

## CHAPTER 1

# Conclusions

Our project’s aim was to produce a Venture Capital (VC) investment screening system that is efficient, robust and powerful. Our system uses online data sources and machine learning to identify startups that are likely to receive additional funding or exit in a given forecast window. While this is a challenging problem even for experienced VC investors, we achieved classification results that have practical application for VC firms.

### 1.1 Evaluation of Criteria

#### 1.1.1 Efficiency

Our system intends to replace manual investment screening (referral, Google search, industry papers, and manual search of startup databases). Our automated system should be more efficient than these methods because it requires less human input. We also evaluated our system’s efficiency based on dataset size and time profile. Our system is relatively robust to dataset size, because it optimises the classification pipeline based on the datasets available. In some cases, our system could use smaller training sets without significant reduction in predictive power. However, our system’s ability to use smaller training sets was related to the length of the desired forecast window, and the breadth of the target outcome. An indicative implementation of our system takes 46 hours to run, which is reasonable for a process that is not time-sensitive in this industry. The majority of this time is due to the pipeline optimisation process. However, when placed into production, this could be run less frequently with minimal reduction in performance.

### 1.1.2 Robustness

Our system must be robust in its performance with respect to time so investors can rely on its predictions. The Venture Capital (VC) industry is concerned that predictive models trained on historical data will not accurately predict future trends and activity. This has been identified as a key barrier to the adoption of automated systems by the VC industry [1]. We evaluated our system across a number of historical datasets, forecast windows, and even multiple evaluation metrics. We found variance across all evaluation metrics to be very low, with slightly more variance over shorter forecast windows. When we explored the feature weights for each model developed on different historical datasets, we found only slight variance. This suggests that our system produces highly robust models, suitable for forward-looking investment screening.

### 1.1.3 Predictive Power

Our system must be accurate at identifying a variety of high-potential investment candidates. We evaluated the systems' predictive power based on its ability to predict over different forecast windows (e.g. 2-4 years), for target companies at different developmental stages (e.g. Seed, Series A etc.), and for different target outcomes (e.g. predicting additional funding rounds). Forecast window has an impact on our system's performance. Our system produced F1 Scores of 0.36, 0.48 and 0.55 for forecast windows of 2, 3 and 4 years. In the future, it would be interesting to also explore forecast windows that are closer in duration to a typical VC investment horizon (5-8 years). Our system's performance has a positive relationship with developmental stage, producing F1 Scores ranging from 0.33 for Pre-Seed companies through to 0.62 for Series C companies. When we fit models separately to each stage, we see a performance improvement (+0.03), and particularly for Series D+ companies (0.53 to 0.66). Finally, our system varies in its performance at predicting target outcomes, producing F1 Scores ranging from 0.51 for predicting additional funding through to 0.24 for predicting an IPO.

## 1.2 Future Research

### 1.2.1 Network Analysis

### 1.2.2 Temporal Analysis

### 1.2.3 Semantic Text Analysis

### 1.2.4 Systems Integration

We developed a connector to CrunchBases API that provided real-time and comprehensive access to their database. Although we abandoned this approach because of the time constraints of this research project, it deserves further investigation. In terms of producing a system that is practical for Venture Capital (VC) firms to use day-to-day, continuous data integration is an important feature. The dataset could be continuously updated and incrementally fed into the classification pipeline to a) ensure that the dataset is up-to-date and b) provide more granular temporal analysis of trends in the dataset.

## 1.3 Summary

We produced a Venture Capital (VC) investment screening system that is efficient, robust and predictive. 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 translated into systems that assist investors directly. This project addressed these issues and lays the groundwork for an industry-ready VC investment screening tool.

# Bibliography

- [1] Stone, T. R. “Computational analytics for venture finance”. Unpublished Ph.D. dissertation. UCL (University College London), 2014.