

# Towards Automated Venture Capital Screening

**Author:** Mark Shelton

**Supervisors:** Melinda Hodkiewicz, Tim French

**Program:** Bachelor of Philosophy (Honours)

**Date:** 01 June 2017



# Background

## Venture Capital & Startups

AustralianSuper

**first**  
state super



Superannuation  
Funds

AirTree

 Blackbird  
VENTURES

**BlueSky**  
Alternative Thinking

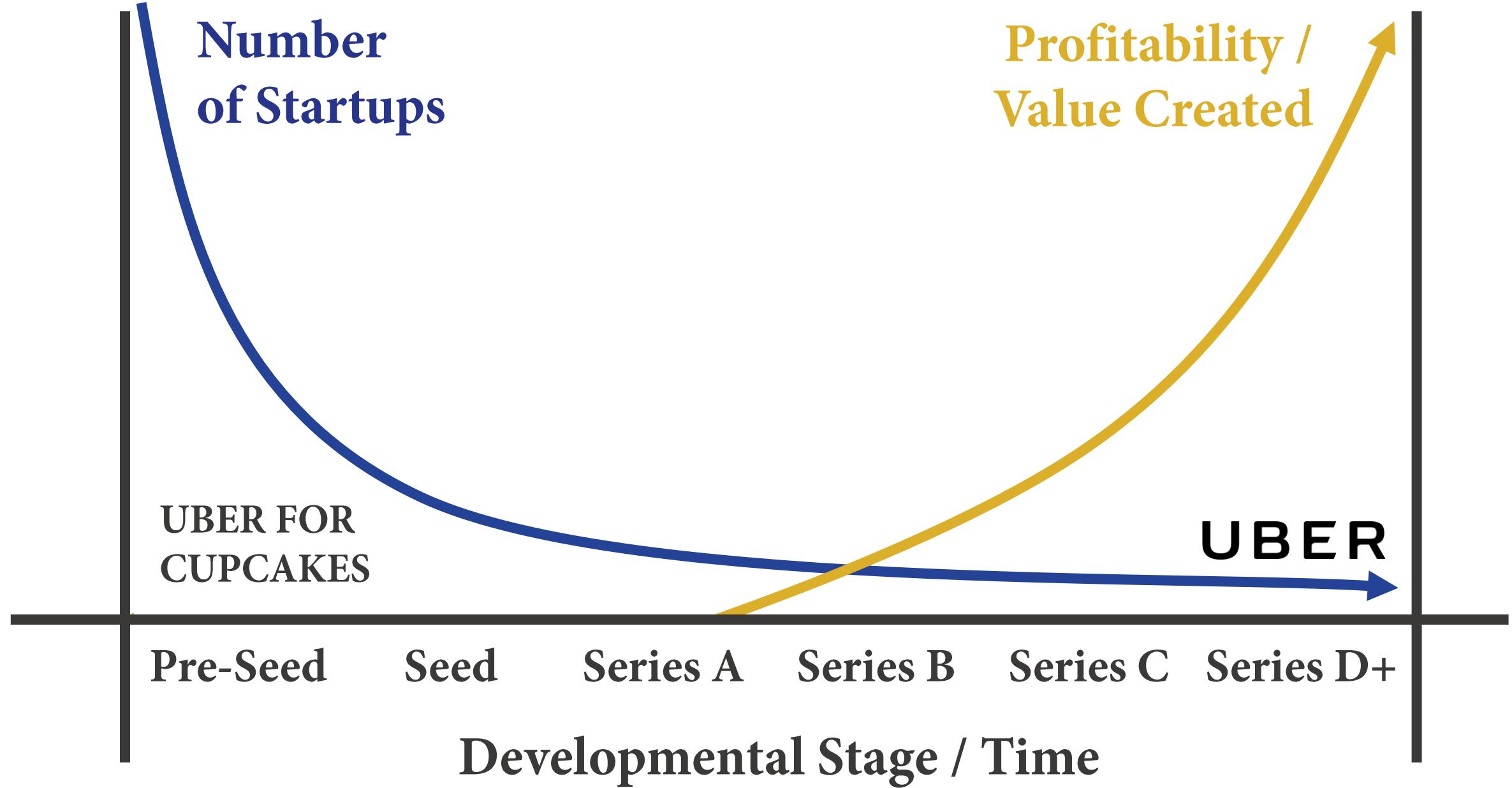
Venture Capital  
Firms

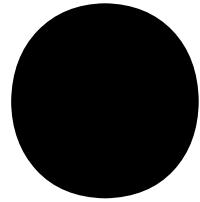
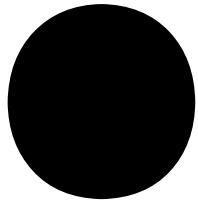
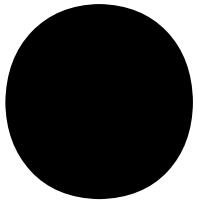
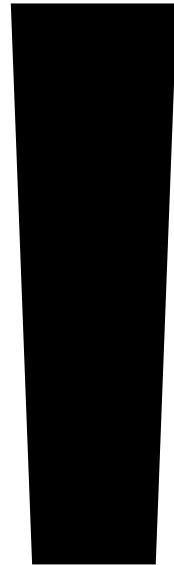
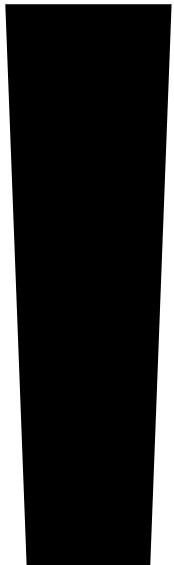
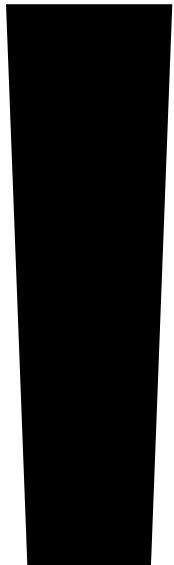
facebook.

 Atlassian

 HealthEngine

Startups





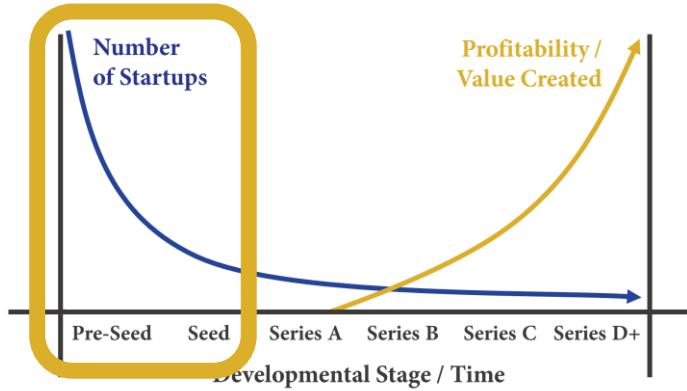
Many investment  
opportunities

Little information  
available

Low chance  
of success

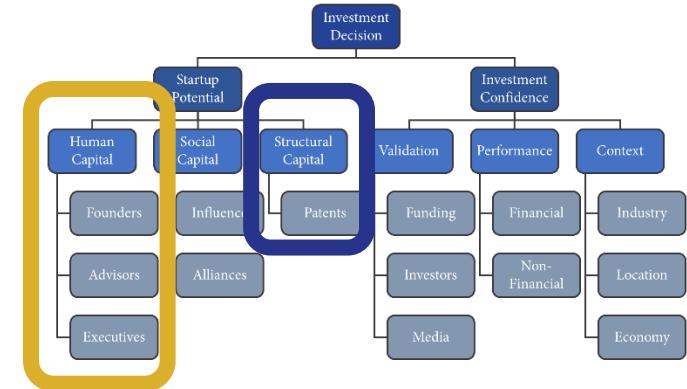
Can we make life  
easier for VC firms  
so they can help more  
startups succeed?

# Limitations of techniques in the literature



Small  
Sample Size

Early-Stage  
Focus



Narrow  
Feature Set

# Investment screening system criteria

Practicality

- Minimal user input, saves effort and time
- Processes new data in a timely fashion

Robustness

- Minimal variance in predictions over time
- Adapts to changing data sources over time

Versatility

- Highly configurable

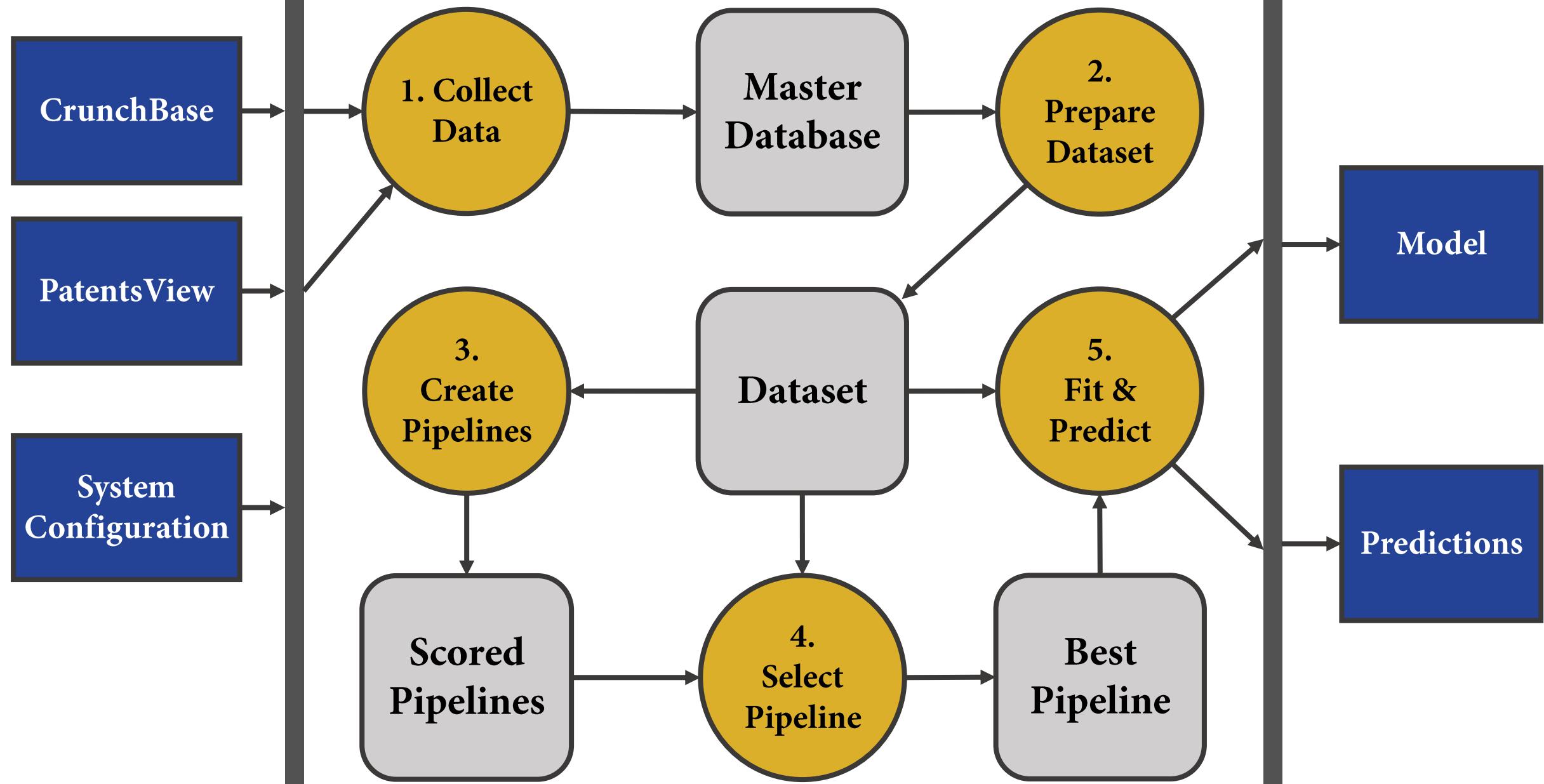
**This seminar  
is organised  
as follows:**

