

# Introduction

Venture Capital (VC) firms face the challenge of identifying a few outstanding investments from a sea of opportunities. The VC industry requires better systems to manage labour-intensive tasks like investment screening. Previous approaches have common limitations: small, private datasets, a focus on early-stage investment, and limited feature sets.

## Aims & Criteria

We aimed to develop a VC screening system that is:

- **Efficient:** Less effort than the status quo (referrals, search) and able to run in reasonable time.
- **Robust:** Reliably makes forward-looking predictions based on historical data.
- **Powerful:** Consistently identifies successful investments across a large domain of parameters.

## System Design

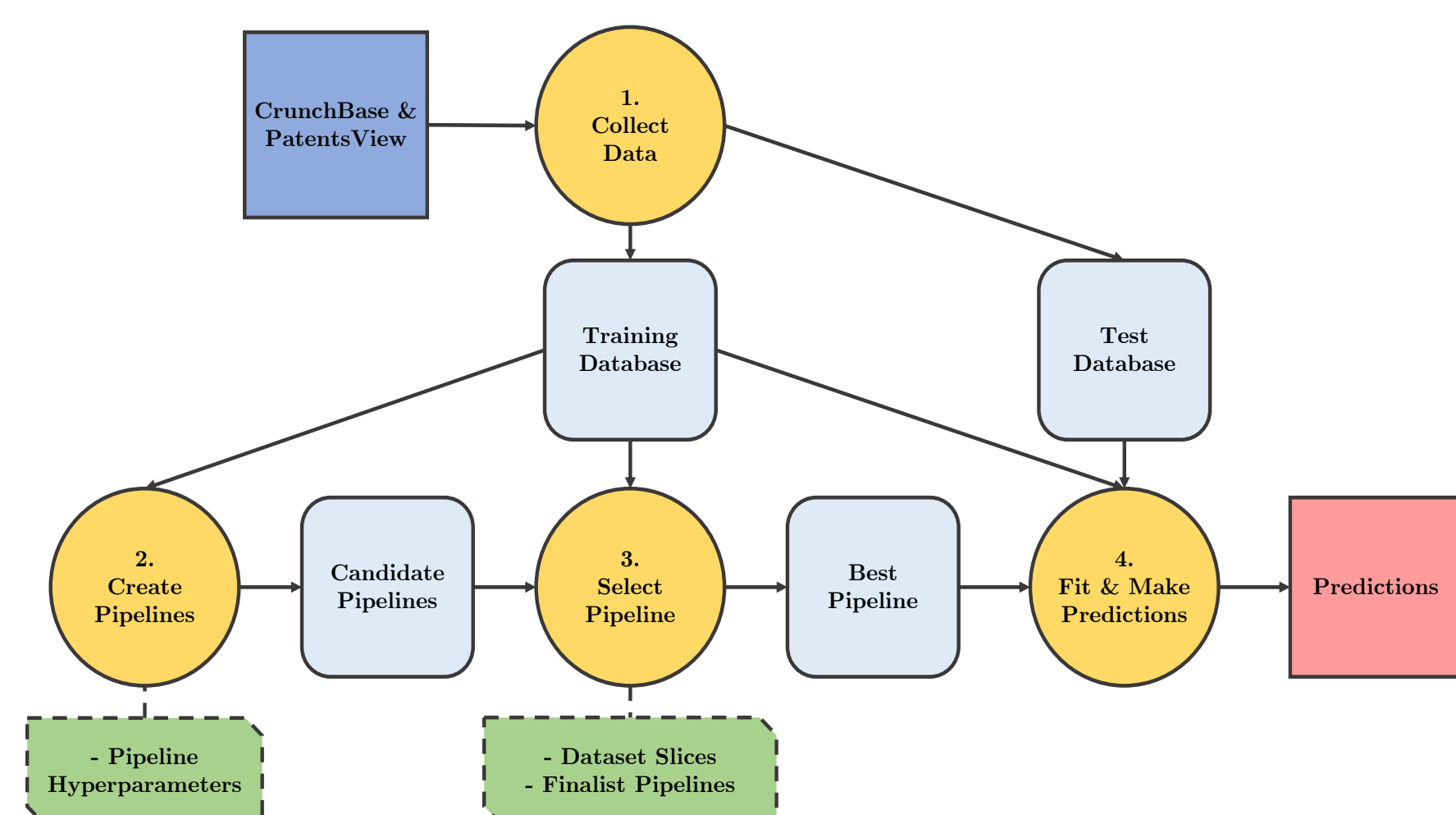


Figure 1. System architecture flowchart.

