



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

Jessica Birkholz¹ 

Accepted: 6 September 2022 / Published online: 30 September 2022
© The Author(s) 2022

Abstract

It is of key importance to identify the degree of novelty and probability of incorporation of business ideas in an early stage, so that targeted support of these different types of entrepreneurship is possible. Selection of business ideas for investments and support programs rely on quantitative and qualitative metrics. The qualitative assessment, however, is biased by subjective impressions and experiences of the decision-maker. Therefore, this paper examines the narrative of business idea descriptions to improve the identification of the degree of novelty and to enhance the estimation of the incorporation probability by advancing the objectivity of qualitative metrics. The paper aims to answer two questions: (1) Are there differences in topic prevalences in novel and non-novel business ideas?, and (2) Does the composition of topics related to a business idea influence its incorporation probability? Structural topic modelling 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 novel and non-novel business ideas with higher incorporation prospects.

Keywords Business idea · Machine learning · Incorporation

JEL Classification O30 · L26 · O38

1 Introduction

Entrepreneurship is widely acknowledged to be of key importance for economic development and tackling societal challenges (Asarkaya & Keles Taysir, 2019; Stuetzer et al., 2013). Treating entrepreneurial activities as a homogenous phenomenon fails to do justice to the broad range of activities and effects these have on their environment (Friar & Meyer, 2003; Szerb et al., 2019; Wong et al., 2005). Entrepreneurial activities with a high degree of novelty, as defined by Dahlqvist and Wiklund (2012), are characterised by affecting a

✉ Jessica Birkholz
jessica.birkholz@uni-bremen.de

¹ Faculty of Business Studies and Economics, University of Bremen, Max-von-Laue-Straße 1, 28359 Bremen, Germany

great geographical area and a wide range of customers in contrast to non-novel activities that either focus on an existing market or a small group of customers. Resulting from these essential variations between entrepreneurial activities, the generation and refinement of novel and non-novel business ideas require imagination of distinct future scenarios and derivation of actions (Kier & McMullen, 2018). The quality of these imagination processes determines the resulting business idea narrative which lays the foundation of the subsequent development of the entrepreneurial endeavour and, thus, its incorporation success (Kier & McMullen, 2018). Therefore, entrepreneurial activities with high and low degree of novelty are pursued based on different motivations and antecedents as well as the path to incorporation differs (Asarkaya & Keles Taysir, 2019; Friar & Meyer, 2003).

To maximise the impact of different kinds of entrepreneurship, the challenge is to identify at an early stage the degree of novelty and incorporation prospects of entrepreneurial activities (Minola et al., 2017). The first phase of decisions, which entrepreneurial endeavour to support or to invest in, is often the screening of entrepreneurial business ideas on quantifiable metrics such as the characteristics of the targeted market, the entrepreneurial team, and resources (Rasmussen & Sorheim, 2012; Song et al., 2008). In a second stage of this process decision-makers, merely rely on qualitative criteria to base their investment decision on, e.g. the presentation and communication of the business idea (McAdam & Marlow, 2011). This step is known to be highly influenced by subjective judgements of the decision-makers (Drnovsek et al., 2018). This is why many authors call for more research to extent objective metrics in this process (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 novel ideas that have a high likelihood to be successfully incorporated (McAdam & Marlow, 2011; Meyskens & Carsrud, 2013; Pandey et al., 2017; Pare et al., 2011). Because the incorporation success of business ideas relies on multiple factors and their interplay, the evaluation of such ideas is featured by a high degree of complexity (Minola et al., 2017; Zacharakis & Meyer, 2000). Research focussing on factors separately does not take into account this complex nature of incorporation processes, leading to the need to analyse the interdependencies of factors that influence entrepreneurial activities.

The purpose of this study is to examine the complex relation between the narrative of novel and non-novel 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 narrative, the written description of a business idea, contains different topics (e.g. customer segment, main product) that are uniquely combined with each other. 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 novel business ideas and 182 text descriptions of non-novel business ideas were handed in to the competition. Machine learning techniques are applied to answer two questions: (1) Are there differences in topic prevalences in novel and non-novel business ideas? and (2) Does the composition of topics influence the incorporation probability? To answer these, structural topic modelling (Roberts et al., 2014) and classification tree analysis (James et al., 2013; Strobl et al., 2009) are employed.

The analyses yield four main results. First, differences in topic prevalences, the frequencies in which every topic appears in a text corpus, in novel business ideas and non-novel ones are identified. Topics connected to the products and key processes for revenue generation, purpose and orientation type of the business occur in different frequency in novel and non-novel business ideas. Second, the interdependence of topics and their importance for incorporation probability differ depending on the degree of novelty of the business ideas.

Third, novel 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 novel 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 more aspects. First, it delivers conclusions about the underexplored pre-foundational stage of entrepreneurial activities and a more detailed understanding of the differences between novel and non-novel business ideas. Second, the pre-foundational information is connected to incorporation probability rather than entrepreneurial intentions to start a new venture. This translates into an application-oriented approach, providing valuable insights for the design of targeted entrepreneurship policies, support measures and investment decisions. For example, early stage investors and managers of accelerators and incubators can apply these results and the methodological procedure for the selection of promising entrepreneurial projects in the screening phase. 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. Overall, the results of the present study suggest that the narrative of business ideas has the capability to complement the entrepreneurial intention as the most prominent indicator for new venture creation. Business ideas can be understood as plans of actions to implement an imagined venture into a venture operating on the market and, thus, is likely to follow entrepreneurial intentions and bridge the gap to entrepreneurial action.

The remainder of this paper is organised as follows: the next chapter discusses the theoretical background of different types of entrepreneurial activities and business idea narratives. 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 modelling 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 Novel and non-novel entrepreneurial activities

The differentiation of entrepreneurship into two types is found by former research to be especially distinct in their dynamics – novel and non-novel entrepreneurship endeavours (e.g. Szerb et al., 2019; Wong et al., 2005). The first type of entrepreneurial activities provides services or products with a high degree of novelty which may translate into innovations. Entrepreneurial opportunities that are characterised by the possibility to provide new services and products to a broad geographical area and to a broad range of customer groups build the basis for this type of entrepreneurial activities (Dahlqvist & Wiklund, 2012). The second type of entrepreneurial activities is characterised by a lower degree of novelty and makes use of arbitrage possibilities, relying on the mobilisation of resources (Dahlqvist & Wiklund, 2012).

Advancements in economic development stem from highly innovative and technology based entrepreneurial endeavours and cannot be attributed to non-novel new ventures (Autio & Yli-Renko, 1998; Shane, 2009). Two dynamics lead to this phenomenon. First, evidence suggests that most new ventures try to gain access to an existing market with already existing products and services, thus, 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). This lack of economic aspirations is also correlated with a lower probability that business ideas are turned into new ventures (Hossinger et al., 2021). 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).

2.2 Decision-making processes to select business ideas

Entrepreneurial opportunities that are characterised by a high degree of novelty, imply that the possible return on investment is likely to be high, because an entrepreneur that acts upon this possibility is addressing a broad scope of geographical areas and customer groups (Dahlqvist & Wiklund, 2012). Thus, it is suspected that business ideas in this vein are of key interest for investors, policy-makers, managers of accelerators and incubators, although the uncertainties connected to novel business ideas are considerable caused by the need of investments prior to market entry (McAdam & Marlow, 2011).