- 0. ~~Background~~
- 1. System Design
- 2. System Performance
- 3. Model Evaluation
- 4. Conclusions

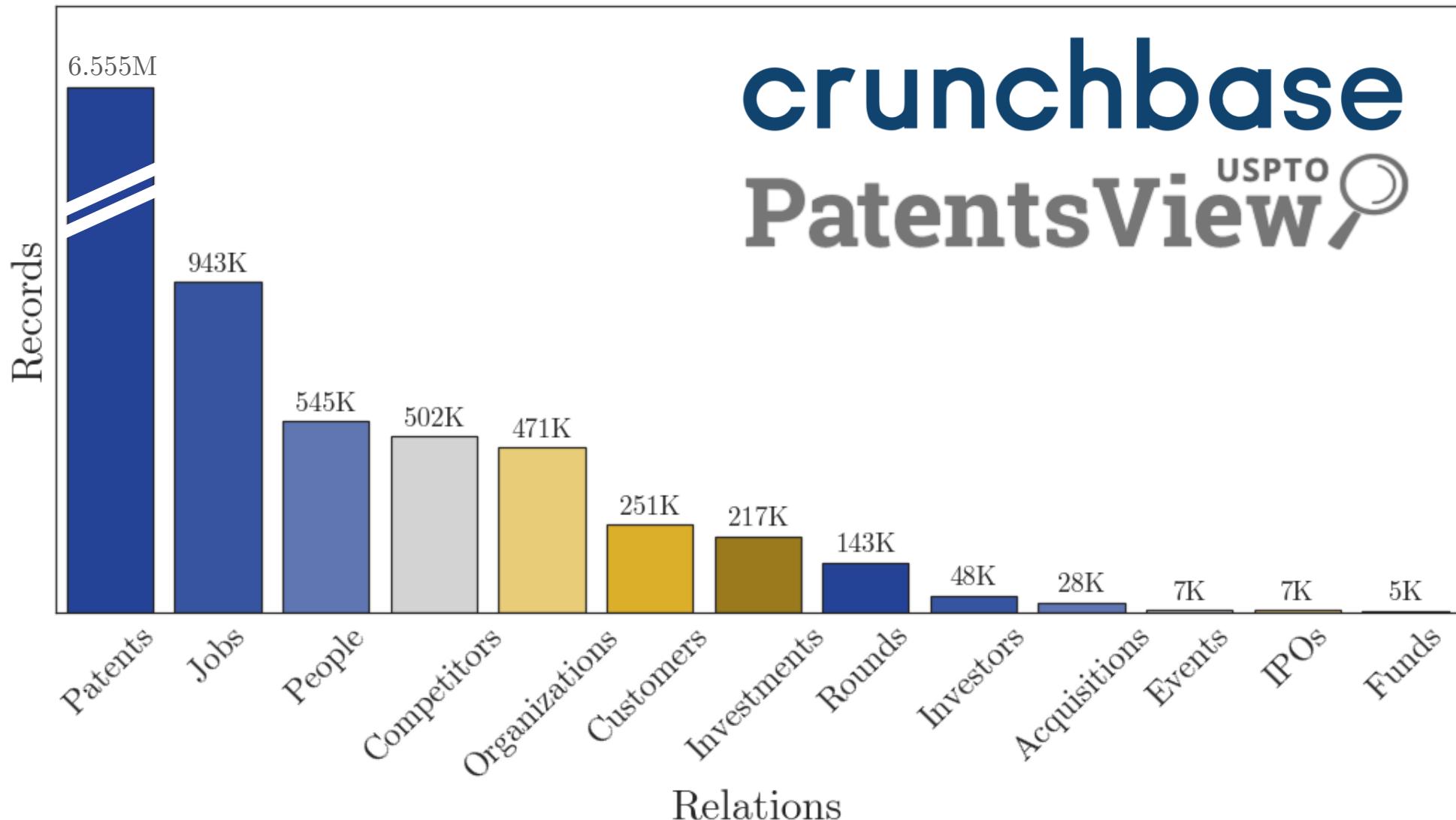
# System Design



# Input

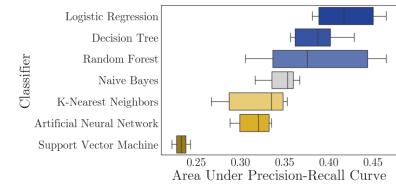
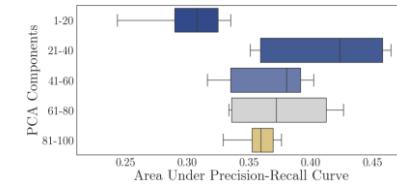
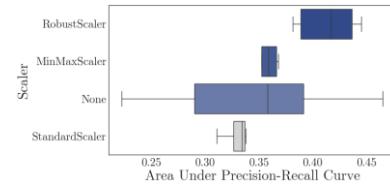
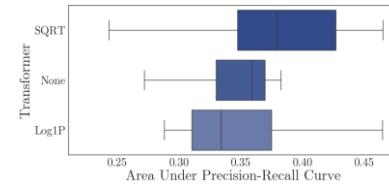
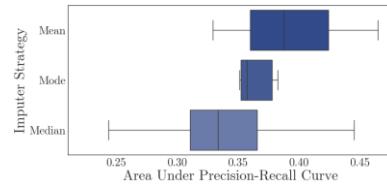
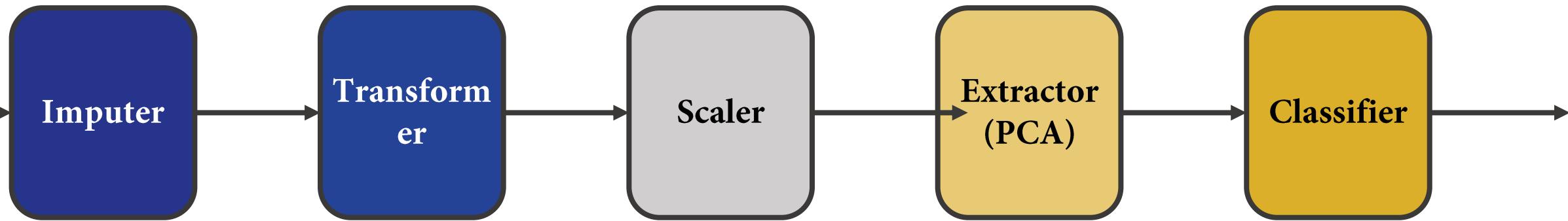


# Design: Data Collection



# Design: Pipeline Creation

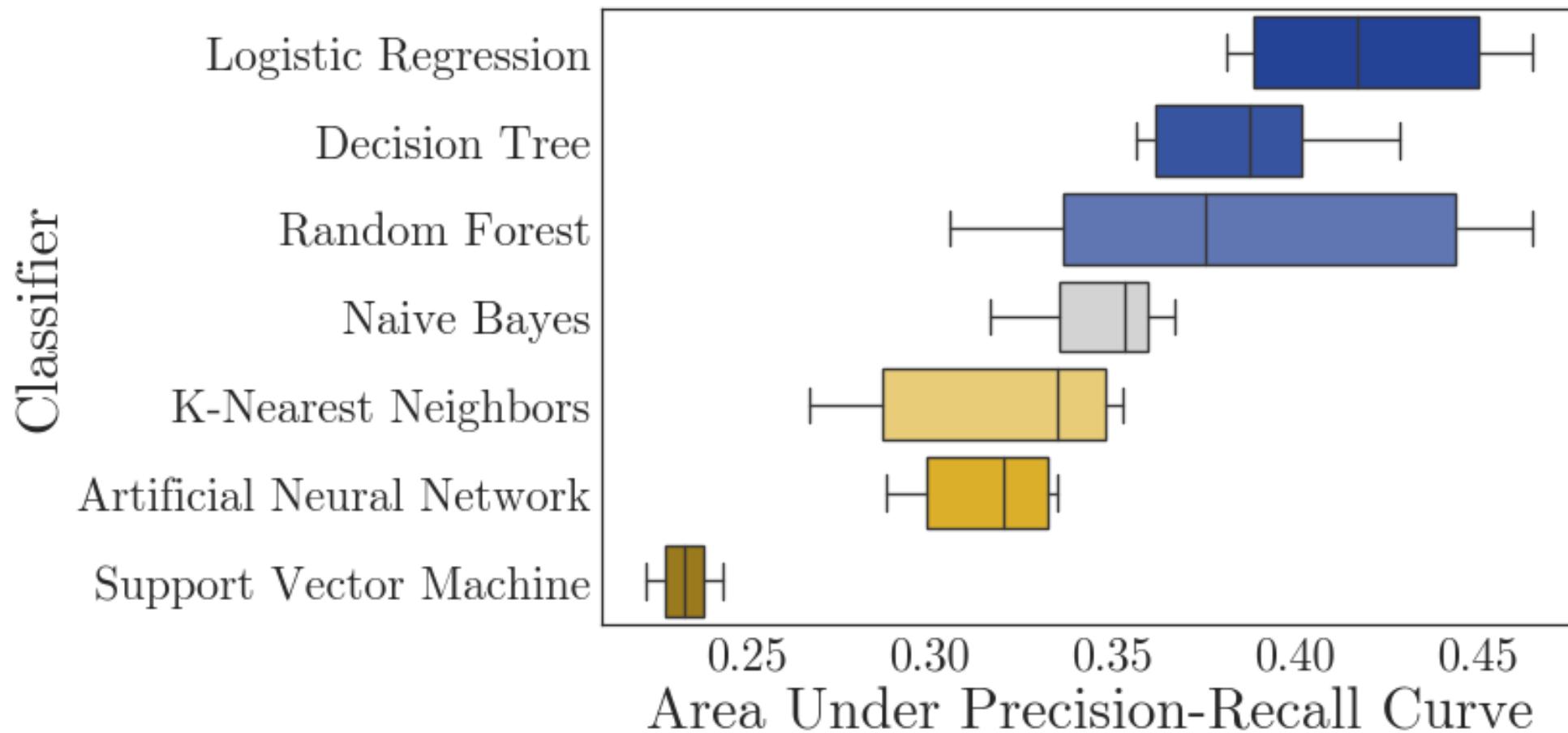
Practicality



Configuration: Forecast Window – 2 Years, Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Design: Pipeline Creation (Classifier)

