St. Petersburg University
Graduate School of Management
Master in Corporate Finance

SUCCESS FACTORS OF RUSSIAN CROWDFUNDING PROJECTS: EMPIRICAL STUDY OF BOOMSTARTER.RU PLATFORM

Master's Thesis by the 2nd year student
Concentration — Corporate Finance
Anna Petrova

Research advisor:
Associate Professor, Vitaliy L. Okulov

St. Petersburg
2018

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, Петрова Анна Андреевна, студентка второго курса магистратуры направления «Менеджмент», заявляю, что в моей магистерской диссертации на тему «Факторы успеха российских краудфандинговых проектов: эмпирическое исследование платформы Boomstarter.ru», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».

_____ (Подпись студента)

24.05.2018 (Дата)

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF THE MASTER THESIS

I, Petrova Anna, second year master student, program «Management», state that my master thesis on the topic «Success factors of Russian crowdfunding projects: empirical study of Boomstarter.ru platform», which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

I am aware that according to paragraph 9.7.1. of Guidelines for instruction in major curriculum programs of higher and secondary professional education at St.Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Education Saint-Petersburg State University «a student can be expelled from St.Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

_____(Student's signature)

24.05.2018 (Date)

АННОТАЦИЯ

Автор	Петрова Анна Андреевна
Название магистерской диссертации	Факторы успеха российских краудфандинговых проектов: эмпирическое исследование платформы Boomstarter.ru
Факультет	Высшая Школа Менеджмента
Направление подготовки	38.04.02 “Менеджмент” (Профиль: Корпоративные финансы)
Год	2018
Научный руководитель	Окулов Виталий Леонидович
Описание цели, задач и основных результатов	Это первый опыт исследования факторов успеха российских краудфандинговых проектов. Основная цель данной магистерской диссертации - выявить ключевые факторы успеха проектов на платформе Boomstarter.ru и их общую предсказательную способность в определении успеха краудфадинговых проектов. В рамках данной работы были рассмотрены не только финансовые факторы, но и факторы, связанные с характеристиками основателей проектов, с уровнем взаимодействия с общественностью и с качеством оформления проекта. Было обнаружено, что ключевыми факторами успеха являются средняя сумма, известированная спонсорами, и факторы, связнные с уровнем взаимодействия с общественностью. Кроме того, в рамках исследования были построены две предсказательные модели, основанные на алгоритме машинного обучения Extreme Gradient Boosting Trees, с целью выявления общей предсказательной способности всех факторов. Эти модели позволяют основателям проектов оценивать прогресс краудфандинговой кампании через вероятность успеха и, тем самым, получать обратную связь и в начале, и во время процесса сбора денег. Таким образом, проведенное исследование показало не только высокую предсказательную силу собранных факторов в определении успеха проекта, но важность междисциплинарного подхода к изучению, казалось бы, чисто финансовой проблемы привлечения капитала.
Ключевые слова	Краудфандинг, факторы успеха, Boomstarter

ABSTRACT

Master Student's Name	Anna A. Petrova
Master Thesis Title	Success factors of Russian crowdfunding projects: empirical study of Boomstarter.ru platform
Faculty	Graduate School of Management
Main field of study	38.04.02 “Management” (specialization: Master of Corporate Finance)
Year	2017
Academic Advisor's Name	Vitaliy L. Okulov
Description of the goal, tasks, and main results	This is a pioneered paper in studying success factors of Russian crowdfunding projects. The main research goal of this paper is to identify key success factors of projects on Boomstarter.ru and their strength in determining crowdfunding project success. In the scope of this paper, proxy factors of financial, founder-related, social-communication, and description and design groups of factors were considered. It was found that key success factors are an average amount pledged by backers and social-communication factors. Moreover, in the scope of the study, two predictive models were built in order to identify success factors strength altogether based on Machine Learning algorithm Extreme Gradient Boosting Trees. These models enable founders to measure the campaign progress in terms of probability of success and, thus, to get a feedback to project’s founders at the beginning and during the money collection process. To sum up, the conducted study showed not only high predictive power of collected factors in determining project success, but also the importance of an interdisciplinary approach in studying purely financial problem of raising capital.
Keywords	Crowdfunding, success factors, Boomstarter

TABLE OF CONTENTS

Introduction 6
Chapter 1. Crowdfunding as a way to attract capital 9
1.1. Start-up financing sources 9
1.2. Crowdfunding: overview 11
1.2.1 Types of crowdfunding platforms 12
1.2.2 Crowdfunding functions and goals of participants 15
1.3. Crowdfunding: Russian market 17
1.3.1 The appearance of crowdfunding in Russia 18
1.3.2 Largest Russian crowdfunding platforms 19
1.4. Review of empirical researches 26
Summary 29
Chapter 2. Success factors on Boomstarter.ru 30
2.1. Research method 30
2.2. Data 36
2.3. Results 39
2.3.1 Single Factor Analysis 39
2.3.2 Multiple Factor Analysis 41
2.4. Findings and discussion 45
2.5. Managerial implications 47
2.6. Research limitations 49
Summary 50
Conclusions 51
References 53
Appendices 61
Appendix A. Data collection script on Python 61
Appendix B. Variables description statistics 67
Appendix C. Variables correlation matrix 68
Appendix D. Success rate according to factors 69
Appendix E. Success rate according to category 74
Appendix F. Success rate according to city 76
Appendix G. Main program on R 77

INTRODUCTION

As known, the most vital resource for launching any project is financing (Gompers and Lerner, 2004; Gorman and Sahlman, 1989; Kortum and Lerner, 2000; Mollick, 2014). Even a project with a genius idea behind may face difficulties with debt and equity raising (Belleflamme et al, 2014; Berger and Udell, 1995; Cassar, 2004; Cosh et al., 2009). Many startups remain unfunded: sometimes entrepreneurial ventures cannot offer anything substantial in return to investors, sometimes new firms just cannot find a way to persuade investors (Belleflamme et al, 2014; Casamatta and Haritchabalet, 2011; Chen et al., 2009; Hellmann, 2007; Kirsch et al., 2009; Shane and Cable, 2002).

One relatively new way to raise capital is crowdfunding. This method of money collection in itself implies raising many small amounts of money from a large number of people, typically via the Internet. Crowdfunding is used to raise funds for various types of projects and helps project's founders not only to find initial investments without the participation of banks, venture capitalists or stock exchanges, but also to attract first customers which could be interested in the project realization.

In the scope of this master thesis only success of money raising stage of a project is considered. Undoubtedly, not all projects are successfully funded on crowdfunding platforms. According to statistics of the biggest crowdfunding platform Kickstarter, only 36% of projects are successfully funded (collected at least 100% of initial goal amount). Experts claim that crowdfunding project success depends not only on the idea behind, but also on the way this idea is introduces to investors and, certainly, on the level of the crowdfunding campaign advertising (Mollick, 2014). Due to this fact, it is necessary for project's founder to understand what key success factors of achievement a target amount of money on crowdfunding platform are.

Currently, there are no studies based on Russian crowdfunding platforms which may help project's founders to succeed in collecting money process. According to the statistics of one of the biggest Russian crowdfunding platform Boomstarter.ru, average projects' success rate on Boomstarter.ru is only 18%. Due to the fact that the percent of successful crowdfunding projects on one of the biggest Russian crowdfunding platforms is very low relatively to the success rate of the worldwide platform Kickstarter.com, the topic of this master thesis is crucial for Russian founders, which would be better off knowing factors which influence crowdfunding project success on Boomstarter.ru. How to help project's founders? What are main success factors of crowdfunding projects on Boomstarter.ru? In the scope of this master thesis crowdfunding is

considered from the point of view of project's founder, whose only aim is to attract capital for project's realization. In this paper the project is considered successful only if at least 100% of an initial goal amount of money is collected in a predetermined time period.

The main research goal of this paper is to identify key success factors of projects on Boomstarter.ru and their strength in determining crowdfunding project success. As possible factors, not only financial, but also social-communication, founder-related, and description and design factors are considered. In order to achieve the main goal of the study following objectives were set:

1. Analyze literature about crowdfunding projects success factors,
2. Collect data directly from the Boomstarter.ru web-site,
3. Identify key success factors of projects on Boomstarter.ru,
4. Build a predictive model which would get probability of the project success based on factors which can be obtained at the start of money collection process,
5. Build a predictive model which would get probability of the project success based on factors which can be obtained at the end of money collection process.

In order to measure strength of success factors in determining crowdfunding project success, the predictive power measured by Gini coefficient was used. In order to identify success factors strength altogether, two predictive models were build based on Machine Learning algorithm Extreme Gradient Boosting Trees, which is considered as one of the best among classifiers building algorithms (Chen and Guestrin, 2016).

In this master thesis both theoretical and practical contributions are expected to be made. The main theoretical value is that this paper is the first paper which identifies success factors of any Russian crowdfunding projects. Moreover, such research is an interdisciplinary research which takes into account not only financial factors, but also social-communication, founder-related and design and descriptions ones. The main practical value is that two predictive models which are build based on the factors obtained at the beginning and at the end of money collection process help not only to identify factor's strength in determining crowdfunding project success, but also to get a feedback to project's founders at the beginning and during the money collection process on crowdfunding platform Boomstarter.ru.

The research thesis consists of introduction, two main chapters and conclusions. The first chapter starts from an overview of various fund raising methods for start-ups, introduction to

crowdfunding with its definition, types, functions and goals of its participants. This part is followed by an overview of Russian crowdfunding platforms including its appearance in the country. The final part of this chapter is dedicated to a review of empirical researches about success factors of crowdfunding projects. The second chapter includes data description along with its collection process description, research methods used to achieve thesis goal, results, findings and discussion, managerial implications and research limitations.

CHAPTER 1. CROWDFUNDING AS A WAY TO ATTRACT CAPITAL

1.1. Start-up financing sources

The question of where to find money for the implementation of the project in order to move things off the ground is traditionally one of the most difficult stage of most start-ups development. Currently there are various financing sources which may help the start-up project founders who does not have an opportunity to raise capital independently (Andenes and Pendegraft, 2016; Kovačić, 2011; Klačmer Čalopa et al., 2014): 3F (i.e. Friends, Family and Fools), Bank Loans, Venture Capital, Business Angels, Government Financing, and Crowdfunding.

3F (i.e. Friends, Family and Fools)

Before turning to other formal external sources usually project's founders try to raise initial capital from the closest people such as family members and friends. Such initial investments are often extremely risky and due to the fact that the majority of start-up ventures fail within first three years, such investors are often called "Fools" (Klačmer Čalopa et al., 2014). And even though such financing method is thought to have the simplest process, it itself implies not only a risk of possible conflicts that may occur between friends or family members (McKinsey, 2007), but also a limited financial resources. Due to the latter, such kind of raising funds is the most convenient for seed investments – investments for a very early stage of project realization process. Thus, such type of investment is significant for any start-up even before turning to influential investor. Such funds collected from friends and family show that project founder(s) and their closest people are ready to sacrifice their own money and since it – strongly believe in the success.

Bank Loans

This way of raising capital can be considered as one of the oldest formal external financial source. Most banks view start-up projects only as potential future customers and in fact the majority of start-up project founders try to avoid bank loans in the earliest stages of the development since they are usually in itself implies complex procedures and are given based on individual's or firm's credit history and founder's property. Experts claim that banks just cannot take such a high risk of lending money into an early-stage venture without collateral entails (Andenes and Pendegraft, 2016). Usually projects founders are young and therefore do not have grounds to get a bank loans (Klačmer Čalopa et al., 2014). In fact, banks treat different start-ups differently. According to statistics high-tech ventures are rare to use bank loans and have much

more difficulties with this type of fundraising in comparison to start-up projects operating in other industries (Brown et al., 2012).

Venture Capital

Venture Capital investment which is also called risk capital investment is a way to attract capital from individual investors, companies, investment banks, funds or other financial institutions. Such investors are seeking to maximize their investments return and therefore they generally provide an intellectual capital as well as a financial one (Andenes and Pendegraft, 2016). The main difference between Venture Capital and Bank Loan is that by investing venture funds one seek for a part of venture equity and therefore for a partial ownership of the start-up. At the same time while banks make an agreement for a predetermined period of time, have predefined interest rates and therefore burden start-up's cash flows by repayments, venture capital does not create any costs and does not affect a venture's cash flow (Klačmer Čalopa et al., 2014; Rakar, 2006). Most of the time venture capital investors believe in the long-term growth potential of a start-up they finance. Some experts thought that the fact that venture investors do not only give external recommendations for the business, but also take an active role in the company's decisions is a disadvantage of this method of fund raising (Andenes and Pendegraft, 2016).

Business Angels

Business angels which are also called informal investors are investors who help start-up projects founders not only by financing the venture, but also by sharing their business experience, skills, knowledge and contacts. Such angel investors help not only new ventures, but also established companies which may face temporary difficulties. At first sight Business Angels investments and Venture Capital investments are similar, but in fact they differ. Angel investors are usually high-income individuals who invest their own money and focus on helping ventures rather than on possible monetary benefit they can get from the start-up business. Moreover they usually prefer to invest in business industries which they know and understand (Gompers, 2002). Even though business angels invest funds in exchange for an equity ownership or convertible debt, the reasons behind such investments may be nonfinancial (Andenes and Pendegraft, 2016). Internal investors are often willing to invest in the projects which are of a great importance to the business angel personally. Due to such project involvement, business angels may have a significant impact on the project successful development. Therefore, choice of an angel investor may play as crucial role as an actual financing (Gompers, 2002)

Government Financing

Most of the time government programs provide financing through grants and subsidies. Government usually finance projects which cannot be financed by the market. It often provide special projects which goal is to help young talented entrepreneurs (Gompers, 2002). Usually government claims that the reasons are rooted in social-economic benefits. First of all, star-ups create new places for employees. Secondly, the start-ups financing may force whole market for a faster development, since star-ups are also related to innovations, from which not only society, but also other companies may benefit (Andenes and Pendegraft, 2016). Overall, government help young ventures who seek problems with fund raising on the initial stages to fill such funding gaps (Gompers, 2002).

Crowdfunding

Crowdfunding is a relatively new way to attract capital through raising many small amounts of money from a large number of people, typically through special crowdfunding platforms on the Internet. Such small contributions may constitute a significant amount of funds and may help start-up founders not only to raise essential capital, but also to attract first customers to the venture. Its more detailed description is provided further.

1.2. Crowdfunding: overview

Crowdfunding is relatively new scientific area, which is gaining more and more popularity in recent years. Due to this fact, even though there are numerous definitions of the crowdfunding academic concept, there is no one that would be approved by the world community of scientists (e.g. Belleflamme et al., 2014; Bouncken, Komorek, & Kraus, 2015; Tomczak & Brem, 2013; Schwienbacher and Larralde, 2010). The main reason why scientists cannot come to a conclusion is that crowdfunding is a broad concept, any definition of which can be easily supplemented since crowdfunding presences in a wide range of areas. Several definitions considered, one can come to a conclusion that crowdfunding is a way to raise capital from the public, represented by a group of people, usually in exchange for some reward through internet platforms (e.g. Kraus et al. , 2016; Mazzola & Distefano, 2010; Schwienbacher and Larralde, 2010; Ribiere & Tuggle, 2010; Yang, Adamic, & Ackerman, 2008.). At the same time there are those who argue that the definition of crowdfunding should be on the contrary narrow (Mollick, 2014). For example, Mollick, E. (2014) proposes the following one:

"Crowdfunding refers to the efforts by entrepreneurial individuals and groups – cultural, social, and for-profit – to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the internet, without standard financial intermediaries."

There are more than 2,000 crowdfunding platforms which operate all over the world. And even though there are different business models behind platforms, basic system is usually the following. There are three participants: project founder who is seeking to attract capital, project's sponsors (generally called "backers") who pledge money into a project and a crowdfunding platform which bridge them together (Ordanini et al., 2011). Project founder places a description of his or her idea on the crowdfunding platform web-site. Most of the time the textual description is supplemented with photos and videos in order to get backers a better understanding of the project idea. Apart from that, the founder sets up a goal, target amount of money, and usually a predetermined time period for money collection process. In case of project success, the definition of which differs from platform to platform, the founder takes invertors' money and has to keep the word by sending backers what was promised. Usually crowdfunding platforms monetize the business by charging percentage fee from successful projects and by providing additional service to help founders with a project description design.

1.2.1 Types of crowdfunding platforms

Not surprisingly, crowdfunding platforms do not have exactly the same rules and can be classified by fund-raising models they use ("All or Nothing", "Keep What You Raised", "Tipping Point", "Free Donations" and "In Demand") and by different types of reward their projects offer (Reward-Based, Equity-Based, Debt-Based and Donation-Based).

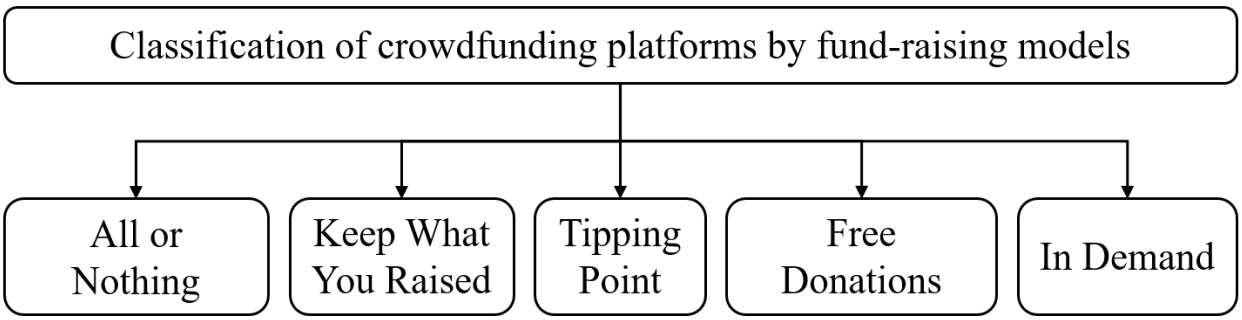

Fig. 1 Classification of crowdfunding platforms by fund-raising model

Source: Kuzmenko, 2016.

“All or Nothing” is a classical scheme crowdfunding, which is used by many popular sites, such as Kickstarter, the most popular worldwide platform, and Boomstarter, one of the biggest Russian crowdfunding platform. The main idea of this model is that collected funds are transferred to the founder only if the goal amount of fund-raising was achieved for a predetermined period of time. If the goal is not achieved, then money is returned to sponsors.

Another fund-raising model is “Keep What You Raised”, where collected funds are transferred to the author of the project, regardless of whether the amount of fund-raising was achieved. This model can be considered as more preferable for founders since they can any amount of money they raised, but as less preferable for backers who want to be sure that to be sure that the founder will realize his or her project.

Tipping Point model can be considered as the main competitor to “All or Nothing” model of crowdfunding money collection. In this model collected funds are transferred the project founder in case an amount of pledged money overcomes the tipping point. Most of the time the turning point is more than 50% of target goal. This model is used by another Russian crowdfunding platform – Planeta.ru. In this model platform’s fee depends on the percentage of the target goal pledged. For example, on Planeta.ru in case the project raised less than 100% money collected but more than 50%, platform allows money transfer to the founder but with higher fee of 15% in contrast with 10% in case of fully collected target goal.

Apart from these models, there are those which do not have any rewards and time limits. “Free Donations” model allows backers to determine the amount of the contribution themselves. Often this fund-raising model works for an already created product. “In Demand” allows founders not to set up any time specific deadlines for money raising so that financing becomes permanent, but can have specifically the necessary amount.

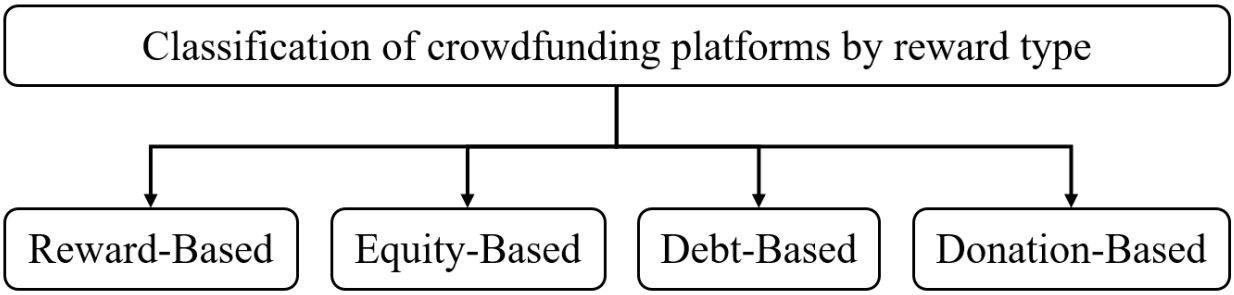

Fig. 2 Classification of crowdfunding platforms by reward type

Source: Forbesa and Schaefera, 2017.

Reward-Based Crowdfunding

On Reward-Based Crowdfunding platforms backers pledge usually a small amount of money in exchange for a reward. So, along with description information a founder sets up "reward levels" – different rewards for the backer in return for different amounts of money that the backer invest in the project. Usually backers sponsor projects in exchange for gifts or products. Most of the time founder combines different kinds of reward such as gratitude, mention on the site or on the packaging the end product, the opportunity to participate in the creation of the product (for example, take part in shooting a movie as an extras), end product or digital copy of the it (for example, film, book, digital music album). The biggest reward-based crowd-funding platform is Kickstarter based in the United States. Two biggest Russian crowdfunding platforms, Boomstarter and Planeta, are both reward-based platforms.

Equity-Based Crowdfunding

Usually Equity-Based Crowdfunding platforms are used to raise capital for growth or launch of a company. On such platforms investors usually pledge a large amount of money in exchange for a small part of firm's equity (Vulkan et al., 2016). This type of crowdfunding is sometimes called crowdinvesting. As a reward on such platforms backers can get usually get shares of the enterprise, along with dividends, or the right to vote at the meetings of the main shareholders (Kuzmenko, 2016). The main players on the Equity Crowdfunding market are AngelList and Crowdfunder based in the United States. Unfortunately, currently this method of financing is banned in Russia due to the lack of legislative and legal framework. Since Equity-Based Crowdfunding platforms are subject to different kinds of regulation (Heminway and Hoffman, 2010), in the world market this kind of crowdfunding platforms are also rare relatively to other types of such platforms. According to research of Massolution (2013) Equity crowdfunding had less than 5% at that moment.

Debt-Based Crowdfunding

Debt-Based Crowdfunding, which is also called Peer-to-Peer Lending or Marketplace Lending, is also a type of crowdfunding where lenders give loans not in exchange of reward or equity, but with the expectation to receive back the principal and interest. This type of crowdfunding sometimes called Debt Crowdfunding since it is similar to bank loan but with one main difference – one borrows not one large amount of money from one bank, but lots of smaller amounts from different people. This enables borrowers to save money and lenders to earn money.

This king of platforms can be used for many reasons such as for example personal needs like credit card refinancing or company's needs like a new project launch. One of the biggest Debt Crowdfunding platform which helps both individuals and small businesses is LendingClub.

Donation-Based Crowdfunding

Donation-Based Crowdfunding, which is also known as charitable giving, where investors, which are called donors in this case, receive nothing in reward except from the information and feedback about the good deeds that the campaign has enabled. Such platforms as Causes, Crowdrise or Just Giving serve as examples of such kind of platform.

1.2.2 Crowdfunding functions and goals of participants

As was discussed above, the main function of crowdfunding is financing, however, similar to other ways of startup financing (Ferrary and Granovetter, 2009), crowdfunding has a number of hidden functions (Kuzmenko, 2016) such as: PR-tool, pre-sale tool and testing the idea tool.

Financing tool

Crowdfunding is considered as a feasible tool to raise funds (Schwienbacher and Larralde, 2010), which supports innovative solutions and start-ups and allows traditional intermediaries to find successful and talented people. Crowdfunding allows founders to get an essential capital not only to finance one-time projects, but also to start a new enterprise (Evans and Leighton, 1989; Mollick, 2014). However, experts claim that crowdfunding cannot fully substitute traditional investors, who usually offer not only essential funds, but also an advice and guidance (Ferrary and Granovetter, 2009; Gompers and Lerner, 2004; Gorman and Sahlman, 1989; Hsu, 2004; Mollick, 2014).

