CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Does Language Drive the Crowd? Case of Czech Reward-Based Crowdfunding

Master's thesis

Author: Bc. Tereza Hudcová

Study program: Economics and Finance

Supervisor: Mgr. Petr Polák, MSc.

Year of defense: 2020

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 3, 2020

Tereza Hudcová

Abstract

This thesis analyses the biggest reward-based crowdfunding platform in the Czech Republic using textual tools on uniquely collected microdata. The research question investigates which of the attributes of project campaigns (including the language style of project descriptions) have a significant impact on successful funding. Empirical analysis combines results of Bayesian Model Averaging and logistic regression. Results reveal that firstly, language style of project descriptions does not possess any significant predictive power. Secondly, that utilization of a video, size of pledging goal, or the number of contributors have a significant effect on the campaign's success, which is in line with current literature. Thirdly, it has proven to be true that project categorization plays an important role as well. On the contrary, the findings do not imply any causal claims, such as whether those factors persuade contributors to donate money.

JEL Classification G23
Keywords crowdfunding, rewards, success determinants, natural language processing

Author's e-mail t.hudcova@gmail.com
Supervisor's e-mail polakpet@gmail.com

Abstrakt

Tato diplomová práce analyzuje unikátní mikrodata největší reward-based crowdfundingové platformy v České republice s využitím nástrojů textové analýzy. Práce zkoumá, jaké atributy crowdfundingových kampaní (spolu se stylem jazyka v popiscích projektů) výrazně ovlivňují jejich úspěšné zafinancování. Empirická analýza kombinuje výsledky Bayesovského průměrování modelů a logistické regrese. Výsledky odhalují, že styl jazyka v popiscích projektů nemá významnou výpovědní hodnotu. Dále výsledky analýzy indikují, že krátké video, velikost požadované částky nebo počet přispěvatelů u každé z kampaní signifikantně ovlivňují zafinancování projektu, což je v souladu s existující literaturou. Co se dále prokázalo jako pravdivé, je fakt, že kategorie, do které je projekt zařazen, hraje významnou roli. Je však nutné zdůraznit skutečnost, že výsledky neimplikují kauzalitu, tzn. že výše identifikované faktory nestojí za rozhodnutím přispěvatelů danou kampaň financovat.

Klasifikace JEL G23
Klíčová slova crowdfunding, odměny, determinanty úspěchu, natural language processing

E-mail autora t.hudcova@gmail.com
E-mail vedoucího práce polakpet@gmail.com

Acknowledgments

The author is grateful especially to Czech crowdfunding platform HitHit, as it kindly provided the dataset used for the analysis.

In addition, the author truly appreciates the joint collaboration with text analytics company Geneea, not only for the access to the natural language processing software platform free of charge, but also for generous help with programming tools.

A big thank you also belongs to Petr Polák, whose suggestions and comments created an invaluable contribution to the content of this thesis.

Last, but not least, the deepest gratitude goes to author's family and close friends for warm encouragement and constant support during the whole period of studying.

Typeset in LATEXusing the IES Thesis Template.

Bibliographic Record

Hudcová, Tereza: *Does Language Drive the Crowd? Case of Czech Reward-Based Crowdfunding.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 75. Advisor: Mgr. Petr Polák, MSc.

Contents

List of Tables

List of Figures

Acronyms

AI	Artifical Intelligence
BMA	Bayesian Model Averaging
CCAF	Cambridge Centre for Alternative Finance
CF	Crowdfunding
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
NER	Named Entity Recognition
NLP	Natural Language Processing
P2B	Peer-to-business Lending
P2P	Peer-to-peer Lending
PIP	Posterior Inclusion Probability
PMP	Posterior Model Probability
UIP	Unit Information Prior
UK	The United Kingdom
US	The United States of America

Master's Thesis Proposal

Author	Bc. Tereza Hudcová
Supervisor	Mgr. Petr Polák, MSc.
Proposed topic	Does Language Drive the Crowd? Case of Czech Reward-Based Crowdfunding

The idea of crowdfunding can be understood as a subset of the larger concept of crowdsourcing, which enables the enterprises to use the crowd to receive ideas, feedback and solutions in order to develop corporate activities (Belleflamme et al., 2014). Specific feature of crowdfunding stems from the fact that entrepreneurs use an Internet platform to get in touch with like-minded individuals, who are willing to contribute to the venture a small amount of money (Valanciene and Jegeleviciute, 2013). Due to increasing frustration on traditional financial markets, crowdfunding has gained a momentum, particularly for businesses at the initial stages, often experiencing difficulties in attracting sources of capital.

Over the past decades, crowdfunding has become a global phenomenon that significantly impacts not only finance, but also many other fields. As it influences many different areas, the academic research is very extensive and fragmented. Consequently, there are relatively few literature reviews in this field. Thus, there is a big room for summary not only of the current state and future prospects of crowdfunding, but also for overview of different methods and approaches that current researchers adopt.

As stated before, the projects are published on dedicated platforms, together with detailed description of intended action. Any member of the platform has then opportunity to become a project funder by contributing pre-determined amount of money. In exchange for this contribution, s /he is promised a certain form of reward. However, only if the funding goal is reached, i.e. the desired amount of money is collected, the funders are obliged to fullfil agreed funding. On the other hand, meaning that the funding goal is not reached, project cannot be realized.

Overall, the percentage of projects fully financed on crowdfunding platforms is relatively low. Therefore, the determination of what actually affects the likelihood of project being funded, is crucial. As determinants of project's success belong to frequently investigated topics, there will BE a big room for comparison of how results stemming from the Czech Republic differ among various countries, platforms or sample sizes.

Methodology Firstly, in order to provide reader with fundamental characteristics of crowdfunding, a general outline of crowdfunding concept, together with history and recent trends in this field, will be provided. Emphasis will be also put on the state in the Czech Republic. In addition, the most-popular crowdfunding models, as suggested by Stasik et al. (2017), will be listed and described. In order to capture magnitude of this phenomenon, growth rates and amount of funds collected over the past years, will be summarised as well.

As it was stated before, crowdfunding is a phenomenon dispersed across many fields. Consequently, identification of the relevant knowledge is not only difficult but also time-consuming. This is why there exist only few literature reviews (Macht et al. (2015), Moritz et al. (2016) or Stasik et al. (2017)). The complexity of this field also made researches to use different tools and methods for analysis of relevant questions. To get a basic notion about those methods, concise summary of typology of current studies will be provided. This is expected to strengthen theoretical foundations for empirical analysis.

Moving on to the empirical part of this Master's thesis, the main goal is to trace the determinants of successful project funding. The analysis will be performed on unique dataset provided by HitHit, the largest crowdfunding platform in the Czech Republic. This dataset will be further enriched by textual analysis of project descriptions. Based on existing research works, key variables that are expected to have a significant impact on the success of the campaign will be selected. This will be done with help of Bayesian Model Averaging. Afterwards, logistic regression will be estimated, primarily as a robustness check. The dependent variable stands for funding success. Exploring the significance of particular success factors in the funding process creates room for analysing Czech pledgers' motivations.

Expected Contribution To the best of author's knowledge, the Czech crowdfunding scene lacks analysis that captures also textual point of view. Usage of dataset tracing the history of all projects which occurred on HitHit might help to fill this gap. The empirical results are expected to identify the main drivers, that play role in funders' decision-making process. It will be also revealed, whether the language used in projects' description has some predictive power. This is the expected added value of the thesis, as the findings can be further compared with existing results from different regions or works using different sample sizes. Consequently, the distinctions of Czech crowdfunding market can be identified.

All in all, this pioneer work is expected to build foundations for further researchers of Czech crowdfunding platforms. In addition, the results might be highly relevant for Czech project founders, as they can subsequently concentrate on the accuracy and quality of the significant factors already in the stage of campaign creation.

Outline

1. Introduction
2. Concept of Crowdfunding – History, Current Trends, Different Models
3. Literature Review – Typology of Research Works
4. Data Description
5. Theoretical Framework and Empirical Model
6. Discussion of Results, Further Implications
7. Conclusion

Core bibliography

Belleflamme, P., Lambert, T., Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. Journal of Business Venturing, 29(5): 585-609.

Crosetto, P., & Regner, T. (2014). Crowdfunding: Determinants of success and funding dynamics (No. 2014-035). Jena Economic Research Papers.

Greenberg, M. D., Pardo, B., Hariharan, K., & Gerber, E. (2013, April). Crowdfunding support tools: predicting success & failure. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 1815-1820). ACM.

Macht, S. A., Weatherston, J. Academic research on crowdfunders: What's been done and what's to come? Strategic Change 24.2 (2015): 191-205.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal of business venturing, 29(1), 1-16.

Moritz, A., & BLOCK, J. H. (2013). Crowdfunding and crowdinvesting: a review of the literature. SSRN Working Paper.

Moritz, A., & Block, J. H. (2016). Crowdfunding: A literature review and research directions. Crowdfunding in Europe (pp. 25-53). Springer, Cham.

Mudambi, S. M., Schuff, D. (2010). What Makes a Helpful Online Review? Study of Customer review on Amazon.com. MIS Quarterly 34(1), 185-200.

Stasik, A., Wilczydďż˝ska, E. (2017). How do we study crowdfunding? An overview of methods and introduction to new research agenda. Journal of Management and Business Administration. Central Europe 26.1 (2018): 49-78.

Tu, T. T. T., Anh, D. P., & Thu, T. T. H. (2018). Exploring Factors Influencing the Success of Crowdfunding Campaigns of Startups in Vietnam. Accounting and Finance Research, 7(2), 19.

Valanciene, J., Jegeleviciute, S. (2013). Valuation of Crowdfunding: Benefits and Drawbacks. Economics and Management, 18(1): 39-48.

Author Supervisor

Chapter 1

Introduction

A common milestone that many entrepreneurs need to overcome at the initial stage of their business is raising of capital. Most frequently, the financing for the venture is provided by banks, shareholders, angel investors or by venture capital funds. In the recent years, the new form of attracting funds has emerged.

Crowdfunding (also interchangeably referred to as crowdfinancing) has become phenomenon of rising significance. Compared to the traditional forms of financing, crowdfunding is a procedure where individuals or groups have the opportunity to collect small contributions from the community of the large number of the Internet users without the intermediation of financial institution (Mollick 2014). However, the project backers are obliged to pay the promised funding only if the pre-determined funding goal is reached. Taking into account that the share of fully funded actions of such platforms is relatively low (see for example, Koch & Siering (2015)), it should be of great interest of project founders to search for the determinants of successful funding.

Consequently, over the past decade, reward-based crowdfunding has gained undeniable attention from researchers all over the world. So far, there exist several studies examining characteristics of a campaign, which impact successful funding (such as Chan *et al.* (2018)). Next stream aims to understand the dynamics of contributions during the running of a campaign (Kuppuswamy & Bayus (2018) and Mollick (2014)). Academia has also investigated motivations behind backers' contributions (please refer to Zhang & Chen (2019) or Leimeister & Bretschneider (2017)), as well as exploitation of geographical and social reach (Agrawal *et al.* 2015). However, the impact of language on the crowdfunding success still remains to large extent undisclosed, as it is discussed only by few research works (see Sun *et al.* (2016) or Mitra & Gilbert (2014)).

Based on the unique dataset of 2,870 projects from HitHit, the largest platform in the Czech Republic, this thesis investigates, which factors have the decisive impact on successful project funding. Besides standard predictors of project campaign identified by current literature, this thesis investigates the impact of project descriptions' content on successful funding by means of Natural Language Processing (NLP). More specifically, it is analysed, whether sentiment of language used in descriptions has an impact on the overall result of the campaign. It is also inspected, whether the presence of some particular words (most frequently occurring nouns) or phrases (commonly used relations with adjectives or verbs) in project descriptions impacts the probability of success. To the best of author's knowledge, this is a pioneer approach applied in the Czech reward-based crowdfunding environment, as vast majority of extant studies concentrates on common project attributes like monetary goal, length of a campaign, number of rewards or contributors. Consequently, the impact of Czech language on the crowdfunding success remains to large extent undisclosed, taking into account not only the smaller sample size, but also the fact that Czech platform HitHit uses different policies than US-based Kickstarter, the biggest platform in the world.

As there are many explanatory variables to consider, the thesis uses Bayesian Model Averaging method (so-called BMA), as it deals with arising model uncertainty. The main objective of this technique is to identify, which regressors perceive strong explanatory power in terms of campaign success. Afterwards, logistic regressions with variables identified by BMA is estimated as a robustness check.

What is more, most of the existing studies on determinants that foster funding success have been conducted by using data from well-emerged crowdfunding markets – US, UK or Western Europe. Thus, this academic work contributes to the existing research by provision of coherent summary not only of the biggest reward-based crowdfunding platform in the Czech Republic, but also depicts Czech crowdfunding environment as a whole.

The rest of this thesis is structured as follows. Firstly, the concept of crowdfunding together with brief insight into its history is provided in Chapter 2. Afterwards, light is shed not only on business models that have developed over time, but also on the current state of this field worldwide as well as in the Czech Republic. Chapter 3 presents concise literature review, primarily listing the key issues addressed in the category of reward-based crowdfunding, as it lies in the scope of this paper. In addition, it also summarizes conducted research works discussing the role of language in this field. Chapter 4 focuses on descriptive statistics of analysed dataset. Theoretical framework, applied methodology as well as estimated results can be found in Chapter 5. Consequently, Chapter 6 discusses the key findings, limitations of adopted technique as well as suggestions for future researchers. Chapter 7 concludes.

Chapter 2

Concept of Crowdfunding and Natural Language Processing

2.1 Crowdfunding

2.1.1 Historical Overview and Current Trends

In order to better understand the concept of Crowdfunding (CF), it is essential to gain the basic knowledge not only about its development over time, but also about the recent trends and challenges that are characteristic for this field.

Many specialists and researchers claim that CF is not entirely a new concept. The idea, that individual or entrepreneur collects money from large number of investors, has been historically realised several times. For example, in 1985, funds of 1,400 investors were raised in order to film the Crocodile Dundee movie (Guilliatt 1988). One may also think of charity organisations, which heavily rely on small contributions from donors (Fleming & Sorenson 2016).

A turning point that is essential for the emergence into the current state of CF, is the enabling of widespread access to the Internet. In addition, as stated by Ordanini *et al.* (2011), even more dynamic growth occurred after the introduction of so-called Web 2.0. Both previously mentioned factors heavily contributed to the rise of dedicated platforms.