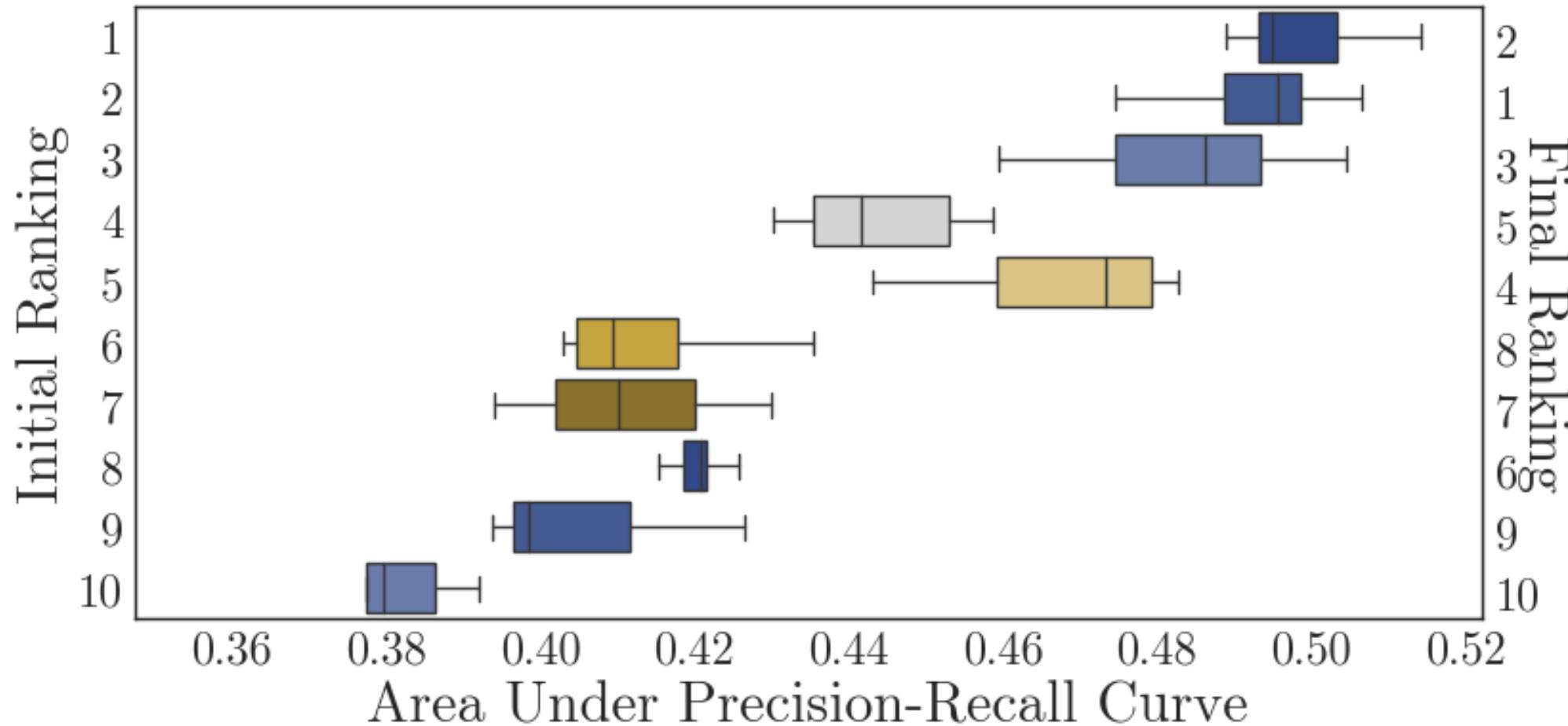
Practicality



**Configuration:** Forecast Window – 2 Years, Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Design: Pipeline Selection

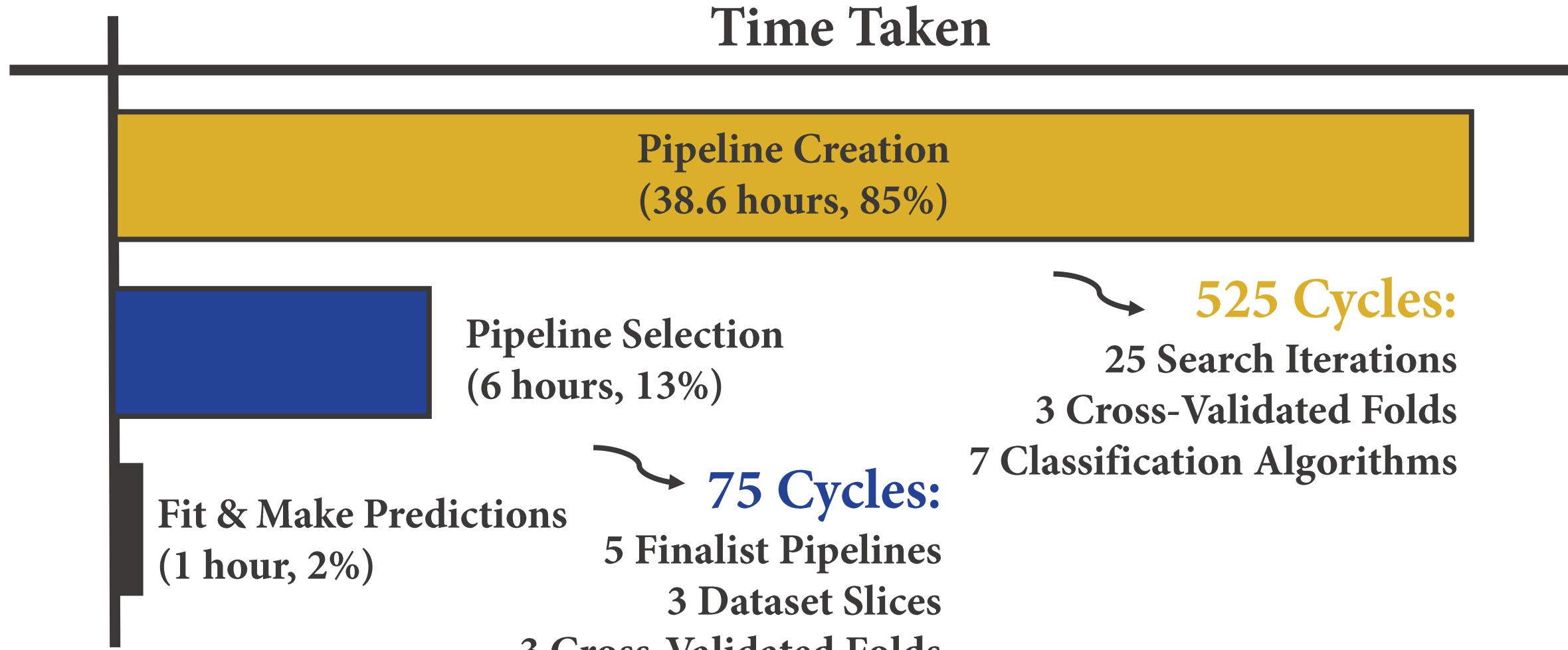
Practicality



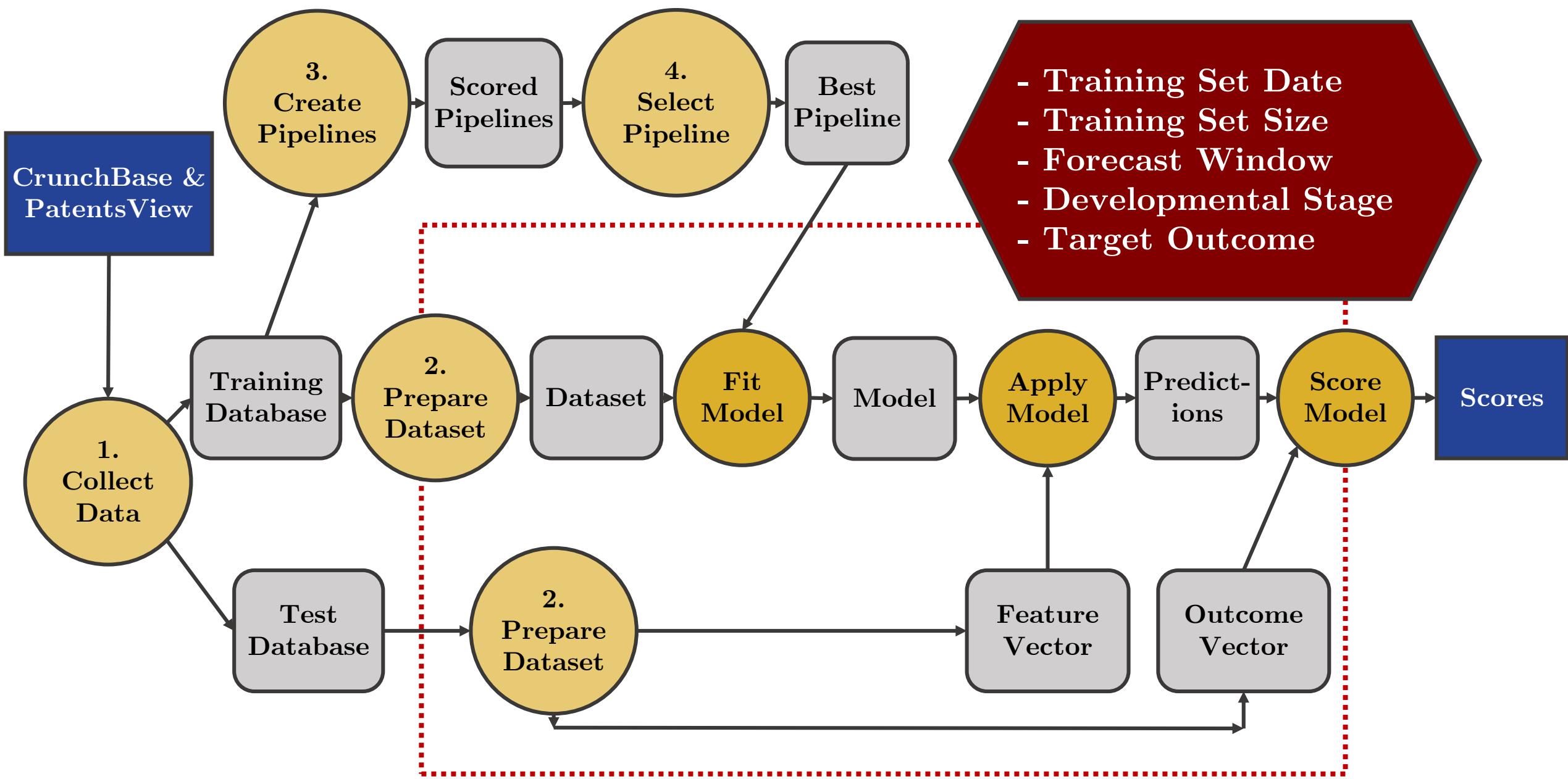
Configuration: Forecast Window – 2 Years, Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Design: Time Profile

