Entrepreneurial Finance:
Emerging Approaches Using Machine Learning and Big Data
Other titles in Foundations and Trends® in Entrepreneurship

*Developments in Strategic Entrepreneurship*
B. Casales Morici and I. Zander
ISBN: 978-1-68083-710-0

*Taxes and Entrepreneurship: A Literature Review and Research Agenda*
Donald Bruce, Tami J. Gurley-Calvez and Alex Norwood

*Returnee Entrepreneurs: A Systematic Literature Review, Thematic Analysis, and Research Agenda*
Jan Henrik Gruenhagen, Per Davidsson and Sukanlaya Sawang
ISBN: 978-1-68083-664-6

*Entrepreneurship as Trust*
Tomasz Mickiewicz and Anna Rebmann
ISBN: 978-1-68083-642-4

*The Evolution of Entrepreneurship as a Scholarly Field*
Hans Landstrom
ISBN: 978-1-68083-626-4

*Financing Entrepreneurship and Innovation in China*
Lin William Cong, Charles M.C. Lee, Yuanyu Qu and Tao Shen
ISBN: 978-1-68083-598-4
Entrepreneurial Finance: Emerging Approaches Using Machine Learning and Big Data

Francesco Ferrati
School of Entrepreneurship (SCENT)
Department of Industrial Engineering
University of Padova
Italy
francesco.ferrati@unipd.it

Moreno Muffatto
School of Entrepreneurship (SCENT)
Department of Industrial Engineering
University of Padova
Italy
moreno.muffatto@unipd.it
Editorial Scope

Topics

Foundations and Trends® in Entrepreneurship publishes survey and tutorial articles in the following topics:

- Nascent and start-up entrepreneurs
- Opportunity recognition
- New venture creation process
- Business formation
- Firm ownership
- Market value and firm growth
- Franchising
- Managerial characteristics and behavior of entrepreneurs
- Strategic alliances and networks
- Government programs and public policy
- Gender and ethnicity
- New business financing:
  - Business angels
  - Bank financing, debt, and trade credit
  - Venture capital and private equity capital
  - Public equity and IPOs
- Family-owned firms
- Management structure, governance and performance
- Corporate entrepreneurship
- High technology:
  - Technology-based new firms
  - High-tech clusters
- Small business and economic growth

Information for Librarians

Foundations and Trends® in Entrepreneurship, 2021, Volume 17, 4 issues. ISSN paper version 1551-3114. ISSN online version 1551-3122. Also available as a combined paper and online subscription.
## Contents

1. **Introduction**  
   3

2. **An Introduction to Machine Learning**  
   8

3. **Data-Driven Venture Capital Firms**  
   13

4. **Crunchbase**  
   16

5. **Scope of the Study**  
   20

6. **Classification of the Research Objectives in Using Machine Learning and Crunchbase**  
   21
   6.1 Predicting the Exit Event of a Company  
       23
   6.2 Predicting Funding Events in a Given Period of Time  
       25
   6.3 Predicting the Next Event That a Company Will Achieve  
       26
   6.4 Predicting Investment Relationship Between Companies and Investors  
       27
   6.5 Generating Investment Recommendations for Investors  
       28
   6.6 Predicting a Company Valuation  
       29
   6.7 Classifying Companies by Industry  
       30
7 Integrating Different Machine Learning Modules to Support Investors' Decision-Making 32

8 Features' Classification 35

9 The Algorithms Used 38

10 Comparing the Performances of Different Models 43

11 Discussion 52

12 Conclusion and Future Research 55

Acknowledgment 57

Appendices 58

Appendix A: Classification of the Features Provided by Crunchbase 59

Appendix B: Classification of the Algorithms 87

Appendix C: Classification of the Performance Metrics 91

References 93
Entrepreneurial Finance: Emerging Approaches Using Machine Learning and Big Data

Francesco Ferrati¹ and Moreno Muffatto²

¹School of Entrepreneurship (SCENT), Department of Industrial Engineering, University of Padova, Italy; francesco.ferrati@unipd.it
²School of Entrepreneurship (SCENT), Department of Industrial Engineering, University of Padova, Italy; moreno.muffatto@unipd.it

ABSTRACT

For equity investors the identification of ventures that most likely will achieve the expected return on investment is an extremely complex task. To select early-stage companies, venture capitalists and business angels traditionally rely on a mix of assessment criteria and their own experience. However, given the high level of risk with new, innovative companies, the number of financially successful startups within an investment portfolio is generally very low. In this context of uncertainty, a data-driven approach to investment decision-making can provide more effective results. Specifically, the application of machine learning techniques can provide equity investors and scholars in entrepreneurial finance with new insights on patterns common to successful startups.