We developed a multi-stage VC investment screening system (Fig 1). This system is near-autonomous and adapts as the datasets change. The core of the system is an optimisation process that generates a supervised learning classification pipeline. This pipeline is applied to data from large public online data sources (CrunchBase and PatentsView).

## System Performance

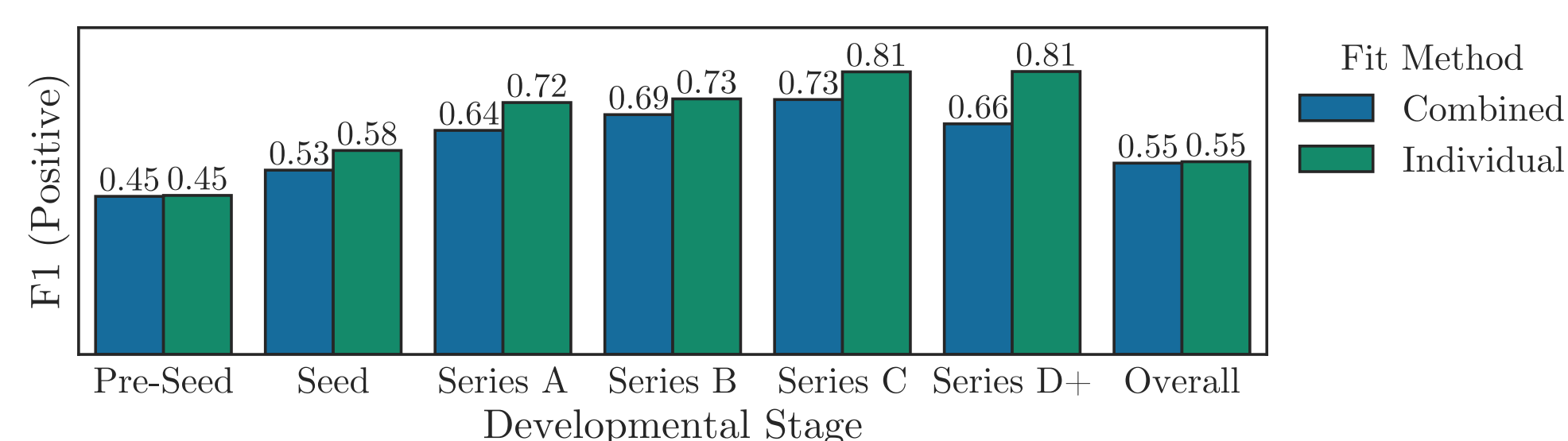


Figure 2. System performance by developmental stage and fit method (forecast window of 4 years).

Our system's performance is better or comparable to previous results from the literature. Performance was related to longer forecast window (2-4 years) and later developmental stage (e.g Series C). Training on individual stages led to improved prediction of later stages (Fig 2).

## Model Evaluation

We developed and evaluated a conceptual framework for startup investment. Models generated by our system are robust to time (for 2012-16) and forecast window (over 2-4 years), and vary with developmental stage and target outcome (Fig 3).