To allocate investments and support measures efficiently between business ideas, the most promising ideas need to be identified (Minola et al., 2017; Zacharakis & Meyer, 2000). Examples of allocation processes are business plan competitions and crowdfunding campaigns, in which resources are awarded, or accelerators and incubators, which have a limited number of spaces in their programs to be assigned. The rationale in such situations is to allocate resources to those that have a business idea that is likely to turn into a new venture operating on the market (Minola et al., 2017; Zacharakis & Meyer, 2000). In these cases, decisions are usually made in two main steps that require the application of different criteria the business ideas have to fulfil to pass (Minola et al., 2017; Rasmussen & Sorheim, 2012). The first stage of the decision-making process is the screening of many proposals to identify those that deserve further consideration (Minola et al., 2017; Tyebjee & Bruno, 1984; Zacharakis & Meyer, 2000). This procedure focusses on quantifiable criteria as for example the sustainability and the growth potential of a business idea (McAdam & Marlow, 2011). There is extensive research on which, mostly quantifiable, factors investors base their decision on. Factors classified in the categories related to the market, to the product, to the managerial competences of the entrepreneurial team, external barriers, resource capacities, competitive strategies, the characteristics of customers, and financial characteristics of the venture or more specifically the potential revenue of the investment are listed in former studies (Pare et al., 2011). Song et al. (2008) comprised these factors in a meta-analysis to the categories market and opportunity, entrepreneurial team, and resources.

The second stage of the decision-making process often relies on qualitative criteria as for example the vision of the entrepreneur, to convey the essence of the business opportunity (McAdam & Marlow, 2011), preparedness of the entrepreneurs (Chen et al., 2009; Rasmussen & Sorheim, 2012), a promotion focus and perceived feasibility of the business

idea (Drnovsek et al., 2018). Intangible assets in regard to communication are of crucial importance for the success of a business idea in this second stage of the process, especially because entrepreneurs at this early stage typically cannot refer to a convincing track record of achievements (Rasmussen & Sorheim, 2012; van Werven et al., 2019). However, the effect of the communication of the same business idea can vary depending on the fit between investors and entrepreneurs (Cox et al., 2017) and the cognitive scheme that govern the attention and information evaluation of the investors (Drnovsek et al., 2018). Hence, the decision-making process in which business ideas are evaluated is characterised by a large degree of subjectivity.

As the screening and reviewing of business ideas is known to be a time-consuming task, former studies aimed at reducing the effort needed to perform this procedure by automatising it (Minola et al., 2017; Roure & Keeley, 1990; Zacharakis & Meyer, 2000). Actuarial models not only reduce the time needed to identify promising business ideas, but also overcome the subjectivity of human assessment by weighting all information cues equally whereas experts are biased by their cognitive scheme and experiences (Zacharakis & Meyer, 2000). Moreover, these models are shown to often outperform experts in the fields of bank lending and psychological evaluations (Zacharakis & Meyer, 2000). It is assumed that this is also relevant for the case of highly novel entrepreneurial projects because these are more difficult to communicate and the estimation of their potential depends more on subjective evaluations, thus, information asymmetries are still a burden to access external funding (Proksch et al., 2017). Although this approach of actuarial models is not new, one part of business idea descriptions was neglected in former models – a formerly understated factor for success or failure of business ideas – their narrative (Kier & McMullen, 2018; McAdam & Marlow, 2011; Rasmussen & Sorheim, 2012). An incorporation of the narratives of business ideas into an actuarial model provides the benefit that the qualitative assessment of business ideas can be complemented by an unbiased model that enables experts and investors to classify the idea more objectively as promising or not and, thus, opens the way for more efficient targeted entrepreneurship support.

2.3 Importance of business idea narratives

Business idea descriptions provide a rich resource for information in early phases of the entrepreneurial process (Pare et al., 2011). New venture ideas, as defined by Davidsson (2015), are imagined future ventures. Related to this, Vogel (2017) defines venture concepts as the nascent stage of a business model and a vague understanding of its components. Thus, a business idea can be defined as a description of new venture ideas and concepts of how this imagined business works and how revenue is generated (Klofsten, 2005). This perspective 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). All these aspects establish a desired endpoint, i.e. an imagined business in various facets, which is then complemented by events and characters, linked in a temporal sequence, that are relevant to develop the business towards this desired endpoint. This plan of actions to implement the imagined venture can be referred as narrative (Cunliffe et al., 2004; Pentland, 1999; van Werven et al., 2019). Thus, business ideas and their narratives provide information, which is likely to lay the foundation for subsequent developments and performance of the incorporation (Davidsson, 2015; O'Connor, 2002; Pandey et al., 2017; Pare et al., 2011; Vogel, 2017).

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 early entrepreneurship support by enabling the identification of business ideas with higher prospects (Chattopadhyay & Ghosh, 2008). Although the business idea may be subject to change over time, entrepreneurs have to consider the relational and temporal commitments related to the initial business idea, thus, changing the core concept comes with restrictions (Berends et al., 2021). Despite the key importance of the narratives of business ideas, studies on these are rare. Thus, this study has two purposes. The first aims to identify differences in the narrative between novel and non-novel business ideas. The second purpose is dealing with the influence the narrative structure has on the incorporation probability of business ideas.

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

To be able to make use of the business idea narrative as an unbiased indicator for novel ideas, it is of key importance to identify topics, i.e. subparts of the narrative, within the business ideas which allow a distinction between novel and non-novel business ideas. Former research finds evidence for differences between entrepreneurial projects with high and low novelty 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) are found to differ. These findings lead to the assumption that novel and non-novel business ideas aim to enter different industries, markets and have different customer groups which is connected to the underlying entrepreneurial opportunity structure (Dahlqvist & Wiklund, 2012). Thus, the assumption is that there are topics central to novel business ideas, which play only a minor role in non-novel business ideas and vice versa. This is likely to translate into different topic proportions, i.e. the narrative, in the business ideas. In addition, it can be argued that topics are combined with each other in a different manner in novel or non-novel business ideas. Therefore, it is suspected that the relations between topics differ with the degree of novelty of the business ideas. Thus, two proposition are formulated as follows:

Proposition 1 *Topic proportions in novel and non-novel business ideas differ.*

Proposition 2 *Correlations between topics associated with novel and non-novel 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 and their interplay to the incorporation probability of the business idea in each of these types—novel and non-novel (McAdam & Marlow, 2011; Meyskens & Carsrud, 2013; Pandey et al., 2017).

