#2109 Bremen Papers on Economics & Innovation

Do not judge a business idea by its cover: The relation between topics in business ideas and incorporation probability

Jessica Birkholz*

December 2021

Abstract
It is of key importance to identify innovative business ideas in an early stage, so that funding resources can be adequately allocated according to their economic potential. Traditional indicators do not reliably discriminate business ideas with high degree of innovativeness and high incorporation chances from those with low degree of innovativeness and low incorporation prospects. Therefore, this paper examines the content of business idea descriptions to improve the estimations of the incorporation probability. The paper aims to answer two questions: 1) Are there differences in topic prevalence in innovative and non-innovative business ideas?, and 2) How does the composition of topics related to a business idea influence its incorporation probability? Structural topic modeling and classification tree analysis are applied on business idea descriptions from a competition in Bremen, Germany, from 2003 until 2019. The results show that business idea descriptions are a rich source of information to identify both innovative ideas and those with higher incorporation prospects.

Keywords
Innovative entrepreneurship, business idea, machine learning, incorporation

JEL Classifications
O30; L26 ; O38

* University of Bremen, Max-von-Laue-Straße 1,
28359 Bremen, Germany
Corresponding author: Jessica Birkholz
e-mail: jessica.birkholz@uni-bremen.de
ORCID: 0000-0003-4252-2184

1. Introduction

Entrepreneurship is widely acknowledged to be of key importance for economic development. Thus, it is not surprising that the stimulation of entrepreneurship is on top of the agenda for policy makers. Believing all entrepreneurial activities support economic welfare, policy measures to stimulate entrepreneurship follow a non-selective funding practice. For example, in Germany programs to foster entrepreneurship are targeted at broad groups of people ranging from unemployed to university graduates to young companies. Moreover, all stages of the entrepreneurial process - ranging from the ideation to the operational phase - are supported (see for example BMWi, 2018). These programs include funding, credits and consultation services.

Although these policy measures are appreciated by experts evaluating the German ecosystem for entrepreneurial activities (Sternberg et al., 2020), the assumption that this funding practice leads to the best entrepreneurship outcomes is flawed. Not all entrepreneurs contribute to economic development in the same proportion. Only a small fraction of entrepreneurial activities are responsible for increases in labour opportunities, wealth, and the development of entrepreneurial role models (Morris, 2011). Thus, a targeted stimulation of promising entrepreneurial activities is desirable.

Thus, the challenge is to identify at an early stage the entrepreneurial activities that lead to innovation and successful incorporation. Former research concentrated on characteristics of the entrepreneur or of the general business model to explain differences in innovation and incorporation probability. However, many indicators (e.g., characteristics of the entrepreneur) cannot reliably be attributed to a specific type of business (Morris, 2011). Thus, an ex-ante identification of promising business ideas cannot be achieved based solely on these meta-indicators. This is why many authors call for more research in this vein (Gundry et al., 2016; McAdam & Marlow, 2011; Meyskens & Carsrud, 2013; Pandey et al., 2017). Especially, this research call focusses on business ideas itself and a more diversified set of indicators to identify successful and innovative ventures (McAdam & Marlow, 2011; Meyskens & Carsrud, 2013; Pandey et al., 2017; Pare et al., 2011).

The purpose of this study is to examine the relation between content of innovative and non-innovative business ideas and the probability of a business idea turning into a new venture operating on the market, which will be referred to as the incorporation probability in the following. The content, the written description of a business idea, contains different topics (e.g., customer segment, main product). To fulfil the purpose of this study, data from a business idea competition in Bremen, Germany in the time period from 2003 until 2019 is used. 247 text descriptions classified as innovative business ideas and 182 text descriptions of non-innovative business ideas were handed in to the

competition. Machine learning techniques are applied to answer two questions: 1) Are there differences in topic prevalence in innovative and non-innovative business ideas? And 2) How does the composition of topics influence the incorporation probability? To answer these, structural topic modeling (Roberts et al., 2014) and classification tree analysis (James et al., 2013; Strobl et al., 2009) are employed.

The analysis yields four main results. First, differences in topic prevalences, the frequencies every topic appears in a text corpus, in innovative business ideas and non-innovative ones are identified. Topics connected to the products and key processes for revenue generation, intention and orientation type of the business occur in different frequency in innovative and non-innovative business ideas. Second, the interdependence of topics and their importance for incorporation probability differ depending on the degree of innovativeness of the business ideas. Third, innovative entrepreneurship is not a homogeneous group of entrepreneurial activities. Besides technology entrepreneurship, there exists a hybrid entrepreneurship that combines the development of new technologies with social purposes. Fourth, the conventional wisdom that innovative business ideas have lower incorporation rates due to higher risks or challenges, cannot be confirmed by this study. Rather, the data suggests that the incorporation probability depends on the structure and combination of topics. Even more, the topic structure of a business idea proves to be more important than traditional indicators like, for example, team size.

Besides these results, the study contributes in three aspects. First, it delivers conclusions about the underexplored pre-foundation stage of entrepreneurial activities and a more detailed understanding of the differences between innovative and non-innovative business ideas. Second, the pre-foundational information is connected to incorporation probability rather than entrepreneurial intentions. This translates into an application-oriented approach, providing valuable insights for the design of targeted entrepreneurship policies. Third, it develops a strategy to gain insights about the incorporation potential directly out of non-numeric business idea descriptions. Traditional indicators do not – or only indirectly – measure this potential and profit from being complemented by this information.

The remainder of this paper is organized as follows: the next chapter discusses the theoretical background of the entrepreneurship stimulation policies and the propositions of this study are presented. In Chapter three the business idea competition data base for the analysis is described. The methodological approach taken by this study, consisting out of structural topic modeling and classification tree analysis, is outlined in Chapter four. The empirical findings are presented in Chapter five. The paper entails a discussion of the results in Chapter six and concludes in Chapter seven.

2. State of the art

2.1. The problem of funding entrepreneurship

There is good reason to support entrepreneurial activities. Entrepreneurs often take on issues that the government cannot tackle and open new development paths to counteract societal and environmental challenges. Furthermore, they create wealth and jobs (Asarkaya & Keles Taysir, 2019; Audretsch & Keilbach, 2004; Autio & Yli-Renko, 1998). Despite these positive effects, the goal of many policy measures to stimulate as many entrepreneurial activities as possible is unlikely to contribute to economic development (Shane, 2009).

The approach of a non-selective funding practice of entrepreneurial activities has a main difficulty: Advancements in economic development stem from highly innovative and technology based entrepreneurial endeavours and cannot be attributed to the typical new venture (Autio & Yli-Renko, 1998; Shane, 2009). Three observations underline this statement. First, evidence suggests that most new ventures try to gain access to an existing market with already existing products and services entering competitive industries with high failure rates (Hurst & Pugsley, 2011; Johnson, 2004; Shane, 2009). This is also reflected in current findings on the German case that highlight that economic aspirations like increased income, which is a driving force for profitability and growth of ventures play only a minor role in motivating potential entrepreneurs (Sternberg et al., 2020). Second, the typical new venture does not innovate by any measurable margin (Hurst & Pugsley, 2011). Although a large share of innovative activities are done by small businesses, it is not the case that the typical new and small venture is innovative (Hurst & Pugsley, 2011). In Germany only 10.3% of all entrepreneurial projects can be classified as innovative (Sternberg et al., 2020).[1] Third, jobs created by new ventures only cover a small fraction of total job creation compared to established firms (Shane, 2009). Because new ventures often fail to survive the first years, the jobs created by them are also less stable and offer less payment than the ones created by established ventures (Shane, 2009; Wagner, 1997). In Germany only a quarter of new ventures are in their growth stage and thus create a considerable amount of jobs (Kollmann et al., 2020). These observations mount up to the macroeconomic phenomenon that countries with faster economic growth show declining firm formation rates (Shane, 2009). In summary, policy measures that aim at increasing only the amount of entrepreneurial activities disproportionally support entrepreneurs without the prospect of contributing to economic development (Shane, 2009).

[1] An entrepreneurial project is defined as innovative when it aims at sectors with middle to high technology intensity (Sternberg et al., 2020).

2.2. The problem to classify entrepreneurial activities ex-ante

Former literature identified factors influencing the probability of the success or failure of entrepreneurial endeavours. Individual characteristics of the entrepreneur, available resources, and intangible assets were under investigation (Pare et al., 2011; Rasmussen & Sorheim, 2012; Song et al., 2008).

Characteristics of successful entrepreneurs are higher experienced self-efficiency (Klofsten, 2005; Markman et al., 2002), an effectuation mindset (Klofsten, 2005; Sarasvathy, 2003) and a balanced set of skills (Stuetzer et al., 2013). A balanced set of skills is connected to an entrepreneurial personality profile and an early interest in an entrepreneurial career (Stuetzer et al., 2013). Interestingly, balanced skills are more important in innovative entrepreneurial activities than in non-innovative ones (Stuetzer et al., 2013). Work and industry experience appear more frequently in the case of innovative entrepreneurial activities (Friar & Meyer, 2003). Although these criteria may hint at the innovativeness of an entrepreneurial project and its potential to be incorporated, these indicators have two difficulties: 1) some are difficult to assess (self-efficiency, effectuation mindset, balanced skills, entrepreneurial personality profile) and would require to collect considerable amount of data 2) and the easy-to-measure traditional human capital indicators do not show a strong link to outcomes in the case of nascent entrepreneurship (Davidsson & Gordon, 2012; Stuetzer et al., 2013).

Available resources are especially important for innovative entrepreneurial endeavours to counteract the liability of newness. Founding a venture within a team is a strategy to cope with constraints in resource availabilities (Munoz-Bullon et al., 2015). It has been shown that innovative entrepreneurial activities are more frequently developed by teams rather than solo entrepreneurs (Friar & Meyer, 2003). Besides a work environment that fosters creativity, the team may be extended by partners bringing in new perspectives and ideas (Gundry et al., 2016; Munoz-Bullon et al., 2015; Shalley et al., 2009; West, 2007). Nevertheless, the ability to attract external partners is not a useful success predictor in early phases of the entrepreneurial process but rather comes into play at later stages in the business idea development. Moreover, the presence of partners and a large team size does not necessarily contribute to venture development or innovation because, for example, internal conflicts are more likely to arise in bigger teams (Meyskens & Carsrud, 2013).

Especially when external funding is needed, intangible assets in regard to communication are required for the success of a business idea. Here, the vision of the entrepreneur, to convey the essence of the business opportunity as well as the sustainability and growth potential of the business idea, plays an important role (McAdam & Marlow, 2011). Central aspects in this regard are preparedness of the entrepreneurs (Chen et al., 2009; Rasmussen & Sorheim, 2012), a promotion focus and perceived feasibility (Drnovsek et al., 2018). These indicators influence the investment decisions of

investors into potential and small ventures. However, in case of innovative entrepreneurial projects these indicators are more difficult to communicate and to estimate, thus, information asymmetries are still a burden to access external funding (Proksch et al., 2017).

The aforementioned indicators have considerable shortcomings. The subjectivity of the indicators makes it difficult to measure them reliably. Moreover, these factors may differ in their importance for innovative business ideas compared to typical business ideas and are even more difficult to assess due to the liability of newness and smallness (McAdam & Marlow, 2011). Although their widely acknowledged importance, criteria concerning markets and opportunities are not thoroughly investigated in former research (Pare et al., 2011; Song et al., 2008).

2.3. Why ideas matter and can be used as a source of information

The ex-ante identification of business ideas which are likely to turn into new innovative ventures is hardly possible in a reliable manner by analysing the indicators presented in the last section. What is required is information attributable to a specific business idea in its core rather than indicators describing the framework of a business idea. This emphasises a formerly understated factor for success or failure of innovative business ideas – their content (McAdam & Marlow, 2011; Rasmussen & Sorheim, 2012). The content, the written description of business ideas, complements the framework indicators to achieve a more reliable prediction of business developments.

Business idea descriptions provide a rich resource for information in early phases of the entrepreneurial process (Pare et al., 2011). A business idea can be defined as a description of how a business works and how revenue is generated (Klofsten, 2005). Thus, it is very closely related to the vision of the entrepreneur and mission statements of organisations, which play a central role in organisational strategy (Kirk & Nolan, 2010; Pandey et al., 2017), stakeholder perceptions (Pandey et al., 2017), and activities and ideology of the organisation (Fox, 2006; Pandey et al., 2017). Thus, business ideas provide information, which is likely to lay the foundation for subsequent developments and performance of the incorporation (Pandey et al., 2017; Pare et al., 2011). Moreover, business idea descriptions are a decisive factor for funding allocations. A combination of business idea characteristics and the observation if the business idea is incorporated to a later point in time, provides application-oriented results for designing early entrepreneurship funding programs (Chattopadhyay & Ghosh, 2008). However, studies on the content of business ideas are rare. Thus, this study addresses two research questions. The first aims to identify differences in the content between innovative and non-innovative business ideas. Research question two is dealing with the influence the content structure has on the incorporation probability of business ideas.