PR-tool

Crowdfunding can be an effective PR-tool for creative and business projects that do not have an audience (Belleflamme et al., 2014; Gerber and Hui, 2013). By placing the project on crowdfunding platform and by its further promotion, the founder can receive a tremendous benefit not only in attracting funds, but also in creating an audience interested in the project beyond the platform. Crowdfunding campaign can also help to get a press attention, which may significantly boost project audience and number potential backers. Apart from that, such marketing can significantly help projects which are going to create an ecosystem by using complimentary products. For example, success of "Pebble and Ouya" videogame console campaign pushed some

independent developers to create applications for future projects even before the official release. This can serve as an example when crowdfunding campaign helped to build a competitive advantage even before the official release.

Pre-sale tool

Most of the time founders offer final products in return for financial support of the audience. The product is usually offered with a special discount or another benefit. This opportunity of pre-selling is used mostly by entrepreneurial ventures who produce some kind of a product. Due to such kind of reward crowdfunding platform can be considered as an internet shop with pre-sales which can serve as a great start for any startup.

Testing the idea tool

Successful or unsuccessful completion of crowdfunding money collection can be an excellent indicator of the relevance of the idea. Due to the fact that crowdfunding platform users support projects that they like or consider as socially useful, an founder can evaluate his or her ideas and get understanding whether such project is needed or not. This allows founders to save lot of money and effort in case of low demand (Agrawal et al., 2014; Belleflamme et al., 2014, 2014; Mollick, 2014). Apart from that, success of the project can serve as a demonstration of the product demand which may help founders to attract more capital from traditional sources. For example, initially, Pebble Smart Watch did not manage to convince investors in the product success, but successful crowdfunding campaign on Kickstarter demonstrated an existing product demand which helped to attract large amount of venture capital funds from private investors (Dingman, 2013; Mollick, 2014).

Overall, a founder goal is not only in attracting money for one-time project or for new startup. Some founders use crowdfunding platforms to test their ideas, some seek to demonstrate demand of the product to huge investors, some want to attract press attention and boost the audience of the projects and some consider such platforms as a pre-sale internet web-site.

But apart from that, crowdfunding creates a transparent and effective mechanism for the peoples' projects financing, which is vital not only for founders, which most of the time cannot find support in banks, but also for backers, who want to support ideas safely with minimal doubts about what money will be spend on. From the point of view of the backer, crowdfunding platform can be considered as the ability to invest directly into the production of interesting product or innovation idea at the initial stage.

Founders and backers relationships differ from platform to platform and from project to project (Belleflamme et al., 2014). These lead to different goals of backer to support projects. Due to the fact that the context in which backers fund projects sometimes overlap, campaigns can allow individuals to achieve several goals simultaneously.

Projects launched on Donation-Based Crowdfunding platforms do not offer any kind of reward for their donors. It means that the main backer's goal in this case is a satisfaction he or she derives from the project realization. Such kind of projects follow a patronage model (Mollick, 2014). In contrast, at first sight, by investing in projects on Debt Crowdfunding platforms lenders may pursue only financial benefit – the rate of return from their microloans. However, many lenders support projects because of the idea of the campaign behind, which means that it can be also considered as a patronage model (Mollick, 2014).

When it comes to Reward-Based Crowdfunding platforms, it is clear that the most probable reason for any backer to sponsor a project is the fact that he or she would like to be a future customer of the startup. Another goal behind is to support a good idea of the campaign even if backer is not going to become a venture customer. In case of Equity-Based Crowdfunding, one can suggest that the main goal of any investor is to get a financial benefit. Investors may tend to support projects not because of the inking to it, but because of belief in the startup success, which will provide a profit on future.

But in fact goals of backers are not so obvious. Some of them may support a project because it was founded by their friends or relatives, some of them consider a project as socially important, some people just want to contribute to feel meaningful and some of backers may even do it as a kind of joke. But even though backers' motivation is vary a lot, all people tend to believe that the project they sponsor will be successful. All crowdfunding backers can be considered as investors who make a decision whether to support a project or not based on their expectation about campaign's success and the personal appeal to the project idea behind (Agrawal et al., 2010). But what is worth mentioning, the funds on crowdfunding platforms are attracted mainly to high quality projects (Burtch et al., 2011).

1.3. Crowdfunding: Russian market

In contrast to crowdfunding abroad, which transformed in the current appearance just before the crisis of 2008, in Russia such crowdfunding platforms appeared with a delay in 2012. At the beginning, many experts were skeptical about the prospect of Russian crowdfunding market development. Majority of them claimed that Russian people have different mentality which would

prevent people to allow the crowdfunding system development as it happened in the West. Some of them thought that various schemes of fraud such as financial pyramids in 90's made people suspicious. And even some of them considered crowdfunding another fraud. Nevertheless, during last 5 years the industry has settled and proved its consistency with growth of more than 200% per year (TJournal, 2017). Since 2012 more than 30 different platforms for financing creative, social, business projects have been launched.

1.3.1 The appearance of crowdfunding in Russia

In fact, analogous of crowdfunding appeared in Russia in the 19th century. At that time it called "Dutch treat" and was used for different needs: from local to the country ones. Whole villages collected money for the general festive tables and for the acquisition of new tools. In November 1808, Alexander II issued the Imperial Decree on the funds collection all over Russia for the construction of the monument to Minin and Pozharsky. At 1878, after the war with Turkey, government collected money for the construction of a voluntary fleet and the acquisition of overseas ships. In the period of the World War II, entire villages collected food for the soldiers at the front.

But even though some analogous of crowdfunding appeared in Russia many years ago, similar to crowdfunding web-sites appeared only in 2008. First platform, Kroogi.ru, idea of which is based on crowdfunding was launched in 2008. It is an internet web-site based on the system "Pay what you want". The portal allows authors to post their music for free to listen and download, while the user can support the author, with any amount and in a way convenient for him. The site also features a full-fledged social network: users could exchange messages, "like" and comment musical composition, join groups, participate in voting. One of the first musicians to post his release on this site was Boris Grebenshchikov, who published the album "Live at The Royal Albert Hall 2008", which was exclusive and posted only on Kroogi.ru. Since 2009, not only musicians joined the project, but also photographers, artists and writers.

According to the data provided by site owners in February 2009, 700 musicians have already joined the project and the audience of the site was 20,000 unique users. In 2009, every sixth site visitor downloaded the album and paid for it. What is also worth mentioning is that the majority of the platform users was Russian whose average contribution was $1. As of January 2010, the maximum one-time payment from an unknown user for music was $1,000. As the creators of the site note, 85% of payments from Russia were realized mainly via SMS or through WebMoney electronic payment system.

Later, in 2010, the Internet portal ThankYou.ru was launched, which distributed content based on the same system “Pay what you want”. Currently, there are two sections presented on the web-site: literature and music. A distinctive feature of the site is a special method of content placement. Any musician gets on the platform directory only if special art board of the platform approve his or her application for the music placement. Only these board decides what content to be place on the web-site. At the beginning, ThankYou.ru focused on the stars - the platform cooperate with such famous people as Zakhar Prilepin, Victor Erofeev, Anna Kozlova, Mikhail Tarkovsky and such music bands as Lyapis Trubetskoy, Noize Mc, Time Machine, Mummy Troll, etc. The lowest payment in the history the site amounted to 0.11 rubles, the largest was 20,000 rubles. Average donation of the platform is 212 rubles.

The first full-fledged crowdfunding platform in Russia which name is "S miru po nitke" was launched in 2010. The most resonant story in the beginning of their work was the collection of funds for the installation in Novosibirsk of the monument to Steve Jobs. Even though the project attracted only 80,000 out of 128,000 rubles, it enabled the platform to gain fame. Nevertheless, on March 26, 2015 the site finished its work. Later on, a founder of the platform published an official statement in which he announced the work on the new project, although there is no information about it since 2015. The founder, Alex Dunaev, notes: “We did this platform in 2010, when there were no crowdfunding in Russia. We originally considered the project as a kind of social experiment. It was important to understand whether this would become a significant economic phenomenon or not. This did not happen. The project played a role of the crowdfunding initiator on Russian market, but at one day we realized that this is not what is needed.”

The summer of 2012 was a landmark for the Russian crowdfunding: with the difference in a couple of months, the two biggest crowdfunding platform were launched in Russia – Planeta.ru and Boomstarter.

1.3.2 Largest Russian crowdfunding platforms

From the very beginning of the origin of crowdfunding in Russia till today there are only two major platforms, which set trend on Russian market and reflects tendencies of the whole crowdfunding market – Planeta.ru and Boomstarter.

Planeta

In 2011 the musical group "Bi-2" who made an attempt to raise funds through crowdfunding, and in six months managed to collect one million rubles. Bass player Max Lakmus

was so inspired of the campaign success, that on June 7, 2012 Planeta.ru was launched. Initially the platform was focused on “people of art” who experienced difficulties in the era of the digital revolution. However, later it was decided to support all good projects, and eventually other projects appeared on the platform. Since the creators of Planet.ru had extensive connections in show business, the platform immediately attracted famous people.

Since the launch, the platform has enabled its authors use the system “Keep What You Raised”. Under such fund-raising model, all funds received during the campaign (except for commissions) are sent to project founder, despite whether the goal was achieved or not. By the way, the second most popular crowdfunding platform in the world, Indiegogo, uses the same fund-raising model.

In addition to crowdfunding, other services were developing on the platform. For example, an online broadcast and an online store for products where one can find books, disks with autographs and products from already completed projects. At 2014, the resource was awarded the Runet Prize in the nomination “Economics, Business and Investments”. In 2015 Planet together with commercial corporations MegaFon, RUSAL and Lipton launched several long-term special projects to support charity and social entrepreneurship. Joint charitable program of the mobile operator MegaFon and Planeta.ru was recognized at the Digital Communication Awards 2015, as the best in the nomination “Digital-project and strategy”.

The program “MegaFon helps” was launched on Planeta.ru in the spring 2015. The purpose of the project was to attract public attention to the social initiatives and effective ways to address them by financial support. The main idea of the project was that MegaFon increased any campaign contribution four times which meant that for any project it was enough to collect 25% of the final goal, the remaining amount was added by MegaFon. Overall, MegaFon supported 58 charity projects, for a total amount of about 180,000,000 rubles. Also in February 2016 Planeta.ru and MegaFon announced the launch of a new season special project “MegaFon helps” with exactly the same rules.

From the very beginning crowdfunding platform Planeta.ru attracted not only popular people to raise funds, but also social and charitable projects. Thus, gaining loyal audience, the web-site actively increased the total number of attracted funds for its projects. During the first full-time 2013 year Planeta.ru launched about 500 projects and attracted about 10 million rubles. The most funded project was founded by Harry Bardeen who collected money on a cartoon “Three Melodies” who found a support from 1,033 backers. The cartoon collected 2 251 681 rubles which

were just 2% above an initial goal amount. During 2014 year Planeta launched more than 1500 projects which was 3 times higher than the amount of projects during 2013 year. Already two years after the platform launch, attracted funds exceeded 166 million rubles. Many projects were completed successfully in 2014 in the sections: charity (25 projects), music (20 projects), society (20 projects), and photography (10 projects). The temporary record by the end of 2014 year was set by a TV show “Petrushka” which collected 5 865 800 rubles from 2477 people, who decided to support the TV show.

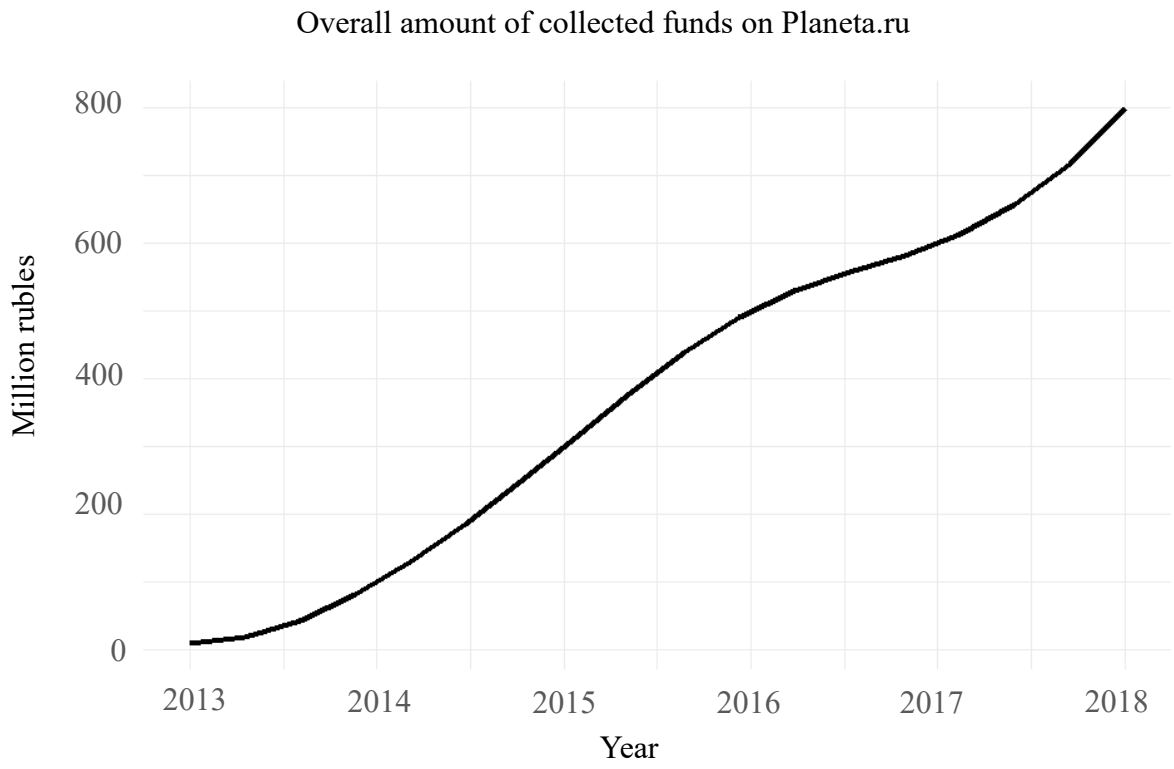

Fig. 3 Overall collected funds on crowdfunding platform Planeta.ru

Source: Made by the author based on the data from Planeta.ru platform

During the next 2015 year Planeta launched about 2000 projects and attracted over 167 million rubles which was 2 times more than the platform attracted since its beginning during the previous 3.5 years of web-site operation. Such success can be explained by a huge number of “star projects” and an increasing PR-activity of the platform. One of such start projects was money collection for a new album of rock bands “Akvarium” which attracted 7 303 803 rubles and by which set a temporary record. What is worth mentioning, during 2015 year there was another projects which collected a significant amount of money. For example, a founder of the first record which was set on the platform in 2014, Harry Bardeen, attracted 6 150 000 rubles for his new cartoon “Slushaya Bethovena”. Another example is a project which were founded by another Russian musician – Boris Grebenshikov. His project collected more than 6 180 000 rubles.

However, his record was soon beaten by a project which raised funds for the film version of the novel Pelevin V. “Empire V” and collected 7 331 006 rubles.

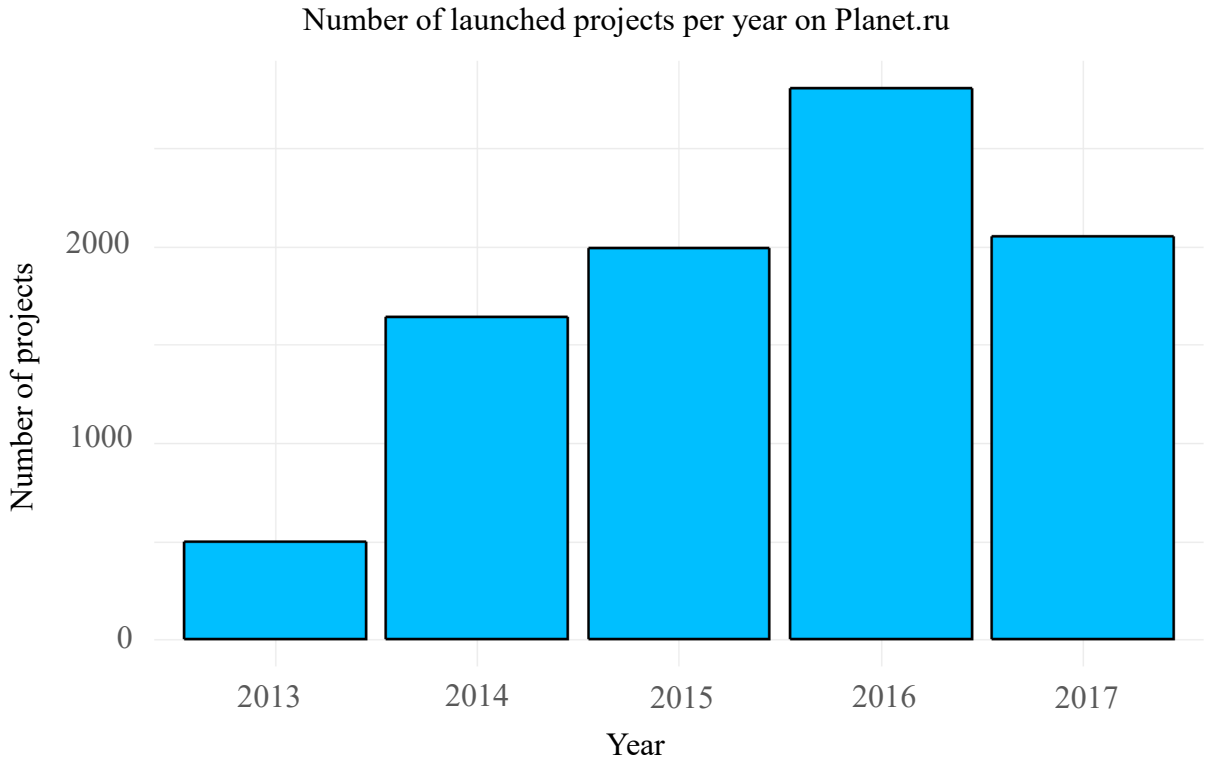

Fig. 4 Number of launched projects per year on crowdfunding platform Planeta.ru

Source: Made by the author based on the data from Planeta.ru platform

By the 2016 year the platform attracted almost 600 million rubles and launched more than 2 800 projects. What is also interesting, the year brought new record which were again founded by another Russian rock band “Alisa”. More than 4500 backers pledged their money and overall collected 11 333 777 rubles. This is still the most funded project on the platform. The most attractive project of 2017 year was again founded by Russian rock band “Nochnie Snayperi”. Even though the project did not break a record, it collected a significant amount of funds – 5 707 577 rubles. What is also worth mentioning, 2017 year brought another type of a record – the platform operated with 780 current projects 28th of April.

By 2018 Planeta.ru attracted over 800 million rubles and launched more than 9 000 projects. As it was already mentioned, the platform uses a fund-raising model “Keep What You Raised”. Under which all funds received during the campaign (except for commissions) are sent to project founder, despite whether the goal was achieved or not. By 2018 year the platform recognized about 3 000 projects to be successful, which by definition on Planeta mean that projects

collected over 50% of the target goal and the founders was ready to realize a project with less than target amount of money.

Boomstarter

Another Russian crowdfunding platform, Boomstarter, is an analogue of the most famous world-wide crowdfunding platform Kickstarter.com. Like the western original, the Russian project followed the same crowdfunding fund-raising model – “All or Nothing”. In this case, a founder receives money only if the project collects at least all the claimed goal amount for a certain period of time. Like a Kickstarter, Boomstarter is a Reward-Based crowdfunding platform, which means that in case of successful campaign, backers receive girts in exchange for money pledged. Such gifts vary significantly from a thank you letter to an acting training. The post popular gift is of course the final product if it exist in the scope of a crowdfunding platform.

Boomstarter was officially launched on August 21, 2012 by two Russian entrepreneurs, Ruslan Tugushev and Eugene Gavrilin. In the same 2012, the web-site won in the nomination “The best socially significant start-up” of the “Startup of the Year” award, established by the business incubator of the Higher School of Economics. The main social mission of Boomstarter is to change Russian audience’s perception about crowdfunding.

Boomstarter is also a pioneer of joint special programs with major brands such as Nokia, Rexona, MTS and even with government. It was this site that showed all advantages of the use of crowdfunding by large companies and government not only for financing purposes, but also as a powerful PR-tool. One of the very first large special programs on the web-site was a cooperation with MTS, as part of their PR-program "wowmoscow". According to the rules of this program, MTS was ready to support the most interesting projects. MTS sponsored them by 30%, provided that the projects’ founders will independently collect 70% of the declared goal amount on Boomstarter. Each project could apply for the program. The main idea was that founders had to add a special sponsorship package of MTS as one of rewards, which intended to brand the founder’s product. At the time when program participating project was funded on 70%, the operator bought out special rewards, and, thus, has financed the project. According to the program statistics there were 9 successfully funded projects out of 32 projects participated in it.

Boomstarter organized two programs with government cooperation: with Public Chamber of the Russian Federation and with Department of Culture of Moscow. The second one was launched in 2014. The project was aimed at creating favorable conditions for the implementation

of public projects, together with residents of Moscow. Under the terms - for each ruble pledged by a backer, the Moscow Department added one more. Within this special project, 14 crowdfunding campaigns were successfully funded and in total attracted 14 989 445 rubles.

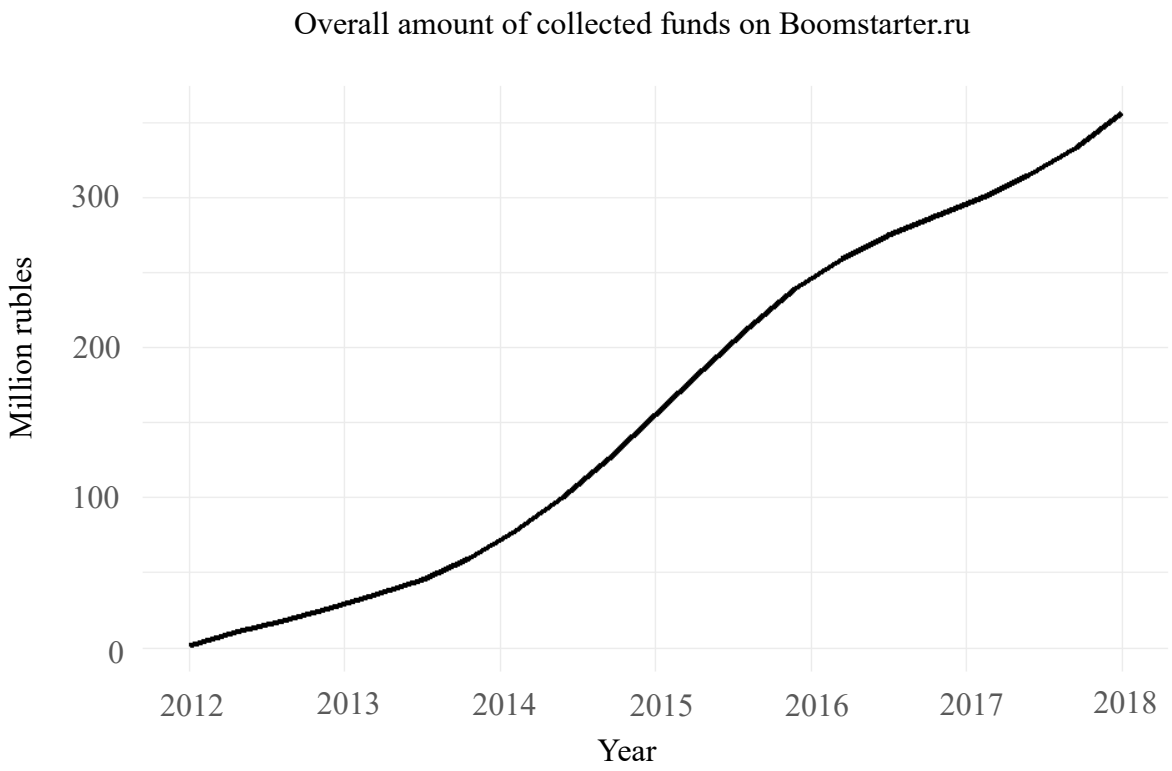

Fig. 5 Overall collected funds on crowdfunding platform Boomstarter.ru
Source: Made by the author based on the data from Boomstarter.ru platform

In 2015 Boomstarter decided to launch the program “Become a shareholder”. According to this program, any user could purchase an option that allowed in the future to buy shares at a fixed price after Initial Public Offering of Boomstarter. The seller of the option was obligated to make a sale of securities in accordance with the terms of the option agreement. According to the option, one option corresponded to one share of Boomstarter, which cost 1 000 rubles and in turn corresponded to 0.00005% of the company equity. Among other things, the shareholder of Boomstarter was promised to have a priority when being placed in the list of sponsors on the each sponsored project page. A total of 100 000 Boomstarter options were put up for sale, which corresponded to 5% of the Boomstarter equity capital. Finally the project have not received a necessary support from the society. Overall, only 3 356 of options were sold and the own project of Boomstarter was considered as unsuccessful. By now, the company is still a private organization.