As the enactment of novel entrepreneurial opportunities is often connected to high-technology or growing industries and a broader spectrum of customers that are addressed than it is the case for non-novel endeavours (Dahlqvist & Wiklund, 2012), it is hypothesised that these activities require different behaviours and visions of the entrepreneurs to be

successfully incorporated. This is underlined by theoretical considerations made by Schumpeter (1934) and Kirzner (1973). The innovative entrepreneur described by Schumpeter (1934) is characterised by specific personality traits and attitudes that enable the entrepreneur to go against the odds and break with routines. In contrast to this, the non-innovative entrepreneur as described by Kirzner (1973) is interpreted as an individual attentive to its environment and mobilising resources to use arbitrage possibilities. This highlights that the entrepreneurial activities as well as the actors pursuing those, differ depending on the degree of novelty, implying that the mechanisms of incorporation differ as well. 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 narratives serve as determinants of incorporation probability of business ideas of the two types. Successfully incorporated business ideas are hypothesised to be able to provide insights about which narratives correspond to an existing opportunity that the entrepreneurial project acted upon (Vogel, 2017). 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 novel and non-novel 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 submitted business ideas are evaluated by an expert jury that consist of members of the entrepreneurial ecosystem in Bremen and Bremerhaven. A jury of six to eight experts is staffed with members of the tertiary institutions involved in the ‘BRIDGE’ network,¹ members of the chamber of commerce and companies that sponsor the competition that also act as investors. Long-standing cooperation between these partners facilitated a stable composition of actors of the local entrepreneurship ecosystem in the expert jury. The jury was instructed to evaluate the business idea submission based on a questionnaire, including aspects as novelty and creativity of the business idea, the conclusiveness of the idea, the degree of a realistic implementation, and the economic potential of the idea. These aspects were surveyed in all years under

¹ These tertiary institutions include the University of Bremen, the City University of Applied Sciences Bremen, the University of Applied Sciences in Bremerhaven and the Jacobs University Bremen. Jury members from these institutions were professors in the field of entrepreneurship or deputy rectors. Moreover, the Bremer Aufbau-Bank is member of the ‘BRIDGE’ network.

investigation in a consistent manner.² 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 and complete submissions are considered in the analysis.³ 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 Vitae 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 lemmatised. Words classified as adposition, auxiliary, coordinating conjunction, determiner, numeral, particles, pronouns, subordinating conjunction, adverb, interjection and auxiliary verbs⁴ were added to the list of stop words.⁵ 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 the results of the topic model (Birkholz et al., 2021). The retrieved unigrams build the basis for the subsequent topic model.⁶

Table 1 shows the mean and median number of words per submission differentiated by their degree of novelty. These indicate longer submission texts for novel business ideas than for non-novel 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 novel which results in a proportionally higher number of total words considered.

3.2 Degree of novelty of business ideas

Submissions from 2005 onward are rated by an expert jury evaluating their degree of novelty. This rating serves as a basis to classify the submissions. A rating above average in

² The quantitative scale used to measure the fulfilment of these criteria were consistently surveyed by using a 5-point Likert scale. Regarding the wording of the criteria, only in the case of novelty and creativity of the business idea changed slightly. In the years 2005 until 2008 this criterion was surveyed by referring to the novelty of the concept and from 2009 onwards this was extended by creativity of the concept. Other criteria that changed across these years in their quantitative assessment (changes in the scale) and, thus, were not considered for analysis.

³ 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.

⁴ Including modal verbs as subgroup of auxiliary verbs.

⁵ Since lemmatisation for the German language is not entirely classifying words in a correct manner, these words were checked manually.

⁶ 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).

Table 1 Descriptive statistics of the text descriptions based on the submitted business ideas

	Novel business ideas	Non-novel 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	434.014	435.653
Sum of words over all submissions	184,342	127,303
Number of submissions	247	182

Source Calculation based on business ideas submitted to ‘CampusIdeen’ competition from 2005 to 2019 with a jury rating of novelty with higher than average score (novel business ideas) and lower than average score (non-novel business ideas)

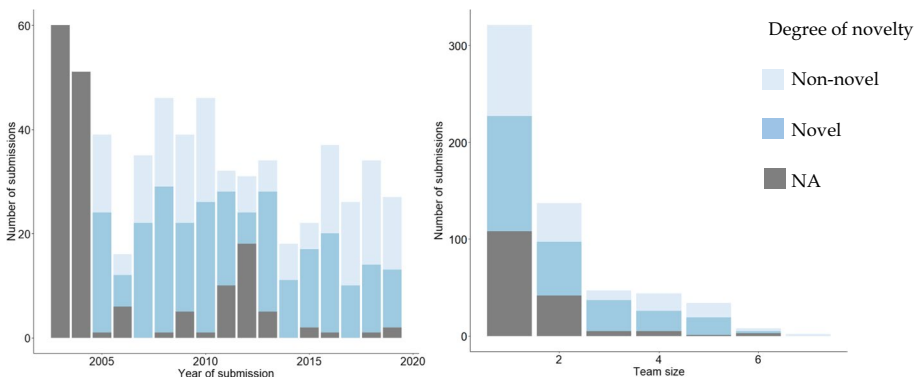


Fig. 1 Number of novel and non-novel submissions to the business idea competition per year (left figure) and per team size (right figure). *Source* ‘CampusIdeen’ competition submissions 2003–2019

newness is interpreted as novel and a rating below average in this criterion is associated with non-novel business ideas.⁷

Figure 1 illustrates the number of submissions to the competition, separated in novel and non-novel, in each year and for each team size. No clear time trend in the number of submissions, or their degree of novelty is evident.⁸ The vast majority of submissions are

⁷ 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 novel. The alternative to use the median value (3 on the Likert-scale) to split the data set into novel and non-novel business ideas is not taken into consideration here, because this approach would lead to the loss of 88 business ideas that have an evaluation equal to the median value. Thus, an assignment to one group (novel or non-novel) is for these business ideas not possible. However, when the analyses of Chapter 5.2, 5.3, and 5.4 are conducted with the median split sample, only slight changes in the results are evident, showing the robustness of the findings. The results of the robustness checks are available upon request.

⁸ The NAs in the Fig. 1 correspond to missing data on the rating of novelty. Since all submissions can be used in the first analysis step, independently of the availability of the rating of the novelty, these data points are also displayed in the figures.

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 novel ideas.

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.⁹ The incorporation rate is 24% in the case of non-novel business ideas and 27% in the case of novel ones.

4 Methodological approach

In order to test the propositions, a four-step analytical procedure is applied. The research design proceeds after the data filtering process with structural topic modelling to reveal latent topics in business ideas. Since the topic modelling approach does not require information about the degree of novelty, all submission texts, amounting to 593 documents with a total of 400,854 words, are utilised for this step. In a second step, novel and non-novel 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 novel and non-novel 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 novel ideas and one explaining the foundation success of non-novel 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.

4.1 Topic modelling

Topic modelling 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 (Lee & Kang, 2018; Storopoli, 2019). Due to the rising importance of text data, topic modelling is applied in economic analyses. For example Ambrosino et al. (2018) applied topic modelling to uncover topics discussed in economic scientific literature and illustrated the development of the discipline over time. A similar methodological approach was used by Lee and Kang (2018) to identify topics in the field of technology and innovation management studies. Other than using the topic modelling approach to examine the scientific literature, Arora et al. (2020) analysed website data to examine dynamic capabilities of new ventures. Moreover, this analytical approach is transferable for the analysis of open ended survey questions as shown by Roberts et al. (2014).

⁹ 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).

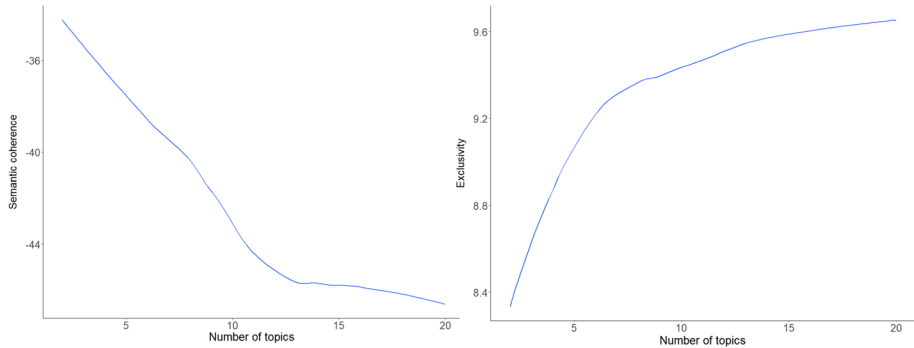


Fig. 2 Semantic coherence (left figure) and exclusivity (right figure) per number of topics. *Source* Own calculation

Latent Dirichlet Allocation (LDA) is one of the most common algorithms to perform topic modelling 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 modelling (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 smartphone and tablet applications 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 modelling 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 exclusivity¹⁰ are considered to determine an appropriate number of topics. Figure 2 shows the indicators over a range of number of topics.¹¹ Both plots show a salient point around ten topics. Plotting the semantic coherence and exclusivity of each topic in each topic model against each other is applied to identify the topic model

¹⁰ 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 (Chang et al., 2009; Silge 2018).