2.3.1. Differences between innovative and non-innovative business ideas in regard to prevalent topics

To be able to make use of the business idea content as an unbiased indicator for innovative ideas, it is of key importance to identify topics, i.e. subparts of the content, within the business ideas which allow a distinction between innovative and non-innovative business ideas. Former research found evidence for differences between innovative and non-innovative entrepreneurial projects in entrepreneur´s characteristics like balanced skill set (Stuetzer et al., 2013), industry experiences (Friar & Meyer, 2003) and mindset (Klofsten, 2005). Furthermore, venture´s characteristics like team size (Friar & Meyer, 2003) and cooperation with external partners (Munoz-Bullon et al., 2015) were found to differ. These findings support the assumption that there are topics central to innovative business ideas, which play only a minor role in non-innovative business ideas and vice versa. This is likely to translate into different topic proportions in the business ideas. In addition, it can be argued that topics are combined with each other in a different manner in innovative or non-innovative business ideas. Thus, it is suspected that the relations between topics differ with the degree of innovativeness of the business ideas. Thus, two proposition are formulated as follows:

Proposition 1: There are differences between topic proportions in innovative and non-innovative business ideas.

Proposition 2: Correlations between topics associated with innovative and non-innovative business ideas differ.

2.3.2. Relation of topics and incorporation probability of business ideas

The second unanswered issue is the relevance of each topic to the incorporation probability of the business idea (McAdam & Marlow, 2011; Meyskens & Carsrud, 2013; Pandey et al., 2017). It can be hypothesised that determinants of incorporation differ with the degree of innovativeness, since Schumpeter (1934), describing the innovative entrepreneur, and Kirzner (1973), describing the entrepreneur aiming at profit through arbitrage, focus on different characteristics. The innovative entrepreneur described by Schumpeter (1934) is characterised by specific personality traits and attitudes. In contrast to this, the not-innovative entrepreneur as described by Kirzner (1973) is interpreted as an individual attentive to its environment. These considerations highlight that the two types of entrepreneurship have different requirements to be successful. Although widely acknowledged that there is not a one-size-fits-all solution to transform a business idea into a new venture, it remains unclear which topics serve as determinants of incorporation probability depending on a specific composition of a business idea. Thus, two propositions are stated as follows:

Proposition 3: The combination of topics influence the incorporation probability of business ideas.

Proposition 4: The determinant topics for the incorporation probability differ for innovative and non-innovative business ideas.

3. Data

This paper makes use of data gathered by "BRIDGE", which provides consultation for anyone interested in business foundation coming from the academic environment of Bremen and Bremerhaven. "BRIDGE" offers individual or team coaching as well as workshop series. Once a year, they organise a business idea competition called "CampusIdeen". This event is taking place since 2003 continuing until 2019. Submissions to this event contain a detailed description of business ideas along guiding questions provided by "BRIDGE". This text data offers a description of ideas before their incorporation and therefore offers rarely existing pre-foundation information. The consistent design of the competition and its long-time of existence favour building a coherent data base for further analysis based on submissions.

3.1. Data preparation

The texts submitted to the "CampusIdeen" competition were filtered in a four-step procedure. First, the submissions were checked manually to ensure only serious submissions are considered in the analysis.[2] In addition, multiple submissions of an idea of the same entrepreneur in multiple years were deleted from the data set, keeping only the first handed in. Second, Curriculum Vitaes were removed from the submission texts since the structure of these cannot be compared to continuous text and, therefore, may bias the results. To avoid overrepresentation of words and formulations appearing in the submission form of "BRIDGE", guiding questions as well as headlines were removed from the submission texts. Third, stop words, i.e. words delivering no meaning, are removed. To do so, the ISO stop word list of the German language was applied. The remaining text data was lemmatized. Words classified as adposition, auxiliary, coordinating conjunction, determiner, numeral, particles, pronouns, subordinating conjunction, adverb, interjection and auxiliary verbs[3] were added to the list of stop words.[4] Fourth, words appearing in over 90% of the submissions and words appearing in less than 10% of the business ideas are removed, because otherwise these could bias

[2] A submission is not considered serious, when the text consists of less than half a page and not all questions in the submission form are answered.

[3] Including modal verbs as subgroup of auxiliary verbs.

[4] Since lemmatisation for the German language is not entirely classifying words in a correct manner, these words were checked manually.

the results of the topic model (Birkholz et al., 2021). The retrieved unigrams build the basis for the subsequent topic model.[5]

Table 3.1 shows the mean and median number of words per submission differentiated by their degree of innovativeness. These indicate longer submission texts for innovative business ideas than for non-innovative ones. The similarity of the standard deviation in both cases suggests that the difference in submission length is not caused by greater variety of word usage. More submissions are classified as innovative which results in a proportionally higher number of total words considered.

Table 3.1: Descriptive statistics of the text descriptions of the submitted business ideas
Source: Calculation based on business ideas submitted to "CampusIdeen" competition from 2005-2019 with a jury rating of newness with higher than average score (innovative business ideas) and lower than average score (non-innovative business ideas)

	Innovative business ideas	**Non-innovative business ideas**
Mean count of words per submission	746.324	699.467
Median count of words per submission	708	658
Standard deviation of the word count per submission	434.0143	435.6531
Sum of words over all submissions	184,342	127,303
Number of submissions	247	182

3.2. Degree of innovativeness of business ideas

Submissions from 2005 onward are rated by an expert jury evaluating their degree of newness. This rating serves as a basis to classify the submissions. A rating above average in newness is interpreted as innovative and a rating below average in this criterion is associated with non-innovative business ideas.[6]

Figure 3.1 illustrates the number of submissions to the competition, separated in innovative and non-innovative, in each year and for each team size. No clear time trend

[5] The alternative use of n-grams would come with the risk to create artificial word combinations resulting from removing stop words from the text corpus. Moreover, bigrams and unigrams appear to be highly correlated to each other in former research, meaning that the use of bigrams does not provide a substantial improvement compared to the conservative approach to make use of unigrams (Braga et al., 2009).

[6] The submissions were rated on a 5-point-Likert scale from one or more jury members. The arithmetic mean of all available jury ratings was used in case of more than one evaluation. The mean value (2.95 on the Likert-scale) across all submissions in all years was used to classify the business idea submissions as above or below average innovative.

in the number of submissions, or their degree of innovativeness is evident.[7] The vast majority of submissions are handed in by individuals and teams of two. In neither of these two graphics a year or a specific team size is associated with especially high or low shares of innovative ideas.

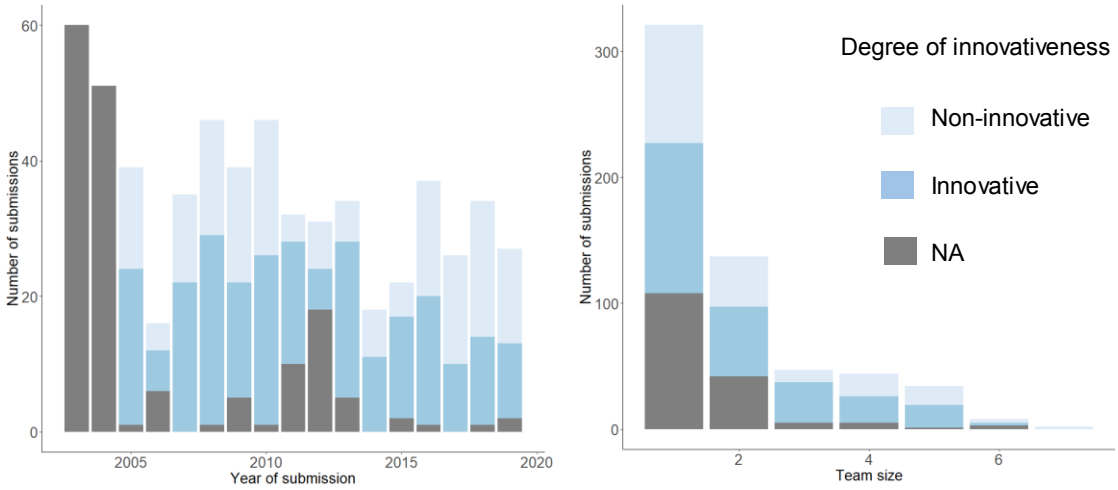


Figure 3.1: Number of innovative and non-innovative submissions to the business idea competition per year (left Figure) and per team size (right Figure)
Source: "CampusIdeen" competition submissions 2003-2019

3.3. Incorporation probability of business ideas

Whether a business idea came into existence or not is retrieved by manually checking company websites, social media profiles of the entrepreneurs and information the transfer office had. The data collection was done in September 2020, which corresponds to one year after the last business idea competition.[8] The incorporation rate is in both cases – innovative and non-innovative business ideas – approximately 25% as shown in Figure 3.2.

[7] The NAs in the Figure correspond to missing data on the rating of newness. Since all submissions can be used in the first analysis step, independently of the availability of the rating of the newness, these data points are also displayed in the Figures.
[8] The timeframe of one year is chosen because it becomes unlikely for a business to be founded after this period (Capelleras & Greene, 2008; Carter et al., 1996).

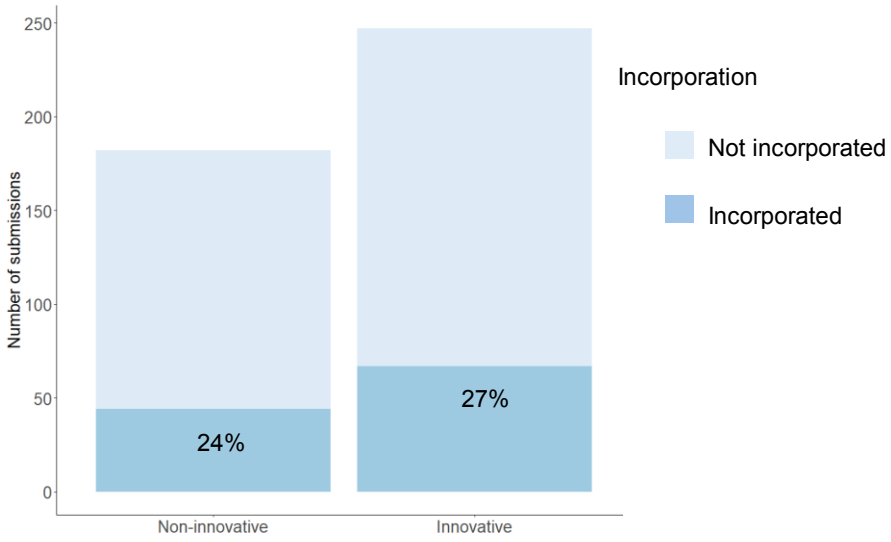


Figure 3.2: Incorporation probability of the business ideas in the data set
Source: "CampusIdeen" competition submissions 2005-2019 complemented with incorporation information as retrieved in September 2020

4. Methodological approach

In order to test the propositions, a four-step analytical procedure is applied. The research design as illustrated in Figure 4.1 proceeds after the data filtering process with structural topic modeling to reveal latent topics in business ideas. Since the topic modeling approach does not require information about the degree of innovativeness, all submission texts, amounting to 593 documents with a total of 400,854 words, are utilised for this step. In a second step, innovative and non-innovative business ideas are compared to identify differences in topic proportions. As this step requires complete information about the meta-data of a business idea, due to incompleteness of submission data, in total 429 documents can be considered. Third, the correlations between the topics are analysed to reveal the topic compositions of innovative and non-innovative business ideas. The data used for this step is equivalent to the former step. Fourth, the results of the topic model are employed to estimate two classification trees – one explaining the foundation success of innovative ideas and one explaining the foundation success of non-innovative ideas. This last step makes usage of meta-data of the business ideas similarly as in the second and third step, thus again 429 documents can be analysed.

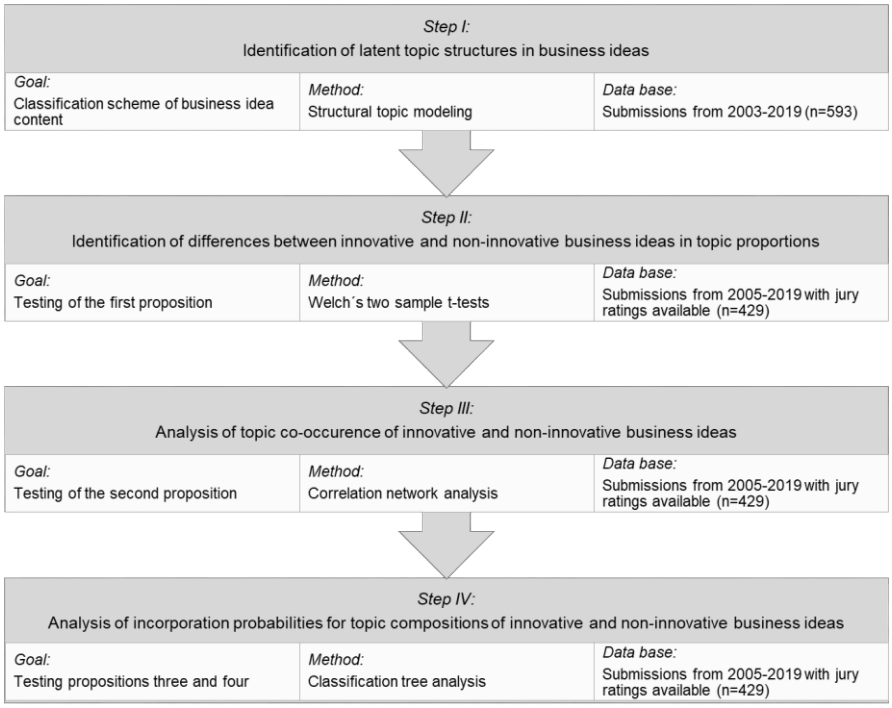


Figure 4.1: Research design of the study
Note: The data base in step two and three starts with the year 2005 because in the years 2003 and 2004, no jury ratings are available in the data set

4.1. Topic Modeling

Topic modeling as an unsupervised machine learning technique, makes identification of thematic structures in text data possible without the need of human assessment and thus is able to generate text classifications uninfluenced by individual researchers (Storopoli, 2019). Latent Dirichlet Allocation (LDA) is one of the most common algorithms to perform topic modeling and was introduced by Blei, Ng and Jordan in 2003. The procedure is based on the co-occurrence of words in documents without taking grammatical structures into account. If two words are co-occurring together more often than they would in a random collocation of words, the probability based algorithms evaluate these words as connected to each other (Birkholz et al., 2021). This approach is extended by Roberts et al. (2014) to structural topic modeling (STM). In contrast to the LDA algorithm, the STM procedure enables an incorporation of meta-data of the analysed documents.