Those platforms act as a convenient intermediaries, as they significantly decrease the costs of running a campaign. This is done by exploitation of geographical as well as social reach of the Internet to connect project founders with dispersed backers. Introduction of secure online payments together with increased usage of online debit/credit cards played an important role as well.

ArtistShare[1], established in 2001 in the US, is widely recognised as the very first online crowdfunding platform in the world. It connects artists and fans – it allows them to contribute to new creative artistic works of their choice. Its model became a blueprint for more popular platforms, such as Kickstarter, Wefunder, PledgeMusic or IndieGoGo[2].

A summary of current state of crowdfunding is a formidable task to do. Even though there exist several resources, such as various statistics, articles, devoted websites or academic papers, the data is often not comparable (across regions, types of crowdfunding or time), is incomplete or does not stem from reliable information source.

One of the possibly valid streams can be considered Cambridge Centre for Alternative Finance (CCAF). This institution was created in order to examine, how the field of alternative finance is evolving, as it faces several challenges. For instance, digitalisation or other social or economic shifts. Since 2015, it has already published the total of 4 reports capturing the current situation on the European alternative finance market.

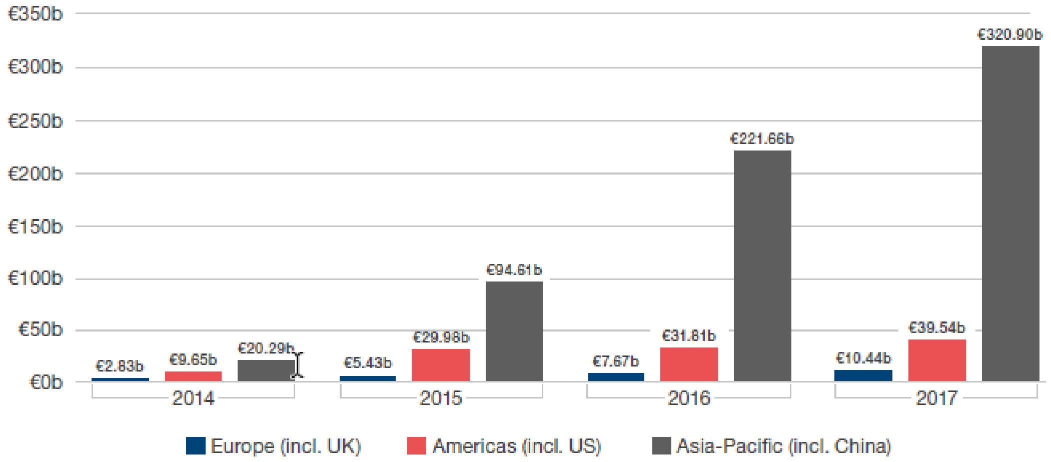
The latest study published in 2019 collects data of 269 crowdfunding, P2P lending and other alternative finance platforms in 45 European countries. It provides a coherent summary of these platforms' operations reported in 2017.

In order to put European numbers into broader context, CCAF also conducts a global benchmarking research programme, which tracks transactions on this market at a global scale. Figure 2.1 shows the evolution of alternative finance markets in the world. The Asia-Pacific region can be considered as an ultimate leader. It is important to emphasize, that Chinese market accounts for 99% of the volume of this region. Therefore, the growth can be attributed solely to platforms operating in China. Compared to the rest of the world, European region is obviously the smallest one.

[1] http://www.artistshare.com

[2] http://www.kickstarter.com, http://www.wefunder.com, http://www.pledgemusic.com, http://www.indiegogo.com

Figure 2.1: Online Alternative Finance Market Volumes in 2017 Globally



Notes: The bar charts depict the development of alternative finance market worldwide in years 2014, 2015, 2016 and 2017. Recall that Chinese platforms comprise 99% of the Asia-Pacific region.

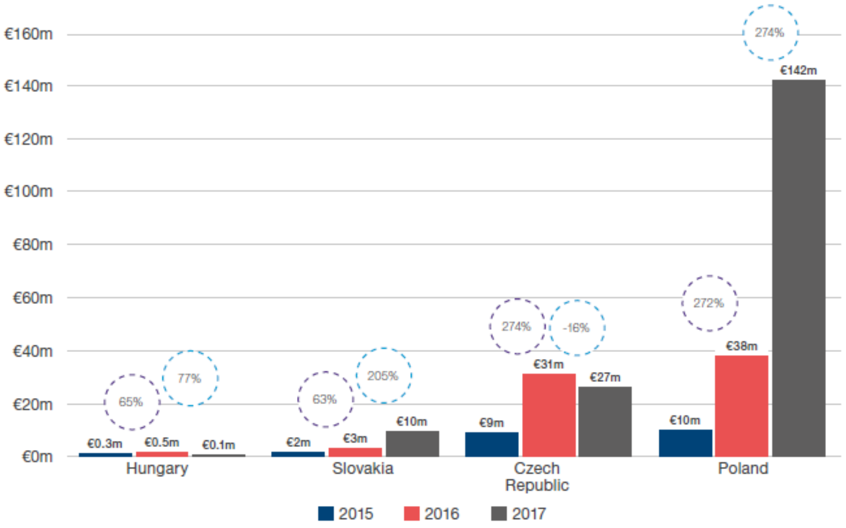
Source: (Ziegler *et al.* 2018)

Now it will be proceeded to the European online alternative finance industry. Based on the CCAF findings, when encompassing the UK, the market grew by 36% to reach the amount of EUR 10.44 billion in year 2017. Even though the United Kingdom still perceives the leading position in terms of the size of individual market, its market share declined from 73% in 2016 to 68% in 2017. When omitting the UK, the European alternative finance market increased by 63% (from EUR 2 billion to EUR 3.4 billion) over the respective period. The rate of growth is considerably smaller compared to previous years – for example, 102% in 2016 (Ziegler *et al.* 2018).

It is also important to emphasize, that when excluding the UK, P2P consumer lending market accounts for the greatest market share, as it represents almost half of the size of this market (namely 41%). On the other hand, reward-based crowdfunding, which is in the main scope of this paper, accounts for only minor fraction of the European alternative finance (5%).

The Alternative Finance Industry Report also provides more refined picture of Eastern Europe. The cluster of Eastern Europe consists of Poland, the Czech Republic, Slovakia and Hungary. This regional market grew from EUR 70 million in 2016 to EUR 179 million in 2017, experiencing growth rate of 153% (Ziegler *et al.* 2018).

Figure 2.2: Online Alternative Finance Market Volumes 2017 in Eastern Europe



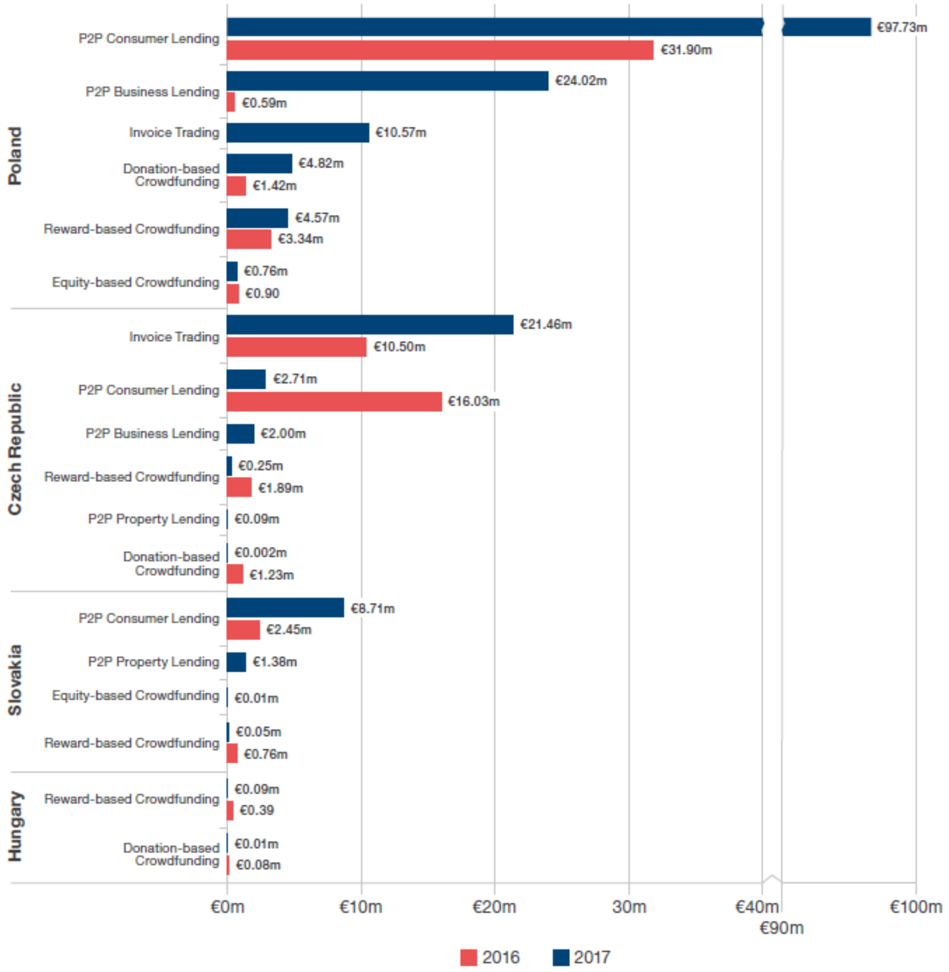
Notes: The bar charts demonstrate the growth of alternative finance market in years 2015, 2016 and 2017, respectively. The percentage in dashed bubbles indicates the market growth from 2015 to 2016 and from 2016 to 2017 for given country.

Source: (Ziegler *et al.* 2018)

From Figure 2.2 it can be concluded, that Poland exhibited significant growth in 2017, as its market almost tripled in size. Note that Czech alternative finance market declined over the respective period. However, this may not reflect an actual volume of the market itself, as in 2017, there were fewer platforms participating in the survey.

Figure 2.3 demonstrates the detailed breakdown of segments of Alternative Finance Industry, concentrating on the cluster of Eastern Europe.

Figure 2.3: Online Alternative Finance Market: 2017 Breakdown of Eastern Europe



Source: (Ziegler *et al.* 2018)

There is no doubt that P2P lending is still the leading segment within the industry in selected countries. However, reward-based crowdfunding platforms (being the analysed object of this thesis) have definitely gained significant importance as well. Note that the Czech market experienced rather declining trend in 2017, which was very likely caused by fewer number of platforms which participated in the survey, as already mentioned before.

From the previous breakdowns it is apparent, that rapid growth and fast development of this field caused the split into different types of crowdfunding. To describe this variation in more detail, the division into specific categories based on academic works as well as on typology of already existing platforms can be found in the next subsection.

2.1.2 Crowdfunding Business Models

The following list presents the most common crowdfunding business models, as suggested by many research papers, for instance, Kshetri (2015), Fleming & Sorenson (2016), Kuti & Madarász (2014) or Alegre & Moleskis (2016). There exist four main categories, namely:

- Donation-based crowdfunding
- Reward-based crowdfunding
- Debt- or credit-based crowdfunding
- Equity-based crowdfunding

Donation-based crowdfunding takes place when projects of humanitarian, artistic or personal nature are funded. In this specific form, contributors do not expect any financial returns or other rewards of non-financial nature, so the concept heavily relies on voluntary contributions. As stated by Belleflamme *et al.* (2015), contributors to donation-based projects can be rather viewed as philantropists. Some examples of particular platforms include GoFundMe, Givology or Kopernik. New donation-based platform Donio[3] has recently gained attention in the Czech Republic.

Reward-based crowdfunding is characteristic by project funders who are (similarly as in the previous case) not interested in financial return in exchange for their contribution. The platform offers individuals an opportunity to donate to project in exchange for some product. Contributors effectively pre-purchase the product, which significantly reduces risk from the project founder's perspective. Consequently, the funders of successful projects are awarded by tangible, but non-financial benefits (Kuppuswamy & Bayus 2018). This is the case of so-called *all-or-nothing* principle, which is more common. Here, project funder receives the collected amount of money only if the pre-defined target amount is reached. Otherwise, the money is returned to contributors. The second type is *keep-it-all* model, whose occurrence is not that frequent. In this case, project initiator can keep the collected amount of money, even though the goal was not reached.

A common feature of donation-based and reward-based category is that there is a substantial taste heterogeneity of their funders (Belleflamme *et al.*

[3] http:/www.donio.cz

2015), compared to the following two business models. Popular platforms of this kind include Kickstarter, Polak Potrafi or IndieGoGo. Analysed Czech platform HitHit also belongs to this category.

Debt- or credit-based crowdfunding is commonly referred to as Peer-to-peer Lending (P2P) or Peer-to-business Lending (P2B). Bachmann *et al.* (2011) defines this breed of market interaction as unsecured loans between lenders and borrowers on particular online platforms, who act as middlemen, while intermediation of financial institution is not required. Well-established intermediaries include European platforms, such as Zopa, Twino, Lending Club or US-based Prosper. In the Czech Republic, the most popular P2P lending provider is platform Zonky.

Equity-based crowdfunding platforms' funders can be described as individuals, who need to assess the risk of undertaken investment carefully. Some streams of academic literature (see Belleflamme *et al.* (2015), for example) distinguish two types of this model. First, equity-based crowdfunding, where investors may acquire equity stakes in corporations for a small amount of money. In the second type, so-called loyalty-based crowdfunding, a fraction of earned profit is offered to participating investors. Some examples include French equity-based platform Smart Angels or CrowdCube from the United Kingdom.

Even though previously listed models belong among those of being well-established, there exist many more other smaller categories, that are often mixed. Here is a proposition of some distinct examples, as suggested by Stasik & Wilczyńska (2018):

- **Pre-purchase crowdfunding** is understood as a subset of reward-based concept. Project funders are promised to receive the product that entrepreneur is making, for example, a music album.

- **Litigation crowdfunding** takes place when the third party which is not linked to the lawsuit offer financial resources to accuser in exchange for certain percentage of financial compensation from the verdict. This category can be found mainly in the UK, the US or Australia.

- **Invoice crowdfunding** is a commonly used model where enterprises sell outstanding receivables or invoices (either separately or bundled together) to pool of investors through online auctions (EC 2016).

As reward-based form of crowdfunding is at the centre of this thesis' research focus, the emergence of this particular category in the Czech Republic will be discussed in the proceeding subsection. This will be followed by comprehensive literature review in Chapter 3, summarising the key trends identified by current researchers.

2.1.3 Reward-based Crowdfunding in the Czech Republic

Fondomat (established in August 2011) can be considered as the very first reward-based platform in the Czech Republic. However, its management moved quickly from Prague to London and the company ceased to exist already in 2015.