In the beginning, during half of 2012 year Boomstarter attracted only 1 million rubles from 167 launched projects. During the next 2013 year the platform launched 8 times more projects than in the first year and attracted 21 times more funds, amounted for about 30 million rubles. In 2014 year Boomstarter launched even more projects. Almost 2000 campaigns attracted about 50 million rubles. According to the statistics, in 2015 Boomstarter attracted twice as much as a year before. During that year the platform funded 17.8% out of 2 232 projects and by which attracted over 150 million rubles.

Next 2016 year was a boom period for Boomstarter in terms of the rate of successfully funded projects. About 30% of 860 launched projects received more than the predefined goal amount of money and overall collected 91 million rubles, which is more than was raised in 2014 despite of the rapid decrease in launched project's number. 2017 was not as favorable. Number of launched projects continue to decrease and stated for 571 projects. Number of attracted funds also dropped twice are amounted for about 50 million as in 2015.

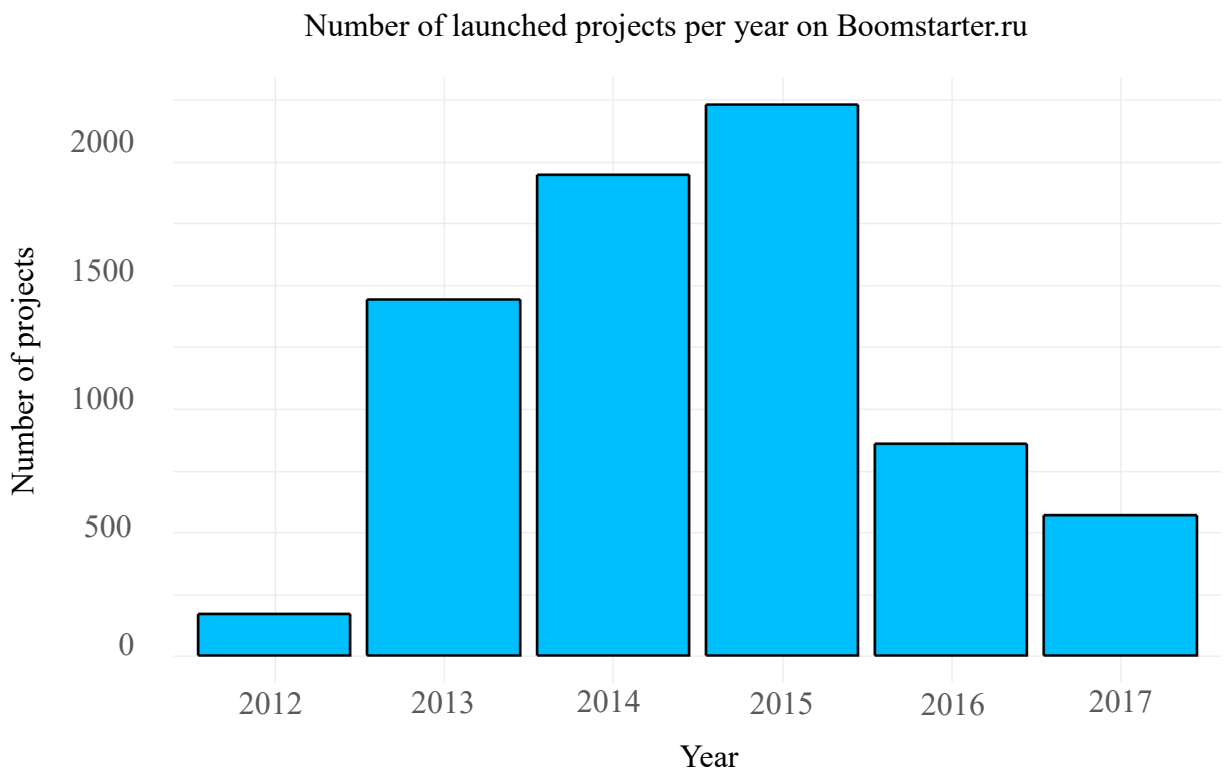

Fig. 6 Number of launched projects per year on crowdfunding platform Boomstarter.ru

Source: Made by the author based on the data from Boomstarter.ru platform

Currently the most funded project is a campaign which collected money for a table game "Serp". Its founder raised 6 949 000 rubles, pledged by 770 backers in the beginning of 2017 year. What is also important, the project's goal was only 100 000 rubles, which means that the project

exceeded the target amount about 7 times. By the way, overall category “Games” attracted 36 622 575 rubles, which represents 10% of total collected amount of money on Boomstarter. The second most funded category is “Films and videos” which also collected about 10% of all attracted funds on Boomstarter. The most popular project in this category is a film “28 panfilovtsev”, which attracted 3 190 995

1.4. Review of empirical researches

Crowdfunding has motivated a growing body of academic literature over the last decade. There is a growing number of research papers which vary from theoretical papers to empirical ones. The majority of empirical studies have examined success factors of crowdfunding projects (e.g. Beier and Wagner, 2015; Kunz et al., 2017; Mollick, 2014; Parhankangas and Renko, 2017; Yuan et al., 2016).

One of the pioneered work belongs to Mollick (2014), who describes fundamental dynamics of crowdfunding project success based on the data of the biggest worldwide crowdfunding platform Kickstarter. The study claims that the project’s quality and the personal social network are related with crowdfunding projects’ success. To be more precise, it was found that goal amount of money and project’s duration have negative influence on the campaign success, while such quality signals as presence of a video, the project update within first three days, the presence of the project on the crowdfunding platform home-page and number of friends in social network Facebook have positive relation with crowdfunding project success. Courntey et al. (2017), Parhankangas and Renko (2017) and Kuppuswamy and Bayus (2015) confirmed these findings. Moreover, Mollick E. has noted that the there is a geographic influence on the crowdfunding projects nature and its success. Agrawal et al. (2011) also examine the geographic origin of backers and found the reduced role for spatial proximity of backers and founders.

One of the most important success factors that should be carefully chosen is project’s goal amount of money. According to many research paper, one of which, Mollick (2014) is already mentioned, campaign’s goal has negative effect on the project success (e.g. Colombo et al., 2015; Gleasure and Feller, 2014; Liao et al., 2015; Zheng et al., 2014). What is also important, it is found that project’s goal has influence on number of backers and the average amount of money pledged by them. To be more precise, it was found that number of backers increase with the target goal amount, while the average amount of pledged money in contrast – decrease (Colombo et al., 2015b).

Apart from project's goal there is another important factor – campaign's duration. What is worth mentioning, several research papers, that analyzed the relation between duration and project success, came to different conclusions. According to studies of Mollick (2014) and Frydrych et al. (2014), there is a negative relation between the factor and campaign's outcome, while according to the study of Liao et al. (2015), the longer the project's duration, the higher the probability of its success. Such a discrepancy in results can be due to geographic differences of crowdfunding platforms analyzed in these research papers: China in Liao et al. (2015), United States in Mollick (2014) and Frydrych et al. (2014). In favor of this argument, Zheng et al. (2014) have found that there is a negative relation between campaign's duration and its success on the United States crowdfunding market, while positive one on Chinese crowdfunding market.

Information about crowdfunding projects also play an important role in the project success. Several empirical research papers investigated that quality and amount of the provided information of the campaign's web-page have positive relation with crowdfunding project success (Burtch et al., 2015; Gleasure and Feller, 2014). For instance, the chances for crowdfunding campaign to be successful are higher when the links of external project's web-pages are attached to the project's description (Buttice et al., 2017; Colombo et al., 2015). Moreover, crowdfunding campaigns which founders frequently update the project have higher probability of positive outcome (Gleasure and Feller, 2014; Koch and Siering, 2015; Kunz et al., 2017; Mollick, 2014; Xiao et al., 2014; Xu et al., 2014). The same is also true for equity based crowdfunding campaigns (Beckwith, 2016; Block et al., 2016).

Another significant success factor which was already mentioned is the presence of a video (Koch and Siering, 2015; Mollick, 2014; Zvilichovsky et al., 2015). Apart from the video, number of images included on the project's web-page description has positive correlation with success (Colombo et al., 2015; Koch and Siering, 2015). According to research paper of Dushnitsky and Marom (2013), effective videos should be short, well-prepared and aimed for potential backers. Another study of Parhankangas and Renko (2017) concluded that social projects, which videos are with very understandable, simple language, are more likely to be successfully funded. What is also worth mentioning, Colombo et al. (2015) found that presence of the already received contributions increases probability of positive outcome.

Moreover, number of comments left by backers on the project's web-page has positive relation with the campaign success (Gleasure and Feller, 2014; Kim et al., 2017; Kunz et al., 2017; Li and Jarvenpaa, 2015; Xiao et al., 2014). However, when number of comments are too much the relation turns to negative one. Apart from that, the length of text description has positive relation

with campaign success (Koch and Siering, 2015). Although, too much information many have an opposite, negative, relation with project success (Xu et al., 2014). Overall, the positive relation between information provided and project success depends on the reduction of the information asymmetry between backers and founders.

Another factor that has positive relation with campaign success is project's category. A significant number of projects are related to technology development (Colombo et al., 2015; Mollick, 2016). The probability of success among such projects depend on the type of innovation behind (Chan and Parhankangas, 2017). Authors found that projects that can be considered to have greater radical innovativeness are less likely to positive outcome in terms of project success, while projects that feature greater incremental innovativeness has positive impact on the project success. The negative effect of radical innovativeness on the campaign outcome can be explained by the fact that such projects seems to be much more risky and are more difficult for backers to understand. Apart from technological crowdfunding projects, many projects are oriented to sustainability and, according to the study of Calic and Mosakowski (2016) are more likely to be successful. Many projects are related with art (Galuszka and Bystrov, 2014), for example with dance, photography, different kinds of movies (Sorensen, 2012), theater (Beaulieu and Sarker, 2013; Boeuf et al. 2014; Josefy et al., 2017) and videogames. Other projects are related to fashion, design (Beaulieu and Sarker, 2013), agriculture (Liao et al., 2015),food and journalism (Jian and Usher, 2014; Jian and Shin, 2015).

Apart from that, empirical findings show that the presence of the campaign on the crowdfunding platform home page has positive influence on the project success (Qiu 2013). Moreover, attached personal social network page (Agrawal et al. 2013), high number of connections in social online networks (Giudici et al. 2018) and number of links to social networks pages related to the project (Mollick 2014; Thies et al. 2014) have positive relation with campaign's outcome.

Apart from above-mentioned crowdfunding projects features, there is another one, which is also play a significant role in money attraction – rewards. According to research studies of Gerber and Hui (2013) and Boeuf et al. (2014), types and number of rewards have an impact on the project success. There are many types of rewards which depend on the project type. For example, Boeuf at al. (2014) divided rewards into two groups: symbolic rewards such as for example thank-you email, list name on a web-site, credits of a move (Buttice and Colombo, 2017) and material rewards such as for instance gifts or final product.

Apart from information about the project, founder-related features are also important when it comes to project success (Zvilichovsky et al., 2013). It was found in the study of Koch and Siering (2015) that founder's experience in crowdfunding as a backer has positive relation with his or her project success, while interestingly the founder's experience in launching projects did not have a significant influence on the campaign success. Moreover, the communication between founders and backers may plan a significant role in the project success. Wnag et al. (2017) found out that, apart from backer's comments quantity and sentiment, the founder's reply length and speed are positively related with campaign's outcome.

Summary

Undoubtedly, the new way of financing – crowdfunding – attracts everyone's attention. Not only the attention of potential founders and backers, but also of researchers. Even though there are several empirical researches which analyze success factors of crowdfunding projects, the success criteria are still not clear.

Crowdfunding on Russian market are gaining popularity among population. In contrast to crowdfunding abroad, in Russia crowdfunding platforms appeared with a delay in 2012. From the very beginning of the origin of crowdfunding in Russia till today there are only two major platforms, which set trend on Russian market and reflects tendencies of the whole crowdfunding market – Planeta.ru and Boomstarter. Both of them are reward-based crowdfunding platforms. In contrast to foreign countries, there are still no research papers that would make an attempt in identifying main success factors of Russian crowdfunding campaigns and therefore help crowdfunding project's founders to successfully attract capital on Russian crowdfunding platforms.

CHAPTER 2. SUCCESS FACTORS ON BOOMSTARTER.RU

The research object is Boomstarter.ru – one of two biggest crowdfunding platforms on Russian market. The platform is classified as Reward-Based, which means that backers pledge usually a small amount of money in exchange for a reward. Boomstarter follows "All or Nothing" fund-raising model, which in turn means that collected funds are transferred to the founder only if the goal amount of fund-raising was achieved for a predetermined period of time. If the goal is not achieved, then money is returned to backers. Thus, in this paper the project is considered successful only if at least 100% of an initial goal amount of money is collected in a predetermined time period.

2.1. Research method

The majority of research papers, which have a goal to identify success factors of crowdfunding projects on different platforms, use Logistic Regression model since it is the easiest model, which can classify observations between two categories (e.g. Koch, Siering 2015, Wang et al. 2018). While in order to identify the strengths of collected factors in predicting crowdfunding project success on Boomstarter more complicated classifying models can be used. For example, in the paper of Greenberg et al. written in 2013 year different machine learning algorithm which solve a binary classifying problems were considered to predict crowdfunding project success based on the data of Kickstarter. Authors tested different algorithms and can to a conclusion that algorithms which are based on decision trees provide the best results, and run the fastest. Taking into account an experience of previously conducted researches, the practical part was divided in two steps: Single Factor Analysis (SFA) and Multiple Factor Analysis (MFA). What is also worth mentioning, in the scope of this master thesis the strength of success factors in determining crowdfunding project success was measured by their predictive power, which in turn was measured by Gini coefficient.

Single Factor Analysis

In order to identify key success factors of projects on Russian crowdfunding platform Boomstarter.ru, the Single Factor Analysis for each variable was conducted. As a result of SFA the predictive power of each variable was identified separately. The predictive power of each variable was measured by Gini coefficient, which shows the overall quality of any scoring function, i.e. an output of a predictive model which gives an estimate of a binary dependent variable. The binary scoring function was calculated from the Single Factor Logistic Regression, which model for any $Factor_i$ is the following:

$$Success\ state = \frac{1}{1 + \exp\left(-(\alpha + \beta \times Factor_i)\right)}$$

Gini calculation is based on Receiver Operating Characteristic (ROC) curve. ROC curve can be successfully used to show the discriminatory power of the any binary scoring function, i.e. its ability to identify successful and unsuccessful projects in our case. Each point on the curve represents some value of a given score. An x-axes correspond to False Positive rate, which is the rate of wrongly identified unsuccessful project, and a y-axes – to a True Positive rate, which is the rate of correctly identified successful projects. If one assume this value to be the cut-off value, one can read the proportion of rejected unsuccessful and successful projects. An example of a ROC curve is given in Figure 7. It can be seen from the graph that 20% of wrongly identified unsuccessful projects lead to 60% of correctly identified successful projects.

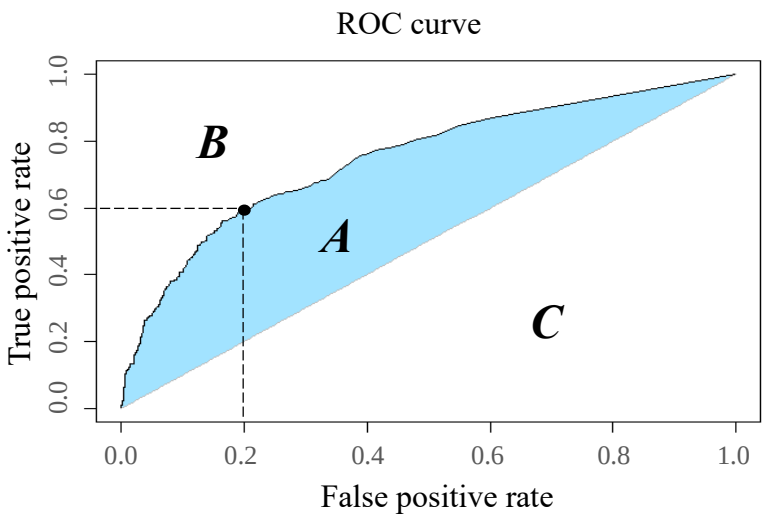

Fig. 7 ROC curve example

Source: Made by the author

Gini coefficient takes values between -1 and 1. The ideal model, i.e. scoring function that perfectly separates successful and unsuccessful projects, has a Gini coefficient equal to 1. On the other hand, a model that represents a random score has a Gini coefficient equal to 0. Negative values correspond to a model with reversed meanings of scores. Overall, Gini measures the classifier's advantage over a purely random one. Using Figure 7 the Gini coefficient can be defined as in the following formula:

$$Gini = \frac{A}{A + B} = 2A$$

Gini coefficient is closely related to another predictive power measure – Area Under the Curve (AUC) coefficient. AUC is an area under ROC curve and can be interpreted as the probability that a classifier will rank a randomly chosen successful project higher than a randomly chosen unsuccessful one. Thus AUC takes values from 0, which corresponds to models with reversed meanings of scores, to 1, which in turn corresponds to a perfectly separating binary classifier. Random predictive will show AUC equal to 0.5. AUC is calculated by the following equation according to the ROC curve showed on Figure 7:

$$AUC = A + C$$

Gini can be considered as a reformulation of AUC with more intuitive scale. Gini is a linear extension of an AUC coefficient, which positive values correspond to model which perform better than a random classifier and in turn negative values – to classifiers which are worse than a random one. Formula of the linear relationship between Gini and AUC is the following:

$$Gini = 2 \times AUC - 1$$

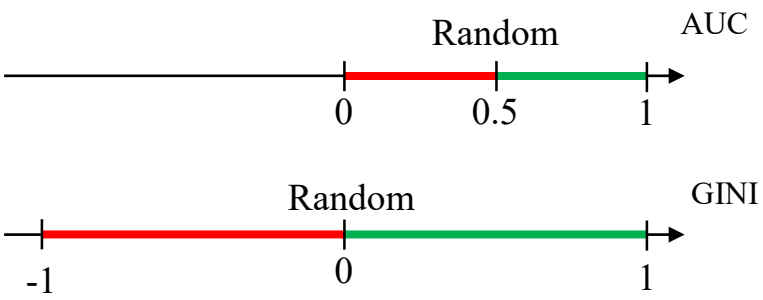

Fig. 8 Scale of AUC and Gini values

Source: Made by the author

Gini coefficient was chosen for several reasons. First are foremost is that it shows the predictive power without depending on one particular threshold. Secondly, it is wildly used in practice of banking clients score modeling, which idea is very similar to that in this paper – built to predict probability that a particular client will be able to repay the debt (Rezac et al., 2011).

Multiple Factor Analysis

In order to identify strengths of collected factors in predicting crowdfunding project success on Boomstarter, a machine learning algorithm Extreme Gradient Boosting Trees (XGBoost) was used. This method was chosen since it is currently one of the best binary classifying algorithm (Nielsen, 2016). The majority of programmers, who wins machine learning competitions, base their solutions on this algorithm (Chen and Guestrin, 2016).

Boosting was firstly formulated in such research papers as Freund (1995), Freund and Schapire (1996, 1997) and Schapire (1990). The main idea of boosting process is that a classifying model with predicting power slightly higher that a random one can be significantly improved by iteratively adding another classifying models that compensate for the previously occurred error. Later, in 1997, gradient boosting algorithm was firstly introduced by Breiman, who specified that boosting is an algorithm which uses an idea of gradient descent. In other words gradient boosting algorithm provides a linear combination of weak classifying models which serve as base learners, when on each iteration a new model is trained so that it searches for the negative gradient direction in order to minimize the error. When it comes to Extreme Gradient Boosting Trees algorithm, its main idea is to reinforce the predictive power of multiple decision trees models which serve as weak classifying algorithms, which are slightly better that random ones. The idea of gradient boosting application to decision trees was firstly provided by Friedman J. in 2001.

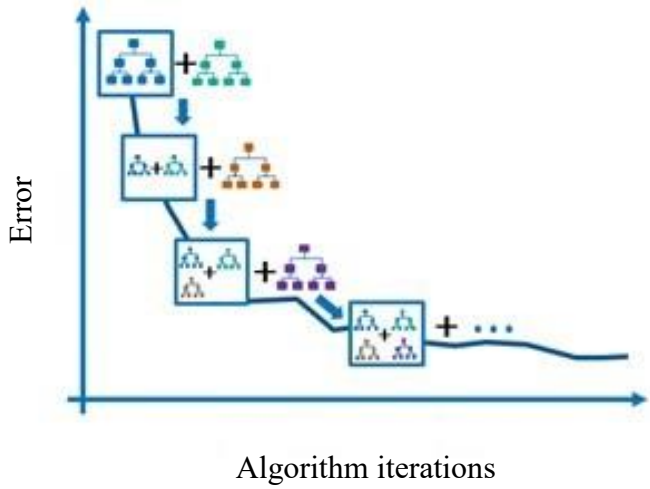

Fig. 9 Extreme Gradient Boosting Trees algorithm illustration

Source: Data science blog www.datascienceblog.pw

Decision tree is a non-parametric supervised machine learning method, which is widely used in different of machine learning algorithms for classification and regression. The goal is to create a model that predicts the value of a target dependent variable by learning a simple decision rules inferred from the data features. Decision tree is a top-down tree, which includes two types of nodes: decision nodes and end nodes. Each tree typically starts with a single decision node, which branches into another two nodes which can be both, decision and end one. This gives it a treelike shape. Each decision node separates observations between two branches by a logical rule which can be true or false for each observation. Therefore, each observation can be run through a decision tree and as a result correspond to one particular end node. Each end node, which is also called leafs, in turn gives a score which is also called leaf weights and can be either binary or not,

depending on an algorithm which uses this decision tree. In case of continuous leaf weights a decision tree is called a regression tree. In case of Extreme Gradient Boosting Trees algorithm the regression tree is used.

In order to illustrate a regression tree model, let's consider an example of a tree which goal it to identify whether a person likes computer games or not based on a following data set $D = \{(x_i, y_i)\}$ with 4 examples (people in the scope of the example) and 2 features: age (numerical) and gender (Boolean: 1 – man, 0 – woman). The data can be seen on Fig. 10, where x_i= *(age, gender)* $y_i = \begin{cases} 1, does\ like\ computer\ games \\ 0, NOT\ like\ computer\ games \end{cases}$

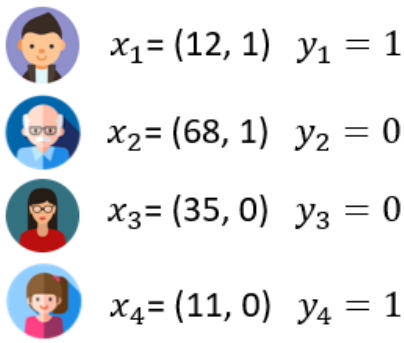

Fig. 10 Data set for a given example

Source: Made by the author

One of possible regression trees is shown on Figure 11. For a man which age is 68 the algorithm will give a score of -1, which will correspond to a negative impact on his likelihood to like computer games, while a boy of 12 years old will get a score of 2 which will correspond to a positive impact on his likelihood to like computer games.

