

CHAPTER 1

Literature Review

In this chapter, we first review the startup investment literature to develop criteria to evaluate our investment screening system. We then turn our focus to determining the best techniques to use to create this system, which we break down into three intercorrelated areas: feature selection, data sources and classification algorithms.

1. **Criteria Selection.** Venture capital firms review many potential investment candidates to shortlist for investment. Traditional screening methods involve referral, networking and Internet search. These screening methods are highly time-consuming and subject to human selection biases. Based on our review, we believe that a superior system can be assessed on the basis of its efficiency, robustness, and predictive power.
2. **Feature Selection.** Venture capital investment is a key driver of startup development but our understanding of factors that influence those investment decisions and the subsequent performance of those investments is incomplete. Based on our review of the literature, we propose a hierarchical framework that includes a variety of features that have been previously indicated to be relevant to this investment screening problem. At a high-level, our framework incorporates determinants of startup potential and signals that influence investment confidence.
3. **Data Sources.** Startup performance is a multi-faceted problem and different data sources provide insights into different actors, relationships and attributes. Our review focuses on new online data sources which have the potential to transform entrepreneurship and venture capital research. Preliminary evidence suggests that the startup databases CrunchBase and AngelList are promising and likely to provide a comprehensive feature set that can form the basis of our system.
4. **Classification Algorithms.** Predicting startup performance is a difficult problem for humans. After all, a high percentage of even VC-backed star-

tups still fail. However, machine learning techniques have been used in other areas of finance (e.g. in the public markets) with some success. We cross-reference the characteristics of our intended dataset with the characteristics of common supervised classification algorithms. Our analyses suggest that we should expect Random Forests, Support Vector Machines and Artificial Neural Networks to be most suitable for our system.

1.1 Criteria Selection

Venture capital financing has lagged behind other forms of high finance (e.g. bond trading, loan applications, insurance) in adopting computational analytics to aid decision-making. Banks are now able to evaluate personal loan requests in minutes while VC firms take far longer to put together deals, sometimes months. While these are markedly different forms of finance (VC has a longer return period, larger investments, higher risk profiles), a more data-informed and analytical approach to venture finance is still foreseeable.

In this section, we provide an introduction into VC firm strategy and review the existing state of the VC investment process. We find that analytical tools are nascent and use of analytics in industry is limited. To date only a small handful of VC firms have publicly declared their use of computational analytical methods in their decision making and investment selection process. We explore why the use of data mining in the VC industry is limited and we develop criteria by which we can judge a VC investment screening system to be successful.

1.1.1 Venture Capital Industry

Early-stage investment is a key driving force of technological innovation and is vitally important to the wider economy, especially in high-growth and technology intensive industries (e.g software, medical and agricultural technologies). Venture capital (VC) is a form of private equity, a medium to long-term form of finance provided in return for an equity stake in potentially high growth companies [16]. Reported US venture capital investments in 2015 totalled US\$60 billion [16].

Figure 1.1 illustrates the typical structure of a Venture Capital Fund. A typical fund is managed by a Venture Capital Firm (legally referred to as a General Partner) consisting of several investment partners. The fund itself (the Limited Partnership) is essentially an investment fund raised from various institutional investors such as pension funds, university endowments and family offices (legally referred to as Limited Partners). Beyond fundraising the main responsibilities

of investment partners (also referred to as General Partners) are sourcing investment opportunities, making investment decisions and taking board membership to assist the management of investee private companies (also referred to as Portfolio Companies).

Figure 1.1:

1.1.2 Venture Capital Firm Strategy

Typically, VC firms are reliant on a small number of high-risk investments to produce outsized returns through successful exit events. A common rule-of-thumb is that given a portfolio of ten startup companies: three will fail entirely, three will remain active but will not be very profitable, returning less than the invested capital, three will be active and profitable, returning just the invested capital and one highly successful startup will provide the investor with a multiple return on all of the investments [stone2014]. In comparison to other traditional investment classes, VC financing is heavily biased towards control at the expense of risk mitigation. Although VC firms tend not to take majority stakes in startups, they exert their influence through significant minority stakes, board membership, their relative seniority to the company's founders, and through leveraging their business networks [9].

