Towards Automated Venture Capital Screening

THE UNIVERSITY OF WESTERN AUSTRALIA

Author: Mark Shelton (21151978) **Discipline:** Computer Science and Software Engineering **Supervisors:** Melinda Hodkiewicz, Tim French

Introduction

Venture Capital (VC) firms invest in high-potential, high-growth technology companies (startups). VC firms face the challenge of identifying a few outstanding investments from a sea of opportunities. Previous approaches to help VC firms efficiently screen investment opportunities have common limitations: small datasets [1, 2], a focus on early-stage investment [1, 2, 3, 4, 5], and narrow feature sets [3, 6].

Aims & Criteria

We sought to develop a VC screening system that is:

- **Practical:** More efficient than status quo (referrals, search) and processes new data in a timely fashion.
- **Robust:** Reliably makes predictions based on historical data without over-fitting to short-term trends.
- **Versatile:** Consistently identifies successful investments across a large domain of possible tasks.

System Design

We developed a VC investment screening system that applies machine learning to data from two large public online data sources, CrunchBase and PatentsView (Figure 1).

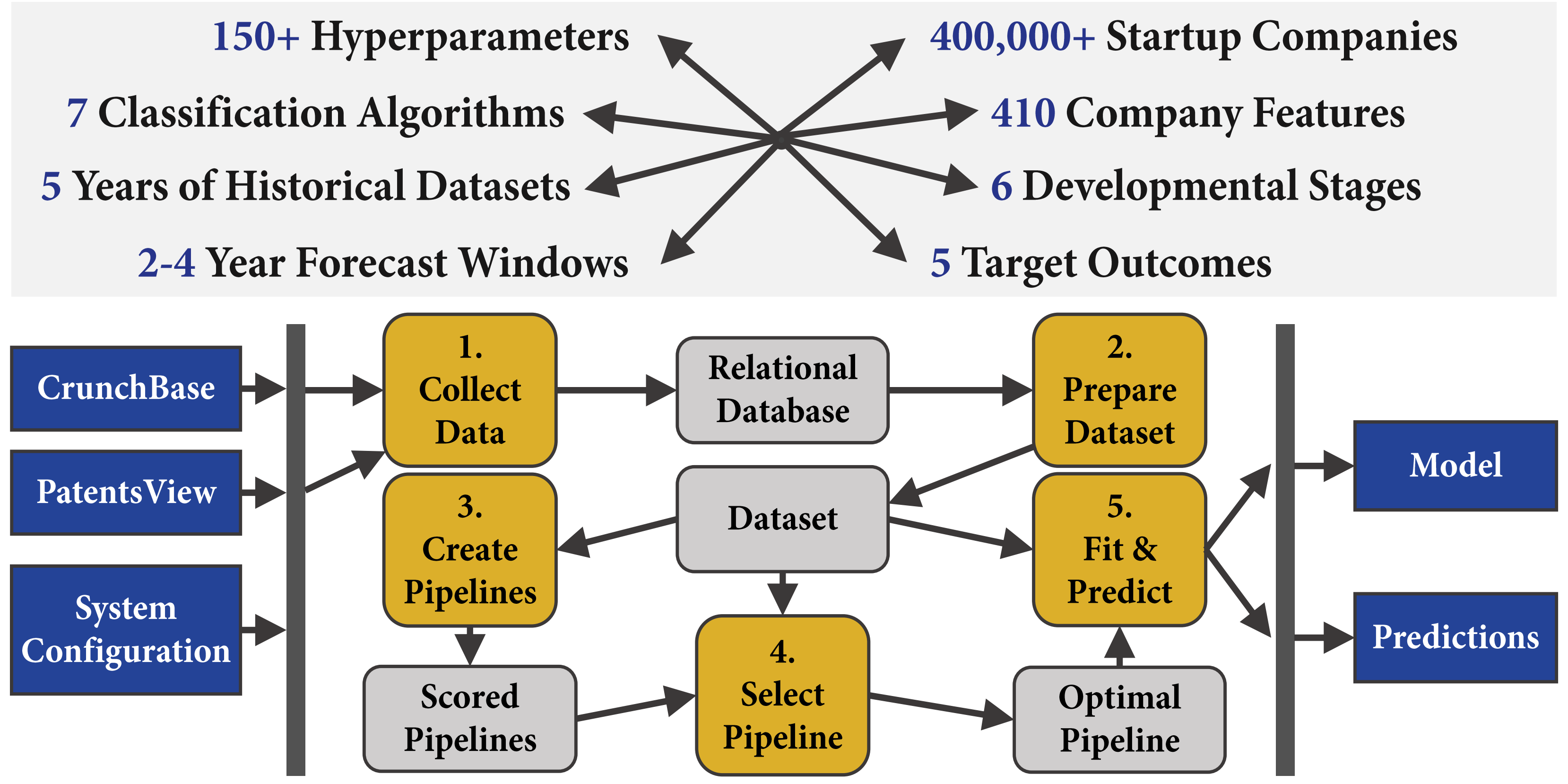

Figure 1. *System characteristics and architecture.*

The core of our system is a two-stage optimisation process that generates a supervised classification pipeline. The classification pipeline is composed of an imputer, transformer, scaler, extractor (PCA), and classifier. Classifier tuning had the greatest effect on performance. Random Forest and Logistic Regression were the best performing classifiers.