For the creation of the topic model, this paper controls for two aspects. First, the year of submission is taken into account. It is reasonable to assume that specific topics are driven by trends. For example, business ideas in context of the development of apps may be rare in earlier years of the competition, growing in importance only in the latest years. The prevalence of such a topic is not especially insightful since it depends more on the technological progress coming along with the years rather than a shift in

entrepreneur's minds. Second, the team size is considered when creating the topic model. Business ideas written by a bigger team, are likely to describe the synergies between team members and their individual qualifications more extensively than individually written submissions.

Although topic modeling does not rely on many a priori defined indicators, it is necessary to determine a suitable number of topics. Following the advice of Chang et al. (2009), semantic coherence and exclusivity[9] are considered to determine an appropriate number of topics. Figure 4.2 shows the indicators over a range of number of topics.[10] Both plots show a salient point around ten topics. In Figure 9.1, in the appendix, the semantic coherence and exclusivity of each topic in each topic model is plotted against each other. High exclusivity and high semantic coherence is considered to be ideal (Chang et al., 2009). The topic model containing eight topics fulfils this to the largest extent.[11]

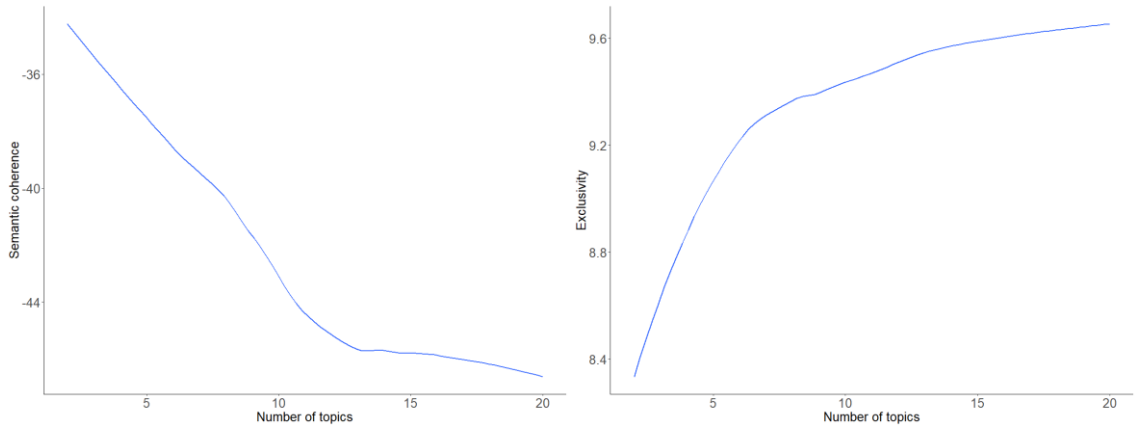


Figure 4.2: Semantic coherence (on the left side) and exclusivity (on the right side) per number of topics
Source: Own calculation

In order to validate the first and the second proposition, topic proportions within innovative business ideas are compared to these in non-innovative ones. Welch´s two sample t-tests (Welch, 1947) are applied to identify significant differences in proportions.

[9] Semantic coherence reflects if a human being would divide the topics in a similar way and exclusivity expresses the uniqueness of the words a topic consist of (Silge 2018; Chang et al. 2009).

[10] A range from two up to twenty topics were considered. More topics than that did not seem to be appropriate in light of the sample size of 593 business ideas. Furthermore, the trend in the Figure 4.2 appears to show a clear trend, thus the inclusion of more topics would lead to the same result.

[11] The displayed topic models in the plots in Figure 9.1 refer to the pareto optimal one out of 50 runs as incorporated by the R package "stm", function "manyTopics". See Figure 9.2 in the appendix for the distribution of the topics over the documents, Table 9.2 for the correlation of topics in their occurrence and Table 9.1 for the similarity of topics measured by the cosine index.

In a second step, correlation networks are calculated to reveal the topic structures of innovative and non-innovative business ideas.

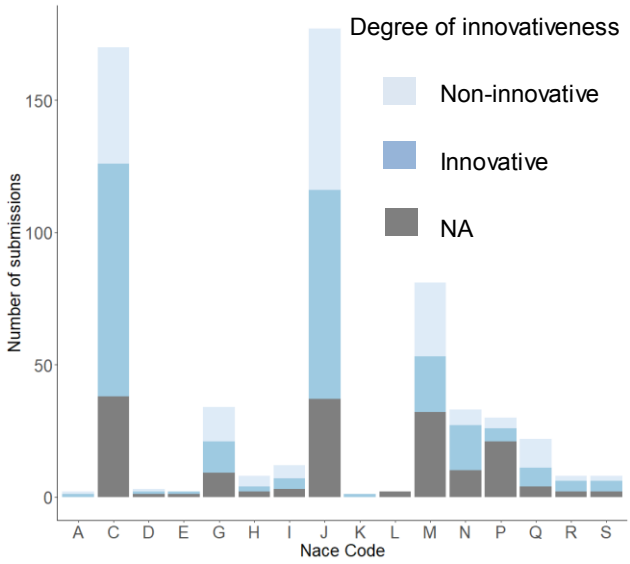
4.2. Classification tree analysis

The eight topics identified by the former analytical step, are implemented in the estimation of classification trees to retrieve their impact on the incorporation probability of business ideas. Every business idea has a probability to contain a topic. For example a business idea can have a 0.10 expected share of topic 1, 0.25 expected share of topic 2, and so on amounting over all topics to 1. These data points are now merged with meta-data of the business idea to obtain a full data set. In order to validate the third and fourth proposition, one classification tree is estimated for the subset of innovative business ideas and one classification tree for the subset of non-innovative ideas.

Logistic and probabilistic regression models have a shortcoming. The result of these models allows the reader to identify the factors with the biggest influence on the outcome, but it does not reflect human decision making in an easily transferable manner and, thus, does not clearly lead to advice on policy design (James et al., 2013; Strobl et al., 2009). Classification trees are an approach to address this issue. By dividing the predictor space into non-overlapping regions guided by the classification error rate, the classification tree approach aims at the identification of combinations of factors leading in this study to a prediction of incorporation or no incorporation of business ideas with this combination (Milanovic Glavan et al., 2015). Additionally, classification trees are able to handle qualitative predictors without the requirement of dummy variables.

4.2.1. Control variables

To reveal the relationship between topics and incorporation probability, control variables describing the industrial environment are included. The distribution of the business idea submissions over the NACE classification of industries is given in Figure 4.3. The most prominent domains are manufacturing (C), and information and communication (J).



Legend of NACE Codes

(A) Agriculture, forestry and fishing
(C) Manufacturing
(D) Electricity, gas, steam and air conditioning supply
(E) Water supply; sewerage, waste management and remediation activities
(G) Wholesale and retail trade; repair of motor vehicles and motorcycles
(H) Transportation and storage
(I) Accommodation and food service activities
(J) Information and communication
(K) Financial and insurance activities
(L) Real estate activities
(M) Professional, scientific and technical activities
(N) Administrative and support service activities
(P) Education
(Q) Human health and social work activities
(R) Arts, entertainment and recreation
(S) Other service activities

Figure 4.3: Distribution of business ideas over industries classifications (NACE Codes)
Note: The industry classification are assigned to the business ideas with manual evaluation of each submission text
Source: "CampusIdeen" competition submissions 2005-2019

A further set of control variables are jury ratings describing the quality of the business ideas. The data shows that innovative business ideas reach higher mean scores in conclusiveness, realistic implementation, and economic potential than the non-innovative business ideas, as shown in Table 4.1. Furthermore, the variety in innovative business ideas is higher reflected by the higher standard deviations in all variables.

Table 4.1: Descriptive statistics of jury ratings of business ideas
Source: "CampusIdeen" competition submissions 2005-2019

Statistic	N	Mean	St. Dev.	Min	Pctl (25)	Pctl (75)	Max
Non-innovative business ideas							
Conclusiveness	182	2.363	0.835	1	2	3	4.5
Realistic Implementation	182	2.565	0.924	1	2	3	5
Economic Potential	182	2.052	0.736	1	1.5	2.5	4
Innovative business ideas							
Conclusiveness	247	2.963	0.924	1	2.3	3.5	5
Realistic Implementation	247	2.991	0.937	1	2.3	3.7	5
Economic Potential	247	2.694	0.877	1	2	3.1	5

5. Results

5.1. Identification of latent topic structures in business ideas

Table 5.1 illustrates the result of the topic model. Topic 3 has the highest expected share in the overall text corpus of 20%. With roughly 8% share of the overall text corpus, Topic 1 is the least appearing topic. The most common words per topic and illustrative text excerpts[12] are the basis for the topic labels, which describe the main theme of the business ideas.

The result shows that the topics are meaningful and very distinct from each other, which speaks for the previous consideration for coherence and exclusivity. The topic “Development of technical devices and software” reflect the description of the creation of a prototype of devices or software solutions (e.g. “software”, “device”, “development”). “Consultancy for individuals” is thematically concentrated on the labour market and individual qualifications (e.g. “labour”, “university”, “possibilities”). “Customer orientation” focuses on customer needs and market niches in B2C markets (e.g. “product”, “customer”, “target group”). The topic “Social entrepreneurship” describes local projects with societal impact (e.g. “project”, “children”, “location”). “Process development” deals with supportive consultancy services for process development for firms (e.g. “development”, “demand”, “service”). “Business consultancy” aims at activities to provide firms with consultancy in various departments (e.g. “company”, “euro”, “market”). “Project orientation” are descriptions focussing on planning and execution of projects as a service not restricted to any domain (e.g. “implementation”, “creation”, “service”). The topic “Social media platforms” deals with the exchange of different types of media with new or existing social contacts (e.g. “app”, “platform”, “media”).

The topics can be clustered into three categories: First, products and key processes for revenue generation, which the idea builds upon, are represented. This is especially evident for the topics “Development of technical devices and software”, “Consultancy for individuals”, “Process development”, “Business consultancy” and “Social media platforms”. Second, the orientation type of a business describing the way the entrepreneurs approach the market, are covered with the topics “Customer orientation” and “Project orientation”. Business ideas with a higher share of the topic “Customer orientation” rely on knowledge about their customers to align the business with the demands to provide value. “Project orientation”, on the other hand, is informed by an iterative process of business development through the acquisition of projects. Third, the topics not only reflect the main focus of the business ideas, but also hint at the intention of the entrepreneurs. This is especially present in the case of the topic “Social entrepreneurship”, showing a focus on societal impact of the activities.

[12] The text excerpts cannot be shown here due to confidentiality of the data.

Table 5.1: Topics uncovered by structural topic modeling analysis based on all business idea submissions Source: Own calculation

Topic cluster	**Topic Label**	**Expected topic share**	**Most common 10 terms (English translation and *German word stems*)**
Products and key processes for revenue generation	Development of technical devices and software (Topic 3)	20.1%	Product, system, software, figure, device, market, simple, idea, existing, development *produkt, system, softwar, abbildung, geraet, markt, einfach, ide, besteht, entwicklung*
Products and key processes for revenue generation	Consultancy for individuals (Topic 2)	14.4%	Bremen, offer, students, work, city, bremish, university, possibilities, working, offered *bremen, angebot, studenten, arbeit, stadt, bremer, universitaet, moeglichkeiten, arbeiten, angeboten*
Orientation type of a business	Customer orientation (Topic 6)	13.9%	Product, customers, Germany, target group, design, customer, online, important, competition, offer *produkt, kunden, deutschland, zielgrupp, design, kund, onlin, wichtig, konkurrenz, angebot*
Intention	Social entrepreneurship (Topic 8)	12.2%	Idea, project, concept, children, find, interest, theme, contact, providing, location *ide, projekt, konzept, kinder, finden, interess, thema, kontakt, bietet, ort*
Products and key processes for revenue generation	Process development (Topic 7)	11.9%	Development, demand, data, network,service, goal, see, information, concept, additionally *entwicklung, bedarf, daten, netzwerk, servic, ziel, sieh, informationen, konzept, zusaetzlich*
Products and key processes for revenue generation	Business consultancy (Topic 4)	10.6%	company, customers, euro, collaboration, service, GmbH, market, Germany, market, costs *unternehmen, kunden, euro, mitarbeit, dienstleistung, gmbh, markt, deutschland, market, kosten*
Orientation type of a business	Project orientation (Topic 5)	8.9%	Field, implementation, planning, let, creation, context, offer, service, market, innovation *bereich, umsetzung, geplant, lassen, erstellung, rahmen, anbiet, dienstleistungen, markt, innov*
Products and key processes for revenue generation	Social media platforms (Topic 1)	7.9%	App, user, platform, internet, online, website, media, market, information, facebook *app, nutzer, plattform, internet, onlin, websit, media, market, informationen, facebook*

5.2. Identification of differences between innovative and non-innovative business ideas in topic proportions

In order to identify peculiarities of innovative business ideas in the topic structure, the documents of business ideas are divided in two groups. One group containing innovative business ideas, and the other group consisting out of non-innovative business ideas. By comparing the expected shares of the former mentioned topics in these two groups, it is possible to identify the key topics present in innovative business ideas. Figure 5.1 illustrates the distribution of topic shares in the two groups of business ideas.