This study presents a comprehensive overview of the applications of machine learning algorithms to the Crunchbase database. We highlight the main research goals that can
be addressed and then we review all the variables and algorithms used for each goal. For each machine learning algorithm, we analyze the respective performance metrics to identify a baseline model. This study aims to be a reference for researchers and practitioners on the use of machine learning as an effective tool to support decision-making processes in equity investments.

**Keywords:** decision making; startup; investments; venture capital; machine learning; Crunchbase.
In recent decades the digital revolution is spreading in all sectors of economic and social life. The increased computational capacity, the improvement of the algorithms as well as the large amounts of available data have given a new boost to artificial intelligence applications (Agrawal et al., 2018). Progress has been so significant that we can claim to have been entered a “second machine age”, driven by artificial intelligence and big data (Brynjolfsson and McAfee, 2014). The digital revolution is radically transforming the way people operate and various areas of management are increasingly benefiting from these profound changes. Intelligent systems are applied in production and management processes, having a strong impact, for example, within decision making activities. In this context of change, entrepreneurship is also inevitably transforming itself, as much as it can be said that artificial intelligence and big data started a new era in entrepreneurship (Obschonka and Audretsch, 2019). While artificial intelligence has received increasing attention in a variety of research and application fields, such as economics, economic policy, innovation and management (Cockburn et al., 2019; George et al., 2014), not so much research has been done in entrepreneurship so far. Interestingly, Obschonka and
Audretsch (2019) place emphasis on the co-evolution and reciprocity of the fields of research and entrepreneurial practice. Smart machines and algorithms will not only inspire and enhance a new generation of research in this field, but will also reshape the current real phenomenon of entrepreneurship. It remains to be understood how artificial intelligence and big data can contribute to a transformation of both research and practice, perhaps giving rise to “smarter forms of entrepreneurship”.

Being a branch of entrepreneurship, entrepreneurial finance (Cumming and Vismara, 2017) is also affected by this transformation. Similarly, artificial intelligence and big data can change entrepreneurial finance research and practice and can let us in an era of “smarter entrepreneurial finance”.

In this context, the main stakeholders in artificial intelligence are equity investors, such as venture capitalists (VCs) and business angels. VCs inject capital into startups in exchange for equity, with the aim of supporting the venture’s growth and increasing its value in order to make a capital gain when they sell their equity stack. In general, to achieve an adequate return on investment (ROI), VCs look for companies that grow exponentially. The high ROI sought by investors is justified by the high risk of failure associated with technology-driven startups. The average ROI per investment is generally very low and VCs need to apply portfolio design strategies so that a small number of companies are able to cover the losses of others and create an overall profit (Metrick and Yasuda, 2011). In order to maximize the chances of generating a profit, VCs should be able to estimate the investment outcome as accurately as possible. The prediction of future events related to a company, such as its ability to make an exit (in the form of a merge and acquisition or an initial public offering) plays an important role in the investment selection process. Prediction is therefore a key element in the decision-making process as the ability to make reliable predictions can reduce uncertainty.

The evaluation of investment proposals, especially for early-stage technology-driven startups, represents a very difficult task. In fact, as in entrepreneurial action (Sarasvathy, 2009), decision-making in entrepreneurial finance is characterised by high level of uncertainty and
information asymmetry. Unlike assessments of small and medium-to-large enterprises, the evaluation of investment opportunities in startup companies cannot be based on economic and financial considerations, because of the lack of historical financial data (Miloud et al., 2012). Traditional evaluation techniques are therefore less suitable for startup companies, given the specificity of this type of venture (Silva, 2004). To improve the financial performance of an investment fund, VCs must implement strategies to improve every phase of the investment process, starting with refining their decision-making process (Zacharakis and Shepherd, 2007). Over the last 30 years, many studies have tried to codify investors’ decision-making into a set of effective criteria (Ferrati and Muffatto, 2019). The variables considered by investors are many and often with subtle nuances. It was also found that there is no consensus as to which variables are more relevant, and that selection methods have so far been based primarily on human judgment and experience often with what is normally referred to as “gut feeling”.