¹¹ 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 Fig. 2 appears to show a clear trend, thus the inclusion of more topics would lead to the same result.

with high exclusivity and high semantic coherence, which is considered to be ideal (Chang et al., 2009). The topic model containing eight topics fulfils this to the largest extent.¹²

In order to validate the first and the second proposition, topic proportions within novel business ideas are compared to these in non-novel ones. Welch's two sample t-tests (Welch, 1947) are applied to identify significant differences in proportions. In a second step, correlation networks are calculated to reveal the topic structures of novel and non-novel 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 novel business ideas and one classification tree for the subset of non-novel ideas.¹³

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 design for support (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. Because of these characteristics of classification trees, this approach was applied to identify turning points in business process orientation (Milanovic Glavan et al., 2015) and to estimate the economic success of an entrepreneur (Chattopadhyay & Ghosh, 2008).

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 'Nomenclature générale des Activités économiques dans les Communautés Européennes' (NACE) classification of industries is given in Fig. 3. The most prominent domains are manufacturing (C), and information and communication (J).

A further set of control variables are jury ratings describing the quality of the business ideas. The data shows that novel business ideas reach higher mean scores in conclusiveness, realistic implementation, and economic potential than the non-novel business ideas, as shown

¹² The topic models refer to the Pareto optimal one out of 50 runs as incorporated by the R package 'stm', function 'manyTopics'. See Fig. 7 in the appendix for the distribution of the topics over the documents, Table 4 for the correlation of topics in their occurrence and the similarity of topics measured by the cosine index.

¹³ Please see Table 5 and Table 6 in the appendix for the correlation matrices of all variables considered in the analyses.

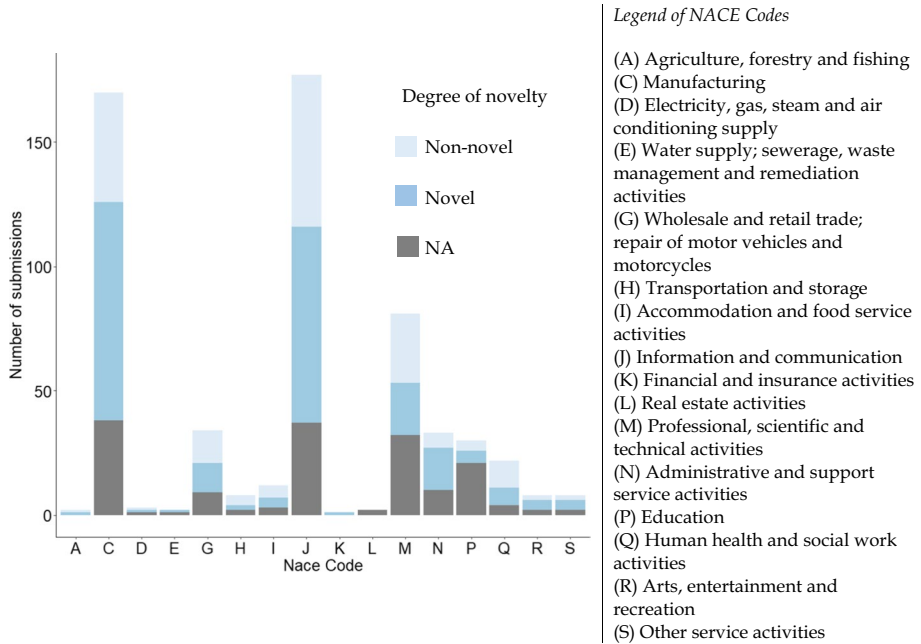


Fig. 3 Distribution of business ideas over industry classifications (NACE). Note: The industry classification is assigned to the business ideas with manual evaluation of each submission text. Source ‘CampusIdeen’ competition submissions 2005–2019

Table 2 Descriptive statistics of jury ratings of business ideas

Statistic	N	Mean	St. Dev.	Min	Pctl (25)	Pctl (75)	Max
<i>Non-novel business ideas</i>							
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
<i>Novel business ideas</i>							
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

Source ‘CampusIdeen’ competition submissions 2005–2019

in Table 2. Furthermore, the variety in novel business ideas is higher reflected by the higher standard deviations in all variables.

Table 3 Topics identified by structural topic modelling analysis based on all business idea submissions

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
Purpose	Social entrepreneurship (Topic 8)	12.2%	idea, project, concept, children, find, interest, theme, contact, providing, location ide, projekt, konzept, kinder, finden, interest, 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, zusatzlich
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, markt, 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, anbieter, 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

Source Own calculation

5 Results

5.1 Identification of latent topic structures in business ideas

Table 3 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 common topic. The most common words per topic and illustrative text excerpts¹⁴ 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 of coherence and exclusivity. The topic 'Development of technical devices and software' reflects 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 *purpose* of the entrepreneurs. This is especially present in the case of the topic 'Social entrepreneurship', showing a focus on societal impact of the activities.

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

In order to identify peculiarities of novel business ideas in the topic structure, the documents of business ideas are divided into two groups. One group containing novel business ideas, and the other group consisting of non-novel 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 novel business ideas. Figure 4 illustrates the distribution of topic shares in the two groups of business ideas.

¹⁴ The text excerpts cannot be shown here due to confidentiality of the data.

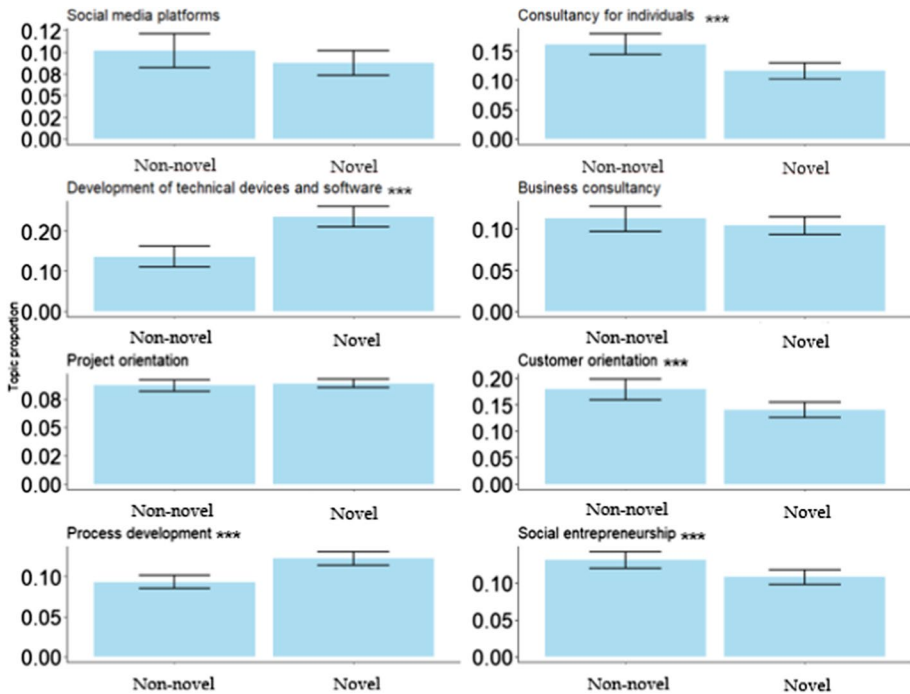


Fig. 4 Comparison of topic proportions in novel and non-novel business ideas. Note: The error bars corresponds to a confidence interval of 95%; ***corresponds to a significant difference with a p -value lower than 0.01. *Source* Own calculation

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 novelty and, thus, is a building block for the understanding of novel business ideas. Based on this result, the first proposition can be confirmed.