The expected shares differ significantly in five of eight topics. Thus, the main finding of this analysis is that the proportion of topics in business ideas has the power to reveal the degree of innovativeness and, thus, is a building block for the understanding of innovative business ideas. Based on this result, the first proposition can be confirmed.

Two topics are significantly more associated with innovative business ideas compared to non-innovative business ideas. These topics are “Development of technical devices and software” and “Process development”. They correspond to the generation of new or the improvement of technologies and processes, taking a proactive approach and the exploration of new ways of doing something.

Three topics are significantly more prominent in business ideas rated as non-innovative compared to shares in innovative business ideas. The first topic is “Customer orientation”. This topic reflects a demand driven approach aiming at incremental satisfaction of customer needs rather than producing novelties. The second topic is “Social entrepreneurship”. This confirms the traditional assignment of “Social entrepreneurship” to a non-profit and non-innovative sector (Wilson & Post, 2013), since this topic is concerned with problem solving rather than developing innovations. The last topic is “Consultancy for individuals”. These topics are similar in the sense that they focus on the identification and adaptation to demands, societal issues, and changing labour market conditions. They represent more a reactive than a proactive approach to the business context, which explains their link to non-innovative ideas.

For the topics “Social media platforms”, “Business consultancy”, and “Project orientation” there are no significantly different expected shares between the two groups. Since these topics are more general, it is likely that these complement the former topics in diverse ways.

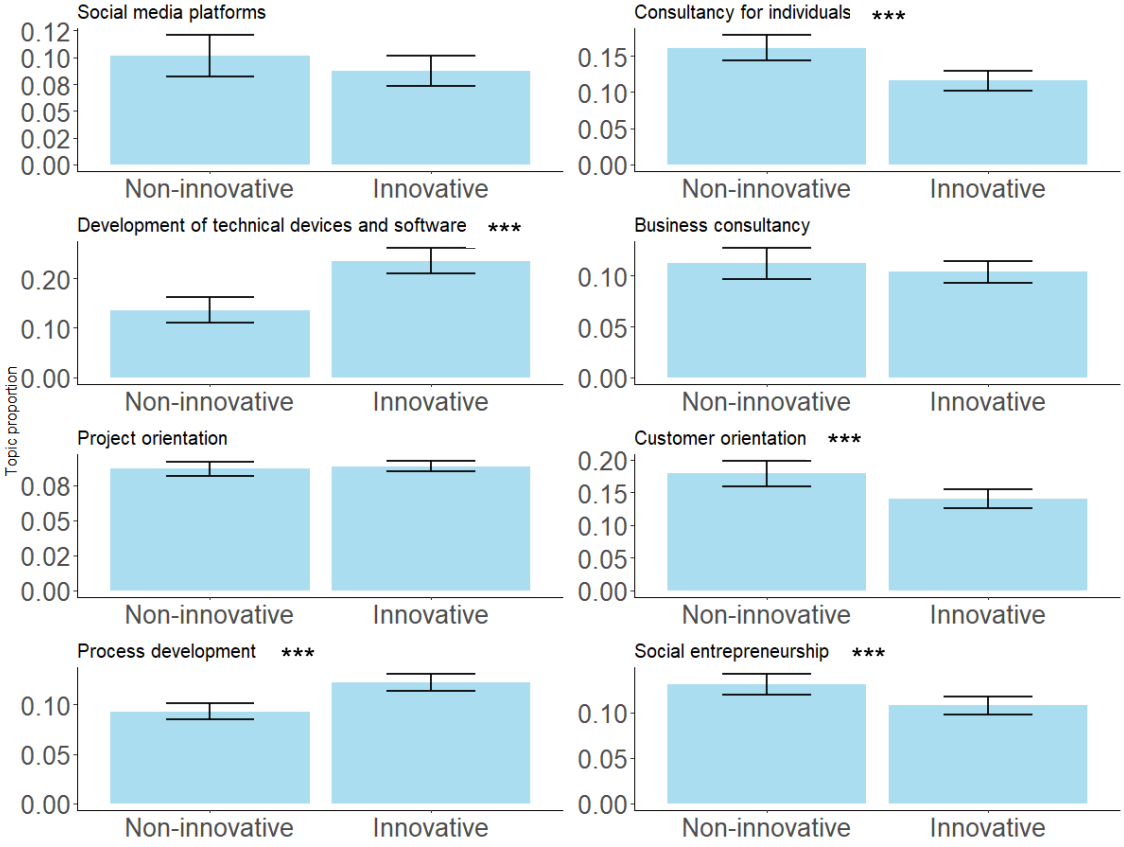


Figure 5.1: Comparison of topic proportions in innovative and non-innovative business ideas
*Note: The error bars corresponds to a confidence interval of 95%; *** corresponds to a significant difference with a p-value lower than 0.01*
Source: Own calculation

5.3. Analysis of topic interdependencies of innovative and non-innovative business ideas

To properly discriminate innovative and non-innovative business ideas, not only the proportion per topic is relevant, but also how the topics are interrelated to each other. For this purpose, the correlation structures of topics are examined for innovative and non-innovative business ideas as illustrated in Figure 5.2.

A main result from this analysis is that topics correlate with each other and the co-occurrence structures of topics differ between innovative and non-innovative business ideas. Thus, the second proposition is confirmed.

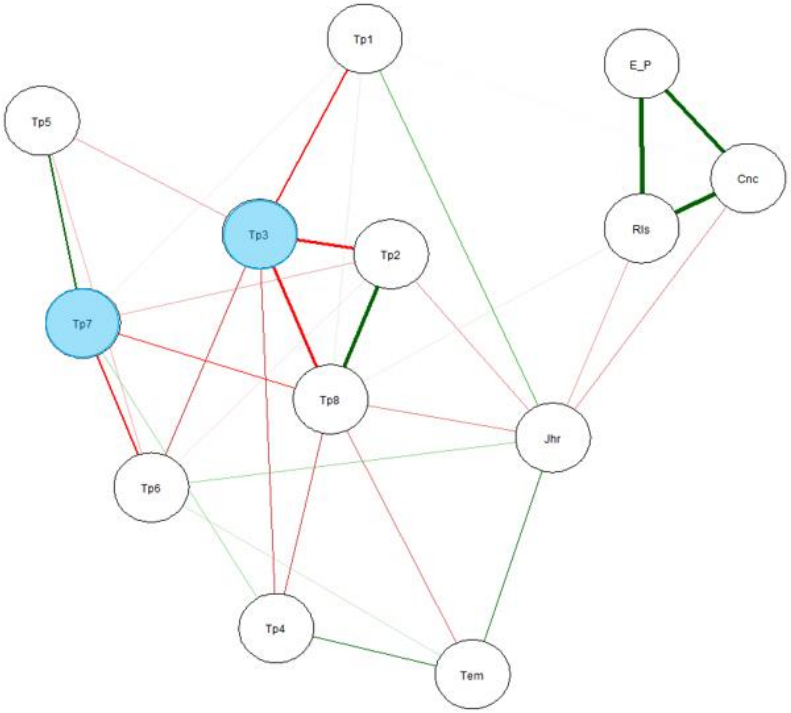
The relations between the topics in the case of innovative business ideas reveal that “Development of technical devices and software” (Tp3) serves as an indicator for innovative ideas when this topic is central to the idea and not combined with other topics. The second topic associated with innovative business ideas “Process development” (Tp7) can be used as an indicator for innovativeness when combined with “Business consultancy” (Tp4) and “Project orientation” (Tp5). The degree of innovativeness is likely to be diminished by combining “Process development” (Tp7) with “Customer orientation” (Tp6), “Social entrepreneurship” (Tp8), “Social media platforms” (Tp1), and “Consultancy for individuals” (Tp2).

Business ideas focussing on the topics “Consultancy for individuals” (Tp2) and “Social entrepreneurship” (Tp8) are likely to be non-innovative. Moreover, ideas solely emphasising “Customer orientation” (Tp6) can be assumed to be non-innovative.

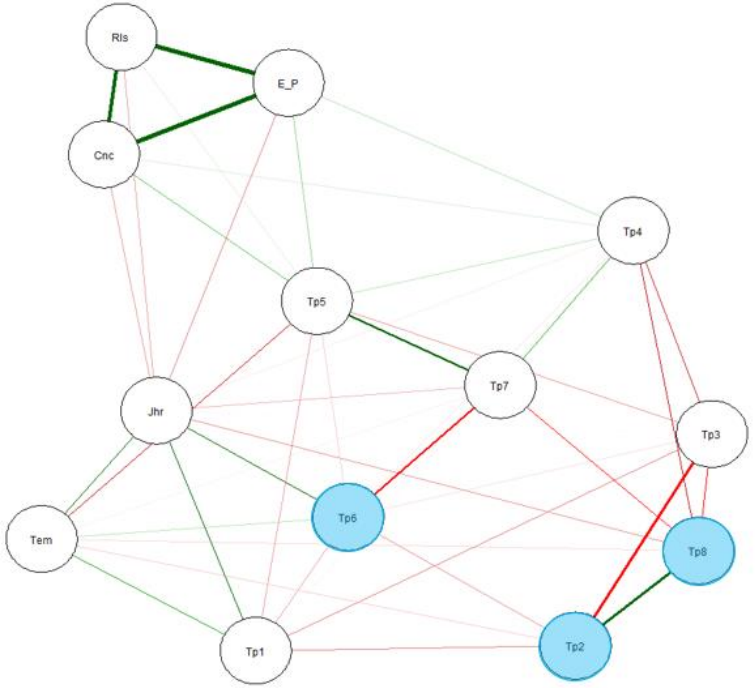
The meta-variables “Economic Potential” (E_P), “Realistic Implementation” (RIs) and “Conclusiveness” (Cnc), in particular, show more links between themselves and the topics in the case of non-innovative business ideas.[13] These findings could be explained by the difficulty to evaluate innovative business ideas precisely due to their liability of newness. Also other context factors of the business ideas, namely team size and year of submission, show an impact on the topical structure.

[13] This finding indicates that the expert jury of the business idea competition attests non-innovative business ideas dealing with “Business consultancy” (Tp4) higher “Economic Potential” (E_P) and “Conclusiveness” (Cnc). Moreover, the jury evaluated non-innovative business ideas featuring a “Project orientation” (Tp5) more positively in terms of economic prospects (E_P), conclusiveness (Cnc) and realistic implementation (RIs).

Innovative business ideas



Non-innovative business ideas



Legend of topics

(Tp1) Social media platforms
(Tp2) Consultancy for individuals
(Tp3) Development of technical devices and software
(Tp4) Business consultancy
(Tp5) Project orientation
(Tp6) Customer orientation
(Tp7) Process development
(Tp8) Social entrepreneurship

Figure 5.2: Correlation networks of innovative and non-innovative business ideas
Note: The colour of the links correspond to negative (red) and positive (green) correlations. The transparency of the edges reflect the Pearson correlation coefficients. For correlation coefficients please see Tables 9.3 and 9.4 in the appendix. The blue nodes in the graphic represent the topics identified to be specific for innovative and non-innovative business ideas.
Source: Own calculation

5.4. Analysis of incorporation probabilities for topic compositions of innovative and non-innovative business ideas

Classification trees are applied to examine the interplay of topic structures of business ideas and their impact on incorporation probability. The first performed split corresponds to the most important variable to explain the incorporation probability. The value depicted behind the split variable is equal to the threshold dividing the business ideas into the right and the left branch of the tree which makes the most precise prediction. In case the condition is fulfilled, the reader follows the left branch of the splitting criteria. If the condition is not fulfilled, the reader follows the right branch. The terminal nodes of the classification trees show the prominent prediction ("No Foundation" or "Foundation") for the path leading there. Beneath this result the ratio of observations fulfilling either "No Foundation" or "Foundation" as outcome is displayed for this terminal node. Both decision trees show splits even if the prediction in both resulting terminal nodes is "No Foundation". This is because the algorithm performs a split if this improves the purity of the terminal node.

Two main findings can be drawn from the analysis. First, the topics are the main drivers of the classification trees and not the meta-data as team size, year, economic potential, or realistic implementation.[14] Additionally to the topics, only industry classifications and the conclusiveness of the business idea are of importance. Thus, it can be confirmed that the topic structure of business ideas contains worthy information about the incorporation probability, and therefore, the third proposition is valid. Second, the splitting criteria cover topics concerning the base product or process of the business idea as well as the business orientation and entrepreneurial intention. The data suggests that the evaluation of business ideas and their prospect of incorporation requires a detailed investigation of these aspects of the business idea. This finding confirms the fourth proposition.

