

CHAPTER 1

Introduction

Venture capital (VC) is financial capital provided to early-stage, high-potential, high-growth startup companies. VC firms are behind many successful high-tech companies, such as Google, Apple, Microsoft and Alibaba. VC firms fund startup companies with cash in exchange for an equity stake but unlike investors in the public markets, VC firms often take a more active role in managing their investments, providing expertise and advice in both managerial and technical areas. VC firms have two primary roles: as scouts, able to signal the potential of new startups, and as coaches, able to help startups realise that potential [3]. In this way, the experience, skills and networks of the VC firms who invest in a startup directly influence the startups trajectory. The widespread adoption of the Internet and inexpensive, ubiquitous computing has decreased the cost of starting a business and transformed the venture funding landscape – companies require less funding to launch but require more to scale in highly competitive markets [graham2013]. There is now an impetus for VC firms to change the way they do business.

VC firms face the challenge of choosing a few outstanding investments from a sea of hundreds of thousands of potential opportunities. VC firms seek investments in companies that can provide a liquidity event that returns many times their investment value within the time frame of their fund. For startups, a liquidity event (also commonly referred to as an ‘exit’) is typically either an initial public offering (IPO) or an acquisition by a larger competitor. Most VC firms expect their investments to reach a liquidity event within 3–8 years, in accordance with their fund timeframe [CITE]. When compared to traditional investors, this would be considered a long-term investment strategy. However, when compared to the trajectory of most companies, this period is particularly short – not many companies, even successful companies, are capable of maturing at this rate. This makes the VC investment screening process particularly difficult. In addition, there are many potential investment candidates, most of which ultimately fail, and traditional metrics of performance (e.g. cashflow, earnings) often do not exist or are uncertain [12]. Traditionally, investment opportunities are either referred

or identified through technology scans (e.g. Google searches, patent searches). These manual search processes are time-consuming for VC firms.

The VC industry requires better systems and processes to efficiently manage labour-intensive tasks like investment origination and screening. Existing approaches in the literature to predict startup performance have three common limitations: small sample size [1, 9, 8, 10, 14, 2, 13, 7], a focus on early-stage investment [stone2014, 4, 1, 6, 15, 7], and sparse use of features [1, 2, 6, 7, 13, 9]. Although individual studies address some of these limitations, none attempt to synthesise their findings into a standalone study and software design. In addition, there is little evidence that previous research has been translated into systems that are able to assist investors directly. The popularity of online databases like AngelList and CrunchBase, which offer information on startups, investments and investors, is evidence of the VC industrys desire for better, more quantitative methods of assessing startup potential. By 2014, over 1,200 investment organisations (including 624 VC firms) were members of CrunchBase’s Venture Program, mining CrunchBase’s startup data to help inform their investment decisions [11]. There is preliminary evidence that mining these data sources may address previous limitations and make investment origination and screening more efficient and effective [stone2014, 5].

We believe it is now possible to address previous limitations in this domain and produce a VC investment screening system that is efficient, robust and powerful. Our system is based around identifying startup companies that are likely to receive additional funding or have a liquidity event (exit) in a given forecast window. This system can generate statistics and make recommendations that may assist VC firms to efficiently and effectively screen investment candidates. To be useful in this context, the implementation of the system must meet the following criteria:

1. Efficiency. Our system must be more efficient than traditional, manual investment screening. A technique to achieve this is autonomously collecting data from readily-available, online data sources. In this project, we focus primarily on the CrunchBase online database. We test whether this source provides enough observations to provide meaningful statistics. We also test whether we can produce a comprehensive and generalisable feature set, one which would allow investors to complement or replace their data sources over time.
2. Robustness. Our system must be robust enough to be reliable over time and agnostic to specific data sources. The system must provide a generalised, robust solution for investors that does not require significant technical knowledge to use, is not specific to a time-period, and is not reliant on

a single data source. The parameters and features that the system selects should be invariant to time so investors can have reasonable confidence in its predictions.

3. **Predictive Power.** Our system must be consistently accurate at identifying a variety of high-potential investment candidates. The system provides a broad first screening process for investors so it is important that it is highly sensitive (i.e. it excludes very few positive results). The system should be robust to different forecast windows (i.e. exit in three years from now) as VC firms make investment decisions with different periods so they can strategically manage the investment horizons of their funds. Similarly, the system should accurately assess companies at any developmental stage.