Despite VC firms, often significant, influence on the trajectory of their investments, they are still highly selective of the companies that they invest in. Although rarely reported, a small number of studies show VC investment rates vary between 1.5-3.5% of proposals considered [stone2014]. Accordingly, traditional venture finance is a very labour intensive and time consuming process involving extensive due diligence on behalf of the investor [fried1994]. The VC investment process involves several main stages: deal origination, screening, evaluation, structuring (e.g., valuation, term sheets), and post investment activities (e.g., recruiting, financing).

1.1.3 Current Venture Capital Systems

Early-stage investment is characterised by a large number of investment candidates, high degree of uncertainty; a lack of reliable data on company performance (particularly financial performance); and a high time-cost of undertaking due diligence. This makes for a complicated origination and screening process. While

referral from trusted sources (e.g., entrepreneurs, accountants, lawyers, other investors) is often used to screen opportunities, as the cost of starting businesses dramatically decreases investors are faced with an increasingly large number potential businesses and investment opportunities to assess and evaluate. Such a proliferation has led to an information overload problem in venture finance.

Despite evidence that VC firms could benefit from increased use of data mining, it appears few are interested in advanced data analytics. Stone [stone2014] interviewed Fred Wilson of Union Square Ventures who said: “We have not been able to quantify [startup potential]. We havent even tried. Although I am sure someone could do it and they might be very successful with it. To us, the ideal founding team is one supremely talented product oriented founder and one, two, or three strong developers, and nothing else.” Likewise, when asked, Chris Dixon of Andreessen Horowitz said: “Ive seen a few attempts to do it quantitatively but I think those are often flawed because the quantitatively measurable things are either obvious, irrelevant, or suffer from overfitting (finding patterns in the past that dont carry forward in the future”.

Similarly, while recently new software tools have been developed to assist VC firm, there is limited evidence of their adoption.

1.1.4 Proposed Criteria

Based on our review of the VC industry and current VC origination and screening processes, we have developed criteria on which we can evaluate our proposed system.

1. Efficiency. Our system must be more efficient than traditional, manual investment screening by referral and technology scan (e.g. Google search, media, databases). This means that it needs to be able to provide enough information – both observations and features – to be able to meet similar levels of accuracy.
2. Robustness. Our system must be robust enough to be reliable over time. The system must provide a generalised, robust solution for investors that does not require significant technical knowledge to use, and is not overfitted to a specific time-period or data source.
3. Predictive Power. Our system must be consistently accurate at identifying a variety of high-potential investment candidates. 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.

1.2 Feature Selection

Our understanding of the factors that influence venture capital investment decisions and the subsequent performance of those investments is incomplete. We believe a diverse range of features is critical to developing accurate models of startup performance and investment decisions.

Prior work focuses on basic company features (e.g. the headquarters' location, the age of the company) for startup investment predictive models [4, 10]. Semantic text features (e.g. patents, media) [11, 21] and social network features (e.g. co-investment networks) [19, 6, 20] may also predict startup investment. We expect a model that includes semantic text and social network features alongside basic company features could lead to better startup investment prediction.

We propose a conceptual framework that builds upon previous work to ensure that we include a comprehensive and relevant set of features in our investment screening system. Ahlers et al. [1] developed a conceptual framework for funding success on equity crowdfunding platforms. Their framework has two factors: venture quality and level of uncertainty. The first factor is based on work by Baum and Silverman [3] that suggests key determinants of startup potential are human capital, alliance (social) capital, and intellectual (structural) capital. The second factor is based on investors' confidence in their estimation of startup potential.

We seek to generalise Ahlers' framework [1] beyond equity crowdfunding. While the first factor of Ahlers' framework (venture quality) applies to startups of all stages, Ahlers operationalise their second factor with respect to whether startups offer an equity share in their crowdfunding, and whether they provide financial projections. These features are specific to equity crowdfunding. We propose an extension of Ahlers' framework that generalises and develops this second factor. We describe investment confidence as a product of third party validation, historical performance and contextual cues. Our proposed framework is depicted in Figure 1.2.