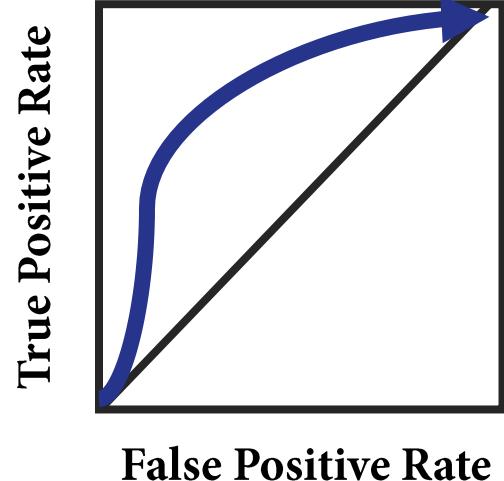
Practicality



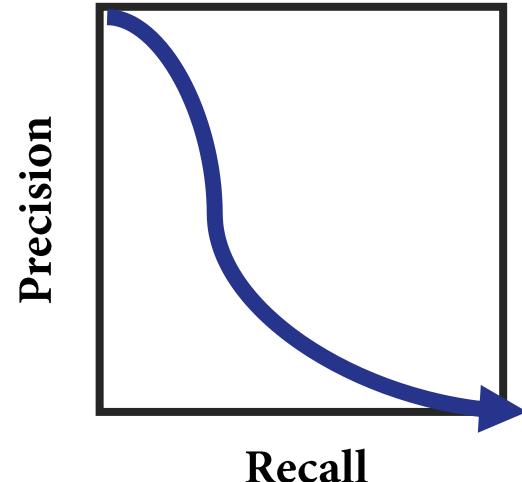
# System Performance



Receiver  
Operating  
Characteristic  
(ROC)  
Curve



Precision-  
Recall  
(PR) Curve



Matthews  
Correlation  
Coefficient  
(MCC)

TP	FN
FP	TN

F1 Score  
(Positive)

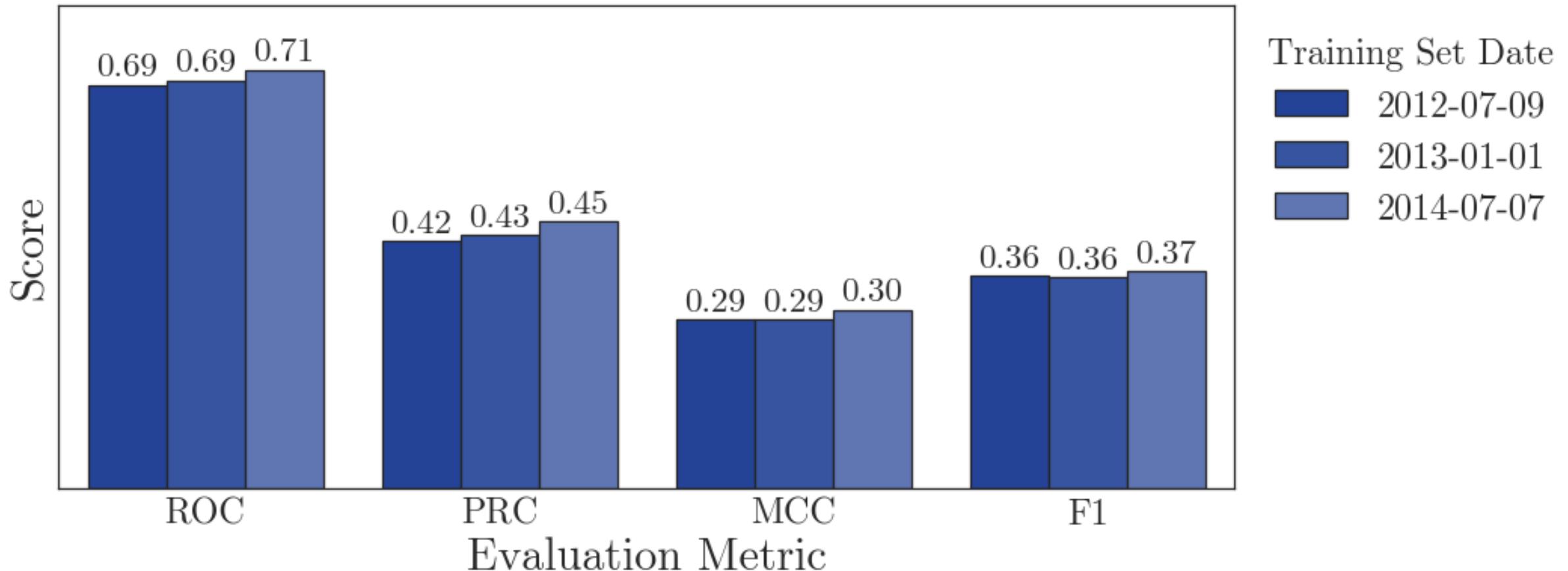


$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Recall

# Performance: Training Set Date

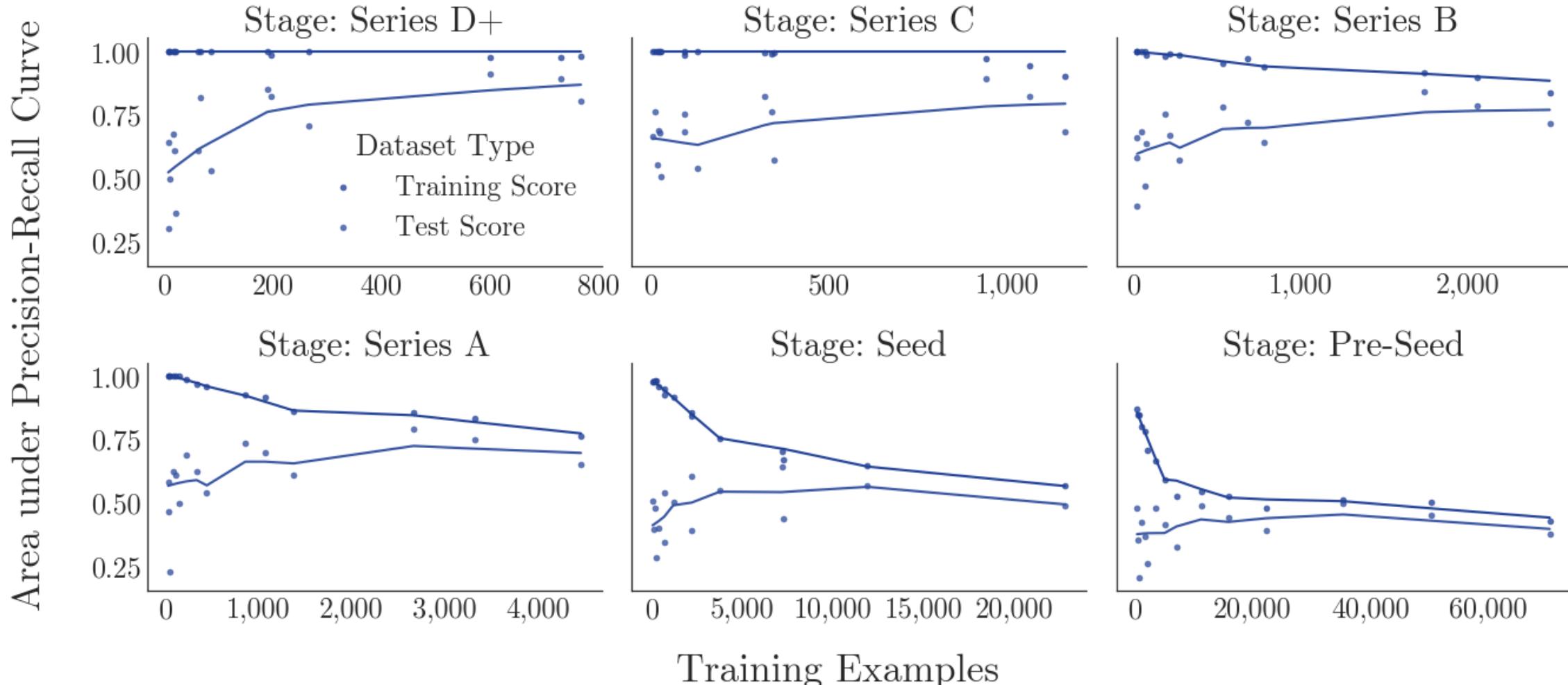
Robustness



Configuration: Forecast Window – 2 Years, Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Performance: Training Set Size

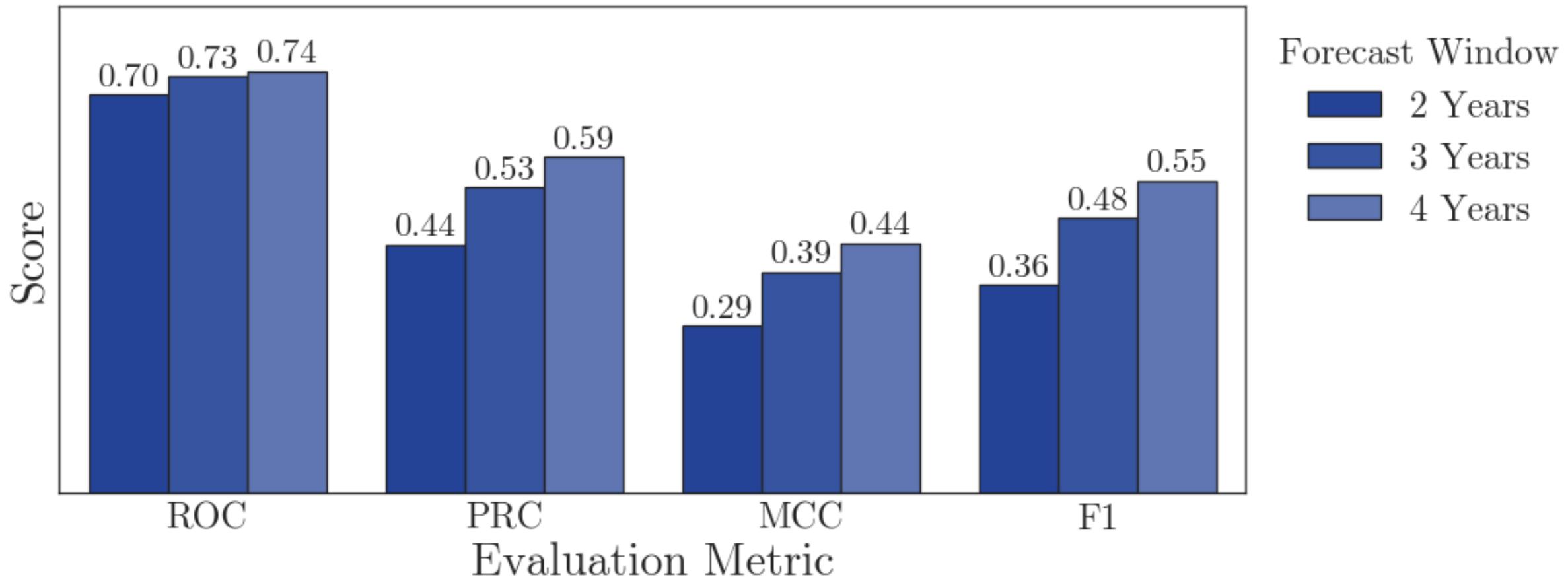
Robustness



Configuration: Dataset Slices – 9 (2012-16), Forecast Window – 2-4 Years, Target Outcome – Acquisition, IPO or Extra Funding.