Additionally, in both trees the prediction is "No Foundation" in most cases. This finding is expected because the majority of business ideas do not get incorporated.[15] Moreover, the classification tree for non-innovative business ideas follows a simpler structure than the tree for innovative ones. This could stem from a lower sample size for

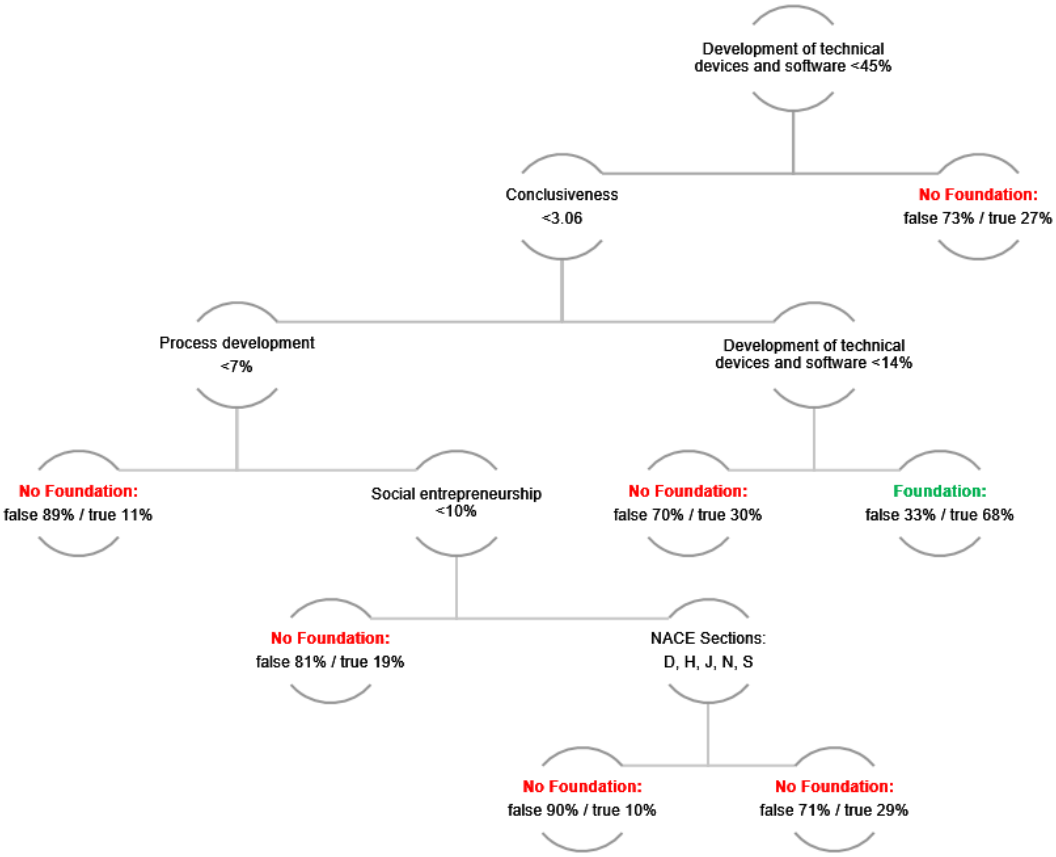
[14] Using a random forest based upon decision trees is performed as a robustness check, but the result still holds true. For this purpose all business ideas were applied in one analysis to ensure that the effect is not driven by the degree of innovativeness or by the sample with a jury rating. The variable importance, as the main result of the random forest analysis, is illustrated in Figure 9.3 in the appendix.

[15] Misclassification rate is not reported in the graphs as this measure is not insightful. This is because the majority of business ideas are not incorporated. Thus, it is expected that most predictions of the decision trees are "No Foundation" and misclassifies the majority of incorporated business ideas.

the former classification tree and the condition that each leave must at least contain 30 business ideas.

5.4.1. Innovative business ideas

Innovative business ideas



Sample size	247

Figure 5.3: Results of the classification tree analysis for innovative business ideas
Note: sample size corresponds to number of submitted documents
Source: Own calculation

Figure 5.3 illustrates the classification tree for innovative business ideas. The most important splitting criteria for innovative business ideas is “Development of technical devices and software” (Topic 3). In case a business idea´s description with a higher share than 45% assigned to this topic, this business idea is not likely to be incorporated. Not even a third of business ideas showing this topic structure is successful. However, when a high score in “Conclusiveness” of the business idea is given, a share of “Development of technical devices and software” (Topic 3) between 14% and 45% leads to the highest share of incorporations (68% incorporations). Hence, potential entrepreneurs aiming at the development of technical solutions should be very explicit about their plan to develop a technical device or software. However, a too narrow

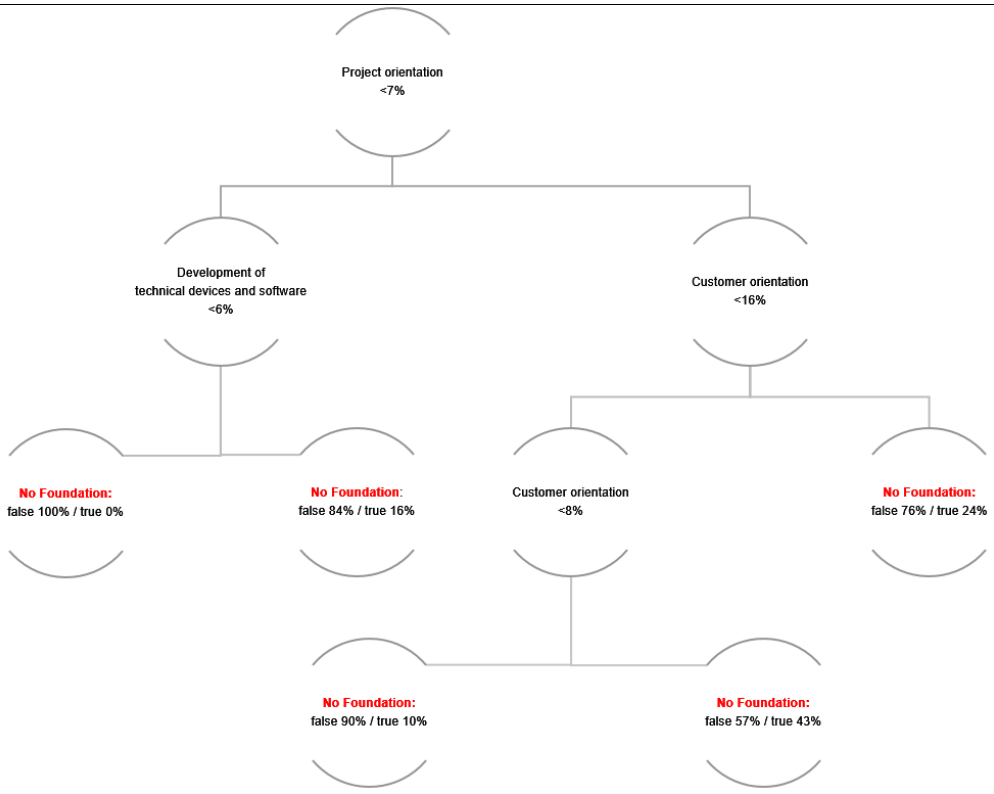
focus on this also hampers incorporation probability. Therefore, it seems advisable to not only focus on the development of the technical solution, but also keep other factors, like e.g. market orientation, in mind.

A different path visible in this classification tree is when technical solutions show a lower share than 45% of the business idea and are less conclusive. In this case the involvement of “Process development” and “Social entrepreneurship” helps to avoid high rates of incorporation failures. Nevertheless, when the business idea aims at industries in agriculture, forestry and fishing (A), transportation and storage (H), information and communication (J), administrative and support service activities (N) or other service activities (S) the incorporation probability is limited to 10%.

While the first path focusses solely on the development of technical devices and software, the second path shows that innovative business ideas profit from combination of various topics.

5.4.2. Non-innovative business ideas

Non-innovative business ideas



Sample size	182

Figure 5.4: Results of the classification tree analysis for non-innovative business ideas
Note: sample size corresponds to number of submitted documents
Source: Own calculation

Figure 5.4 reflects the patterns within the non-innovative business ideas. “Project orientation” (Topic 5) divides the business ideas into two paths. Business ideas with a share of “Project orientation” lower than 7% overall show lower chances of incorporation compared to those focussing on “Project Orientation”. Non-project oriented, and non-innovative, ideas should avoid the “Development of technical devices and software” (Topic 3) because when taken into consideration all business ideas were not incorporated. Following the track of higher shares in “Project orientation” (Topic 5), either a higher share than 16% in “Customer orientation” (Topic 6) or a lower share than 8% yields higher incorporation chances. This pattern indicates that non-innovative business ideas require a clear focus on one market - either B2C (as reflected in Topic 6) or other clients.

6. Discussion & Implications

This study analyses the topics appearing in innovative and non-innovative business ideas. The topics identified in this study, reflect three aspects of entrepreneurial activity – entrepreneurial intentions, business orientation and main products or processes. Thus, the application of topic modeling as a technique to analyse content of texts, contribute to the existing literature by assessing not only objective goals but also entrepreneurial motivations and operation modes. This provides so far unused, information to differentiate entrepreneurial activities (Hurst & Pugsley, 2011; Shane, 2009).

The first research question of this study, if topic structures differ between innovative and non-innovative business ideas, can be answered with a yes. The content of innovative and non-innovative business ideas differ in their topic proportions. In addition, the correlations between the topics and meta-data provide a differentiated perspective on the content of business ideas. The topic proportions already give a hint at how to identify innovative business ideas, which is further enhanced by taking the correlations between the topics into account. This enables a more distinct promotion of innovative entrepreneurship activities and, thus, answers the call of former researchers (Hurst & Pugsley, 2011; Shane, 2009). Summarising the empirical evidence, it cannot only be confirmed that the basis of the business idea, the product or process, is of critical relevance for determining the degree of innovativeness (McAdam & Marlow, 2011), but also that entrepreneurial intentions and business orientation are important differences between innovative and non-innovative business ideas. Moreover, not only topic proportions differ in these two groups of business ideas, but also their correlation with external factors. Non-innovative business ideas show higher correlations between the topic structure and jury ratings in economic potential, conclusiveness, and degree of realistic implementation. In contradiction to the findings of Pare et al. (2011), the team size does not have a relation to the meta-variables “Economic Potential” (E_P), “Realistic Implementation” (RIs), and “Conclusiveness” (Cnc), i.e. greater team sizes do not

improve the overall quality of the business idea. However, the result indicates a positive correlation between team size and “Customer orientation” and a negative relation to “Social entrepreneurship”. The latter finding corresponds to notions that social entrepreneurship stems from highly individual experiences that may cancel each other out in a bigger team (Asarkaya & Keles Taysir, 2019; Mair & Noboa, 2003). The data suggests that team size influences ideas and perspectives embedded in the topic structure (Munoz-Bullon et al., 2015; West, 2007).

The second research question of this study, i.e. if the topic composition of a business idea relates to its incorporation probability, can be also answered with yes. The incorporation probability of business ideas is related to their particular topic structure. This finding not only confirms former observations from entrepreneurial promotion programs that the base product or process is a main determinant (Klofsten, 2005; McAdam & Marlow, 2011), but goes beyond that. The results of this study show that not only base products and processes are of key importance, but also entrepreneurial intention as well as business orientation. This holds true for innovative and non-innovative business ideas. However, the importance of each specific topic differs depending on the degree of innovativeness of each idea.

Innovative business ideas show two distinct paths in the analysis. The first string of successful innovative business ideas focusses on technology and software development. The second path is characterised by a broader approach taking social aspects and the improvement of processes into account. These insights suggest that there are two types of innovative business ideas – specialised on a technical solution or a broader approach for process improvement. In the case of innovative entrepreneurship, the results underline former research, that focussing on technical solutions and neglecting the marketing side and the balance between supply and demand leads to lower incorporation probabilities (Friar & Meyer, 2003; Teece, 2010; Youtie et al., 2012). This finding goes in line with the requirement of innovative entrepreneurs to communicate their idea in a concise manner to be able to access necessary external, initial funding (Rasmussen & Sorheim, 2012). In comparison to technical solutions, the development of processes seems to require a broader perspective involving as well social and environmental (e.g., industry) aspects. This type of entrepreneurial activities could be interpreted as hybrids joining social purposes with economic rationality (Wilson & Post, 2013). Following the hybrid entrepreneurship path, business ideas should avoid aiming at service driven industries. These industries may be easier to enter but also have higher failure rates due to more competition (Johnson, 2004; Shane, 2009). This study confirmed this to be even true in the pre-foundational stage of the entrepreneurial process.

In case of non-innovative business ideas, project-oriented services with a clear focus on a customer segment lead to the highest incorporation rate. This result points to the relevance of market- and customer orientation, also highlighted by former research

(Meyskens & Carsrud, 2013). Entrepreneurial endeavours with a low degree of innovativeness serve a purpose through the transfer of knowledge and skills and use opportunities for arbitrage.

To sum up the empirical findings, business ideas are a multifaceted phenomenon and a distinction in only innovative or non-innovative may not be enough to design valuable policies. Therefore, the results suggest that the topic structure of business ideas is a valuable source to target promotion measures of entrepreneurial activities (Hurst & Pugsley, 2011; Shane, 2009). Based upon the results of this study, a more differentiated promotion of innovative business ideas becomes possible. First, topic proportions in business idea descriptions can hint at the degree of innovativeness. This may be used in consultancy of potential entrepreneurs and the allocation of resources in programs aiming at the promotion of innovative entrepreneurship. Second, to attract innovative entrepreneurial endeavours to apply for funding programs, the call for applications may use words appearing often in the desired topics.

