Analyzing startup performance: How methods of machine learning and multi-criteria optimization may help to uncover successful startups

Master thesis

of

Andreas Levin

At the Department of Economics of

Business Engineering

Submitted on 03.11.2014

at the Institute for Entrepreneurship, Technology-Management and Innovation

of the Karlsruhe Institute of Technology

Reviewer: Prof. Dr. Orestis Terzidis

Advisor: Dr. André Presse

External advisors: Prof. Dr. Kalyanmoy Deb; Prof. Dr. Sanaz Mostaghim

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht zu haben und die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis beachtet zu haben.

Karlsruhe, den 3. November 2014

Abstract

Scientists and politicians agree that entrepreneurship is a driving force for economic development, determining its speed and direction. To ensure future growth it is critical to understand how entrepreneurial activity can be supported in a way that it creates sustainable value and additional employment. For this reason the study of startup performance is a topic of interest to both scientists and politicians. Recent publications reveal three obstacles in startup performance research. First, the prevailing assumption of a generally positive growth-profit-relationship is empirically not supported. Second, research often uses the terms performance and success synonymously even though they are semantically different. Third, the analysis of simple relations between performance measures and independent variables did not uncover performance rules with consistent effects and high predictive power.

This study tries to assess how research on startup performance may benefit from the application of modern methods of computational science, specifically multi-criteria optimization and machine learning. In order to do so we apply a resource-based view approach, enhanced by dynamic capabilities, to obtain a theoretical framework that allows us to analyze startup performance. Based on this framework we build a coherent analytical model that enables us to test the effectiveness of algorithms. Using artificially generated data we test the problem-solution capabilities that the algorithms non-dominated sorting and random forests, exemplarily representing the areas of multi-criteria optimization and machine learning, have in the context of startup performance analysis.

Our results confirm that the criterion of pareto-optimality, obtained through non-dominated sorting, is a property that allows us to compare various performance measures simultaneously and independent of their interdependencies. Our study further demonstrates that pareto-optimality can be used to define a clear distinction between performance and success, whenever reasonable assumptions about preferences can be made. Moreover our study shows that the random forest algorithm is able to detect multivariate performance rules in RBV-based frameworks and can provide predictions on startup performance.

Acknowledgements

This study has been conducted at the Institute of Entrepreneurship, Technology Management and Innovation of the Karlsruhe Institute of Technology (KIT) and in cooperation with the College of Engineering of the Michigan State University.

I could not have accomplished this thesis without the support and commitment of my supervisors Prof. Dr. Kalyanmoy Deb, Dr. André Presse and Prof. Dr. Sanaz Mostaghim. I pose in deep gratitude towards them for allowing me to work on such an interdisciplinary topic and for openly sharing their experience, knowledge and time with me. Further I owe a special debt of gratitude to Prof. Dr. Kalyanmoy Deb, who gave me the chance to be part of his outstanding research group at MSU, which is full of brilliant, innovative, and inspiring people. Additionally I would like to thank the DAAD for the financial aid that I received through a PROMOS-scholarship. Without this support my stay at MSU wouldn't have been possible.

Finally, I sincerely want to thank my family for always supporting me, both spiritually and financially, throughout my studies. I thank my parents Monika and Martin, who had to send me off to other places more often than they wanted to, and Tobias for being both my brother and my best friend.

Table of contents

List of figures

List of tables

List of abbreviations

RBV	Resource-based view
DC	Dynamic capabilities
GPR	Growth-profitability-relationship
SME	Small- and medium-sized companies
STEM	Science, technology, engineering and mathematics
ZDT 1	Zitzler, Deb, and Thiele
MSE	Mean-squared error

1 Introduction

Entrepreneurship is an important part of the puzzle, that is, economic and social development (Rezaei et al. 2012). It is the driving force of economic evolution, determining its speed and direction (Terjesen & Wang 2012). Facing a slowed world economy, in the search for new growth, entrepreneurship has received an increasing interest from policy makers around the world. Exemplary for this, in 2002 Romano Prodi as President of the European Commission announced that promoting entrepreneurship would be critical to ensure future growth on the old continent (Audretsch 2007). But not only politicians emphasize the importance of entrepreneurship for economic development, so do scientists. In a recent interview Nobel Prize winning economist Ronald Coase stressed that entrepreneurship is key for achieving economic growth, pleasantly noticing that the topic finally has become a global phenomenon (Terjesen & Wang 2012). As we can see, both, politicians and scientists, are interested in understanding how entrepreneurial activity can be supported in order to stimulate growth and create employment (Delmar et al. 2013).

To answer this question more effectively, however, further research is required. Especially startup performance has become a topic of interest in entrepreneurship research, as it addresses the reasoning behind firm failure, assesses levels of firm performance and tries to enable the prediction of future success. While being one of the most studied fields in entrepreneurship, startup performance remains among its least understood (Mckelvie & Wiklund 2010). While there is a large consensus that startup performance is a multidimensional phenomenon, research widely assumed these dimensions to be positively correlated, recent findings however challenge this assumption (Delmar et al. 2013; Davidsson et al. 2009; Wiklund & Shepherd 2005; Murphy et al. 1996). Furthermore, reflecting past studies, scholars realize they haven't been able to identify variables that have a consistent effect on startup performance (Mckelvie & Wiklund 2010; Davidsson et al. 2009). Finally, when analyzing startup success, scholars seem to have paid insufficient attention to the preferential structures necessary to interpret levels of performance as success (Delmar et al. 2013; Mckelvie & Wiklund 2010).

While entrepreneurship research faces mentioned challenges, the analytical capabilities of computational science may bear the potential to resolve them. Computational science is a highly interdisciplinary field of study concerned with building mathematical models and procedures to solve problems in various scientific disciplines (Maxville 2013; Rocha et al. 2010; Wang et al. 2009). Regarding the research gaps in startup performance analysis, specifically the areas of computer-based multi-criteria optimization and machine learning may be able to enhance research, as they are able to process many variables simultaneously, obtaining information about relationships and trade-offs.

The objective of this study is to evaluate the potential that computational methods in the area of multi-criteria optimization and machine learning have regarding the analysis of startup performance. For that purpose we introduce two exemplary algorithms, non-dominated sorting and random forest, and evaluate their benefit for startup performance analysis.

In order to elaborate on this research question, our study is structured as follows. With this introduction constituting the first chapter, in the second chapter we will present a theoretical framework that allows us to analyze startup performance and point to current gaps in research. Additionally, the second chapter will provide an overview on multi-criteria optimization and machine learning, detailing their potential to enhance startup performance analysis. In chapter three we proceed to build a multifactorial model that enables us to perform a quantitative analysis of the startup-performance-relationship. Following this, in chapter three we also introduce two algorithms, non-dominated sorting and random forest, respectively representing the areas of multi-criteria optimization and machine learning. In chapter four we analyze the performance of these algorithms in the context of startup performance analysis, using our multifactorial model and coherent artificial data. In chapter six we discuss the results obtained, while chapter seven constitutes the conclusion of this thesis and provides an outlook on future research.

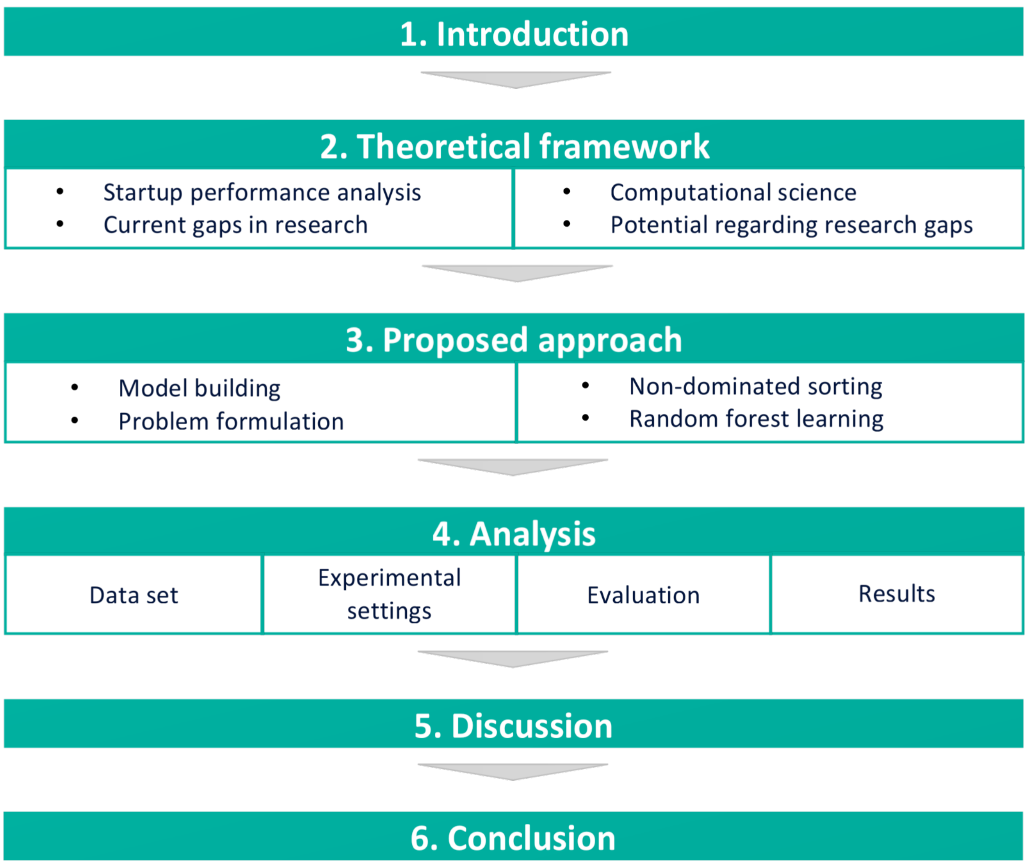

Figure 1 – Structure of this thesis

2 Theoretical framework

The aim of this study is to combine two areas of research. More precisely the idea is to enhance research on startup performance in the entrepreneurial field through the application of methods of multi-criteria optimization and machine learning, which originate from the field of computational science. Following this structure the theoretical framework of this study is divided into two parts. First we intend to establish a common understanding of the entrepreneurial context, explaining how startup performance can be analyzed and pointing to current gaps in research. Then we introduce selected methods of multi-criteria optimization and machine learning, indicating their potential to reduce mentioned research gaps.

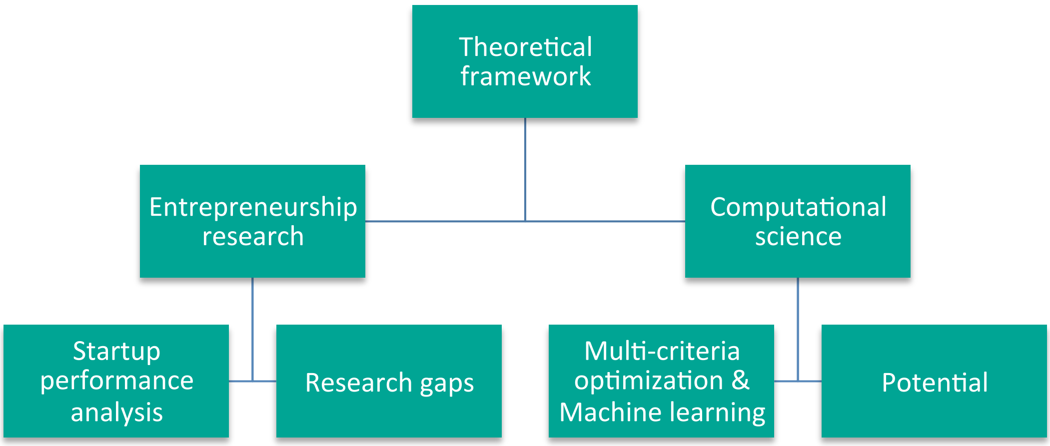

Figure 2 – The theoretical framework of this thesis

2.1 Startup performance analysis and current research gaps

In order to understand how startup performance can be analyzed one must first understand what a startup is, what its performance is and how an explanatory link between these two concepts can be established. For that matter we will proceed to define the startup term, to specify startup performance and to introduce the resource-based view as an explicative theory able to connect a startup to its performance.

Startup: A concept hard to define

As our study aims to analyze the relationship between startups and their performance, we first have to establish a common understanding of what a startup is. Unfortunately defining the startup term is not as straightforward as one would hope. While entrepreneurship is one of the most vital and dynamic fields in management science, it is also characterized by a vast heterogeneity regarding its approaches, methodologies and definitions

(Audretsch 2012; Wiklund et al. 2011). Looking into high ranked entrepreneurial journals the term startup is being used synonymously alongside terms such as new technology based firms, high-tech new ventures, young innovative companies or small and medium-sized companies (Czarnitzki & Delanote 2012; Ganotakis 2012; Gronum et al. 2012; Gimmon & Levie 2010; Kakati 2003). Each one of these terms appears to relate to a specific subset of startups, however, there seems to be a flaw in research as the term startup itself is rarely being defined (Huang et al. 2012; Davila et al. 2003). Keeping Schumpeter's theory of creative destruction in mind it is obvious that there must be not one, but a combination of properties that qualify a company to be a startup. Some of them may be hard to capture, like innovativeness, while others may be easier to determine, like age or size. Nevertheless our study eventually aims to perform a quantitative analysis on startup performance and therefore requires a clear, quantifiable definition for the term. However our work also should not miss out on the bigger picture of what a startup is made of. To cope with these requirements, in order to define the startup term we proceed as follows: First we describe the startup phenomenon on a broader, more qualitative level following the approach of Eric Ries. After having understood the broader implications of the concept we then derive a more quantitative definition based on entrepreneurial research literature.

In his book "The Lean Startup" Eric Ries (2011) provides a holistic, coherent, yet practical description of the nature of a startup: "*A startup is a human institution designed to deliver a new product or service under conditions of extreme uncertainty*". The definition of Ries captures three aspects that are essential to the startup phenomenon: The social aspect, innovation and uncertainty. The social aspect is reflected by mentioning the human as being central to the institution. A startup is more than its ideas and products. It's about the people involved. Larger, well-established competitors may have similar ideas and products, but the social context of their organization is different. Missing hierarchies, quick decision-making and a prevailing entrepreneurial spirit among staff characterize the social context of startups, enabling them to adjust faster to environmental change than their larger competitors (Rosenbusch et al. 2011). Besides the social aspect, Ries definition mentions innovation to be another characteristic for startups. This argument is coherent to Schumpeter's theory of creative destruction. As existing markets are dominated by large companies, startups use innovation to disrupt these markets or even to create entirely new markets or niches. Subsequently innovation allows them to establish competitive advantage or temporary monopolies (Rosenbusch et al. 2011). While innovation is key to being a startup, it is important to understand that it does not necessarily have to occur on product level. Startups may also be innovative because they use new market channels, different payment schemes or appeal to a new group of customers. Finally, Ries definition mentions uncertainty to be a third characteristic for startups. This uncertainty arises because startups, while trying to bring innovation to the market, at the same time typically lack of resources (Stucki 2013; Davila et al. 2003; Lee et al. 2001). While larger organizations dispose of excess resources that enable them to absorb possible failure of innovative ideas, for startups this failure poses an existential risk to their organization. While low means of resources may encourage startups to act more efficiently than larger competitors it also makes them more vulnerable to external influences, threatening their business model (Rosenbusch et al. 2011).

After having understood the characteristics of a startup on a broader, more qualitative level, for the purpose of our study, we now want to seek for a more formal and quantifiable definition. Research literature suggests defining the startup term from an institutional point of view, characterizing it either by its organizational context, its outcome or its behavior. As our study aims to explain different levels of startup performance, we reject a performance-oriented definition of the term, as it would put unnecessary limitations to the scope of our study. Further, as we do not intend to explain performance through behavioral patterns but rather based on resource configurations, following a resource-based view approach, a contextual definition seems most suitable for our purposes. Based on their organizational context startups can be described through institutional features such as age, size, ownership or legal status. Reviewing prevailing entrepreneurship literature Audretsch (2012) finds that for contextual definitions the criteria firm age and size are predominantly used. Both criteria are not only coherent to Schumpeter's view of creative destruction but their combination also enables to exclude companies that fulfill only one criterion, as it would be the case for a company small in size but too old to be a startup (Audretsch 2012).

Within this study we will define the startup term using the institutional criteria size and age. For the size criterion we follow the example of Stucki (2013) and Huang et al. (2013) and limit the company size of startups to companies of equal to, or smaller than medium-size. To ensure compatibility to available databases we define the terms medium-sized company following the standard set by the European Union (see appendix A1). For the age criterion we follow the example of Visintin & Pittino (2014) limiting startups to companies of an age equal to or smaller than 15 years. Keeping the qualitative definition of Ries in mind, we combine Audretsch's findings from his entrepreneurship literature review with practices of recent publications of top ranked journals to obtain our definition of the startup term:

A startup is a novel institution, not older than 15 years, of moderate size, with an headcount lower than 250 employees and an annual turnover lower than 50 million euros a year, that aims to deliver new products or services to a market while incurring high levels of uncertainty.

Performance: A concept that is multidimensional

In order to analyze the relationship between startups and their performance, it's not sufficient to establish a common understanding of what a startup is, but also of what startup performance is and how it can be measured (Murphy et al. 1996; Chandler & Hanks 1993).

Semantically performance is a concept that describes *"the capability of an instance"* (The Oxford Dictionary of English 2010, 3rd ed.), expressing what its able to accomplish. Performance indicators thus describe the level of outcome a startup is able to deliver, providing a feedback to stakeholders on where the venture is heading.

While measuring performance can be simple in some contexts, for startups this does not seem to be the case. Whereas appropriate performance frameworks have been developed for large public companies, only little progress has been made in developing similar schemes for startups (Santos et al. 2002). In fact the heterogeneous use of performance measures in past studies does not only limited their comparability, but is also widely assumed to be the main cause for the conflicting results on startup performance and its driving mechanisms (Delmar et al. 2003; Murphy et al. 1996).

Recognizing these unfavorable circumstances we take them as an impulse to be even more fastidious, defining what startup performance is. So far research has revealed that startup performance is not a single- but a multidimensional phenomenon (Rauch et al. 2009; Wiklund & Shepherd 2003; Murphy et al. 1996; Dess et al. 1997; Chandler & Hanks 1993). Research further identifies three different measures to be relevant for evaluating startup performance: Growth, profitability and survival (Delmar et al. 2013). Moreover research comes to the conclusion that it's inappropriate to observe startup performance based solely on a subset of these three measures. The reason for this is twofold: First of all, each measure is considered to reflect only one dimension of the overall performance and second of all there is no empirical evidence supporting these measures are symmetrical or generally positively correlated (Delmar et al. 2013; Mckelvie & Wiklund 2010).

As startup performance is a concept reflected not by one but by combination of measures, we will continue to introduce each one of them explaining how they can be determined, why they are seen as an indicator of performance and what their respective limitations are.

Growth is a concept used by many studies to assess startup performance. It is typically determined as either change in sales or number of employees over a period of time (Audretsch 2012). Both bases, sales and number of employees, have strengths and weaknesses for reflecting growth. While sales are sensitive to inflation and exchange rates, employment isn't. Employment however is being affected by productivity increases and automatization, while sales aren't (Delmar et al. 2003). Whether observed through employment or sales, growth is generally perceived as a positive indicator of startup performance. High growth indicates that a company is fitting to its environment, meeting market needs and having a competitive advantage (Delmar et al. 2013). It also is a sign that the market has recognized and accepted a company's products, signaling increasing demand (Markman & Gartner 2002). Furthermore growth is a sign for larger size, implying the company operates closer to the minimum efficient scale, being more cost efficient and getting better access to capital markets (Delmar et al. 2013). In addition to that, growth is often associated with a decreasing likelihood of firm failure as well as an increase of wealth creation and a broader economic development (Markman & Gartner 2002). While growth is largely perceived as a positive indicator of performance, there are also arguments supporting that growth can have negative effects on the overall startup performance (Davidsson et al. 2009). High growth might swamp management, disturb its focus and create distortions in the organization. A rapidly increasing number of employees for example may slow down internal knowledge transfer, making an organization less flexible, impairing its original culture and its entrepreneurial spirit (Markman & Gartner 2002). Furthermore growth often times requires investments, which usually have a negative impact on short-term profitability, financial independence and thus on a firms likelihood of survival (Delmar et al. 2013). In addition to that, the use of growth as performance criteria is also limited by the observation that some entrepreneurs rather opt to secure sound levels of profitability than to increase their growth rate.