Next, we must operationalise this conceptual framework into features that we can incorporate into our machine learning model. Table 1.1 shows a review of features tested in previous studies of startup investment. In Appendix A, we describe each of these features and outline theoretical and empirical evidence that justify their inclusion in our conceptual framework.

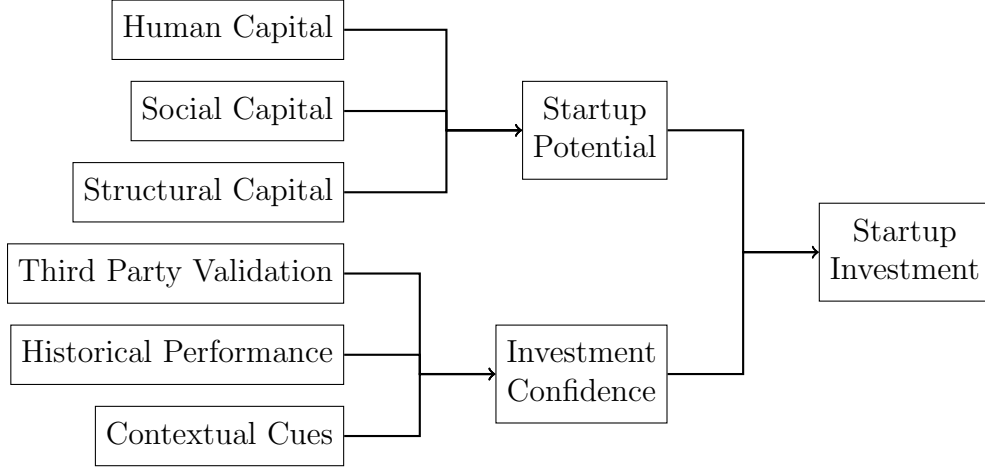


Figure 1.2: Proposed conceptual framework for startup investment. We adapt the framework proposed by Ahlers et al. [1], originally based on work by Baum and Silverman [3]. For an extended version of this framework, please refer to Figure ??.

Feature selection is critical to the success of our proposed conceptual framework. In this section, we have built on the framework proposed by Ahlers et al. [1] in several ways. First, our framework generalises the “Investment Confidence” factor for startups seeking any type of investment (not just equity crowdfunding). Second, our framework has greater depth. Where Ahlers uses one or two features for each factor in their model (e.g. “% Nonexecutive board” represents “Social (alliance) capital”), we perform a review of many features employed in this area and perform a higher degree of classification. For example, in our proposed framework “Social (alliance) capital” is composed of “Social influence” and “Strategic alliances”, each of which will also be composed of several features (e.g. “Twitter followers”, “Average Tweets per day”).

1.3 Data Sources

Predicting startup investment and performance is a complex and difficult task. There are many features that can influence startup investment decisions. Capturing the diversity of these features is critical to developing accurate models. Accordingly, this task will likely involve data collection from multiple data sources. Appropriate selection of these data sources is important because different data sources provide insights into different actors, relationships and attributes.

Previous studies in this field have been limited by data sources restricted in

Features	Results from Studies	
	Significant	Non-Significant
Startup Potential		
Human Capital		
Founders' Capabilities	[4, 2, 10]	[17, 7]
NED Capabilities	[3]	[1, 2]
Staff Capabilities	[4, 2, 7]	[1]
Social Capital		
Strategic Alliances	[3]	-
Social Influence	[4, 2, 6, 20]	-
Structural Capital		
Patent Filings	[11, 12, 3]	[1, 10]
Investment Confidence		
Third Party Validation		
Investment Record	[1, 4, 8, 11, 7]	-
Investor Reputation	[2, 19, 12]	[11]
Media Coverage	[4]	[2]
Awards and Grants	[1]	-
Historical Performance		
Financial Performance	[4, 3]	-
Non-Financial Performance	[2, 10]	[11]
Contextual Cues		
Competitor Performance	[17, 8, 10]	[4, 7]
Broader Economy	[4, 8, 11, 7, 12]	[17, 1]
Local Economy	[17, 4, 8, 10, 11]	-