The following work is presented in three chapters:

1. **Literature Review.** We review the theoretical background of startup performance and VC investment and evaluate previous attempts at using data mining in this domain. We determine that the VC industry requires better systems to efficiently manage labour-intensive tasks like investment screening. Existing approaches in the literature to tackle similar business problems have three common limitations: small sample size, a focus on early-stage investment, and incomplete use of features. Preliminary evidence suggests that online data sources and machine learning techniques may allow us to address previous limitations and produce an investment screening system that is efficient, robust and powerful.
2. **Design.** We outline the design of our system architecture. Our system uses data from the CrunchBase online database with some supplementation from PatentsView (US Patents Office). We use two datasets collected in September 2016 and April 2017 for training and testing respectively. We develop a classification pipeline using the popular Python-based machine learning library Scikit-learn [pedregosa2011]. Our pipeline includes imputation, feature transformation and scaling, extraction and classification. We optimised these steps using cross-validated randomized hyperparameter search.
3. **Evaluation.** We evaluate our proposed system against three criteria: efficiency, robustness and predictive power. Firstly, we evaluate efficiency by exploring the learning curves of our classification techniques and whether there is sufficient data to produce reliable statistics. Secondly, we evaluate robustness by evaluating our models against multiple reverse-engineered

historical datasets and measuring their variance. Thirdly, we evaluate predictive power by testing different forecast windows outcomes. Finally, we discuss our findings more broadly and their implications for investors and future research into startup investment and performance.

Bibliography

- [1] Ahlers, G. K., Cumming, D., Gunther, C., and Schweizer, D. “Signaling in equity crowdfunding”. In: *Entrepreneurship Theory and Practice* 39.4 (2015), pp. 955–980.
- [2] An, J., Jung, W., and Kim, H.-W. “A Green Flag over Mobile Industry Start-Ups: Human Capital and Past Investors as Investment Signals”. In: *PACIS 2015 Proceedings*. AIS Electronic Library, 2015.
- [3] Baum, J. A. and Silverman, B. S. “Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups”. In: *Journal of Business Venturing* 19.3 (2004), pp. 411–436.
- [4] Beckwith, J. “Predicting Success in Equity Crowdfunding”. Unpublished thesis. Joseph Wharton Research Scholars. Available at http://repository.upenn.edu/joseph_wharton_scholars/25. 2016.
- [5] Bhat, H. and Zaelit, D. “Predicting private company exits using qualitative data”. In: *Advances in Knowledge Discovery and Data Mining*. Ed. by Huang, J., Cao, L., and Srivastava, J. Vol. 6634. Lecture Notes in Computer Science. Berlin: Springer, 2011, pp. 399–410.
- [6] Cheng, M., Sriramulu, A., Muralidhar, S., Loo, B. T., Huang, L., and Loh, P.-L. “Collection, exploration and analysis of crowdfunding social networks”. In: *Proceedings of the Third International Workshop on Exploratory Search in Databases and the Web*. ACM. 2016, pp. 25–30.
- [7] Croce, A., Guerini, M., and Ughetto, E. “Angel Financing and the Performance of High-Tech Start-Ups”. In: *Journal of Small Business Management* (2016).
- [8] Dixon, M. and Chong, J. “A Bayesian approach to ranking private companies based on predictive indicators”. In: *AI Communications* 27.2 (2014), pp. 173–188.
- [9] Gimmon, E. and Levie, J. “Founder’s human capital, external investment, and the survival of new high-technology ventures”. In: *Research Policy* 39.9 (2010), pp. 1214–1226.
- [10] Hoenen, S., Kolympiris, C., Schoenmakers, W., and Kalaitzandonakes, N. “The diminishing signaling value of patents between early rounds of venture capital financing”. In: *Research Policy* 43.6 (2014), pp. 956–989.

- [11] Patil, A. *CrunchBase's Venture Program Members Are Making Startup Data Better Than Ever*. Ed. by Crunchbase.com. <https://info.crunchbase.com/2015/01/crunchbases-venture-program-members-are-making-startup-data-better-than-ever/>. Online; accessed 18 05 2015. Jan. 2015.
- [12] Shane, S. and Cable, D. "Network ties, reputation, and the financing of new ventures". In: *Management Science* 48.3 (2002), pp. 364–381.
- [13] Werth, J. C. and Boert, P. "Co-investment networks of business angels and the performance of their start-up investments". In: *International Journal of Entrepreneurial Venturing* 5.3 (2013), pp. 240–256.
- [14] Yu, Y. and Perotti, V. "Startup Tribes: Social Network Ties that Support Success in New Firms". In: *Proceedings of 21st Americas Conference on Information Systems*. 2015.
- [15] Yuan, H., Lau, R. Y., and Xu, W. "The determinants of crowdfunding success: A semantic text analytics approach". In: *Decision Support Systems* 91 (2016), pp. 67–76.