7] and there are big expectations for crowdfunding to accelerate sustainability [8]. Having this in mind, we classified projects with environmental characteristics and tested for differences among characteristics of the project with environmental concepts.

2. MATERIALS AND METHODS

For the purpose of this study, we used data on Kickstarter campaigns from the "Kaggle.com" open-source repository. The initial full dataset from the repository contained data on 430,938 Kickstarter campaigns in the period 2009-2019. The dataset provides essential information on crowdfunding campaigns (title of the campaign, project goal, funding goal as the amount of money a creator needs to complete the project, blurb as short description displayed under the name of the project and on the browse part of the platform page, pledged funds, as the amount of money the project raised, backers, as the number of people that have supported/invested in the project, state of the project as successful, failed, cancelled, live or suspended, country and city of origin of the campaign creator, currency, category and similar).

We applied filters to the initial dataset, before running statistical tests. Data is typically collected (crawled) in multiple iterations, with certain campaigns appearing more than once, or the program itself can enter the same campaign in the database multiple times. As we downloaded the dataset from the kaggle.com open repository, we do not know how the data was collected. Accordingly, we had to check for the double entries and eliminate them. We have performed full-data matching to confirm that entries are completely identical. If we identified such data, we eliminated them from the sample, hence such could not have contributed additionally to this research. Parallel to this we, excluded projects with a status of unknown outcomes as in [9]. We labelled these projects as cancelled or suspended. Next, we eliminated campaigns from 2011 to 2019, as there was a small number of projects. Finally, in line with [9-12], we excluded projects with too small or too big goals (with a value below $5,000 and over $500,000). This led to a final dataset of 121,437 project campaigns.

We wanted to test whether environmental campaigns have distinguished characteristics from other projects. Therefore, we formulated our main research hypothesis based on the fact that significant difference between the environmental and non-environmental campaigns exist.

In line with this, we had to identify campaigns that have environmental elements. There are few approaches to differentiate environmental projects. The simplest one is to use categories in case the platform enabled this option. Essentially, the CF platform divides projects into categories, with the environment being one of them (as used in [13]). Only a few platforms, however, have integrated environmental features in their classification. Text analysis is the next approach, which aims to find predefined words [10, 14-15]. This approach is simple to use and employable with large datasets. However, it potentially introduces a bias in the classification because it uses a dictionary of keywords arbitrarily defined by the authors [12]. Frequently, individual words can have multiple meanings. The term "green" is an example of a keyword used to classify environmental projects. In-

stead, project creators can use the word green to allude to a characteristic of the campaign, referring to the colour. Certain authors use machine-learning algorithms to define automatically the dictionary of keywords used to discriminate between green and non-green campaigns [12, 16]. In this paper, we used text analysis to define projects with environmental characteristics. This way we got 2,698 campaigns in the related sample, that have environmental elements, approximately 2.2% of the total sample.

In Table 1, we have reported the descriptive statistic of the related sample used in this study. The percentage of successful projects is in line with the general statistic retrieved from Kickstarter. According to the statistic from the official Kickstarter webpage, up to June 2021, 38.82% of campaigns have successfully been financed, whereas in our sample the number of successful campaigns is less than one per cent higher. The average number of backers/investors in the campaign is 213.41. Only 166 projects have attracted more than 10,000 investors, even though the number of backers for one project can reach 87,143.

Characteristic	Total Sample	Min	Max
No. of projects	121,437	/	/
Environmental projects (%)	2,698 (2.2)	/	/
Successful projects (%)	48,186 (39.7)	/	/
Average no of backers mean (median)	213.41 (24.00)	0	87,143
Average funding goals (in 000) mean (median)	27.76 (12,000)	$5,000	$500,000
The average amount of pledged (in 000) mean (median)	19.65 (2,034.00)	$0	$13,285,226
Duration of the campaign mean (median)	34.32 (30.00	1 day	90 days
Staff Pick (%)	17,707 (14.6)	0	1
Spotlight (%)	48,186 (39.7)	0	1

Table 1 – Descriptive statistic of the related sample

The average amount of funds set as a goal ($27,757.83) is much higher than the quantity of money pledged ($19,646.56). Because we limited our sample, the funding goal ranges from $5,000 to $500,000. On the other hand, the amount of pledged funds varies from 0 (for the projects that have not collected any amount of money) up to $13,285,226. However, only three projects in the sample have pledged more than $10 million, and 206 projects more than $million.

The duration of the campaign is limited by the platform and can be from 1 to 90 days. In our sample, the average duration of the campaign is 34.32 days.

Over 14% or 17,707 campaigns were designated by Kickstarter team members as a "favourite" while they were active, measured by the indicator staff pick. Accordingly, 48,186 campaigns are marked as spotlighted allowing creators to make a home for their project on Kickstarter after they have been successfully funded.

3. RESULTS AND DISCUSSION

To test the differences among characteristics of the environmental and non-environmental campaigns, we used a t-test as an inferential statistic used to determine if there is a significant difference between the means of two groups (environmental and non-environmental campaigns), which may be related in certain features. We used the same variable as defined in the descriptive statistic, with a modification related to the values of goal and pledged funds. Due to the high skewness of the distribution data related to these values, we used the logarithm of the goal and pledged funds.