# Performance: Forecast Window

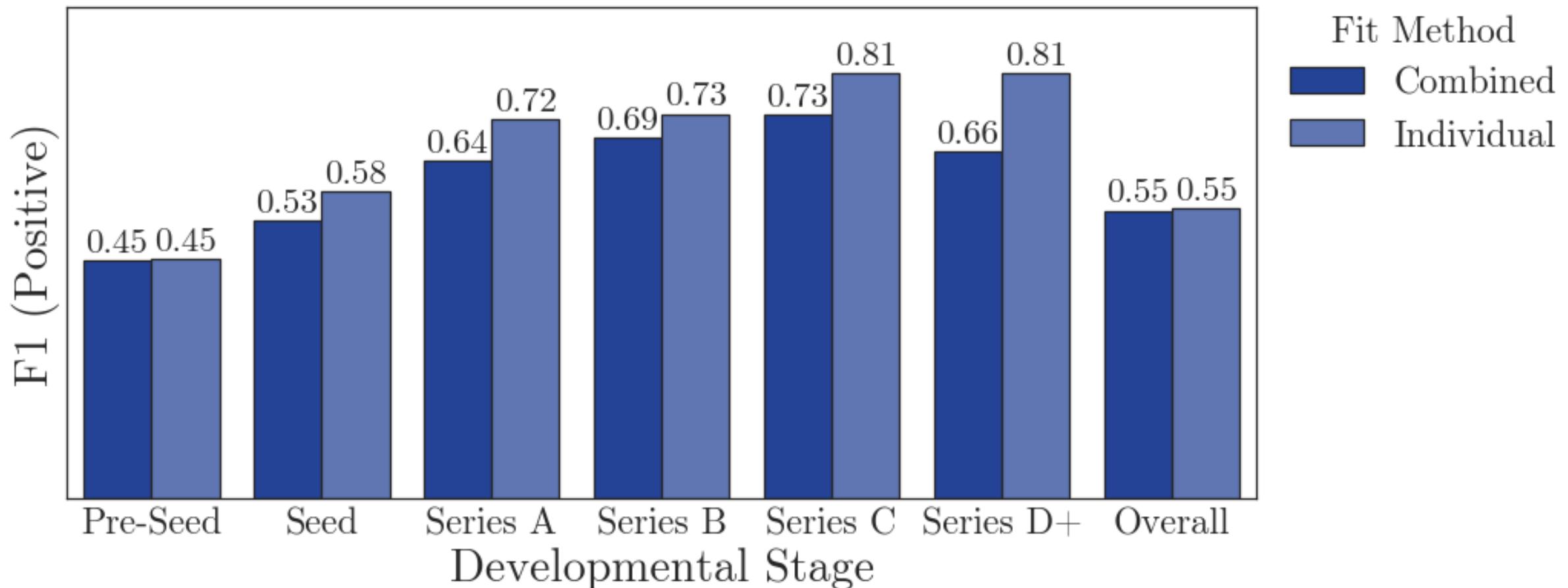
Versatility



Configuration: Dataset Slices – 9 (2012-16), Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Performance: Developmental Stage

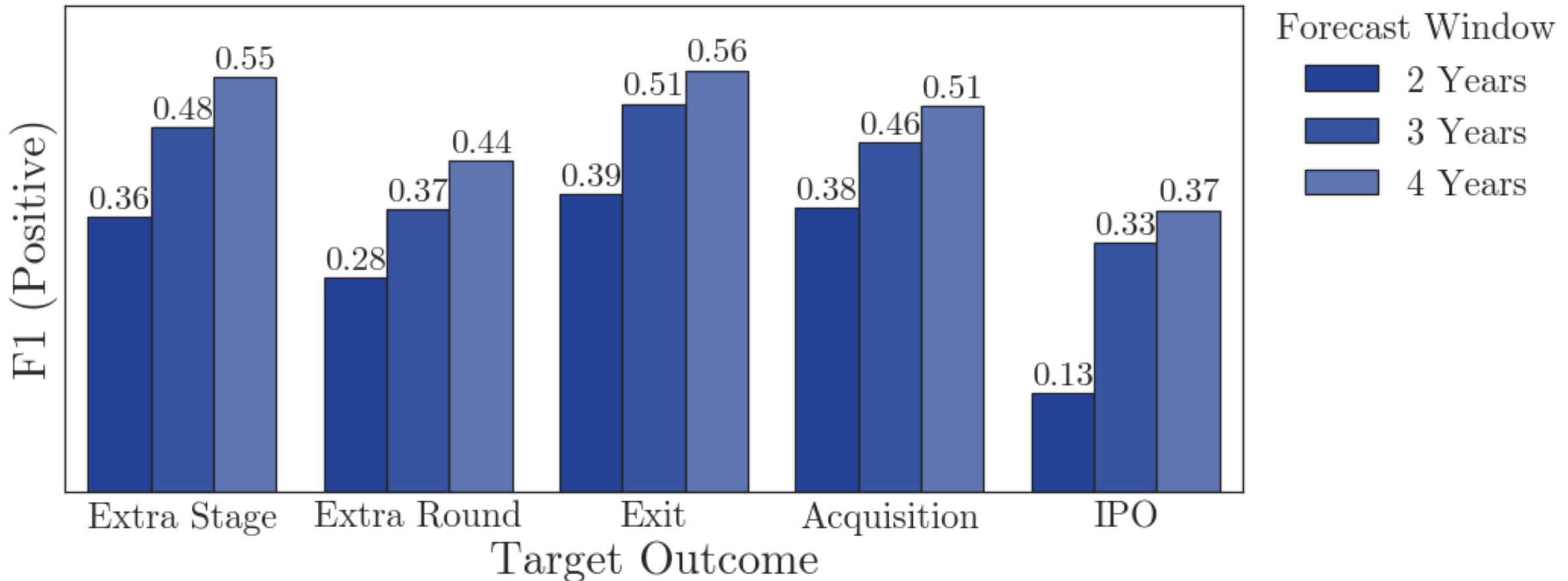
Versatility



Configuration: Dataset Slices – 3, Forecast Window – 4 Years, Target Outcome – Acquisition, IPO or Extra Funding.

# Performance: Target Outcome

Versatility



Configuration: Dataset Slices – 9 (2012-16), Developmental Stages – All.

# Performance: Comparison to Literature

Versatility

Beckwith  
(2016)

- Predicting equity crowdfunding success
- Best score – F1 (Positive) of 0.33

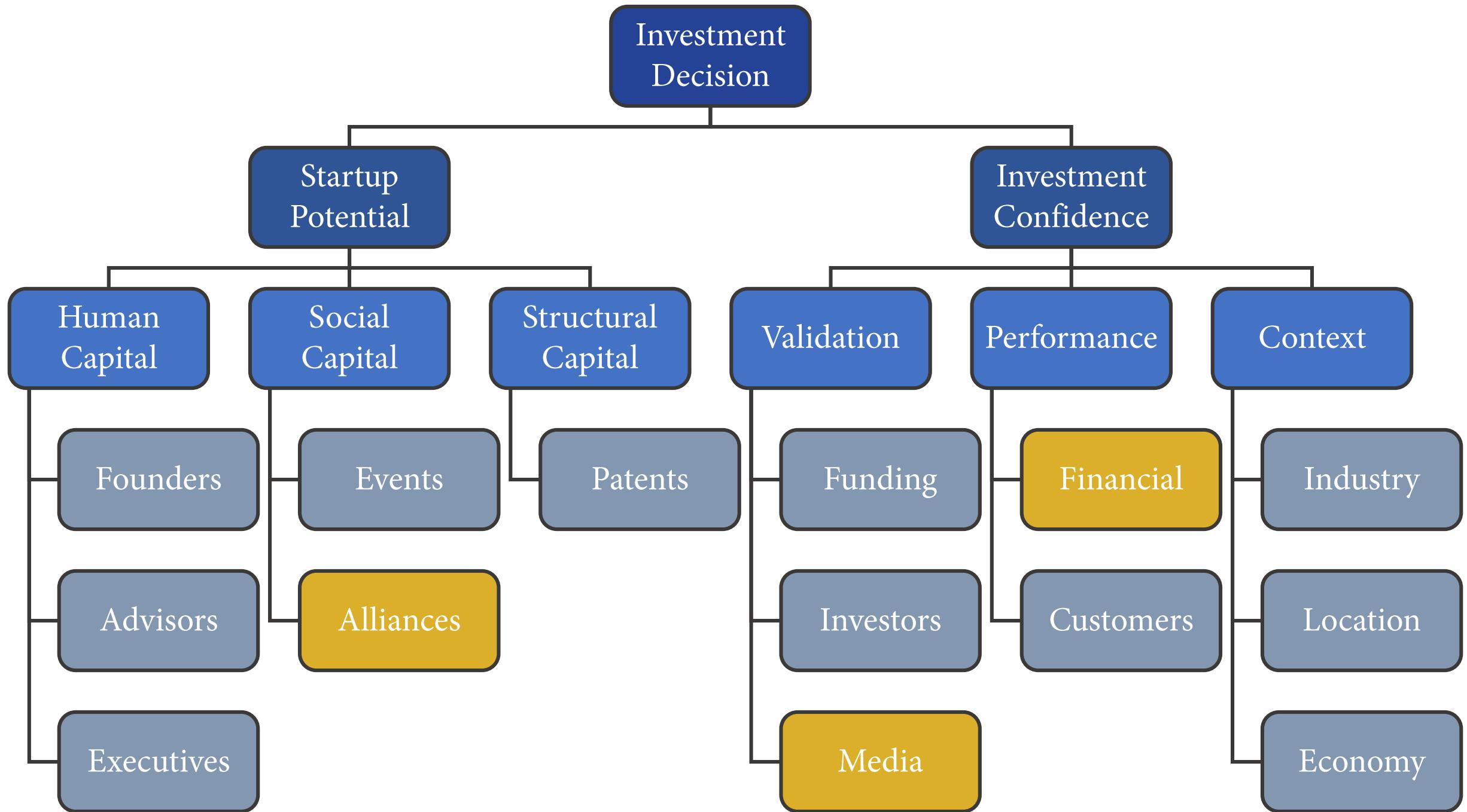
Stone (2014)