Now let's formulate an algorithm for a given data set with n examples (projects in the scope of current goal) and m features: $D = \{(x_i, y_i)\}$, where $|D| = n$, $x_i \in R^m$ – independent variables values of i_{th} project, $y_i \in R$ – the target success state of i_{th} project (Boolean: 1 – successful project, 0 – unsuccessful one). When it comes to a formal definition of a tree used in this algorithm, let's begin with a space of regression trees, which is also known as CART. It is defined as $F = \{f(x) = w_{q(x)}\}$, where $q: R^m \to T$ represents the structure of each decision tree that maps an example to the corresponding leaf index $i \in 1..T$ and $w \in R^T$ are leaf weights, where w_i correspond to a score on a particular leaf i. In its turn f_k corresponds to an independent decision tree structure q and leaf weights w. As it was already mentioned, in the scope of Extreme Gradient Boosting Trees algorithm each leaf has a continuous score. Let's illustrate a given definition on

the example of regression tree which can be seen on Figure 11. For a given example, number of leaves $T = 3$, leaf weights are $w_1 = 2$, $w_2 = 0.5$ and $w_3 = -1$, structure of a given tree is so that $q(x_1) = 1$, $q(x_2) = 3$, $q(x_3) = 3$, $q(x_4) = 2$ and therefore $f(x_1) = w_1$, $f(x_2) = w_3$, $f(x_3) = w_3$, $f(x_4) = w_2$.

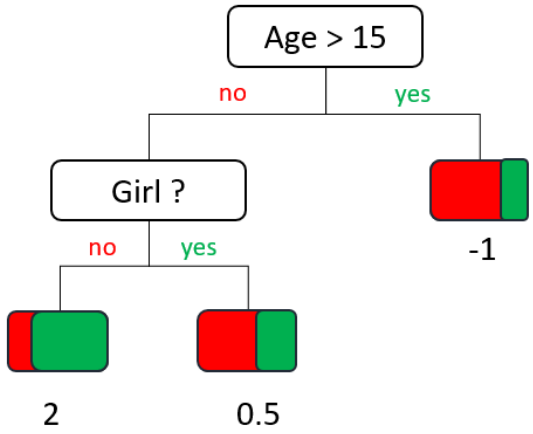

Fig. 11 Example of a decision tree

Source: Made by the author

The algorithm consists of a predefined number of steps – N. On each step k the prediction of a target success state vector $\hat{y}^{(k)} = (\hat{y}_1^{(k)}, \dots, \hat{y}_n^{(k)})$ is calculated. In order to do it, on each step Extreme Gradient Boosting Tree algorithm builds a new tree t_k, which takes into account errors accrued on the last step. To get f_k, function which gives value in corresponded to i_{th} project leaf in tree t_k, the following regularized objective is minimized for each decision node split until all nodes become end ones:

$$\min_{f_k} L^{(k)} = \min_{f_k} \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)\right) + \Omega\,(f_k),$$
$$where\ \Omega\,(f_k) = \gamma T + 1/2\,\lambda\,||w||^2$$

Here l is a loss function that measures the difference between the target success state y_i and its prediction $\hat{y}_i$. The loss function used in Extreme Gradient Boosting Trees algorithm is an approximation of Taylor series expansion of a function $\log(1 + \exp(-y_i\hat{y}_i))$ to second-order derivatives:

$$l\left(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)\right) = \log(1 + \exp(-y_i\hat{y}_i)) + (\hat{y}_i - y_i)f_k(x_i) + 1/2\hat{y}_i(1 - \hat{y}_i)f_k^2(x_i)$$

The second term Ω is a penalty function which controls the complexity of the model which correspond to regression trees. To be more precise, the first term penalties for the tree size which is represented by number of leafs, while the second term helps to smooth the final learnt weights

to avoid over-fitting – when the predictive algorithm perfectly work on a train sample, but does not provide good results on other test data sets.

After the tree t_k is built, the final prediction $\hat{y}_i^{(k)}$ on k_{th} iteration for each i_{th} project is calculated by summing up the score in the corresponding leaves of all projects by the following formula:

$$\hat{y}_i^{(k)} = \sum_{j=1}^{k} f_j(x_i) \ = \hat{y}_i^{(k-1)} + f_k(x_i)$$

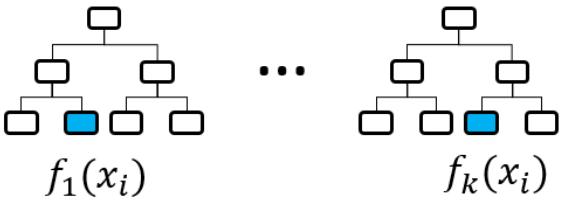

Figure 12. Illustration of i_{th} project classification into scores by decision trees on step k
Source: Made by the author

On the last step N after the final prediction $\hat{y}_i^{(N)}$ on N_{th} iteration for each i_{th} project is calculated, the final probability of success is derived by the following formula:

$$p(x_i) = \frac{1}{1 + \exp(-\hat{y}_i^N)}$$

In order to build two predictive models in the scope of this master thesis an already realized Extreme Gradient Boosting Trees algorithm in the programming language R was used. What is also worth mentioning, in order to avoid over-fitting, not only initial sample was separated into train and test datasets with 3:1 ratio, but also a cross-validation was used. It in itself implies that on each step algorithm randomly takes a predefined percentage of a train sample (70% in this case) and due to it does not allow a model to be over-fitted on a train sample.

2.2. Data

If one think what can be real success factors of crowdfunding project success, first intuitive factors which are likely to come to a possible backer's mind are (1) the attractiveness of the idea behind, (2) the quality of the crowdfunding project which includes clear idea presentation and well thought-out project and (3) a sufficient advertising level. Unfortunately, it is impossible to measure such intuitive success factors directly and due to this fact only proxy factors could be measured.

In order to identify success factors of crowdfunding projects and their strength in determining crowdfunding project success, the cross-sectional data about projects from Russian crowdfunding platform Boomstarter was collected. Since there is no open access data of Russian crowdfunding platform Boomstarter.ru, the data was collected directly from the web-site by a special script written by me in the programming language Python. The code can be found in Appendix A.

Data of all ever launched projects was collected on 9th of February 2018. Initial sample consists of 7600 projects, including 1294 successfully funded. Data was cleaned from all current projects. Final sample included observation of 7303 projects, including the same number of successful campaigns.

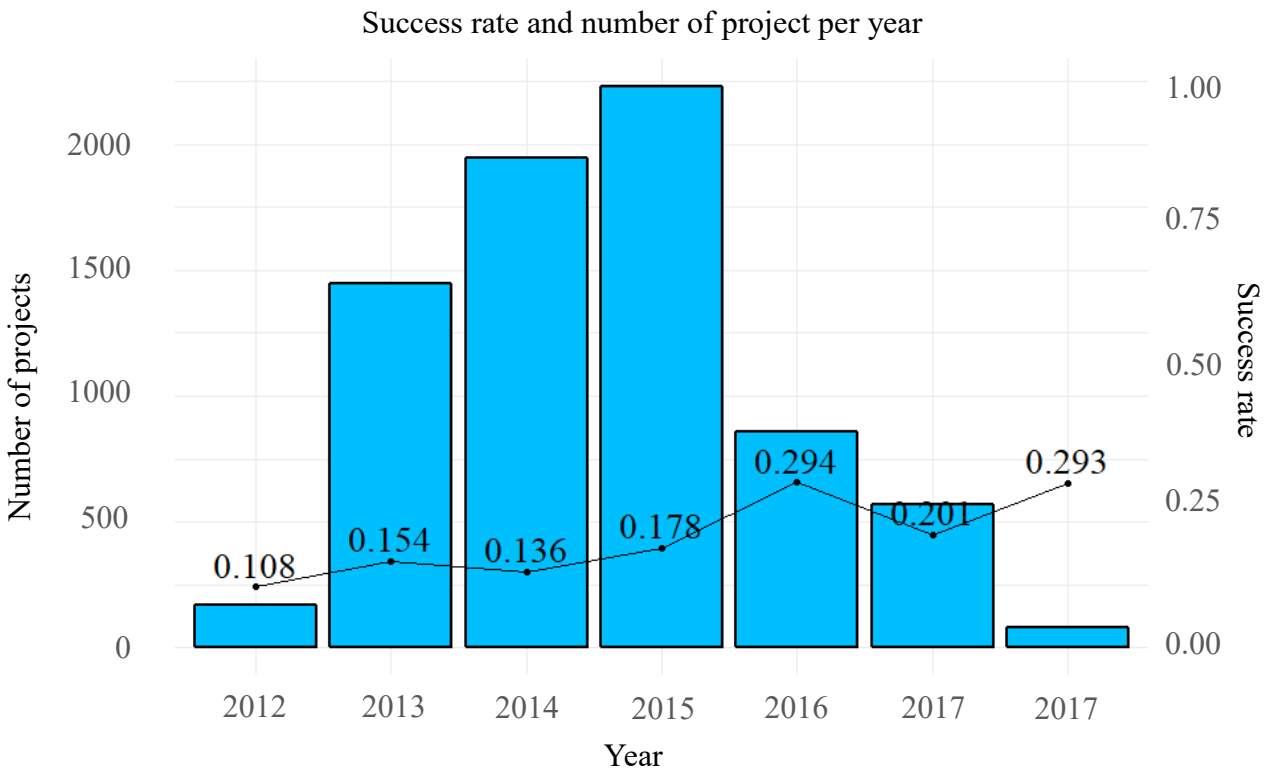

Fig. 13 Success rate according to each year

Source: Made by the author based on the data from Boomstarter.ru platform

Collected data can be separated in several categories: Financial factors, Founder-related factors, Social Communication factors and Description & Design factors. Financial data collected includes information about project success, goal amount of money raising, final amount of money pledged, final progress rate of money collection process, final number of backers, and information about reward levels (number of levels and rewards prices). Social Communication data includes

number of posted news on the project campaign page, number of comments written by backers, and number of social reposts in social networks Vkontakte and Facebook. Founder-related data consists of founder name, number of projects previously launched on the platform by founder, number of backed project by founder, number of friends on social networks Vkontakte and Facebook and number of attached web-sites about the founder and the project. Description data includes category of each project, founder location, text description length in number of abstracts and symbols. Design data includes information of pictures and videos presented in the project description.

From the collected data 17 variables were formed. Only one of these variables represents the final result of the project success and is included in the model as dependent variable – Success.state, which is equal to 1 if the project was successfully funded and to 0 otherwise. Description of all variables according to four abovementioned categories are presented in the Tables 1 – 4.

Table 1 Financial Factors: variables description

Variable	**Description**	**Type**
Success.state	Final status of the project: 1 if project raised at least 100% of the pledging goal, 0 otherwise	Boolean
Goal	The amount of money founders are willing to raise	Numerical
Pledged.by.backer	Average amount of money pledged by backer	Numerical
Rewards	Number of backer levels	Numerical

Table 2 Founder-related Factors: variables description

Variable	**Description**	**Type**
Founder.as.backer	Number of projects sponsored by the founder	Numerical
Founder.projects	Number of projects launched by the founder	Numerical
Founder.friends	Maximum number of friends founder has on Facebook and Vkontakte	Numerical
Founder.sites	Number of web-sites links attached by founder	Numerical
Founder.text.length	Number of symbols in the founder’s description	Numerical

Table 3 Social Communication Factors: variables description

Variable	**Description**	**Type**
News	Number of project's updates by founders	Numerical
Comments	Number of comments by backers	Numerical
Reposts	Sum of reposts in social networks Vkontakte and Facebook	Numerical

Table 4 Design & Description Factors: variables description

Variable	**Description**	**Type**
Category.Rank	Percentage of the number of projects in each category which was successfully funded	Numerical
Has.video	Availability of title video: 1 – there is video, 0 – not	Boolean
Images	Number of pictures provided in the project's description	Numerical
Text.length	Number of symbols in the project's description	Numerical
Text.abstracts	Number of abstracts in the project's description	Numerical

2.3. Results

2.3.1 Single Factor Analysis

In order to identify success factors of crowdfunding projects, the relationship between variables and success was analyzed separately. First of all, the graphical analysis was conducted. A range of each variable values was cut into several buckets. After, the percentage of successfully finished crowdfunding projects in each category was calculated. Graphical analysis results are useful for better understanding of the variables nature and demonstrate the right signs of estimates obtained by the logistic regression. Graphical representation of relationship between each considered variable and crowdfunding project success on Boomstarter can be found in Appendix C, while the graphical analysis result for "Goal" variable can be also seen on Fig. 14.

According to this graphical analysis, all variables except from "Goal" have positive relationship with crowdfunding project success rate on Boomstarter. "Goal" is the only of 16 independent variables, which is negatively related with crowdfunding project success state. As seen from Figure 14, the higher an initial project goal, the lower the success rate in the corresponding basket. At the same time, even though the majority of variables have positive relation with success rate, no for all of them the increase in value will definitely lead to higher

success rate of a corresponding value basket. Variable “Founder.text.length”, which represents length of the founder description text, shows that despite the positive relationship with project success, the success rate of a basket, which contains of projects with the highest description length, has a success rate lower than it of the previous basket with projects with lower founder description length. It means that length of founder description has positive relation with success state until a certain level.

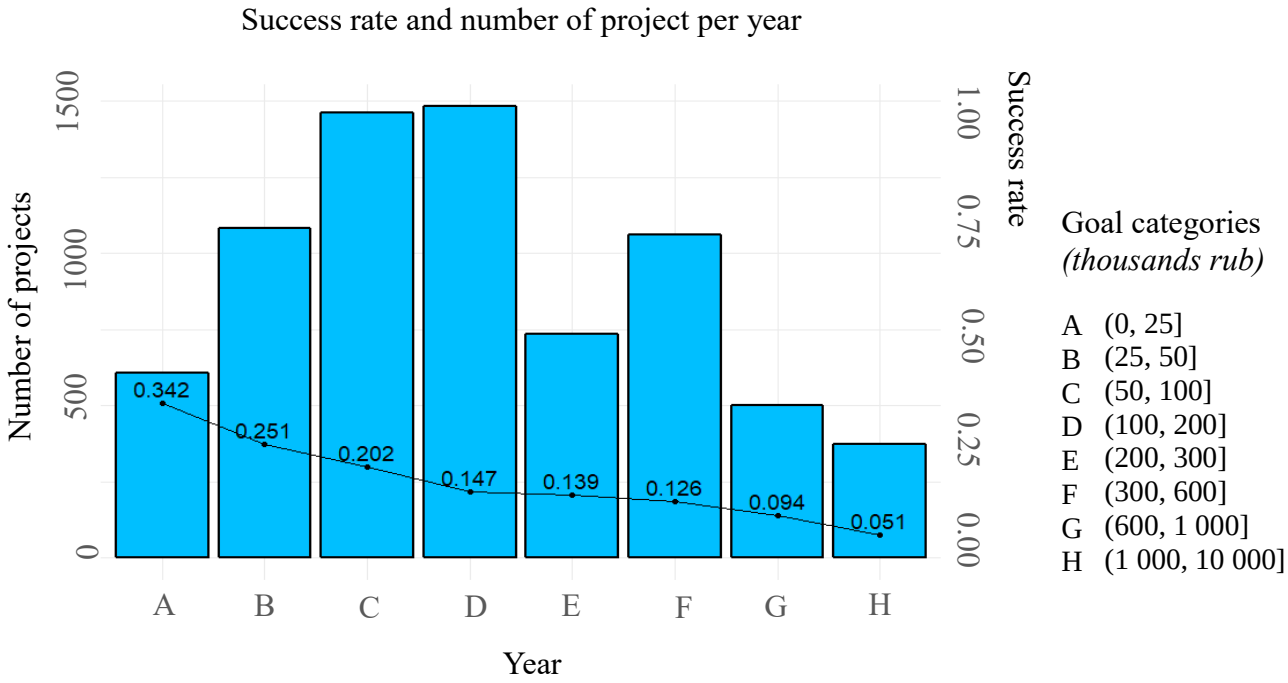

Fig. 14 Success rate according to Goal

Source: Author’s calculations

Second step of Signal Factor Analysis is run Logistic Regression Model for each variable and calculate the predictive power measure, Gini coefficient. As seen from results presented in Table 5, “Pledged.by.backer” and “News” have extremely high predictive power. Such variables as “Comments”, “Reposts” have Gini coefficient of about 50% which mean that these variables also have high predictive power. “Goal”, “Friends”, “Founder.projects”, “Founder.as.backer” and “Founder.sites” have bit worse but still quite good predictive power. What is interesting, these variables represent the information about project founder. Weaker predictive power was showed by such variables as “Rewards”, “Founder.text.length”, “Has.video” and “Images”. The predictive power of “Text.abstracts” and “Text.length” is similar to a random score. Overall, one can conclude that:

- Social Communication factors are the most important,
- Category is one of key factors,

- Founder-related factors are important,
- Main financial factor is an average pledge amount by backer,
- Description & Design factors are the least important.

Table 5 Single Factor Analysis results

Variable	Coefficient	Stand. Error	p_value	AUC	GINI
Pledged.by.backer	1,39361	5,16E-02	1,96E-160	90,99%	81,98%
News	0,29305	1,03E-02	1,84E-176	87,34%	74,69%
Comments	0,23891	1,21E-02	3,83E-87	75,05%	50,10%
Reposts	0,00348	2,03E-04	9,13E-66	74,90%	49,80%
Category.Rank	6,17340	3,71E-01	3,23E-62	70,23%	40,46%
Founder.projects	0,13609	1,16E-02	1,16E-31	68,47%	36,95%
Founder.Friends	0,13197	1,64E-02	8,71E-16	63,85%	27,69%
Founder.as.backer	0,39693	3,29E-02	1,34E-33	62,90%	25,80%
Goal	-0,40868	2,94E-02	6,44E-44	62,70%	25,41%
Founder.sites	0,18563	1,80E-02	6,85E-25	61,45%	22,90%
Rewards	0,06979	7,47E-03	9,45E-21	58,68%	17,36%
Founder.text.length	0,15024	2,66E-02	1,66E-08	58,64%	17,28%
Images	0,02789	3,80E-03	2,13E-13	56,93%	13,86%
Has.video	0,47476	7,53E-02	2,88E-10	56,63%	13,25%
Text.abstracts	0,00353	1,15E-03	2,16E-03	54,59%	9,18%
Text.length	0,19315	4,24E-02	5,33E-06	54,01%	8,01%

Design & Description factors

Financial factors

Social Communication factors

Founder-related factors

Source: Author's calculations

2.3.2 Multiple Factor Analysis

In order to identify strengths of collected factors in predicting crowdfunding project success on Boomstarter, two predictive models were built. First model was built based only on factors which can be obtained at the beginning of the money collection process. These factors are: Category.Rank, Founder.projects, Founder.Friends, Founder.as.backer, Goal, Founder.sites,

Rewards, Founder.text.length, Images, Has.video, Text.abstracts, and Text.length. Second model was built based on all available factors at the end at money collection process. It means that its input data also included the following factors: News, Comments, Reposts, Pledged.by.backer. What is worth mentioning, these are top 4 success factors due to Single Factor Analysis.

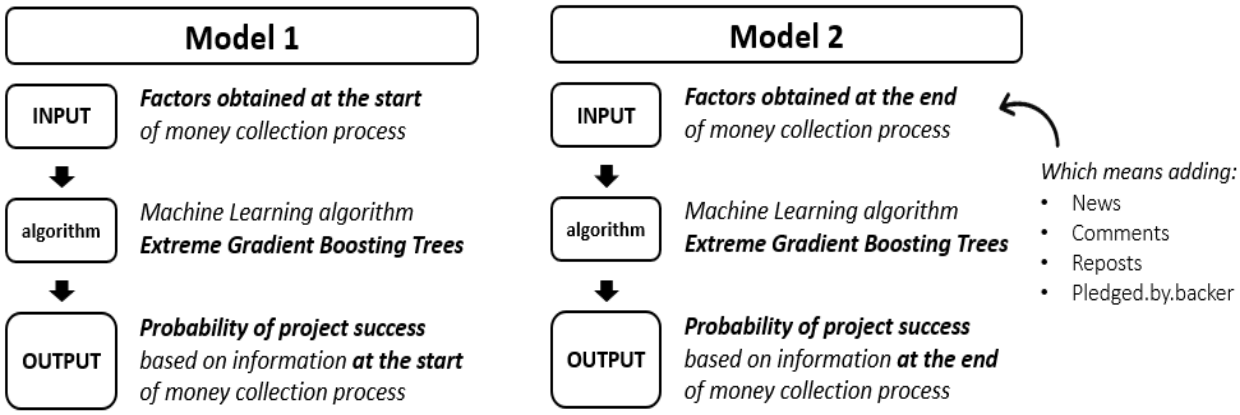

Fig. 15. Description of predictive models

Source: Made by the author

Both models were built by using machine learning algorithm Extreme Gradient Boosting trees which was realized on the programming language R in the library CARET. The sample of 7303 projects was randomly separated into train (75% of the data) and test (25% of the data) samples. Moreover, in order to avoid over-fitting, which is too perfect prediction on a train sample, which may lead to weak prediction power of other data, in particular test sample, the cross-validation with 70% was used. As it was mentioned in the research method section, it means that on each step algorithm randomly takes a 70% of a train sample and due to it does not allow a model to be over-fitted.

Table 6 Multiple Factor Analysis Results: Gini

Predictive power measure	**Model 1**	**Model 2**
Gini	70%	97%

Source: Author's calculations

Both models show a strong predictive power of collected proxy success factors. The first model showed Gini coefficient equal to 70% on the test sample, while the second model apparently outperformed the first one and showed Gini coefficient of 97%. The result of the first model shows

that factors which can be obtained at the beginning of the money collection process perform much better all together and show a good level of predictive power. But at the same time, the model performs still worse than such factors as News and Pledged.by.backer, which showed Gini equal to 74.69% and 81.98% respectively.

Gini of the second model shows that the classifier performs very closely to a perfect one and outperform a random one a lot. Gini of 97% corresponds to AUC equal to 98.5%, which means that probability that the classifier ranks a randomly chosen successful project higher than a randomly chosen unsuccessful one is almost 1.

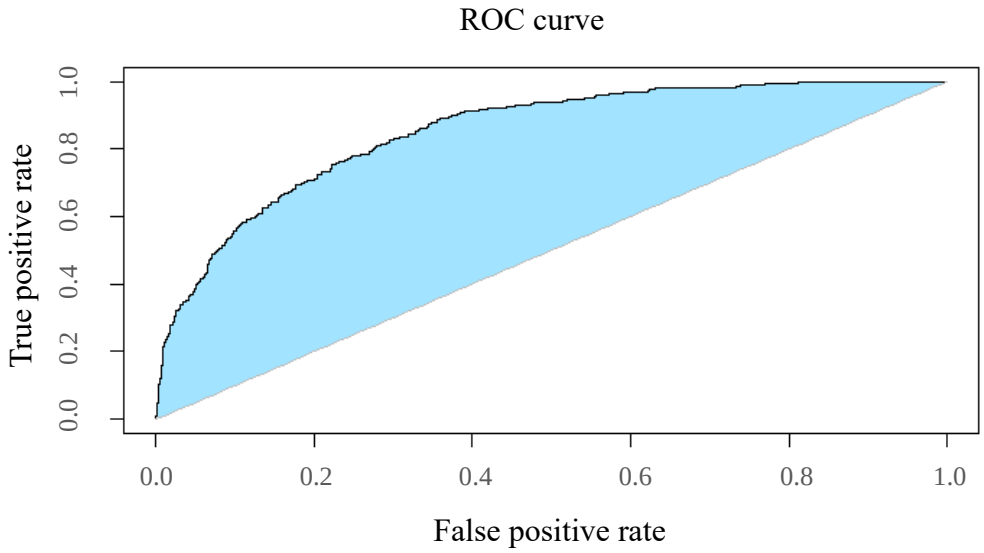

Fig. 16 ROC curve of Model 1

Source: Author's calculations

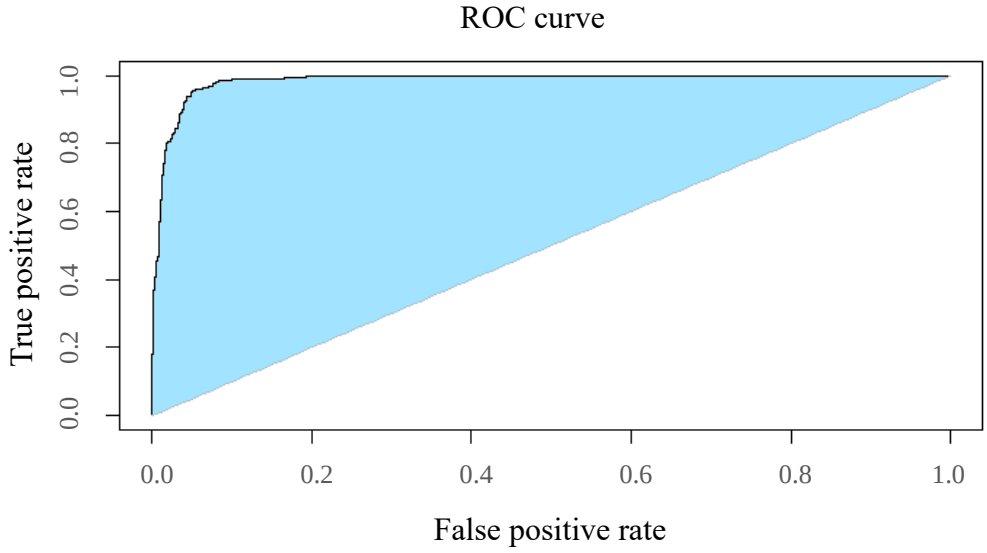

Fig. 17 ROC curve of Model 2

Source: Author's calculations

Apart from that, other predictive power measures such as Accuracy, Sensitivity and Specificity were calculated. All these measures depend on a particular threshold, which can be chosen based on the maximization of a target statistical measure of the performance. The results can be seen in Table 7.