Table 1.1: Features relevant to startup investment. We review thirteen empirical studies that investigate drivers of startup investment. For each study, we note whether included features have a significant effect on the startup investment model. We classify identified features according to our proposed conceptual framework.

sample size. Most studies have samples of fewer than 500 startups [1, 10], or between 500 and 2,000 startups [11, 20, 2, 19, 8], and only a few have large scale samples (more than 100,000 startups) [17, 6]. Sample size is more critical to model development than the sophistication of machine learning algorithms or feature selection [5]. Startups databases (e.g. CrunchBase) and social networks (e.g. Twitter) offer larger data sets than those previously studied. We expect data collected from these sources will lead to the discovery of additional features

and higher accuracy in startup investment prediction.

In Table 1.2, we outline the characteristics of relevant data sources and how they could contribute to our chosen features. In this section, we describe desirable characteristics of data sources for this task, review potentially relevant data sources, and ultimately determine which data sources are most likely to suit the characteristics of this task.

1.3.1 Source Characteristics

Entrepreneurship research is transforming with the availability of online data sources: databases, websites and social networks. Entrepreneurship studies have historically relied on surveys and interviews for data collection. Measures of human capital (e.g. founders' capabilities), strategic alliances, and financial performance are difficult to capture elsewhere. However, the trade-off for access to these features is that surveys and interviews are time-consuming and costly to implement. While online surveys address some of these issues, it is still difficult to motivate potential participants to contribute. Online data sources like startup databases and social networks are efficient because collecting data is a secondary function of users interacting with these sources. Researchers can also collect data from these sources automatically and at scale. For these reasons, we only consider online data sources for inclusion in this study, specifically crowd-sourced startup databases (e.g. CrunchBase, AngelList), social networks (e.g. Twitter, LinkedIn), government patent databases (e.g. PatentsView) and private company intelligence providers (e.g. PrivCo). In Appendix B we review the characteristics of each of these data sources commonly used in entrepreneurship research.

1.3.2 Source Evaluation

Entrepreneurship and venture capital research is primed to take advantage of the availability of new online data sources. We evaluated relevant data sources for their suitability to predicting startup investment. Startup databases CrunchBase and AngelList provide the most comprehensive set of features. There are small differences between the features recorded by each. CrunchBase has slightly more coverage and tracks media better but lacks AngelList's social network. At least one startup database should be used and either are satisfactory. Of the other data sources we review, PatentsView is the most promising. PatentsView provides comprehensive patent information, though it could prove difficult matching identities to other sources. Other data sources are less promising because of access issues. LinkedIn cannot be easily collected now the API is deprecated.

Properties	Startup Databases			Social Media		Other Sources	
	CrunchBase	AngelList	LinkedIn	Twitter	PatentsView	PrivCo	
Features							
Startup Potential							
Human Capital							
Founders' Capabilities	✓	✓	✓✓	✗	✗	✗	
NED Capabilities	✓	✓	✓✓	✗	✗	✗	
Staff Capabilities	✓	✓	✓✓	✗	✗	✗	
Social Capital							
Social Influence	✓	✓✓	✓✓	✓✓	✗	✗	
Strategic Alliances	✓	✓	✗	✗	✓	✗	
Structural Capital							
Patent Filings	✗	✗	✗	✗	✓✓	✗	
Investment Confidence							
Third Party Validation							
Investment Record	✓✓	✓✓	✗	✗	✗	✓	
Investor Reputation	✓	✓✓	✓	✗	✗	✗	
Media Coverage	✓✓	✓	✗	✓	✗	✗	
Awards and Grants	✓	✗	✗	✗	✗	✗	
Historical Performance							
Financial Performance	✗	✗	✗	✗	✗	✓✓	
Non-Financial Performance	✓✓	✓✓	✓	✗	✗	✓	
Contextual Cues							
Competitor Performance	✓	✓	✗	✗	✗	✗	
Broader Economy	✓	✓	✗	✗	✗	✗	
Local Economy	✓	✓	✗	✗	✗	✗	
Ease of Use							
Cost Effective	✓	✓✓	✓	✗	✓✓	✗	
Time Efficient	✓✓	✓✓	✗	✓✓	✓✓	✗	
Accurate Data	✓	✓	✓✓	✓✓	✓✓	✓✓	
Large Data Set	✓✓	✓✓	✓✓	✓✓	✓✓	✓	