Two topics are significantly more associated with novel business ideas compared to non-novel 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-novel compared to shares in novel 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-novel ideas.

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

5.3 Analysis of topic interdependencies of novel and non-novel business ideas

To properly discriminate novel and non-novel business ideas, not only the proportion per topic is relevant, but also how the topics are interrelated with each other. For this purpose, the correlation structures of topics are examined for novel and non-novel business ideas as illustrated in Fig. 5. The dots in the graphics represent the variables and the blue coloured ones highlight the variables that were identified in the former step as mainly appearing in novel or non-novel business ideas. Green lines between the variables indicate positive correlations and red ones correspond to negative relations. The transparency of the edges reflect the significance of correlation.

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

The relations between the topics in the case of novel business ideas reveal that ‘Development of technical devices and software’ (Tp3) serves as an indicator for novel ideas when this topic is central to the idea and not combined with other topics. The second topic associated with novel business ideas ‘Process development’ (Tp7) can be used as an indicator for novelty when combined with ‘Business consultancy’ (Tp4) and ‘Project orientation’ (Tp5). The degree of novelty 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-novel. Moreover, ideas solely emphasising ‘Customer orientation’ (Tp6) can be assumed to be non-novel.

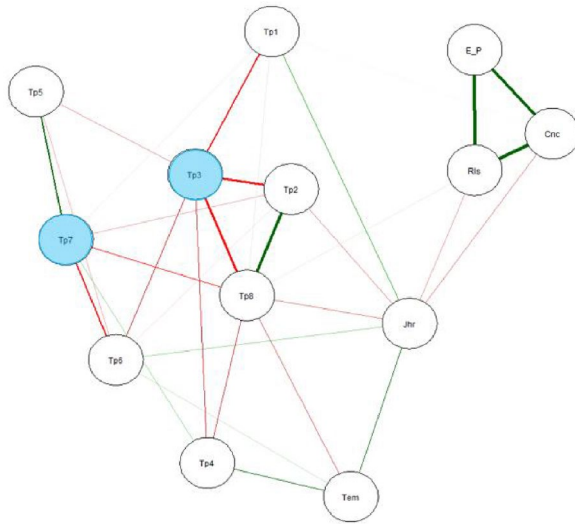
The meta-variables ‘Economic Potential’ (E_P), ‘Realistic Implementation’ (RIs) and ‘Conclusiveness’ (Cnc), in particular, show more links between themselves and the topics in the case of non-novel business ideas.¹⁵ These findings could be explained by the difficulty to evaluate novel business ideas 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.

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

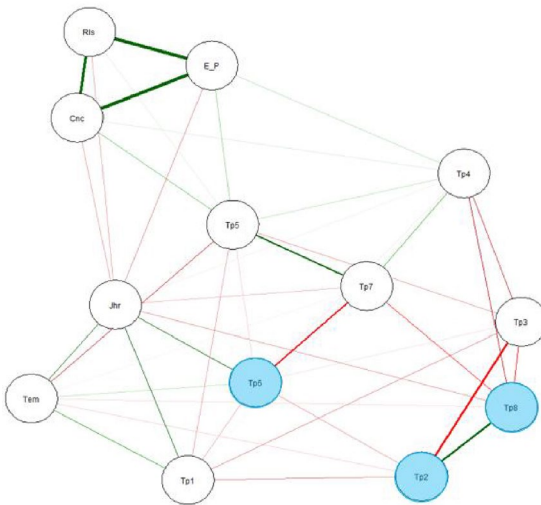
Classification trees are applied to examine the interplay of topic structures of business ideas and their impact on incorporation probability. The results are illustrated in Fig. 6. The first performed split corresponds to the most important variable to explain the incorporation probability. The value 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

¹⁵ This finding indicates that the expert jury of the business idea competition attests non-novel business ideas dealing with ‘Business consultancy’ (Tp4) higher ‘Economic Potential’ (E_P) and ‘Conclusiveness’ (Cnc). Moreover, the jury evaluated non-novel business ideas featuring a ‘Project orientation’ (Tp5) more positively in terms of economic prospects (E_P), conclusiveness (Cnc) and realistic implementation (RIs).

Novel business ideas



Non-novel 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

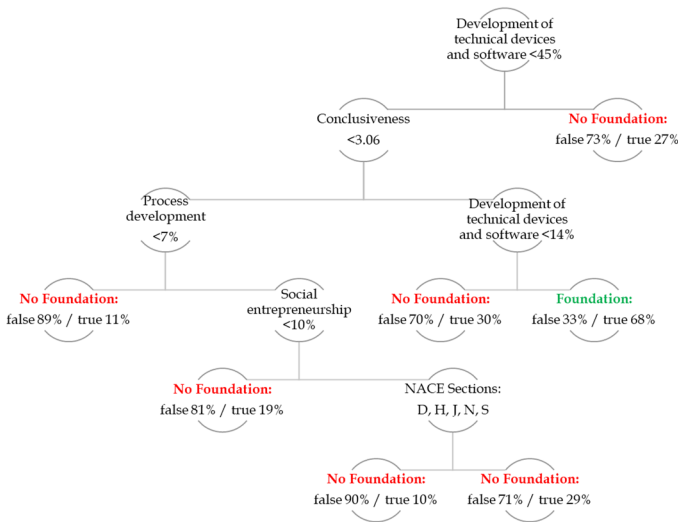
(Tp6) Customer orientation

(Tp7) Process development

(Tp8) Social entrepreneurship

Fig. 5 Correlation networks of novel and non-novel 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 5 and 6 in the appendix. The blue nodes in the graphic represent the topics identified to be specific for novel and non-novel business ideas (Color figure online). *Source* Own construction

Novel business ideas



Non-novel business ideas

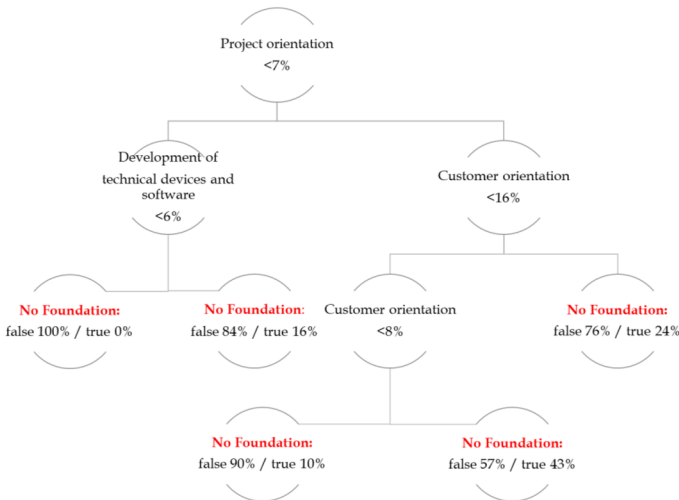


Fig. 6 Results of the classification tree analyses for novel and non-novel business ideas. Note: The sample sizes correspond to the number of submitted documents that are classified as innovative or non-innovative (number of novel observations: 247; number of non-novel observations: 182). *Source* Own calculation

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.¹⁶ 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 purpose. 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.¹⁷ Moreover, the classification tree for non-novel business ideas follows a simpler structure than the tree for novel ones. This could stem from a lower sample size for the former classification tree and the condition that each leave must at least contain 30 business ideas.