If one set a threshold of 50%, the accuracy of the first model will be 85%. In fact, accuracy is not a good measurement of the prediction models performance in case of considered dataset, since the percentage of unsuccessful projects is very high – 81.3%. It means that an accuracy of a model which classify each project as an unsuccessful one will be the same 81.3%. Accuracy of the first model shows that it performed better than the described classifier. For the threshold of 50% the specificity is quite well and equals to 97%, while sensitivity in this case is only 30%. It shows that the model defined successful projects only in 30%. If to set a threshold lower, for example, 30%, the first model will identify successful projects in 54%, while will make more mistakes in identifying unsuccessful one.

Based on these threshold dependent measures, the second model also performs much better than the first one does. Based on the second model, the classifier gives high accuracy of 95% in case of both considered thresholds. When it comes to true positive rate, in case when threshold equals to 50% the classifier identifies successful projects with the rate of 81%, while in case when threshold equals to 36% – in 91% cases. True positive rate is high is case of both thresholds and shows better results with the higher cut-off value.

Table 7 Multiple Factor Analysis Results: Other predictive power measures

Predictive power measures	**Model 1**		**Model 2**	
	Threshold 50%	**Threshold 30%**	**Threshold 50%**	**Threshold 36%**
Accuracy	85%	84%	95%	95%
Specificity = True negative rate	97%	90%	98%	96%
Sensitivity = True positive rate	30%	54%	81%	91%

Source: Author's calculations

2.4. Findings and discussion

The goal of this master thesis was to identify key success factors of projects on Boomstarter.ru and their strength in determining crowdfunding project success.

The Single Factor Analysis, which goal was to identify key success factors of projects on Boomstarter.ru, showed that an average amount of money pledged by backer has the best predictive power. Along with this factor, social communication factors such as (1) number of news posted by project founder, (2) number of comments left by backers and (3) number of reposts on social networks Vkontakte and Facebook showed strong positive relation with project success. Therefore these group of factors can be considered as the most important one. This result is not surprising, since it shows importance not only the level of communication between the audience and the founder, but also the degree of the interest which the audience shows.

Moreover, the conducted analysis shows that such factor as project category matters for its success. In the considered models this factor was included as a percentage of a successful campaigns in each category. According to Boomstarter data, category “table games” had the highest success rate amounted at 40%. Slightly less success rate have “graphical design” and “illustrations” categories – 38%. Another category which have relatively high success rate is “society” – 31%. Projects of this category vary from a new leader school construction to an ecological online game creation. Following success rate of about 30% have “journalism” and “children's literature”. The most “successful” musical category is “jazz”, which success rate 40%. Rock music projects have success rate slightly less than 30% – 27%.

Founder-related factors show medium strength in predicting project success. The best predictive power out of this category of factors has a number of a previously launched projects by a founder, which shows his or her experience in crowdfunding projects realization. Another factor, which represents number of founder’s friends in social networks, can be also thought to be social communication factor and is in fact really important relatively to others. It is not a rare case when friends and relatives are ready to support founder’s idea. Such type of financing which is sometimes called “Friends, Family and Fools” may be realized through crowdfunding as well. What is also interesting, next founder related factor represents a number of projects which a founder has sponsored. Such a positive relation between number of sponsored projects and the project success can be explained by founder’s experience and interest in crowdfunding. Another founder-related factor is a number of funder’s web-sites attached to a description. Most of the time apart from his or her social page links an author adds web-sites with the project web-page, presence

of which may serve as a signal of a good quality project for a backer. A factor which correspond to a founder's description length has a weak relationship with project success. Undoubtedly, this factor cannot represent its content, which in this case should be more important for a potential backer. But at the same time, the provided graphical analysis shows that this factor still has a positive relation with success. Overall, founder-related factors are likely to reflect founder's experience in crowdfunding and a backer's feeling of trust to a project and, in particular, founder.

When it comes to financing success factors, the second predictive power rate after top-1 factor of an average amount pledged by backer is a goal amount of money which the founder wants to collect. It is the only factor which has negative relation with project success. This relationship is not surprising, since it is always much easier to collect smaller amount of funds, since it may need less backers and less average amount pledge by them. The weakest prediction power among financing factors show factor which correspond to number of reward levels which can be chosen by a backer.

Description and design proxy factors showed the weakest prediction power, which was slightly better than the prediction power of a random classifier. However, among this category factors which are related to a project design, such as number of images included in the campaign description and a presence of a video, have a bit better predictive power than factors directly related to a description text. According to a graphical analysis, the majority of projects include a video and in turn have a higher success rate within the group than those which does not have it. So, this binary factor cannot serve as a good classifier and may only add strength to another factors. Number of images is definitely makes description more attractive and easier to understand. However, its quality is not measured, thus this fact may explain such a weak predictive power of the factor. Such variables as a text length and a number of text abstracts showed the worst relationship with project success, although according to a graphical analysis show positive relation with the success.

To sum up, social communication factors along with an average amount of money pledged by backer showed the best predictive power. What is worth mentioning, these four factors can be obtained only at the end of money collection process. All of them, except from factor related to a number of news, represent the interest of an auditory in the project's idea and therefore can tell something about the level of advertisement.

The Multiple Factor Analysis showed that all factors which can be obtained only at the start of money collection process identify project success with much higher predictive power than

they have separately. The predictive power of this model can be classified as a high one, since its Gini equals to 70%. However, this model does not outperform such factors as an average amount pledged by backer and number of news posted by a founder.

The second multiple factor model which included all collected factors performes almost as a perfect classifier which Gini coefficient of 97% and an accuracy rate of 95% which means that it identifies project success/unsuccessful wrongly only in 5% of cases. Overall, the results of the Multiple Factor Analysis show a high predictive power of collected proxy success factors in the question of predicting project success on Boomstarter.

2.5. Managerial implications

First of all, this research will give founders a better understanding of factors which influence project success. But in fact this research gives founders much more than just information about success factors. Two built predictive model can serve as predictive tool and can enable founder (1) to measure the campaign progress in terms of probability of success and therefore (2) to get a feedback on their project's designs at the beginning of money collection process and (3) to get feedback on current state of affairs during money collection process.

Imagine a founder who is going to launch a crowdfunding project on Boomstarter. First of all, he or she can get an estimated probability of his or her crowdfunding project success by running the first predictive model. This information may serve as a ground to improve the crowdfunding project design and the idea presentation of it before the money collection process starts. Let's consider an example of a crowdfunding campaign "The Voronezh Alphabet" which was finished on Boomstarter on 25th of April in 2014 year. The probability of the project success according to the first predictive, which is based only on factors obtained at the beginning of money collection process, is 14%. Depending on a situation, this result may have lead the founder to change something in the project's idea presentation.

Moreover, he or she can test different values of four main success factors, which can be obtained only by the end of money collection process, and estimate probability of success based on all considered factors. For example, assuming that the project considered will have 20 News, 10 Comments, 500 Reposts and an average amount pledged by a backer equal to 1 000 rubles by the end of considered crowdfunding campaign, the probability of success will boost to 87%.

Apart from that, second model may help founders to measure campaign progress in terms of probability of success and thus get a feedback on current state of affairs, assuming that values

of top four success factors, which can be obtained only by the end of money collection process, will not change anymore. More precisely, second model results for a particular project can be illustrated at the 3D plot. By fixing one of four key success factors, for example news, which is directly controlled by the founder, he or she can analyze the prospects of his or her crowdfunding project success by looking at the 3D plot which shows probability of success depending on the remaining factors such as Comments, Reposts and Pledged.by.backer. In the considered example, assuming a value of News of 2, the founder may use a 3D plot presented on Figure 18 to get a feedback on the current state of affairs during the money collection process. What is worth mentioning, even though project's founder can control number of posted news, he or she should not thoughtlessly increase number of such updates, since it may not definitely lead to increase of the probability of campaign success.

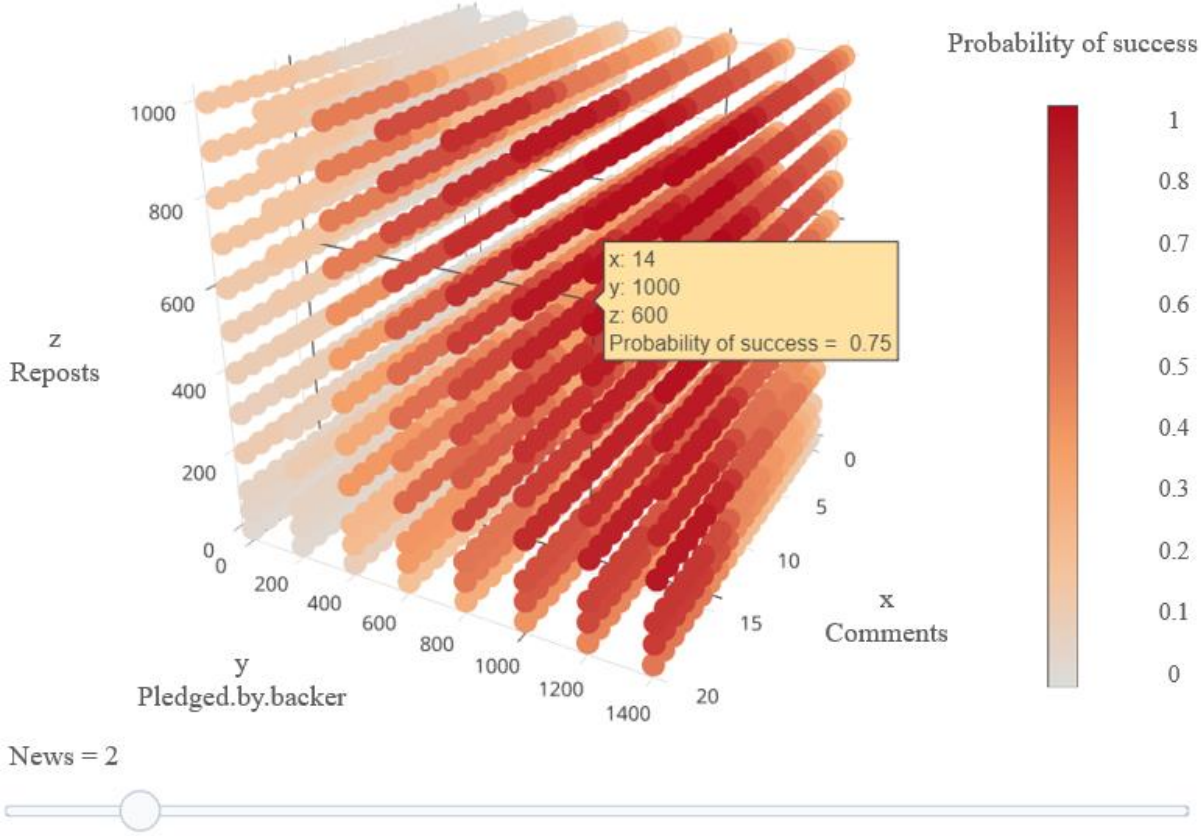

Fig. 18 3D plot of probability of success depending on Comments, Reposts and Pledged.by.backer factors with fixed News equal to 2
Source: Author's calculations

Apart from managerial implication for founders, predictive tools, which were built in the scope of this study, can help backers as well. Imagine a situation when a backer has limited sources and several interesting to him or her projects. Probability of campaign success may serve as one more criteria for a project investment.

2.6. Research limitations

When it comes to limitations of this master thesis, one can state that it is a limited range of factors analyzed. In fact, there are many factors which obviously would be important to include in the models.

First variable which come to the mind is a campaign duration. This factor was not impossible to collect since the end date of each ended project was not available on Boomstarter. This factor can significantly enforce strength of a factor which represents a goal amount of money which a founder wants to raise on the crowdfunding platform. It is so due to the fact that even though a goal amount of a project is not too high, the duration of the money collection period is so small so that it leads to an excessively high average amount of money which should be attracted per day. Apart from that, it will be interesting to analyze whether too long duration period lead to a possible backer's conclusion that a project founder does not believe in a projects success so that he or she afraid not to collect money.

Another important factor which was not measured is a position of a campaign on Boomstarter web page. Whether the campaign was presented on the platform's main web-page of not, if it was so, for how many hours is was mentioned on it. Another possible metrics was a number of mouse clicks which should be done in order to find each project from the main page. Moreover, when it comes to advertising, it is worth to measure the level of advertising not only by social network reposts but also through some articles in popular magazines.

Some projects are closely related to a temporary popular topics. For example, such topic as crypto-currency is highly popular nowadays and projects which are related to similar to this topic may have higher probability of success. However, there are many projects which topic are not popular but still finds its demand. Therefore, the factor which identifies whether project topic is in trend or not may enforce the predictive power of collected factors.

As was mentioned, description and design factors showed weak success predictive power. One of the possible reasons of such poor performance can be the fact that such factors does not measure the quality of the idea presentation sufficiently. For example, the quality of an attached video, its filming and the overall attractiveness may show much better results. And when it comes to description related factors, one can consider text complexity and text's semantic characteristics such as for example "mood" (whether it is positive, neutral or negative).

Another factor which could be included in the model is a diversification a rewards offered for backers. This factor may also have a positive influence on the project success since backers may take care of a present they would like to get in exchange for their money pledged.

Overall, the main limitation of the paper is that mainly superficial factors were considered in the research. But even though there are plenty of factors which may actually influence crowdfunding project success, the collected proxy factors are enough to give to predict project success with a very high accuracy.

Summary

In order to achieve the main goal of the research paper, the data which was collected directly from Boomsarter.ru, was analyzed to identify main success factors of crowdfunding projects on Boomstarter. Factors strength was measured by Gini coefficient. Firstly, Single Factor Analysis was provided. Calculations showed that main success factors are an average amount of money pledged by backer and three social-communication factors.

Then, two models were built based on Machine Learning algorithm Extreme Gradient Boosting Trees in order to identify strength of collected factors altogether. These predictive models show high predictive power. Thereby, Multiple Factor Analysis showed a high predictive power of collected proxy success factors in the question of identifying project success on Boomstarter.ru.

CONCLUSIONS

The following research paper was aimed to identify key success factors of projects on Boomstarter.ru and their strength in determining crowdfunding project success. In order to achieve this goal, firstly, literature analysis was conducted. According to the existing pool of research papers, there are no studies based on Russian crowdfunding platforms which may help project's founders to succeed in collecting money process.

Second step was a data collection process. Unfortunately, it is impossible to measure such intuitive success factors as an idea behind, the quality of its presentation on crowdfunding project web-page and the level of advertising directly. Due to this fact, only proxy factors were derived directly from one of the largest Russian crowdfunding platforms Boomstarter.ru by a special script written on a programming language Python.

Key success factors are an average amount pledged by backers and social-communication factors such as (1) number of news posted by project founder, (2) number of comments left by backers and (3) number of reposts in social networks Vkontakte and Facebook. This result is not surprising, since it shows not only the importance of communication level between the audience and the founder, but also the degree of the interest which the audience shows. What is worth mentioning, these four main factors are the only collected factors which can be obtained only at the end of money collection process.

Moreover, the conducted analysis showed that such factor as project category and founder-related factors are important in terms of project success as well. Medium prediction power of founder-related factors may lead to a conclusion that founder's experience in crowdfunding and a backer's feeling of trust to a project and, in particular, to a founder matters for the project success. The least prediction power was showed by description and design proxy factors. One of the possible reasons of such poor performance can be the fact that such factors does not measure the quality of the idea presentation properly.

In order to identify collected success factors strength in determining crowdfunding project success, two predictive models were built by using a machine learning algorithm Extreme Gradient Boosting Trees. The analysis showed that all factors, which can be obtained only at the start of money collection process, identify project success with much higher predictive power than they have separately. The predictive power of this model can be classified as a high one, although this model does not outperform such factors as an average amount pledged by backer and number of news posted by a founder. The second multiple factor model, which included all collected factors,

perform almost as a perfect classifier with Gini coefficient of 97% and an accuracy rate of 95%, which means that it identifies project success state wrongly only in 5% of cases. Overall, the results of the Multiple Factor Analysis show a high predictive power of collected proxy success factors in the question of identifying project success on Boomstarter.

The main theoretical value is that this paper is pioneered paper which identifies success factors of crowdfunding projects on Russian crowdfunding platform. Apart from that, such research is an interdisciplinary research which takes into account not only financial factors, but also social-communication, founder-related and design and descriptions ones. Doubtless, there is a room for improvement regarding a wider range of factors considered. This research limitation may serve as a possible gap for future researches.

The main practical value is that two predictive models which are build based on the factors obtained at the beginning and at the end of money collection process enable founders to measure the campaign progress in terms of probability of success. These model helped not only to identify factor's strength in determining crowdfunding project success, but also to get a feedback to project's founders at the beginning and during the money collection process on crowdfunding platform Boomstarter.ru.

To sum up, the conducted study is pioneered paper in studying of success factors of Russian crowdfunding projects, which showed not only high predictive power of collected factors, but also the importance of an interdisciplinary approach studying purely financial problem of raising capital.

REFERENCES

Agrawal, A., Catalini, C., Goldfarb, A., 2010. The geography of crowdfunding. *SSRN Electronic Journal*.

Agrawal, A., Catalini, C. and Goldfarb, A. 2014. Some Simple Economics of Crowdfunding. *NBER Innovation Policy & the Economy*. Vol. 14(1), 63 – 97.

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

Andenes, S., Pendegraft, E., 2016. Financing of Startups: A Comparative Study of Norway and the USA. Master thesis. Norwegian School of Economics.

Bartlett, MS, 1947. The Use of Transformations. *Biometrics*. Vol. 3, 39 – 52.

Beaulieu, T., Sarker, S. 2013. Discursive meaning creation in crowdfunding: A socio-material perspective. *AIS Electronic Library*.

Beckwith, J. J. 2016. Predicting Success in Equity Crowdfunding. *Joseph Wharton Scholars.*

Beier, M., Wagner, K. 2015. Crowdfunding success: a perspective from social media and e-commerce. *AIS Electronic Library*.

Belleflamme, P., Lambert, T., Schwienbacher, A., 2014. Crowdfunding: Tapping the Right Crowd. *Journal of Business Venturing*. Vol. 29(5), 585 – 609.

Berger, A., Udell, G., 1995. Relationship lending and lines of credit in small firm finance. The *Journal of Business*. Vol. 68, 351 – 381.

Block, J.H., Hornus, L., Moritz, A. 2016. Which updates during an equity crowdfunding campaign increase crowd participation? *SSRN Electronic Journal*.

Boeuf, B., Darveau, J., Legoux, R. 2014. Financing creativity: Crowdfunding as a new approach for theatre projects. *International Journal of Arts Management*. Vol. 16(3), 33.

Boddy, R., Smith, G. 2009. Statistical methods in practice: for scientists and technologists. 1th ed. *New York: John Wiley & Sons.*

Breiman, L., 1997. Arcing the edge. *Technical Report 486*. Department of Statistics, University of California, Berkeley.

Breiman, L., 1998. Arcing classifiers (with discussion). *The Annals of Statistics*. Vol. 26(3), 801 – 849.

Breiman, L., 1999. Prediction games and arcing algorithms. *Journal of Neural Computation*. Vol. 11, 1493 – 1517.

Breiman, L., 2000. Some infinite theory for predictor ensembles. *Technical Report 579*. Department of Statistics, University of California, Berkeley.

Brown, M., Degryse, H., Hoewer, D., Penas, M. 2012. How Do Banks Screen Innovative Firms? Evidence from startup Panel Data. *ZEW – Centre for European Economic Research*. Discussion Paper No. 12 – 032.

Burtch, G., Ghose, A., Wattal, S., 2011. An empirical examination of the antecedents and consequences of investment patterns in crowd-funded markets. *Information Systems Research*. Vol. 24(3), 499 – 519.

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

Butticè, V., Colombo, M. G., & Wright, M. 2017. Serial crowdfunding, social capital, and project success. *Entrepreneurship Theory and Practice*. Vol. 41(2), 183 – 207.

Butticè, V., Colombo, M. G., Franzoni, C., & Rossi-Lamastra, C. 2014. Signaling to the Crowd. Heterogeneous tastes and the response to reputational signals in film crowdfunding. *14th Annual REER Conference*, 1 – 28.

Butticè, V., Franzoni, C., Rossi-Lamastra, C., & Rovelli, P. 2017. The road to crowdfunding success: a review of extant literature. *Oxford University Press*.

Calic, G., Mosakowski, E. 2016. Kicking off social entrepreneurship: how a sustainability orientation influences crowdfunding success. *Journal of Management Studies*. Vol. 53(5), 738 – 767.

Casamatta, C., Haritchabalet, C., 2011. Dealing with venture capitalists: shopping around or exclusive negotiation. Unpublished Working Paper.

Cassar, G., 2004. The financing of business start-ups. *Journal of Business Venturing*. Vol. 19. 261 – 283.

Chen T., Guestrin C., 2016. XGBoost: A Scalable Tree Boosting System. International *Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, USA, 785 – 794.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. *Cornell University Library.*

Chen, X.-P., Yao, X., Kotha, S., 2009. Entrepreneur passion and preparedness in business plan presentations: a persuasion analysis of venture capitalists' funding decisions. *Academic Management Journal*. Vol. 52, 199 – 213.

Colombo, M.G., Franzoni, C., Rossi-Lamastra, C. 2015. Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship: Theory and Practice*, Vol. 39, 75 – 100.

Cosh, A., Cumming, D., Hughes, A., 2009. Outside entrepreneurial capital. *The Economic Journal*. Vol. 119, 1394 – 1533.

Courtney, C., Dutta, S., Li, Y., 2017. Resolving information asymmetry: Signaling, endorsement, and crowdfunding success. *Entrepreneurship: Theory and Practice*. Vol. 41, 265 – 290.

Dingman, S. (2013). Canadian smart-watch startup VC funding. *The Globe*. https://www.theglobeandmail.com/technology/business-technology/ canadians-smartwatch-startup-matches-record-15-million-in-vc-funding/article11965214/

Dushnitsky, G., & Marom, D. (2013). Crowd monogamy. *London Business School Review*. Vol. 24(4), 24 – 26.

Ferrary, M., Granovetter, M., 2009. The role of venture capital firms in Silicon Valley's complex innovation network. *Economy and Society*. Vol. 18, 326 – 359.

Frank, E., Harrell, Jr., 2001. Ordinal Logistic Regression. *Regression Modeling Strategies*, 331 – 343.

Friedman, J., Hastie, T., and Tibshirani, R. 1998. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*. Vol. 28(2), 337 – 407.