Profitability itself is another measure to assess startup performance. It is often determined by the return on assets and indicates the financial benefit that an organization provides to its shareholders (Lin & Wu 2014; Davidsson et al. 2009; Steffens et al. 2009). Profitability is generally assumed to be a positive indicator for startup performance. It is interpreted as a sign of a satisfactory market demand, of meeting market needs and of having built a competitive advantage (Davidsson et al. 2009). Furthermore profitability provides a company access to new resources, limiting its dependency on external financing, weakening external control and lowering its likelihood of exiting the market (Delmar et al. 2013). Despite these undoubtedly positive aspects of profitability, there are also reasons limiting its use as an indicator of performance. In industries like the high-tech sector it is common that startups initially can't generate any sales or profits and rather develop in terms of growth in number of employees. In these scenarios profitability is not a suitable criterion to analyze startup performance (Delmar et al. 2003). Even if the environment allows startups to operate profitably, the use of the profitability criteria is limited as we observe that some entrepreneurs rather opt for growth investing their resources to build economies of scale or first mover advantages, deliberately diminishing short or medium term profitability.

Last but not least, the criterion of survival is a third measure to assess startup performance. Survival is considered to be an important measure of performance as mortality rates among young, early stage companies are high (Stucki 2013). Surviving an extended amount of time shows that a startup is suited to its market environment, gathered business experience and developed skills adapting to changes of market forces (Thornhill & Amit 2003). Unlike growth or profitability, survival cannot be interpreted as something negative, however, its expressiveness is also limited. As opposed to growth and profitability, survival it is not a continuous but a binary indicator, which therefore does not allow to identify or compare more than two levels of performance (Delmar et al. 2003). In addition to that, classifying non-survivors can become a rather complicated task as compared to measuring return on assets or sales growth, since not all market exits correspond to firm failure, as companies can also be split, merged or taken over by other entities (Delmar et al. 2013).

Concluding our findings on startup performance we notice that startup performance is a concept hard to address because its nature is multidimensional. We introduced the most commonly accepted performance criteria growth, profitability and survival, analyzing their rational as well as their limitations for explaining overall startup performance. Based on the state of knowledge we recognize that, in order to determine a startups performance in its entirety, the three measures growth, profitability have to be considered simultaneously.

The resource-based view: A theory for explaining startup performance

After having established a common understanding of what a startup is and how its performance can be assessed, for the purpose of our study, we now require a theory that can constitute a link between a startup and its performance. Looking at recent studies in entrepreneurship and reviewing literature on strategic management we learned that the resource-based view (RBV) is a concept suitable to establish this link.

The resource-based view is one of the most widely accepted theories of strategic management and a common framework for explaining differences in performance of companies within the same industry (Lin & Wu 2014; Newbert 2008; Eisenhardt & Martin 2000). The RBV has been successfully applied in numerous entrepreneurial studies examining the growth-profitability relationship of startup companies (Lin & Wu 2014; Visintin & Pittino 2014; Mckelvie & Wiklund 2010; Ganotakis 2012; Davidsson et al. 2009; Newbert 2008).

To comprehend the RBV it is important to understand that it is a theory, which evolves from an internal perspective of the firm. According to the RBV, companies can be described as the set of resources and capabilities at their disposal. While resources are defined as tangible or intangible assets owned and controlled by the firm, capabilities are described as a firm's capacity to deploy its resources in order to implement its strategy.

Typical examples for company resources include a firm's cash reserves or its patent rights. Capabilities on the other hand can be reflected in a company's process of planning and coordination or its culture of continuous improvement. While modeling companies as resource-capability combinations, the RBV simultaneously assumes resources and capabilities to be heterogeneously distributed across companies and to be imperfectly mobile.

When comparing company performances, the RBV attributes superior performance to competitive advantage and identifies superior resource-capability combinations to be the main cause of this advantage. More precisely, according to RBV, a company will attain competitive advantage whenever it possesses a resource-capability combination that is valuable, rare, inimitable and non-substitutable. Competitive advantage then subsequently allows a company to improve its short and long-term performance, as resource-capability combinations are assumed to be imperfectly mobile (Lin & Wu 2014; Davidsson et al. 2009; Newbert 2008; Thornhill & Amit 2003; Lee et al. 2001; Eisenhardt & Martin 2000; Wernerfelt 1984).

For entrepreneurs the RBV implies that a business should only pursue opportunities that match its resource-capability advantage (Davidsson et al. 2009). An advantage that can only be based on resources or capabilities that generate value, are not available to competitors and can neither be imitated nor substituted by them. For entrepreneurial research the RBV implies that differences in startup performance within industries can be explained by observing the companies resource-capability configurations.

Having explained the RBV theory and its implications in detail, we now want to point to some empirical evidence supporting its ability to explain differences in startup performance. Thornhill & Amit (2003) found strong support for the RBV when analyzing Canadian corporate bankruptcies. Studying more than 5000 Swedish and Australian SME's Davidsson et al. (2009) similarly conclude that startup profitability is indicative of having built a resource-based competitive advantage. In addition to that, Lin & Wu (2014) found sound evidence that valuable, rare, inimitable and non-substitutable resources have a significant effect on competitive advantage and firm performance.

Despite this supportive evidence there are also critics to the RBV. Their criticism is usually directed towards the fact that the RBV implements a rather static view on companies and their environment, which is often seen as a harsh contradiction to an economic reality that is characterized by dynamic change (Eisenhardt & Martin 2000). Due to this static point of view critics complain that RBV would only be able to explain competitive advantage and performance at given points in time. As useful as the RBV thus seems to be, mentioned criticism presents a valid argument. Fortunately this argument can be debilitated by enhancing the RBV with the dynamic capabilities theory (DC).

Dynamic capabilities – Extending the RBV to dynamic environments

As previously identified RBV reflects a static view on companies and their environment, thus making it inadequate for explaining competitive advantage and performance in situations of change (Eisenhardt & Martin 2000). To account for this deficiency, scholars often use the RBV in a broader sense, extending its notion by the concept of dynamic capabilities (Lin & Wu 2014; Davidsson et al. 2009; Eisenhardt & Martin 2000). Dynamic capabilities are defined as *"firm processes and routines that use resources to integrate, reconfigure, gain and release resources in order to match market change"* (Eisenhardt & Martin

2000). An example for a dynamic capability would be a company's process for strategic decision-making or its ability to build alliances. When introducing dynamic capabilities to the RBV framework, they immediately become an integral part for explaining firm performance, as they enable companies to manipulate their resource-capability configurations in order to meet the changing requirements of dynamic environments.

Apart from the theoretical reasoning there is also empirical evidence supporting the notion of dynamic capabilities. In a study on 1000 high-tech startups in Taiwan for example Wu (2007) found that DC mediate between entrepreneurial resources and performance, enhancing the overall comprehension of RBV on startups. Moreover, studying 217 companies in China, Li & Liu (2014) confirmed that dynamic capabilities have a significantly positive affect on competitive advantage and company performance.

Summing up the theoretical framework of this study: In order to establish an explanatory theory that links startups and their performance we first introduced the RBV. Recognizing its static view on companies and their environment we further presented the notion of DC. Finally, we decide to combine RBV and DC to obtain a theoretical framework that is able to explain startup performance in both static and dynamic environments. To confirm the legitimacy of this approach we further referred to scholars who successfully applied the same theoretical framework and even highlighted empirical evidence supporting its use.

Research gaps: Impulses for this study

After having established a theoretical framework for analyzing the startup-performance relationship and having build a common understanding of all elementary concepts involved, following the present state of knowledge, we now want to point to current gaps in startup performance research. In order to do so, we revised the top ten entrepreneurship journals according to ISI SCI ranking, as identified by Sassmannshausen (2012), searching for the terms {performance OR success}. As a result of our literature review we identified three major research gaps that we will proceed to explain individually.

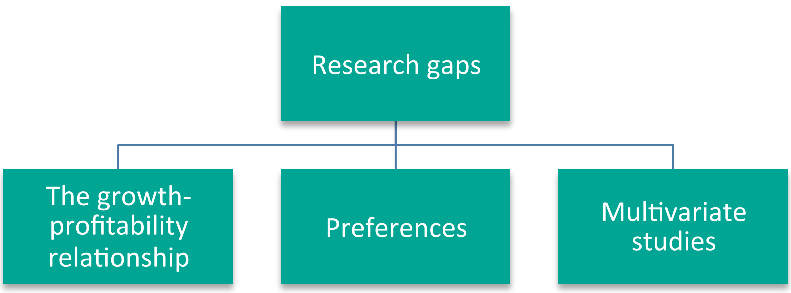

Figure 3 – Current research gaps in startup performance analysis

<u>The growth-profitability relationship – From consensus to conflict</u>

Startup performance, while being one of the most studied fields in entrepreneurship, also remains among its least understood (Mckelvie & Wiklund 2010). Recent publications on the topic indicate that scholars specifically have a growing interest in understanding the relationship between growth rate dynamics and profitability (Federico & Capelleras 2014). This observation is interesting as the assumption of a generally positive relationship between growth and profitability is at heart of many entrepreneurship theories. Recent publications give reason to believe that there might be a gap between this assumption and empirical evidence (Delmar et al. 2013). Growth and profitability may evolve and interact in a more complex and multidimensional way. The assumption of a generally positive relationship between growth and profit is being challenged by recent studies, both logically and empirically (Federico & Capelleras 2014; Delmar et al. 2013; Mckelvie & Wiklund 2010; Davidsson et al. 2009; Steffens et al. 2009).

Logically, growth and profitability are not symmetric (Davidsson et al. 2009). We can underline this statement with a simple example: Few things could be easier than increasing the sales growth of a company. One would only have to buy products at market price and then resell them for a significantly lower price. As a consequence one would expect sales growth to increase significantly. However this growth would neither be profitable nor sustainable and subsequently have a negative impact on a company's profitability as well as on its likelihood of survival. This admittedly extreme example shows that growth and profitability are logically not symmetric. Even though the idea of a positive relation between growth and profitability is based on well-known principals like economies of scale, experience effects, first mover advantage and network externalities, these arguments are being undermined (Davidsson et al. 2009). Recent studies show that in many industries the minimum efficient scale is reached at rather smaller size, making economies of scale a low barrier of entry (Davidsson et al. 2009). Moreover economies of scale seem to be transient as growing firms have an increasing need for coordination affecting its overall efficiency having a negative impact on profitability (Federico & Capelleras 2014). From a competitive point of view high growth, when accompanied by low levels of profitability, can be interpreted as a lack of competitive advantage (Davidsson et al. 2009). In such case, a company has to compete with equally attractive alternatives where growth can only be achieved through costly efforts like marketing or price cuts, both likely to have a negative impact on profitability. Or putting our argument the other way around, assuming a company that poses significant competitive advantage, one would think it would serve its most profitable costumers first (Steffens et al. 2009). If so, subsequent growth could only be achieved by serving less profitable customers, which again would have a negative impact on profitability.

Empirically the evidence on the link between growth and profitability remains mixed. There are reports on correlations measuring from weakly positive, to statistically insignificant, to negative (Delmar et al. 2013; Mckelvie & Wiklund 2010). Delmar et al. (2013) identify a positive relationship between profitability and growth and so did Coad (2010) and Davidsson et al. (2009). Markman and Gartner (2002) however found no relationship between growth and profitability. Brännback et al. (2009) also didn't find evidence supporting that sales growth is positively associated with profitability. In an earlier study, Reid (1995) actually found a negative relation between profitability and growth. More importantly, while analyzing the meta-analysis of Capon et al. (1990), Davidsson et al. (2009) found that positive associations between growth and financial performance were only found in a cross-industry studies, while in studies controlling for industry these effects were in fact very small.

Studying the growth-profitability relationship (GPR) scholars currently have two promising ideas. As growth and profitability only seem to move together in some cases, while having a neutral or even negative relation in other instances, scholars propose to redirect research into looking at growth patterns, rather than trying to find an oversimplified overall relationship (Steffens et al. 2009). By focusing on growth patterns, observing how the GPR relationship may evolve over time, scholars expect to gain better insights into the causal mechanisms behind growth (Mckelvie & Wiklund 2010; Steffens et al. 2009; Delmar et al. 2003). Profits may serve as a prerequisite for a sustained growth trajectory and growth may reinforce a firms' profits (Federico & Capelleras 2014). The second promising idea considering future growth-profitability studies is to make sure that they consistently control their results for industry influence. Davidsson and other scholars found evidence supporting the assumption that the GPR varies among industries (Federico & Capelleras 2014; Delmar et al. 2013; Delmar et al. 2003). Past research did not consistently control for this factor, which may have partially caused the conflicting results observed in the past.

All in all current findings suggest future studies to

1. consider growth and profitability simultaneously, but separately (Delmar et al. 2013; Davidsson et al. 2009; Steffens et al. 2009)
2. adopt a view that explicitly incorporates the intricate relationship between growth and profitability (Delmar et al. 2013; Davidsson et al. 2009; Steffens et al. 2009)
3. consistently control for industry effects (Federico & Capelleras 2014; Delmar et al. 2013; Davidsson et al. 2009)

<u>Preferences – From performance to success</u>

Looking into literature one cannot help but notice that the term success is often times used by scholars as a synonym for the performance measures under observation, see e.g. Visintin & Pittino (2014), Rosenbusch et al. (2011) or Ganotakis (2012). This use of term however seems inconsistent to its actual definition. Oxford dictionary defines success as "*the accomplishment of an aim or purpose*", which implies that success is a concept that includes preferences. Thus, by using performance and success synonymously when investigating startup growth and profitability, previous studies implicitly assumed all entrepreneurs opt for financial performance, weighing growth and profitability as equally important.

Logically this seems to be a harsh assumption as the importance of growth and profitability as indicators of success may differ among founders (Delmar 2008). Some founders may opt primarily for growth trying to realize effects of scale and secure first mover advantage, while others may opt for profitability trying to build up financial independence, perceiving growth to be a risky option possibly leading to firm failure (Delmar et al. 2013). In fact it's most likely that the importance of growth and profitability does not only vary among founders but also may change over the lifetime of an organization (Mckelvie & Wiklund 2010; Stuart & Abetti 1987). To give an example: For an entrepreneur the main goal could be as simple as the survival of his firm. However, once a venture capitalist invests into his company, increasing profitability is going to become a major objective for him (Stuart & Abetti 1987).

Empirical work also does not support the assumption that all entrepreneurs opt for growth and profitability in equal measures. For example there is evidence that some founders don't want their company to grow if this would imply negative effects on the well-being of their employees, on the financial independence of the company or its likelihood of survival (Wiklund et al. 2003). Likewise Delmar et al. (2013) came to the conclusion that it's not correct to assume all entrepreneurs opt for growth. More precisely, looking at growth and profitability, Davidsson et al. (2009) found that both indicators shouldn't be portrayed as equally important for firm performance. Following the same line, Federico & Capelleras (2014) conclude that some firms pursue growth opportunities baring the risk of destroying value, while others may enjoy superior profits and refuse to grow. Finally, all these studies are consistent to Chandler and Hanks (1993) findings that different people may not be equally satisfied with the same level of performance.

Concluding our findings we observe that literature has paid insufficient attention to the preferential structures necessary to interpret levels of performance as success (Mckelvie & Wiklund 2010). In order to determine whether a company is successful or not, its not sufficient to observe its levels of performance, but one must further compare this performance to its original goal setting (Stuart & Abetti 1987). Even if this information is available there still is the problem that different companies have different goals, each of which may change over time, making cross company comparison very difficult (Murphy et al. 1996).

All in all we reach to the conclusion that there are two possible ways to address startup success in future studies:

1. Either by including information about the startups preferences regarding performance measures into the analysis
2. Or by eliminating any assumptions on the importance of performance indicators, examining them jointly but as separate dimensions, leaving the interpretation of success to the observer.

Multivariate studies – From simple to complex relations

Even after hundreds of studies scholars have not been able to identify variables that have a consistent effect on growth and profitability (Mckelvie & Wiklund 2010; Davidsson et al. 2009). One possible reason for this lack of progress could be that researches missed to understand the dynamic relationship between growth and profitability. Another possible reason is that previous studies largely failed to account for differences on company as well as industry level.

Regarding the industry level, taking a look at past studies on growth and profitability Delmar et al. (2013) found out that a possible reason for their conflicting results is that most of them failed to account for differences in the type of industry. An earlier study by Delmar et al. (2003) already indicated that cross-sectional evaluations might be problematic, following the findings of Chandler and Hanks (1993) who concluded that industry level factors have a significant influence on the relationship among performance measures. However, looking at differences at both industry and company level Federico & Capelleras (2014) found that company specific attributes seem to be even more important in explaining differences in performance than industry level factors. Especially a company's access to resources seems to be a key factor in explaining patterns of growth and profitability (Delmar et al. 2013). Looking not only at their access to resources but also at their configurations, Delmar et al. (2003) conclude that firm growth is a very complex phenomenon, where various firm factors are differentially related to different forms of growth. All together these studies suggest the implementation of a RBV as a way to determine more consistent effects among independent variables and performance measures. While designing and executing precise RBV studies is openly admitted to be a challenging task (Delmar et al. 2013), even though a perfect design might be impossible to obtain, multivariate approaches are still considered to allow investigating more complex relations and expected to have hire predictive power (Davidsson et al. 2009; Dess et al. 1997).

All in all these findings suggest future studies to

1. openly acknowledge the complexity of the growth-profitability relationship (Czarnitzki & Delanote 2012; Delmar et al. 2003)
2. apply a resource-based view approach (Davidsson et al. 2009)
3. focus on inter-firm heterogeneity within particular environments (Federico & Capelleras 2014)

2.2 Computational science and its potential

Now that we have introduced a theoretical framework for analyzing startup performance and identified current research gaps in this field of study, we want to advance and explain how computational science might help to reduce these gaps. For this matter we will first introduce relevant concepts within the area of computational science and then point to their potential to solve research gaps in startup performance analysis.

Computational science is a wide and dynamically evolving area of research concerned with building mathematical models and quantitative analysis techniques for solving problems in various scientific disciplines (Maxville 2013). Within the scope of our study we only introduce selected concepts of computational science that are relevant to our approach. In order to do so, we will advance in the following manner: First we will familiarize ourselves with the concept of multi-criteria optimization describing the criteria of pareto-optimality and explaining the difficulty of multi-criteria decision-making. Then we will introduce the concept of machine learning and provide a more detailed description on how ensemble methods work.

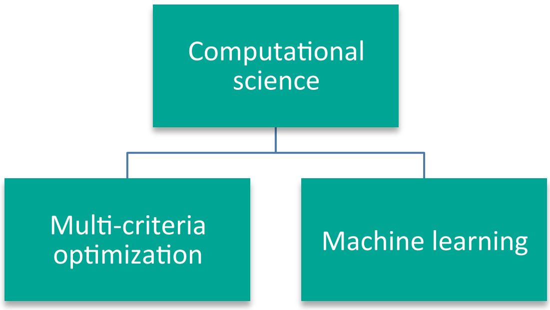

Figure 4 – Multi-criteria optimization and machine learning as areas of interest

Multi-criteria optimization: A way to handle conflicts

Optimization is generally defined as the task of finding one or more feasible solutions to a mathematically formulated problem minimizing or maximizing one or more objectives. Optimization is a topic that is relevant for any agent, involved in one or more decision processes, who is concerned to utilize available resources in an efficient way. In order to optimize a real world scenario there are three sub-tasks to be fulfilled: Model building, application of an optimization procedure and decision-making. The first step, building an appropriate model, is as important as the optimization procedure itself, as an optimal solution must always be seen in the context of the model it originated from (Branke et al. 2008). Following popular annotation an optimization problem can be formulated as:

Definition 1 (Optimization problem)

$$Minimize\ \{f_1(x), f_2(x), \ldots, f_M\}$$

$$Subject\ to\ x \in S$$

Where $f_i(x)$ represent one or more objective functions aimed to be minimized simultaneously. S is the nonempty feasible region defined by a set of constraints modeling the problem environment. A feasible solution can be described by its decision vector x, which again is part of S. Objective vectors are images of decision vectors and consist of objective values $z = f_1(x), f_2(x), \dots, f_M(x)$. Furthermore the image of the feasible region in the objective space is called feasible objective region. At this point it should also be mentioned that restricting problem formulation to minimization problems does not pose a limitation to modeling, as any maximization task can easily be turned into a minimization task by simply negating their objective functions (Ehrgott 2005).

As S is nonempty, solving a single objective optimization problem ($M = 1$) may result in various optimal points but will only result in one global optimal objective value. In the single objective case decision-making therefore is relatively easy, as often times only few decision vectors correspond to the global optimal objective value. In situations of multi-criteria optimization however decision-making becomes more complicated. Assuming a two-dimensional minimization problem and looking at the vectors $f(x) = (1,0)$ and $f(y) = (0,1)$ one has difficulty to decide which of these represents a better solution to the problem. $f(x)$ is better regarding the second objective but $f(y)$ is better regarding the first objective. However, when comparing both solutions to the vector $f(z) = (2,2)$ its obvious $f(x)$ and $f(y)$ dominate the latter by being better in both objectives.