5.4.1 Novel business ideas

The classification tree on the left side of Fig. 6 is based on the novel business ideas. The most important splitting criteria for novel 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),

¹⁶ 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 novelty or by the sample with a jury rating. The variable importance, as the main result of the random forest analysis, is illustrated in Fig. 8 in the appendix.

¹⁷ 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.

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 novel business ideas profit from combination of various topics.

5.4.2 Non-novel business ideas

The classification tree on the right side of Fig. 6 reflects the pattern within the non-novel 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-novel, 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-novel business ideas require a clear focus on one market—either B2C (as reflected in Topic 6) or other clients.

6 Discussion

Business ideas bridge the gap between entrepreneurial intentions and entrepreneurial actions that translate into economic development. The factors that enable an efficient usage of capabilities of individuals and innovative or problem-solving ideas is of crucial importance to steer the transfer of skills and solutions from individuals to the economy. Therefore, this study focusses on the information business idea descriptions provide and their relation to novelty and likelihood of incorporation. In the following the contributions to research, practical implications and avenues for future research are discussed.

6.1 Contributions to research

This study analyses the topics appearing in novel and non-novel business ideas. The topics identified in this study, reflect three aspects of entrepreneurial activity – entrepreneurial purpose, business orientation and main products or processes. Thus, the application of topic modelling 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 novel and non-novel business ideas, can be answered with a yes. The narratives of novel and non-novel business ideas differ in their topic proportions. In addition, the correlations between the topics and meta-data provide a differentiated perspective on the narratives of business ideas. The topic proportions already give a hint at how to identify novel business ideas, which is further enhanced by taking the correlations between the topics into account. 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 novelty (McAdam & Marlow, 2011), but also that entrepreneurial purpose and business orientation are important differences between novel and non-novel business ideas. Moreover, not only topic proportions differ in these two groups of business ideas, but also their correlation with external factors. Non-novel 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', 'Realistic Implementation', and 'Conclusiveness', 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). This corresponds to findings of former literature, stating that the heterogeneity of a team influences the entrepreneurial orientation of a business idea (Diáñez-González & Camelo-Ordaz, 2016).

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 purpose as well as business orientation. This holds true for novel and non-novel business ideas. However, the importance of each specific topic differs depending on the degree of novelty of each idea.

Novel business ideas show two distinct paths in the analysis. The first string of successful novel 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 novel business ideas – specialised on a technical solution or a broader approach for process improvement. In the case of novel entrepreneurship, the results underline former research, according to which 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 is in line with the requirement of novel 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-novel 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 novelty serve a purpose through the transfer of knowledge and skills and use opportunities for arbitrage.

6.2 Practical implications

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. However, a non-targeted support of entrepreneurial activities neglect

the heterogeneity of their impact. Especially initiatives for supporting academic entrepreneurship are characterised by conflicting goals, leading to a non-selective funding practice (Sandström et al., 2018). This implies, that the evaluation of the effectiveness of these programs is hardly possible and the consultation is not tailored to the specific needs of the entrepreneurial endeavours (Rasmussen & Wright, 2015; Sandström et al., 2018).

The distinction of entrepreneurial activities depending on their degree of novelty, enables support programs to aim at either opening new economic development paths or to tackle societal challenges by mobilising resources. This study contributes to the identification of novel and non-novel business ideas and their incorporation probability. 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). Thus a more distinct promotion of novel entrepreneurship activities is enabled and, therefore, serves as a response to former researchers (Hurst & Pugsley, 2011; Shane, 2009).

Further, the results of this study may not only inform public support programs, but also early stage investors, managers of accelerators and (corporate) incubators. First, topic proportions in business idea descriptions can hint at the degree of novelty. This may be used in consulting of potential entrepreneurs and the allocation of resources in programs aiming at the promotion of novel entrepreneurship. Second, to attract novel entrepreneurial endeavours to apply for funding programs, the call for applications may use words appearing often in the desired topics.

6.3 Limitations and future research directions

Finally, there are some limitations of this study paving the way for further research. First, incorporation probability is the output variable the study opts for. 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, especially variables that describe the team member characteristics. The data set applied in this study does not provide full information about all team members and their characteristics. However, for a subset of the solo-entrepreneurs the relations between individual characteristics and degree of novelty and incorporation success of the business ideas are only weak.¹⁸

Second, this study differentiates between novel and non-novel business ideas. However, novel entrepreneurship might not be the only desirable outcome as entrepreneurship is a multifaceted phenomenon, but also entrepreneurial endeavours transferring knowledge and skills or tackling a societal challenge. Therefore, the analysis of other types of entrepreneurial activities is a valuable avenue for further research.

Third, the definition of novelty may be subject to change in the time period 2005 until 2019 due to technological progress. Incorporation of more indicators of novelty 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. Especially, a multi-item measurement of this variable, e.g. aligned with the one proposed by Dahlqvist and Wiklund (2012), would be beneficial to gain more fine-grained information on how the jury arrives at their assessment.

¹⁸ For this robustness check, the phi-coefficient was calculated. The values were between -0.152 and 0.295. The variables considered in this step were gender, being a scientific employee, affiliation with an University, affiliation with an University of Applied Sciences, educational background in economics and business administration, educational background in natural sciences, educational background in engineering, possession of an academic degree and possession of a PhD.

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. Moreover, it has been shown that the perceived feasibility that an incorporation of a business idea depends on the knowledge of the entrepreneurs. Since in this study all business ideas are handed in by entrepreneurs with a background in tertiary education, the evaluation of the feasibility of business ideas is not assumed to be dispersed by this fact (Bergmann, 2017).

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. In this line, the long period covered in this study (2003–2019) could come with the limitation that technological progress changed the relation between topics and incorporation probability. This study provides an averaged picture over the whole time period under investigation. Sampling data from diverse competitions would allow the detection of the most recent trends.

Sixth, the advancement of stemming and lemmatisation algorithms in the German language seems to be valuable for future research. So far, not every word is correctly stemmed and lemmatised in the German language. However, these procedures are necessary because these allow for the unification of most of the terms (Birkholz et al., 2021).

7 Conclusion

To foster innovation and wealth generation as well as finding solution for societal challenges, the stimulation of entrepreneurial activities is of great importance. However, a non-targeted support for different types of entrepreneurial activities prevents these to unfold their full potential. Selection of business ideas for investments and support programs rely on quantitative and qualitative metrics. The qualitative assessment, however, is biased by subjective impressions and experiences of the decision-maker. To advance the objectivity of the assessment of business ideas, the purpose of the present paper is to examine the topic structures of novel and non-novel 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 modelling 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 novel from non-novel business ideas with high and low incorporation probabilities. These findings help to design targeted entrepreneurship support and the selection of promising entrepreneurial projects for early stage investors and managers of accelerators and incubators.