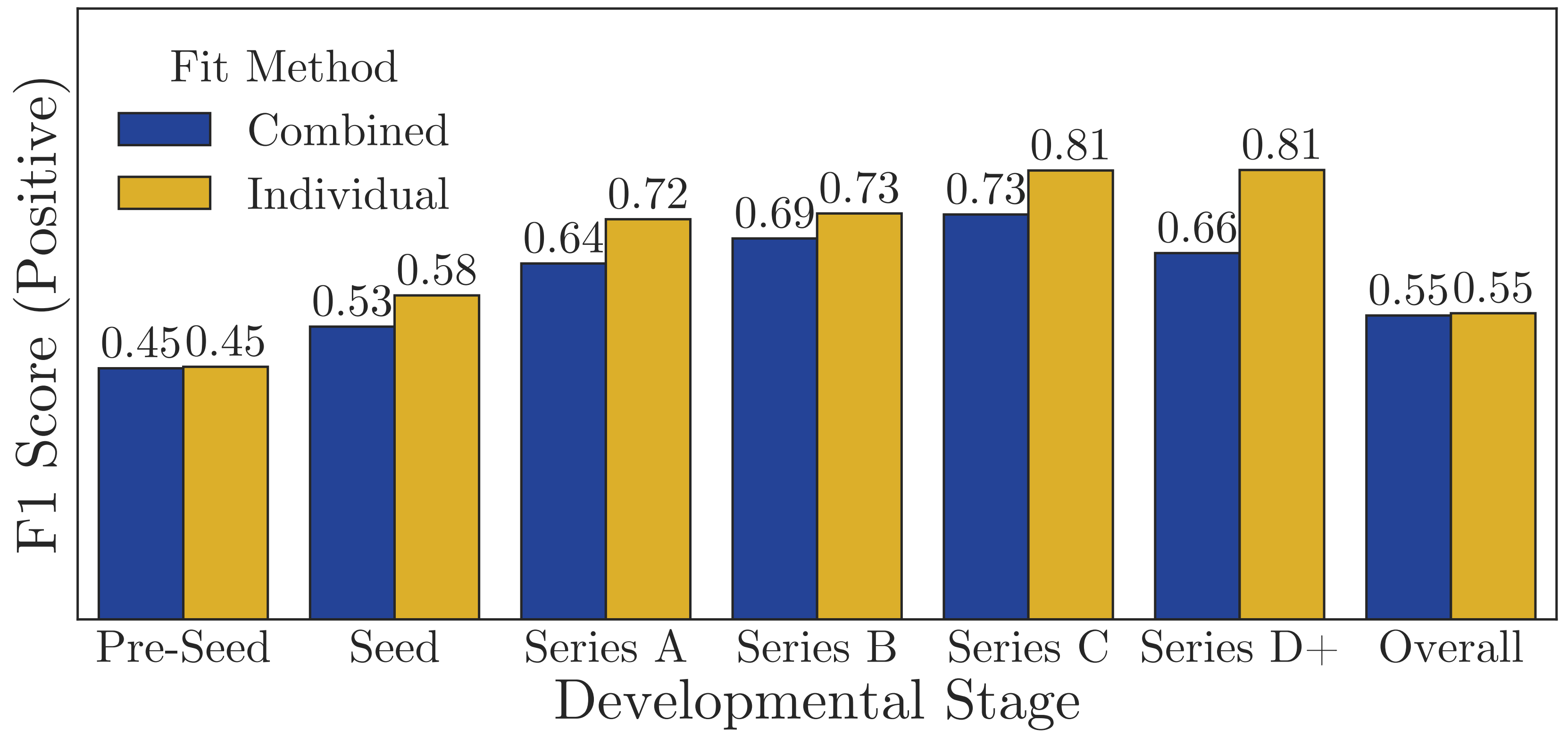

Figure 2. *System performance by startup developmental stage and fit method. Forecast window of 4 years.*

System Performance

Our system's performance is better or comparable to previous results from the literature [4, 5, 6]. We evaluated our system using F1 Score, a measure that provides a balanced assessment of Precision and Recall on a scale of 0 - 1. Evaluation of the positive class (startups that achieved the target outcome) is most critical in screening. The system performed best over longer forecast windows (up to 4 years) and later developmental stages (e.g Series C). Training on individual stages led to better prediction of later stages (Figure 2). Our system's optimisation process suggests we can expect its performance to improve as its data sources grow.

Model Evaluation

Previous work in the literature has studied relatively narrow feature sets [3, 6]. We developed a 14-factor conceptual framework for VC investment based on work by Ahlers et al. (2015) [1]. We evaluated 11 factors from the framework.