Frydrych, D., Bock, A. J., Kinder, T., & Koeck, B. 2014. Exploring entrepreneurial legitimacy in reward-based crowdfunding. Venture Capital. Vol. 16(3), 247 – 269.

Galuszka, P., & Bystrov, V. 2014. The rise of fanvestors: A study of a crowdfunding community. *First Monday*. Vol. 19(5).

Gerber, E. M., Hui, J. 2013. Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction*. Vol. 20(6), 34:1 – 34:32.

Giudici, G., Guerini, M., Rossi-Lamastra, C. 2018. Reward-based crowdfunding of entrepreneurial projects: the effect of local altruism and localized social capital on proponents' success. *Small Business Economics*. Vol. 50(2), 307 – 324.

Gleasure, R., Feller, J. 2014. From the wisdom to the Wealth of Crowds: A Metatriangulation of Crowdfunding research. TOTO Research Project working paper.

Gleasure, R., Feller, J. 2016. Does Heart or Head Rule Donor Behaviors in Charitable Crowdfunding Markets? International Journal of Electronic Commerce. Vol 20(4), 499 – 524.

Gompers, P., Lerner, J., 2004. The Venture Capital Cycle. *MIT Press*.

Gompers, P.A. and Sahlman, W.A. 2002. Entrepreneurial Finance: A Case Book. *New York: John Wiley & Sons.*

Gorman, B. 2017. A Kaggle Master Explains Gradient Boosting. http://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/

Gorman, M., Sahlman, W.A., 1989. What do venture capitalists do? *Journal of Business Venturing*. Vol. 4, 231 –248.

Greenberg, M. et al., 2013. Crowdfunding Support Tools: Predicting Success & Failure. *Conference on Human Factors in Computing Systems*, 1815 – 1820.

Hellmann, T., 2007. Entrepreneurs and the process of obtaining resources. *Journal of Economics and Management Strategy*. Vol. 16, 81 – 109.

Heminway, J., Hoffman, S., 2010. Proceed at your peril: crowdfunding and the securities act of 1933. *Tennessee Law Review*. Vol. 78, 879 – 972.

Jian, L., Shin, J. 2015. Motivations behind donors' contributions to crowdfunded journalism. *Mass Communication and Society*. Vol. 18(2), 165 – 185.

Jian, L., Usher, N. 2014. Crowd-funded journalism. *Journal of Computer-Mediated Communication*. Vol. 19(2), 155 – 170.

Josefy, M., Dean, T. J., Albert, L. S., Fitza, M. A. 2017. The Role of Community in Crowdfunding Success: Evidence on Cultural Attributes in Funding Campaigns to "Save the Local Theater". *Entrepreneurship: Theory and Practice*. Vol. 41(2), 161 – 182.

Kim, Y., Shaw, A., Zhang, H., Gerber, E. 2017. Understanding trust amid delays in crowdfunding. *ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2017*. Association for Computing Machinery.

Kirsch, D., Goldfarb, B., Gera, A., 2009. Form or substance: the role of business plans in venture capital decisions making. *Strategic Management Journal*. Vol. 30, 487 – 515.

Klačmer Čalopa, M., Horvat J., Lalić M., 2013. Analysis of financial sources for start-up companies. *Journal of Management*. Vol. 19, 19 – 44.

Koch, J. A., & Siering, M. 2015. Crowdfunding success factors: the characteristics of successfully funded projects on crowdfunding platforms. *Twenty-Third European Conference on Information Systems*, Münster, Germany.

Kortum, S., Lerner, J., 2000. Assessing the contribution of venture capital to innovation. RAND *Journal of Economics*. Vol. 31, 674 – 692.

Kotha, R., & George G. 2012. Friends, family or fools: Entrepreneur experience and its implications for equity distribution and resource mobilization. *Journal of Business Venturing*. Vol. 27, 525 – 543.

Kovačić, I., 2011. Short introduction into startups for people with ideas. http://www.ucionica.net/internet/kratki-uvod-u-svijet-startupa-za-sve-s-idejama-1427/

Krausa, S., Richterb, C., Bremc, A., Cheng, C.F., Chang, M.L. 2016. Strategies for reward-based crowdfunding campaigns. *Journal of Innovation & Knowledge*. Vol. 1, 13 – 23.

Kuppuswamy, V., Bayus, B. 2015. Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. *UNC Kenan-Flagler.* Research Paper No: 2013 – 15.

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

Kuzmenko, V., 2016. Crowdfunding market in Russia: Analysis, History and Prospects. [In Russian]. Lomonosov Moscow State University.

Li, Z., & Jarvenpaa, S. 2015. Motivating IT-Mediated Crowds: The effect of goal setting on project performance in online crowdfunding. *Thirty Sixth International Conference on Information Systems*.

Liao, C., Zhu, Y., Liao, X. (2015). The Role of Internal and External Social Capital in Crowdfunding: Evidence from China. *Revista de Cercetare si Interventie Sociala*. Vol. 49, 187 – 204.

Massolution. 2013. The crowdfunding industry report.

McKinsey & Company, 2007. From idea to new business: A manual for new growth companies. [In Norwegian] Oslo.

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

Mollick, E. 2016. Containing Multitudes: The Many Impacts of Kickstarter Funding. *SSRN Electronic Journal*.

Nielsen, D. Tree Boosting With XGBoost: Why Does XGBoost Win "Every" Machine Learning Competition? Master Thesis. Norwegian University of Science and Technology.

Ordanini, A., Miceli, L., Pizzetti, M. and Parasuraman, A. 2011. Crowdfunding: transforming customers into investors through innovative service platforms. *Journal of Service Management*, Vol. 22(4), 443 – 470.

Parhankangas, A., Renko, M., 2017. Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Ventures*. Vol. 32, 215 – 236.

Qiu, C. (2013). Issues in crowdfunding: Theoretical and empirical investigation on Kickstarter. *SSRN Electronic Journal*.

Rezac, M., & Rezac, F. 2011. How to measure the quality of credit scoring models. *Finance a Uver*. Vol. 61(5), 486.

Shane, S., Cable, D., 2002. Network ties, reputation, and the financing of new ventures. *Journal of Management Science*. Vol. 48, 364 – 381.

Sørensen, I. E. 2012. Crowdsourcing and outsourcing: the impact of online funding and distribution on the documentary film industry in the UK. *Media, Culture & Society*. Vol. 34(6), 726 – 743.

Thies, F., Wessel, M., Benlian, A. 2014. Understanding the Dynamic Interplay of Social Buzz and Contribution Behavior within and between Online Platforms – Evidence from Crowdfunding. Proceedings of the 35th International Conference on Information Systems.

Tucci, C., Afuah, A., Viscusi G. 2018. Creating and capturing value through crowdsourcing. *Oxford: Oxford University Press.*

Vulkan, N., Astebro, T., Sierra, M. F. (2016). Equity crowdfunding: a new phenomena. *Journal of Business Venturing Insights*. Vol. 5, 37 – 49.

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

Xiao, S., Tan, X., Dong, M., Qi. J. 2014. How to Design Your Project in the Online Crowdfunding Market? Evidence from Kickstarter. *Proceedings of the Thirty Fifth International Conference on Information Systems*. Auckland: New Zealand.

Xu, A., Yang, X., Rao, H., Fu, W. T., Huang, S. W., Bailey, B. P. 2014. Show me the money! An analysis of project updates during crowdfunding campaigns. *32nd annual ACM conference on human factors in computing systems*, 591 – 600.

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

Zheng, H., Li, D., Wu, J., Xu, Y. 2014. The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Information and Management*. Vol. 51(4), 488 – 496.

Zvilichovsky, D., Inbar, Y., & Barzilay, O. 2013. Playing both sides of the market: Success and reciprocity on crowdfunding platforms. *Thirty Fourth International Conference on Information Systems*, Milan.

Zvilichovsky, D., Inbar, Y., & Barzilay, O. 2015. Playing both sides of the market: Success and reciprocity on crowdfunding platforms. *SSRN Electronic Journal*.

Crowdfunding Hacks website. http://crowdfundinghacks.com/different-types-of-crowdfunding/

Crowdfunding platforms. https://boomstarter.ru/, https://planeta.ru/

Huffingtonpost news portal. https://www.huffingtonpost.com/david-drake/2000-global-crowdfunding-_b_8365266.html

Platform for predictive modelling and analytics competitions. https://www.kaggle.com/

Russian Internet edition and news aggregator. https://tjournal.ru/58082-mnogo-ili-malo-pyatiletka-rossiyskogo-kraudfandinga

APPENDICES

Appendix A. Data collection script on Python

```
# STEP 1
# Get all projects links from the web-site

j = 0
ar = {}
for letter in ["а", "у", "о", "и", "я", "ю", "е"]:
    print(letter)
    for i in range(1, 2000):
        url = 'https://boomstarter.ru/projects/search?button=&page='
        +str(i)+'&q=' + letter + '&utf8=%E2%9C%93'
        r = requests.get(url)
        r.encoding = 'utf-8'
        tree = html.fromstring(r.text)
        projects = tree.xpath("//h5/a")
        if len(projects)==0:
            break
        for project in projects:
            j=j+1
            pr_url = project.attrib["href"]
            ar.update({pr_url:''})

with open('ALL LINKS FINAL.csv', 'w', newline="") as csvfile:
    fieldnames = ['ID','link']
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()
    i = 0
    for item in ar.keys():
        i = i+1
        writer.writerow({'ID': i, 'link': str(item)})

# STEP 2
# Get all descriptive information and save to Data.csv

df = pd.read_csv('ALL LINKS FINAL.csv')
all_links = df.link
all_links = list(all_links)

with open('DATA.csv', 'w', newline="") as csvfile:

    fieldnames = ['ID', 'Link', 'Title', 'Location', 'Category',
                  'Description', 'Pledged', 'Goal', 'Backers',
                  'Progress', 'News', 'Comments', 'Current',
                  'State', 'State all', 'End Date', 'Start Date',
                  'Duration', 'Founder projects', 'Founder as backer',
                  'VK friends', 'FB friends', 'Founder sites',
                  'Founder text length', 'Has video', 'Imagies',
                  'Rewards', 'Levels', 'Text length', 'Paragraph']
    writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
    writer.writeheader()

    i = 0
    for pr_url in all_links:

        i = i + 1

        # read file
        r = requests.get(pr_url)
        r.encoding = 'utf-8'
```

```
project = html.fromstring(r.text)
filename = pr_url.split('/')[5]

# create dictionary
filerow = {}

# ID
filerow.update({'ID':str(i)})

# LINK
filerow.update({'Link':pr_url})
link = pr_url

# LOCATION
try:
    filerow.update({'Location':project
    .xpath("//li[@class='location']/a/text()")[0]})
except:
    print('location')
    print(link)

# CATEGORY
try:
    filerow.update({'Category':project
    .xpath("//li[@class='category']/a/text()")[0]})
except:
    print('category')
    print(link)

# PLEDGED
try:
    filerow.update({'Pledged':project
    .xpath("//div[@class='pledged']/text()")[0].replace("\n","")
    .replace("\r","").replace(" ", "").replace("₽","")})
except:
    print('PLEDGED')
    print(link)

# GOAL
try:
    filerow.update({'Goal':project
    .xpath("//div[@class='goal']/text()")[1]
    .replace('\n', '').replace("из","")
    .replace("\r\n","").replace(" ","")})
except:
    print('GOAL')
    print(link)

# BACKERS
try:
    filerow.update({'Backers':project
    .xpath("//div[@class='backers']/text()")[0]
    .split(" ")[0]})
except:
    print('BACKERS')
    print(link)

# PROGRESS
try:
    filerow.update({'Progress':project.xpath
    ("//li[@class='percentage']/text()")[0]})
except:
    print('PROGRESS')
```

```
            print(link)

        # NEWS
        if (project.xpath("//li[@class='js-tab'][1]/a/span/text()")):
            filerow.update({'News':project
            .xpath("//li[@class='js-tab'][1]/a/span/text()")[0]})
        else:
            filerow.update({'News':'0'})

        # COMMENTS
        if (project.xpath("//li[@class='js-tab'][2]/a/span/text()")):
            filerow.update({'Comments':project
            .xpath("//li[@class='js-tab'][2]/a/span/text()")[0]})
        else:
            filerow.update({'Comments':'0'})

        # CURRENT: current? 2 inf 0 old 1 current
        if (project.xpath("//li[@class='timer']/text()")[0]
        .replace('\n', '').replace("\r\n", "")):
            if (project.xpath("//li[@class='timer']/text()")[0]
            .replace("\r\n", "")=='До цели'):
                filerow.update({'Current':'2'})
            else:
                filerow.update({'Current':'0'})
        else:
            filerow.update({'Current':'1'})

        # STATE: win = 1, not = 0, progress = 0
        if (filerow['Current']=='0' and
float(filerow['Progress'].strip('%'))>=100):
            filerow.update({'State':'1'})
        else:
            filerow.update({'State':'0'})

        # STATE ALL: win = 1, not = 0, progress = 1|0
        if (float(filerow['Progress'].strip('%'))>=100):
            filerow.update({'State all':'1'})
        else:
            filerow.update({'State all':'0'})

        # DATES
        if (filerow['Current']=='0'):
            filerow.update({'End Date':project
            .xpath("//li[@class='timer']/text()")[0]
            .replace("Завершен","").replace("\n","")})

        if (filerow['Current']=='1'):
            filerow.update({'End Date':project
            .xpath("//div[@class='hint']/span/text()")[1]
            .replace("в воскресенье ","").replace("в понедельник ","")
            .replace("во вторник ","").replace("в среду ","")
            .replace("в четверг ","").replace("в пятницу ","")
            .replace("в субботу ","")[:-12]})

        if (filerow['Current']=='2'):
            filerow.update({'End Date':project
            .xpath("//div[@class='hint']/span/text()")[0]
            .replace("Проект начался ","")
            .replace(" и завершится", "")})

        if (filerow['Current']=='1'):
            filerow.update({'Start Date':project
            .xpath("//div[@class='hint']/span/text()")[0]
```

```
            .replace("Проект начался ","").replace(" и завершится", "")})

        # FOUNDER PROJECTS
        if (len(project.xpath("//li[@class='achievements']/a/text()"))==4):
            filerow.update({'Founder projects':project
            .xpath("//li[@class='achievements']/a/text()")[0]
            .replace('спонсор ','')})
        else:
            filerow.update({'Founder projects':0})

        # FOUNDER AS BACKER
        try:
            filerow.update({'Founder as backer':project.xpath
            ("//li[@class='achievements']/a/text()")[1].replace('автор
','')})
        except:
            print('FOUNDER AS BACKER')
            print(link)

        # FOUNDER NAME
        try:
            filerow.update({'Author name':project.xpath
            ("//li[@class='name js-about-creator']/text()")[0]})
        except:
            print('FOUNDER NAME')
            print(link)

        # FOUNDER SITES
        try:
            filerow.update({'Founder sites':
            len(project.xpath("//ul[@class='sites']/li"))-1})
        except:
            print('FOUNDER SITES')
            print(link)

        # FB FRIENDS
        if (project.xpath("//ul[@class='public-
info']/li[@class='facebook']/span/text()")):
            filerow.update({'FB friends':project.xpath
            ("//ul[@class='public-
info']/li[@class='facebook']/span/text()")[0]
            .replace(' друга','').replace(' друзей','')})
        else:
            filerow.update({'FB friends':0})

        # VK FRIENDS
        if (project.xpath("//ul[@class='public-
info']/li[@class='vkontakte']/span/text()")):
            filerow.update({'VK friends':project
            .xpath("//ul[@class='public-
info']/li[@class='vkontakte']/span/text()")[0]
            .replace(' друга','').replace(' друзей','')})
        else:
            filerow.update({'VK friends':0})

        # HAS VIDEO
        if (project.xpath("//div[@class='video-wrapper']/img")):
            filerow.update({'Has video':0})
        else:
            filerow.update({'Has video':1})

        # IMAGIES
        try:
```

```
        filerow.update({'Imagies':len(project
        .xpath("//div[@class='description']")[0].xpath("//img"))})
    except:
        print('IMAGIES')
        print(link)

    # TEXT
    try:
        discription = project
        .xpath("//div[@class='description']")[0]
        .text_content()
        discription_list = discription.split('\n')

        file = open('boomDataTEXT/' + filename[:150]
        + '.txt',"w", encoding='utf-8')
        j = 0
        for s in discription_list:
            j = j + 1
            if (j == 1):
                file.write(s)
            else:
                file.write('\n' + str(s))
        file.close()

        # LENGTH
        filerow.update({'Text length':len(discription)})

        # PARAGRAPH
        filerow.update({'Paragraph':len(project
        .xpath("//div[@class='description']/p"))})

    except:
        print('TEXT')
        print(link)

    # FOUNDER TEXT
    try:
        discription = project
        .xpath("//div[@class='biography']")[0]
        .text_content()
        discription_list = discription.split('\n')

        file = open('boomDataTEXTfounder/' + filename[:150]
        + '.txt',"w", encoding='utf-8')
        j = 0
        for s in discription_list:
            j = j + 1
            if (j == 1):
                file.write(s)
            else:
                file.write('\n' + str(s))
        file.close()

        # FOUNDER TEXT LENGTH
        filerow.update({'Founder text length':len(discription)})
    except:
        print('FOUNDER TEXT')
        print(link)

    try:
        writer.writerow(filerow)
    except:
        writer.writerow({'ID':str(i)})
```

```
            print(link)

# STEP 3
# Get rewards levels and social reposts from VK, FB and save to Data 2.csv

def save_social_reposts(path, links):
    with open(path + 'DATA 2.csv', 'w', newline="") as csvfile:
        driver = webdriver.Chrome()
        fieldnames = ['ID', 'link', 'vk', 'fb', 'Rewards', 'Levels']
        writer = csv.DictWriter(csvfile, fieldnames=fieldnames)
        writer.writeheader()
        i = 0
        for item in links:
            driver.get(item)
            time.sleep(2)
            # create dictionary
            filerow = {}

            # ID
            i = i + 1
            filerow.update({'ID':str(i)})

            # Filename
            filerow.update({'link': str(item)})

            #VK
            try:
                vk = driver.find_element_by_xpath
                ("//*[@class='social-likes__counter social-
likes__counter_vkontakte']").text
            except NoSuchElementException:
                vk = 0
            filerow.update({'vk': vk})

            # FB
            if i%30==0:
                time.sleep(400)

            try:
                fb = driver.find_element_by_xpath
                ("//*[@class='social-likes__counter social-
likes__counter_facebook']").text
            except NoSuchElementException:
                fb = 0
            if fb!=0:
                print(fb)
            filerow.update({'fb': fb})

            #levels
            amounts = driver.find_elements_by_xpath("//*[@class='amount']")
            levels = ''
            k = 0
            for a in amounts:
                if (k == 0):
                    levels = levels + amounts[k].text.replace('a',
'').replace(' ', '')
                else:
                    levels = levels + '&' + amounts[k].text.replace('a',
'').replace(' ', '')
                k = k + 1
            filerow.update({'Levels': levels})
            #write a row
            writer.writerow(filerow)
```

Appendix B. Variables description statistics

Name	Mean	SD	Min	Max
Goal	1 763 767,72	111 309 203,47	73	9 500 000 000
News	2,78	6,13	0	86
Comments	7,12	334,22	0	28 152
Reposts	98,24	426,85	0	22 012
Rewards	7,76	4,39	0	69
CategoryRank	0,18	0,09	0	0,67
Pledged.by.backer	4 919,57	74 822,85	0	2 984 320
Friends	554,96	1 104,90	0	9 997,00
Founder.projects	1,33	5,71	0	195
Founder.as.backer	1,39	1,38	0	19
Founder.sites	1,85	1,82	0	20
Founder.text.length	509,22	769,18	0	13 297
Has.video	0,58	0,49	0	1
Images	12,48	8,42	4	110
Text.abstracts	28,90	27,75	0	437
Text.length	3 636,02	3 918,81	0	226 444

Appendix C. Variables correlation matrix

	State	Log.Goal	News	Comments	Reposts	Rewards	Log.Pledged.by.backer	CategoryRank	Log.Friends	Founder.projects	Founder.as.backer	Founder.sites	Log.Founder.text.length	Has.video	Images	Paragraph	Log.Text.length
State	1,00																
Log.Goal	-0,19	1,00															
News	0,54	-0,01	1,00														
Comments	0,01	0,01	0,03	1,00													
Reposts	0,23	0,11	0,29	0,01	1,00												
Rewards	0,13	0,20	0,20	0,00	0,18	1,00											
Log.Pledged.by.backer	0,43	0,01	0,33	0,02	0,19	0,24	1,00										
CategoryRank	0,24	-0,04	0,20	-0,01	0,08	0,05	0,17	1,00									
Log.Friends	0,12	-0,09	0,09	-0,02	0,06	0,09	0,14	0,03	1,00								
Founder.projects	0,17	0,00	0,25	0,01	0,07	0,08	0,10	0,07	0,04	1,00							
Founder.as.backer	0,22	-0,13	0,13	0,00	0,05	-0,02	0,09	0,13	0,01	0,08	1,00						
Founder.sites	0,14	0,02	0,14	0,02	0,09	0,14	0,17	0,05	0,16	0,11	0,08	1,00					
Log.Founder.text.length	0,09	0,06	0,09	0,01	0,06	0,11	0,13	0,05	0,19	0,03	0,05	0,24	1,00				
Has.video	0,09	0,11	0,12	0,01	0,09	0,22	0,19	0,01	0,12	0,03	-0,05	0,17	0,13	1,00			
Images	0,10	0,23	0,18	0,01	0,16	0,39	0,20	0,03	0,01	0,08	0,00	0,09	0,04	0,09	1,00		
Text.abstracts	0,05	0,14	0,10	-0,01	0,06	0,23	0,11	0,02	0,03	0,01	-0,03	0,08	0,06	0,08	0,40	1,00	
Log.Text.length	0,06	0,21	0,11	0,02	0,08	0,15	0,11	0,03	-0,01	0,05	0,03	0,07	0,09	0,04	0,28	0,33	1,00