Appendix

See Figs. 7 and 8.

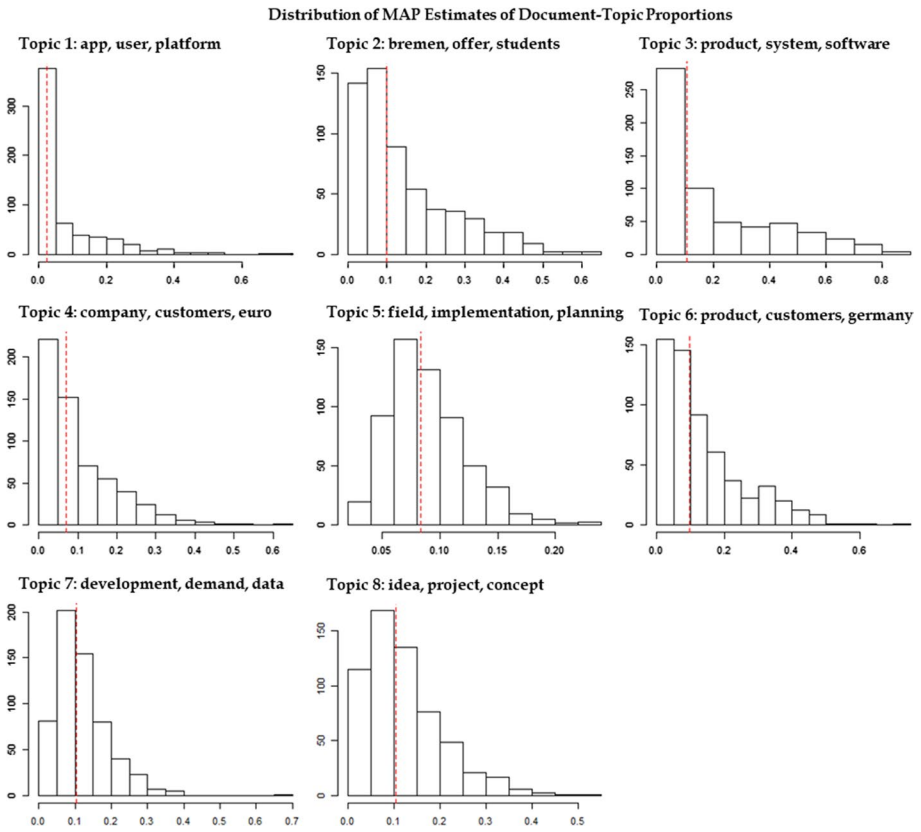


Fig. 7 Distribution of topics over submissions to the ‘CampusIdeen’ competition from 2003 to 2019. *Source* Own calculation

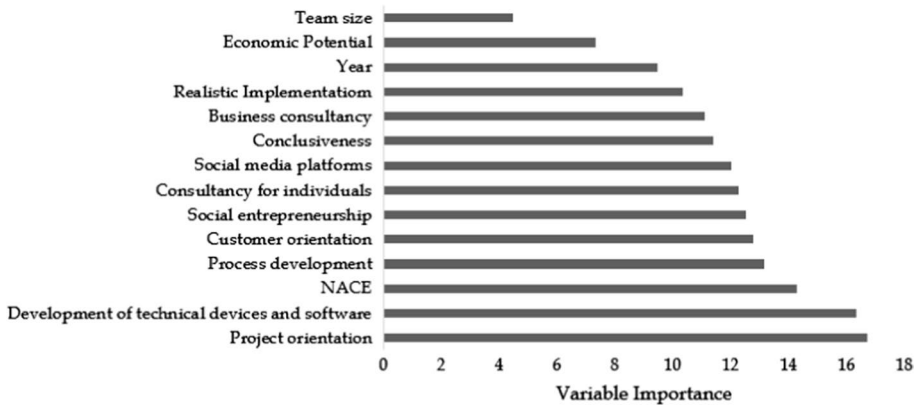


Fig. 8 Variable importance as revealed by the random forest analysis based upon all submitted business ideas. Note: The importance was measured by the mean decrease in Gini-index and the sample size is $n=593$. *Source* Own calculation

See Tables 4, 5 and 6.

Table 4 Similarity and correlation of topics

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
Topic 1								
Topic 2	0.133 (-0.148)							
Topic 3	0.122 (-0.313)	0.104 (-0.586)						
Topic 4	0.089 (-0.098)	0.065 (-0.051)	0.105 (-0.294)					
Topic 5	0.059 (-0.119)	0.078 (0.131)	0.116 (-0.244)	0.050 (0.119)				
Topic 6	0.123 (-0.025)	0.125 (-0.179)	0.420 (-0.228)	0.278 (-0.097)	0.060 (-0.186)			
Topic 7	0.134 (-0.167)	0.094 (-0.157)	0.158 (0.083)	0.051 (0.155)	0.033 (0.414)	0.066 (-0.454)		
Topic 8	0.111 (0.000)	0.047 (0.582)	0.177 (-0.526)	0.088 (-0.253)	0.059 (-0.064)	0.060 (0.028)	0.152 (-0.322)	

Note: The similarity of topics is measured by the cosine index and the correlation is measured by the Pearson coefficient (in parentheses)

Source Own calculation

Table 5 Correlation matrix of included variables into the classification tree analysis for novel business ideas

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Year	Team	Conclusiveness	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	

Note: The sample size is n = 247

Source Own calculation

Table 6 Correlation matrix of included variables into the classification tree analysis for non-novel business ideas

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Year	Team	Conclusiveness	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	

Note: The sample size is n = 182

Source Own calculation

Acknowledgements The author wants to thank the transfer office and BRIDGE at the University of Bremen for the data provision and support. Further, the author wants to express gratitude for helpful comments from participants at the following conferences and seminars: the external research seminar at the Maastricht University, the 18th ISS Conference, the doctoral colloquium at the Friedrich-Schiller University of Jena, and a presentation at the ierp seminar. Moreover, sincere thanks to Dr. Ivan Savin, David Karpa, Dr. Matheus Eduardo Leusin, the editor and three anonymous reviewers.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received for conducting this study.