<u>Definition 2 (Domination)</u>

$$let\ the\ points\ x, y \in S\ be\ given\ x\ dominates\ y, expressed\ as\ x \succ y, if\ f_i(x) \leq f_i(y)$$

$$for\ all\ i = 1, \dots, M\ with\ strict\ inequality\ for\ at\ least\ one\ i$$

As our simple example already indicated, in multi-criteria optimization one does not typically find only one optimal objective value but a set of equally optimal objective values, that can be identified with the help of definition 2. Each solution in the optimal set is characterized by the fact that its objective value is not being dominated by any other point in the feasible objective region. As a consequence these solutions are also referred to as non-dominated, efficient, non-inferior or pareto-optimal solutions (Deb 2011; Branke et al. 2008; Deb 2003).

<u>Definition 3 (Pareto optimality)</u>

$$A\ point\ x_p \in S\ is\ called\ pareto-optimal\ if\ no\ x\ \in S\ exists\ so\ that\ f_i(x) \leq f_i(x_p)$$

$$for\ all\ i = 1, \dots, M\ with\ strict\ inequality\ for\ at\ least\ one\ i$$

In multi-criteria optimization decision-making is difficult because the pareto-optimal set contains not only one, but various solutions, where none can be said to be better than any other. However these solutions imply trade-offs among them, where one must sacrifice on one objective in order to get a gain in another objective.

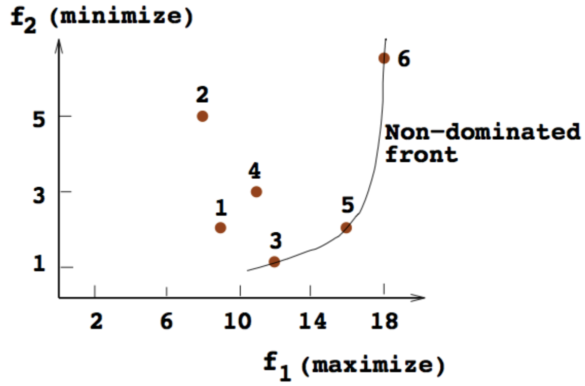

Figure 5 – The non-dominated front reveals trade-offs (Deb 2011)

Taking Figure 5 as an example, we can see that solutions number 3,6 and 5 are all part of the set of optimal solutions, indicated by the non-dominated front. While comparing these three solutions, however, one realizes that solutions number 3 and 6 are situated on steeper parts of the non-dominated front as compared to solution number 5. This means that, compared to solution number 5, the solutions number 3 and 6 had to sacrifice relatively large amounts of one objective in order to realize a relatively small improvement in the other objective. Even though all three solutions are non-dominated, solution number 5 thus looks like a better trade-off between the two objectives. As this example shows, even though decision-making is difficult in multi-criteria optimization, analyzing the trade-offs among objectives in the pareto-optimal set can provide valuable insights to decision makers and ease the decision-making process.

So far, we found out that pareto-optimality is a criterion, which allows us to identify sets of optimal solutions considering multiple objectives simultaneously and that the solution to be selected from this set is determined by the decision maker, who's decision process can be analytically supported by looking at trade-offs within the optimal set (Deb 2003; Branke et al. 2008).

In order to determine the set of optimal solutions, to both single- and multiple objective problems, a variety of optimization procedures exist. The main approaches to be distinguished are deterministic- and stochastic-based methods. While deterministic-based methods, like sequential quadratic programming or the αbb-method, use gradient-based information and follow mathematical principles to determine optimal solutions, stochastic-based methods, like evolutionary algorithms or simulated annealing, are random based, usually don't use gradient-based information and follow some natural or physical principle to determine their solutions. Which type of method and algorithm to use, however, highly depends on the type of problem, the computational power at hand as well as the individual preferences regarding accuracy and speed (Shukla & Deb 2007).

Regarding its application, we can say that multi-criteria optimization is a concept increasingly adopted by other areas of research, specifically because real-world problems usually involve multiple objectives (Wallenius et al. 2008; Zitzler et al. 2004). In engineering science methods of multi-criteria optimization are used to improve aerodynamic designs (Olhofer & Sendhoff 2001), while medical science uses multi-criteria optimization tools to enhance molecular docking processes (Mackerell 2004) and design radiology treatment therapies (Shao & Ehrgott 2008). In management science current applications of multi-criteria optimization are to be found in the area of advanced production planning and scheduling (Dickersbach 2008) as well as land use planning (Janssen et al. 2008). Moreover, the observation that professional management journals increasingly recognize the importance of multi-criteria decision-making for their area of research further justifies the approach of our study (Wallenius et al. 2008).

Machine Learning: A way to perform complex multivariate analysis

Machine learning is an area of research dedicated to the formal study of learning systems. It is a highly interdisciplinary field that combines ideas from statistics, computer science and optimization to build algorithms able to learn from data (Wang et al. 2009; Rocha et al. 2010). It is an area of research experiencing increasing attention, as modern technologies allow us to create, store and provide enormous amounts of data, commonly referred to as big data. Having access to such vast amounts of data, a major problem becomes how to extract relevant information and knowledge from it. Methods of machine learning are one way to accomplish this. Learning specifically describes the process of transforming outside information into knowledge. Machine learning is the automatization of this process through methods of computational science. It is a field of study that develops computational approaches able to recognize complex patterns in observed data and to build coherent models allowing to make predictions on unobserved examples (Wang et al. 2009). Its main task, learning, can be expressed with a help of a target function $f(x)$, where x is the vector describing the input object under observation and $f(x)$ is the objective value or label describing the phenomenon to be learned (Rocha et al. 2010). The task of a machine learning procedure is to approximate $f(x)$ by building a model, or learner, on a set of training examples, which subsequently can be used to make predictions on outcomes of yet unobserved input objects. Depending on the target function $f(x)$, learning problems can be categorized into classification and regression problems. Depending on the type of data available to determine $f(x)$, learning procedures can be distinguished into supervised and unsupervised learning (Zhang & Tsai 2002). Learning is considered to be supervised when labeled training data is used to determine the approximation of $f(x)$, while unsupervised learning approaches approximate this function without the use of label information (Rocha et al. 2010). Examples for unsupervised learning methods are k-means clustering, neural networks and ensembles. Supervised methods include linear discriminant analysis, support vector machines, neural networks, and ensemble methods.

Ensemble methods – Strong by combining weak learners

Ensemble methods are machine learning techniques that can be used for both supervised and unsupervised learning. Ensembles are further able to perform classification and regression problems (Kulkarni & Sinha 2012). Compared to other machine learning techniques, the strength of the ensemble approach is its flexibility. Ensembles can solve classification and regression problems of single-task, multi-task or multi-target learning (Brown & Wyatt 2005).

The main idea of an ensemble is to combine various weak learners to construct one strong learner. Therefore an ensemble uses simple learning algorithms to create a set of individually trained learners and obtains its decision by combining their decisions through a voting scheme (Dietterich 1999). Research shows that the result is often times better than the individuals composing it, achieving higher levels of accuracy (Kulkarni & Sinha 2012; Kotsiantis 2010). The key principle driving ensemble success is diversity (Brown & Wyatt 2005). To achieve higher levels of accuracy base learners have to be diverse and exhibit different patterns of generalization. This is a prerequisite for ensemble success, as combining a million identical learners obviously would not bring any improvement. While being a prerequisite, the claim for diversity is also a conflict to the individual error rate of each learner (Webb & Zheng 2004). The key for building successful ensembles therefore is to master this trade-off and increase learner diversity while maintaining acceptable levels of individual error.

Types of ensembles can be distinguished by their voting scheme, the type of base learner they use and the data subsets they select to train base learners. As research considers the data selection to be the main contributor to ensemble diversity, we will briefly introduce the three methods proven to be most successful: Bagging, boosting and random subspace selection (Banfield et al. 2007; Bernard et al. 2007). The idea of bagging is to train each base learner on a bootstrap replicate (sample with replacement) of the training data set (Kulkarni & Sinha 2012). Boosting algorithms build learners in series adaptively changing the distribution of the training set, based on the accuracy of the previously created classifiers, paying more attention to misclassified instances (Kotsiantis 2010). While bagging and boosting algorithms produce diversity by aiming at the distribution of the data set, random subspace algorithms aim to produce diversity by only using a subset of the training data's features for learner construction and randomizing its selection (Bernard et al. 2007). Research on methods of data selection has shown that boosting is more effective at reducing bias than bagging, while bagging is more effective at reducing variance (Kulkarni & Sinha 2012; Webb & Zheng 2004). Regarding noise, an important factor for real world scenarios, research shows it has a negative impact on boosting's accuracy, while it increases the diversity achieved through bagging (Dietterich 1999). While random subspaces is an approach that can easily be combined with both, bagging or boosting, depending on its implementation, it might not be suitable for data sets that have only a small amount of features to begin with (Dietterich 1999).

Having introduced the main idea of the ensemble approach, we now want to point to some real world applications. So far ensembles have been applied to a variety of problems, where descriptive or predictive data mining was at need. In computer science ensembles have been successfully trained to perform spam or malware detection (Alam & Vuong 2009), in meteorology they are being used for rainfall forecasting (Wu 2009) and biomedicine applies ensembles to provide diagnostic and treatment suggestions to professionals (Gavrishchaka et al. 2010).

The potential: How these methods may enhance startup performance analysis

To make a statement on the potential that methods of multi-criteria optimization and machine learning might bear for the research on startup performance, we first have to review the extent to which they already have been applied in this area of research. In order to do so, we revised the top ten entrepreneurship journals according to ISI SCI ranking, as identified by Sassmannshausen (2012), searching for the terms {machine learning OR data mining} as well as {optimization OR multi-criteria}. Regarding the topic machine learning we were able to identify four significant results. All four of them applied methods of text-mining to patent-related databases in order to study trends in innovation. We found no study that applied machine learning techniques to study aspects of startup performance. Regarding the area of multi-criteria optimization, we identified eight significant results. Five of them used methods of optimization to study the decision-making process between self-employment and employment or the problem of managing time between one's venture and a wage job. The remaining three studies included applications for determining optimal locations, for building optimal knowledge networks as well as for identifying the optimal strength for patent rights. We found no publication that applied multi-criteria optimization to study the relation between performance measures or their relation to other independent variables. Based on this literature review we conclude that methods of machine learning and multi-criteria optimization rarely have been used in entrepreneurship research at all and even less for the study of startup performance. This impression is consistent to Lévesque's (2004) earlier findings that publications in entrepreneurship journals seldom use mathematical approaches for theory development. In the study "*mathematics, theory and entrepreneurship*" Lévesque however demonstrates that mathematical modeling would in fact be an effective tool for developing entrepreneurship theory. Within the areas of mathematics Lévesque especially mentions the potential of optimization for the study of entrepreneurial decision-making processes and its trade-offs.

Our literature review on real-world applications of multi-criteria optimization and machine learning as well as related applications in entrepreneurship research, combined with the findings of Lévesque, give reason to believe that the application of multi-criteria optimization and machine learning may have a significant potential to enhance entrepreneurial performance research.

To present the potential that we identified with respect to the previously mentioned research gaps we will proceed as follows: First, we will argue how multi-criteria optimization may help to reduce the research gap regarding the GPR. Then we will present reasoning how its criteria of pareto-optimality may help to establish a more clear distinction between startup performance and startup success. Finally, we will address how methods of machine learning may enable us to discover multivariate relations in RBV-frameworks and to make predictions on startup performance and success.

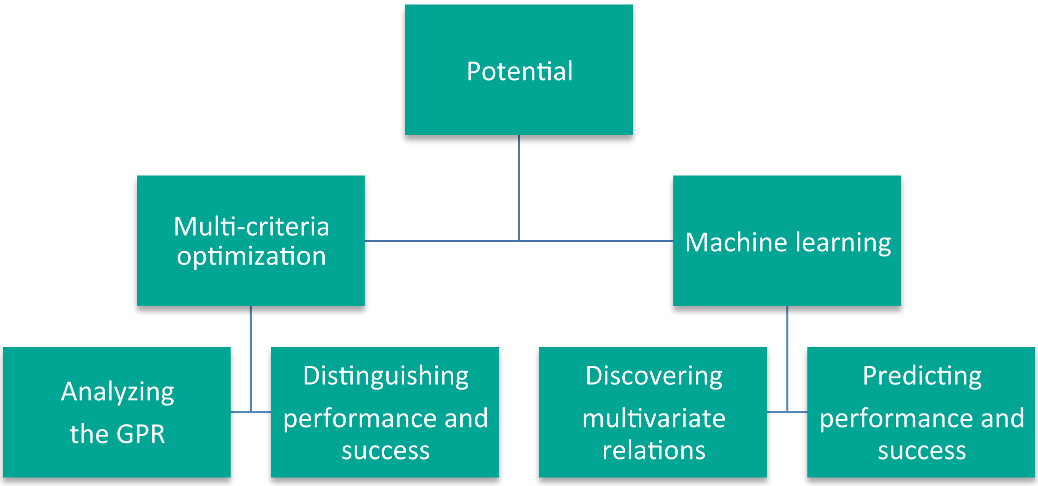

Figure 6 – The potential of multi-criteria optimization and machine learning

<u>The growth-profitability relationship – From consensus to conflict</u>

As previously elaborated, logical reasoning and empirical evidence indicate that the GPR cannot be assumed to be generally positive. There are scenarios in which the objectives growth and profitability relate in a conflicting manner. Since startup performance is a multidimensional phenomenon that includes both growth and profitability, when analyzing it, we have to consider the underlying entrepreneurial decision-making process to be a problem of multi-criteria optimization (Federico & Capelleras 2014; Delmar et al. 2013; Mckelvie & Wiklund 2010; Brännback et al. 2009; Davidsson et al. 2009; Steffens et al. 2009; Markman & Gartner 2002; Reid 1995).

When analyzing startup performance, methods of multi-criteria optimization could help to determine optimal sets in real data. Pareto-optimality is a property that allows to identify optimal companies considering multiple performance measures simultaneously, independent of whether they relate in a conflicting manner or not. The visualization of this pareto-optimal front could further reveal additional insights on the trade-offs among performance measures.

In addition to that, multi-criteria optimization provides a framework that allows us to model relationships between independent variables and conflicting performance measures accordingly. The major problem of this modeling, however, is that the objective functions, putting variables and performance measures into relation, are unknown. This is the point where an integrated approach of multi-criteria optimization and machine learning would reveal additional analytic potential. Based on real data, methods of machine learning could approximate the objective functions of interest and provide them to the corresponding optimization problem. Provided with such approximations, methods of multi-criteria optimization then could determine theoretically optimal solutions to performance-related decision problems.

All in all, multi-criteria optimization could enable startup performance research by allowing to consider various performance measures simultaneously, by enabling to obtain information about trade-offs among performance measures and by offering to determine theoretically optimal solutions regarding performance measures, based on models that include approximated objective functions.

Preferences – From performance to success

As we identified earlier, the term startup success has not been used adequately across studies on startup performance. Success describes the accomplishment of an aim, while performance describes only the accomplishment. The difference between performance and success thus lies in the inclusion of an aim. Hence, in order to distinguish whether a startup is successful or not we require information about its preferences with respect to performance measures. Preference information however is nontrivial and difficult to obtain, as performance is multidimensional, offering a variety of different aims. Even if the specific information was available at some point in time, there still is the difficulty of different companies having different goals and that these may change over time (Federico & Capelleras 2014; Visintin & Pittino 2014; Delmar et al. 2013; Ganotakis 2012; Rosenbusch et al. 2011; Mckelvie & Wiklund 2010; Davidsson et al. 2009; Delmar 2008; Wiklund et al. 2003; Chandler & Hanks 1993; Stuart & Abetti 1987).

When analyzing startup performance, pareto-optimality is a property that could define startup success without requiring complex preference information. For each objective one would only need to know or assume whether all agents generally desire to maximize or minimize this objective. Regarding performance measures this information should not be difficult to obtain, as it seems fairly intuitive that most entrepreneurs will desire to maximize their profitability rather than to minimize it or try to maximize their likelihood of survival rather than to minimize it. Given this type of general preferential information, pareto-optimality is a criterion to define success when comparing sets of startup companies: Provided with a free choice, no rational agent would pick a company outside the pareto-optimal set, as for each of them there is at least one other startup in the set dominating it, hence offering more preferable performance measures. Thus no agent would consider the companies outside the pareto-set to be the successful ones but the ones inside the set.

All in all, given general preferential information about performance measures, pareto-optimality could serve as a definition for success when comparing startup performance.

Multivariate studies – From simple to complex relations

Current literature reveals that startup performance is a complex phenomenon influenced by factors on industry and company level. In order to determine more consistent effects among independent variables and performance measures, scholars suggest the implementation of extensive RBV models. Such frameworks would allow to investigate more complex relations between variables and performance measures, expected to offer a higher predictive power (Federico & Capelleras 2014; Li & Liu 2014; Lin & Wu 2014; Delmar et al. 2013; Davidsson et al. 2009; Delmar et al. 2003; Thornhill & Amit 2003; Dess et al. 1997)

Being able to discover linear and non-linear relationships between many variables, methods of machine learning have the potential to uncover complex, multivariate relations in RBV-based frameworks. As some algorithms can handle both classification and regression problems for even multiple tasks, these could potentially detect relationships regarding multiple qualitative or quantitative performance measures (Brown & Wyatt 2005; Breiman 2001). Especially random forests are able to handle both, qualitative and quantitative input variables simultaneously and thus seem to be well suited to process the data that describes a business and its environment. In addition to that, most machine learning techniques are able to analyze large amounts of input variables at a time, enabling them to coop with extensive RBV models.

All in all, within RBV-frameworks methods of machine learning, when given a set of data, could be able to detect complex, multivariate rules considering multiple performance measures. Moreover, they could build models that approximate the objective functions of respective performance criteria, enabling research to make predictions on the performance of unobserved startups.

Conclusion – How different stakeholders may benefit

Having analyzed the potential of multi-criteria optimization and machine learning to minimize research gaps in startup performance analysis, we now want to point out how this would benefit researchers, entrepreneurs and investors.

For researchers, when comparing startups, the application of pareto-optimality would allow them to determine optimal sets of companies considering various conflicting objectives simultaneously, while offering a clear definition for success. Such optimal sets would further allow researchers to study trade-offs and relationships among performance measures by simply analyzing the shape of the pareto-optimal front. Besides this, multi-criteria optimization could provide them with a framework for building mathematical models that can consider conflicting objectives and link independent variables to performance measures. Based on these models, machine learning techniques could help researchers to identify multivariate relations in real data and provide them with approximations for objective functions, allowing them to make predictions on startup performance. By integrating these approximations into optimization problems researchers could further determine theoretically optimal configurations for startup companies and derive insights on the relations between independent variables and performance measures.

For entrepreneurs, determining optimal sets in real data via the criterion of pareto-optimality would allow them to perform coherent benchmark analysis and thus offer them a better understanding of their positioning with respect to other startups. The analysis of pareto-optimal fronts could further increase the entrepreneurs understanding of the trade-offs among performance measures. In addition to that, depending on the quality of the approximated objective functions, multi-criteria optimization could also allow to formulate and solve performance related decision-making problems, which might ultimately help to improve real life entrepreneurial decisions.

For investors, the application of multi-criteria optimization and machine learning methods could allow them to identify sets of successful startups within large data, considering multiple performance measures simultaneously. In addition to that, the models obtained through machine learning could enable investors to evaluate startups based on predicted future performance, potentially increasing their control of investment risks.

3 Proposed approach

In the theoretical framework of this study we identified three gaps in startup performance analysis. First, the GPR does not seem to be generally positive as often assumed. Second, scholars used the terms performance and success synonymously, implying all entrepreneurs have equal preferences towards performance measures. Third, previous studies only studied few variables at a time looking for simple relations, while complex rules in multivariate studies are expected to have higher predictive power.

Our proposed approach aims to work on these issues. First, we want to demonstrate how the concept of pareto-optimality allows us, while comparing startups, to determine optimal sets of companies considering growth and profitability jointly but separate, adopting a view that explicitly incorporates their intricate relationship. Based on this we also want to show how pareto-fronts can visualize this intricate relationship. Second, we want to suggest the use of pareto-optimality as a definition for separating successful from unsuccessful companies when comparing startups based on various performance measures. Third, we want to conduct a proof of concept on whether or not the machine learning technique random forest is able to recognize complex rules in multivariate, RBV-based studies and whether or not it can predict future startup performance as well as pareto-optimality or success. In order to do so, our approach develops as follows. First, we use the current state of knowledge to build a multivariate, RBV-based model that allows us to analyze startup performance in an appropriate manner. Then we present the respective algorithms that we want to apply, before we transition to the analytical part of this thesis.