- Predicting funding for Pre-Seed companies
- Best score – ROC (AUC) of 0.57

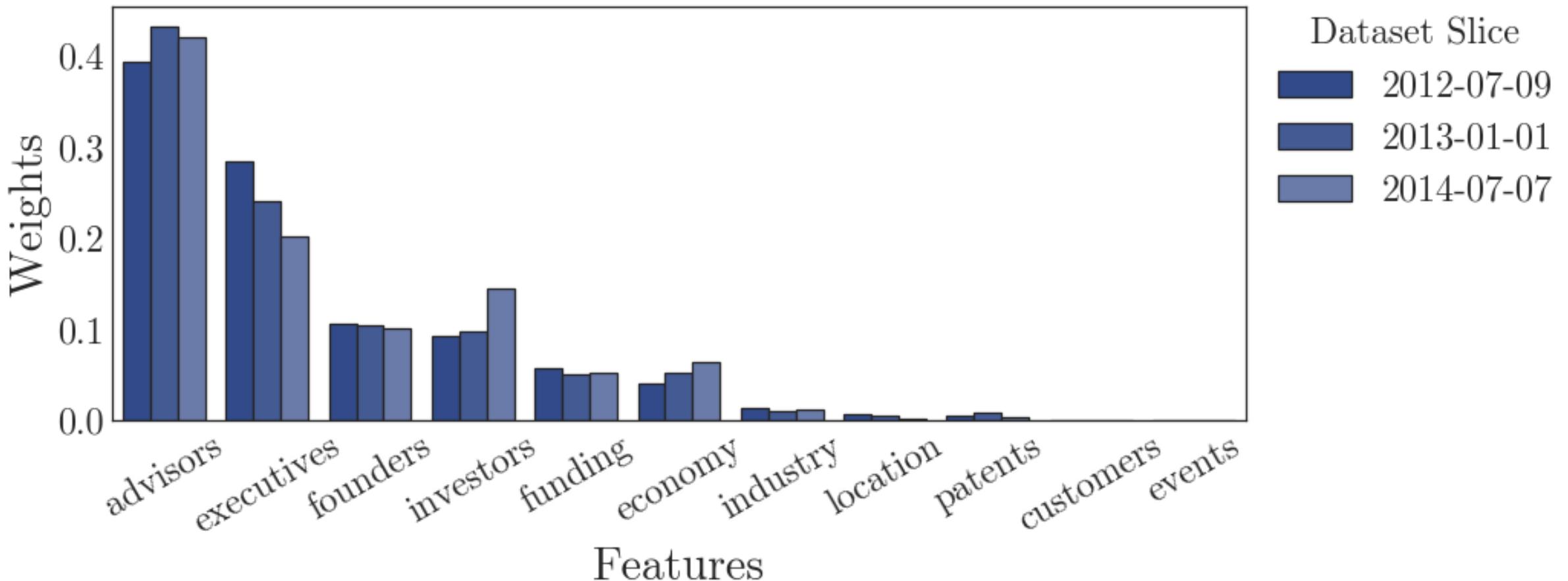
Bhat (2011)

- Predicting exits for Series C+ companies
- Best score – F1 (Positive) of 0.75

# Model Evaluation

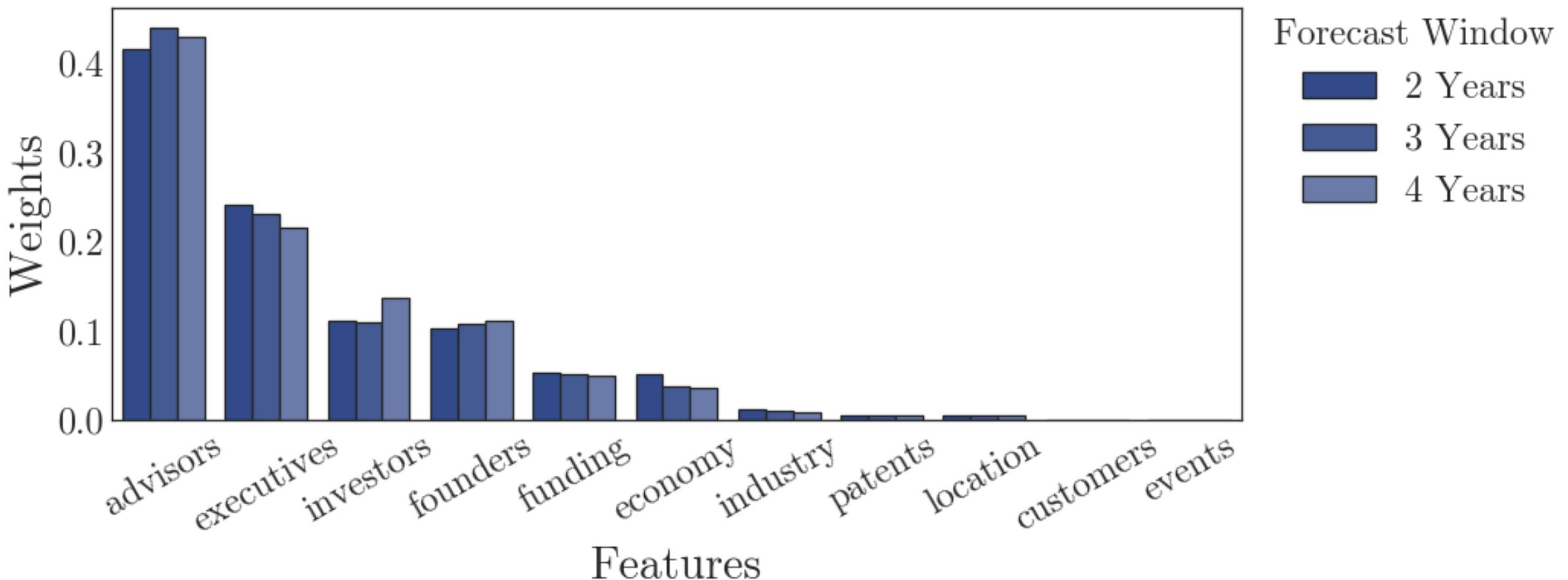


# Features: Training Set Dates



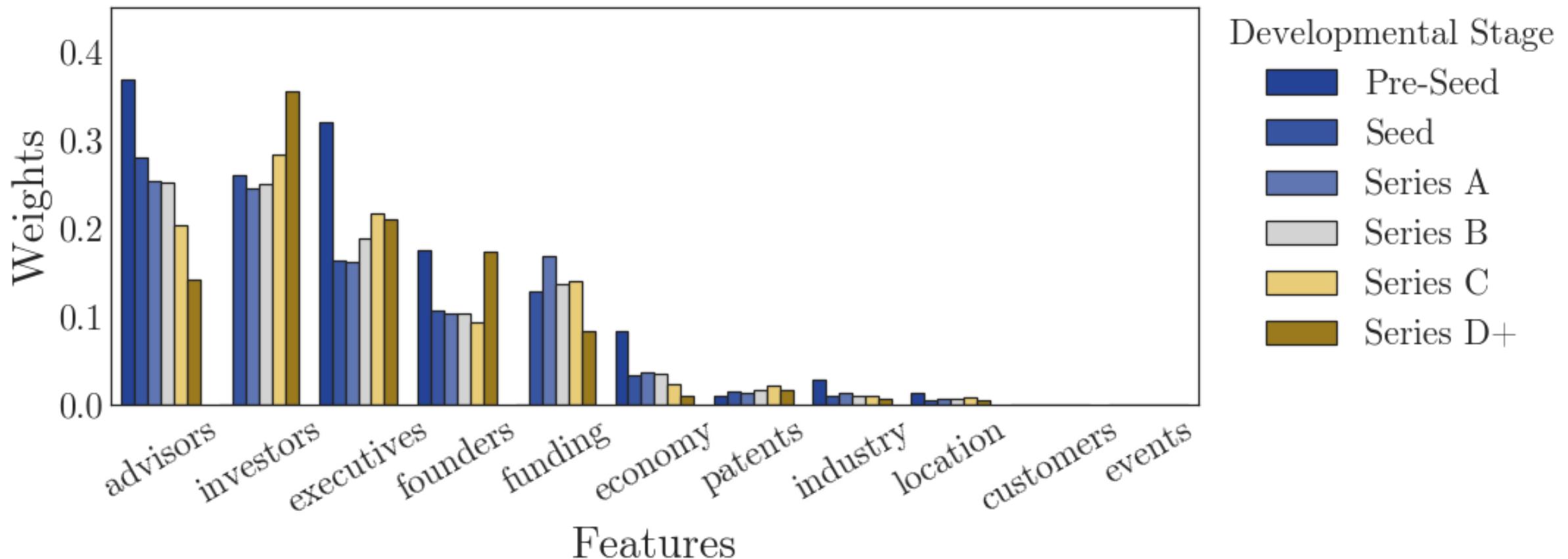
Configuration: Forecast Window – 2 Years, Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Features: Forecast Window



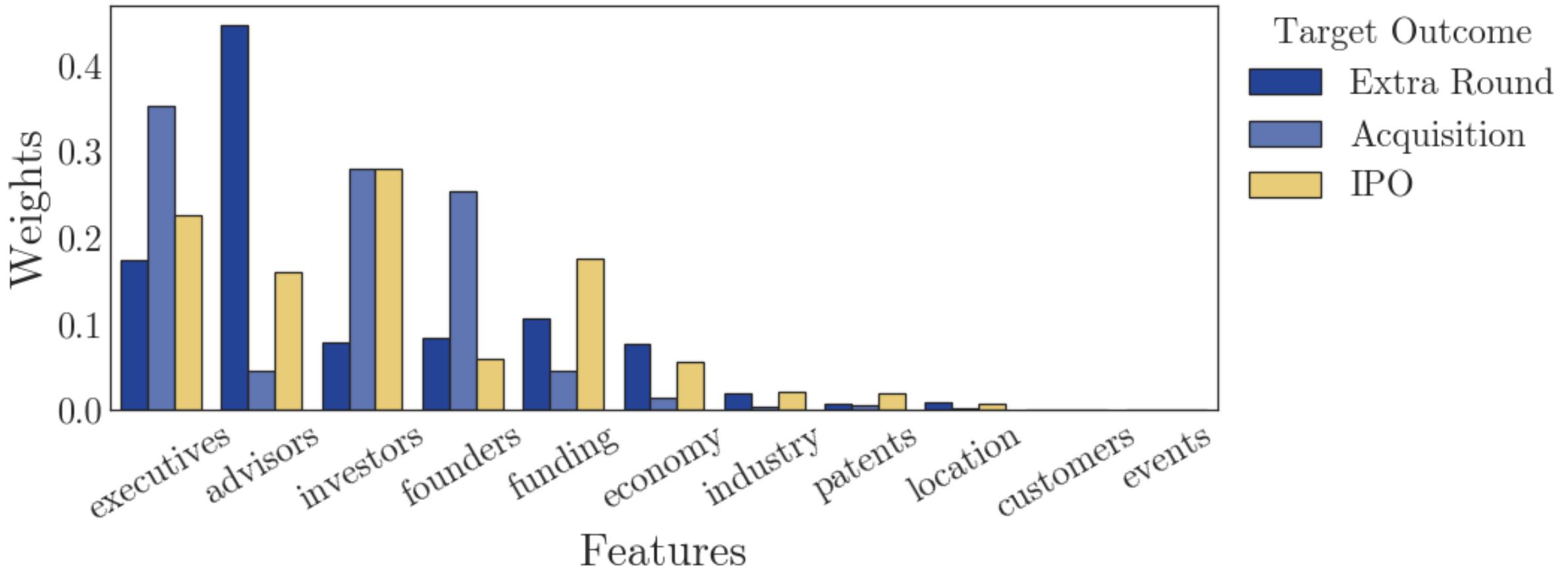
Configuration: Dataset Slices – 9 (2012-16), Developmental Stages – All, Target Outcome – Acquisition, IPO or Extra Funding.