Source: Author's calculations

Appendix D. Success rate according to factors

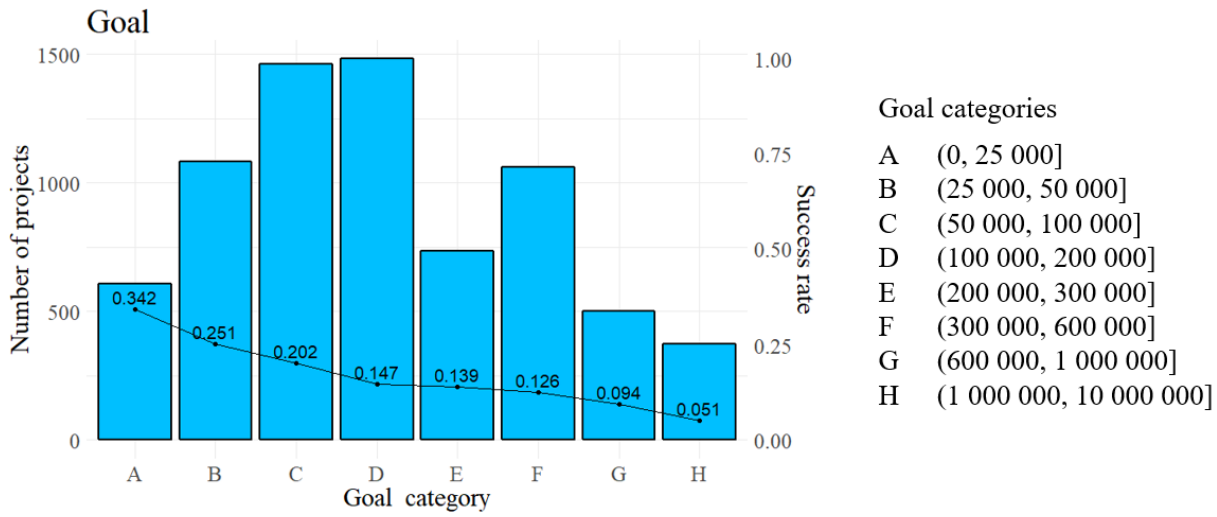

Figure 18.1 Success rate according to Goal

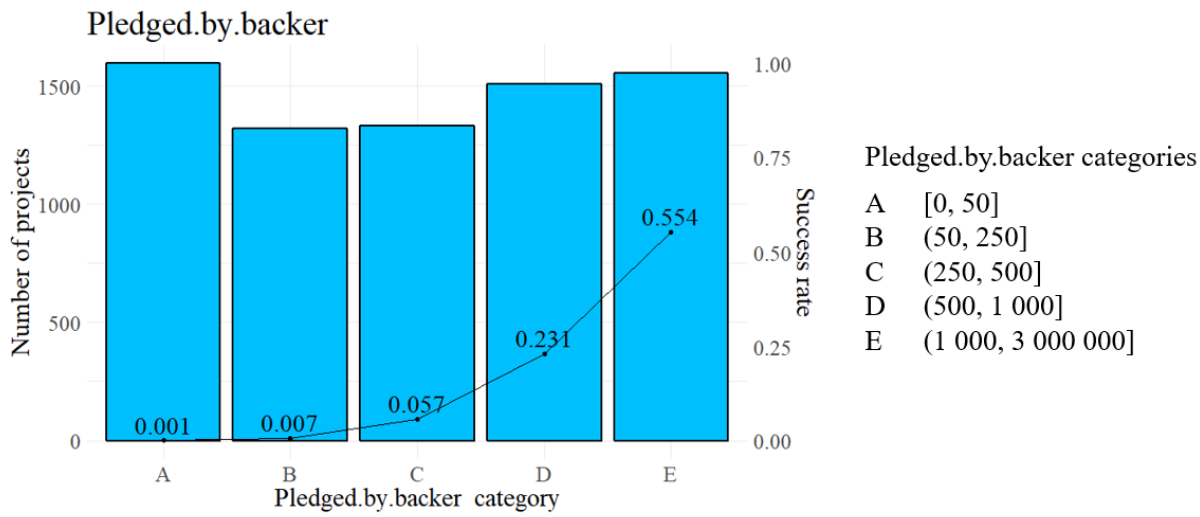

Figure 18.2 Success rate according to Pledged.by.backer

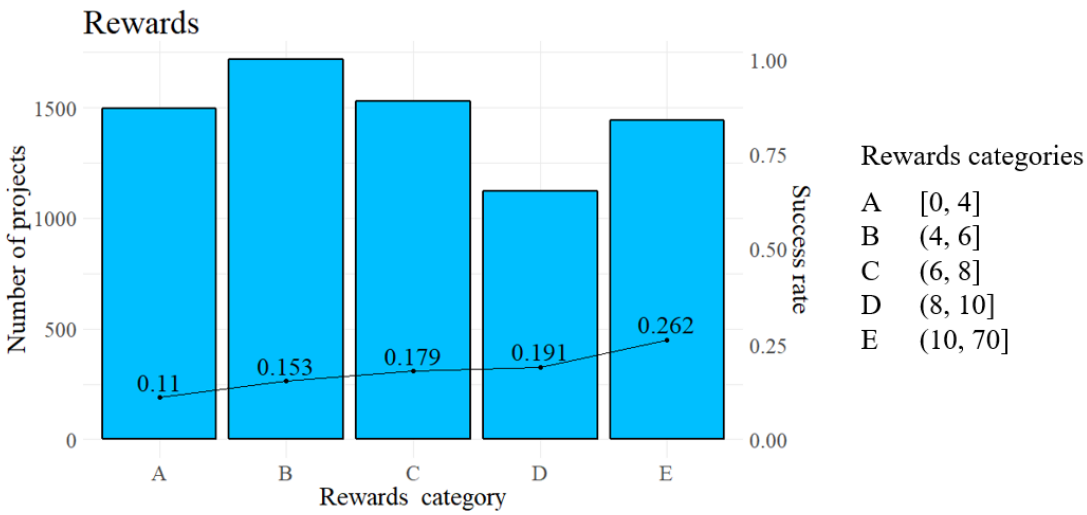

Figure 18.3 Success rate according to Rewards

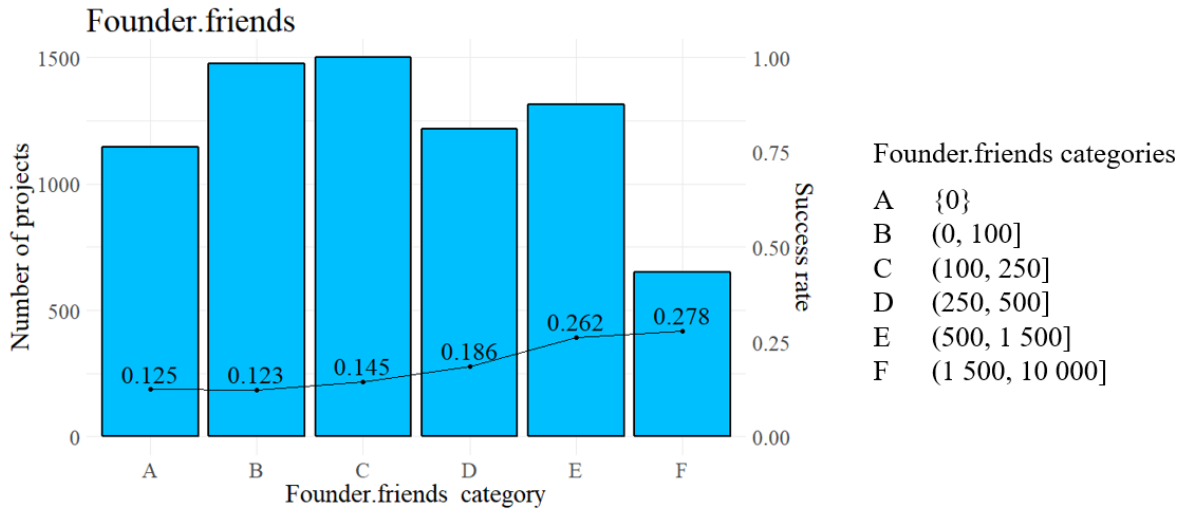

Figure 18.4 Success rate according to Founder.friends

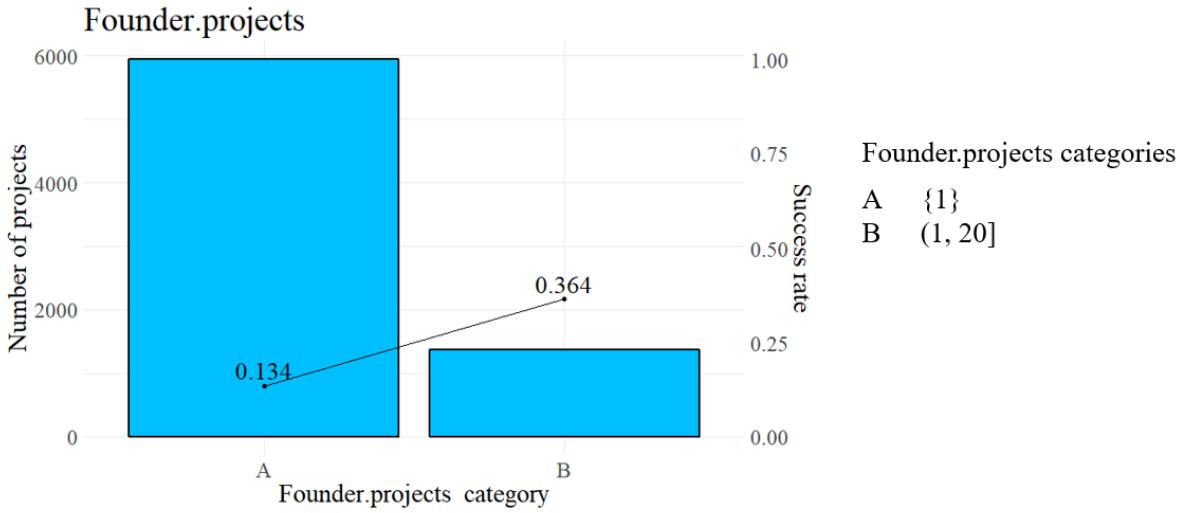

Figure 18.5 Success rate according to Founder.as.backer

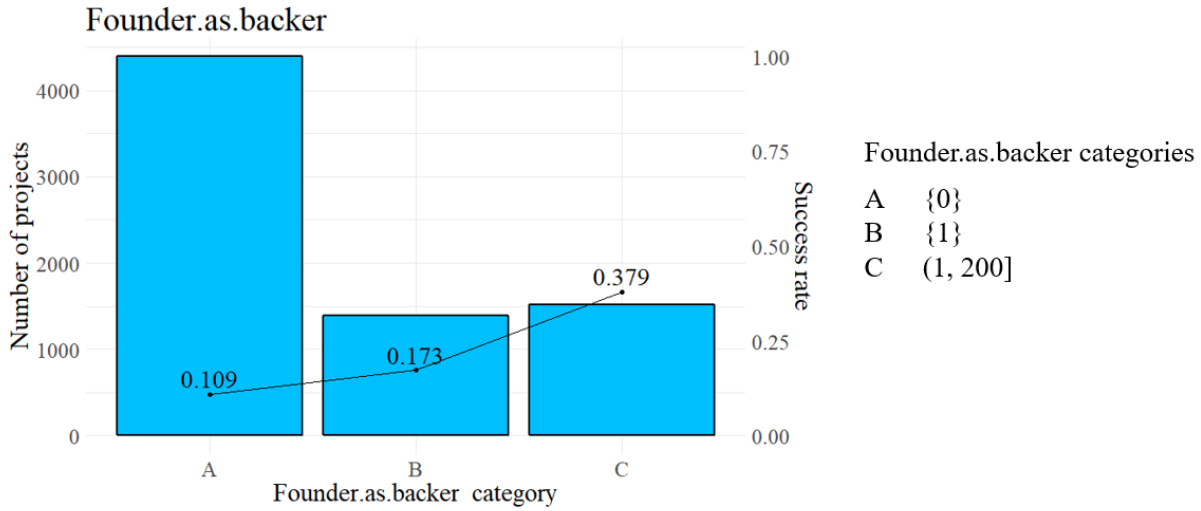

Figure 18.6 Success rate according to Founder.projects

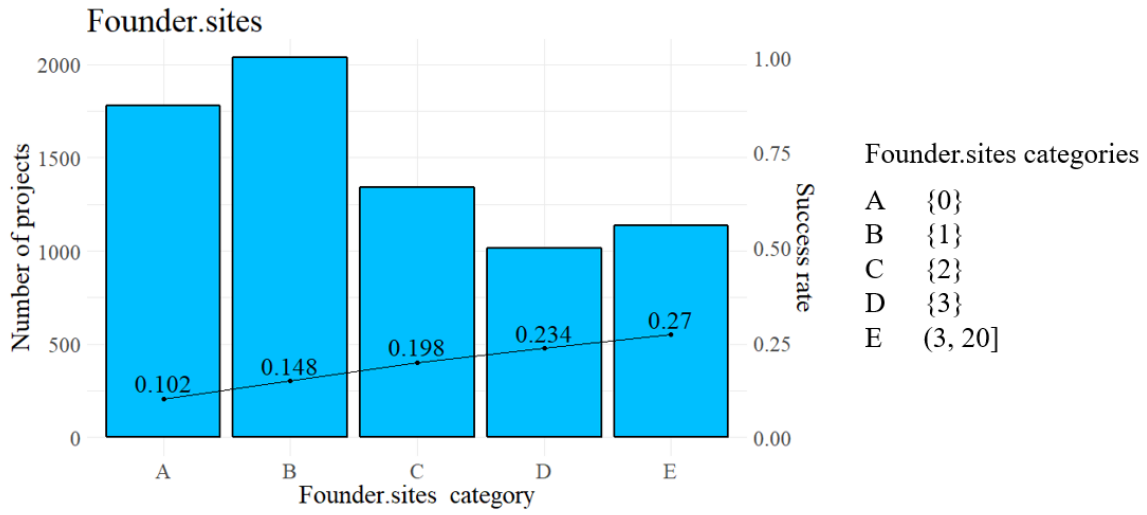

Figure 18.7 Success rate according to Founder.sites

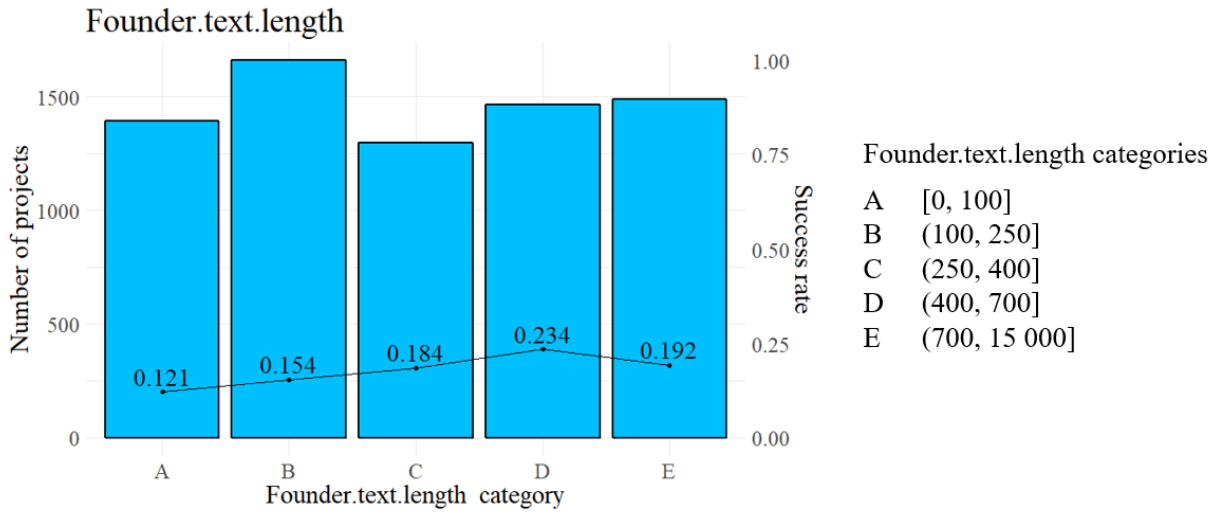

Figure 18.8 Success rate according to Founder.text.length

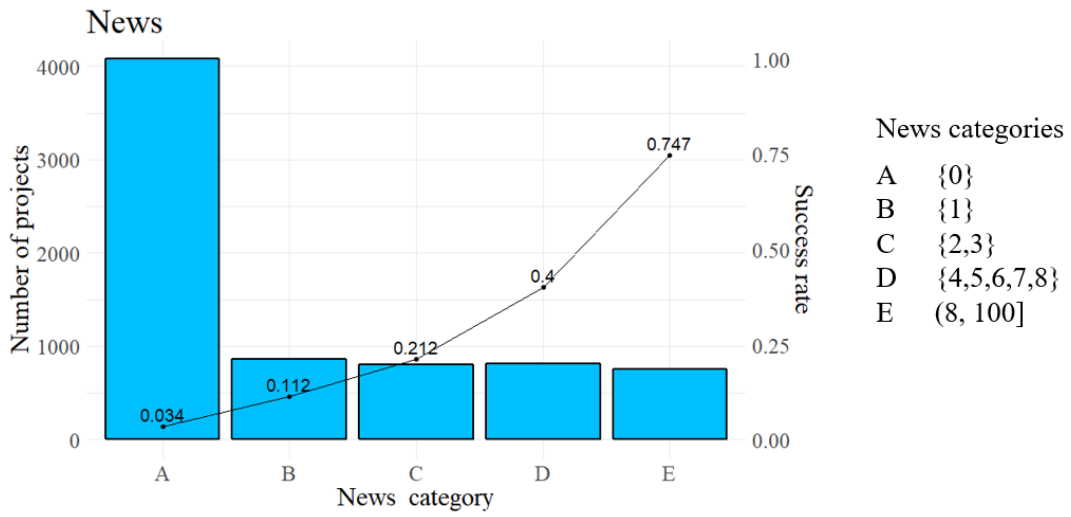

Figure 18.9 Success rate according to News

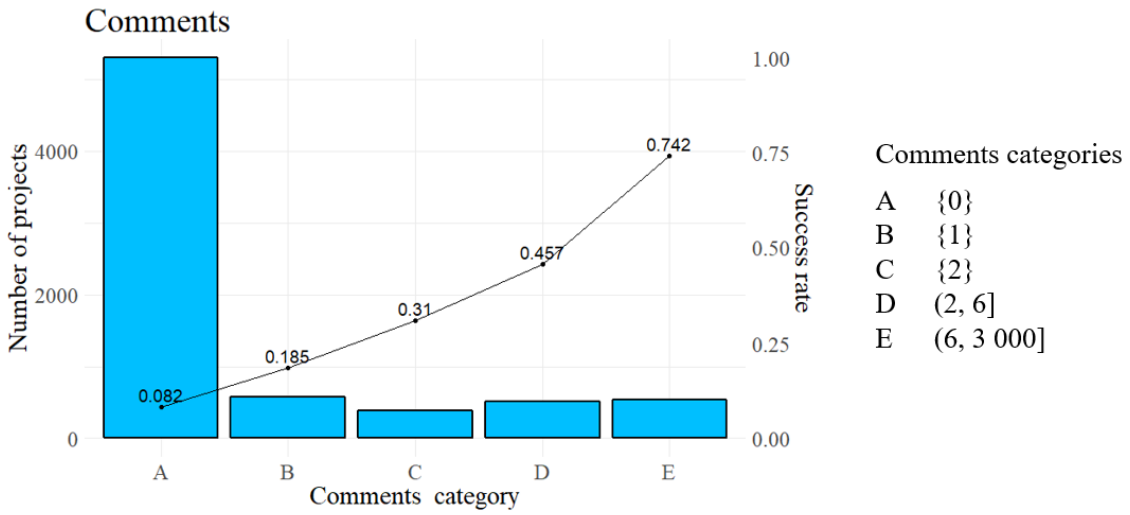

Figure 18.10 Success rate according to Comments

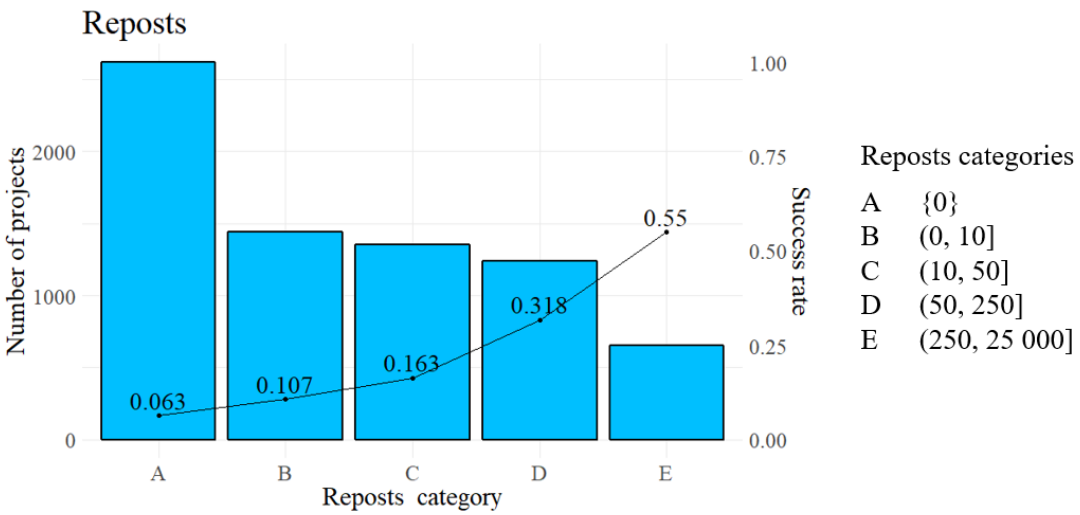

Figure 18.11 Success rate according to Reposts

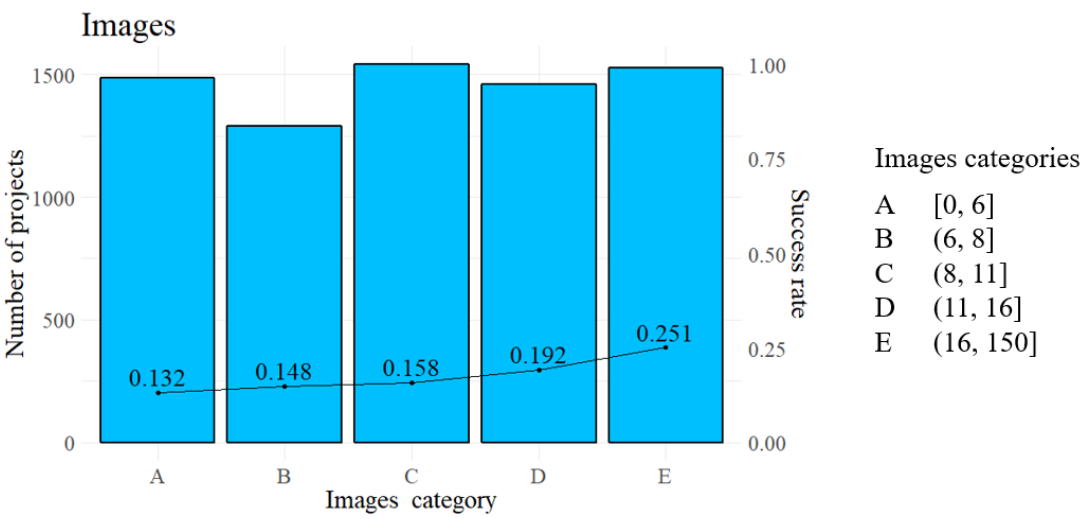

Figure 18.12 Success rate according to Images

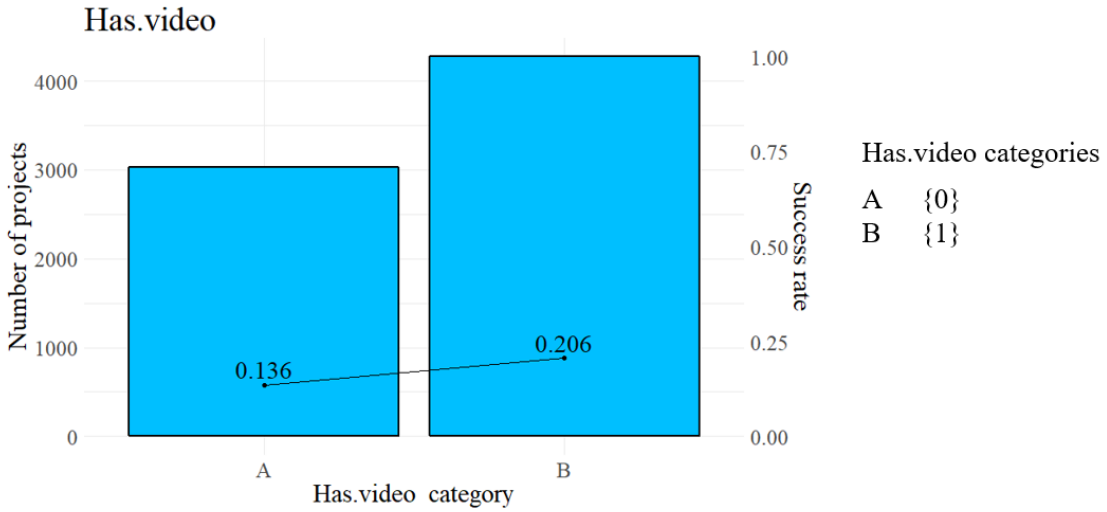

Figure 18.13 Success rate according to Has.Video

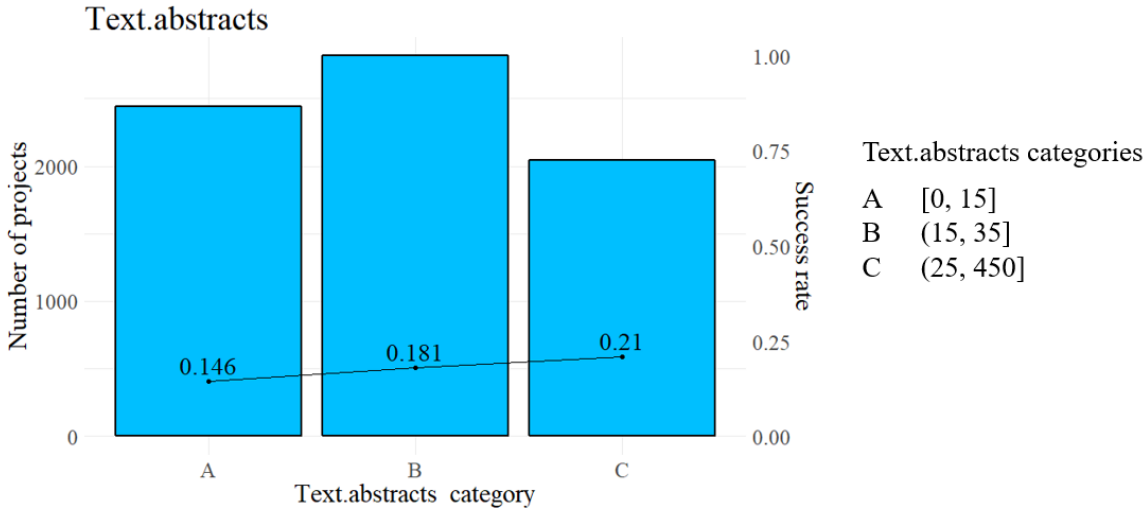

Figure 18.14 Success rate according to Text.length

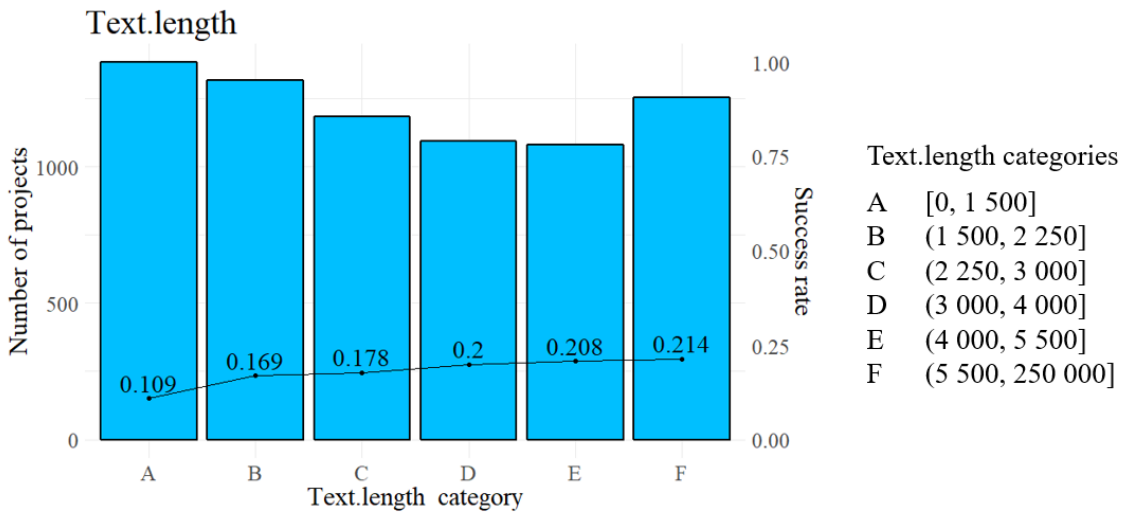

Figure 18.15 Success rate according to Text.abstracts

Appendix E. Success rate according to category

Category	Success rate
Table games	40%
Jazz music	40%
Illustrations	39%
Graphical design	37%
Society	31%
Children literature	30%
Journalizm	28%
Rock music	27%
Comics	26%
Illustrated publications	24%
Theater	23%
#WOWMOSCOW	23%
Documentaries	23%
Public art	23%
Scientific and popular science literature	21%
Publications	21%
Food	21%
Fiction	21%
Poetry	20%
Sculpture	19%
Events	18%
Didgital art	18%
Other Music	17%
Design	17%
Filmimg	17%
Pop-music	16%
Sport	16%
Industrial design	16%
Painting	15%
Short films	15%

Category	Success rate
Animation	15%
Choreography	14%
Films and videos	14%
Indi-rock	12%
Show	11%
Conceptual art	11%
Video clip	11%
Sporting	11%
Hand-made	10%
Entertaining	10%
Art	9%
Technology	9%
Other games	9%
Electronic music	9%
Photos	9%
Equipment	9%
Videogames	8%
Software	7%
TV series	6%
Hip-hop music	5%
Fasion	5%
Periodicals	5%
Business	2%
Classical music	0%

Source: Author's calculations

Appendix F. Success rate according to city

City	Number of projects	Success rate
Irkutsk, Russia	76	42%
Saratov, Russia	76	21%
Novosibirsk, Russia	152	20%
Yekaterinburg, Russia	204	20%
Saint-Petersburg, Russia	926	19%
Moscow, Russia	2809	18%
Omsk, Russia	67	18%
Voronezh, Russia	62	16%
Kazan, Russia	101	16%
Volgograd, Russia	66	15%
Tomsk, Russia	60	15%
Nizhny Novgorod, Russia	81	15%
Chelyabinsk, Russia	106	13%
Perm, Russia	69	13%
Rostov-on-Don, Russia	86	12%
Krasnoyarsk, Russia	87	11%
Samara, Russia	80	10%
Kaliningrad, Russia	52	10%
Krasnodar, Russia	107	9%
Others*	2326	15%

** Others includes cities with less than 50 launched projects*

Source: Author's calculations

Appendix G. Main program on R

```
library(dplyr)
library(ggplot2)
library(AUC)
library(reshape)
library(gridExtra)
library(scales)
library(woe)
library(ROCR)
library(rpart)

setwd("C:/Users/Анечка/Documents/11 ВШМ/ДИПЛОМ/R studio")

### Reading the data
boomstarter <- read.csv('DATA TRANSFORMED.csv', sep=',')

# GRAPH
graph <- function(temp, v){

  temp$rank <- temp$GOOD/temp$TOTAL
  temp$value <- temp$TOTAL

  AXIS1_MIN = 0
  AXIS1_MAX = max(temp$value)

  scale_to_value1 <- function(values) (values*AXIS1_MAX)
  scale_to_value2 <- function(values) (values/AXIS1_MAX)

  g <- ggplot() +
    geom_bar(aes(x = cat, y = value, name = "Success rate"), data = temp,
stat="identity",colour="black", fill = "deepskyblue",size=1) +
    geom_line(aes(x = cat, y = scale_to_value1(rank),  group=1), data = temp,
size=0.7) +
    geom_point(aes(x = cat, y = scale_to_value1(rank)), data = temp,
size=1.3) +
    scale_y_continuous(limits=c(0, max(temp$value)),
                       sec.axis = sec_axis( ~ scale_to_value2(.), name =
"Success rate")) +
    xlab(paste(v,' category'))+ylab('Number of projects')+ ggtitle(v)
+theme_minimal() +
    geom_text(aes(x = cat, y = scale_to_value1(rank), label = round(rank,
digits = 3)), data = temp, vjust=-0.4, color="black", size=7, family="serif")
+
    theme(plot.title = element_text(size = 28, family="serif"),
          axis.text=element_text(size=18, family="serif"),
          axis.title=element_text(size=22, family="serif"),
          text=element_text(family="serif"))

  return(g)
}

### DATA PREPARATION
dt <- boomstarter
dt <- data.frame(dt)
dt$Images <- dt$Imagies
dt$Founder.friends <- dt$Friends
temp_new <- dt$Founder.as.backer
dt$Founder.as.backer <- dt$Founder.projects
dt$Founder.projects <- temp_new
dt$Text.abstracts <- dt$Paragraph
```

```
dt$Pledged.by.backer <- sapply(1:nrow(dt), function(x)
{ifelse(dt$Backers[x]==0,0,dt$Pledged[x]/dt$Backers[x])})
dt$Reposts <- sapply(1:nrow(dt), function(x) {sum(dt$vk[x],dt$fb[x])})

# DATA
dt1 <- list()
dt2 <- list()
dt3 <- list()
dt1$State <- dt$State
dt2$State <- dt$State
dt3$State <- dt$State
dt1$ID <- dt$ID
dt2$ID <- dt$ID
dt3$ID <- dt$ID
dt1 <- data.frame(dt1)
dt2 <- data.frame(dt2)
dt3 <- data.frame(dt3)

### VARIABLES TRANSFORMATION
Categories <- list()
SFA_tables <- list()
SFA_sum <- list()
SFA_roc <- list()
test_auc <- list()
test_gini <- list()

### BREAKS
Breaks <- list()
My_levels <- list()
Graphs <- list()

Breaks$Goal <-
c(0,25000,50000,100000,200000,300000,600000,1000000,15000000000)
My_levels$Goal <- c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H')

Breaks$News <- c(-1,0,1,3,8,100)
My_levels$News <- c('A', 'B', 'C', 'D','E')

Breaks$Comments <- c(-1,0,1,2,6,30000)
My_levels$Comments <- c('A', 'B', 'C', 'D','E')

Breaks$Reposts <- c(-1,0,10,50,250,100000)
My_levels$Reposts <- c('A', 'B', 'C', 'D','E')
max(dt$Reposts)
Breaks$Founder.projects <- c(-1, 1,20)
My_levels$Founder.projects <- c('A', 'B')

Breaks$Founder.as.backer <- c(-1,0,1,200)
My_levels$Founder.as.backer <- c('A', 'B', 'C')

Breaks$Founder.friends <- c(-1,0,100,250,500,1500,10000)
My_levels$Founder.friends <- c('A', 'B', 'C', 'D', 'E', 'F')

Breaks$Founder.sites <- c(-1,0,1,2,3,21)
My_levels$Founder.sites <- c('A', 'B', 'C', 'D','E')

Breaks$Founder.text.length <- c(-1,100,250,400,700,15000)
My_levels$Founder.text.length <- c('A', 'B', 'C', 'D','E')

Breaks$Has.video <- c(-1,0,1)
My_levels$Has.video <- c('A', 'B')

Breaks$Images <- c(-1,6,8,11,16,150)
```

```
My_levels$Images <- c('A', 'B', 'C', 'D', 'E')

Breaks$Text.length <- c(-1,1500,2250,3000,4000,5500,250000)
My_levels$Text.length <- c('A', 'B', 'C', 'D','E', 'F')

Breaks$Text.abstracts <- c(-1,15,35,450)
My_levels$Text.abstracts <- c('A', 'B', 'C')

Breaks$Rewards <- c(-1,4,6,8,10,70)
My_levels$Rewards <- c('A', 'B', 'C', 'D','E')

Breaks$Pledged.by.backer <- c(-1,50,250,500,1000,3000000)
My_levels$Pledged.by.backer <- c('A', 'B', 'C', 'D','E')


### WOE TRANSFORMATION
variables <- c('Goal', 'News', 'Comments', 'Founder.friends',
"Founder.projects",
               "Founder.as.backer", "Founder.sites", "Has.video", "Imagies",
"Text.length", "Text.abstracts")
for (v in variables){
  # Add variable and cut it
  dt1[[v]] <- dt[[v]]
  dt2[[v]] <- cut(dt[[v]], Breaks[[v]])

  # Get woe
  Categories[[v]] <- woe(Data = dt2, v, F, "State", length(Breaks[[v]]), Bad
= 0, Good = 1)
  Categories[[v]]$cat <- My_levels[[v]]
  Categories[[v]]$cat <-  factor(Categories[[v]]$cat, levels =
My_levels[[v]], ordered = T)

  # Graph
  Graphs[[v]] <- graph(Categories[[v]], v)

  # Write woe in dt3
  dt3[[v]] <- dt2[[v]]
  dt3[[v]] <- lapply(dt3[[v]], as.character)
  for (i in c(1:length(Categories[[v]]$BIN))) {
    category <- Categories[[v]]$BIN[i]
    woe <- Categories[[v]]$WOE[i]
    dt3[[v]][dt3[[v]] == category] <- rep(woe, length(dt3[[v]][dt3[[v]] ==
category]))
  }
  dt3[[v]] <- unlist(dt3[[v]], recursive=FALSE)
}

for (v in variables){
  print(Graphs[[v]])
}

for(v in variables){

  ### Splitting the dataset on train and test smaller datasets
  smp_size <- floor(0.75 * nrow(dt3))
  ### Set the seed to make calculations reproductible
  set.seed(1)
  ### Splitting the dataset
  train_ind <- sample(seq_len(nrow(dt3)), size = smp_size)
  train <- dt3[train_ind,]
  test <- dt3[-train_ind,]

  train$f <- train[[v]]
```

```
  test$f <- test[[v]]

  ### Fitting the model
  model_log <- glm(formula = State ~ f,
                   family = binomial(link = 'logit'),
                   data = train)
  SFA_tables[[v]] <- summary(model_log)$coefficients

  ### Prediction on train dataset
  pr_test<-predict(model_log, test)
  test_auc[[v]]<-auc(roc(pr_test,factor(test$State)))