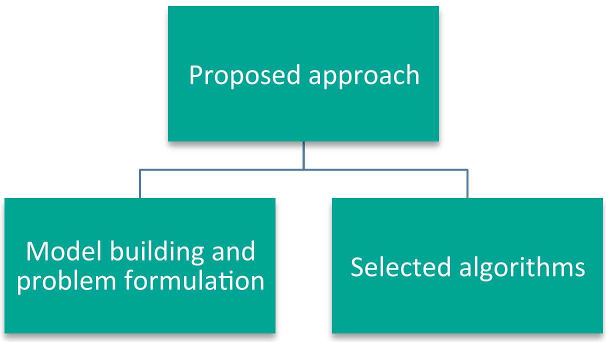

Figure 7 – Structure of the proposed approach

3.1 Model building and problem formulation

Building an appropriate model to analyze startup performance is a key task for fulfilling our research goal. To do so we proceed as follows: First we clarify the objectives for our model, then we introduce the methodology that we apply to develop this model and then we give a detailed explanation on our final result, mentioning its restrictions and limitations.

Our objective was to build an RBV-based model, found on the current state of knowledge, that observes a variety of factors relevant for explaining startup performance. Furthermore the model should allow us to analyze interfirm-heterogeneity within particular environments looking at the objectives growth and profitability simultaneously but separate. In addition to that, the model should be compatible to a standardized questionnaire, developed by our institute, to ensure its applicability to real data. Furthermore our aim was to build a model that we could formulate in mathematical terms so that it may be used as a framework for analyzing optimization problems regarding startup performance.

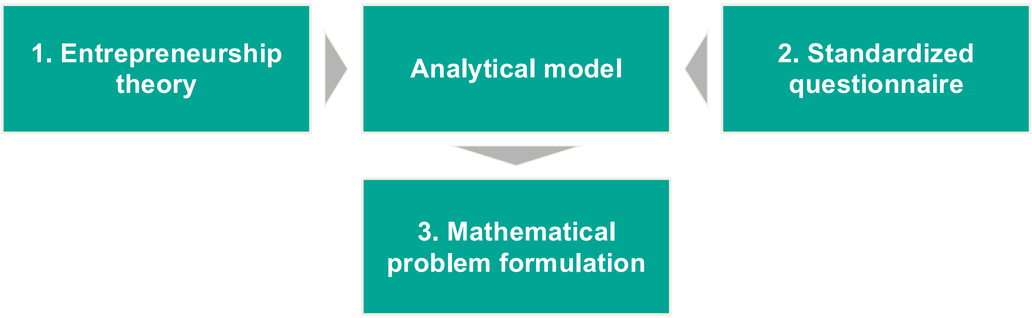

Figure 8 – How we derived the analytical model and its mathematical formulation

To develop our model we advanced in three steps. First, we performed a literature review to identify the factors one ideally would have to observe in order to describe a startup and its environment, when examining its performance. For that purpose we analyzed the top ten entrepreneurship journals, according to ISI SCI ranking, searching for the terms {performance OR success OR growth OR factors OR resource-based view}. Second, we matched the resulting set of factors with the items contained in our standardized questionnaire to obtain a subset of factors, for which we are able to gather real world data. Third, we used standard notation to formulate the resulting model in mathematical terms.

While Figure 9 and Figure 10 provide an overview on our model, appendix A2 displays it in all detail.

	Success factors – Independent variables						
Factor	Human capital	Technological capabilities	Financial resources	Strategy	Networks	Comp. environment	Control variables
Variables	$x_1, \dots, x_8$	x_9, x_{10}, x_{11}	x_{12}, x_{13}	x_{14}	x_{15}	x_{16}, x_{17}	x_{18}, x_{19}, x_{20}

Figure 9 – The analytical model: Success factors

	Performance measures – Dependent variables		
Factor	Profitability	Growth	Survival
Variables	y_1	y_2	y_3

Figure 10 - The analytical model: Performance measures

As you can see our model covers seven different types of independent variables, or success factors, including human capital, the startups degree of innovation, its financial resources, its networks, its competitive strategy as well as its market conditions alongside the control variables size, age and industry. With the help of our questionnaire we are able to measure these factors through a total of 20 variables (see appendix A2). To provide a better understanding of our model we will proceed to introduce each of its factors, justify its use and explain how we are able to measure it through our questionnaire items. To complete our comprehension of the model we also provide its mathematical formulation and finally look into its restrictions and limitations.

Independent variables: Describing a startup and its environment

RBV and DC theory imply that differences in firm performance within industry can be explained by observing a company's resource-capability configurations. Our model is build on this theory. Its independent variables aim to capture the resources and capabilities of a startup, while also observing characteristics of its environment, in order to explain its performance. To better understand the variables it includes we will introduce each one of them, describing its influence by pointing to empirical evidence and explaining how we measure it using our questionnaire items.

<u>Human Capital – Founder's characteristics are relevant for startup performance</u>

Many entrepreneurship scholars suggest that the human capital of a founding team is a critical factor for determining startup performance (Visintin & Pittino 2014; Gimmon & Levie 2010). This view is consistent to the RBV, which classifies human capital as a valuable resource and possible origin for competitive advantage (Ganotakis 2012). Thus, human capital has become one of the most common factors used to predict startup performance (Mckelvie & Wiklund 2010). However it is important to understand that the term human capital itself refers to a broad range of personal aspects attributed to an entrepreneur or a founding team. These aspects are most commonly grouped into educational characteristics and working experience, with working experience being specified in terms of technical skills, managerial skills and industry experience (Coad et al. 2014; Ganotakis 2012; Edelman et al. 2005). Higher levels of education are believed to increase a person's communication skills and learning abilities, helping them to better recognize and exploit opportunities in their environment (Stucki 2013; Ganotakis 2012). Technical and managerial skills are supposed to increase the ability to establish routines and handle employees, allowing people to build better business processes and achieve superior levels of performance (Coad et al. 2014; Edelman et al. 2005). Industry experience is seen as an indicator for existing relationships to suppliers and customers, thus being a signal for better market knowledge and a positive influence on performance (Ganotakis 2012). All in all founding teams with greater human capital are expected to have better judgment in their business environment (Cassar 2014). Since the founding teams judgment has a direct effect on the startups behavior, human capital is believed to have a significant influence on its performance (Visintin & Pittino 2014). As human capital consists of a wide variety of skills, scholars further suggest that it is more likely to encounter high levels human capital within a group of people than in a single person. Hence, they expect firms founded by a group of people to enjoy advantages over firms founded and lead by a single person (Ganotakis 2012). Besides these positive aspects scholars also suspect there is a negative side to human capital. Cognitive theory indicates that very high levels of education and experience may diminish performance as agents might perceive their knowledge to be sufficient to succeed, ignoring relevant outside information (Ganotakis 2012). Moreover, previous experiences might be misleading as knowledge transfer is restricted as each business is somewhat unique (Cassar 2014).

Empirical evidence supports the hypothesis of human capital having an influence on startup performance. In his study Lee et al. (2001) found a significant association between technical experience and performance. While Choi (2004) was able to observe that entrepreneurs with managerial abilities are more likely to identify opportunities and successfully introduce products to markets, Cassar (2014) showed that entrepreneurs possess an advantage evaluating business opportunities in industries where they have vocational experience. In addition to that, Colombo and Grilli (2010) were able to identify that entrepreneurial teams, compared to single founders, are more likely to possess the capabilities required to achieve a fit between technology and strategy. Following this path of study Visintin and Pittino (2014) showed that a balanced demographic structure is a relevant aspect for building well-performing entrepreneurial teams. Besides this evidence Ganotakis (2012) was able to show the negative side of human capital by identifying a significant U-shaped relationship between general experience of founders and new technology based ventures.

To measure the human capital factor in our model it embraces eight different variables. Four of those determine the number of members in the founding team, how many of them are holding a university degree and how many of those degrees were in the field of STEM or social science. The remaining four variables measure the teams working experience, managerial experience, startup experience and experience in research and development using a five-level Likert items.

Financial resources – Their availability is a prerequisite for startup performance

The financial resources of a startup are a factor that is often considered when analyzing startup performance (Mckelvie & Wiklund 2010). RBV however does not consider financial resources to be valuable, rare, inimitable or non-substitutable, implying they have a lesser influence on firm performance (Lin & Wu 2014). As contrary as it may seem entrepreneurship scholars are not opposed to this view. Their rational for observing financial resources is not that they expect financial resources to enhance performance but that its presence is a prerequisite for positive levels of performance (Stucki 2013). Many startups run short of financial resources during early years because they generate only limited cash flows and amount of seed capital is often limited (Lee et al. 2001). Therefore many startups require external capital sources like bank financing or venture capital in order to survive. Unfortunately their access to this capital is limited and even if they attain access they have to pay significant premiums to obtain it, as they have no business history and thus are considered to be extremely risky investments (Stucki 2013; Lee et al. 2001). For entrepreneurship scholars observing equity investments seems a relevant factor for explaining startup performance as it does not only provide financial aid, but also may impose an external institution of control or induce management know-how or even engender legitimacy to other stakeholders (Davila et al. 2003; Lee et al. 2001).

Empirical evidence supports the hypothesis of financial resources being relevant for explaining startup performance. Using data on Swiss startups Stucki (2013) found that financial constraints have a persistently negative effect on firm survival. In a previous study Hvide and Møen (2010) found financial constraints to have a negative effect on new firm success. The findings of Davila et al. (2003) underline this, suggesting that startups may delay their growth due to lack of financial resources. Furthermore he concludes that venture capital funding is a relevant factor when trying to explain differences across startup companies.

To measure startups financial resources our model includes two variables. The first one assesses the total amount of equity capital raised by a startup since its foundation. The second variable is a binary indicator displaying whether a startup has received any investment from a venture capitalist or not.

Networks – A resource that influences performance

Networks are an aspect often considered to be a key element of entrepreneurship and a factor often examined with respect to startup performance (Hoskisson et al. 2011). This habit is supported by the RBV as it considers networks to be valuable, rare, inimitable and non-substitutable resources. The reasoning behind this is that the creation and maintenance of networks is seen as a mechanism for accessing new knowledge, which can potentially lead to competitive advantage (Lin & Wu 2014; Gronum et al. 2012; Eisenhardt & Martin 2000). As most organizations, startups often cover only a small part of the value chain and therefore critically depend on the interaction with other companies. Thus an extended and well-functioning business network can help them to lower transaction costs, obtain strategic assets or facilitate organizational innovation (Huang et al. 2012; Lee et al. 2001). A business network thereby consists of two types of links: Cooperative, bilateral relationships and supporting unilateral relationships. Cooperative bilateral relationships typically exist between startups and a venture capitalists or other enterprises or a universities, while supporting, unilateral relations are mostly formed between startups and governmental agencies (Lee et al. 2001).

Empirical evidence supports the hypothesis of networks having a significant influence on startup performance. In a study on dynamic capabilities Døving et al. (2008) suggest that networks provide SMEs with access to resources, complementary skills, capabilities, and knowledge that are internally not available. A related study of McFadyen et al. (2009) also comes to the conclusion that networks benefit a firm's knowledge creation. In addition to that, Hansen (1995) found a positive association between entrepreneurial networks and organizational growth. Coherent to that Lee et al. (2001) found that networks have a positive influence on organizational performance in terms of sales growth. However Gronum et al. (2012) suggest that the connection between networks and firm performance might be more complex and is being mediated by innovation.

To measure a startups network our model includes one variable. It is a binary variable assessing whether a startup has a university cooperation or not. Besides this, one might argue that the variable determining the presence of venture capital mentioned earlier also conveys a type of network information.

<u>Innovation – A dynamic capability that creates competitive advantage</u>

Innovation is a relevant factor when observing startup performance because it is a phenomenon closely connected to the very essence of entrepreneurship. Some scholars even argue that its innovation, which distinguishes entrepreneurship research from other fields of studies like business administration. In fact within the entrepreneurial community the assumption that entrepreneurs have to be innovative in order to be successful and compete against bigger rivals is widely spread (Rosenbusch et al. 2011). The assumption is rooted on Schumpeters (1934) reasoning that innovation reflects an opportunity for entrepreneurs to establish temporary monopolies, thus allowing them to obtain profits. This argument is consistent to the RBV, which considers innovation to be a dynamic capability that can lead to competitive advantage and performance (Eisenhardt & Martin 2000; Teece et al. 1997). From a company's perspective innovation can be perceived as either an output or a process. Innovation as an output is of course the result of the innovation process implemented by a firm. Innovation as an output can appear in the form of new patents, new technological findings or production skills, thereby representing valuable and inimitable resources. Innovation as a process however describes the way in which people interact within an organization in order to form their innovative ideas into outcomes, hence representing a form of dynamic capabilities (Gronum et al. 2012; Lee et al. 2001). While having many positive effects, we also have to mention that innovation is largely assumed to require substantial amounts of resources. Looking from that angle, innovation may also imply increased uncertainty or even existential risks to startups that are typically resource-scarce (Rosenbusch et al. 2011).

Empirical evidence supports the hypothesis that innovation has a significant influence on startup performance. Studying manufacturing and service firms in Australia, Prajogo (2006) found a positive relationship between innovation and business performance, in terms of sales growth, profitability and market share. Analyzing US business service firms Mansury and Love (2008) also found that innovation has a positive effect on both sales and employment growth. However the relationship does not seem to be that simple. In a meta-analysis on 42 empirical studies Rosenbusch et al. (2011) found that while innovation has a positive effect on performance the innovation–performance relationship seems to be context dependent. Factors like firm age, industry, and type of innovation seem to affect the impact of innovation on firm performance. Investigating the context specificness of the innovation-performance relationship Thornhill (2006) supports this hypothesis by revealing that knowledge, industry dynamism and innovation interact in the way they influence sales growth

To measure a startups innovation our model includes three different variables. The first one determines the total number of own new patents a startup uses or has used. The second assesses this number with respect to purchased patents and the third variable is concerned with the total number of license a startup uses or has used.

Strategy – Achieving a fit between resources and environment is critical

The strategy of a firm is a factor often concerned when trying to explain startup performance (Mckelvie & Wiklund 2010). The strategy reflects the dynamic capabilities of a company, indicating its ability to manage its resources in order to meet the requirements of its environment. The strategy is considered to be an important mediator for connecting company resources and success. While resources enable a firm to implement strategies that lead to performance, it finally is the strategy that defines in which way a company utilizes these resources to relate to its environment. For that matter the quality of a firm's strategy can never be judged independently of its base of resources. Previous studies indicate that a company strategy provides a generative mechanism, which transforms resources into performance. This mechanism is very important for startups, as they typically can't keep bigger companies from entering their market. For them it is less likely to achieve a competitive advantage based on resources alone, but rather on a proper combination of these resources and their strategy (Edelman et al. 2005).

Empirical evidence supports the opinion that a startups strategy has a significant influence on its performance. In their study on 192 small firms Edelman et al. (2005) conclude that neither resources nor strategies alone explain performance but the internal co-alignment of both factor does. Specifically the co-alignment of human capital and strategy had a significant effect on performance. In an earlier study on 123 new ventures McDougall et al. (1994) found a direct relationship between strategy and firm performance that depends on the industry a firm operates in. Furthermore, the study of Delmar et al. (2013) suggests that the relationship between the performance measure growth and survival is also mediated by a firms strategic orientation.

To measure the factor strategy our model includes one variable. This variable assesses whether a startups strategy is best described as either price, quality or innovation leadership.

<u>Competitive environment – Pressure relates to performance</u>

The competitive environment is a relevant factor when analyzing startup performance, as there is substantial agreement that businesses are inseparable from their environment (Dodge 1994). Changes in the environment cause issues that require management decisions and thus possibly affect an organizations ability to survive and grow. A company's environment can be described through its dynamism, hostility and competitive rivalry. While dynamism refers to the continuity of change in an environment, hostility indicates the extent to which an environment is seen as unfavorable to a company's goals and mission. Competitive rivalry however describes the number of competitors on the market and the nature of their competitive dynamics. Competitive rivalry can cause both dynamism and hostility, however, its effect on a company's behavior may vary (Zahra 1993). Some scholars assume that intense competition drives the emergence of new ideas and leads to innovation. When rivalry is tough, companies may feel an increased need to innovate their products and processes in order to differentiate them again from competition (Plummer & Acs 2014). On the other hand rising competition may also increase environmental hostility forcing companies to conserve their resources, which omits to exploit new opportunities (Plummer & Acs 2014; Zahra 1993). With respect to performance, competition may have harmful effects on startup survival, in particular when a large number of relatively small competitors create high levels of rivalry (Pe'er & Keil 2013). This assumption can be underlined by the fact that small companies generally lack sufficient resources to buffer environmental impacts (Lee et al. 2001). Likewise, when facing intense competition, startups may find it more difficult to build close relationships with purchasers and customers or may have to increase salaries as employees find it easier to change jobs frequently (Pe'er & Keil 2013).

Empirical evidence supports the hypothesis that the competitive environment has a significant influence on startup performance. Using data from 102 companies Zahra (1993) examined the association between a firm's external environment, corporate entrepreneurship and financial performance and found that environmental characteristics have a significant influence on entrepreneurial activities. The work further revealed the importance for entrepreneurs to fit their actions to environmental settings in order to be financially successful. In a subsequent study Zahra and Covin (1995) were able to show that hostility has a moderating effect on the corporate entrepreneurship-performance relationship. Based on 645 small businesses Dodge (1994) empirically revealed that competition is the dominant dimension producing change in the relative importance of business problems. Moreover, focusing on the issue of localized competition, Plummer and Acs (2014) found competition to negatively moderate the relationship between new knowledge and entrepreneurial activity.

To measure the competitive environment our model includes two variables. The first one indicates the current number of competitors in a startups environment and the second one assesses the development of this number over time, indicating whether competition is increasing, decreasing or remaining without change.

Control variables – Ensuring reasonable performance analytics

Within any area of research the use of control variables is necessary to ensure the comparability of results among studies. This scientific guideline however has not been consistently addressed by previous studies on startup performance. Some scholars even identify this flaw to be the main cause for the conflicting results on startup performance that have been observed in the past (Delmar et al. 2003). Revising previous studies and the RBV we identify the factors age, industry and size to be the most commonly used control variables in startup performance analysis (Delmar et al. 2013; Gronum et al. 2012; Mckelvie & Wiklund 2010; Edelman et al. 2005). Firm size and age have shown to affect both firm growth and survival. The reasoning behind this is that older companies and companies of bigger size potentially have stronger market positions, better access to resources and more developed management routines than their smaller or younger counterparts (Delmar et al. 2013). With respect to industry context studies indicate that growth processes are not symmetrical across industries. Apparently each industry has its own competitive environment and logic, thus offering different levels of benignity for new firm growth and profitability (Thornhill & Amit 2003).

To account for these control variables our model includes three different items. The first one determines a startups age in years, the second one assesses the industry a startup belongs to and the third one determines its size based on eight broad categories of annual turnover within the last fiscal year.

Dependent variables: Measuring startup performance

As already explained in the theoretical framework of this study there are three different measures to assess startup performance: Growth, profitability and survival. Previous studies have shown that each of these measures only represents one dimension of the overall performance. Hence, all three measures have to be considered in order to obtain a holistic view on a startups performance.

Coherently our model includes three measures to assess startup performance. The first one determines the return-on-assets within the last fiscal year, the second one measures percentage growth in number of employees within the last fiscal year and the third one assesses whether the company is still existent or not.

Mathematical formulation: A formal specification of our model

After explaining our model in detail we want to complete its presentation by using it as a framework to deliver a formulation for the entrepreneurial decision problem regarding performance.

Definition 4 (The entrepreneurial decision problem regarding performance)

$Minimize\ \{-G(x), -P(x), -S(x)\}$

$Subject\ to$:

(1) $x = \{x_1, x_2, \dots, x_{20}\}$

(2) $x_2 \leq x_1$

(3) $x_3 \leq x_2$

(4) $x_4 \leq x_2$

(5) $x_{20} \in \{1,2,\dots,17\}$

(6) $x_{13}, x_{15} \in \{0,1\}$

(7) $x_{14}, x_{17} \in \{0,1,2\}$

(8) $x_5, x_6, x_7, x_8 \in \{0,1,2,3,4\}$

(9) $x_1, x_2, x_3, x_9, x_{10}, x_{11}, x_{12}, x_{16}, x_{19}, x_{20} \in \mathbb{N}$

Definition 4 describes the multi-criteria optimization problem that entrepreneurs face regarding the objectives growth $G(x)$, profitability $P(x)$ and survival $S(x)$. Each of these functions defines a relationship between a startup x and its respective performance measure. As constraint number (1) indicates x is a vector that consists of the 20 variables that our models uses to describe startups and their environment, implementing RBV and DC. In addition to that, constraints number (5) to (9) then define domains for each of these variables, which originate from its measuring item in the underlying questionnaire. The remaining constraints number (2), (3) and (4) represent logical conditions to our input data. Constraint number (2) states that the number of employees with a university degree in a company cannot be higher than its total number of employees. Following the same idea constraints number (3) and (4) ensure that the number of employees holding a specific type of degree cannot be higher than the total number degree holders. If an input object fulfills constraints (1) to (9) it's considered to be a valid description of a startup. The resulting decision problem for the entrepreneur is how to configure his startup x in order to achieve optimal levels of performance, when trying to maximize profitability, growth and survival. Unfortunately the objective functions to this problem are not determined. However, by evaluating many combinations of startups x and their performance $G(x), P(x), S(x)$, methods of machine learning may be able to approximate these functions.