Finally, there are some limitations of this study giving way to further research. First, incorporation probability is the output variable the study opts at. Reasons why an idea is not incorporated are manifold and cannot be solely explained by variables included in the analysis. The reasons why business ideas fail to be incorporated may be subject of future research. This also holds true for additional control variables not available in the used data set. Second, this study differentiates between innovative and non-innovative business ideas. However, innovative entrepreneurship might not be the only desirable outcome, but also entrepreneurial endeavours transferring knowledge and skills or tackling a societal challenge. Third, the definition of innovativeness may be subject of change in the time period 2005 until 2019 due to technological progress. Incorporation of more indicators of innovativeness other than expert opinions may be beneficial, especially because the expert committees in the investigated business idea competition may not be representative for general expert assessments. Fourth, the incorporation of a business idea is primarily the decision by the entrepreneurs and does not represent a purely external assessment of the business idea. Thus, future research could be done by collecting more information on market success to avoid bidirectional causality. However, since it would restrict the sample in a severe manner and an incorporation without any external approval is unlikely, this study concentrates on this stage of the entrepreneurial process. Fifth, this study builds upon a sample that represents only business ideas in Bremen and Bremerhaven, Germany. Although Bremen is a federal state which is known to represent Germany well and is often used as a test market for products, future research could extend the data base. By sampling more data from other business idea competitions, more sophisticated methods to build classification trees with training and test data would become possible (e.g. pruning and cross validation techniques). Such a procedure would allow a more accurate determination of the model accuracy, without the need to restrict the decision tree growing process to a minimum sample size of 30 at the terminal nodes, as done in this

study. Sixth, the advancement of stemming algorithms in the German language seems to be valuable for future research. So far, not every word is correctly stemmed in the German language. However, stemming is necessary because it allows for the unification of most of the terms (Birkholz et al., 2021).

7. Conclusion

To foster innovation and wealth generation, the stimulation of entrepreneurial activities is of great importance for policy makers. However, the policies often follow a non-selective funding practice and, thus, distribute resources over a broad range of potential entrepreneurs and diverse business ideas. The problem with this approach is that only a small fraction of new ventures generate innovations and wealth. Thus, this non-selective funding practice is criticised and researchers call for targeted entrepreneurship policies. Traditional indicators (e.g., characteristics of the entrepreneurs, resources available, and external investments) do not provide a reliable source of information to properly assess the incorporation potential of a business idea. To close this gap, the purpose of the present paper is to examine the topic structures of innovative and non-innovative business ideas and their effect on the incorporation probability. Texts describing business ideas submitted to a business idea competition are the basis for the topic modeling and classification tree analyses conducted in this study. In summary, the results show that the topic structure of business ideas provide a source of information to discriminate innovative from non-innovative business ideas and can be applied to predict the incorporation probability of ideas. These findings help to target policies to enable innovation and wealth generation by new ventures.

Acknowledgements
The author wants to thank the transfer office and BRIDGE at the University of Bremen for the data provision and support.

References

Asarkaya, C., & Keles Taysir, N. (2019). Founder's Background as a Catalyst for Social Entrepreneurship. Nonprofit Management and Leadership, 30(1), 155–166. https://doi.org/10.1002/nml.21353

Audretsch, D. B., & Keilbach, M. (2004). Entrepreneurship Capital and Economic Performance. Regional Studies, 38(8), 949–959. https://doi.org/10.1080/0034340042000280956

Autio, E., & Yli-Renko, H. (1998). New, technology–based firms as agents of technological rejuvenation. Entrepreneurship & Regional Development, 10(1), 71–92. https://doi.org/10.1080/08985629800000005

Birkholz, J., Günther, J., & Shkolnykova, M. (2021). #2101 Bremen Papers on Economics & Innovation Using Topic Modeling in Innovation Studies: The Case of a Small Innovation System under Conditions of Pandemic Related Change. Bremen Papers on Economics & Innovation, 2101, 31. https://doi.org/10.26092/elib/451

BMWi (2018). Förderlotse. Wegweiser durch die Programme zur Gründungsförderung. München.

Braga, I. A., Monard, M.-C., & Matsubara, E. T. (2009). Combining Unigrams and Bigrams in Semi-Supervised Text Classification. 12.

Capelleras, J.-L., & Greene, F. J. (2008). The Determinants and Growth Implications of Venture Creation Speed. Entrepreneurship and Regional Development, 20(4), 317–343. https://doi.org/10.1080/08985620701855683

Carter, N. M., Gartner, W. B., & Reynolds, P. D. (1996). Exploring Start-up Event Sequences. Journal of Business Venturing, 11(3), 151–166. https://doi.org/10.1016/0883-9026(95)00129-8

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. 10.

Chattopadhyay, R., & Ghosh, A. K. (2008). Entrepreneurial Intention Model-Based Quantitative Approach to Estimate Entrepreneurial Success. Journal of Small Business and Entrepreneurship, 21(1), 1–21. https://doi.org/10.1080/08276331.2008.10593410

Chen, X.-P., Yao, X., & Kotha, S. (2009). Entrepreneur Passion And Preparedness In Business Plan Presentations: A Persuasion Analysis Of Venture Capitalists' Funding Decisions. Academy of Management Journal, 52(1), 199–214. https://doi.org/10.5465/amj.2009.36462018

Davidsson, P., & Gordon, S. R. (2012). Panel studies of new venture creation: A methods-focused review and suggestions for future research. Small Business Economics, 39(4), 853–876. https://doi.org/10.1007/s11187-011-9325-8

Drnovsek, M., Franic, S., & Slavec, A. (2018). Exploring Antecedents of Business Angels' Intention to Invest. Zbornik Radova Ekonomskog Fakulteta u Rijeci: Casopis Za Ekonomsku Teoriju i Praksu/Proceedings of Rijeka School of Economics: Journal of Economics and Business, 36(2), 701–734. https://doi.org/10.18045/zbefri.2018.2.701

Fox, R. (2006). Using Corpus Linguistics to Describe Corporations' Ideologies. Tourism and Hospitality Management, 12(2), 15–24. https://doi.org/10.20867/thm.12.2.2

Friar, J. H., & Meyer, M. H. (2003). Entrepreneurship and Start-Ups in the Boston Region: Factors differentiating High-Growth Ventures from Micro-ventures. Small Business Economics, 21(2), 145–152.

Gundry, L. K., Ofstein, L. F., & Monllor, J. (2016). Entrepreneurial Team Creativity: Driving Innovation from Ideation to Implementation. Journal of Enterprising Culture, 24(01), 55–77. https://doi.org/10.1142/S0218495816500035

Hurst, E., & Pugsley, B. W. (2011). What Do Small Businesses Do? Brookings Papers on Economic Activity, 73–118.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Johnson, P. S. (2004). Differences in Regional Firm Formation Rates: A Decomposition Analysis. Entrepreneurship Theory and Practice, 28(5), 431–446. https://doi.org/10.1111/j.1540-6520.2004.00056.x

Kirk, G. R., & Nolan, S. B. (2010). Nonprofit mission statement focus and financial performance. Nonprofit Management and Leadership, 20(4), 473–490. https://doi.org/10.1002/nml.20006

Klofsten, M. (2005). New venture ideas: An analysis of their origin and early development. Technology Analysis & Strategic Management, 17(1), 105–119. https://doi.org/10.1080/09537320500044776

Kollmann, T., Jung, P. B., Kleine-Stegemann, L., Ataee, J., & de Cruppe, K. (2020). Deutscher Startup Monitor. Inovation statt Krise.

Mair, J., & Noboa, E. (2003). Social Entrepreneurship: How Intentions to Create a Social Enterprise Get Formed. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.462283

Markman, G. D., Balkin, D. B., & Baron, R. A. (2002). Inventors and New Venture Formation: The Effects of General Self–Efficacy and Regretful Thinking. Entrepreneurship Theory and Practice, 27(2), 149–165. https://doi.org/10.1111/1540-8520.00004

McAdam, M., & Marlow, S. (2011). Sense and Sensibility: The Role of Business Incubator Client Advisors in Assisting High-Technology Entrepreneurs to Make Sense of Investment Readiness Status. Entrepreneurship and Regional Development, 23(7–8), 449–468. https://doi.org/10.1080/08985620903406749

Meyskens, M., & Carsrud, A. L. (2013). Nascent Green-Technology Ventures: A Study Assessing the Role of Partnership Diversity in Firm Success. Small Business Economics, 40(3), 739–759. https://doi.org/10.1007/s11187-011-9400-1

Milanovic Glavan, L., Bosilj Vuksic, V., & Vlahovic, N. (2015). Decision Tree Learning for Detecting Turning Points in Business Process Orientation: A Case of Croatian Companies. Croatian Operational Research Review, 6(1), 207–224.

Morris, R. (2011). 2011 High-Impact Entrepreneurship Global Report.

Munoz-Bullon, F., Sanchez-Bueno, M. J., & Vos-Saz, A. (2015). Startup Team Contributions and New Firm Creation: The Role of Founding Team Experience. Entrepreneurship and Regional Development, 27(1–2), 80–105. https://doi.org/10.1080/08985626.2014.999719

Narver, J. C., Slater, S. F., & Tietje, B. (1998). Creating a Market Orientation. Journal of Market Focused Management, 2, 241–255.

Pandey, S., Kim, M., & Pandey, S. K. (2017). Do Mission Statements Matter for Nonprofit Performance? Insights from a Study of US Performing Arts Organizations. Nonprofit Management and Leadership, 27(3), 389–410. https://doi.org/10.1002/nml.21257

Pare, J.-L., Rédis, J., & Hikkerova, L. (2011). The Influence of Organizational Capital on the Conception of the Enterprise Project. International Journal of Business, 16(4), 315–339.

Plomaritou, E. I., Plomaritou, V., & Giziakis, K. (2011). Shipping Marketing and Customer Orientation: The Psychology and Buying Behavior of Charterer and Shipper in the Tramp and Liner Market. Management: Journal of Contemporary Management Issues, 16(1), 57–89.

Proksch, D., Stranz, W., Röhr, N., Ernst, C., Pinkwart, A., & Schefczyk, M. (2017). Value-Adding Activities of Venture Capital Companies: A Content Analysis of Investor's Original Documents in Germany. Venture Capital, 19(3), 129–146. https://doi.org/10.1080/13691066.2016.1242573

Rasmussen, E., & Sorheim, R. (2012). Obtaining Early-Stage Financing for Technology Entrepreneurship: Reassessing the Demand-Side Perspective. Venture Capital, 14(2–3), 77–89. https://doi.org/10.1080/13691066.2012.667908

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses: STRUCTURAL TOPIC MODELS FOR SURVEY RESPONSES. American Journal of Political Science, 58(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Sarasvathy, S. D. (2003). Entrepreneurship as a science of the artificial. Journal of Economic Psychology, 24(2), 203–220. https://doi.org/10.1016/S0167-4870(02)00203-9

Shalley, C. E., Gilson, L. L., & Blum, T. C. (2009). Interactive Effects of Growth Need Strength, Work Context, and Job Complexity On Self-Reported Creative Performance. Academy of Management Journal, 52(3), 489–505. https://doi.org/10.5465/amj.2009.41330806

Shane, S. (2009). Why Encouraging More People to Become Entrepreneurs Is Bad Public Policy. Small Business Economics, 33(2), 141–149.

Silge, J. (2018): Training, evaluating, and interpreting topic models https://juliasilge.com/blog/evaluating-stm/ [accessed on 13.09.2020].

Song, M., Podoynitsyna, K., van der Bij, H., & Halman, J. I. M. (2008). Success Factors in New Ventures: A Meta-analysis. The Journal of Product Innovation Management, 25(1), 7–27. https://doi.org/10.1111/j.1540-5885.2007.00280.x

Sternberg, R., Gorynia-Pfeffer, N., Wallisch, M., Baharian, A., Stolz, L., & von Bloh, J. (2020). Global Entrepreneurship Monitor—Unternehmensgründungen im weltweiten Vergleich. Länderbericht Deutschland 2019/20. RKW Kompetenzzentrum.

Storopoli, J. E. (2019). Topic Modeling: How and Why to Use in Management Research. Iberoamerican Journal of Strategic Management, 18(3), 316–338.

Stropel, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods, 14(4), 323-348. https://doi.org/10.1037/a0016973

Stuetzer, M., Obschonka, M., & Schmitt-Rodermund, E. (2013). Balanced Skills among Nascent Entrepreneurs. Small Business Economics, 41(1), 93–114. https://doi.org/10.1007/s11187-012-9423-2

Teece, D. J. (2010). Business Models, Business Strategy and Innovation. Long Range Planning, 43(2–3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003

Wagner, J. (1997). Firm Size and Job Quality: A Survey of the Evidence from Germany? Small Business Economics, 9(5), 411–425.