  test_gini[[v]] <- 2*test_auc[[v]] - 1

  SFA_roc <- roc(pr_test,factor(test$State))
}

SFA_sum <- data.frame()
for (i in c(1:length(variables))){
  v <- variables[i]
  SFA_sum[i,"Variable"] <- variables[i]
  SFA_sum[i,"Coefficient"] <- SFA_tables[[v]][2,1]
  SFA_sum[i,"p_value"] <- SFA_tables[[v]][2,4]
  SFA_sum[i,"AUC"] <- test_auc[[v]]
  SFA_sum[i,"GINI"] <- test_gini[[v]]
}
SFA_sum

#write.csv(SFA_sum, "SFA_sum.csv")

for(v in variables){

  ### Splitting the dataset on train and test smaller datasets
  smp_size <- floor(0.75 * nrow(dt1))
  ### Set the seed to make calculations reproductible
  set.seed(1)
  ### Splitting the dataset
  train_ind <- sample(seq_len(nrow(dt1)), size = smp_size)
  train <- dt1[train_ind,]
  test <- dt1[-train_ind,]

  train$f <- train[[v]]
  test$f <- test[[v]]

  ### Fitting the model
  model_log <- glm(formula = State ~ f,
                   family = binomial(link = 'logit'),
                   data = train)
  SFA_tables[[v]] <- summary(model_log)$coefficients

  ### Prediction on train dataset
  pr_test<-predict(model_log, test)
  test_auc[[v]]<-auc(roc(pr_test,factor(test$State)))
  test_gini[[v]] <- 2*test_auc[[v]] - 1

  SFA_roc <- roc(pr_test,factor(test$State))
}

SFA_sum1 <- data.frame()
for (i in c(1:length(variables))){
  v <- variables[i]
  SFA_sum1[i,"Variable"] <- variables[i]
  SFA_sum1[i,"Coefficient"] <- SFA_tables[[v]][2,1]
  SFA_sum1[i,"p_value"] <- SFA_tables[[v]][2,4]
```

```
  SFA_sum1[i,"AUC"] <- test_auc[[v]]
  SFA_sum1[i,"GINI"] <- test_gini[[v]]
}
SFA_sum1



# Models building
### Reading the data
boomstarter <- read.csv('DATA TRANSFORMED.csv', sep=',')

### DATA PREPARATION
dt <- boomstarter
dt <- data.frame(dt)
dt <- dt[complete.cases(dt), ]

#dt <- dt[dt$End.Year>2014,]

# DATA
dt1 <- list()
dt2 <- list()
dt3 <- list()
dt1$State <- dt$State
dt2$State <- dt$State
dt3$State <- dt$State
dt1 <- data.frame(dt1)
dt2 <- data.frame(dt2)
dt3 <- data.frame(dt3)

### VARIABLES TRANSFORMATION
Categories <- list()
SFA_tables <- list()
SFA_sum <- list()
SFA_roc <- list()
test_auc <- list()
test_gini <- list()

dt$Pledged.by.backer <- sapply(1:nrow(dt), function(x)
{ifelse(dt$Backers[x]==0,0,dt$Pledged[x]/dt$Backers[x])})
dt$Reposts <- sapply(1:nrow(dt), function(x) {sum(dt$vk[x],dt$fb[x])})
dt$Text.abstracts <- dt$Paragraph


# MODEL 1
variables0 <- c('Goal', "Rewards", "CategoryRank",'Friends',
 "Founder.projects", "Founder.as.backer", "Founder.sites",
               "Founder.text.length", "Has.video", "Images",
               "Text.abstracts", "Text.length")

for (v in variables0){
  dt2[[v]] <- dt[[v]]
}

dt0 <- dt2
## Splitting the dataset on train and test smaller datasets
smp_size <- floor(0.75 * nrow(dt0))
### Set the seed to make calculations reproductible
set.seed(1)
### Splitting the dataset
train_ind <- sample(seq_len(nrow(dt0)), size = smp_size)
train_0 <- dt0[train_ind,]
test_0 <- dt0[-train_ind,]
```

```
drops <- c("State")
xtrain_0 <- train_0[ , !(names(train_0) %in% drops)]
xtrain_0 <- as.matrix(xtrain_0)
ytrain_0 <- as.matrix(train_0$State)
xtest_0 <- test_0[ , !(names(test_0) %in% drops)]
xtest_0 <- as.matrix(xtest_0)
ytest_0 <- as.matrix(test_0$State)
ytrain1_0 <- ytrain_0
ytrain1_0 <- sapply(1:nrow(ytrain1_0), function(x)
{ifelse(ytrain_0[x]=='1','success','fail')})
ytest1_0 <- ytest_0
ytest1_0 <- sapply(1:nrow(ytest1_0), function(x)
{ifelse(ytest_0[x]=='1','success','fail')})

xgb_grid_1 <- expand.grid(
  nrounds = c(800),
  max_depth = c(6), #6
  eta = c(0.105), #0.09.0.095
  gamma = c(6) ,
  colsample_bytree = c(0.1),
  min_child_weight = c(12),
  subsample=0.7
)

# pack the training control parameters
xgb_trcontrol_1 = trainControl(
  method = "cv",
  number = 5,
  verboseIter = TRUE,
  returnData = FALSE,
  returnResamp = "all", # save losses across all models
  classProbs = TRUE, # set to TRUE for AUC to be computed
  summaryFunction = twoClassSummary,
  allowParallel = TRUE
)

set.seed(1)
#apply(data.matrix(X_train[,-1]),2,as.numeric),
# train the model for each parameter combination in the grid,
# using CV to evaluate
xgb_tr_1 <- train(
  x = xtrain_0,
  y = ytrain1_0,
  trControl = xgb_trcontrol_1,
  tuneGrid = xgb_grid_1,
  metric="ROC",
  method = "xgbTree" #,weight = w
)

pred0 <- predict(xgb_tr_1$finalModel,xtest_0)
th <- 0.5
test_auc<-auc(roc(1-pred0,factor(ytest_0)))
test_gini <- 2*test_auc - 1
err <- 1-mean(as.numeric(1-pred0 > th) != ytest_0)
pred01 <- ifelse(1-pred0 < th, 0, 1)
f1 <- F1_Score(y_pred = pred01, y_true = ytest_0, positive = "1")
conf_matrix <- table(as.numeric(1-pred0 > th),ytest_0)
sens <- sensitivity(conf_matrix)
spec <- specificity(conf_matrix)

test_gini
err
f1
```

```
sens
spec

# ROC CURVE

plot(roc(1-pred0,factor(ytest_0)))

# MODEL 2
variables2 <- c('Goal', "Rewards", "CategoryRank",
                'Friends', "Founder.projects", "Founder.as.backer",
                "Founder.sites", "Founder.text.length",
                "Has.video", "Imagies", "Text.abstracts", "Text.length",
                'News', 'Comments', "Reposts", 'Pledged.by.backer')

for (v in variables2){
  dt3[[v]] <- dt[[v]]
}

final <- dt3[-train_ind,]

dt0 <- dt3

train <- dt0[train_ind,]
test <- dt0[-train_ind,]

drops <- c("State")
xtrain <- train[ , !(names(train) %in% drops)]
xtrain <- as.matrix(xtrain)
ytrain <- as.matrix(train$State)

xtest <- test[ , !(names(test) %in% drops)]
xtest <- as.matrix(xtest)
ytest <- as.matrix(test$State)

ytrain1 <- ytrain
ytrain1 <- sapply(1:nrow(ytrain1), function(x)
{ifelse(ytrain[x]=='1','success','fail')})

ytest1 <- ytest
ytest1 <- sapply(1:nrow(ytest1), function(x)
{ifelse(ytest[x]=='1','success','fail')})

xgb_grid_1 <- expand.grid(
  nrounds = c(800),
  max_depth = c(6), #6
  eta = c(0.105), #0.09.0.095
  gamma = c(6) ,
  colsample_bytree = c(0.1),
  min_child_weight = c(12),
  subsample=0.7
)

# pack the training control parameters
xgb_trcontrol_1 = trainControl(
  method = "cv",
  number = 5,
  verboseIter = TRUE,
  returnData = FALSE,
  returnResamp = "all", # save losses across all models
  classProbs = TRUE, # set to TRUE for AUC to be computed
  summaryFunction = twoClassSummary,
  allowParallel = TRUE
)
```

```
set.seed(1)
#apply(data.matrix(X_train[,-1]),2,as.numeric),
# train the model for each parameter combination in the grid,
# using CV to evaluate
xgb_tr_2 <- train(
  x = xtrain,
  y = ytrain1,
  trControl = xgb_trcontrol_1,
  tuneGrid = xgb_grid_1,
  metric="ROC",
  method = "xgbTree" # ,  weight = w
)

pred2 <- predict(xgb_tr_2$finalModel,xtest)
th <- 0.50
test_auc<-auc(roc(1-pred2,factor(ytest)))
test_gini <- 2*test_auc - 1
err <- 1-mean(as.numeric(1-pred2 > th) != ytest)
pred21 <- ifelse(1-pred2 < th, 0, 1)
f1 <- F1_Score(y_pred = pred21, y_true = ytest, positive = "1")
conf_matrix<-table(as.numeric(1-pred2 > th),ytest)
sens <- sensitivity(conf_matrix)
spec <- specificity(conf_matrix)

test_gini
err
f1
sens
spec

# ВЗЯЛИ ДАННЫЕ ОДНОГО ПРОЕКТА
project <- as.matrix(test_0[1,!(names(test_0) %in% drops)])

# p1 ПОСЧИТАЛИ ДЛЯ ЭТОГО ПРОЕКТА
round(1-predict(xgb_tr_1$finalModel,project), digits=2)

values1 <- list()
values1[["News"]] <- c(0,2,5,10,25)
values1[["Comments"]] <- c(0,1,2,3,4,5,6,7,8,9,10,11,
                           12,13,14,15,16,17,18,19,20)
values1[["Reposts"]] <- c(0,50,100,200,300,400,500,600,
                          700,800,900,1000)
values1[["Pledged.by.backer"]] <- c(0,200,400,600,800,
                                    1000,1200,1400)

a <- expand.grid(values1[["News"]], values1[["Comments"]],
values1[["Reposts"]],values1[["Pledged.by.backer"]])

colnames(a) <- c('News', 'Comments', "Reposts", 'Pledged.by.backer')


project_all <- list()
l <- nrow(a)
for (v in variables0){
    project_all[[v]] <- rep(project[,v],l)
}
project_all <- data.frame(project_all)

project_all <- cbind(project_all, a)


# p2
```

```
project_pr <- project_all
project_pr$p2 <- round(1-predict(xgb_tr_2$finalModel,
as.matrix(project_all)), digits=2)

project_pr <- project_pr[order(project_pr$p2),]



data_plot <- project_pr[project_pr$News == 2,]



p <- plot_ly(data_plot, x = ~ Comments,
             y = ~ Pledged.by.backer, z = ~ Reposts,
             marker = list(color = ~ p2,
             colorscale = c('#FFE1A1', '#683531'), showscale = TRUE),
             text = ~ paste('Probability of success = ',
                            round(p2, digits = 2)),
             hoverinfo = "x+y+z+text",
             hoverlabel = list(bgcolor  = '#FFE1A1')) %>%
  add_markers() %>%
  layout(scene = list(xaxis = list(title = 'Comments'),
                      yaxis = list(title = 'Pledged.by.backer'),
                      zaxis = list(title = 'Reposts')),
         annotations = list(
           x = 1,
           y = 1,
           text = 'Probability of success',
           showarrow = F
         ))
p
```