Boundaries: Restrictions and limitations to our model

Now that we have presented our model we also have to mention its restrictions and limitations. First of all, our model is based on a combination of RBV and DC theory. As any theory can only considered to be valid until falsified, the validity of our model is limited to the validity of its underlying theories. Even if we assume these theories to be valid we can identify two additional issues that limit the legitimacy of our model: The number of factors considered within the model and the items per factor to determine its specification. Even though, according to current studies, our model covers the five most relevant factors for describing a startup, looking from a theoretical point of view, it could include more than just these five. In a study on success criteria of high-tech new ventures for example Kakati (2003) suggests to include product characteristics as an additional factor to describe startups in an RBV. Chorev and Anderson (2006) on the other hand propose to include organizational aspects as well as the overall economic situation when investigating upon startup performance. As we can conclude from this, our model is limited by the amount of factors it considers. Furthermore the number of items it uses to determine the specification of each factor poses another limitation to our model. One could easily define additional items and possibly obtain a more precise measurement of the factors concerned. However defining an applicable model is a trade-off between its specificity and the availability of compatible real world data. As its not a prerequisite for our study to build a perfect theoretical model but to design a model that is applicable to available real world data, we preferred building a simpler model for which we can ensure data availability, than building a complex model and gradually adjusting it in the hope of making it applicable.

With respect to our study, our main goal is to deliver a proof of concept whether selected methods of computational science have the potential to improve startup performance analysis. Furthermore we want to carry out this proof of concept in a framework that enables a direct application of its methods to real world data. Regarding this purpose we consider our model to be well-suited as it was developed on a broad theoretical foundation, includes the most important factors for analyzing startup performance as identified by recent studies and is connected to a real data collection initiative that ensures its applicability.

3.2 Selected Algorithms

Combining RBV and DC into the theoretical backbone of our study we have successively developed a model that allows us to analyze startup performance within particular environments. Having obtained this model we now may use it in order to analyze how methods of multi-criteria optimization and machine learning may enhance startup performance analysis. Before we can proceed to the proof of concept, however, we first have to introduce the algorithms of multi-criteria optimization and machine learning that we plan to apply. We intend to apply two different algorithms. First, we want to use a non-dominated sorting algorithm to analyze the benefit, which the concept of pareto-optimality may provide to the analysis of startup performance. In a second step we intend to apply a random forest algorithm in order to study whether this method of machine learning is able to detect multivariate rules within our framework and make quality predictions on startup performance. After introducing each of these algorithms we will proceed to perform the analytical part of this study.

Non-dominated sorting: An algorithm to determine pareto-optimality

As we mentioned earlier, multi-objective optimization problems usually don't have a single optimal solution, but a set of various optimal solutions. We further explained how the criteria of pareto-optimality allows us to identify this set, which reveals information about the trade-offs between objectives, helping decision makers to select a solution (Deb 2011; Branke et al. 2008). Referring to recent publications we further indicated that there is significant evidence suggesting growth and profitability interact in a conflicting way over time. Therefore the GPR cannot be analyzed in a way that assumes both measures to be generally positively correlated. As a consequence we have to look at the GPR and the underlying entrepreneurial decision process as a problem of multi-criteria optimization. To account for this situation we want to analyze how algorithms, that determine pareto-optimal sets, can enhance the analysis of startup performance. For that purpose we first want to describe what a non-dominated sorting algorithm does and then introduce the algorithm that we applied in our work.

Formally the task of a non-dominated sorting algorithm can be described as follows. Given an optimization problem Q and provided with a set of solutions P of size N and an objective space of M dimensions the algorithm has to determine a subset of solutions P' for which the following definition holds:

$$\forall \ p' \in P' \colon \nexists \ p \in P \ with \ f_i(p) \leq f_i(p')$$
$$for \ all \ i = 1, \dots, M \ with \ strict \ inequality \ for \ at \ least \ one \ i$$

The specific algorithm that we apply in our study to determine this set can be described as follows.

Algorithm 1: Non-dominated sorting algorithm from Deb (2008)

input: Optimzation problem Q, set of solutions P of size N
begin
  $P' = \{\emptyset\}$
  for each $p \in P \wedge p \notin P'$
    $P' = P' \cup p$
    for each $q \in P' \wedge q \neq p$
      If $p \succ q$
        $P' = P' \setminus \{q\}$
      else if $q \succ p$
        $P' = P' \setminus \{p\}$
      end if
    end for
  end for
end
output: Set of pareto-optimal solutions P'

Random Forest: An algorithm to determine complex, multivariate relations

As we already came to know, current literature identifies startup performance to be a complex phenomenon and suggests studying it through the use of extensive RBV models. Moreover, literature recommends investigating more complex relations between independent variables and performance measures, as they are expected to have higher predictive power. At this point machine learning could potentially enhance entrepreneurial research, as it possesses procedures able to discover complex multivariate relationships regarding multiple objectives handling vast amounts of input data at a time. To deliver a proof of concept on the analytical capabilities of machine learning in the context of startup performance, we choose to apply a tree ensemble, or RF, as an exemplary algorithm. To get a better understanding of this algorithm we will first explain it in detail and then give reasons for why we chose this particular learning algorithm to pursue or research goal.

Random forest is a supervised machine learning technique that combines elements of random subspace sampling and bagging to build ensembles of decision trees, taking decisions based on a majority voting of its base learners (Huang et al., 2013, Breiman, 2001). In more formal terms a random forest can be expressed as a set of tree-structured classifiers $h(x, \theta_i)$, where the $\theta_{i=1 \ldots M}$ are independent identically distributed random vectors, attained through bagging and elements of random subspaces, that classifies input x based on a unit voting among its members (Kulkarni & Sinha, 2012).

In pseudo code the algorithm itself can be described as follows:

1. Draw N bootstrap samples from the training set, where N represents the number of trees in the ensemble.
2. For each bootstrap sample grow a tree with the following modification: At each node, rather than choosing the best split among all M input variables, select $K \ll M$ variables at random and choose the best split among this subset of variables. Where each tree is grown till the minimum size of terminal nodes T.
3. Predict new data by providing its feature values as an input to each tree in the forest and aggregating their N predictions through majority voting (for classification problems) or averaging (for regression problems).

A reference on how to implement this algorithm can further be found at the end of this chapter.

The algorithms three main parameters are: The number N of trees in the forest, the number K of features selected for splitting when building the trees and the number T determining the minimum size of terminals nodes in all trees (Kulkarni & Sinha 2012; Li & Yue 2010; Bernard et al. 2007; Breiman 2001).

Considering our problem of studying multivariate relations between independent variables and various measures of startup performance, random forest offer a variety of beneficial properties. First of all, applied to different benchmarking and real world problems RF proved to be a highly accurate learner for a variety of data sets, being able to recognize both linear and non-linear relationships. Moreover RF is well suited to process data describing a business environment as it can process both qualitative and quantitative input variables simultaneously considering multiple objectives. In addition to that RF is a fast learner, able to handle large numbers of input features and possessing a structure that is easy to parallelize. Furthermore it is an algorithm comparably easy to use since overfitting is a lesser problem, as Breiman (2001) proved that its generalization error converges to a limit as the number of trees become larger. Besides this, RF implements bagging and therefore allows making unbiased estimations on generalization errors and feature importance prior to testing. As each tree in the ensemble is build of a bootstrap sample and each bootstrap sample leaves out about one third of the training data, error estimates can simply be obtained by predicting all training instances via the subset of trees who's samples did not contain them. These so-called out-of-bag samples can further be used to estimate variable importance prior to testing. For this purpose each features values in the out-of-bag samples are randomly disturbed and the influence of this disturbance on the misclassification rate is taken as an evidence for feature importance. Besides these rather technical reasons we also selected RF since it is a technique that already has been successfully applied to various real world problems of different domains. In medical science for example RFs are used for tasks like skin detection (Khan et al. 2010) or cell type classification (Huang et al. 2013). Further applications of RF include problems of economical science as in the study of Luo et al. (2010) on trade balance development or even problems of home-land security such as terrorist profiling (Xu et al. 2009). In computer science RFs are further used for text and image recognition (Bernard et al. 2007) with a famous commercial application being Microsoft's body motion recognition system Kinect.

Algorithm 2: Random forest algorithm from Sirikulviriya & Sinthupinyo (2011)

input: Training data D
begin
  To generate a forest of N trees
  for i = 1 to N **do**
    Randomly sample the training data D with replacement to produce D_i
    Create a rood node, N_i containing D_i
    Call BuildTree(N_i)
  end for
  BuildTree(N)
  If N contains only instances of one class **then**
    return
  else
    Randomly select K of the M possible splitting features
    Select the feature F with the highest information gain to split on
    Create f child nodes of N, $N_1, \ldots, N_f$, where F has f possible values ($F_1, \ldots, F_f$)
    for i = 1 to f **do**
      Set the contents of N_i to D_i , where D_i is all instances in N that match F_i
      Call BuildTree(N_i)
    end for
  end if
end
output: Trained random forest

4 Analysis

So far our study identified recent gaps in startup performance research and pointed out the potential of methods of computational science to close these gaps. Using RBV and DC we developed a theoretical framework to analyze the startup-performance relationship, built a coherent multifactorial data model and introduced a set of algorithms expected to realize parts of mentioned potential. With this being done, at this point we want to precede and deliver a proof of concept that is threefold: First, we want to show that pareto-optimality, achieved through non-dominated sorting, is able to determine optimal sets in real data on startup performance, considering multiple performance measures simultaneously, independent of whether they are conflicting or not. Second, we want to demonstrate that, when provided with general preferential information on performance measures, pareto-optimality can serve as a definition for success. Third, we want to investigate whether, when provided with a set of data, the machine learning algorithm random forest is able to detect multivariate rules in RBV-frameworks, enabling us to make predictions on future startup performance.

In order to carry out this proof of concept we first define a set of experimental settings, generate appropriate artificial data and introduce our methods of evaluation. Based on this we then present the results obtained, laying the groundwork for further discussion.

Figure 11 – Structure of the analysis of this study

4.1 Data set

As already mentioned before, we used artificial data to verify the performance of the algorithms under observation. At this point we want to explain first why we chose to use artificial data for testing and then how we generated this data.

Using artificial data for verification testing is a common practice in computational science. The main reason for this is not the absence of real data, but the observation that real data often does not provide all information required for a thorough assessment. Artificial data in contrast provides a controlled testing environment whose advantages are threefold (Albuquerque et al. 2011; Jeske et al. 2005; Scott & Wilkins 1999). First of all, artificial data allows investigators to know about structural regularities of the data prior to its testing. Knowing about these regularities is essential whenever one wants to assess the ability of an algorithm to uncover regularities in data. Using real data, however, prior to testing often little is known about structures hidden in the data set. A second argument in favor of testing on artificial data is, that it allows us to systematically vary the degree of data difficulty while testing, thus enabling us to assess algorithm robustness. Using real data this is rarely possible (Scott & Wilkins 1999). A third argument supporting the use of artificial data for testing is that arbitrarily large data sets can be obtained within short time. Real data in contrast is usually difficult to obtain because of its commercial value, privacy issues as well as the time and cost associated with its collection (Albuquerque et al. 2011; Jeske et al. 2005; Scott & Wilkins 1999).

With respect to our study we decided to test on artificial data due to the following reasoning: Since literature review couldn't identify any study that combined techniques of machine learning with a RBV-based multivariate model to analyze startup performance, our approach seems to be somehow unique. Therefore the main goal of our study should be to deliver a proof of concept. For a proof of concept the use of artificial data is more suitable than real data, as it offers a more controlled testing environment enabling us to define all structures hidden in the set, to gradually adjust its degree of difficulty and to test on arbitrarily large data sets.

Methodology: The origins of our data generation procedure

After having explained why we chose to test on artificial data, in the following we want to present the methodology that we applied in order to create this data. To do so, we start with a brief introduction on the overall method, before describing each step of the data generation process in detail.

Our objective was to implement a procedure that is able to generate artificial data coherent to our analytic model, that allows us to control the performance patterns in the data and is reusable for future studies. The methodology that we applied to achieve this goal was based on an approach of Scott & Wilkins (1999), whose three step process is fairly simple to explain:

1. The independent variables are generated through random sampling
2. A set of rules is defined that implements relationships between independent and dependent variables
3. The dependent variables are generated through the application of mentioned rules and an adjustable amount of noise

This approach enabled us to control the structural regularities in the data through the number of rules applied, the complexity of the rules applied and a noise parameter. While implementing this methodology we made further assumptions on our test data that are important to understand:

1. All artificially generated startups are assumed to be from the same industry, of similar size and age, which is why we refrain from initiating the control variables
2. All artificially generated startups are assumed to be survivors, which is why we limit the performance measures we generate to growth and profitability
3. The performance measures growth and profitability are assumed to be conflicting

After having explained the methodology we used to create the test data and pointing out the assumptions made while generating it, we now want to proceed to describe in detail how we initiated each of the independent and dependent variables.

Data generation: How we initialized the independent variables

Following Scott and Wilkins procedure, we initialized each independent variable of our model as a continuous variable, normally distributed within its domain. Then we discretized each variable by rounding it off and adjusting it to the constraints of its domain. Analyzing completely random sampled data, however, we noticed that some of the variables that we claimed to be independent were in fact dependent, as they had logical connections among each other. An example for such a condition is the variable x_2, which indicates the number of founding team members that hold a university degree. The value of this variable cannot be higher than the value of x_1, the variable indicating the total number of founding team members. To cope with these logical interdependencies we made further adjustments to the mean values and domains of the concerned variables. Referring to the previously mentioned example, we solved the problem by defining the normal distribution of x_2 in terms of x_1 and adjusting its domain accordingly. To provide a transparent view on our data generation process variable distributions and adjustments are displayed in appendix A3. The table shows how we initialized each variable, indicating its mean, its variance and adjustments made. Having no further information about variable distributions we defined μ as the center point of each variables domain and σ as $\frac{\mu}{2}$. For the subset of model variables that have an open interval domain $(x_1, x_{10}, x_{11}, x_{12})$ we chose a reasonable, but somehow arbitrary μ.

As a result we obtained a procedure able to generate arbitrary amounts of normally distributed independent variables that describe startup configurations and are not only compatible to our theoretical model but also logically consistent.

Data generation: : How we initialized the dependent variables

Following the approach of Scott and Wilkins, in order to generate the dependent variables' values we had to define a set of rules that describe relationships between independent and dependent variables, or in our case between success factors and performance measures. To do so, we proceeded in two steps: First we defined a set of rules that have an influence on startup performance. Then we defined how startups would use this performance to allocate it among the conflicting performance measures growth and profitability.

While defining the rule set our goal was to create structural regularities that would reflect patterns of reality. Therefore we conducted a literature review looking for empirical studies in the entrepreneurial field that analyzed relationships between independent variables and performance measures. Based on their findings we then formulated mathematical rules expressing the relations encountered in these studies. The resulting rule set is displayed in Table 1.

Table 1 – Artificial data generation: Rules to initiate the dependent variables

Rule #	Empirical finding	Mathematical formulation of the condition	Source
1	Heterogeneous skills increase performance	$if\left(\frac{x_2}{x_1}=0{,}5\right)$	Visintin & Pittino, 2014
2	Heterogeneous skills increase performance	$if\left(\frac{x_3}{x_4}=1\right)$	Visintin & Pittino, 2014
3	General experience increases performance	$if(5<(x_5+x_6+x_7+x_8)<10)$	Ganotakis, 2012
4	Experience diversity increases performance	$If(x_5>0\ AND\ x_6>0\ AND$ $x_7>0\ AND\ x_8>0)$	Ganotakis, 2012
5	The size of the founding team has an influence on startup performance	$if(3<x_1<7)$	Ganotakis, 2012
6	Strategies that fit the resource profile lead to performance	$if(x_7>2\ AND\ x_{14}<0{,}66)$	Edelmann, 2005
7	Strategies that fit the resource profile lead to performance	$if(0{,}66<x_{14}<1{,}23\ AND$ $x_7\ >2\ AND\ x_{15}>0{,}5)$	Edelmann, 2005
8	Strategies that fit the resource profile lead to performance	$if(x_6>2\ AND\ x_{14}>1{,}23)$	Edelmann, 2005

Table 1 presents the eight rules that we defined in our rule set, revealing the conditions that have to be fulfilled in order to comply with each of them and referring to the study they are taken from. Each rules originates from an empirical study whose findings we translated into a mathematical formulation, using the factors of our model. Even though some may argue that the specific mathematical formulation, which we chose to express each finding is somewhat arbitrary, which is true, its meaning however reflects a relation observed in reality. Visintin and Pittino (2014) for example studied 103 Italian startups, revealing that founding teams differentiating and integrating academic and non-academic profiles exhibited higher levels of performances. This insight is reflected by two rules in our set: Rule number one implies that equal shares in academics and non-academics in the founding team benefits a startups performance. Rule number two implies that an equal share in founding team members with STEM degree and founding team members with a social science degree benefits a startups performance. Rules number three, four and five are inspired by a study of Ganotakis (2012) who found that the level of experience in entrepreneurial teams has an inverted U-shaped relationship to startup performance. In his study he found that the combination of various skills enhances startup performance and that high levels of human capital are more likely to be encountered in entrepreneurial teams than among single founders. Rules number six, seven and eight, in contrast, follow a study of Edelman et al. (2005) who concluded that the quality of a firms strategy cannot be judged independently of the resources it is based on. While rule number six implies that startups benefit from a price strategy whenever their founding team has sufficient managerial experience, rule number seven implies that an innovative strategy is beneficial to a startup when combined with a university cooperation and sufficient managerial experience. Last but not least rule number eight implies that a quality-based strategy benefits from an entrepreneurial team having sufficient levels of technical experience.

Having defined a set of rules to establish realistic relations between independent and dependent variables, we then had to define how these would translate into specific values for growth and profitability. The main idea that we implemented can be explained as follows: First of all, we assumed that all startups have a basic level of performance in both growth and profitability. Depending on the experimental settings, this basic level of performance may or may not be disturbed by random noise. Fulfilling a performance rule is assumed to improve a startups performance beyond its basic level. A startup can use this additional, rule-based performance to improve growth or profitability. How a startup decides to distribute its rule-based performance among the performance dimensions growth and profitability is assumed to depend on its strategy. Startups implementing a price strategy are assumed to primarily opt for growth, while firms that follow a quality strategy are assumed to rather opt for profitability. In any case the two performance measures growth and profitability are assumed to be conflicting. This conflict was realized by implementing a mathematical function referred to as the ZDT 1 problem (Zitzler et al. 2000).

A mathematical formulation of our procedure to generate the dependent variable values can be expressed as follows:

$$(1)\ P_i = B + \sigma + p_i\left(x_{i,14}\right) * \sum_{j}^{8} R * \omega_j * v_{ij}$$

$$(2)\ G_i = B + \sigma + g_i\left(x_{i,14}\right) * \sum_{j}^{8} R * \omega_j * v_{ij}$$

$With:$

$P_i = Profitability\ of\ company\ i$

$G_i = Growth\ of\ company\ i$

$B = Basic\ level\ of\ performance$

$\sigma = Level\ of\ noise$

$p_i(s) = \ Preference\ of\ company\ i\ to\ opt\ for\ profitabiltiy$

$g_i(s) = \ Preference\ of\ company\ i\ top\ opt\ for\ growth$

$R = Rule\ based\ performance$

$\omega_j = Importance\ of\ performance\ rule\ j$

$v_{ij} = Variable\ indicating\ whether\ company\ i\ fulfills\ performance\ rule\ j$

$i = Number\ of\ companys\ generated$

$g_i(x_{i,14}) = \frac{x_{i,14}}{2}$

$p_i(x_{i,14}) = 1 - \sqrt{\frac{x_{i,14}}{2}}$

$x_{i,14} \in [0,2]$

$v_{ij} \in \{0,1\}$

$\omega_j \in [0,1]$

$B, \sigma, R \in \mathbb{R}$

As equations (1) and (2) indicate a startups growth and profitability consist of the basic level of performance B attributed to each company, a level of noise σ and an influence depending on how many of the j performance-rules are fulfilled. For that matter v_{ij} indicates if company i fulfills rule j, ω_j represents the relative importance of rule j, R displays the magnitude of rule-based performance and $p_i(x_{i,14})$ or $g_i(x_{i,14})$ indicate the extent to which a company utilizes this rule-based performance to improve profitability or growth. Thereby $p_i(x_{i,14})$ and $g_i(x_{i,14})$ implement a conflicting relation as defined in the ZDT 1 problem and depend on $x_{i,14}$, a variable reflecting the strategy a company perceives. Ultimately its the values B, σ, R and ω_j that are adjustable and thus to be defined individually for each experimental setting.