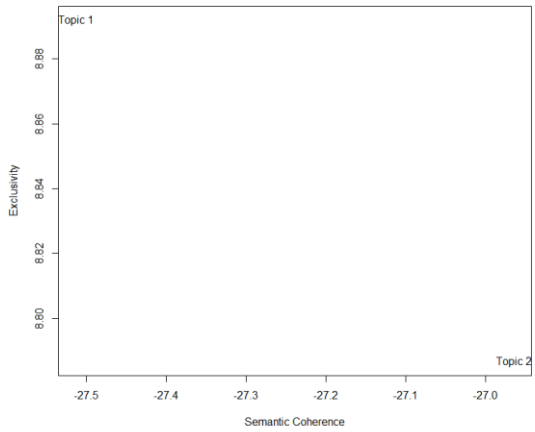
Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika, 34(1/2), 28-35. https://doi.org/10.2307/2332510

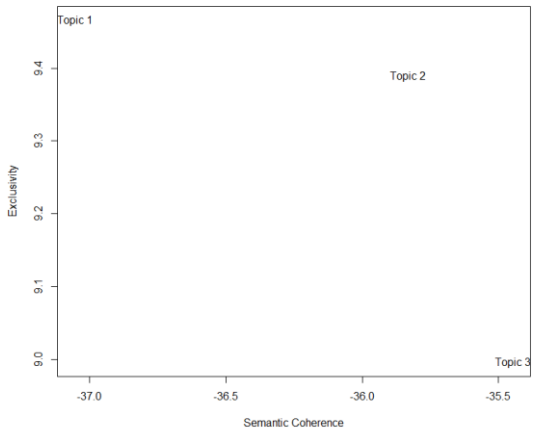
West, G. P. (2007). Collective Cognition: When Entrepreneurial Teams, Not Individuals, Make Decisions. Entrepreneurship Theory and Practice, 31(1), 77–102. https://doi.org/10.1111/j.1540-6520.2007.00164.x

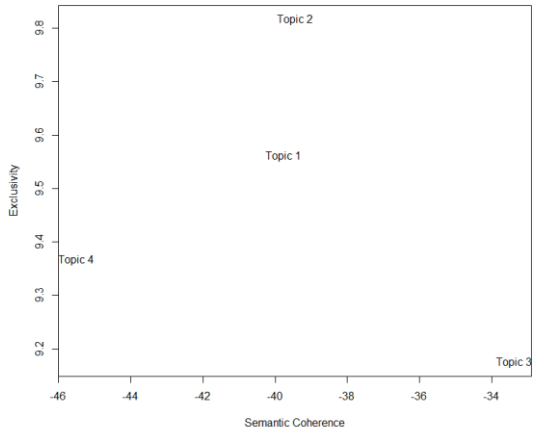
Wilson, F., & Post, J. E. (2013). Business Models for People, Planet (and Profits): Exploring the Phenomena of Social Business, a Market-Based Approach to Social Value Creation. Small Business Economics, 40(3), 715–737. https://doi.org/10.1007/s11187-011-9401-0

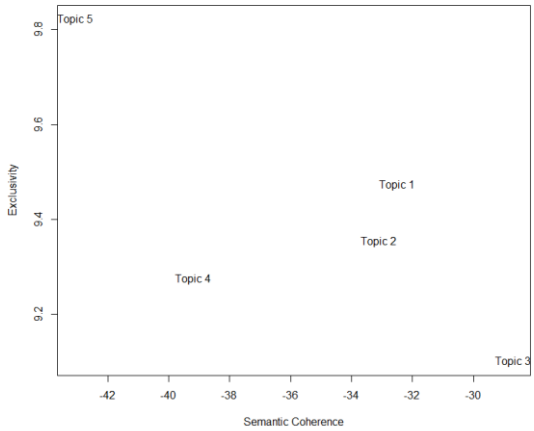
Youtie, J., Hicks, D., Shapira, P., & Horsley, T. (2012). Pathways from Discovery to Commercialisation: Using Web Sources to Track Small and Medium-Sized Enterprise Strategies in Emerging Nanotechnologies. Technology Analysis and Strategic Management, 24(10), 981–995. https://doi.org/10.1080/09537325.2012.724163

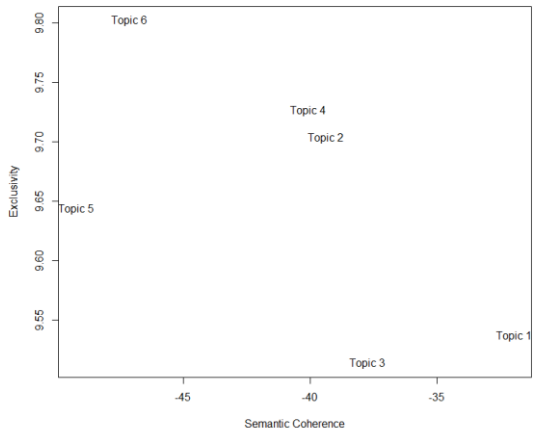
Appendix

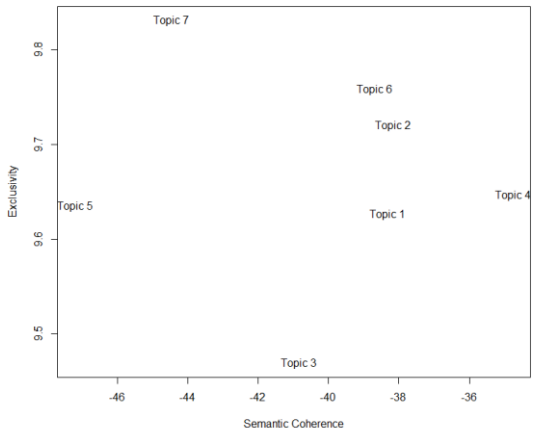

Topic 1
Topic 2
Exclusivity
Semantic Coherence

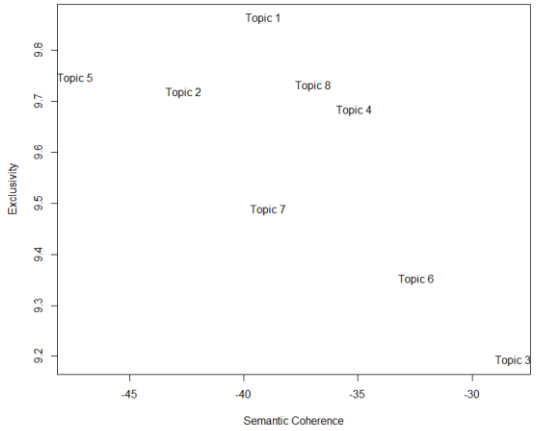

Topic 1
Topic 2
Topic 3
Exclusivity
Semantic Coherence

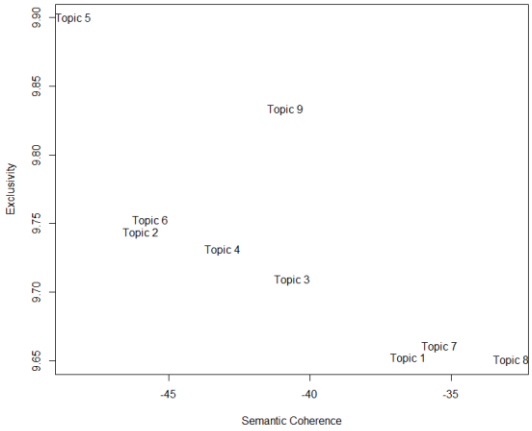

Topic 2
Topic 1
Topic 4
Topic 3
Exclusivity
Semantic Coherence

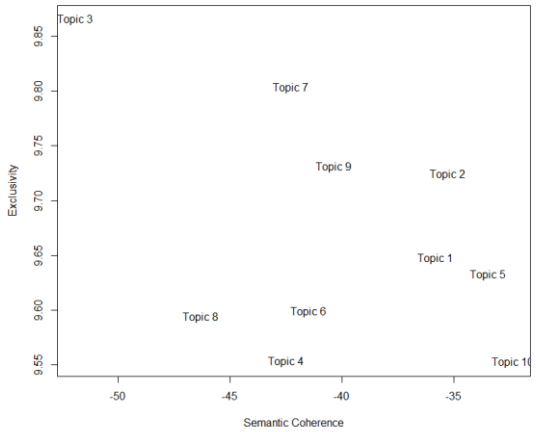

Topic 5
Topic 1
Topic 2
Topic 4
Topic 3
Exclusivity
Semantic Coherence

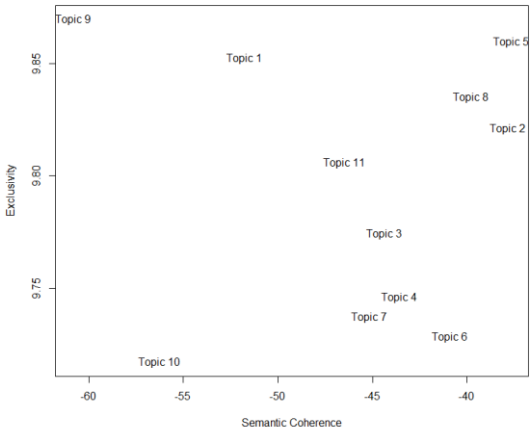

Topic 6
Topic 4
Topic 2
Topic 5
Topic 1
Topic 3
Exclusivity
Semantic Coherence

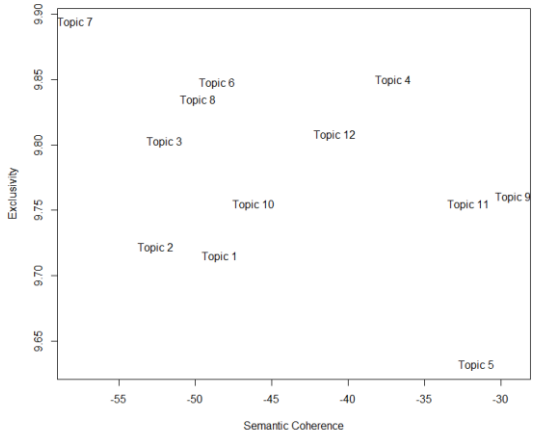

Topic 7
Topic 6
Topic 2
Topic 4
Topic 5
Topic 1
Topic 3
Exclusivity
Semantic Coherence

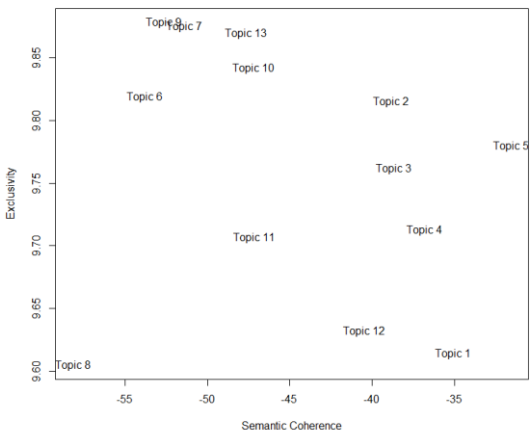

Topic 1
Topic 5
Topic 2
Topic 8
Topic 4
Topic 7
Topic 6
Topic 3
Exclusivity
Semantic Coherence

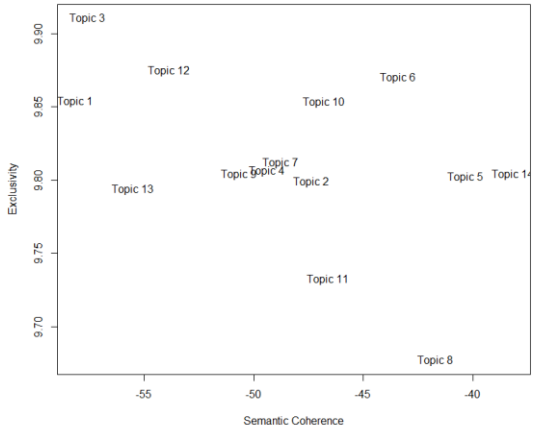

Topic 5
Topic 9
Topic 6
Topic 2
Topic 4
Topic 3
Topic 7
Topic 1
Topic 8
Exclusivity
Semantic Coherence

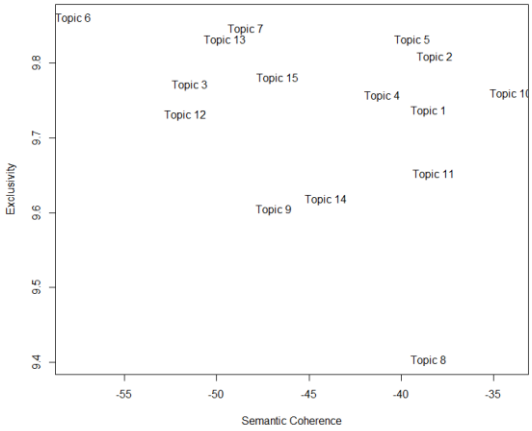

Topic 3
Topic 7
Topic 9
Topic 2
Topic 1
Topic 5
Topic 8
Topic 6
Topic 4
Topic 10
Exclusivity
Semantic Coherence

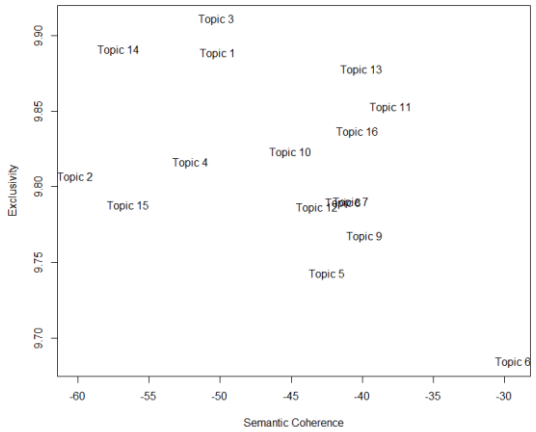

Topic 9
Topic 5
Topic 1
Topic 8
Topic 2
Topic 11
Topic 3
Topic 4
Topic 7
Topic 6
Topic 10
Exclusivity
Semantic Coherence

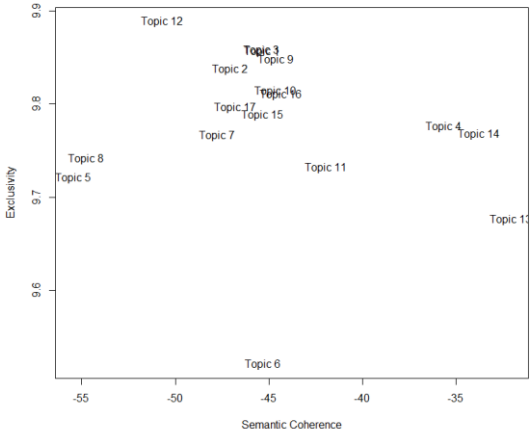

Topic 7
Topic 6
Topic 8
Topic 4
Topic 3
Topic 12
Topic 10
Topic 11
Topic 9
Topic 2
Topic 1
Topic 5
Exclusivity
Semantic Coherence