Table 1.2: Data sources relevant to startup investment. We review six data sources commonly used in entrepreneurship research for their suitability for our startup investment task. We evaluate data sources for their ability to provide relevant features for our analyses and for their ease of use in data collection. We exclude offline sources from our analyses. Ratings are: ✗ = poor, ✓ = satisfactory, ✓✓ = good.

Twitter provides social network topology and basic profile information through its free API but does not provide access to historical tweets. Financial reports are too expensive for the purposes of this study.

1.4 Classification Algorithms

Predicting startup performance is a difficult problem for humans. Computational analytics have been heavily deployed in high finance and we believe there is scope for applying related techniques to improve upon investment decision making in the domain of venture finance. Machine learning is characterised by algorithms that improve their ability to reason about a given phenomenon given greater observation and/or interaction with said phenomenon. Mitchell provides a formal definition of machine learning in operational terms: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .” [15].

Machine learning algorithms can be classified based on the nature of the feedback available to them: supervised learning, where the algorithm is given example inputs and desired outputs; unsupervised learning, where no labels are provided and the algorithm must find structure in its input; and reinforcement learning, where the algorithm interacts with a dynamic environment to perform a certain goal. These algorithms can be further categorised by desired output: classification, supervised learning that divides inputs into two or more classes; regression, supervised learning that maps inputs to a continuous output space; and clustering, unsupervised learning that divides inputs into two or more classes.

We evaluated common machine learning algorithms with respect to their suitability for predicting startup investment. In Table 1.3, we rank these algorithms by cross-referencing their assumptions and properties with the task characteristics. In the following sections, we describe the characteristics of the startup investment prediction task, review common machine learning algorithms, and determine which algorithms are most likely to suit the characteristics of this task.

1.4.1 Task Characteristics

Machine learning tasks are diverse. Our investigation into startup investment is a task that suits supervised machine learning algorithms. We will manipulate the data we collect into a single labelled data set. Startups will be labelled

Criteria	Machine Learning Algorithms						
	NB	LR	KNN	DT	RF	SVM	ANN
Data Set Properties	2	4	6	2	1	5	7
Missing Values	✓✓ [13]	✓ -	✗ [13]	✓✓ [13]	✓✓ [18]	✓ [13]	✗ [13]
Mixed Feature Types	✓✓ [13]	✓✓ -	✓✓ [13]	✓✓ [13]	✓ [18]	✓ [13]	✓ [13]
Irrelevant Features	✗ [13]	✗ [14]	✓ [13]	✓✓ [13]	✓✓ [18]	✗ [13]	✗ [13]
Imbalanced Classes	✓✓ -	✓✓ -	✗ -	✗ [13]	✓ [18]	✓✓ [13]	✓ [13]
Algorithm Properties	2	1	5	5	2	5	2
Predictive Power	✗ [5]	✓ [5]	✓ [5]	✗ [13]	✓✓ [5]	✓✓ [5]	✗✓ [5]
Interpretability	✓✓ [13]	✓✓ [14]	✗ [13]	✓✓ [13]	✓ [14]	✗ [13]	✗ [13]
Incremental Learning	✓✓ [13]	✓✓ -	✓✓ [13]	✓ [13]	✓ -	✓ [13]	✓✓ [13]
Overall	2	2	6	4	1	5	6