This procedure allowed us to project the relations defined in our rule set, into values for growth and profitability. Moreover, these values further implement a conflicting relation between growth and profitability and result from a set of eight different rules that display real world empirical findings. The procedure also allowed us to control the structural regularities and the difficulty of the data by modifying the number of rules considered, their relative importance among each other as well as the amount of noise induced. Finally, even though the generation procedure is random based, to gain further control over the structures present in the data set, our algorithm also includes the option to define the percentage of data points τ in the set, that should fulfill minimum one of the activated performance rules.

Boundaries: Limitations to the validity of our data

Despite all efforts the artificial data we generated also has deficiencies. Even though the rules it bases on are inspired from real world studies, its structure does not reflect reality. The number of rules in the set, their complexity and the level of noise induce structure into the data, however it is not expected to be as diverse as real data. Keeping these deficiencies in mind we still consider the data to be suitable for our proof of concept.

To demonstrate the benefits of the non-dominates sorting algorithm the artificial data has to implement a conflicting relationship between performance measures, a criterion that it fulfills. To proof whether RF is able to detect patterns in entrepreneurial data, it is important that the data incurs structures, a criterion our data also fulfills.

4.2 Experimental settings

Having identified gaps in startup performance research and having elaborated how selected methods of computational science might minimize these gaps, our study aims to investigate two issues. One issue is to analyze how non-dominated sorting might enhance the analysis of startup performance, as it is able to determine pareto-optimal sets considering various performance measures simultaneously. The second issue is to test to what extent the machine learning algorithm random forest is able to detect performance rules in startup data and make predictions on future startup performance.

Non-dominated sorting: Introducing two test cases

With respect to the non-dominated sorting algorithm there were two test cases that we observed. First, we wanted obtain a pareto-optimal front in an artificially generated data set where noise was absent. Second, we wanted to obtain a pareto-optimal front in a data set where noise was present. Following the mathematical formulation of our data generation process, Table 2 offers a detailed description of the testing data we created for each of these test cases. Assuming that all startups prefer higher growth and profitability to less, we then run the non-dominated sorting algorithm, as previously indicated, to determine the pareto-optimal solutions.

Table 2 – Artificial data generation for the non-dominated sorting testing

Test case	Description	B	σ	$w_{i\ \ i=1,\ldots,8}$	R	τ	# of companies
1	No noise	3,5%	0	0,125	7%	70%	1000
2	10% noise	3,5%	0,35%	0,125	7%	70%	1000

Random Forest: Introducing four test cases

With respect to the random forest algorithm there are two properties we wanted to test. First, we wanted to evaluate RF's performance regarding the regression problem of predicting future startup performance. The second property what we wanted to test was the RF's capability regarding the classification problem of predicting whether a startups performance would going to be pareto-optimal or not. For both, the regression and classification problem, we tested how the RF's performance is affected by the factors data difficulty and noise. In order to do so we defined four different testing scenarios as indicated by Figure 12.

		Level of noise	
		None	*10%*
Difficulty of data	*Low*	*Scenario 1*	*Scenario 2*
	High	*Scenario 3*	*Scenario 4*

Figure 12 – Scenarios for random forest testing

To adjust the degree of difficulty of the testing data we made use of three parameters: The number of independent variables in the data set, the number of rules present in the data set and the complexity of the rules present in the set. Test scenarios 1 and 2 had a low level of difficulty, as only two rules were active in their sets, implementing simple performance patterns based on only five of the 17 variables of our model. In addition to that we reduced the number of input variables provided to the RF by eliminating all variables of the model that were irrelevant for explaining the performance patterns. Test scenarios number 3 and 4 had a higher degree of difficulty as eight rules were present in their data sets and the RF input consisted of all 17 independent variables of the model. For the test scenarios 2 and 4 we further induced noise to the data. For that matter we used Breiman (2001) as a point of reference and chose a level of 10% noise, originating from a normal distribution with $\sigma = 0{,}35$ and $\mu = B = R = 3{,}5$.

Following the mathematical formulation of our artificial data generation process Table 3 and Table 4 provide a detailed description of testing data of each scenario.

Table 3 – Artificial data generation for RF testing: Test scenarios 1 and 2

Scenario	Description	B	σ	$w_{i\ \ i=1,2}$	R	τ	# of companies
1	Difficulty low, no noise	3,5%	0	0,5	7%	70%	1000
2	Difficulty low, 10% noise	3,5%	0,35%	0,5	7%	70%	1000

Table 4 – Artificial data generation for RF testing: Test scenarios 3 and 4

Scenario	Description	B	σ	$w_{i\ \ i=1,\dots,8}$	R	τ	# of companies
3	Difficulty high, no noise	3,5%	0	0,125	7%	70%	1000
4	Difficulty high, 10% noise	3,5%	0,35%	0,125	7%	70%	1000

Having defined the scenarios, we then combined them into test cases in order to study the effects of data difficulty and noise on the RF. To do so we performed pairwise comparisons of scenarios that only differed in one of the two attributes. Figure 13 shows the four resulting test cases, the effect they test for and the scenarios they consist of.

Test case	Effect under observation	Scenarios
1	The influence of noise on the RF performance	1 and 2
2	The influence of noise on the RF performance	3 and 4
3	The influence of data difficulty on the RF performance	1 and 3
4	The influence of data difficulty on the RF performance	2 and 4

Figure 13 – Random forest testing: The four test cases to be observed

Having explained our test cases, the scenarios they consisted of and the data configuration they are build on, we also have to specify the configuration of the RF. As mentioned before there are three main parameters to configure a RF: The number N of trees in the forest, the number K of features selected for splitting and number T determining the minimum size of terminals nodes in all trees. Following the suggestions of Breiman (2001) we set $N = 100$, $T = 10$, $K = (x)^{1/2}$ for the classification problem and $K = \frac{x}{3}$ for the regression problem, with x being the total number of independent variables provided to the RF.

System: Describing the hardware we tested on

Finally we also want to specify the configurations of the system that we used to run our tests. Our experiments were run on a laptop with the Intel Core i5 (I5-3427U) 1,8GHz processor with 8GB of RAM. The amount of RAM dedicated to MATLAB was increased to 4GB due to the computational intensity of our task.

4.3 Evaluation methods

Having explained the data generation procedure we used, the scenarios we tested on, the algorithmic configuration as well as the system that we test on, we now want to introduce the methods we applied in order to evaluate the results obtained.

Non-dominated sorting: How we evaluated dominated and non-dominated sets

To evaluate the benefit that non-dominated sorting and the resulting pareto-optimal fronts may have for startup performance research we chose a simple, visual approach. By plotting exemplary data sets and their pareto-fronts, determined through non-dominated sorting, we wanted to envision the rather theoretical concept of pareto-optimality as defined by definition 3. To do so we generated two-dimensional scatterplots that display pareto-optimal and not pareto-optimal points jointly but distinguished.

Random Forest: How we evaluated regression and classification problem

Choosing appropriate methods to evaluate learning algorithms is an important topic for machine learning and thus for our study. As a variety of different measures have been defined in the literature we will proceed giving a detailed description of the process and the methods that we selected to evaluate the performance of the RF.

First we divided each data set into two independent subsets. A training set that we used to build our model and a test set that we used to evaluate the models performance. Following common research practice we performed the partitioning of the data through a random sampling assigning 70% of the data for training and 30% for testing (Han et al. 2006).

To evaluate the regression problem of the RF, predicting startup performance, we chose the mean-squared error (MSE), as it is a common measure for evaluating numerical predictors (Luo et al. 2010; Ferri et al. 2009; Han et al. 2006; Breiman 2001).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2$$

$$a_i = actual\ value\ of\ instance\ i$$

$$p_i = predicted\ value\ for\ instance\ i$$

The MSE indicates by how much the predictions of a model deviate from the true values, penalizing strong deviations through a square term. In order to provide a better understanding of the magnitude of this error, with respect to the problem, we further chose to put the MSE into relation with the mean value of the true value distribution.

To evaluate the classification problem of the RF, predicting whether a startups performance is pareto-optimal or not, we chose four different measures widely employed in research: Accuracy, Precision, Recall and the F-Measure (Alam & Vuong 2009; Huang et al. 2013; Ferri et al. 2009; Han et al. 2006; Buja et al. 2005).

$$Accuracy = \frac{Correctly\ classified\ instances}{Total\ number\ of\ instances}$$

Accuracy is one of the most common and simplest measures to evaluate classifiers. It indicates the percentage of correctly classified instances. Despite being popular, accuracy is a measure known to be inappropriate when evaluating imbalanced data sets. If a data set is extremely skewed, using accuracy as the only measure of performance, a system can appear to be high performing by simply deeming all instances negative. However, labeling all instances as negative might be completely unsatisfying. To account for this deficiency we further included precision, recall and the F-measure into our evaluation scheme. In order to understand how these measures work we first have to explain the different types of errors a classifier can make. To do so we will use a binary classification problem as an example. When confronted with a binary problem a classifier can produce four types of outcomes that can be displayed in a confusion matrix Figure 14.

		Predicted class	
		Yes	*No*
Actual Class	*Yes*	*True positive*	*False negative*
	No	*False positive*	*True negative*

Figure 14 – The confusion matrix

In case the predictor delivers a correct classification it can either be a true positive or true negative result. In case the prediction is a false classification it can either be a false negative or a false positive result. Distinguishing between these types of classification results precision, recall and F-measure can provide a more detailed measure of the classifier performance.

$$Precision = \frac{True\ positives}{True\ positives + False\ positves}$$

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

$$F - Measure = \frac{2 * Recall * Precision}{Recall + Precision}$$

While precision describes the probability of a positive classification being truly positive, recall corresponds to the probability of a truly positive result to be classified correctly. The advantage of assessing these two indicators is that, depending on the application, one might be more important than the other. For example when designing a classifier that should diagnose whether a patient has AIDS or not high recall might be more important than precision, as false negative result causes more harm than a false positive one. The F-Measure for that matter is a simple combination of precision and recall weighing both objectives as equally important (Huang et al. 2013; Ferri et al. 2009).

Having introduced evaluation measures for both regression and classification problems at last we need a procedure to measure whether the effects of noise and data complexity have a significant influence on learner performance. To do so, we implemented paired t-tests on the results of 10 valuations with a statistical significance corresponding to $\alpha = 0{,}05$, marking significant differences with an astersik. An approach often used in literature when statistical significance is reported (Ferri et al. 2009; Banfield et al. 2007; Dietterich 1999).

4.4 Results

Having identified gaps in startup performance research we elaborated how selected methods of computational science may help to close these gaps. We further synthesized a theoretical framework that can serve as a foundation for performance analysis and built a coherent multifactorial model to describe startup companies. After identifying algorithms of interest we built an artificial data generator and defined a set of test cases to evaluate their benefit for startup performance research. In the following part we want to present the results obtained from these tests. In order to do so, we will start with our findings on the non-dominated sorting application and then proceed to the results regarding the RF.

Non-dominated sorting: Describing the results obtained

Analyzing the outcomes of our non-dominated sorting application there are two test cases we have to evaluate.

Regarding the non-noise test case, Figure 15 shows the resulting plot of dominated and non-dominated startup performances. First of all, this plot clearly reflects the way, in which our data generation method works. In the plot we observe seven different fronts of performance, each of them displaying the convex shape of the ZDT 1 problem. As in this test case all performance rules are assumed to be of equal importance, companies of different fronts differ in the number of rules they fulfill. Evaluating the set of pareto-optimal points, as determined by non-dominated sorting, we observe that it reveals both the existence of different performance fronts as well as the shape of the conflicting relationship between profitability and growth. In addition to that, it becomes clear that the pareto-optimal set offers us an easy way to compare and evaluate large amount of companies regarding multiple performance measures simultaneously. Moreover the result of this comparison is simple to visualize, thus easing its communication. Each startup symbolized by a grey point is not part of the pareto-optimal set and thus at least dominated by one other startup. If we further induce simple preferential information to this situation, by assuming that all startups in the set prefer more growth and profitability to less, pareto-optimality moreover becomes a property to distinguish between successful and unsuccessful startups. Comparing a defined set of startups solely based on the criteria growth and profitability, as it is the case in figure 15, provided with free choice no entrepreneur will consider one of the dominated companies to be successful within the set, as there is at least one other startup in the same set showing higher levels of performance, thus making it more preferable and hence a successful one.

Figure 16 shows the resulting plot of the noisy test case. Comparing the plot with Figure 15 we notice how the influence of noise blurred the lines between the different fronts of performance. Moreover we observed that Figure 16 also does not reveal the shape of the conflicting relationship between growth and profitability that clearly anymore. Looking at the non-dominated set we noticed that it does not even reflect the convex shape, as induced by the ZDT1 problem. This being so, the non-dominates set however still indicated the highest performing startups within the set. Moreover, even in this case pareto-optimality is a criterion to distinguish successful and unsuccessful startups, once simple preferential information are induced by assuming that all startups in the set prefer more growth and profitability to less.

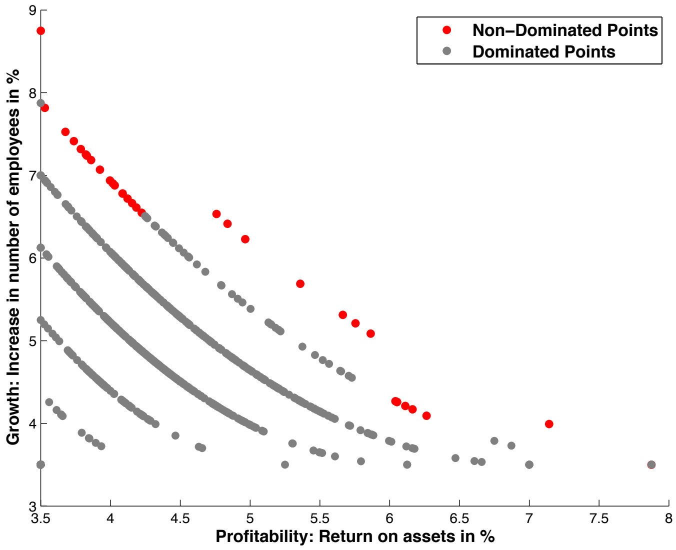

Figure 15 – Results of the non-dominated sorting: Noise free test case

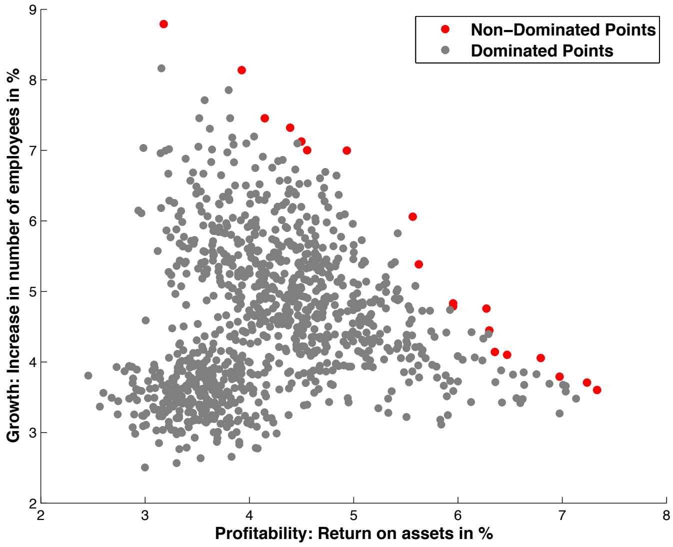

Figure 16 – Results of the non-dominated sorting: Noisy test case

Random Forest: Describing the results obtained

Following our research goal we tested RF's problem-solution capability with respect to two problem statements. In a first step we assessed the RF's performance regarding the regression problem of predicting future startup performance. In a second step we tested to what extent RF is able to handle the classification problem of predicting whether a startups performance will be pareto-optimal or not. In both cases we further assessed what kind of influence the factors data difficulty and noise had on the quality of the RF's solutions. In the following part we will present the numerical results we obtained.

<u>Regression problem – Discovering rules and predicting future performance</u>

Table 5 and Table 6 contain the results of the four test cases that we used to assess the RFs capability to detect multivariate performance rules in startup data and to make numerical predictions on future performance based on them. Looking at the results it seems as if RF is generally able to detect performance rules in entrepreneurial data. Observing the noise free test cases three out of four prediction series showed a MSE that was lower than 25% of true value mean. However, looking at the range of all results we noticed that in worst case the MSE was almost 50% of the true value mean, while in best case we measured a MSE lower than 6% of the true value mean. This insight indicates that while RF seems able to detect patterns in data on startup performance, this ability does not seem to be stable but varies across different testing environments. This is a finding is further supported by the unexpected result that the predictive power of the RF was almost consistently higher for profitability than for growth.

Analyzing the influence of noise on the RF's predictive performance we observed that an increase in the level of noise lead to an increase in the MSE in both corresponding test cases. Regarding the measure of profitability this increase was statistically significant on the $\alpha = 0{,}05$ level, for growth however it was not.

Table 5 – Results of the RF regression: The influence of noise

Scenario	Description	Profitability			Growth		
		MSE	μ_{True}	$1 - \frac{MSE}{\mu_{True}}$	MSE	μ_{True}	$1 - \frac{MSE}{\mu_{True}}$
1	Difficulty low, no noise	0,3195	3,8745	91,753%	1,4879	5,4494	72,696%
2	Difficulty low, 10% noise	1,4476*	3,3895	56,407%	1,6430	4,2562	61,398%
3	Difficulty high, no noise	0,2127	3,7769	94,368%	0,9595	4,0907	76,544%
4	Difficulty high, 10% noise	1,0010*	3,3793	70,379%	0,9238	4,1633	77,811%

Analyzing the influence of data difficulty on the RF's predictive performance we observed that an increase in the level of data difficulty lead to a decrease of the MSE. This difference was consistent and statistically significant on the $\alpha = 0{,}05$ level for growth and profitability in both corresponding test cases.

Table 6 – Results of the RF regression: The influence of data difficulty

Scenario	Description	Profitability			Growth		
		MSE	μ_{True}	$1 - \frac{MSE}{\mu_{True}}$	MSE	μ_{True}	$1 - \frac{MSE}{\mu_{True}}$
1	Difficulty low, no noise	0,3406	3,8814	91,225%	1,3484	5,4123	75,086%
3	Difficulty high, no noise	0,1979*	3,7785	94,763%	0,94*	4,9489	80,998%
2	Difficulty low, 10% noise	1,5846	3,3925	53,291%	1,603	4,2408	62,196%
4	Difficulty high, 10% noise	0,9926*	3,3732	70,574%	0,8911*	4,1382	78,467%

<u>Classification problem – Discovering rules and predicting pareto-optimality</u>

Table 7 and Table 8 present the results of the four test cases that we used to assesses, to what extent RF is able to predict whether a startups performance going to be pareto-optimal or not. Looking at the overall results it seems as if RF was highly able to learn this distinction in the course of our study. The lowest precision rate we encountered across our test series was 97,1% and the lowest recall rate that we observed was 99,7027%. Overall, accuracy was within the lines of 97,47% and 99,53%. These observations indicate that, in the course of our study, RF's capability to detect patterns of pareto-optimality were extremely high and consistent across all test cases. This examination is specifically interesting as the portion of pareto-optimal points in the test sets was only around 10% in noise free test cases and about 2% in noisy test cases.

Analyzing the influence of noise on the RF's classification performance regarding pareto-optimality, we observed that an increase in the level of noise lead to a decrease in accuracy, precision, recall and the F-Measure. These decreases were statistically significant at the $\alpha = 0{,}05$ level. Additionally the negative influence of noise on the predictive performance of the RF is consistent to the findings we made regarding the regression problem.

Table 7 – Results of the RF classification: The influence of noise

Scenario	Description	Pareto-optimality			
		$Accuracy$	$Precision$	$Recall$	$F-Measure$
1	Difficulty low, no noise	99,2667	99,5945	99,7420	99,6673
2	Difficulty low, 10% noise	$98{,}0667^*$	$98{,}1333^*$	100^*	$99{,}0575^*$
3	Difficulty high, no noise	97,4667	97,1000	100	98,5255
4	Difficulty high, 10% noise	$98{,}4667^*$	$98{,}4000^*$	100	$99{,}1932^*$

Analyzing the influence of data difficulty on the RF's predictive performance we obtained inconsistent results. Increasing the level of data difficulty in a noise free environment lead to a significant decrease of the predictive performance across all measures. In a noisy environment however an increase in data difficulty lead to a significant increase in accuracy, precision and the F-Measure. These results are further inconsistent to the observations we made with respect to the regression problem, where an increase in data difficulty had a persistently positive effect on predictive performance.