The t-test statistic shows a statistically significant difference among four major variables: the success of the project, average funding goal, average amount of pledged and staff pick.

Environmental campaigns are on average more successful than non-environment (0.414 against 0.396, p-value < 0.01). Campaign success depends on several factors, among which the most important are a well-presented and media-placed idea, the choice of platform on which it is presented, the ways and forms in which funds are raised for invested financial resources. Numerous studies point to the most important determinant of environmental campaign success. Campaign goal, length of the funding period, staff-pick, quality and complexity of the project, number of social contact and friends, comments, updates and similar factors stand out [10-12, 17-22].

In addition, creators of environmental campaign set higher goals (4.261 against 4.187, p-value < 0.01), and attract more pledged money (3.270 compared to 3.188, p-value < 0.01).

Finally, they are more often chosen as a staff pick (0.17 opposite to 0.15, p-value < 0.01), meaning that the Kickstarter team tagged it as a "favourite" while it was active.

Other variables included in the analysis are not statistically significant, but all point to higher values for the environmental campaigns. The average number of backers is slightly higher than in the non-environmental campaigns. We may argue that if the amounts pledged are higher, the number of backers would also be higher. On the other hand, we can also stipulate, that investors prefer campaigns with environmental elements, and thus these campaigns receive more attention and support.

Characteristic	Total Sample	Environmental Sample
Characteristic	Sample	Oriented towards the circular economy
No. of projects	118,739	2,698
Successful projects (%)	0.396*** (0.489)	0.414*** (0.492)
Average no of backers	213.384 (1220.883)	214.397 (833.942)
Average funding goals (log)	4.187*** (0.418)	4.261*** (0.439)
The average amount of pledged (log)	3.188*** (1.319)	3.270*** (1.296)
Duration of the campaign	34.32 (10.915)	34.30 (10.391)
Staff Pick (%)	0.15*** (0.352)	0.17*** (0.379)
Spotlight (%)	0.40 (0.489)	0.41 (0.493)

Table 2 – Comparison between two samples
Standard errors are in parentheses
*** Significance level: 0.01.

Our findings, which are based on the simple t-statistics are consistent with most of the mainstream literature. CF are a tool to foster [23] and accelerate sustainability [8], both for environmental and sustainable entrepreneurship [24]. They might be considered as an example of a business model that can help develop and scale up sustainable innovations by bringing together like-minded individuals, firms, and investors. Also, crowdfunding can be a very useful tool to achieve growth based on sustainability [25], as witnessed in Spain. Projects with environmental characteristics have higher survival rates and suggest the creation of healthy sustainability ventures through crowdfunding [26], especially for a project with general goods as the main component. Clearly, cleantech projects are an excellent example, as they deliver more than a product and accelerate the transition to a low-carbon economy [26].

To sum- up, quantitative field studies and big data analysis can offer a deeper analysis of the main characteristics of crowdfunding campaigns. In parallel, the application of classification techniques allows detailed insight into certain campaign types or phenomena,

such as the environmental component or the study of changes in the behaviour of creators and backers during a corona virus pandemic.

4. CONCLUSION

In this paper, we have investigated two issues. The first issue is related to the overview of different techniques that can be applied in the analysis of data from CF platforms, especially bearing in mind that big data can be collected from CF platforms. In addition, we wanted to present one of the methods to test the data. We have opted to test for differences among characteristics of the project with environmental concepts.

In the analysis, we used open-source data from Kaggle.com. Although previous studies also used the same set as we did, we do not have a firm knowledge that the data set is accurate. Therefore, we had to spend a lot of time cleaning up the data. This deficiency can be overcome by using custom search algorithms or scraper robots which crawls CF platforms and collect data on campaigns.

Using the data from Kickstarter, first, we classified campaigns to environmental and non-environmental using the technique of text analysis. As this method has significant drawbacks, future directions of development should include a more complex and accurate machine-learning algorithm that enables more reliable classification.

Next, we used simple t-statistic to the differences between the chosen groups. In general, we found evidence that environmental campaigns are more successful, have a higher goal and attract more funds and investors, supporting our main research hypothesis. They could be, also, considered more tailored to the Kickstarter team taste. Analysis of the characteristics of environmental projects can point to the direction for improving future campaigns in this area and may help potential crowdfunding users to design successful campaigns

This paper has several limitations. The sample used is restricted to only one CF platform, Kickstarter. New technologies enable the processing of a significantly larger amount of data, so future research directions will strive to include as many similar CF platforms as possible.

5. ACKNOWLEDGEMENTS

REFERENCES

[1] X. Cheng, L. Xu, T. Zhang, Y. Jia, M. Yuan and K. Chao., "A novel big data based telecom operation architecture," in *1st International Conference on Signal and Information Processing, Networking and Computers*, 2016.

[2] J. Shi, K. Yang, W. Xu and M. Wang, "Leveraging deep learning with audio analytics to predict the success of crowdfunding projects," *The Journal of Supercomputing*, pp. 1-21, 2021.