Another platform, namely NakopniMě, came into being in 2012 and also belongs to the pioneer companies of Czech crowdfunding scene. This website incorporated standard model of reward-based crowdfunding, where individuals or businesses could have launched projects of any kind and where funds are received if and only if the target amount is reached. As well as the case of Fondomat, this platform's webpage is currently non-existing.

A distinctive idea can be found behind origination of platform Sportstarter (established in 2015). The primary objective of this platform was to provide support to Czech sportsmen and sportswomen. Here, the project could have been established by any professional athlete or sportsgroup in the Czech Republic in order to attract funds for realization of their goals. The website of each project is always accompanied by short video together with list of rewards (in order to assure funders, that it is not a donation-based model). The key difference of this portal was, that it incorporated scenarios when funding goal was not reached, i.e. *keep-it-all* model. The last active post on the webpage was published in 2016 and therefore, it was concluded, that this platform is no longer active.

Startovač, founded in 2015, is the second largest platform in the Czech Republic (but operating in Slovakia as well). Here, project initiators can publish actions of various types - artistic or entrepreneurial, in order to collect financial resources for their realization. Here, charity projects (this means projects without rewards to be delivered in exchange for contribution) are strictly prohibited. Startovač offers different time spans for collection of money - namely 15, 30 or 60 days. The recommendation from platform is intuitive: the greater the size of pledging goal, the longer duration of the campaign should be. Again, platform adopts so-called *all-or-nothing* approach, so the project founder receives the money only if the funding goal is attained. Compared to analysed platform, HitHit, Startovač adopts slightly different approach, which is targeted on greater overall success rate of the projects. That is why the portal allows its users to raise money for projects that are worth for example, only CZK 10,000 (and more). This resulted in success rate being equal to 60% in 2015 (Boček 2015).

Overall, the market in the Czech Republic is still expanding. Here, crowdfunding is not directly regulated. On the contrary, there exist general laws and rules that set guidelines not only for collecting money from the public, but also for the use of funds. In addition, there are also restrictions that aim to protect consumers or prevent money laundering (Šoltés & Štofa 2016).

HitHit

HitHit is the largest platform in terms of volume of successfully financed projects in the Czech Republic. Launched in 2012, HitHit has quickly become the leading reward-based CF intermediary, as the total amount of funded projects until January 2020 has been approximately CZK 244 million.[4] In order to better understand the picture behind the platform's operation, the lifetime of a project, together with platform-specific conditions, will be described below.

HitHit supports actions of any kind, except those with charitable background, which have no rewards to offer. The minimum requested amount for every project is set to be at least CZK 50,000, the maximum size of the pledging goal is not specified. Every project initiator has to agree with platform's terms and conditions in advance. Subsequently, one has to specify contact details, name and short description of the project plus the amount of money being requested. Afterwards, platform evaluates the project. This means that the content should meet some appropriateness and also ethical criteria. This procedure usually lasts around one week. If there are no concerns about project's eligibility, it is published on the website. In the past, the founders could choose the duration of the campaign at the platform's webpage - either 30 or 45 days. During this time, HitHit users can search through active projects on the webpage and decide, which they would like to support by particular amount of money. In addition, they select a reward which they would like to receive in exchange for their contribution. Only if the target amount is reached in pre-

[4] Author's own computations for published projects until January 2020.

specified time, project founder receives the money collected and is obliged to deliver the promised rewards. Furthermore, only in this case he or she pays reservation fee to the platform. For projects under CZK 200,000 it is 9% out of the target amount together with administrative fee of CZK 699. Project founders are also expected to cover costs associated with value-added taxes. On the contrary, if the project is worth more than previously mentioned CZK 200,000, the platform offers individual solutions that should be tailor-made for each initiator. However, regardless of the project being fully financed or not, project founders have to cover transaction costs of the payments.[5]

Special service (compared to other platforms operating in the Czech Republic) that HitHit offers to project founders, who feel uncertain of their marketing skills, is so-called HitHit LAB. This particular service is recommended especially for artists or initiators of creative projects, who wish to set the large target amount and may not feel proficient enough in terms of online communication. HitHit therefore provides various types of advisory services, such as mentoring, professional copywriting, shooting and editing of project videos or other above-standard paid services like social media management. Here, the main goal is to increase project's PR as much as possible.

As already mentioned before, HitHit experiences lower success rate of projects compared to its greatest competitor, Startovač. On the contrary, the funded projects are separately of greater volume.

The biggest amount of money that was pledged, CZK 3,962,664; was for enterprise *Bohempia*. The company focuses on manufacturing of barefoot shoes made of hemp. This project collected 808 % out of target amount in 30 days (the pledging goal of CZK 490,000 was collected in less than 10 hours).

The two biggest players in Czech crowdfunding environment also share another special feature. Sometimes the mobilization of project fans and community is not enough to meet the pre-determined goal. In this situation, both platforms offer the opportunity to initiators to pay the rest of the funds by themselves. The allowance of such practice is not that common. For example, on US server Kickstarter it is strictly prohibited and may be the reason why whole project is cancelled. HitHit and Startovač claim that for authors, the project realization is the key objective, so it does not matter if they contribute partly by themselves. What is more, the success of the campaign is also in the interest of platforms, as only once the goal is reached, they generate profit.

[5] Information extracted on March 26, 2020.

2.2 Natural Language Processing

NLP started to occur around 1950s, as a byproduct of Artifical Intelligence (AI) and linguistics. Despite the fact that at the beginning, NLP did not go hand-in-hand with information retrieval (primarily focused on searching and indexing large amounts of text), those two fields have converged over time, as highlighted by Nadkarni *et al.* (2011).

Generally, in programming, the text is understood as *unstructured data,* due to the fact that the structure of given information is not known to the computer software. Revolution in this field came with Noam Chomsky, who demonstrated, that the language can be analysed by means of mathematical structure. Next turning point occurred in 1980s, with the development of statistical self-learning systems. Those systems are often compared to neuron networks, but there exist several versions. The language is often modelled in algebraical space which has thousands of dimensions which distort and screen into another spaces. As already mentioned, these systems learn from the processed data. For example, the Internet translators are 'trained' on thousands of already translated phrases (Hana 2015).

There exist several definitions and suggestions of what NLP actually is. It can be viewed as the ability of computer software to recognize and analyse human (spoken) language. Nowadays, it is already perceived as a part of AI (Liddy 2001). Within the lexical field of NLP, academic society (such as Rouse (2016)) often distinguishes two main subgroups:

- Syntax
- Semantics

Syntax describes the arrangement of words in a sentence to make logical sense. In this case, sentence is separated into individual components and subsequently, dependency between the terms is inspected. Syntax analyses, which are frequently used, include: parsing (overall grammatical breakdown of a sentence), breaking of sentences (technique placing sentence boundaries in large amounts of texts) or word segmentation. In simpler words, syntax accounts for meaning. As Liddy (2001) emphasizes, order and dependency form the meaning. Sentence *"John likes Kate."* is clearly not the same as *"Kate likes John."*

Semantics, on the other hand, is devoted to the actual purpose and use of words in a text. This is done by application of various algorithms. Some items of semantic analysis include named entity recognition (categorisation of a word into specific group) or word sense disambiguation (determining the meaning of a word from a context of a whole sentence).

As Named Entity Recognition (NER) lies within the scope of this paper, it will be explained in more detail. NER in simple words refers to determination, whether a word (or group of words) belongs to some particular category. Currently, there exist both statistic, as well as linguistic proposals on how to work with texts. For example, NLP processor can be taught to concentrate on details in capitalisation, specific suffixes (etc., Inc.). In addition, the software can be set to spot some characteristic patterns hidden in text. These strategies become even more challenging, when the reviewed text is informal – e.g. when typing errors are present or spoken informal language is being analysed (Falci 2017).

2.2.1 Sentiment Analysis

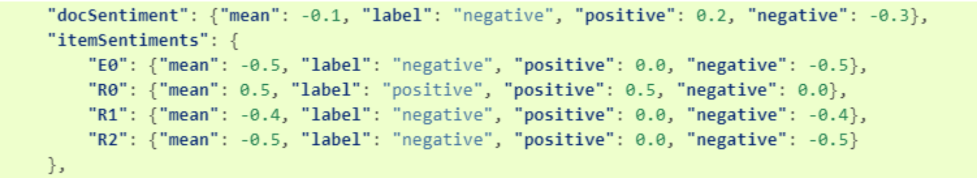
Another tool which is used in this paper is sentiment analysis, also interchangeably named as opinion mining (Sun *et al.* 2016). It was formally introduced at the beginning of this century. Sentiment analysis is understood as computational treatments, which aim to disclose the opinions perceived by opinion holders which are expressed in text. The goal is to infer attitude of a person to some specific topic. This point of view can be person's judgement or assessment (Wang *et al.* 2017). There exist two main approaches in order to infer sentiment from the text - lexical-based approach, which basically extracts opinion words. The second one, which is also used in this thesis, is called machine-learning approach.

For computation of sentiment for this thesis, the software named 'The Interpretor' developed by Geneea Analytics adopts the latter, machine-learning approach. The software is trained several times (the Interpretor has a huge knowledge base containing millions of records) on various types of text. Each recognised opinion word in the database is extracted and labelled with positive/negative score which is defined in the database as well. In addition, this score may change if it is present with some other word in relation, which is also recognised in the database. On top of that, software is capable to distinguish, whether the sentence or particular word is negated – in such case, positivity or negativity is handled vice versa. The software therefore calculates the mean sentiment for the whole sentence or document as follows:

$$meansentiment = pos\_score + neg\_score \tag{2.1}$$

Where pos_score is weighted average of all positive scores of items from the text and neg_score represents the same, but for the negative group. For better illustration, Figure 2.4 depicts the output for analysed sentence: *"The trip to London was amazing. Only the food was weird. Especially the pizza was terrible."*

Figure 2.4: Sentiment Calculation by *The Interpretor* Software



```
"docSentiment": {"mean": -0.1, "label": "negative", "positive": 0.2, "negative": -0.3},
"itemSentiments": {
    "E0": {"mean": -0.5, "label": "negative", "positive": 0.0, "negative": -0.5},
    "R0": {"mean": 0.5, "label": "positive", "positive": 0.5, "negative": 0.0},
    "R1": {"mean": -0.4, "label": "negative", "positive": 0.0, "negative": -0.4},
    "R2": {"mean": -0.5, "label": "negative", "positive": 0.0, "negative": -0.5}
},
```

Notes: The figure depicts sentment calculation in Python, generated for the sentence: *"The trip to London was amazing. Only the food was weird. Especially the pizza was terrible."* The detailed explanation can be found below.

Source: http://www.geneea.com

Here, the software recognized the total of 4 items (i.e. opinion words). E_0 which stands for *pizza*, has taken over the negative sentiment value from the adjective *terrible* (being R_2). R_0 stands for *amazing* with positive sentiment value being equal to 0.5. R_1 stands for *weird,* which is relation with *food* understood as negative term. Then, the overall sentiment of the analysed sentences is equal to -0.1 (term "docSentiment" "mean"), which is a sum of weighed average of all positive terms plus weighed average of all negative terms.

2.2.2 Geneea

Geneea Analytics s.r.o.[6] was established in 2014 in order to provide analyses of complex texts as a paid service for various clients, such as mass media companies, banks and e-shops. Most frequently, it deals with large amounts of customer feedback, newspaper articles or it creates support chatbots, computer programmes that imitate human conversation with Internet users. Currently,

[6] www.geneea.com

it operates mainly in the Czech Republic and Slovakia, but it also has some customers in the UK or the US.

Since 2019, it has entered into the contract with Charles University Innovations Prague. The Institute of Formal and Applied Linguistics, Charles University, developed a software called MorphoDiTa (Morphological Dictionary and Tagger), which is further distributed and sold by Geneea Analytics. It can perform several above-mentioned analyses, such as tagging, named entity recognition, tokenization or parsing of a sentence.

Chapter 3

Literature Review

3.1 Literature Reviews Devoted to Crowdfunding

Because of the fact that crowdfunding influences many academic disciplines, the research scope of existing studies and papers can be described as wide and dispersed. This also explains the lack of comprehensive literature reviews existing in this field.

Gleasure & Feller (2016) incorporate the method of so-called 'metatriangulation' (i.e. building a theory from multiple paradigms) of 120 crowdfunding-centred papers, in order to identify the leading theoretic approaches. This is done for each of existing 4 crowdfunding categories (described in Chapter 2) separately.

However, there also exist research works which adopt process of separation into categories based on different criteria. For example, Moritz & Block (2016) use analysis of current crowdfunding environment based on division in terms of main stakeholders - capital seekers, its providers and platforms.

Stasik & Wilczyńska (2018) created a comprehensive overview of current methodologic approaches by dividing studies into 3 groups, based on typology of the research works. Those include platform-centred studies, studies exploiting the institutional concept of crowdfunding and last, but not least, the role of crowdfunding at microeconomic level.

As stated before, the existing range of academic debate on crowdfunding is too wide for the scope of this thesis. Therefore, some up-to-date topics are discussed in the next subsection. All in all, the strain is put primarily on crowdfunding success factors in line with the main objective of this thesis. In addition, some pioneer works elaborating on the impact of text on the likelihood of successful campaigns are analysed as well.

3.2 Typology of Research Works on Crowdfunding

3.2.1 Goals of the Project Founders

Previous research literature has already examined the topic of motivation of crowdfunding from the project founder's perspective several times (for example, Belleflamme *et al.* (2013) or Hu *et al.* (2015)).

Even though that the funding seems to be the primary goal in the business concept, it might not always be true. For example, crowdfunding can be used for demonstration, that there indeed exists demand for particular product or service. This can in turn lead to attracting funding from traditional financial sources. A case in point are watches Pebble, that gained attention via crowdfunding campaign realised in 2012. Even though it was its first Kickstarter campaign, the team behind the smart watches managed to collect more than USD 10 million, contributed by total of 69,000 backers. As the initial target amount was only 1% of this sum (USD 100,000), it has been shown that there is an excess demand for such products, which boosted Pebble's trademark.

Crowdfunding campaign can be also exploited in terms of marketing purposes. As proposed by Mollick (2014), this is extremely important for the projects in their early stages. Namely for the industries, where projects create opportunities for other complementary goods or services. For illustration, some software producers can begin to develop applications for specific products (assume Pebble watches, for example) even before the actual product has been released.

