Data Science

Research Questions, Focus and Introductions

# Typical Introduction Structure

* Introduce the topic/establish controversy
* Define key concepts and limiting scope of research
* Move toward the gap in research
* State the gap in research
* State the aims of your study/possibly state your hypothesis and signpost to future chapters

## Activity 1: Read through part of the introduction and identify the different sections (where relevant)

## **Dissertation Title: Can Machine Learning Accurately Predict the Success of a Start-Up?**

**1. INTRODUCTION**

*“Being a data-driven investor is a very well-known concept in the hedge fund industry, but in VC is yet not so popular and a lot of work can be done in this space” (Arroyo et al., 2019).*

**1.1 Motivation**

Start-ups are increasingly ubiquitous in the modern world, with cutting edge technology allowing people to innovate, create, and bring to life the previously unimaginable. The acceleration in the conception of start-ups is particularly meaningful since start-ups can bring great economic and social benefit through technological innovation. Economically speaking, there are multiple ways in which start-ups can provide benefit: by advancing technology, opening new markets, boosting production of goods and services, and increasing employment (Corl, 2019). As a result of these economic effects, a multitude of social benefits are provided.

Technological innovation aside, start-ups are particularly attractive from an investment perspective, for those who wish to make money by investing in the start-up itself. There are numerous types of investors who are interested in start-ups, Venture Capital firms (VCs), angel investors, governments (through grants), crowd funders, and more recently, hedge funds. Of these investors, the most significant is VC which on average invests larger amounts into start-ups than any other investor group (Entis, 2013). The importance of VCs to the start-up ecosystem is evident from the sizeable amount of dollars invested by different VC firms over the past decade.1 Funding to start-ups from VCs has increased from approximately $60B in 2011, to over $300B in 2020 (see Figure 1). This growth includes a 4% year-on-year (YoY) increase between 2019 and 2020 even with the global Covid-19 pandemic, indicating the importance of start-up investment (Teare, 2021).

Graphical user interface, application

Description automatically generated

The benefit of investing in start-ups comes from the scale by which they grow – anybody who has seen the film “The Social Network” is aware of how Facebook grew from a college bedroom at Harvard University, to the college dropout house in Silicon Valley, to the multinational technology giant that is Facebook today. VC firms strive to find start-ups that will scale in the same way that Facebook did, and to subsequently invest in these start-ups as early as possible to maximise their return on investment (ROI). The best example of the massive ROI which can be earned through an early investment in a start-up comes from Sequoia Capital, one of the most prominent and successful VCs. Having initially invested $60M in WhatsApp ($8M in a 2011 series A round and $52M in a 2013 series B round), the acquisition of WhatsApp by Facebook for $22B in 2014 returned $3B to Sequoia, a 4900% increase on their initial investment (CB Insights, 2018). This is one of many examples of the exceptional returns that can be generated through VC investment, and this is what makes it such an attractive asset class for investors with a greater appetite for risk, especially in comparison to investment in the public stock market. On average, it was found that VC funds that have closed since 2004 have performed at least as well as the S&P 500, a widely used benchmark for the public stock markets (Harris et al., 2016).

While the colossal ROI can be extremely tempting to investors, investing in start-ups requires both a huge amount of skill (in finding the right companies and then helping to manage them), and a significant amount of due diligence to decide whether to invest in each start-up. The skill requirement is further accentuated by the significant amount of risk involved with investing in start- ups. To illustrate the risk associated with investing in start-ups, one needs look only to the high failure (bankruptcy) rates. It was found that as of 2019, “start-up failure rates are around 90%. 21.5% of start-ups fail in the first year, 30% in the second year, 50% in the fifth year, and 70% in their 10th year” (National Business Capital and Services, 2020). Accordingly, VC investment returns typically follow the 80-20 rule, or the Pareto principle, in that 80% of the wins come from 20% of the deals (CB Insights, 2018). As such, the vast majority of VC investments must either return very little, none, or even lose the VC money, and therein lies the high risk to reward ratio that VCs operate on.

Ultimately, investing in start-ups is one of the riskiest investments that there is in the financial world of asset classes. This risk is amplified by the inherent difficulty of selling ones shares in a failing start-up due to a lack of liquidity and restrictions due to shareholder agreements.2 In order for VCs to maximise their ROI, it is important to identify which start-ups are least likely to go bankrupt, and which start-ups are most likely to scale in size and reach exit, which is typically done through either an IPO (Initial Public Offering) or a merger/acquisition (M&A). However, “... the tools investors currently have available are not robust enough to reduce risk and help them managing uncertainty better” (Arroyo et al., 2019). The lack of tools available to help identify a “good” investment, and subsequently create a high ROI, is an issue prevalent in many VCs. In fact, a well-known Israeli VC investor stated that “Ninety-five percent of VCs aren’t actually returning enough money to justify the risk, fees, and illiquidity their investors are taking on by investing in their funds” (Dean, 2017).

In this vein, the motivation underlying this study is to mitigate some of the risk involved in the VC investment process by way of a data-driven, machine learning approach. “Being a data-driven investor is a very well-known concept in the hedge fund industry, but in VC is yet not so popular and a lot of work can be done in this space” (Arroyo et al., 2019).

**1.2 Research Questions**

Following the motivation, the desire is to use machine learning to see if it is possible to aid the VC investment process, and subsequently mitigate some of the risk in the high-risk investment space within which VCs operate. In this respect, the research questions which this study will endeavour to answer can be defined. The primary research question and main objective of this study is to answer the question:

*“Can machine learning models be used to accurately predict the success of a start-up?”*

In approaching this question, the word “success” will be defined by whether a start-up has exited. Given this definition, the primary research question will be approached by using binary classification models, as such attempting to predict whether a company exited or went bankrupt using a historic data set. Depending on the models used in approaching the primary research question, it is possible that some secondary information regarding the importance of the variables within the models is ascertained. Therefore, the secondary research question of this study is:

*“Are there any variables which are more important in signalling the success of a start-up than others?”*

The purpose of this secondary research question is twofold. Firstly, to allow for VCs to refine their investment process and look at significant start-up performance indicators which may signal success; and secondly, so that founders and entrepreneur can use important signals as performance benchmarks for their start-ups. In answering these questions, this dissertation aims to contribute to three key areas:

1. To the current literature which has researched the effectiveness of machine learning techniques in predicting the success of a start-up.

2. To VCs who may be interested in incorporating a data-driven, machine learning approach into their investment framework.

3. To entrepreneurs who may be interested in using important features of successful start-ups as benchmarks for their own.

**1.3 Structure**

The following study contains four main sections. The first section (Literature Review) will provide necessary definitions, explain the VC investment process, and discuss previous articles and studies which follow similar themes to this study’s discourse. The second section (Methodology) will discuss the methodology which will be used to answer the research questions in the data analysis and modelling section. The third section (Analysis, Results, & Discussion) will show the data analysis and modelling process outlined in the methodology, and the results found through this process. Furthermore, it will discuss these results and their meanings with respect to the literature. The fourth and final section (Conclusions) will discuss the key findings of the study. Moreover, it will contain a discussion of any limitations found and proceed to give direction to future research considering these limitations and the findings of the study. The references used and the appendix can be found following these four main sections.