[3] I. Ljumović and P. Biljana, "Financing of Innovation: Are Crowdfunding and Venture Capital Complements or Substitutes?," in *PaKSoM 2020,* Niš, 2020.

[4] A. Stasik and W. Ewa, "How do we study crowdfunding? An overview of methods and introduction to new research agenda," *Journal of Management and Business Administration. Central Europe*, vol. 26, no. 1, pp. 49-78, 2018.

[5] H. Farag and J. Sofia, "How alternative finance informs central themes in corporate finance," *Journal of Corporate Finance*, vol. 67, p. 101879, 2021.

[6] V. K. Manda and Y. Satya, "Peer-to-peer lending using blockchain," *International Journal Of Advance Research And Innovative Ideas In Education*, vol. 6, pp. 61-66, 2019.

[7] J. Hörisch and I. Tenner, "Crowdfunding sustainable entrepreneurship: What are the characteristics of crowdfunding investors?," *Journal of Cleaner Production*, vol. 290, p. 125667, 2021.

[8] A. Böckel, H. Jacob and T. Isabell, "A systematic literature review of crowdfunding and sustainability: highlighting what really matters," *Management Review Quarterly* , pp. 1-21, 2020.

[9] X. Liang, H. Xiaojuan and J. Jiang, "Research on the Effects of Information Description on Crowdfunding Success within a Sustainable Economy—The Perspective of Information Communication," *Sustainability*, vol. 12, no. 2, p. 650, 2020.

[10] D. J. Cumming, L. Gael and S. Armin, "Crowdfunding cleantech," *Energy Economics,* vol. 65, pp. 292-303, 2017.

[11] E. Mollick, "The dynamics of crowdfunding: An exploratory study," *Journal of business venturing*, vol. 29, no. 1, pp. 1-16, 2014.

[12] V. Butticè, C. Massimo G., F. Elena and O. Carlotta, "Green oriented crowdfunding campaigns: Their characteristics and diffusion in different institutional settings," *Technological Forecasting and Social Change*, vol. 141, pp. 85-97, 2019.

[13] J. Hörisch, "Crowdfunding for environmental ventures: an empirical analysis of the influence of environmental orientation on the success of crowdfunding initiatives," *Journal of cleaner production* , vol. 107, pp. 636-645, 2015.

[14] G. Calic and E. Mosakowski, "Kicking off social entrepreneurship: How a sustainability orientation influences crowdfunding success," *Journal of Management Studies* , vol. 53, no. 5, pp. 738-767, 2016.

[15] S. Vismara, "Sustainability in equity crowdfunding," *Technological Forecasting and Social Change*, vol. 141, pp. 98-106, 2019.

[16] M. Siering, J.-A. Koch and A. V. Deokar, "Detecting Fraudulent Behavior on Crowdfunding Platforms: The Role of Linguistic and Content-Based Cues in Static and Dynamic Contexts," *Journal of Management Information Systems*, vol. 33, no. 2, pp. 421-455, 2016.

[17] F. Thies, A. Huber, C. Bock, A. Benlian and S. Kraus, "Following the crowd—does crowdfunding affect venture capitalists' selection of entrepreneurial ventures?," *Journal of Small Business Management*, vol. 57, no. 4, pp. 1378-1398, 2019.

[18] H. F. Chan, N. Moy, S. Markus and T. Benno, "The effects of money saliency and sustainability orientation on reward based crowdfunding success," *Journal of Business Research*, vol. 125, pp. 443-455, 2019.

[19] M. Wessel, F. Thies and A. Benlian, "The emergence and effects of fake social information: Evidence from crowdfunding.," *Decision Support Systems*, vol. 90, pp. 75-85, 2016.

[20] M. Beier and K. Wagner, "Crowdfunding Success: A Perspective from Social Media and E-Commerce.," in *ICIS*, 2015.

[21] G. L. de Larrea, M. Altin and D. Singh, "Determinants of success of restaurant crowdfunding," *Journal of Hospitality Management International*, vol. 78, pp. 150-158, 2019.

[22] I. Ljumović, V. Pavlović and K. Goranka, "Financing agri-food business in the Mediterranean area through crowdfunding: Do environmental issues matter?," *New Medit*, vol. forthcoming, 2021.

[23] J. Block, L. Hornuf and A. Moritz., "Which updates during an equity crowdfunding campaign increase crowd participation?," *Small Business Economics,* vol. 50, no. 1, pp. 3-27, 2018.

[24] N. Thompson, K. Kiefer and J. G. York., "Distinctions not dichotomies: Exploring social, sustainable, and environmental entrepreneurship," in *Social and sustainable entrepreneurship*, Bingley, Emerald Group Publishing Limited, 2011, pp. 201-229.

[25] G. R. Martín, "Análisis del crowdfunding en España: una nueva herramienta social para fomentar la sostenibilidad," *REVESCO: revista de estudios cooperativos* , vol. 135, pp. 61-70, 2020.

[26] N. Bento, G. Gianfrate and M. H. Thoni, "Crowdfunding for sustainability ventures.," *Journal of Cleaner Production*, vol. 237, p. 117751, 2019.