Overall, similarly to other forms of venture financing, crowdfunding may allow the inflow of additional resources (along with provision of capital), which is certainly one of the biggest motivational byproducts for entrepreneurs in terms of future project realization.

3.2.2 Motivations of the Project Backers

Allison *et al.* (2015) state two elementary categories of these motivations, *intrinsic* (such as altruism) and *extrinsic* (receiving public rewards or other benefits from contribution).

More recent works, such as Zhang & Chen (2019), incorporate slightly different division. Authors distinguish between *other orientation* and *self orientation*. Other orientation is described as motivation to help others, i.e. altruistic case. The latter is characterised as backer's possessive motivation stemming from his/her ego, expecting that the funding will bring him/her satisfaction and the feeling of taking control over the project. The findings imply, that self orientation has stronger effect than the other orientation, when it comes to funding decision. In addition, authors also inspect backer's motivations from the gender perspective. They found that the ties between other orientation and decision to fund is stronger for females.

Steigenberger (2017) incorporates a different technique - he conducts a survey, which was distributed across reward-based crowdfunding supporters. From the data, two groups of supporters can be distinguished. Firstly, funders who are motivated solely by purchasing motive. Even though the second group is motivated by the purchase as well, they additionally care of involvement and altruistic motives. When evaluating the trustworthiness, contributors rely primarily on previous activities of an entrepreneur.

Leimeister & Bretschneider (2017) found similar results. By examining backers' motivations, authors find that they often perceive self-interest motivations to fund – to possess a reward, to be recognized by others or to lobby a certain project in the hopes of its success. But on the other hand, they emphasize that a certain fraction of backers are pro socially oriented and they develop simple feelings of liking for the project idea.

All in all, both types of incentives have to appeal to project founders' motivations (Hossain & Oparaocha 2017). This is the milestone, where the two-sided communication stemming from attractive value proposition is extremely important and can have big impact on campaign success (Belleflamme *et al.* 2014).

3.2.3 Antecendents of Successful Campaign

As the primary objective of entrepreneurs is the realization of their project plan, crowdfunding success factors justly belong to one of the most investigated subject. This is beneficial, as due to the increasing number of studies, one can easily compare, how the results may change in different populations, types of projects or sample sizes.

One of the very first research papers investigating the determinants of successful funding was published by Mollick (2014). From the dataset of both, successful and failed projects, author choses total of 22,651 ideas (worth USD 5,000 or more) for analysis. Along with independent variables that are normally accessible at Kickstarter database, such as target amount of money or number of project backers, Mollick also incorporates additional regressors. To be more precise, he assumes that the following variables serve as a proxy for project's quality (as they indicate its preparedness):

- Utilisation of a video
- Minimum spelling errors
- Regular updates up to three days

Based on the results, Mollick claims, that project campaigns signal higher quality by including previously mentioned explanatory variables. As a result, they experience greater probability to obtain funding. In addition, the study also confirms general trend observed by other researchers as well - that successful initiatives achieve the pre-determined goal only by small margin, whereas unsuccessful project ideas usually fail by big difference. This phenomenon is supported by other research works (see Šoltés & Štofa (2016) or Wang *et al.* (2017)).

Cordova *et al.* (2015) conducted study similar to Mollick (2014). However, they used different sample - sum of 1,127 technology projects from 4 different platforms (IndieGoGo, Kickstarter, Ulule and Eppela). As opposed to Mollick (2014), authors indicate, that project duration is positively related to the likelihood of success of the campaign. What is more, Cordova *et al.* (2015) argue that regressors serving as a signal of quality actually do not have any significant impact on funding success.

Zhou *et al.* (2016) contributed to existing research by analysis of 151,752 projects on platform Kickstarter. Authors include the standard control variables indicated by previous literature, such as pledging goal, campaign duration or number of images/videos. However, authors also aim to investigate crowdfunding in terms of content found in the project description. More precisely, they use three variables related to the content of project description and two of them are describing the characteristics of project's owner. Extending the previously mentioned findings, Zhou *et al.* (2016) indicated that information content has significant and positive effect on successful project funding.

Having a closer look on studies performed on smaller samples, one may spot several differences. For example, Frydrych *et al.* (2014) indicate that duration of the campaign has no significant impact and that smaller pledging goal increases the project's legitimacy. Consequently, it increases chances for successful funding as well. In addition, authors observed, that inclusion of short video clip becomes a common practice. Thus, its utilization has no significant impact on project's success.

Another paper working with a small sample was published by Hobbs *et al.* (2016) who analysed 100 creative projects published on Kickstarter. They introduce interesting coding scheme, which aims to evaluate the quality of each campaign. As results indicate, authors identified two groups - strong and weak predictors of funding success. The money raised, reward quality as well as number of project contributors were identified as variables with predictive power in terms of successful campaign. On the contrary, the second group (weak predictors) consisted of number of rewards, pledging goal as well as results on Google search. The findings are consistent with Frydrych *et al.* (2014), as length of the campaign (i.e. duration) has no significant impact on meeting the target.

3.2.4 Category Effects

An interesting subject of investigation is also provided by academic work concentrating on project categorization. These findings are extremely important for discussion whether some specific groups of projects based on activity (or product) are having greater chance of funding success. So far, there is still no consensus on the impact of respective categories on project success.

In their study, Chan *et al.* (2018) explain why such heterogeneity across categories may exist. Every project can exhibit distinct preferences and investment patterns. A case in point can be technology projects, that are likely to have more sophisticated investors. What is more, realisation of such project probably requires more of founder's effort and time. On the other hand, new initiatives in artistic environment incur less costs during the campaign creation. Authors find, that categorization has only low, but still significant impact on project funding.

For instance, Crosetto & Regner (2014) using probit regression, find that categories *literature, design* and *games* have negative and significant effect on project success. On the contrary, category *music* creates positive and significant effect.

This was also confirmed by Mollick (2014), who emphasized that there exists systematic variation of success rate across project categories.

3.2.5 Textual Analyses of Crowdfunding Campaigns

As the popularity of language processing tools increases across various fields, it naturally started to occur in crowdfunding environment as well. Felipe *et al.* (2017) pointed out, that the evolution of NLP enabled wider spread of research works, for instance, analysis of narratives present in the project descriptions. In addition, authors discuss the opportunity to assess the textual feeling, which can in turn explain investors' behaviour, in case of crowdfunding, the decision to fund.

One of the ground-breaking works was published by Mitra & Gilbert (2014). In their paper, they examined 45,000 projects on Kickstarter, containing the total of 9 million phrases present in project descriptions. After separating 20,000 phrases for more detailed analysis, it was discovered, that the language chosen by project founders has unexpectedly strong predictive power. It explained almost 60% of the variance of the success. A closer view on the phrases disclosed, that they contain the basics of persuasion principles. Together with the study, authors publicly released the set of positive and negative predictive phrases, with the intention to be considered in project descriptions of future campaigns.

Gorbatai & Nelson (2015) examined the research question, whether linguistic content of reward-based crowdfunding campaigns has an impact on fundraising result. In particular, they argue that women founders are more successful than men. The findings reveal, that females use slightly different language and communication, which significantly affects the campaign's outcome. On top of that, they suggest that business model of reward-based crowdfunding could reduce gender inequalities in fundraising area, as women benefit from the style of their communication.

Wang *et al.* (2017) aimed to assess the impact of sentiment factor in crowdfunding project descriptions, which may influence the decision to fund the campaign. Results indicate that (in comparison to the baseline model) texts with positive sentiment in project descriptions, increase the predictive accuracy of the model by 7%. On the other hand, this does not hold for project title.

Next academic paper that explores the persuasive power of project descriptions was published by Zhou *et al.* (2016). The authors inspect, whether the following three variables: length of project description, inferred readability and tone impact the success of the campaign. Findings reveal that those attributes of project descriptions have incremental predictive power. On top of that, authors suggest that project founders should be aware of this impact and could exploit these in order to improve the likelihood of success.

3.2.6 Possible Extensions of Crowdfunding Research

To summarize, studies and research papers which have been published so far often offer only narrow insights, which are applicable to limited extent. Mollick (2014) suggests several areas to which the attention of academic society should be paid. Firstly, Mollick raises question, whether the criteria of CF project funders differ with respect to other forms of fundraising. Secondly, since the crowdfunding has been proven to remove geographic limitations, it is important to determine and understand, what is the role of such factor, if any. Last but not least, author emphasizes, that crowdfunding can serve as a useful model for other nascent ventures at their early stages.

Chapter 4

Data, Descriptive Statistics and Key Variables

4.1 Analysed Dataset

The dataset used for empirical analysis was generously provided by the biggest Czech crowdfunding platform (in terms of volume of funded projects) HitHit. The initial dataset consisted of the total of 7,147 projects, which were created on the platform since its establishment (i.e. November 2012), until January 2020, so that the whole population of its projects will be analysed.

However, the original sample contained data entries about all project initiatives - some of them were created within the creators' interface of the platform's webpage, but in the end, they were not launched as public campaigns. Thus, those unfinished initiatives had to be excluded, as it is meaningful to examine only the campaigns that were completed and thus, can have only two outcomes - successful or not.

After this step, the dataset comprised of 2,870 projects with complete information. Nevertheless, some data entries required further editing. For instance, the projects issued in Euro (usually ones originating in the Slovak Republic), needed to be searched directly at the webpage, to plug in the denomination in Czech Koruna.

Each project is characterised by unique ID, title of the project, its author, project description, size of the pledging goal, money that was actually collected, length of campaign (either 30 or 45 days), utilization of video, number of rewards, number of contributors, category, location and finally, success of the project and the overall percentage (with respect to the target amount) which was funded.

Subsequently, the dataset obtained from HitHit was combined with textual analysis of project descriptions, performed by Geneea Analytics. During this step, numerical value of sentiment was added to each project. Furthermore, the list of most frequently used named entities was created (it was decided to select words occurring more than 60 times in the whole dataset of project descriptions). In addition, as an extension, three most commonly occurring relations of noun and adjective and two relations of verb and noun were incorporated.

R Studio and The Interpretor software developed by Geneea Analytics were used for the following analysis.

4.2 Key Variables

Based on the evidence from existing research literature and data available from HitHit, the following variables were chosen to be included in the proceeding analysis. For the sake of clarity, variables were separated into two groups – so-called *standard* predictors, quantitative variables that characterise the campaign and *textual* predictors, variables which were derived by Geenea Analytics from textual description of each project. Note that not all variables are incorporated into all steps of the analysis.

Standard Predictors

Success

A dependent binary variable being equal to 1 if the campaign of project i was successful (i.e. if the contributed amount of money exceeds the pledging goal) and zero otherwise.

Goal

The amount of money in CZK which is required to be collected for project i in order to be successful.

Collected

The total amount of funds (in CZK) collected for project i during the duration of the campaign.

Percentage Funded

Variable *Percentage Funded* refers to the share of funds, that were collected during the time span of the campaign compared to the pledging goal for each project *i*.

Backers

Number of contributors who decided to fund project *i* during funding cycle.

Rewards

Number of rewards for each project *i*. Every funder chooses one from this set of rewards. If and only if the campaign is successful, he/she will receive this reward as an exchange for his contribution to the campaign.

Duration

A binary variable representing the length of a funding cycle for project *i* which can be equal either 30 or 45 days.

Video

A binary variable being equal to one, if project *i* has a video published at the project's website, and zero otherwise.

Capital

A binary variable stating whether the project *i* is located in the capital city (1) or not (0).

Category

Group of binary variables, being equal to 1, if project *i* belongs to particular category. Each project is assigned to only one of those categories: *Music, Movie, Art, Sport, Games, Fashion, Technology, Writing, Theater, Food, Education, Community, Dance, Photography, Vodafone, Design.*

Textual Predictors

Sentiment

Variable *Sentiment* refers to the overall mood originating from the project description. It ranges from –1 (negative) to 1 (positive). The value of zero refers to neutral position of the author of the text.

Named entities

Dummy variables being equal to one, if particular named entity is recognised in the project *i*'s description. This group of words includes: *Book,* *Child, Movie, Album, Festival, CD, Year, World, Release, Journey* and *Project.*

Adjective Relations

Dichotomous variables being equal to one, if description of project i has those relations included in text. This group consists of: *New Album, Debut Album* and *Collected Money.*

Verb Relations

Binary variables, similarly as in the previous cases, equal to one if the relation with verb and noun is included in project i's description. This set includes: *To Release Book* and *To Need Help.*

4.2.1 Descriptive Statistics

Before the estimation of empirical model, the sample underwent preliminary analysis. The table below shows distribution of project ideas in categories, sorted by frequency.

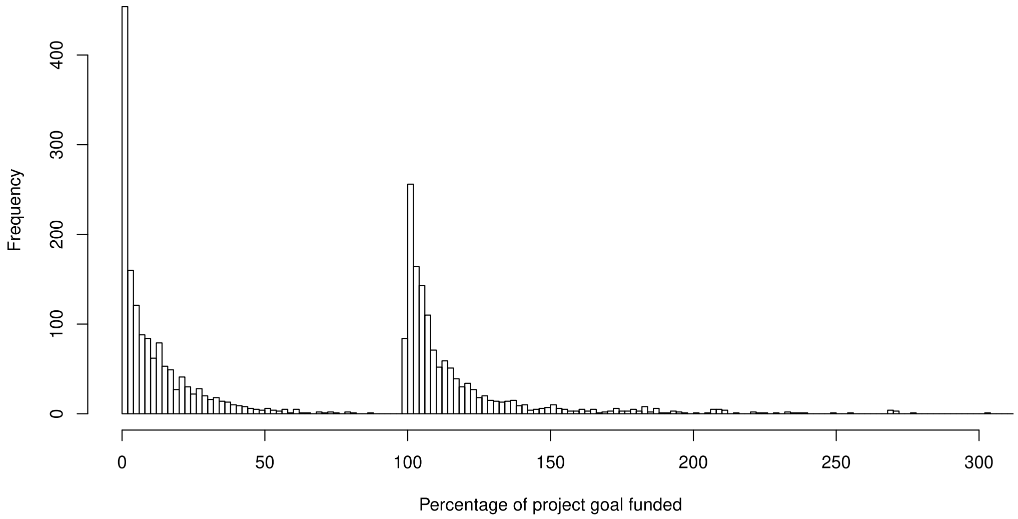
Table 4.1: Projects Grouped by Categories and Result of the Campaign

	#	Successful	Unsuccessful	Success Rate
Music	597	355	242	59%
Writing	418	223	195	53%
Art	331	160	171	48%
Movie	273	139	134	51%
Sport	206	84	122	41%
Education	195	102	93	52%
Food	170	60	110	35%
Technology	153	52	101	34%
Design	140	54	86	39%
Community	137	66	71	48%
Theater	114	63	51	55%
Games	41	14	27	34%
Fashion	32	12	20	38%
Photography	26	4	22	15%
Vodafone*	24	21	3	88%
Dance	13	4	9	31%
Grand Total	2,870	1,413	1,457	45%

*This category comprises project ideas, that were selected to be supported by mobile operator Vodafone, in order to increase the overall well-being of Czech society. Therefore, campaigns had much greater PR and gained more public awareness.