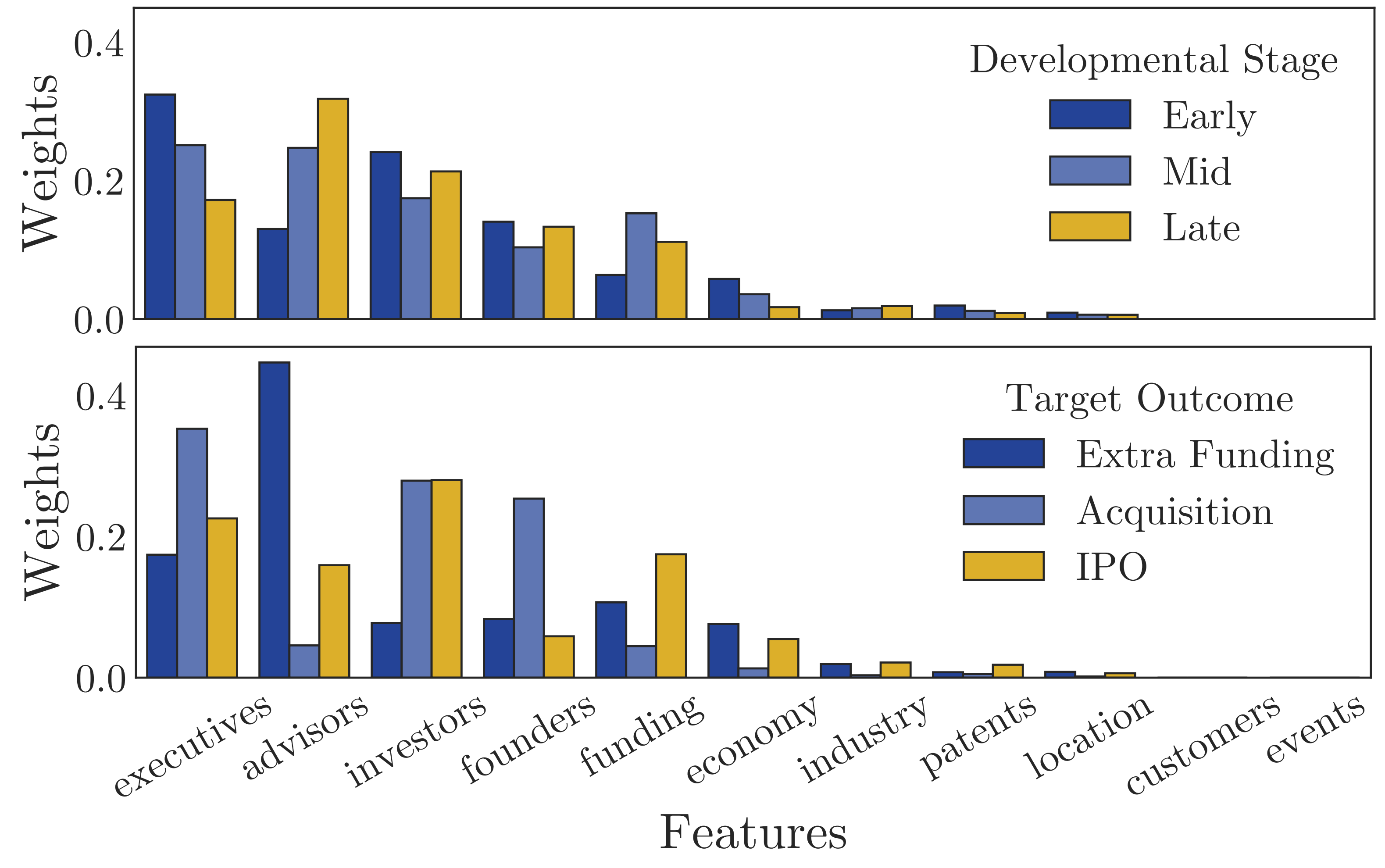

Figure 3. *Normalised feature weights by startup developmental stage (grouped) and target outcome.*

Models generated by our system were robust to time (2012-16) and forecast window (2-4 years), and varied with developmental stage and target outcome (Figure 3). The skills and experiences of a startup's founders, executives and advisors were the strongest predictors of startup performance.

Evaluation of Criteria

☑ **Practical?** Near-autonomous system, adaptable to different datasets, running time of 46 hours.

☑ **Robust?** Minimal variance in performance with respect to training on datasets from different dates.

☑ **Versatile?** Better or comparable results to previous studies across a variety of forecast windows, developmental stages, and target outcomes.

Conclusions

This project makes three primary contributions:

1. A system ready for industry that is near-autonomous, adaptable, and leverages public data.
2. A system with robust performance across a large domain of investment prediction tasks.
3. An empirical study of startup investment more comprehensive than any from the literature.

References: [1] Ahlers, G. K., et al. Signaling in equity crowdfunding. Entrepreneurship Theory and Practice 39, 4 (2015), pp. 955-980. [2] An, J., et al. A green flag over mobile industry start-ups: Human capital and past investors as investment signals. In Pacific Asia Conference on Information Systems. AIS Electronic Library, 2015, p. 67. [3] Yuan, H., et al. The determinants of crowdfunding success: A semantic text analytics approach. Decision Support Systems 91 (2016), pp. 67-76. [4] Beckwith, J. Predicting Success in Equity Crowdfunding. Unpublished thesis. University of Pennsylvania. 2016. [5] Stone, T. R. Computational analytics for venture finance. Unpublished thesis. University College London, 2014. [6] Bhat, H. S., and Zaelit, D. Predicting private company exits using qualitative data. In Pacific Asia Conference on Knowledge Discovery and Data Mining. Springer. 2011, pp. 399-410.