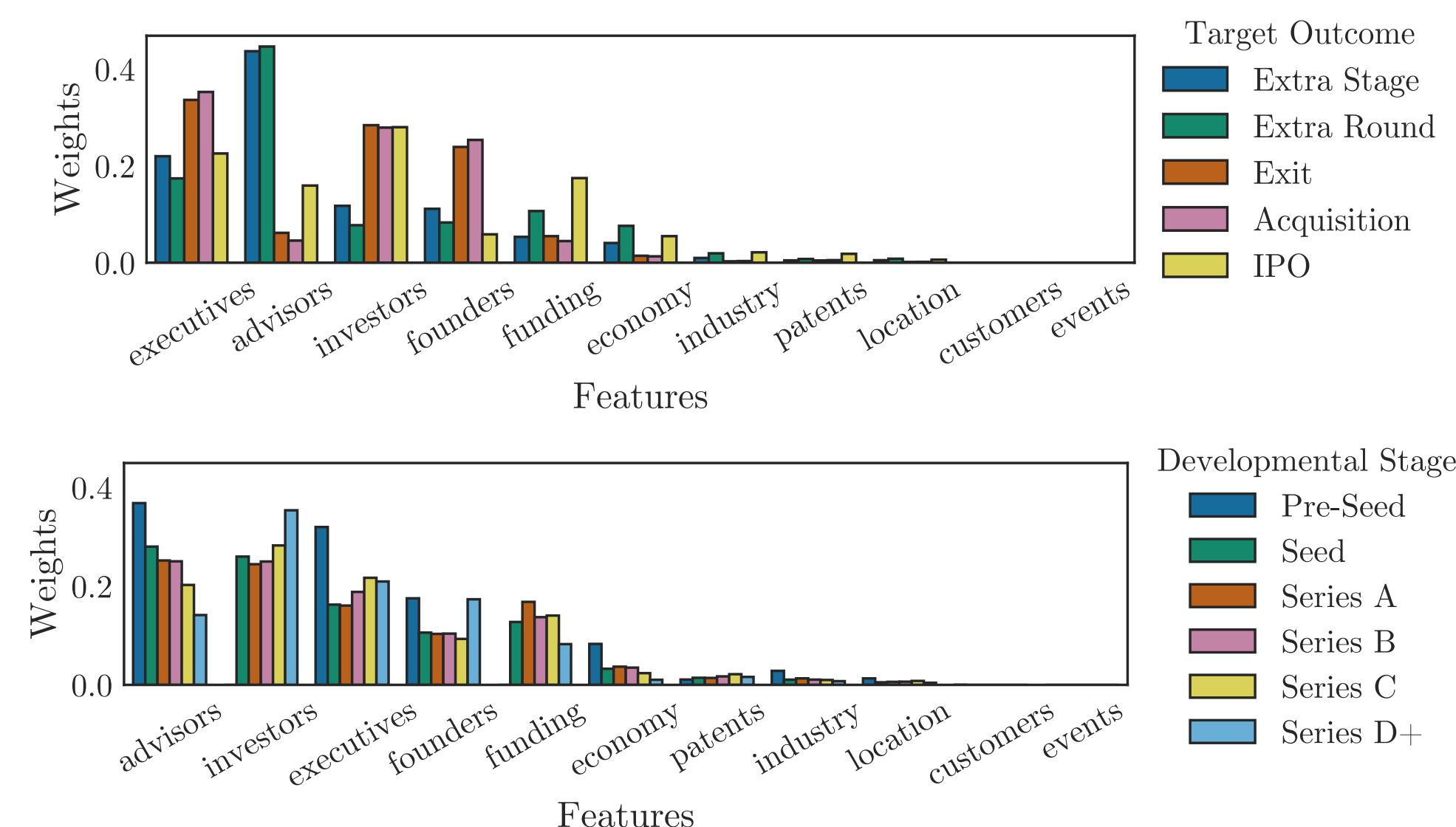


Figure 3. Normalised feature weights by target outcome and developmental stage.

## Evaluation of Criteria

- ✓ **Efficient?** Near-autonomous system, adaptive to different datasets, running time of 46 hours.
- ✓ **Robust?** Minimal variance in performance with respect to training on different dates.
- ✓ **Powerful?** Better or comparable to previous studies. Tested with respect to forecast window, developmental stage, and target outcome.

## Conclusions

This project takes steps towards automation in the VC industry. It makes three primary contributions:

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