The three most common categories are *Music, Writing* and *Art*. On the contrary, the three least represented categories are *Vodafone, Photography* and *Dance*. *Vodafone* is the category experiencing the highest success rate. In addition, approximately every second project idea published in categories *Music, Writing, Art, Movie, Education, Community* and *Theater* ends with favourable outcome, too. Furthermore, it is apparent, that the overall success rate on the platform is notably low, being only 45%. This is consistent with statistics of crowdfunding platforms abroad. So it is meaningful to trace the factors, which have considerable effect on campaign success.

Figure 4.1: Outcomes of Campaigns as a Percentage of Project Goal Funded



Notes: The upper bound of x axis was set to 300, in order to eliminate extreme values and provide clearer picture.

Source: Author's own computations using extracted data, *N=2,870*

Figure 4.1 depicts the distribution of project outcomes, as a percentage of project goal which was funded. It confirms the general trend of crowdfunding campaigns, observed by many preceding academic works (for example, Mollick (2014), or Šoltés & Štofa (2016)). The chart demonstrates, that project initiatives succeed either by negligible margins, or fail by large amounts.

Table 4.2: Descriptive Statistics of Standard Predictors

	Mean	Median	Minimum	Maximum	Std.Dev.
Success	0.49	0	0	1	0.5
Goal	82,399	85,000	10,000	6,352,500	201,402
Collected	85,067	50,890	0	3,962,664	180,765
Percentage Funded	65.51	57	0	809	64.73
Backers	100	44	0	4,567	213
Rewards	14	12	2	82	8
Duration	42	45	30	45	6
Video	0.81	1	0	1	0.39
Capital	0.22	0	0	1	0.42

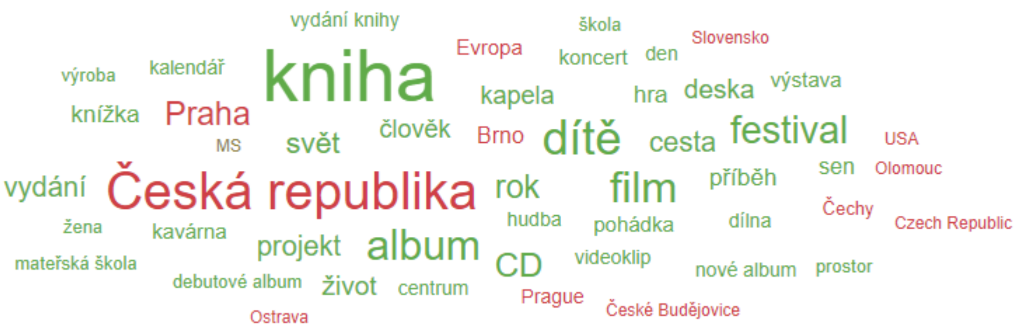
Source: Author's own computations using extracted data, *N=2,870*

Table 4.2 provides summary of standard predictors (except dummy variables for each category). The average value of project's goal is approximately CZK 82,000; slightly below median (CZK 85,000). The amount is in line with the fact that CF belongs to group of microfinancing tools. However, the maximal value of collected funds is considerably larger, compared to the average ones (collected by project *Bohempia* mentioned in Chapter 2). The mean of variable *Percentage Funded* clearly indicates that on average, projects reach only 65.5% of the target goal. For every project initiative, there participate approximately 100 contributors. The most common size of group of rewards is 12. Project with the greatest range of rewards offered choice from 82 alternatives.

4.2.2 Textual Analysis

As stated before, textual description of each project underwent textual analysis using software The Interpretor, developed by Geneea Analytics. In this subsection, interesting results and other insights from this analysis will be summarised. Figure 4.2 shows most common words present in the projects' descriptions. Based on this named-entity-recognition analysis, most frequent words were selected and incorporated into group of textual predictors.

Figure 4.2: Most Frequently Used Words in Project Descriptions



Notes: The size of a text demonstrates, that the word was used more frequently than others. Most commonly used words in project descriptions are: *Book, the Czech Republic, child, festival, Prague, CD, journey, Europe, release* or *world.*

Source: Author's computations using software The Interpretor

Next it was inspected, whether the subset of successful projects contains different types of words or relations, compared to unsuccessful ones. Table 4.3 shows the top 5 items for each category for successful project initiatives.

Table 4.3: NLP Analysis of Successful Project Initiatives

	Term	**Frequency**
Tags	*Book*	129
	Album	67
	Child	65
	Movie	59
	CD	48
Adjective Relations	*New Album*	23
	Debut Album	16
	New CD	16
	Young Person	14
	First Album	14
Verb Relations	*To Release Book*	23
	To Need Help	16
	To Support Release	16
	To Support Creation	14
	To Become a Part of	14

Source: Author's computations using software The Interpretor

Table 4.4 shows the same statistics for unsuccessful initiatives. From the both tables it is apparent, that project descriptions are formulated in a similar manner, regardless of the final outcome.

Table 4.4: NLP Analysis of Unsuccessful Project Initiatives

	Term	**Frequency**
Tags	*Book*	85
	Child	58
	Movie	50
	Festival	42
	Project	36
Adjective Relations	*New Album*	15
	Collected Money	14
	Broad Public	14
	Whole World	13
	Debut Album	13
Verb Relations	*To Release Book*	21
	To Support Release	16
	To Make Dream Come True	14
	To Use Money	13
	To Need Money	12

Source: Author's computations using software The Interpretor

The sample of project descriptions also underwent sentiment analysis. For every project i, the numerical value of sentiment, demonstrating the overall mood of the description, was calculated by *The Interpretor* software. It ranges values between –1 (text with strong negative mood) and 1 (positively formulated description). Table 4.5 provides the reader with descriptive statistics of this variable.

Table 4.5: Descriptive Statistics of *Sentiment*

Mean	Median	Minimum	Maximum	Std.Dev.
0.08	0.1	-0.4	0.5	0.12

Source: Author's own computations using *The Interpretor* software, *N=2,870*

Overall, it can be concluded, that the project descriptions are usually formulated in a neutral manner. Even the descriptions that were evaluated as the most positive, are reaching the value of 0.5. The same holds for the negative statements found in project descriptions.

Another perspective from which one can examine on the sentiment is its distribution across categories. Most of the campaigns (even the most successful ones, namely *Music, Writing* and *Art*) follow approximately normal distribution with mean value approximately around 0.1. However, there are some categories that seem to be more positive in terms of projects' sentiment. Those are: *Dance, Design, Games, Community* and *Food.* Histograms showing the distribution of sentiment for each category can be found in the Appendix.

Chapter 5

Theoretical Framework and Empirical Model

5.1 Bayesian Model Averaging

5.1.1 BMA – Introduction

Statistical models are built upon two pillars of assumptions - **structural** ones (such as inclusion of the variables, functional forms of models or choice of residuals) and the second, assumptions related to **interpretation of parameters**, subject to imposed structural assumptions. Consequently, when the chosen model is estimated, an uncertainty regarding the value of model estimate, arises. And again, it exists at two levels - uncertainty related to the actual value of estimate, conditional on given model. This type is usually addressed by particular study. What is not fully covered is the second type - uncertainty related to the selection of the model itself (Moral-Benito 2013).

Model uncertainty should be of high interest of scientists, as the values of estimated parameters can heavily depend on the particular model. The reason for that is simple - regression is vulnerable towards arbitrary decisions regarding the selection of explanatory variables (Leamer 1978).

A possible approach that deals with model uncertainty is to estimate all possible models from model space. Afterwards, weighted average of all estimates for each X is computed. This technique is called model averaging. It exploits the advantage of making the inference from the whole universe of candidate models.

Suppose we have matrix X of explanatory variables. A natural question arises, which $X_i's$ should be included in the model? If X has K variables, this leads to task of estimating 2^K models, in order to cover all solutions possible.

This thesis follows definition proposed by Zeugner (2011). The weights for each model are derived from posterior model probabilities, which follow Bayes' theorem:

$$p(M_\gamma|y,X) = \frac{p(y|M_\gamma,X)p(M_\gamma)}{p(y|X)} = \frac{p(y|M_\gamma,X)p(M_\gamma)}{\sum_{s=1}^{2K} p(y|M_s X)p(M_s)} \tag{5.1}$$

In this equation, $p(y|X)$ represents integrated likelihood, that is constant for all the models and therefore, it is multiplicative. Posterior Model Probability (PMP) $p(M_\gamma|y,X)$ is then proportional to $p(y|M_\gamma,X)$, which stands for marginal likelihood of the model (i.e. how data is probable given the model M_γ), times prior model probability $p(M_\gamma)$ (how probable is the model from researcher's perspective). Subsequently, the weighted posterior distribution for any statistics θ is equal to:

$$p(\theta|y,X) = \sum_{\gamma=1}^{2^K} p(\theta|M_\gamma,y,X)p(M_\gamma|y,X) \tag{5.2}$$

In simpler words, for every model, BMA computes the PMP which behaves like an information criterion telling he researcher, how well the particular model corresponds to the data. Then, reported coefficients are displayed as a sum – PIP, which is a sum of models where the variable was included. Thus, PIP provides the information on how likely is the variable present in the 'true' model.

The decision about model prior $p(M_\gamma)$ is taken by the researcher and reflects his/her beliefs. As Steel (2017) emphasizes, the assumptions imposed over the choice of priors are crucial, as the weights (derived from posterior model probabilities) depend heavily on the prior assumptions.

All in all, formulas for posterior distributions $p(\theta|M_\gamma,y,X)$ and marginal likelihoods $p(M_\gamma|y,X)$ should reflect selected estimation framework. Another important aspect, which has to be met, is a normal distribution of an error term of every model M_γ. Then, researchers are expected to state their prior beliefs on regression coefficients β_γ. It is usual to assume a prior mean equal to zero, which is rather conservative. This demonstrates, that not much is known about coefficients in the model (Zeugner 2011).

In case there are many explanatory variables to consider (39 in our case), estimation of 2^{39} models appears to be very complicated and more importantly, time consuming. In such case, Markov Chain Monte Carlo (MCMC) samplers offer a convenient solution for BMA estimation. They collect and keep the results of most relevant posterior model distributions and as a result, create a decent approximation. Zeugner (2011) defines the selection algorithm as follows:

At phase i, the sampler works with the model that is currently in use. Let M_i be such a model with PMP of $p(M_i|y, X)$. In the proceeding step $(i + 1)$, new model M_j is proposed to aspire as the 'winning one.' The new model M_j is selected by the sampler if and only if:

$$p_{i,j} = min(1, \frac{p(M_j|y, X)}{p(M_i|y, X)}) \tag{5.3}$$

If model M_j would be refused by the sampler, it moves to the next phase. New model M_k is proposed against M_i. If it would be accepted, it gains the title of 'current' model and has to withstand other models. Using this procedure, the number of times when every model is saved actually converges to the distribution of posterior model probabilities $p(M_i|x, Y)$.

For BMA methodology explained in more detail, please refer to e.g. Steel (2011), Moral-Benito (2013), Steel (2017) or Ley & Steel (2007).

5.1.2 BMA – Specification and Sampling

For the following analysis, package BMS for R Studio was used. It was introduced by Zeugner (2011) and belongs to widely utilized tools for this type of investigation.

In the first step, model prior needs to be determined. Here Beta-Binomial Prior was incorporated. It places majority of the data near prior model size. This particular type was suggested by Ley & Steel (2007) and same as other types of model priors, it requires to choose only expected prior model size. The advantage compared to for example, binomial model prior, is that its usage reduces risk of unintended misleading result, when imposing assumptions about model size.

After the model prior is chosen, hyperparameter g has to be specified as well. It represents researcher's beliefs that coefficients are equal to zero. A small g means that he or she is quite certain that they are indeed 0. A large g signals the opposite. This thesis uses common conservative practice, so-called Unit Information Prior (UIP). It sets $g = N$ (2,870 in our case) for all models.

Therefore, it attributes the same information to the prior as it can be found in one observation.

When specifying the MCMC sampler, so-called *birth-death* sampler was chosen, as it is commonly used in most BMA applications. It adopts the following mechanism: one of K sets of variables is randomly selected and new model M_j is proposed - if the set is already part of 'current' model M_i, then M_j will have the same group of covariates, but not the chosen covariate. If the variable is not in M_i, then candidate model M_j will contain all variables from M_i plus the selected variable.

In order to increase the quality and accuracy of MCMC sampling, one has to specify number of draws that sampler runs through (as it naturally starts at 'some' model which might not be the best one). Consequently, the first sequences of draws might contain models with low PMPs. Therefore, the first set of draws (so-called *burn-ins*) is intended to be left out from the computation results. On the contrary, parameter *iterations* specifies number of proceeding iterations which will be kept.

Table 5.1 shows the set of dependent variable *Success* and 39 explanatory variables defined in Chaper 4 which were considered in the BMA analysis. All variables are in levels, except variable *Goal*. This variable was transformed using natural logarithm in order to deal with variability of the data (note that the name of transformed variable remains unchanged).

Table 5.1: List of Variables for BMA Analysis

Dependent Variable			
Success			
Standard Predictors			
Video	*Goal*	*Rewards*	*Backers*
Duration	*Capital*	*Music*	*Vodafone*
Fashion	*Design*	*Food*	*Theater*
Education	*Dance*	*Games*	*Sport*
Technology	*Movie*	*Writing*	*Photography*
Art	*Community*		
Textual Predictors			
Book	*Project*	*Album*	*CD*
Movie	*Year*	*Journey*	*World*
Festival	*Sentiment*	*ToNeedHelp*	*ToReleaseBook*
NewAlbum	*CollectedMoney*	*Release*	*Child*
DebutAlbum			

The next two tables summarize the results of Bayesian model averaging, which was performed on the set of abovementioned 39 explanatory variables, using MCMC sampling method. Important note is that the sampling was done for the whole set at once (i.e. the two following tables represent one estimated BMA object). However, for the sake of clarity, the summary tables for variables have been presented in the two logical groups as before – standard predictors (quantitative variables together with dichotomous ones for each category) and textual predictors (sentiment, named entities and adjective/verb relations).