Table 1.3: Evaluation of machine learning algorithms for startup investment prediction. We review seven common supervised machine learning algorithms for their suitability for our startup investment task. We evaluate algorithms for their robustness to the structure of the data set and their appropriateness for the constraints of our implementation. We rank the algorithms according to the sum of these measures (in each section and overall) and bold highly-ranked algorithms. Ratings are: ✗ = poor, ✓ = satisfactory, ✓✓ = good. Algorithms are: NB = Naive Bayes, LR = Logistic Regression, KNN = K-Nearest Neighbours, DT = Decision Trees, RF = Random Forests, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

based on whether they are acquired or have had an IPO at a later time. The key objective of machine learning algorithm selection is to find algorithms that make assumptions consistent with the structure of the problem (e.g. tolerance to missing values, mixed feature types, imbalanced classes) and suit the constraints of the desired solution (e.g. time available, incremental learning, interpretability). In the following sections, we outline the characteristics of supervised learning tasks relevant to our startup investment prediction task.

1.4.1.1 Data Set Properties

While data sets can be pre-processed to assist with their standardisation, some types of data sets are still better addressed by particular algorithms. Data set properties like missing data, irrelevant features, and imbalanced classes all have an effect on classification algorithms. Data sets often have missing values, where no data is stored for a feature of an observation. Missing data can occur because of non-response or due to errors in data collection or processing. Missing data has different effects depending on its distribution through the data set. Public data sets, like startup databases and social networks, are typically sparse with missing entries despite their scale. Therefore, robustness to missing values is a desirable property of our algorithm. Despite efforts to only include features that have theoretical relevance, machine learning tasks often include irrelevant features. Irrelevant features have no underlying relationship with classification. Depending on how they are handled they may affect classification or slow the algorithm. We expect irrelevant and non-orthogonal features in our data set because our proposed framework includes features that have not been thoroughly tested in the literature. Therefore, robustness to irrelevant features is a desirable property of our algorithm. Data sets are not usually restricted to containing equal proportions of different classes. Significantly imbalanced classes are problematic for some classifiers. In the worst case, a learning algorithm could simply classify every example as the majority class. Our data set is not dramatically imbalanced overall, but when looking at funding status for different funding rounds it is significantly imbalanced. Therefore, robustness to imbalanced classes is a desirable property of our algorithm.

1.4.1.2 Algorithm Properties

The desired properties of machine learning algorithms are related to the business problems that are being addressed. Predictive power, interpretability and processing speed are all desirable characteristics but involve trade-offs and must be prioritised. Predictive power is the ability of a machine learning algorithm to

correctly classify new observations. Predictive power can be evaluated in many ways. As our data set is likely to have an imbalanced class distribution, we will evaluate predictive power based on balanced metrics like Area under the Receiver-Operator Curve and the F1 Score. If a model has no predictive power, the model is not representing the underlying process being studied. For this reason, predictive power is a desirable property of our algorithm. However, if multiple algorithms provide similar predictive power other selection criteria become significant. Interpretability is the extent to which the reasoning of a model can be communicated to the end-user. There is a trade-off between model complexity and model interpretability. Some models are a “black box” in the sense that data comes in and out but the model cannot be interpreted. For this study, it is a key objective that we improve our understanding of the determinants of startup investment. Therefore, interpretability is a desirable property of our algorithm. Finally, processing speed is another desirable property, especially when handling real-time data or when there is a need to run exploratory analyses on the fly. In this case, processing speed is not critical because generally VC investment decisions are made over weeks and months, though there is some need for the data set to be updated with new information as it becomes available.

1.4.2 Algorithm Characteristics

Supervised machine learning are algorithms that reason about observations to produce general hypotheses that can be used to make predictions about future observations. Supervised machine learning algorithms are diverse, from symbolic (Decision Trees, Random Forests) to statistical (Logistic Regression, Naive Bayes, Support Vector Machines), instance-based (K-Nearest Neighbours), and perceptron-based (Artificial Neural Networks). In Appendix C, we describe each candidate learning algorithm, critique their advantages and disadvantages, and present evidence of their effectiveness in applications relevant to startup investment.