# Features: Developmental Stage

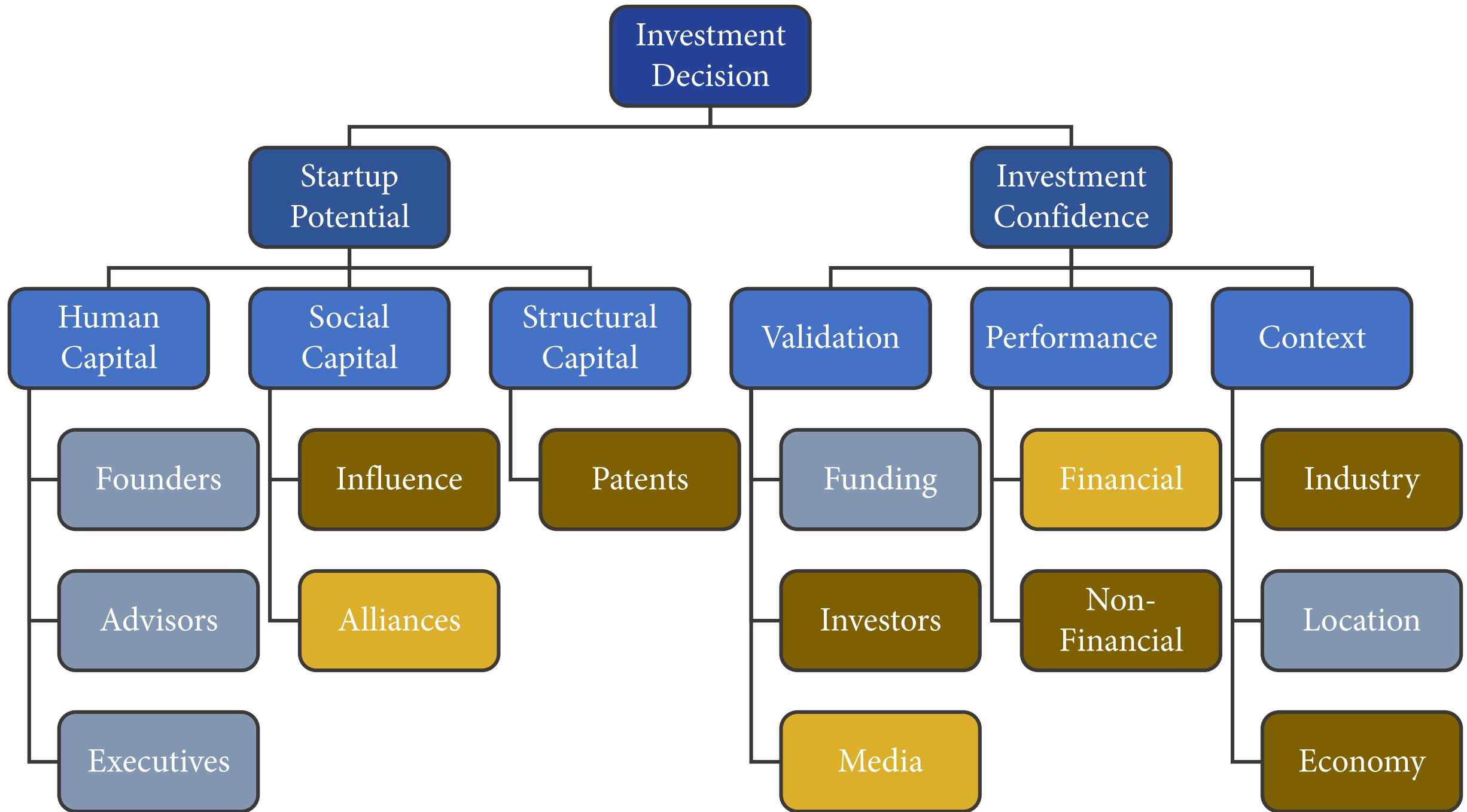


Configuration: Dataset Slices – 9 (2012-16), Forecast Window – 2-4 Years, Target Outcome – Acquisition, IPO or Extra Funding.

# Features: Target Outcome



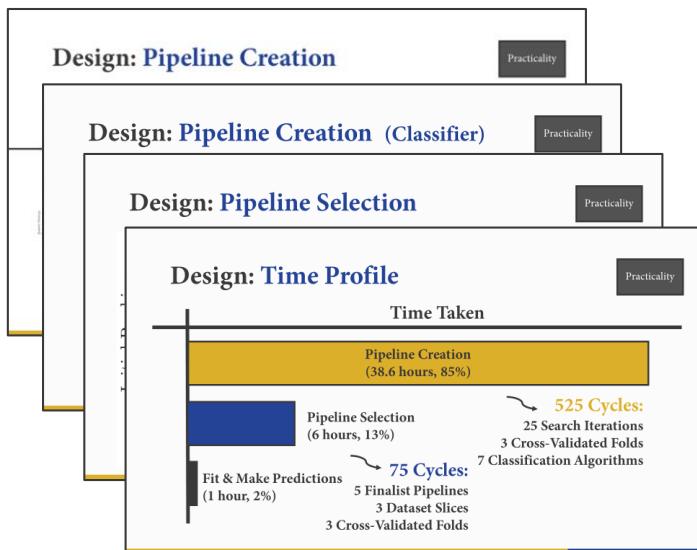
Configuration: Dataset Slices – 9 (2012-16), Forecast Window – 2-4 Years, Developmental Stages – All.



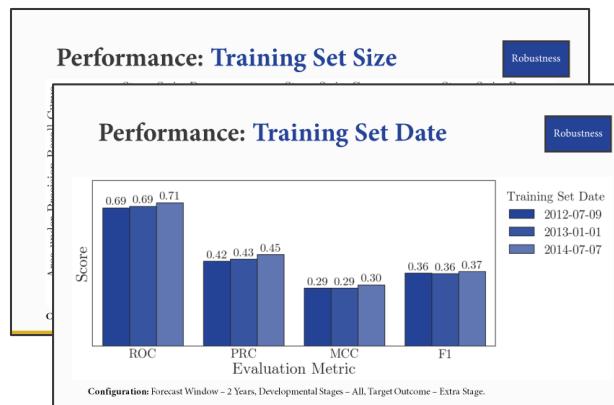
# Conclusions

# Evaluation of Criteria

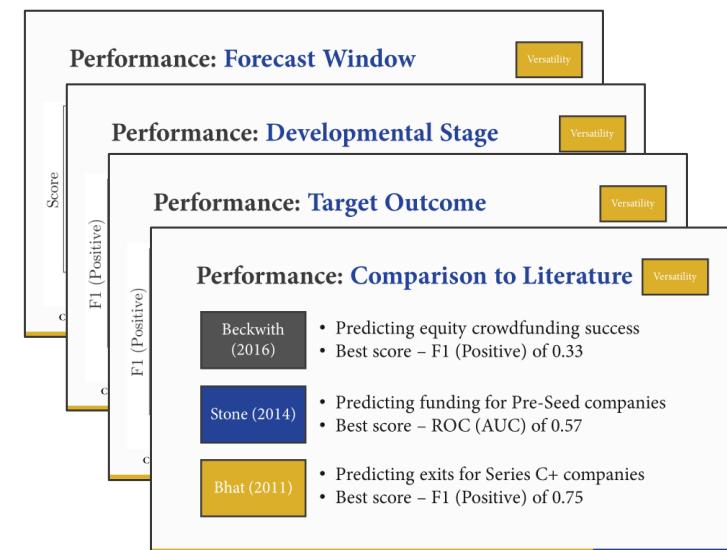
## Practicality



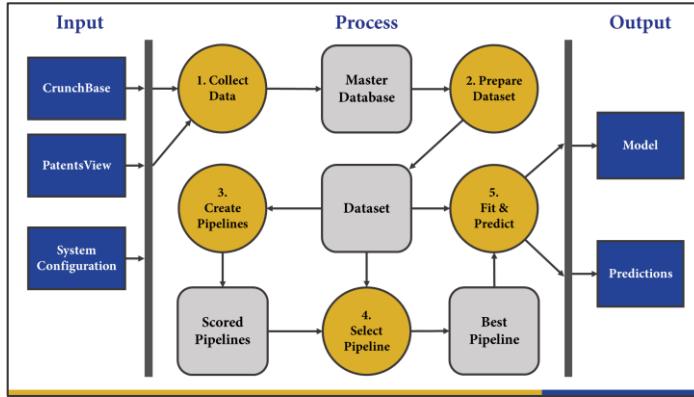
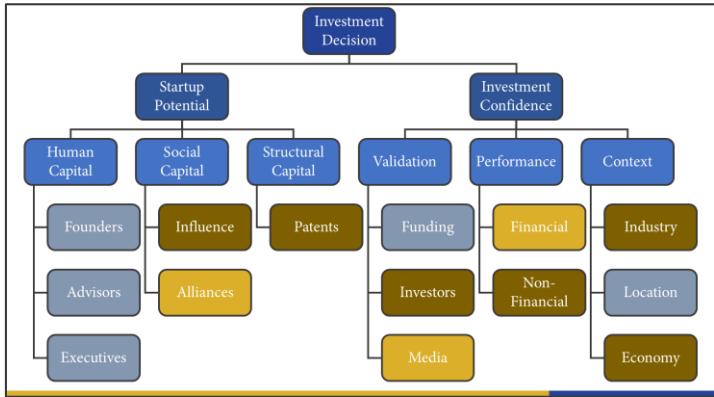
## Robustness



## Versatility



# Future Work



Extend  
Feature Set

Automatic  
Scheduling &  
Data Collection

Commercialisation

# Contributions

1. A VC investment screening system ready for industry: near-autonomous, adaptable, and leverages public data.
2. A VC investment screening system performs better or comparable to previous studies across a large domain.
3. An empirical study of startup performance more comprehensive than any from the literature.