Table 5.2: BMA Coefficients: Standard Predictors

	PIP	**Post Mean**	**Post SD**	**CPS**
Video	1.0000	0.1364	0.0218	1.0000
Goal	1.0000	-0.0012	0.0001	0.0000
Rewards	1.0000	0.0124	0.0011	1.0000
Backers	1.0000	0.0008	0.0004	1.0000
Music	0.9943	0.1064	0.0258	1.0000
Duration	0.9629	−0.0059	0.0014	0.0000
Vodafone	0.9441	0.3405	0.1197	1.0000
Writing	0.5505	0.0521	0.0512	1.0000
Photography	0.2572	−0.0636	0.0797	0.0000
Fashion	0.0744	−0.0129	0.0339	0.0000
Design	0.0656	−0.0054	0.0185	0.0000
Food	0.0276	−0.0016	0.0513	0.0000
Theater	0.0215	0.0012	0.0412	1.0000
Art	0.0169	0.0007	0.0006	1.0000
Education	0.0168	0.0008	0.0217	0.0000
Dance	0.0155	−0.0002	0.0231	0.0000
Capital	0.0087	0.0001	0.0021	0.9850
Games	0.0078	0.0005	0.0008	0.0000
Sport	0.0069	−0.0001	0.0015	0.3280
Community	0.0066	0.0001	0.0005	0.0000
Technology	0.0047	−0.0001	0.0027	0.0000
Movie	0.0033	0.0001	0.0001	0.7794

Notes: Table shows the results of BMA analysis, ordered by PIPs. Estimated PIPs = Posterior Inclusion Probabilities for given variables refer to the sum of probabilities of all models, in which the variable was incorporated. SD = Standard Deviation, CPS = Conditional Positive Sign.

Source: Author's own computations using extracted data, *N=2,870*

Table 5.2 provides an overview about the group of standard predictors. The power of each variable in terms of explaining the data is given by the column **PIP – Posterior Inclusion Probability**. As mentioned before, it is the sum of all PMPs (Posterior Model Probabilities) in which the variable was included. From the summary it can be concluded, that 100% mass of all the models from the model space depend on the size of project goal, number of rewards at each project, number of contributors and short video clip. Next, 99% of the model mass includes category *Music*. 96% of the models also contained dummy variable for the length of the campaign. Category *Vodafone* was included in 94% of the models. Only 55% of the models considered category *Writing* as an important variable. Other standard predictors, like the rest of the categories (*Art, Food, Photography, Theater, Education, Fashion, Technology, Design, Movie, Fashion, Education, Dance, Sport, Community* and *Games*) do not seem to matter to such an extent. This also holds for dichotomous variable *Capital*, marking that the project is located in the capital city.

The second column **Post Mean** provides information about the values of coefficients averaged over all models (it also covers the models where the variable was not included, then the coefficient is equal to zero). In addition, it reveals the sign of the coefficient – while *Video*, *Rewards*, *Backers* and *Vodafone* are probably positive, *Duration* as well as *Goal* have very likely negative sign. **Posterior Standard Deviation** refers to the significance of the coefficient. The information about the sign of a coefficient can be also found in the fourth column – **Cond.Pos.Sign**. Zeugner (2011) defines it as 'posterior probability of a positive coefficient expected value conditional on inclusion'.

Table 5.3 reveals the results for the group of textual predictors. Here, the explanatory importance of variables in the model rapidly decreases. The winning textual predictor was word *Book*, included in the 52% of all of the models. Not only the Named Entities and Verb/Adjective Relations failed to have informative power, but also Sentiment inferred from project descriptions does not seem to have any impact on project success.

Table 5.3: BMA Coefficients: Textual Predictors

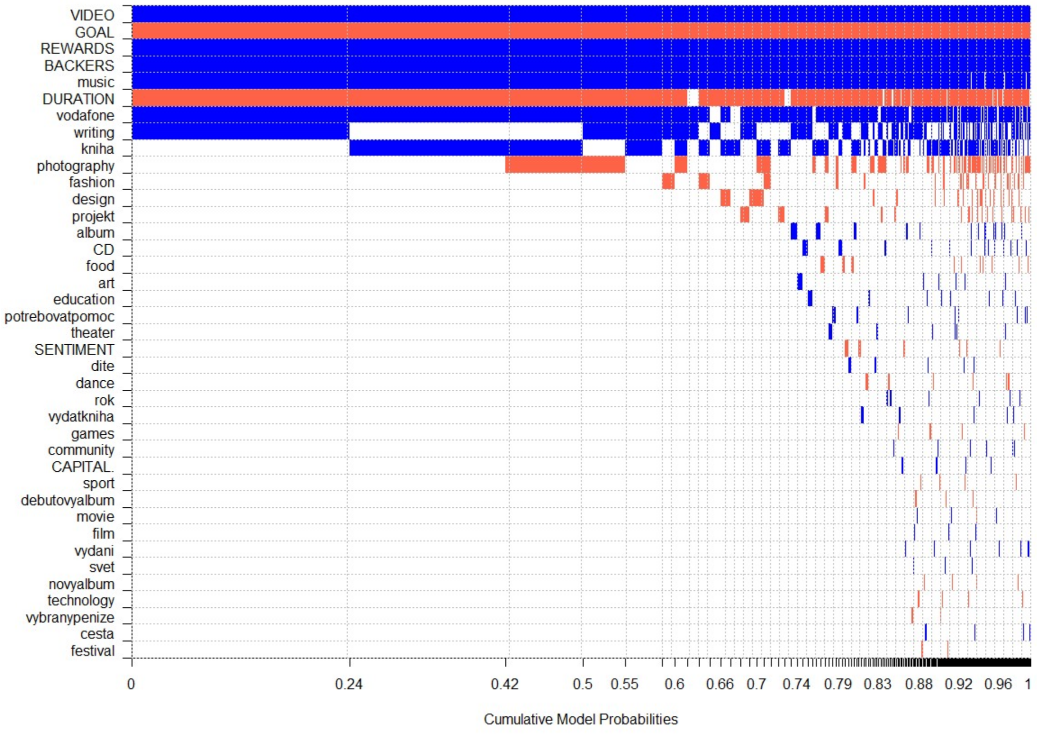
	PIP	**Post Mean**	**Post SD**	**CPS***
Book	0.5220	0.0641	0.0664	1.0000
Project	0.0498	−0.0047	0.0004	0.0000
Album	0.0327	0.0030	0.0263	1.0000
CD	0.0288	0.0022	0.0078	1.0000
ToNeedHelp	0.0201	0.0021	0.0178	1.0000
Child	0.0152	0.0001	0.0094	1.0000
Sentiment	0.0089	−0.0003	0.0059	1.0000
ToReleaseBook	0.0088	0.0006	0.0000	1.0000
Movie	0.0076	0.0000	0.0005	0.9810
Year	0.0065	0.0001	0.0000	1.0000
DebutAlbum	0.0060	0.0000	0.0064	0.8603
NewAlbum	0.0059	0.0001	0.0000	0.3128
CollectedMoney	0.0055	−0.0002	0.0000	0.0000
Journey	0.0050	0.0001	0.0055	1.0000
World	0.0116	0.0001	0.0052	1.0000
Release	0.0097	0.0004	0.0069	1.0000
Festival	0.0008	0.0001	0.0000	0.9937

Notes: Table shows the results of BMA analysis, ordered by PIPs. Estimated PIPs = Posterior Inclusion Probabilities for given variables refer to the sum of probabilities of all models, in which the variable was incorporated. SD = Standard Deviation, CPS = Conditional Positive Sign.

Source: Author's own computations using extracted data, *N=2,870*

Figure 5.1 displays graphical representation of BMA exercise (for the all variables). It is based on the best 1,221 models.

Figure 5.1: Graphical Representation of BMA Procedure



Notes: The dependent variable is *Success*, a dummy variable equal to 1 if the CF campaign was successful and zero otherwise. Rows display names of the variables, ordered by the value of PIP from the left upper corner. Columns show individual models. The darker colour (in black and white view) signals that the variable was incorporated in the model and has positive signs. The lighter colour (again in the grayscale) indicates, that the variable was included in the model as well, however, has negative sign. White fields mean that the variables were not included in the model.

Source: Author's computations using R Studio, *N=2,870*

Interpretation of BMA results is one of the most crucial steps in BMA exercise. This thesis follows Eicher *et al.* (2011), who set the intervals on the values of PIP. Eicher *et al.* (2011) divide the values of PIP based on the evidence for an effect. Values above 99% are marked as decisive, between 95% and 99% as strong and between 75% and 95% as substantial predictors. The values of PIP below 75% are labelled as weak predictors.

Based on suggestions of Eicher *et al.* (2011), decisive, strong and substantial predictors have been evaluated to have some considerable impact on project funding. However, putting the strain only on BMA analysis could result in misleading results. So logistic regression, using selected variables, which are *Video,* *Goal, Rewards, Backers, Music, Duration* and *Vodafone* will be estimated. This should serve as a robustness check of accuracy of used method.

Logit was incorporated due to similar motivations of other researchers in crowdfunding field, who usually use this type of model in their analyses. These include for example, Wang *et al.* (2017), Mitra & Gilbert (2014) or Mollick (2014). Before the estimation of a model, its theoretical framework (following mainly the approach of Wooldridge (2015)) will be presented.

5.2 Logistic Regression

5.2.1 Theoretical Framework

Logit model belongs to the family of binary response models, where the dependent variable takes only two values – 1 (success) and 0 (failure). It serves as a convenient alternative compared to linear probability model, as it ensures that the fitted probabilities lie strictly between 0 and 1. In addition, it handles heteroscedasticity of the error term, which is a feature of linear probability model. In logit, the response probability, defined as $P(y = 1|X) = P(y|x_1, x_2, ..., x_n)$ is of main interest.

In order to avoid limitations of probability models, logit model assumes a cumulative distribution function G, where $\forall z \in \mathbb{R}$ it holds that $0 < G(z) < 1$. For logit model, G is a logistic function, which is equal to

$$G(z) = \frac{exp(z)}{1 + exp(z)} = \Lambda(z) \tag{5.4}$$

It can be also expressed in binary response form:

$$P(y = 1|X) = G(\beta_0 + beta_1 x_1 + ... + \beta_k x_k) = G(\beta_0 + X\beta) \tag{5.5}$$

Adopting latent variable approach, one can derive such models. Let y^* be an unobserved (latent) variable, satisfying the following:

$$y^* = X\beta + e, y = 1[y* > 0] \tag{5.6}$$

Where e is independent of X and probability density function of e is symmetric around zero. G is a cumulative distribution function of e. Consequently, the response probability for y is defined as:

$$P(y=1|X) = P(y^* > 0|X) = p(e > -(X\beta)|X) = 1 - G(-X\beta) = G(X\beta) \tag{5.7}$$

Plugging into equation 5.4 yields:

$$P(y=1|X) = G(X\beta) = \frac{exp(X\beta)}{1+exp(X\beta)} \tag{5.8}$$

Interpretation of marginal effects in logistic regression becomes slightly more complicated, as they depend on X. A change in x_j does not induce a constant change in $p(y = 1|X)$. This can be shown by taking partial derivatives of response probability with respect to x_j. Two cases can be distinguished. If x_j is continuous explanatory variable, then

$$\frac{\partial p(y=1|X)}{\partial x_j} = g(X\beta)\beta_j \quad \text{where} \quad g(z) = \frac{\partial G}{\partial z}(z) \tag{5.9}$$

In the equation above, g is a probability density function. If x_K is a binary explanatory variable, the change in the probability of successful outcome can be estimated directly:

$$\frac{\partial p(y=1|X)}{\partial x_K} = G(\beta_0+\beta_1 x_1+...+\beta_K) - G(\beta_0+\beta_1 x_1+...+\beta_{K-1}x_{K-1}) \tag{5.10}$$

Because $G(.)$ is strictly increasing, then $g(z) > 0 \forall z$. Thus, the partial effect of x_i will always have the identical sign as the coefficient β_i. However, the magnitude of such effect is not that straightforward. In general, there exist two ways how to compute the marginal effects. Firstly, partial effects at average (PEA), which show marginal effect for average individual $g(\bar{X}\hat{\beta})\hat{\beta_K}$ (i.e. all variables are held at their means). This approach is not suitable for the following case, as there are several dichotomous variables and it does not make sense to say that average project has 0.65 videos. The second type average partial effects (APE) is more reasonable, as it takes sample average of the marginal effects across the whole sample.

$$N^{-1}\sum_{i-1}^{N} g(x_i\hat{\beta})\hat{\beta_K} \tag{5.11}$$

For estimation of binary response models, Maximum Likelihood Estimation

(MLE) is adopted. Firstly, density of y given x_i needs to by specified by the following function:

$$f(y|x_i\beta) = [G(x_i\beta)]^y[1 - G(x_i\beta)]^{1-y} \quad , \quad y = 0, 1 \tag{5.12}$$

The log-likelihood of the function for observation i is then

$$\ell_i(\beta) = y_i log[G(x_i\beta)] + (1 - y_i)log[1 - G(x_i\beta)] \tag{5.13}$$

The fact that $G(.)$ lies within $(0, 1)$ interval confirms that $\ell_i(\beta)$ is well-defined $\forall\beta$. Consequently, log-likelihood for the sample size of N is computed as follows:

$$\mathcal{L}(\beta) = \sum_i^N \ell_i(\beta) \tag{5.14}$$

The MLE of β (that maximizes the log-likelihood) is then $\hat{\beta}$ – a logit estimator.

When evaluating goodness-of-fit, binary response models also adopt slightly modified measures. An example is McFadden's Pseudo R-squared, which ranges values between 0 and 1.

$$R^2 = \frac{1 - L_{UR}}{L_0} \tag{5.15}$$

Here, L_{UR} is the log-likelihood for the estimated (unrestricted) model and L_0 is the model with only an intercept. Next alternative how to evaluate the model is percentage correctly predicted. For each i, the predicted probability that $y_i = 1$ given explanatory variables x_i is computed. Then, the percentage of times that predicted y_i matches the actual y_i is the overall percentage correctly predicted. Usually, it is more informative to display percentage correctly predicted for each outcome $y = 0$ and $y = 1$.