Data availability The datasets analysed during the current study are not publicly available due confidentiality but are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ambrosino, A., Cedrini, M., Davis, J. B., Fiori, S., Guerzoni, M., & Nuccio, M. (2018). What topic modeling could reveal about the evolution of economics. *Journal of Economic Methodology*, 25(4), 329–348. <https://doi.org/10.1080/1350178X.2018.1529215>
- Arora, S. K., Li, Y., Youtie, J., & Shapira, P. (2020). Measuring dynamic capabilities in new ventures: Exploring strategic change in US green goods manufacturing using website data. *The Journal of Technology Transfer*, 45, 1451–1480. <https://doi.org/10.1007/s10961-019-09751-y>
- 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>
- Autio, E., & Yli-Renko, H. (1998). New, technology-based firms as agents of technological rejuvenation. *Entrepreneurship and Regional Development*, 10(1), 71–92. <https://doi.org/10.1080/08985629800000005>
- Berends, H., van Burg, E., & Garud, R. (2021). Pivoting or persevering with venture ideas: Recalibrating temporal commitments. *Journal of Business Venturing*, 36(4), 106126. <https://doi.org/10.1016/j.jbusvent.2021.106126>
- Bergmann, H. (2017). The formation of opportunity beliefs among university entrepreneurs: An empirical study of research-and non-research-driven venture ideas. *The Journal of Technology Transfer*, 42(1), 116–140. <https://doi.org/10.1007/s10961-015-9458-z>
- Birkholz, J., Günther, J., & Shkolnykova, M. (2021). 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, 1–31. <https://doi.org/10.26092/elib/451>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Braga, I., Monard, M., & Matsubara, E. (2009). Combining unigrams and bigrams in semi-supervised text classification. In *Proceedings of progress in artificial intelligence, 14th Portuguese conference on artificial intelligence (EPIA 2009), Aveiro* (pp. 489–500).
- 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](https://doi.org/10.1016/0883-9026(95)00129-8)
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288–296), BC. <http://umiacs.umd.edu/~jbg/docs/nips2009-rtl.pdf>
- 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>
- Cox, K. C., Lortie, J., & Gramm, K. (2017). The investment paradox: Why attractive new ventures exhibit relatively poor investment potential. *Venture Capital*, 19(3), 163–181. <https://doi.org/10.1080/13691066.2016.1247982>
- Cunliffe, A. L., Luhman, J. T., & Boje, D. M. (2004). Narrative temporality: Implications for organizational research. *Organization Studies*, 25(2), 261–286. <https://doi.org/10.1177/0170840604040038>
- Dahlqvist, J., & Wiklund, J. (2012). Measuring the market newness of new ventures. *Journal of Business Venturing*, 27(2), 185–196. <https://doi.org/10.1016/j.jbusvent.2010.12.001>
- Davidsson, P. (2015). Entrepreneurial opportunities and the entrepreneurship nexus: A re-conceptualization. *Journal of Business Venturing*, 30(5), 674–695. <https://doi.org/10.1016/j.jbusvent.2015.01.002>
- Diáñez-González, J. P., & Camelo-Ordaz, C. (2016). How management team composition affects academic spin-offs' entrepreneurial orientation: The mediating role of conflict. *The Journal of Technology Transfer*, 41(3), 530–557. <https://doi.org/10.1007/s10961-015-9428-5>
- Drnovsek, M., Franic, S., & Slavec, A. (2018). Exploring antecedents of business angels' intention to invest. *Proceedings of Rijeka School of Economics Journal of Economics and Business*, 36(2), 701–734. <https://doi.org/10.18045/zbfri.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. <https://doi.org/10.1023/A:1025045828202>
- 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>
- Hossinger, S., Block, J., Chen, X., & Werner, A. (2021). Venture creation patterns in academic entrepreneurship: The role of founder motivations. *The Journal of Technology Transfer*. <https://doi.org/10.1007/s10961-021-09904-y>
- Hurst, E., & Pugsley, B. W. (2011). What do small businesses do? *Brookings Papers on Economic Activity*, 43(2), 73–118. <https://doi.org/10.3386/w17041>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 103). Springer. <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>
- Kier, A. S., & McMullen, J. S. (2018). Entrepreneurial imaginativeness in new venture ideation. *Academy of Management Journal*, 61(6), 2265–2295. <https://doi.org/10.5465/amj.2017.0395>
- Kirk, G. R., & Nolan, S. B. (2010). Nonprofit mission statement focus and financial performance. *Non-profit Management and Leadership*, 20(4), 473–490. <https://doi.org/10.1002/nml.20006>
- Kirzner, I. M. (1973). *Competition and entrepreneurship*. University of Chicago Press.
- 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>
- Lee, H., & Kang, P. (2018). Identifying core topics in technology and innovation management studies: A topic model approach. *The Journal of Technology Transfer*, 43(5), 1291–1317. <https://doi.org/10.1007/s10961-017-9561-4>
- 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>
- 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. <https://doi.org/10.17535/crorr.2015.0017>
- Minola, T., Vismara, S., & Hahn, D. (2017). Screening model for the support of governmental venture capital. *The Journal of Technology Transfer*, 42(1), 59–77. <https://doi.org/10.1007/s10961-015-9461-4>
- 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>
- O'Connor, E. (2002). Storied business: Typology, intertextuality, and traffic in entrepreneurial narrative. *Journal of Business Communication*, 39(1), 36–54. <https://doi.org/10.1177/002194360203900103>
- 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.
- Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *The Academy of Management Review*, 24(4), 711–724.
- 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>
- Rasmussen, E., & Wright, M. (2015). How can universities facilitate academic spin-offs? An entrepreneurial competency perspective. *The Journal of Technology Transfer*, 40(5), 782–799. <https://doi.org/10.1007/s10961-014-9386-3>
- 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>
- Roure, J. B., & Keeley, R. H. (1990). Predictors of success in new technology based ventures. *Journal of Business Venturing*, 5(4), 201–220. [https://doi.org/10.1016/0883-9026\(90\)90017-N](https://doi.org/10.1016/0883-9026(90)90017-N)
- Schumpeter, J. A. (1934). *The theory of economic development*. Oxford University Press.
- Sandström, C., Wennberg, K., Wallin, M. W., & Zherlygina, Y. (2018). Public policy for academic entrepreneurship initiatives: A review and critical discussion. *The Journal of Technology Transfer*, 43(5), 1232–1256. <https://doi.org/10.1007/s10961-016-9536-x>
- Shane, S. (2009). Why encouraging more people to become entrepreneurs is bad public policy. *Small Business Economics*, 33(2), 141–149. <https://doi.org/10.1007/s11187-009-9215-5>
- Silge, J. (2018). Training, evaluating, and interpreting topic models <https://juliasilge.com/blog/evaluating-stm/> [Accessed on 24.05.2022].
- Song, M., Podoyrnitsyna, 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.
- Storopoli, J. E. (2019). Topic modeling: How and why to use in management research. *Iberoamerican Journal of Strategic Management*, 18(3), 316–338. <https://doi.org/10.5585/ijsm.v18i3.14561>
- 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>
- Strobl, 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>
- Szerb, L., Lafuente, E., Horvath, K., & Pager, B. (2019). The relevance of quantity and quality entrepreneurship for regional performance: The moderating role of the Entrepreneurial ecosystem. *Regional Studies*, 53(9), 1308–1320. <https://doi.org/10.1080/00343404.2018.1510481>

- 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>
- Tyebjee, T. T., & Bruno, A. V. (1984). A model of venture capitalist investment activity. *Management Science*, 30(9), 1051–1066. <https://doi.org/10.1287/mnsc.30.9.1051>
- van Werven, R., Bouwmeester, O., & Cornelissen, J. P. (2019). Pitching a business idea to investors: How new venture founders use micro-level rhetoric to achieve narrative plausibility and resonance. *International Small Business Journal*, 37(3), 193–214. <https://doi.org/10.1177/0266242618818249>
- Vogel, P. (2017). From venture idea to venture opportunity. *Entrepreneurship Theory and Practice*, 41(6), 943–971. <https://doi.org/10.1111/etap.12234>
- 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>
- Wong, P. K., Ho, Y. P., & Autio, E. (2005). Entrepreneurship, innovation and economic growth: Evidence from GEM data. *Small Business Economics*, 24(3), 335–350. <https://doi.org/10.1007/s11187-005-2000-1>
- 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>
- Zacharakis, A. L., & Meyer, G. D. (2000). The potential of actuarial decision models: Can they improve the venture capital investment decision? *Journal of Business Venturing*, 15, 323–346. [https://doi.org/10.1016/S0883-9026\(98\)00016-0](https://doi.org/10.1016/S0883-9026(98)00016-0)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.