Table 8 – Results of the RF classification: Influence of data difficulty

Scenario	Description	Pareto-optimality			
		$Accuracy$	$Precision$	$Recall$	$F-Measure$
1	Difficulty low, no noise	99,5333	99,8898	99,7027	99,7958
3	Difficulty high, no noise	$97{,}5000^*$	$97{,}7000^*$	100^*	$98{,}8359^*$
2	Difficulty low, 10% noise	97,9333	97,8667	100	98,9217
4	Difficulty high, 10% noise	$98{,}5000^*$	$98{,}2667^*$	100	$99{,}1256^*$

5 Discussion

Discussing our test results there are two types of findings that we have to cover. On one side we have to examine the results obtained through the application of the non-dominated sorting algorithm. On the other side we have discuss the results of the RF application regarding the regression and classification problem.

Looking at the results of our non-dominated sorting application we can confirm that pareto-optimality is a property that allows to compare startups regarding various performance measures simultaneously and independent of whether they are conflicting or not. Our plots further show that, in some cases, pareto-optimal sets are also able to visually reveal the shape of the relationship existing between performance measures. Pareto-optimality also proves to be a quality that is easy to visualize and thus suitable for communicating large amounts of results in short time. In addition to that, our test cases underline how pareto-optimality is a property that allows us to define a distinction between startup performance and startup success, whenever simple but realistic assumptions about the entrepreneurs' preferences regarding performance measures can be made. Mentioning pareto-optimality as a valuable property for startup performance analysis, non-dominated sorting is a simple and well-established method to determine this property. Looking at the results of our study it seems that scientists should consider non-dominated sorting as a method and pareto-optimality as a criteria, when analyzing startup performance regarding multiple performance measures. This would allow them to analyze performance measures jointly but separate, possibly revealing information about their intricate relationship. In addition to that, when analyzing startup performance we suggest using the notion of pareto-optimality in order to establish a distinction between the terms of startup performance and startup success, whenever reasonable assumptions about respective preferences can be made.

Coming to speak of the results of our RF application we have to distinguish between the regression and the classification problem. Resuming our findings on the regression problem we can say that the RF was generally able to detect the multivariate rules defined in our rule set. Especially in noise free test scenarios the MSE was lower than 10% of the true value mean. The ability to detect these patterns however seems to be unstable and influenced by the factors data difficulty and level of noise. Higher levels of noise diminished prediction precision, as one would expect, since the algorithm lacks information to reason the induced deviations. In our test cases it seemed as if the bagging technique implemented by RF was not capable of providing sufficient variance reduction to the noise induced. In contrast to noise, increasing levels of data difficulty lead to more precise predictions of the RF. This finding is somehow counter-intuitive and thus has to be examined. Doing so, we have a simple and adequate explanation for this result, which also reveals an important insight to the reader on how to handle RFs correctly. The data sets of our study labeled with low difficulty did not only apply a small amount of rules, but also consisted of only five out of the 17 variables of our model, where all of these five variables were part of at least one rule in the set. Since RF is an algorithm that applies random subspace sampling, building its underlying decision trees it does not include all variables provided. However, since in our low difficulty cases all variables were highly relevant for pattern recognition, the declined predictive performance is very likely to be caused by the random subspace sampling excluding an unusual high amount of performance-relevant variables. The insight gained through this observation is that, in its default configuration, RF has no difficulty in handling data where not all variables provided are relevant for explaining the patterns to be recognized. If however the opposite should be the case, for example as a result of an effective variable reduction method, one has to adjust the random subspace sampling of the RF accordingly, as leaving out too many relevant variables may lower its predictive power. Resuming our findings on the regression problem, we find that RF seems to be a suitable method for scientists in the entrepreneurial field to detect multiple multifactorial performance rules in startup data. RF confirmed that it's able to handle many variables at a time and can recognize linear as well as non-linear relationships. The input data provided to the algorithm should be extensive in order to increase the likelihood that it includes all relevant variables, thus minimizing the influence of noise. While including large amounts of data, it is not a major problem for RF if part of this data is irrelevant to the patterns to be recognized, as it includes random subspace sampling. Researchers could make use of RF itself in order to determine variable importance and sharpen their models, but subsequently should adjust their random subspace sample accordingly.

Revising the RF's results regarding the classification problem of predicting pareto-optimality, it seems as if the algorithm is highly able to learn and predict this quality. The predictive power of RF was very stable across all our test cases, as the lowest result observed regarding accuracy, precision and recall was still above the 95% rate. This result seems promising, especially knowing that only 2-10% of the data points tested were actually pareto-optimal. Even though the predictive ability of RF seemed stable, the level of noise still had a significantly negative influence on it. This observation followed our expectations and was consistent to what we examined in the regression case. However the effect of increased noise seemed to have a more severe effect on the regression results than on the classification results. One possible explanation for this might be the increased granularity of the regression problem as compared to the classification problem. Looking at the influence that data difficulty had on the classification performance, we observed inconsistent results, which we find difficult to explain. This issue may require further study, its observation however is outshined by the level of precision that we encountered throughout all test cases. It seems as if RF is generally capable of learning patterns that identify pareto-optimal solutions. Thus we can suggest scientist to apply the algorithm to real startup data, in the hope that it can provide accurate predictions on the pareto-optimality of startup performance. To complete the picture on this classification problem we also have to mention that the informational value of the criterion pareto-optimality depends on the number of the performance measures under observation. The more measures one observes the higher the probability of a point being pareto-optimal, which implies that the characteristic of being pareto-optimal becomes less meaningful when observing large amounts of performance measures simultaneously.

As positive as our results may seem, the findings of our study are also limited by certain factors. Regarding the application of non-dominated sorting our study only demonstrated the concept exemplarily and presented arguments supporting its benefit for the analysis of startup performance. However we did not apply non-dominated sorting to specific real world data or deliver a quantitative evaluation of its analytical benefit. Regarding RFs ability to detect multivariate performance rules and predict startup performance we only tested this capacity based on artificial data. While the use of artificial data is highly suggested for validation testing, our results however do not provide evidence on how effective RFs predictive power may be in real-world scenarios. Looking for further limitations to our study we also have to realize that the artificial data we generated did only include specific types of performance rules and that the spectrum of performance rules to be observed in reality possibly is more extensive and diverse.

6 Conclusion

Both scientists and politicians recognize the importance of entrepreneurship for a healthy economic development. Entrepreneurship is a driving force of economic evolution that determines its speed and direction. In search for new growth, it is important to understand how entrepreneurial activity can be supported so that it may create sustainable value and additional employment. Hence, the study of startup performance becomes a topic of interest.

A review of recent publications on startup performance revealed three obstacles in research. First, the prevailing assumption of a generally positive GPR is empirically not supported. Second, research uses the terms startup performance and startup success synonymously even though they are semantically different. Third, analyzing simple relations between performance measures and independent variables did not discover rules with consistently high predictive power. Keeping these findings in mind, we decided to explore what benefits the application of modern methods of computational science, specifically multi-criteria optimization and machine learning, may have for startup performance analysis. Literature review revealed that similar approaches have not been studied so far, even though there is reason to believe they might enhance entrepreneurial theory development significantly. With this presumption in mind, we combined RBV and DC into a theoretical framework that allowed us to analyze startup performance and coherently built a multifactorial model enabling us to test exemplary algorithms. Using artificially generated data we created a variety of test cases to determine the problem-solution capabilities that non-dominated sorting and RF have in the context of startup performance analysis.

Our results confirmed that pareto-optimality is a property that allows to compare various performance measures simultaneously and independent of their interdependencies. In some cases determining the pareto-optimal set even revealed the shape of their intrinsic relationship. Additionally, pareto-optimality is a quality that is easy to visualize and thus suitable for communicating large amounts of results. Besides this, our test examples underlined that pareto-optimality allows us to define a distinction between startup performance and startup success, whenever reasonable assumptions about the general preferences regarding performance measures can be made. Analyzing pareto-optimality, non-dominated sorting is a simple, reliable and well-established algorithm to determine this property. Resuming our findings on the RF algorithm we can say that it was generally able to detect multivariate rules in our RBV-based framework. Regarding the classification problem of predicting the pareto-optimality of a startups future performance, the algorithm seemed highly capable of learning respective patterns. In this case the predictive power was stable across all our test cases with the lowest result regarding accuracy, precision and recall being still above the 95% rate. Regarding the regression problem of predicting future startup performance we can say that RF was partially able to detect the multivariate rules defined in our test cases. Especially in non-noise scenarios its predictive performance was acceptable with a MSE being lower than 10% of the true value mean. Our tests however indicate that the level of noise has a significantly negative influence on RFs predictive power regarding both, the regression and classification task. The level of data difficulty however only had a significant influence on the regression results, where we suspect the random subspace sampling of RF to be the cause of lesser precision in cases with few variable input.

As positive as these results may seem, their validity is also limited. The application of non-dominated sorting was only demonstrated exemplarily using artificial data and did not provide a quantification of its analytical benefit. Regarding the RFs ability to detect rules and predict startup performance we only tested this capacity based on artificial data. While for a proof of concept the use of artificial data is highly suggested, our results however do not provide evidence on how effective RFs predictive power may be in real-world scenarios. However, the fact that our analytical model is connected to real data, as it I compatible to a standardized questionnaire developed by our institute, eases its applicability to real world scenarios in future studies.

Even though there are certain limitations to our findings of this study, future research may still benefit from them. Based on our results, scientists should consider the use of non-dominated sorting as a method and pareto-optimality as a criterion, to analyze startup performance regarding multiple performance measures. This would allow them to analyze performance measures jointly but separate, possibly revealing information about their intricate relationship. In addition to that, they may use the notion of pareto-optimality to establish a distinction between the terms of startup performance and startup success, whenever reasonable assumptions about preferences can be made. Besides this our study also demonstrated that RF is capable of recognizing multivariate patterns in startup data. It is able to handle the many variables of an extensive RBV-framework and to recognize complex relationships, performing regression and classification tasks. Scientist should apply RF, or other comparable machine learning algorithms, to entrepreneurial data in order to discover more consistent rules regarding startup performance, ideally allowing to obtain more accurate predictions on performance. Looking at the broader picture of this study, in the long run, combining the results of machine learning with methods of multi-criteria optimization could even hold additional benefits for entrepreneurial theory development. Methods of multi-criteria optimization could help to build mathematical models linking independent variables and performance measures considering conflicts among objectives. Algorithms of machine learning could provide approximations of their objective functions, allowing to make predictions on unobserved instances. By integrating these approximations into the optimization problem scientists could obtain a formal description of the startup-performance problem. Having this, researchers could then advance and apply optimization techniques to determine theoretically optimal solutions to entrepreneurial decision problems regarding performance.

Having mentioned the long-term perspective implied by our study, in order to continue along its line, there are four distinct aspects future research could investigate upon. First of all, having verified the abilities and benefits of non-dominated sorting and RF within this thesis, consecutive studies should apply these algorithms to real startup data in order to examine their effectiveness in real world scenarios. Second, future studies could focus on machine learning and analyze which algorithms are most suitable to make predictions on startup performance, for example by comparing RF with neuronal networks and support vector machines. Third, depending on the learning algorithm applied, future studies could further focus on the task of extracting explicit performance rules from trained systems in order to improve the understanding of the driving mechanisms behind startup performance. Fourth, subsequent studies could finally aim to combine trained machine learning algorithms with procedures of multi-criteria optimization in order to determine theoretically optimal solutions to performance-related entrepreneurial decision problems.

Bibliography

Alam, M.S. & Vuong, S.T., 2009. Random Forest Classification for Detecting Android Malware. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. IEEE, pp. 663–669. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6682136 [Accessed August 29, 2014].

Albuquerque, G., Löwe, T. & Magnor, M., 2011. Synthetic generation of high-dimensional datasets. *IEEE transactions on visualization and computer graphics*, 17(12), pp.2317–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22034352.

Audretsch, D., 2007. Entrepreneurship capital and economic growth. *Oxford Review of Economic Policy*, 23(1), pp.63–78. Available at: http://oxrep.oxfordjournals.org/content/23/1/63.short [Accessed April 17, 2014].

Audretsch, D., 2012. Entrepreneurship research. *Management Decision*, 50(5), pp.755–764. Available at: http://www.emeraldinsight.com/10.1108/00251741211227384 [Accessed September 2, 2014].

Banfield, R.E. et al., 2007. A comparison of decision tree ensemble creation techniques. *IEEE transactions on pattern analysis and machine intelligence*, 29(1), pp.173–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17108393.

Bernard, S. et al., 2007. Using Random Forests for Handwritten Digit Recognition. In *2007 9th International Conference on Document Analysis and Recognition, ICDAR*. IEEE, pp. 1043 – 1047.

Branke, J. et al., 2008. *Multiobjective optimization: Interactive and evolutionary approaches*, Heidelberg: Springer-Verlag. Available at: http://books.google.com/books?hl=en&lr=&id=N-1hWMNUa2EC&oi=fnd&pg=PA1&dq=Multiobjective+Optimization+interactive+and+evolutionary+approaches&ots=eCBuUu5K3W&sig=v1C7TJdZ_4RjqIrLFYawHn799ps [Accessed May 15, 2014].

Brännback, M. et al., 2009. Growth and profitability in small privately held biotech firms: preliminary findings. *New biotechnology*, 25(5), pp.369–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19491050 [Accessed September 29, 2014].

Breiman, L.E.O., 2001. Random Forests. *Machine Learning*, 45(1), pp.5–32.

Brown, G. & Wyatt, J.L., 2005. Managing Diversity in Regression Ensembles. *Journal of Machine Leraning Research*, 6, pp.1621–1650.

Buja, A., Stuetzle, W. & Shen, Y., 2005. Loss functions for binary class probability estimation and classification: Structure and applications. *Working draft, November*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Loss+Functions+for+Binary+Class+Probability+Estimation+and+Classification+:+Structure+and+Applications#0 [Accessed October 25, 2014].

Capon, N., Farley, J.U. & Hoenig, S., 1990. Determinants of Financial Performance: A Meta-Analysis. *Management Science*, 36(10), pp.1143–1159.

Cassar, G., 2014. Industry and startup experience on entrepreneur forecast performance in new firms. *Journal of Business Venturing*, 29(1), pp.137–151. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902612000948 [Accessed May 4, 2014].

Chandler, G. & Hanks, S., 1993. Measuring the performance of emerging businesses: A validation study. *Journal of Business venturing*, 8(5), pp.391–408. Available at: http://www.sciencedirect.com/science/article/pii/088390269390021V [Accessed September 29, 2014].

Choi, Y., 2004. Entrepreneurs' Decisions to Exploit Opportunities. *Journal of Management*, 30(3), pp.377–395. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0149206303000990 [Accessed August 4, 2014].

Chorev, S. & Anderson, A.R., 2006. Success in Israeli high-tech start-ups; Critical factors and process. *Technovation*, 26(2), pp.162–174. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497205001008 [Accessed May 15, 2014].

Coad, a., 2010. Exploring the processes of firm growth: evidence from a vector auto-regression. *Industrial and Corporate Change*, 19(6), pp.1677–1703. Available at: http://icc.oxfordjournals.org/cgi/doi/10.1093/icc/dtq018 [Accessed September 29, 2014].

Coad, A. et al., 2014. Business experience and start-up size: Buying more lottery tickets next time around? *Small Business Economics*, 43(3), pp.529–547. Available at: http://link.springer.com/10.1007/s11187-014-9568-2 [Accessed September 29, 2014].

Colombo, M.G. & Grilli, L., 2010. On growth drivers of high-tech start-ups: Exploring the role of founders' human capital and venture capital. *Journal of Business Venturing*, 25(6), pp.610–626. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902609000093 [Accessed July 23, 2014].

Czarnitzki, D. & Delanote, J., 2012. Young Innovative Companies: the new high-growth firms? *Industrial and Corporate Change*, 22(5), pp.1315–1340. Available at: http://icc.oxfordjournals.org/cgi/doi/10.1093/icc/dts039 [Accessed September 29, 2014].

Davidsson, P., Steffens, P. & Fitzsimmons, J., 2009. Growing profitable or growing from profits: Putting the horse in front of the cart? *Journal of Business Venturing*, 24(4), pp.388–406. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902608000475 [Accessed August 30, 2014].

Davila, A., Foster, G. & Gupta, M., 2003. Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, 18(6), pp.689–708. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902602001271 [Accessed April 30, 2014].

Deb, K., 2011. Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*. London: Springer, pp. 3–34. Available at: http://link.springer.com/10.1007/978-0-85729-652-8_1 [Accessed September 28, 2014].

Deb, K., 2008. *Multi-Objective Optimization using Evolutionary Algorithms* 1. publ., Chichester; Weinheim: Wiley.

Deb, K., 2003. Unveiling innovative design principles by means of multiple conflicting objectives. *Engineering Optimization*, 35(5), pp.445–470. Available at: http://www.tandfonline.com/doi/abs/10.1080/0305215031000151256 [Accessed September 28, 2014].

Delmar, F., 2008. The Effect of Small Business Managers' Growth Motivation on Firm Growth: A Longitudinal Study. *Entrepreneurship Theory and Practice*, 32(3), pp.437–457.

Delmar, F., Davidsson, P. & Gartner, W.B., 2003. Arriving at the high-growth firm. *Journal of Business Venturing*, 18(2), pp.189–216. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902602000800 [Accessed September 2, 2014].

Delmar, F., McKelvie, A. & Wennberg, K., 2013. Untangling the relationships among growth, profitability and survival in new firms. *Technovation*, 33(8-9), pp.276–291. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497213000199 [Accessed September 2, 2014].

Dess, G., Lumpkin, G.T. & Covin, J.G., 1997. Entrepreneurial strategy making and firm performance: tests of contingency and configurational models. *Strategic management journal*, 18(9), pp.677–695.

Dickersbach, J., 2008. *Supply Chain Management with SAP APO* 3rd editio., Berlin/Heidelberg: Springer.

Dietterich, T., 1999. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2), pp.139–157. Available at: http://link.springer.com/article/10.1023/A:1007607513941 [Accessed September 28, 2014].

Dodge, H.R., 1994. Stage of the organizational life cycle and competition as mediators of problem perception for small businesses. *Strategic management journal*, 15(2), pp.121–134.

Døving, E., Gooderham, P.N. & Dving, E., 2008. Dynamic Capabilities as Antecedents of the Scope of Related Diversification : The Case of Small Firm Accountancy Practices. *Strategic management journal*, 29(8), pp.841–857.

Edelman, L.F., Brush, C.G. & Manolova, T., 2005. Co-alignment in the resource–performance relationship: strategy as mediator. *Journal of Business Venturing*, 20(3), pp.359–383. Available at: http://linkinghub.elsevier.com/retrieve/pii/S088390260400031X [Accessed September 8, 2014].

Ehrgott, M., 2005. *Multicriteria optimization*, Berlin/Heidelberg: Springer-Verlag. Available at: http://link.springer.com/10.1007/3-540-27659-9 [Accessed September 28, 2014].

Eisenhardt, K. & Martin, J., 2000. Dynamic capabilities: what are they? *Strategic management journal*, 21(10-11), pp.1105–1121. Available at: http://doi.wiley.com/10.1002/1097-0266%28200010/11%2921%3A10/11%3C1105%3A%3AAID-SMJ133%3E3.0.CO%3B2-E [Accessed September 29, 2014].

Federico, J.S. & Capelleras, J.-L., 2014. The heterogeneous dynamics between growth and profits: the case of young firms. *Small Business Economics*. Available at: http://link.springer.com/10.1007/s11187-014-9598-9 [Accessed September 29, 2014].

Ferri, C., Hernández-Orallo, J. & Modroiu, R., 2009. An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1), pp.27–38. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167865508002687 [Accessed October 25, 2014].

Ganotakis, P., 2012. Founders' human capital and the performance of UK new technology based firms. *Small Business Economics*, 39(2), pp.495–515. Available at: http://link.springer.com/10.1007/s11187-010-9309-0 [Accessed September 29, 2014].

Gavrishchaka, V., Koepke, M.E. & Ulyanova, O.N., 2010. Ensemble Learning Frameworks for the Discovery of Multi-component Quantitative Models in Biomedical Applications. In *2010 International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. IEEE, pp. 856 – 859. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5421550 [Accessed September 28, 2014].