5.2.2 Logit Model - Estimation Results

Based on the BMA estimation performed before, as well as suggestions from relevant literature, the following model, given by equation 5.16 was introduced

($G(.)$ is a logistic function).

$$\begin{aligned} P(SUCCESS = 1|X) = G(&\beta_0 + \beta_1 Rewards + \beta_2 Backers + \beta_3 Music \\ &+ \beta_4 Video + \beta_5 Duration \\ &+ \beta_6 Vodafone + \beta_7 Goal) \end{aligned} \tag{5.16}$$

Before the application of the model to the whole sample, the chunk of the data was extracted as a testing sample. The model was firstly estimated for this part of the dataset and generated outputs were compared to those for whole sample. As it produced very similar results, it was proceeded to the model which covers all data entries.

Table 5.4: Logistic Regression and BMA Results from Previous Analysis

	Logistic Regression			**BMA**		
	Estimate	Std. Error	APE	PIP	Post Mean	Post SD
Rewards	−0.0217	0.0134	−0.0014	1.0000	0.1364	0.0011
Backers	0.0710***	0.0030	0.0047	1.0000	0.0008	0.0004
Music	0.3721**	0.1657	0.0248	0.9943	0.1064	0.0258
Video	0.6698***	0.0005	0.0448	1.0000	0.1364	0.0218
Duration	−0.0072	0.0119	−0.0005	0.9629	−0.0059	0.0014
Vodafone	2.7147**	1.2162	0.1815	0.9441	0.3405	0.1197
Goal	−0.0171***	0.0012	−0.0011	1.0000	−0.0012	0.0001

Notes: $PseudoR^2 = 0.614, ^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$; $APE =$ Average Partial Effects

Table 5.4 summarizes the results of model estimation merged with BMA analysis' results from the previous steps. Column APE shows the calculated average partial effects. Coefficients for variables *Backers, Video* and *Goal* are highly statistically significant – even at 1% level. Whereas greater number of contributors and utilization of a video has positive effect, the greater size of project goal decreases the likelihood of successful funding. Coefficients at categories *Vodafone* and *Music* are also positive and significant at 5% level. They also experience the same sign as it was indicated by BMA analysis. If a project belongs to category *Music,* the probability of success, ceteris paribus, increases by 2.4%. Utilization of a video also increases successful funding by 4.5%, holding other factors fixed. The greatest impact on successful funding, again ceteris paribus, would have presence in category *Vodafone,* showing that projects in this category increase the probability of successful funding by 18.1%. However, this magnitude (as well as significance) was very likely caused by low representation of this category in the whole sample (only 24 projects out of 2,870) as well as high success rate in this category, caused by increased PR and specific public attention for such projects, as already mentioned before.

Variable *Duration,* indicating that the campaign was 45 days (if it is equal to zero, it lasts only 30 days), is negative, which is in line with BMA analysis. On the contrary, its effect is not significant. Note that coefficient at variable *Rewards* experiences negative sign, whereas the sign of coefficient indicated in BMA exercise was positive. This could be caused by the distribution of error term at logistic regression (as BMA requires normal distribution of error term). Even though the Q-Q plot (see Appendix) signals, that it is approximately normally distributed, there might be some values which deviate from the boundaries of normal distribution (probably at its tails). This could in turn influence, that the outputs from the BMA analysis slightly differ from estimated logit. A case in point is indicated different sign of *Rewards*.

In order to further evaluate the predictive power of estimated model, percentage correctly predicted was computed firstly for the testing sample, which was decided to be 1/3 of the dataset. In this case, out of 957 projects, 90% were predicted correctly. Subsequently, this was done for the whole sample and results are summarised in Table 5.5.

Table 5.5: Percentage Correctly Predicted

Predicted	**True**	0	1
0		1,373	84
1		140	1,273

Source: Author's own computations using extracted data, *N=2,870*

Values indicate, that prediction accuracy of estimated model can be assessed as solid. Out of 2,870 projects, 2,646 were correctly predicted to either succeed or fail. This also serves as a robustness check, that the predictors incorporated into the estimated model are likely to play role in terms of probability of campaign being successful.

Chapter 6

Discussion of Results

Before summarising and concluding the findings, one should be aware of striking difference between causality and correlation. In other words, the analysis revealed, that there exist some forms of relationship between dependent and independent variables considered in the model, but this does not give us any title to claim why such mechanisms exist. The significance of coefficients simply does not directly induce the causal effect on the outcome of the campaign. Thus, it was revealed that the following factors play a role, but it does not mean that they persuade individuals to donate the money.

6.1 Estimated Results

This study performs analysis of 2,870 project initiatives of the biggest Czech crowdfunding platform HitHit from the beginning of its existence (June 2012) until January 2020. On top of that, it exploits NLP tools in order to uncover, whether language style used in project descriptions has some impact on successful funding.

In analysed sample, 81,2% of initiatives (2,334 out of 2,870) utilized a short video clip. Results imply, that inclusion of a video has positive and significant effect on the probability of success. This is in line with several existing works, for instance, Mollick (2014). On the contrary, Frydrych *et al.* (2014) emphasized, that inclusion of video clip becomes a common practice and its effect on favourable outcome has actually no significant impact.

Findings also indicate, that greater number of project funders increase the likelihood of success, but only by a small fraction. Number of rewards at project campaign has negative, but insignifanct effect on campaign's success.

This is partially confirmed by existing research like Hobbs *et al.* (2016), who emphasized that number of rewards has no significant effect.

The length of project campaign (i.e. *Duration*) has shown to induce negative, but insignificant effect on the outcome of the campaign. This is consistent with Frydrych *et al.* (2014) and Hobbs *et al.* (2016), who also claim that length of a campaign is not of particular importance.

What has also appeared to be true, is that classification into project categories has impact on the successful result. It has been shown, that categories *Music* and *Vodafone* significantly and positively impact the favourable outcome. This was pointed out several times by many academic works (see Mollick (2014) or Chan *et al.* (2018)). Furthermore, platform Kickstarter provides statistics about project campaigns, that confirm this phenomenon.[1] Here, the most successful project categories are *Music, Film & Video, Games, Art* and *Publishing.* Except category *Vodafone* being somewhat specific, it matches the significant categories indicated by conducted analysis.

6.2 Remarks on Textual Analysis

As a diligent reader may noticed already during the BMA exercise, textual variables did not succeed in demonstration of predictive power on the outcome of the campaign. In particular, it has been shown, that both sentiment, inferred from the project descriptions, as well as extracted terms (named entities, verb or adjective relations) are left out when tracing for strong indicators of campaign's outcome. This finding is in contrast with Wang *et al.* (2017), who found that project descriptions showing positive sentiment increase the predictive accuracy of the model by 7%. The significance of results can differ due to various reasons. Firstly, the computation mechanism incorporated by The Interpretor software does produce fairly neutral results, as was confirmed by Geneea company. Based on the interviews with Geneea analysts, even very positive texts usually reach maximal value equal to 0.5. On the contrary, sentiment calculated by Wang *et al.* (2017) exhibits greater polarity, stemming from the fact that it is very likely computed in a different manner. Next, English language is usually more emotional in its expression than Czech, which was also confirmed by Geneea based on experience from English customers. And thirdly, the dataset used by Wang *et al.* (2017) was much greater than the one

[1] https://www.kickstarter.com/help/stats?ref=hello

used for this thesis (126,593 vs. 2,870 projects). This could very likely influence the different findings as well.

The results for named entities are in conflict with Mitra & Gilbert (2014), who demonstrates surprising predictive power of language present in project descriptions. Again, the incomparable sample size could cause this difference in results (45,000 vs. 2,870).

In general, there is a room for several possible explanations, why Czech crowdfunding scene may differ.

Firstly, project descriptions are generally formulated in a neutral manner. Here, one does not analyse product reviews or customer feedback, which often tend to be of quite emotional nature. Furthermore, HitHit is not a primary room where charitable (i.e. more emotional) initiatives take place. Secondly, there definitely exist other, unobserved or unmeasured factors which would justly belong to the characteristics of a project, such as the size of social network built by initiator of the campaign, activity on social media, previous experience and overall preparedness of a project. Therefore, the content of the description may fall to the background. And last but not least, as already listed in literature review, contributors are usually motivated by other reasons and therefore, make the decision to fund the project based on attractivity of rewards, altruism or simply by knowing the project initiator, which can be another signal for project's quality.

Even though the textual analysis did not serve its initially intended purpose, it can still offer valuable insights into the background of the platform. In particular, based on the most commonly occurring word, *Book* and verb relation *To Release Book*, one can conclude, that HitHit often operates as a medium for unknown writers to succeed and publish their books. The success rate in this category reveals, that more than half of them (namely 53%) are successful. The similar case are newly published albums by music artists. Thus, textual analysis proved itself as a pillar which supports the hypothesis, that success of a project is likely to be influenced by classification into categories.

The important finding, that textual description is left behind without any particular importance, can also serve as a future recommendation for Czech project founders. In particular, they should concentrate more on previously mentioned 'strong' predictors. Namely, the crucial aspect is to wisely determine the goal of a campaign and definitely, to include a short video clip. On the top of that, they should be aware of other drivers which are not captured by the model. Activity on social networks, other social ties with project contributors or general

PR of the campaign may definitely play role in increasing the probability of success.

6.3 Limitations and Possible Extensions

Empirical works analysing non-experimental data face several limitations and this thesis is not an exception.

From the viewpoint of theoretical background, the relevant drawback to be emphasized is that BMA analysis requires the error term of a model to be normally distributed. Even though the inspection of this condition was undertaken, it is apparent, that the tails of Q-Q plot do not follow exactly normal distribution (see Appendix). Consequently, there might occur individual values that are likely to slightly influence the outputs of BMA exercise (such as differing sign of coefficient at variable *Rewards* at BMA estimation versus at the logit model). On the other hand, vast majority of the BMA results can be assessed as fairly accurate and still remains the best alternative, considering the fact of having large number of candidate variables. As it can be seen from the literature review, BMA indicated the most powerful drivers of campaign's outcome in line with already existing research works.

Next constraint, that could hinder the generalisation of the results, is the different project policy of HitHit. As already mentioned in Chapter 2, if the target amount of money is not collected, platform allows project initiators to repay the rest of funding goal by themselves. This is a striking difference compared to Kickstarter, the largest CF platform, where such act is strictly prohibited and doing so can result in project suspension.[2] As the information on how often is such practice actually exploited on HitHit is missing, one should bear in mind that there can exist a substantial fraction of projects (or more precisely, funds), that were collected from the initiating party. Consequently, such factor could certainly influence the robustness of published results.

Another aspect that might be considered as deficient, is the number of standard explanatory variables contained in the analysed dataset. Even though most of the relevant variables were incorporated, there are still missing many more that could form the clearer picture about the drivers of the campaign. For instance, dataset contained information only about the number of the rewards, not about their form. It could be interesting, to assess if contributors incline

[2] Information is based on author's inquiry from February 25, 2020 on Kickstarter helpdesk.

more to e.g. material things. Next variable to be addressed is number of updates during the duration of project campaign. Both on HitHit (in fact, the number can be found at project's website, but the platform was not able to transfer it into the dataset), as well as on other social media, if the project initiator (such as an artist) operates on Facebook, Instagram or Twitter. And how about the overall recognition of project founder in public? Does it play important role? This blank space definitely provides an opportunity for future researchers to uncover, whether those aspects have some decisive impact on project's success.

In line with previously mentioned project updates, a natural extension of this thesis would lead to the inspection of the dynamics of the campaigns. This topic has been already investigated by foreign academia (see Mollick (2014) or Kuppuswamy & Bayus (2018)). But what does it look like in Czech environment? Is the frequency of contributing constant over time, or does its pattern somewhat change, as the end of the campaign approaches? Is it somehow linked with posting of updates on platform or social media? These concerns are also of big importance and should not be left unanswered.

In current research, most of the academic work concentrates on what happens before or during the stage of campaign. It would be more than meaningful, to trace what happens after collection of funds and distribution of rewards. As backers receive only limited information about campaign's feasibility or technical skills of project founders, they are exhibited to threat of delay in delivery of particular product (this is true mainly for technology projects, where some degree of expertise is crucial). Consequently, the gap between contribution and its possession can be larger than anticipated. So tracing what happens ex-post would certainly bring new perspective. How often are the products or services received later than expected? And does this issue occur more frequently at larger or more popular projects?

Last but not least, the performed analysis provides reader only with snapshot of the largest reward-based platform operating in the Czech environment. There also exist other platforms (such as the second largest Startovač, or other ones listed in Chapter 2) which could be an object for similar investigation. Then it would be possible to compare results across different intermediaries.

Next suggestion hanging in the air is definitely comparison of funders' motivations across different types of crowdfunding. Not only with business models, that are well-established in the Czech Republic (like P2P or P2B lending), but also with nascent ones, like donation-based crowdfunding. Particularly this subset of crowdfunding has gained a tremendous momentum stemming from current COVID-19 crisis, fuelled by solidarity from the public. For illustration, server Donio[3] collected CZK 10 million (approximately EUR 380,000) from Czech public only within one day. The campaign was devoted to production of lung ventilators for Czech hospitals. The final amount over the campaign climbed up to CZK 14,199,818 contributed by 7,483 funders. This shows that even donation-based crowdfunding (without any rewards to offer in return) might be a powerful microfinancing tool and is definitely worth of detailed inspection as well.

[3] http://www.donio.cz

Chapter 7

Conclusion and Further Implications

Reward-based crowdfunding is an innovative way of microfinancing that allows individuals or small ventures to collect funds for their project ideas. This study performs detailed analysis of 2,870 project initiatives from HitHit, the largest Czech crowdfunding platform. The data ranges from the point of platform's creation (November 2012) until January 2020. The goal is to determine which project characteristics (if any) influence the likelihood of successful campaign. On top of that, it also enriches Czech reward-based crowdfunding research by pioneer analysis from the textual point of view. By means of opinion mining and named entity recognition, which are well-recognized tools of textual analysis, this thesis is the first one in the Czech Republic that inspects whether sentiment and frequently used words extracted from project descriptions impact the probability of funding.

As there were 39 possible explanatory variables to consider, this study employs Bayesian Model Averaging method, in order to determine, which set of regressors is anticipated to affect the outcome of the project at most. The estimated results uncovered important findings, which can be transformed into general recommendations for future project creators.

The outcome of the campaign is significantly and positively affected by number of contributors, who decide to fund the campaign. The same holds for inclusion of a short video clip. On the contrary, success of the campaign negatively depends on the size of pledging goal. This is in line with previous findings indicated by foreign researchers concentrating on crowdfunding. Therefore, Czech project initiators should concentrate on adequate determination of project goal and high-quality video clip in order to increase likelihood of successful funding.