Sarasvathy pointed out that uncertainty depends, to a large extent, on the absence of reliable and complete pre-existing data that would help with respect to entrepreneurial tasks (Sarasvathy, 2001). However, nowadays, large databases are increasingly available, making it possible to apply artificial intelligence techniques to reduce uncertainty to some extent. While having a large amount of data at their disposal, using it to make decisions requires investors to analyze a multitude of variables that can be difficult to consider all together by one person or a group of people. For this purpose, machine learning can provide essential support.

The focus of the present research is on investment decisions in technology-based companies by institutional investors. Our goal is to understand how these new tools can bring new perspectives of analysis so far unexplored in this type of investment. Machine learning is a new tool in entrepreneurial finance for which there are still few studies available at the moment. The aim of this study is to analyze the available works in this emerging field and to highlight the possible questions that machine learning can answer. In fact, the research question that we aim to answer in this study is: “For what purposes have machine learning
tools been applied so far to support equity investors in their decision-making process?” For this purpose, we analyzed the available works on the application of machine learning models in investments in technology-based companies. The product of this research can help researchers to clarify the scope of applications and thus make the use of these instruments more viable. Since the feasibility of using machine learning techniques is closely related to the data that is used, in answering the search question, we have selected only the works related to a specific database, Crunchbase. This work is therefore intended to be an analysis of the academic state of the art of the contributions that have applied machine learning approaches using the data provided by Crunchbase.

The choice of Crunchbase as a reference database is due to its growing popularity in research. Before making this choice, the authors also explored the features of other existing datasets potentially suitable for the application of machine learning methods in entrepreneurial finance. In this regard, some existing alternatives focused on startup companies are for example PitchBook, Dealroom, PrivCo and Venture Scanner. However, as far as the authors know, these companies do not allow a complete free access to data for research purposes. On the other hand, in addition to paid access for commercial purposes, Crunchbase is also accessible for free for academic research, through a dedicated license called “Academic Research Access”. Researchers are eligible for full or discounted access to the Crunchbase dataset on a case-by-case basis. Once the researcher has obtained consent to access and after complying with the terms of use, Crunchbase can provide API that allows full access of their data in the CSV and JSON format. Considering the large amount of data collected and the privileged access for research purposes, many researchers have begun to explore the potential of this database, generally to do research in entrepreneurship and economics area. In order to promote the advancement of this field of research, the authors have therefore considered a data source potentially freely accessible to researchers. Considering other paid data sources would in fact limit the progress of research, making the experiments replicable for a limited number of researchers.

However, since this is still an emerging field in research in entrepreneurial finance (Obschonka and Audretsch, 2019), the identified
works have a pioneering character and this research is proposed as a
detailed but still exploratory study on the subject.

This document is organized as follows. Section 2 provides an intro-
duction to machine learning and outlines the main differences from a
traditional statistical approach. Section 3 provides an overview of the
venture capital firms that have already applied a data-driven approach
to their investment decision-making. Section 4 is an introduction to
Crunchbase, one of the most relevant databases on startup companies
and investors. Section 5 describes the scope of this study, focusing on
research contributions that have applied machine learning techniques
to Crunchbase data. Section 6 classifies the studies’ research goals and
describes the various machine learning approaches. Section 7 describes
an example of how the models proposed by previous studies could
be integrated synergistically into investor decision-making. Section 8
synthesizes all the features or variables used, which are obtained either
directly from Crunchbase or through a features engineering process.
Section 9 analyses the algorithms used. In Section 10 we discuss the
results obtained in previous researches in order to establish a baseline
for future research in this field. Finally, Section 11 presents a final dis-
cussion of the applicability of machine learning as a tool for data-driven
investments, while conclusions and future developments are presented
in Section 12. Appendix A details all the features in Crunchbase, while
Appendix B and Appendix C provide, respectively, all the algorithms
and the types of performance metrics considered in previous works.