Gimmon, E. & Levie, J., 2010. Founder's human capital, external investment, and the survival of new high-technology ventures. *Research Policy*, 39(9), pp.1214–1226. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0048733310001411 [Accessed July 25, 2014].

Gronum, S., Verreynne, M.-L. & Kastelle, T., 2012. The Role of Networks in Small and Medium-Sized Enterprise Innovation and Firm Performance. *Journal of Small Business Management*, 50(2), pp.257–282. Available at: http://doi.wiley.com/10.1111/j.1540-627X.2012.00353.x.

Han, J., Kamber, M. & Pei, J., 2006. *Data Mining: Concepts and Techniques* 2nd Editio., San Francisco: Morgan Kaufmann Publ.

Hansen, E., 1995. Entrepreneurial networks and new organization growth. *Entrepreneurship Theory and Practice*, 19(4), pp.7–19. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Entrepreneurial+Networks+and+New+Organization+Growth#0 [Accessed September 29, 2014].

Hoskisson, R.E. et al., 2011. Revitalizing Entrepreneurship: The Search for New Research Opportunities. *Journal of Management Studies*, 48(6), pp.1141–1168. Available at: http://doi.wiley.com/10.1111/j.1467-6486.2010.00997.x [Accessed September 29, 2014].

Huang, H.-C., Lai, M.-C. & Lo, K.-W., 2012. Do founders' own resources matter? The influence of business networks on start-up innovation and performance. *Technovation*, 32(5), pp.316–327. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497211001854 [Accessed September 23, 2014].

Huang, J.-H. et al., 2013. Using random forest to classify T-cell epitopes based on amino acid properties and molecular features. *Analytica chimica acta*, 804, pp.70–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24267065 [Accessed September 28, 2014].

Hvide, H.K. & Møen, J., 2010. Lean and Hungry or Fat and Content? Entrepreneurs' Wealth and Start-Up Performance. *Management Science*, 56(8), pp.1242–1258. Available at: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1100.1177 [Accessed July 9, 2014].

Janssen, R. et al., 2008. Multiobjective decision support for land-use planning. *Environment and Planning B: Planning and Design*, 35(4), pp.740–756. Available at: http://www.envplan.com/abstract.cgi?id=b33071 [Accessed August 19, 2014].

Jeske, D.R. et al., 2005. Generation of synthetic data sets for evaluating the accuracy of knowledge discovery systems. *Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining - KDD '05*, p.756. Available at: http://portal.acm.org/citation.cfm?doid=1081870.1081969.

Kakati, M., 2003. Success criteria in high-tech new ventures. *Technovation*, 23(5), pp.447–457.

Khan, R., Hanbury, A. & Stoettinger, J., 2010. Skin detection: A random forest approach. In *2010 17th IEEE International Conference on Image Processing (ICIP)*. IEEE, pp. 4613–4616.

Kotsiantis, S., 2010. Combining bagging, boosting, rotation forest and random subspace methods. *Artificial Intelligence Review*, 35(3), pp.223–240. Available at: http://link.springer.com/10.1007/s10462-010-9192-8 [Accessed September 28, 2014].

Kulkarni, V.Y. & Sinha, P.K., 2012. Pruning of Random Forest classifiers: A survey and future directions. In *2012 International Conference on Data Science & Engineering (ICDSE)*. IEEE, pp. 64–68. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6282329.

Lee, C., Lee, K. & Pennings, J.M., 2001. Internal capabilities, external networks, and performance: a study on technology-based ventures. *Strategic Management Journal*, 22(6-7), pp.615–640. Available at: http://doi.wiley.com/10.1002/smj.181 [Accessed July 14, 2014].

Lévesque, M., 2004. Mathematics, theory, and entrepreneurship. *Journal of Business Venturing*, 19(5), pp.743–765. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902603000284 [Accessed September 28, 2014].

Li, D. & Liu, J., 2014. Dynamic capabilities, environmental dynamism, and competitive advantage: Evidence from China. *Journal of Business Research*, 67(1), pp.2793–2799. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296312002202 [Accessed September 8, 2014].

Li, Q. & Yue, M., 2010. Correlation Analysis Based on Random Forest in Chinese Alcohol Market. *2010 International Conference on Measuring Technology and Mechatronics Automation*, pp.856–859. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5458756 [Accessed October 25, 2014].

Lin, Y. & Wu, L.-Y., 2014. Exploring the role of dynamic capabilities in firm performance under the resource-based view framework. *Journal of Business Research*, 67(3), pp.407–413. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296312003645 [Accessed August 8, 2014].

Luo, L., Lv, W. & Zhang, R., 2010. An empirical study on balance of trade in China based on random forest regression model. In *2010 5th International Conference on Computer Science & Education*. IEEE, pp. 561–564. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5593550.

Mackerell, A.D., 2004. Empirical force fields for biological macromolecules: overview and issues. *Journal of computational chemistry*, 25(13), pp.1584–604. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15264253 [Accessed July 16, 2014].

Mansury, M.A. & Love, J.H., 2008. Innovation, productivity and growth in US business services: A firm-level analysis. *Technovation*, 28(1-2), pp.52–62. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497207000818 [Accessed September 22, 2014].

Markman, G.D. & Gartner, W.B., 2002. Is Extraordinary Growth Profitable? A Study of Inc. 500 High-Growth Companies*. *Entrepreneurship Theory and Practice*, 27(1), pp.65–75.

Maxville, V., 2013. Introducing: Computational Science. *Procedia Computer Science*, 18, pp.1456–1465. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1877050913004560 [Accessed September 9, 2014].

McDougall, P.P. et al., 1994. The effects of industry growth and strategic breadth on new venture performance and strategy content. *Strategic Management Journal*, 15(7), pp.537–554. Available at: http://doi.wiley.com/10.1002/smj.4250150704.

McFadyen, M.A., Semadeni, M. & Cannella, A. a., 2009. Value of Strong Ties to Disconnected Others: Examining Knowledge Creation in Biomedicine. *Organization Science*, 20(3), pp.552–564. Available at: http://pubsonline.informs.org/doi/abs/10.1287/orsc.1080.0388 [Accessed July 22, 2014].

Mckelvie, A. & Wiklund, J., 2010. Advancing Firm Growth Research: A Focus on Growth Mode Instead of Growth Rate. *Entrepreneurship Theory and Practice*, 34(10), pp.261–288.

Murphy, G.B., Trailer, J.W. & Hill, R.C., 1996. Measuring performance in entrepreneurship research. *Journal of Business Research*, 36(1), pp.15–23. Available at: http://linkinghub.elsevier.com/retrieve/pii/014829639500159X [Accessed September 29, 2014].

Newbert, S., 2008. Value, rareness, competitive advantage, and performance: a conceptual-level empirical investigation of the resource-based view of the firm. *Strategic management journal*, 29(7), pp.745–768. Available at: http://onlinelibrary.wiley.com/doi/10.1002/smj.686/abstract [Accessed September 29, 2014].

Olhofer, M. & Sendhoff, B., 2001. Adaptive encoding for aerodynamic shape optimization using evolution strategies. *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, 1, pp.576–583. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=934443.

Pe'er, A. & Keil, T., 2013. Are all startups affected similarly by clusters? Agglomeration, competition, firm heterogeneity, and survival. *Journal of Business Venturing*, 28(3), pp.354–372. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902612000626 [Accessed September 2, 2014].

Plummer, L. a. & Acs, Z.J., 2014. Localized competition in the knowledge spillover theory of entrepreneurship. *Journal of Business Venturing*, 29(1), pp.121–136. Available at: http://linkinghub.elsevier.com/retrieve/pii/S088390261200095X [Accessed September 29, 2014].

Prajogo, D.I., 2006. The relationship between innovation and business performance—a comparative study between manufacturing and service firms. *Knowledge and Process Management*, 13(3), pp.218–225. Available at: http://doi.wiley.com/10.1002/kpm.259.

Rauch, A. et al., 2009. Entrepreneurial Orientation and Business Performance: An Assessment of Past Research and Suggestions for the Future. *Entrepreneurship Theory and Practice*, 33(3), pp.761–787. Available at: http://doi.wiley.com/10.1111/j.1540-6520.2009.00308.x [Accessed July 13, 2014].

Reid, G.C., 1995. Early life-cycle behaviour of micro-firms in Scotland. *Small Business Economics*, 7(2), pp.89–95. Available at: http://link.springer.com/10.1007/BF01108684.

Rezaei, J., Ortt, R. & Scholten, V., 2012. Measuring entrepreneurship : Expert-based vs . data-based methodologies. *Expert Systems With Applications*, 39(4), pp.4063–4074. Available at: http://dx.doi.org/10.1016/j.eswa.2011.09.091.

Ries, E., 2011. *The Lean Startup* 1.publ. ed., London: Portfolio Penguin.

Rocha, a, Papa, J. & Meira, L. a a, 2010. How Far You Can Get Using Machine Learning Black-Boxes. *2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images*, pp.193–200. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5720365 [Accessed September 28, 2014].

Rosenbusch, N., Brinckmann, J. & Bausch, A., 2011. Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing*, 26(4), pp.441–457. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902609001232 [Accessed July 10, 2014].

Santos, S., Belton, V. & Howick, S., 2002. Adding value to performance measurement by using system dynamics and multicriteria analysis Se. *International Journal of Operations & Production Management*, 22(11), pp.1246–1272.

Sassmannshausen, S., 2012. *99 entrepreneurship journals: A comparative empirical investigation of rankings, impact, and H/HC-index*, Wuppertal. Available at: http://www.econstor.eu/handle/10419/68719 [Accessed September 29, 2014].

Schumpeter, J., 1934. *The theory of economic development*, Cambridge: Harvard University Press.

Scott, P.D. & Wilkins, E., 1999. Evaluating data mining procedures: techniques for generating artificial data sets. *Information and Software Technology*, 41(9), pp.579–587. Available at: http://linkinghub.elsevier.com/retrieve/pii/S095058499900021X.

Shao, L. & Ehrgott, M., 2008. Approximately solving multiobjective linear programmes in objective space and an application in radiotherapy treatment planning. *Mathematical Methods of Operations Research*, 68(2), pp.257–276. Available at: http://link.springer.com/10.1007/s00186-008-0220-2 [Accessed September 28, 2014].

Shukla, P.K. & Deb, K., 2007. On finding multiple Pareto-optimal solutions using classical and evolutionary generating methods. *European Journal of Operational Research*, 181(3), pp.1630–1652. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0377221706005431 [Accessed September 28, 2014].

Steffens, P., Davidsson, P. & Fitzsimmons, J., 2009. Performance Configurations Over Time: Implications for Growth- and Profit-Oriented Strategies. *Entrepreneurship Theory and Practice*, 33(1), pp.125–148.

Stuart, R. & Abetti, P., 1987. Start-up ventures: Towards the prediction of initial success. *Journal of Business Venturing*, 2(3), pp.215–230. Available at: http://www.sciencedirect.com/science/article/pii/0883902687900103 [Accessed September 29, 2014].

Stucki, T., 2013. Success of start-up firms: the role of financial constraints. *Industrial and Corporate Change*, 23(1), pp.25–64. Available at: http://icc.oxfordjournals.org/cgi/doi/10.1093/icc/dtt008 [Accessed May 9, 2014].

Teece, D.J., Pisano, G. & Shuen, A.M.Y., 1997. DYNAMIC CAPABILITIES AND STRATEGIC. *Strategic management journal*, 18(7), pp.509–533.

Terjesen, S. & Wang, N., 2012. Coase on entrepreneurship. *Small Business Economics*, 40(2), pp.173–184. Available at: http://link.springer.com/10.1007/s11187-012-9468-2 [Accessed October 25, 2014].

The Oxford Dictionary of English, 2010. 3rd edition. New York: Oxford University Press.

Thornhill, S., 2006. Knowledge, innovation and firm performance in high- and low-technology regimes. *Journal of Business Venturing*, 21(5), pp.687–703. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902605000509 [Accessed September 29, 2014].

Thornhill, S. & Amit, R., 2003. Learning about failure: bankruptcy, firm age, and the resource-based view. *Organization Science*, 14(5), pp.497–509.

Visintin, F. & Pittino, D., 2014. Founding team composition and early performance of university—Based spin-off companies. *Technovation*, 34(1), pp.31–43. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0166497213001090 [Accessed July 18, 2014].

Wallenius, J. et al., 2008. Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead. *Management Science*, 54(7), pp.1336–1349. Available at: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0838 [Accessed August 22, 2014].

Wang, H., Ma, C. & Zhou, L., 2009. A Brief Review of Machine Learning and Its Application. *2009 International Conference on Information Engineering and Computer Science*, pp.1–4. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5362936.

Webb, G.I. & Zheng, Z., 2004. Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. *IEEE Transactions on Knowledge and Data Engineering*, 16(8), pp.980–991. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1318582.

Wernerfelt, B., 1984. A resource-based view of the firm. *Strategic management journal*, 5(2), pp.171–180. Available at: http://onlinelibrary.wiley.com/doi/10.1002/smj.4250050207/abstract [Accessed September 29, 2014].

Wiklund, J. et al., 2011. The Future of Entrepreneurship Research. *Entrepreneurship Theory and Practice*, 35(1), pp.1–9. Available at: http://doi.wiley.com/10.1111/j.1540-6520.2010.00420.x [Accessed August 14, 2014].

Wiklund, J., Davidsson, P. & Delmar, F., 2003. What Do They Think and Feel about Growth? An Expectancy-Value Approach to Small Business Managers' Attitudes Toward Growth1. *Entrepreneurship Theory and Practice*, 27(3), pp.247–270.

Wiklund, J. & Shepherd, D., 2005. Entrepreneurial orientation and small business performance: a configurational approach. *Journal of Business Venturing*, 20(1), pp.71–91. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0883902604000047 [Accessed July 26, 2014].

Wiklund, J. & Shepherd, D., 2003. Knowledge-based resources, entrepreneurial orientation, and the performance of small and medium-sized businesses. *Strategic management journal*, 24(13), pp.1307–1314. Available at: http://onlinelibrary.wiley.com/doi/10.1002/smj.360/full [Accessed September 29, 2014].

Wu, J., 2009. A Novel Artificial Neural Network Ensemble Model Based on K--Nearest Neighbor Nonparametric Estimation of Regression Function and Its Application for Rainfall Forecasting. In *2009 International Joint Conference on Computational Sciences and Optimization*. IEEE, pp. 44–48. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5193894 [Accessed September 28, 2014].

Wu, L.-Y., 2007. Entrepreneurial resources, dynamic capabilities and start-up performance of Taiwan's high-tech firms. *Journal of Business Research*, 60(5), pp.549–555. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0148296307000197 [Accessed August 12, 2014].

Xu, J., Chen, J. & Li, B., 2009. Random forest for relational classification with application to terrorist profiling. In *2009 IEEE International Conference on Granular Computing*. IEEE, pp. 630–633. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5255041.

Zahra, S. a., 1993. Environment, corporate entrepreneurship, and financial performance: A taxonomic approach. *Journal of Business Venturing*, 8(4), pp.319–340. Available at: http://linkinghub.elsevier.com/retrieve/pii/088390269390003N.

Zahra, S. a. & Covin, J.G., 1995. Contextual influences on the corporate entrepreneurship-performance relationship: A longitudinal analysis. *Journal of Business Venturing*, 10(1), pp.43–58. Available at: http://linkinghub.elsevier.com/retrieve/pii/088390269400004E.

Zhang, D. & Tsai, J., 2002. Machine learning and software engineering. In *Tools with Artificial Intelligence, 2002. (ICTAI 2002). Proceedings. 14th IEEE International Conference on*. Available at: http://link.springer.com/article/10.1023/A:1023760326768 [Accessed September 28, 2014].

Zitzler, E., Deb, K. & Thiele, L., 2000. Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary computation*, 8(2), pp.173–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10843520.

Zitzler, E., Laumanns, M. & Bleuler, S., 2004. A Tutorial on Evolutionary Multiobjective Optimization. In *Metaheuristics for Multiobjective Optimisation*. Ber: Springer, pp. 3–37. Available at: http://link.springer.com/chapter/10.1007/978-3-642-17144-4_1 [Accessed September 28, 2014].

Appendix A

A1. Classification of SMEs according to the definition of the European Union

Enterprise category	Headcount: Annual Work Unit (AWU)	Annual turnover	or	Annual balance sheet total
Medium-sized	< 250	≤ €50 million (in 1996 € 40 million)	or	≤ €43 million (in 1996 € 27 million)
Small	< 50	≤ €10 million (in 1996 € 7 million)	or	≤ €10 million (in 1996 €5 million)
Micro	< 10	≤ €2 million (previously not defined)	or	≤ €2 million (previously not defined)

A2. The analytical model of this study

	Factor	Variable	Item of measurement	Data type	Domain
Performance	Survival	y_3	Company stilll existent	Binary	{0,1}
	Growth	y_2	Groeth in number of employees in %	Real value	ℝ
	Profitab ility	y_1	RoA in %	Real value	ℝ
Success factors	Control variables	x_{20}	Size	Integer	{0,1,2,3,4,5,6,7}
		x_{19}	Age	Integer	ℕ
		x_{18}	Industry	Integer	ℕ ∈ [0,16]
	Comp. environment	x_{17}	Change in competitors since foundation	Integer	{0,1,2}
		x_{16}	# of competitors at present	Integer	ℕ
	Net-works	x_{15}	Cooperation with universities	Binary	{0,1}
	Strat egy	x_{14}	0 = Price strategy	Integer	{0,1,2}
	Financial resources	x_{13}	Equity capital from VC	Binary	{0,1}
		x_{12}	Equital capital raised	Real value	[0, ∞)
	Technological capabilities	x_{11}	# of licenses used	Integer	ℕ
		x_{10}	# of patents purchased	Integer	ℕ
		x_9	# of own patents used	Integer	ℕ
	Human capital	x_8	startup experience within the founding team	Integer	{0,1,2,3,4}
		x_7	managerial experience within the founding team	Integer	{0,1,2,3,4}
		x_6	technical experience within the founding team	Integer	{0,1,2,3,4}
		x_5	working experience within the founding team	Integer	{0,1,2,3,4}
		x_4	# of founding team members with a degree in economics or social sciences	Integer	ℕ
		x_3	# of founding team members with a degree in a STEM field	Integer	ℕ
		x_2	# of founding team members with an university degree	Integer	ℕ
		x_1	# of members in the founding team	Integer	ℕ

A3. Artificial data generation – The initiation of the independent variables

Variable	Meaning	μ	σ	Adjustments
x_1	# of members in the founding team	4	2	$if(x_1 < 1; 1)$
x_2	# of founding team members with a university degree	$\frac{x_1}{2}$	$\frac{x_1}{4}$	$if(x_2 < 0; 0)$ $if(x_2 > x_1; x_1)$
x_3	# of founding team members with a degree in the STEM field	$\frac{x_2}{2}$	$\frac{x_2}{4}$	$if(x_3 < 1; 0)$ $if(x_3 > x_2; x_2)$
x_4	# of founding team members with a degree in economy or social science	$x_2 - x_3$	$\frac{x_3}{4}$	$if\ (x_3 + x_4 > x_2;$ $x_4 = x_2 - x_3)$
x_5	Working experience within the founding team	2	1	$if(x_5 < 0; 0)$ $if(x_5 > 4; 4)$
x_6	Technical experience within the founding team	2	1	$if(x_6 < 0; 0)$ $if(x_6 > 4; 4)$
x_7	Managerial experience within the founding team	2	1	$if(x_7 < 0; 0)$ $if(x_7 > 4; 4)$
x_8	Startup experience within the founding team	2	1	$if(x_8 < 0; 0)$ $if(x_8 > 4; 4)$
x_9	# of own patents used	$0{,}5$	1	$if(x_9 < 0; 0)$
x_{10}	# of patents purchased	$0{,}5$	1	$if(x_{10} < 0; 0)$
x_{11}	# of licenses purchased	$0{,}5$	1	$if(x_{11} < 0; 0)$
x_{12}	Equity capital raised	20.000	20.000	$if(x_{12} < 0; 0)$
x_{13}	Connection to venture capital	$0{,}5$	$0{,}25$	$if(x_{13} < 0; 0)$ $if(x_{13} > 1; 1)$
x_{14}	Strategy	1	$0{,}5$	$if(x_{14} < 0; 0)$ $if(x_{14} > 2; 2)$
x_{15}	Cooperation with university	$0{,}5$	$0{,}25$	$f(x_{15} < 0; 0)$ $if(x_{15} > 1; 1)$
x_{16}	# of competitors at present	$3{,}5$	$1{,}75$	$f(x_{16} < 0; 0)$ $if(x_{16} > 7; 7)$
x_{17}	Change in competitors since foundation	1	$0{,}5$	$if(x_{17} < 0; 0)$ $if(x_{17} > 2; 2)$