Secondly, it has shown to be true that project descriptions do not possess any predictive power in terms of likelihood of project funding. This is in contrast with existing foreign studies (see Mitra & Gilbert (2014) or Zhou *et al.* (2016)), which have proven that language factors considerably increase predictive accuracy of the model. In this viewpoint, Czech reward-based crowdfunding has proven to be distinct.

Thirdly, the findings above suggest, that there might exist other important drivers, which substantially impact the probability of success. Unfortunately, those were not covered by the model. For instance, PR of the project, activity of project initiator on social networks or the degree, to which is the founder well-known by the public. This all could result in some regain of competitive advantage, distinguishing the ideas from other projects.

All in all, these valuable findings create a groundwork for future researchers examining Czech crowdfunding environment. It would be meaningful to gather data capturing the activity of initiators on social networks, updates about the project or tracing what happens after the campaign ends. Future research should also examine the robustness of adopted method, by means of gathering different datasets from the Czech crowdfunding environment, not only reward-based, but also donation-based, which currently experiences steep upsurge.

Bibliography

Agrawal, A., C. Catalini, & A. Goldfarb (2015): "Crowdfunding: Geography, social networks, and the timing of investment decisions." *Journal of Economics & Management Strategy* **24(2)**: pp. 253–274.

Alegre, I. & M. Moleskis (2016): "Crowdfunding: A review and research agenda." .

Allison, T. H., B. C. Davis, J. C. Short, & J. W. Webb (2015): "Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues." *Entrepreneurship Theory and Practice* **39(1)**: pp. 53–73.

Bachmann, A., A. Becker, D. Buerckner, & M. K. Hilker (2011): "Online peer-to-peer lending - a literature review." *Journal of Internet Banking and Commerce* **16(2)**: pp. 1–18.

Belleflamme, P., T. Lambert, & A. Schwienbacher (2013): "Individual crowdfunding practices." *Venture Capital* **15(4)**: pp. 313–333.

Belleflamme, P., T. Lambert, & A. Schwienbacher (2014): "Crowdfunding: Tapping the right crowd." *Journal of business venturing* **29(5)**: pp. 585–609.

Belleflamme, P., N. Omrani, & M. Peitz (2015): "The economics of crowdfunding platforms." *Information Economics and Policy* **33**: pp. 11–28.

Boček, J. (2015): "Projekty na Hithitu a Startovači získaly prvních 100 milionů korun. podívejte se, kdo uspěl." Accessed on February 2, 2020.

Chan, C. S., H. Park, P. Patel, & D. Gomulya (2018): "Reward-based crowdfunding success: decomposition of the project, product category, entrepreneur, and location effects." *Venture Capital* .

Cordova, A., J. Dolci, & G. Gianfrate (2015): "The determinants of crowdfunding success: evidence from technology projects." *Procedia-Social and Behavioral Sciences* **181**: pp. 115–124.

Crosetto, P. & T. Regner (2014): "Crowdfunding: Determinants of success and funding dynamics." *Technical report*, Jena Economic Research Papers.

EC (2016): "Report in crowdfunding in the eu capital markets union." .

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default priors and predictive performance in bayesian model averaging, with application to growth determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Falci, E. (2017): "Debunking natural language processing." .

Felipe, I. J. d. S., W. Mendes-Da-Silva, & C. C. Gattaz (2017): "Crowdfunding research agenda." In "2017 IEEE 11th International Conference on Semantic Computing (ICSC)," pp. 459–464. IEEE.

Fleming, L. & O. Sorenson (2016): "Financing by and for the Masses: An Introduction to the Special Issue on Crowdfunding." *California Management Review* **58(2)**: pp. 5–19.

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

Gleasure, R. & J. Feller (2016): "Emerging technologies and the democratisation of financial services: A metatriangulation of crowdfunding research."

Gorbatai, A. D. & L. Nelson (2015): "Gender and the language of crowdfunding." In "Academy of Management Proceedings," volume 2015, p. 15785. Academy of Management Briarcliff Manor, NY 10510.

Guilliatt, R. (1988): "Australian Dealmaker: John Cornell The Man Who Sold Hollywood on Crocodile Dundee." *The New York Times* .

Hana, J. (2015): "Textové zlatkopectví." Accessed on February 28, 2020.

Hobbs, J., G. Grigore, & M. Molesworth (2016): "Success in the management of crowdfunding projects in the creative industries." *Internet Research* **26(1)**: pp. 146–166.

HOSSAIN, M. & G. OPARAOCHA (2017): "Crowdfunding: Motives, definitions, typology and ethical challenges." *Entrepreneurship Research Journal* **7**.

HU, M., X. LI, & M. SHI (2015): "Product and pricing decisions in crowdfunding." *Marketing Science* **34(3)**: pp. 331–345.

KOCH, J.-A. & M. SIERING (2015): "Crowdfunding success factors: the characteristics of successfully funded projects on crowdfunding platforms." .

KSHETRI, N. (2015): "Success of crowd-based online technology in fundraising: An institutional perspective." *Journal of International Management* **21(2)**: pp. 100–116.

KUPPUSWAMY, V. & B. L. BAYUS (2018): "Crowdfunding creative ideas: The dynamics of project backers." In "The Economics of Crowdfunding," pp. 151–182. Springer.

KUTI, M. & G. MADARÁSZ (2014): "Crowdfunding."

LEAMER, E. E. (1978): *Specification Searches: Ad Hoc Inference With Nonexperimental Data.* New York: Wiley.

LEIMEISTER, J. & U. BRETSCHNEIDER (2017): "Not just an ego trip: Exploring backers' motivation for funding in incentive-based crowdfunding." *Journal of Strategic Information Systems (JSIS)* .

LEY, E. & M. STEEL (2007): "On the effect of prior assumptions in bma with applications to growth regression." *Journal of Applied Econometrics* .

LIDDY, E. D. (2001): "Natural language processing." .

MITRA, T. & E. GILBERT (2014): "The language that gets people to give: Phrases that predict success on kickstarter." *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.* .

MOLLICK, E. (2014): "The dynamics of crowdfunding: An exploratory study." *Journal of business venturing* **29(1)**: pp. 1–16.

MORAL-BENITO, E. (2013): "Model averaging in economics: An overview." *Journal of Economic Surveys* **29**.

MORITZ, A. & J. H. BLOCK (2016): "Crowdfunding: A literature review and research directions." In "Crowdfunding in Europe," pp. 25–53. Springer.

NADKARNI, P. M., L. OHNO-MACHADO, & W. W. CHAPMAN (2011): "Natural language processing: an introduction." *Journal of the American Medical Informatics Association* **18(5)**: pp. 544–551.

ORDANINI, A., L. MICELI, M. PIZZETTI, & A. PARASURAMAN (2011): "Crowd-funding: Transforming Customers into Investors through Innovative Service Platforms." *Journal of Service Management* **22(4)**: pp. 443–470.

ROUSE, M. (2016): "Special report: Artificial intelligence apps come of age."

ŠOLTÉS, M. & T. ŠTOFA (2016): "Crowdfunding–the case of slovakia and the czech republic." *Quality Innovation Prosperity* **20(2)**: pp. 89–104.

STASIK, A. & E. WILCZYŃSKA (2018): "How Do We Study Crowdfunding? an Overview of Methods and Introduction to New Research Agenda." *Journal of Management and Business Administration. Central Europe* **26(1)**: pp. 49–78.

STEEL, M. F. (2011): "Bayesian model averaging and forecasting." *Bulletin of EU and US Inflation and Macroeconomic Analysis* **200**: pp. 30–41.

STEEL, M. F. (2017): "Model averaging and its use in economics." *arXiv preprint arXiv:1709.08221* .

STEIGENBERGER, N. (2017): "Why supporters contribute to reward-based crowdfunding." *International Journal of Entrepreneurial Behavior & Research* .

SUN, S., C. LUO, & J. CHEN (2016): "A review of natural language processing techniques for opinion mining systems." *Information Fusion* **36**.

WANG, W., K. ZHU, H. WANG, & Y. WU (2017): "Impact of sentimental factor on the successful crowdfunding campaigns: A text mining approach." *IET Software* **11**.

WOOLDRIDGE, J. (2015): *Introductory Econometrics: A Modern Approach.* Nelson Education.

ZEUGNER, S. (2011): "Bayesian model averaging with bms." *Tutorial to the R-package BMS* .

Zhang, H. & W. Chen (2019): “Backer motivation in crowdfunding new product ideas: Is it about you or is it about me?” *Journal of Product Innovation Management* **36(2)**: pp. 241–262.

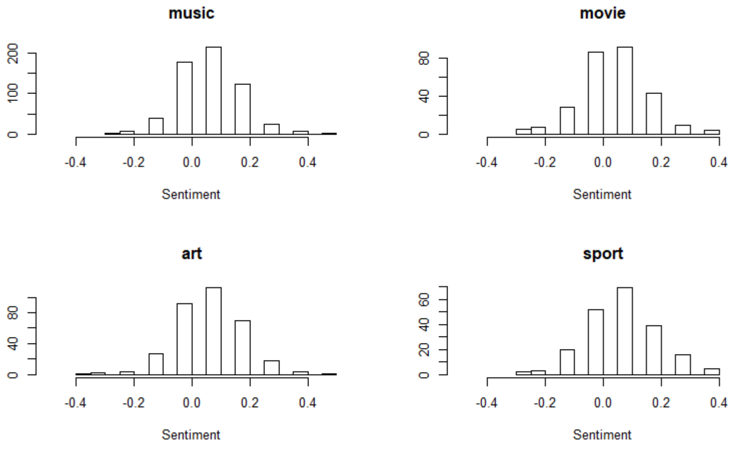
Zhou, M. J., B. Lu, W. P. Fan, & G. A. Wang (2016): “Project description and crowdfunding success: an exploratory study.” *Information Systems Frontiers* **20(2)**: pp. 259–274.

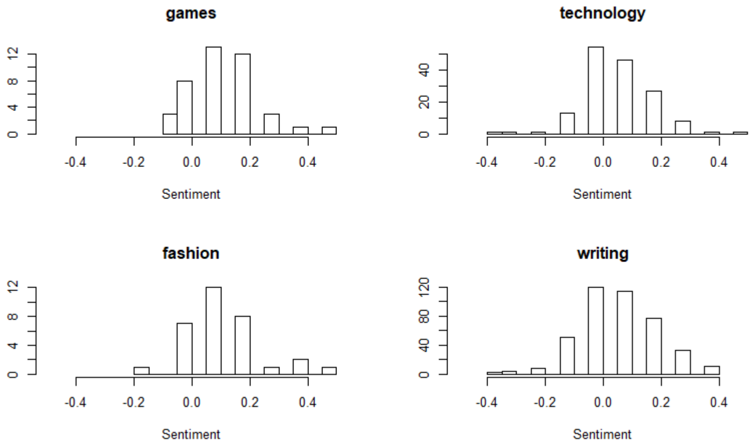
Ziegler, T., R. Shneor, K. Garvey, & K. Wenzlaff (2018): “Shifting paradigms: The 4th european alternative finance benchmarking report.” .

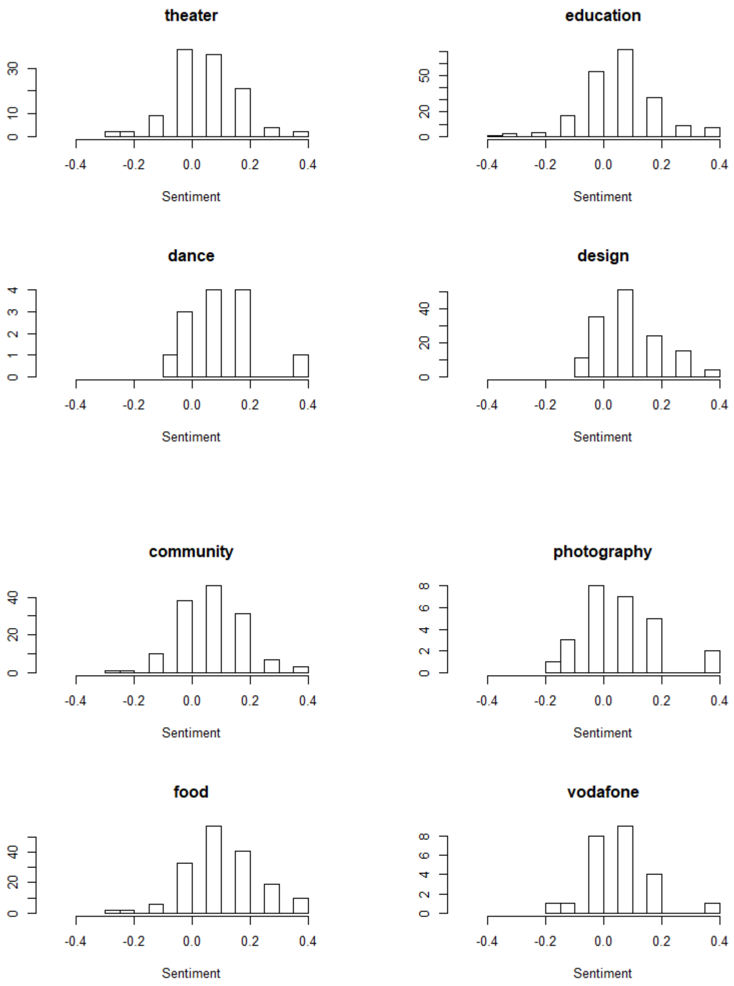
Appendix A

Appendix

Distribution of Sentiment Across Categories

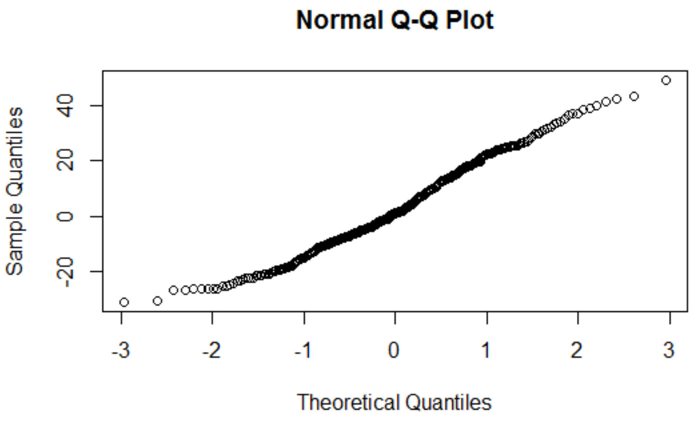






Source: Author's own computations using extracted data, *N=2,870*

Q-Q Plot for Logistic Regression



Source: Author's own computations using extracted data, *N=2,870*