1.4.3 Algorithm Evaluation

We evaluated supervised learning algorithms for their suitability in startup investment prediction. While our evaluation gives us directionality of fit, we hesitate to discard algorithms based on our literature review. Algorithm selection is complex and preliminary testing will provide clarity as to which algorithms should be used. In addition, larger training sets and good feature design tend to outweigh algorithm selection [5]. With those concessions in mind, our findings suggest we

expect Random Forests, Support Vector Machines and Artificial Neural Networks to produce the highest classification accuracies. An ensemble of these algorithms may improve accuracy further, though at the cost of computational speed and interpretability. We may expect Random Forests to outperform the other two algorithms due to robustness to missing values and irrelevant features and native handling of discrete and categorical data. However, Random Forests are not highly interpretable so Decision Trees and Logistic Regression may be preferable for exploratory analysis of the data set.

1.5 Research Gap

The venture capital industry requires better systems and processes to efficiently manage labour-intensive tasks like investment screening. Existing approaches in the literature to predict startup performance have three common limitations: small sample size, a focus on very early stage investment, and incomplete use of features. In addition, there is little evidence that previous research has been translated into systems that are able to assist investors directly. We conducted a literature review to determine how to address these limitations and produce a system that will assist venture capitalists in originating and screening investment candidates.

Firstly, we reviewed the business problem and developed three criteria that will help us evaluate our system: efficiency, robustness and predictive power. Secondly, we developed a conceptual framework of predicting startup performance that incorporates determinants of startup potential and signals that influence investment confidence. This framework informs our feature selection. We then assessed potential data sources and found preliminary evidence that suggests that the startup databases CrunchBase and AngelList are promising and likely to provide a comprehensive feature set that can form the basis of our system. Finally, we reviewed supervised machine learning techniques applied to startup investment and other areas of finance. Our analyses suggested that we should expect Random Forests, Support Vector Machines and Artificial Neural Networks to be most suitable for our system.

Based on this literature review, we believe it is now possible to address previous limitations in this domain and produce an investment screening system that is efficient, robust and powerful. In the next chapter, we will outline the process by which we attempt to develop that system.

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] Caruana, R., Karampatziakis, N., and Yessenalina, A. “An empirical evaluation of supervised learning in high dimensions”. In: *Proceedings of the 25th International Conference on Machine learning*. ACM. 2008, pp. 96–103.
- [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] Conti, A., Thursby, M., and Rothaermel, F. T. “Show Me the Right Stuff: Signals for High-Tech Startups”. In: *Journal of Economics & Management Strategy* 22.2 (2013), pp. 341–364.
- [8] Croce, A., Guerini, M., and Ughetto, E. “Angel Financing and the Performance of High-Tech Start-Ups”. In: *Journal of Small Business Management* (2016).
- [9] Fried, J. M. and Ganor, M. “Agency costs of venture capitalist control in startups”. In: *New York University Law Review* 81 (2006), p. 967.
- [10] 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.

- [11] 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.
- [12] Hsu, D. H. and Ziedonis, R. H. “Patents As Quality Signals For Entrepreneurial Ventures.” In: *Academy of Management Proceedings*. Vol. 2008. 1. Academy of Management. 2008, pp. 1–6.
- [13] Kotsiantis, S. “Supervised Machine Learning: A Review of Classification Techniques”. In: *Informatica* 31.3 (2007).
- [14] Kuhn, M. and Johnson, K. *Applied predictive modeling*. Springer, 2013.
- [15] Mitchell, T. M. *Machine Learning*. New York: McGraw-Hill, 1997.
- [16] National Venture Capital Association. *2016 National Venture Capital Association Yearbook*. <http://www.nvca.org/?ddownload=2963>. Online; accessed 06 Nov 2016. Mar. 2016.
- [17] Shan, Z., Cao, H., and Lin, Q. “Capital Crunch: Predicting Investments in Tech Companies”. Unpublished thesis. Stanford University. Available at <http://www.zifeishan.org/files/capital-crunch.pdf>. 2014.
- [18] Strobl, C., Malley, J., and Tutz, G. “An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.” In: *Psychological Methods* 14.4 (2009), p. 323.
- [19] 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.
- [20] 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.
- [21] 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.