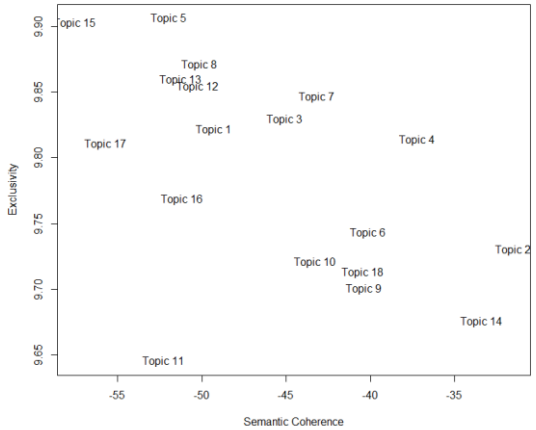

Topic 9
Topic 7
Topic 13
Topic 10
Topic 6
Topic 2
Topic 5
Topic 3
Topic 11
Topic 4
Topic 12
Topic 1
Topic 8
Exclusivity
Semantic Coherence

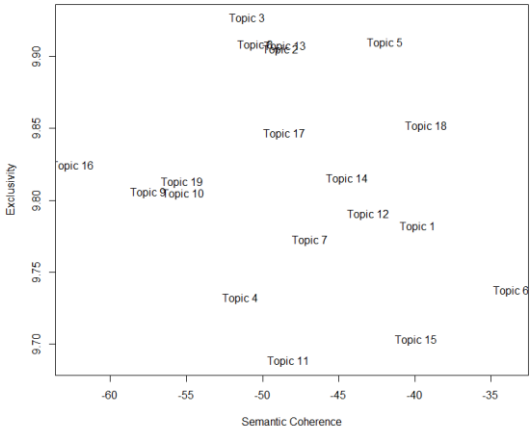

Topic 3
Topic 12
Topic 6
Topic 1
Topic 10
Topic 7
Topic 4
Topic 2
Topic 5
Topic 13
Topic 11
Topic 8
Exclusivity
Semantic Coherence


Topic 6
Topic 7
Topic 13
Topic 5
Topic 2
Topic 3
Topic 15
Topic 4
Topic 12
Topic 1
Topic 11
Topic 14
Topic 9
Topic 8
Exclusivity
Semantic Coherence


Topic 3
Topic 14
Topic 1
Topic 13
Topic 11
Topic 16
Topic 10
Topic 4
Topic 2
Topic 15
Topic 9
Topic 5
Topic 6
Exclusivity
Semantic Coherence


Topic 12
Topic 2
Topic 7
Topic 4
Topic 14
Topic 8
Topic 5
Topic 11
Topic 6
Exclusivity
Semantic Coherence


Topic 5
Topic 8
Topic 7
Topic 3
Topic 1
Topic 17
Topic 4
Topic 16
Topic 6
Topic 2
Topic 10
Topic 18
Topic 9
Topic 14
Topic 11
Exclusivity
Semantic Coherence


Topic 3
Topic 5
Topic 18
Topic 17
Topic 14
Topic 12
Topic 1
Topic 7
Topic 4
Topic 6
Topic 15
Topic 11
Exclusivity
Semantic Coherence

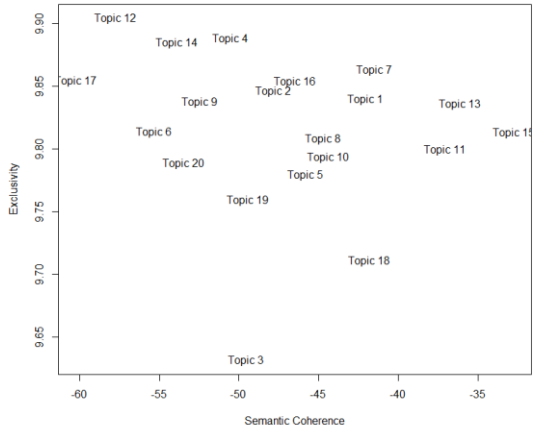


Figure 9.1: Semantic coherence and exclusivity for topic models with 2 to 20 topics
Source: Own calculation

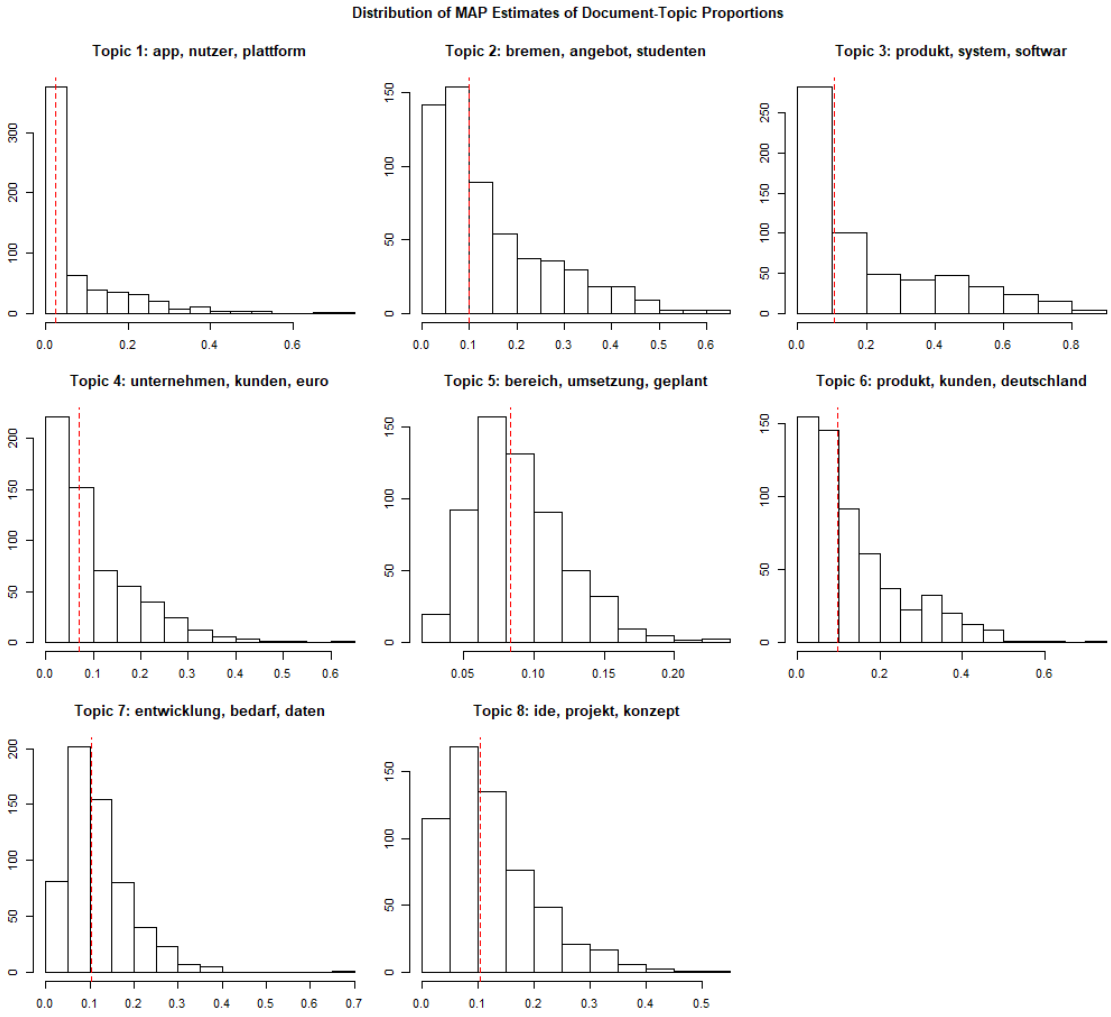


Figure 9.2: Distribution of topics over submissions to the "CampusIdeen" competition from 2003-2019
Source: Own calculation

Table 9.1: Similarity of topics measured by the cosine index
Source: Own calculation

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
Topic 1								
Topic 2	0.1333							
Topic 3	0.1222	0.1044						
Topic 4	0.0892	0.0650	0.1048					
Topic 5	0.0587	0.0779	0.1164	0.0497				
Topic 6	0.1230	0.1253	0.4200	0.2779	0.0601			
Topic 7	0.1343	0.0937	0.1583	0.0509	0.0330	0.0656		
Topic 8	0.1107	0.0468	0.1771	0.0884	0.0590	0.0595	0.1523	

Table 9.2: Correlation of topics measured by the pearson correlation
Source: Own calculation

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
Topic 1								
Topic 2	-0.1479							
Topic 3	-0.3127	-0.5859						
Topic 4	-0.0983	-0.0514	-0.2943					
Topic 5	-0.1186	0.1305	-0.2442	0.1185				
Topic 6	-0.0252	-0.1786	-0.2282	-0.0967	-0.1864			
Topic 7	-0.1672	-0.1567	0.0827	0.1545	0.4140	-0.4544		
Topic 8	0.0000	0.5817	-0.5263	-0.2525	-0.0639	0.0281	-0.3224	

Table 9.3: Correlation matrix of included variables into the classification tree analysis for innovative business ideas (n=247)
Source: Own calculation

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Year	Team	Conclusive-ness	Realistic Implementation	Economic potential
Topic 1													
Topic 2	-0.08												
Topic 3	-0.44	-0.56											
Topic 4	-0.11	0.02	-0.28										
Topic 5	-0.07	0.10	-0.20	0.10									
Topic 6	0.00	-0.15	-0.29	-0.10	-0.18								
Topic 7	-0.14	-0.19	0.08	0.20	0.43	-0.43							
Topic 8	0.15	0.60	-0.54	-0.27	-0.09	0.05	-0.36						
Year	0.27	-0.22	-0.08	0.08	-0.04	0.20	-0.03	-0.22					
Team	0.07	-0.12	-0.11	0.32	0.03	0.17	-0.03	-0.24	0.30				
Conclusiveness	-0.13	0.05	-0.01	0.05	0.03	-0.02	0.02	0.09	-0.22	-0.04			
Realistic Implementation	0.08	0.10	-0.09	-0.01	-0.02	-0.11	0.05	0.15	-0.19	-0.07	0.67		
Economic potential	-0.05	-0.12	0.10	0.06	0.02	-0.07	0.06	-0.08	-0.13	-0.03	0.66	0.68	

Table 9.4: Correlation matrix of included variables into the classification tree analysis for non-innovative business ideas (n=182)
Source: Own calculation

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Year	Team	Conclusive-ness	Realistic Implementation	Economic potential
Topic 1													
Topic 2	-0.22												
Topic 3	-0.22	-0.52											
Topic 4	-0.15	-0.11	-0.28										
Topic 5	-0.20	0.14	-0.21	0.19									
Topic 6	-0.20	-0.20	-0.16	-0.16	-0.17								
Topic 7	-0.14	-0.12	0.11	0.28	0.44	-0.48							
Topic 8	-0.11	0.52	-0.41	-0.29	-0.12	-0.01	-0.31						
Year	0.33	-0.16	-0.15	-0.16	-0.12	0.38	-0.19	-0.21					
Team	0.28	-0.17	0.01	-0.10	-0.30	0.18	-0.15	-0.17	0.38				
Conclusiveness	-0.07	0.06	-0.10	0.17	0.24	-0.07	0.04	0.01	-0.20	0.01			
Realistic Implementation	0.04	0.02	-0.09	0.10	0.17	-0.08	0.00	0.01	-0.20	-0.12	0.66		
Economic potential	-0.11	0.04	-0.06	0.19	0.21	-0.08	0.11	-0.04	-0.22	-0.07	0.70	0.67	

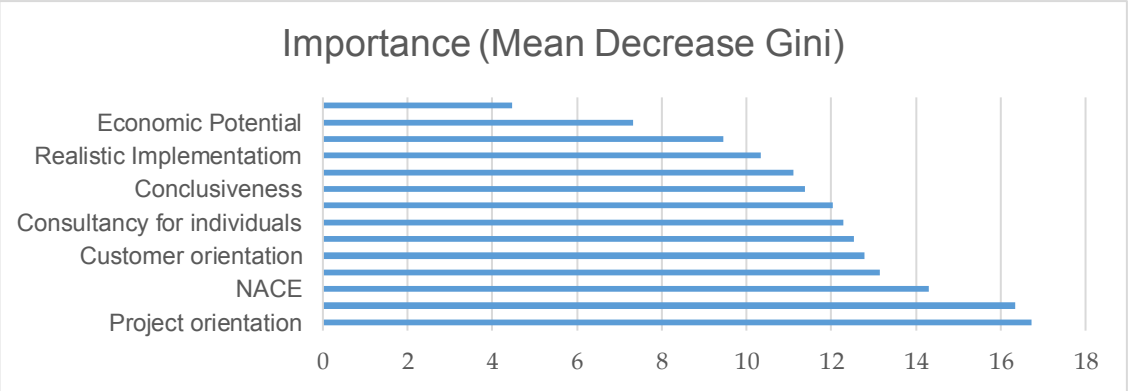


Figure 9.3: Variable importance as revealed by the random forest analysis based upon all submitted business ideas (n=593)
Source: Own calculation

Imprint

Bremen Papers on Economics & Innovation

Published by
University of Bremen, Faculty of Business Studies & Economics,
Institute for Economic Research and Policy (ierp)
Max-von-Laue-Straße 1, 28359 Bremen, Germany

Bremen Papers on Economics & Innovation #2109

Responsible Editor: Prof. Dr. Jutta Günther

ISSN 2629-3994