# Towards Automated Venture Capital Screening

**Author:** Mark Shelton

**Supervisors:** Melinda Hodkiewicz, Tim French

**Program:** Bachelor of Philosophy (Honours)

**Date:** 01 June 2017

# Appendices

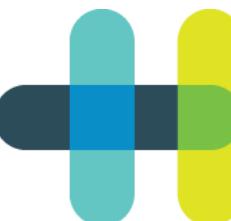
**facebook.**

 **Atlassian**

 **airbnb**

**Canva**

 **Spotify®**

 **HealthEngine**

# What do venture capitalists think?

*“We have not been able to quantify startup potential. We haven't 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”*

---

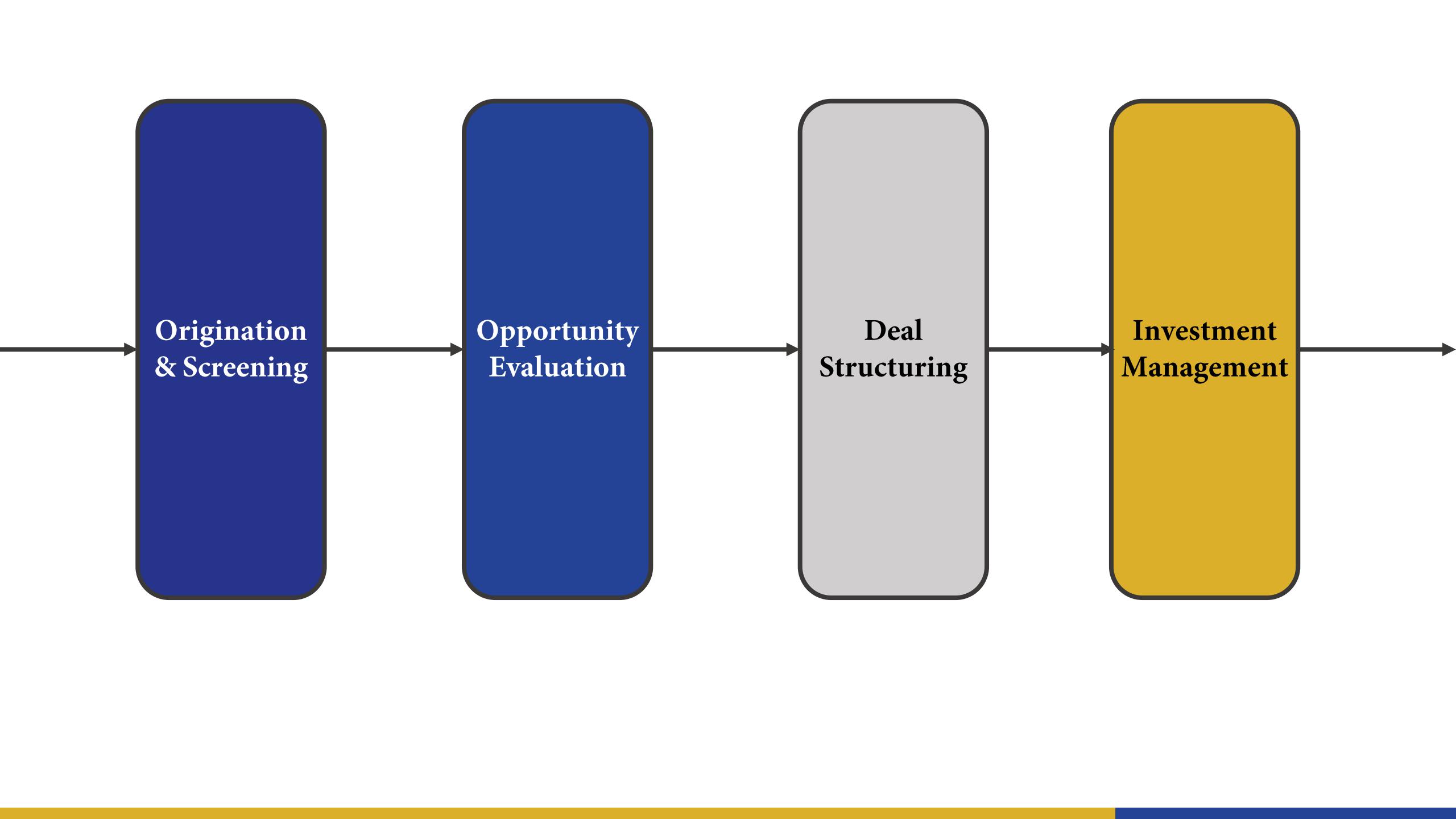
**Fred Wilson, Union Square Ventures**

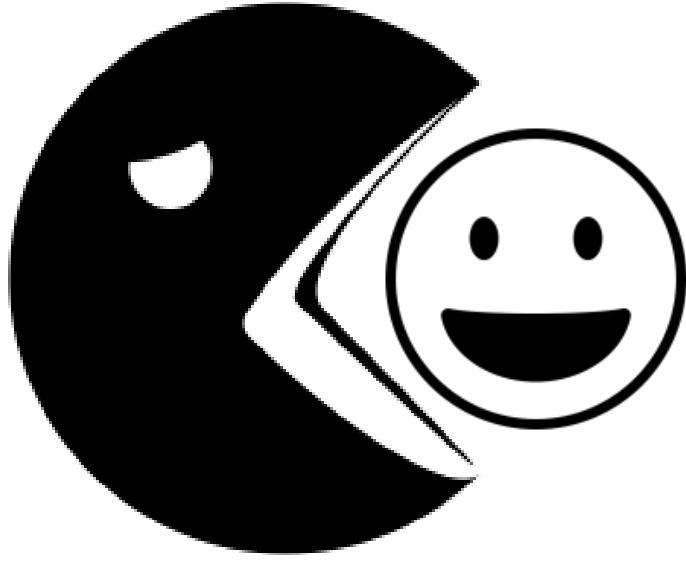
**Chris Dixon, Andreessen Horowitz**

---

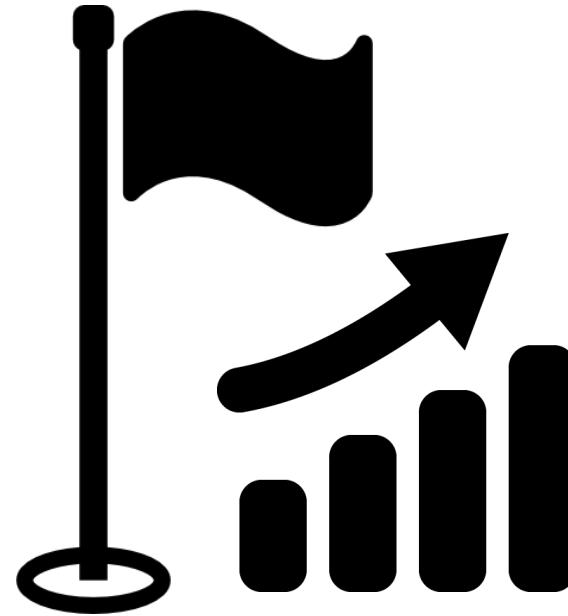
*“One of the main activities of good investors is trying to find ‘accurate contrarian theses’ about what make good startups, markets, founders etc. I've seen a few attempts to do it quantitatively but I think those are often flawed because the measurable things are either obvious, irrelevant, or suffer from over-fitting.”*







Acquired



IPO

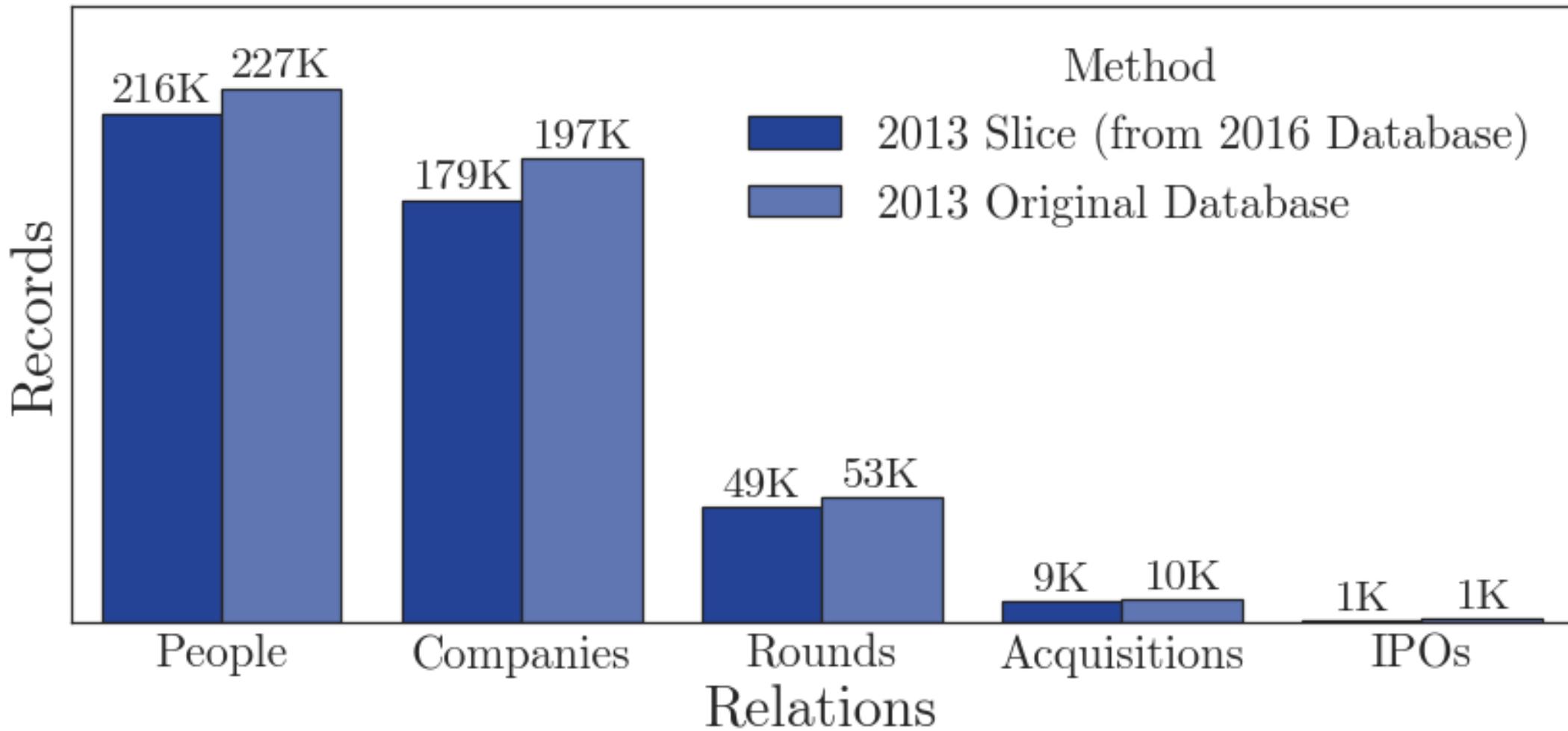


Extra Round

Exit

Extra Stage

# Design: Dataset Preparation





## Feature (2013-April)

Age (Years)	7.4	0.6	4.3	3.8
Funding Raised (\$m)	92.0	0.0	171.0	12.0
Funding Rounds (N)	8	0	8	4
Developmental Stage	Series D+	Pre-Seed	Series C	Series A
Predicted Outcome	✓	✗	✓	✓

## Outcome (2017-April)

Funding Raised (\$m)	96.0	5.0	336.0	77.0
Developmental Stage	Series D+	Series A	Acquired	Series B
Actual Outcome	✗	✓	✓	✓
Correct Prediction	✗	✗	✓	✓

Configuration: Dataset Slices – 3, Forecast Window – 4 Years, Target Outcome – Extra Stage.

# Bibliography

Beckwith, J. Predicting success in equity crowdfunding.  
Honours Thesis. University of Pennsylvania. 2016.

Stone, T.R. Computational analytics for venture finance. PhD  
Thesis. University College London. 2014.

Bhat, H.S., and Zaelit, D. Predicting company exits using  
qualitative data. In Pacific-Asia Conference on Knowledge  
Discovery and Data Mining. Springer. 2011, pp. 399-410