



**Southeast Asia's First
Student-run Venture Fund**

Established in 2017

Data Science and Analytics in Venture Capital

Team 1:

**Augustine Thia, Gordon Ng, Kelvin Sutedja, Marcus Wan,
Tee Zhi Zhang, Wang Dong Yue Felix**

Team Lead: Nicholas Koh, Benedict Chong

The Rise of Technology Adoption in VC



Increasing amount of VC firms in SEA opening up data science and analytics arms...


And usage of data science and analytics is gaining popularity in mainstream media

We're hiring for various technical roles! If you've just started coding and are looking to further your skills with mentorship and business applications, PM me to find out more :-)

**Data Science Intern**
Vertex Holdings • Singapore, Singapore

16

We are hiring! Feel free to reach out to me if you want to know more about this role.

**Data Science - Associate**
Openspace Ventures • Singapore, Singapore

35 • 3 comments

**500 Startups**
500 Startups is a global venture capital firm headquartered in San Francisco

Overview People Culture Funding Jobs **32** Share Follow

Investments Data Analyst

Apply now

The Rise of Intelligent and Data-Driven Venture Investing

Featured Article

VCs double down on data-driven investment models

New efforts could automate the role of an analyst

McKinsey Quarterly

A machine-learning approach to venture capital

June 26, 2017 | Interview

Introduction

Deal Sourcing

Deal Scoring

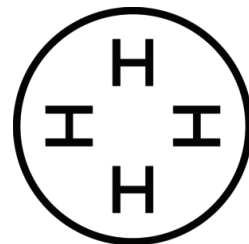
Portfolio Support

Fund Operations

Recommendations

The Rise of Technology Adoption in VC

645
VENTURES



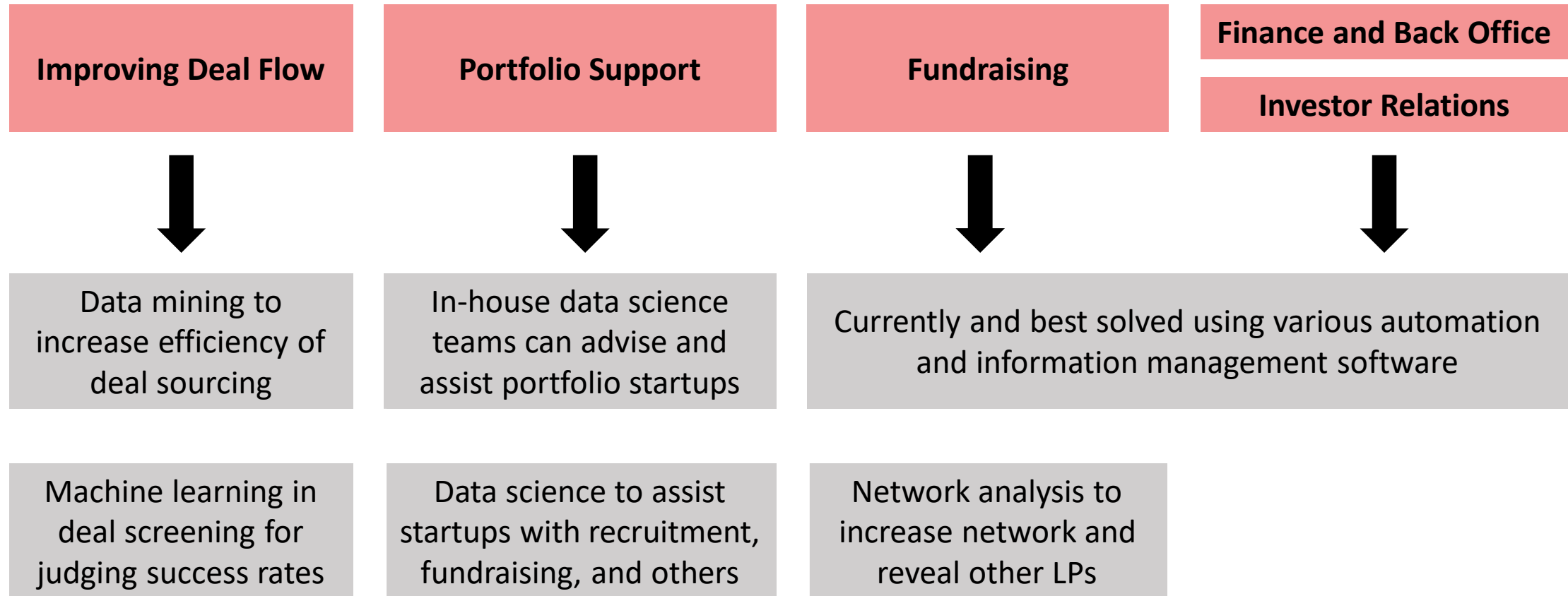
How Data Science and Analytics can aid VCs

Breakdown of VC Operations

Operations	General Description	Potential Bottlenecks
Marketing/ Fundraising	Raising funds from investors and developing outreach of fund	Difficulty in raising funds from LPs
Investor Relations	Regular meetings with investors and managing information to them	Bogged down by content creation, documentations, and meetings
Deal Flow Generation	Sourcing for start-ups, building pipeline and due diligence for start-ups	How to increase higher-quality deal flow and accurately screen them
Portfolio Support	Introductions, interviewing job candidates, general business and strategic advice	Inability to significantly value add to portfolio companies
Finance and Back Office	Fund administration, fund compliance, financial reporting	May be bogged down by administrative challenges and delays

Usage of Data Science and Analytics in VC Ops

Ways Data Science and Analytics can complement VC Firms:



We will briefly go through the technologies and case studies of deal flow, portfolio support and fundraising

Data Science and Analytics for Deal Sourcing

Digital Footprint

First level

- Twitter
- Facebook
- Web traffic
- App store rankings

Second level

- Web scraping/ data crawling
- Identification of keywords or phrases that express positive or negative sentiments

Third level



Assess

- Customer demand
- Traction
- Product-market fit

- Early identification of trends and opportunities
- Quantitative ranking of customer sentiment
- Set alerts to notify of any sentiment-changing events

Case Study: Hone Capital



Description

- VC arm of Chinese PE firm CSC Group
- \$400 mn committed, mandate to invest in hundreds of quality startups as fast as possible
- Founded in 2015, became SV's most active seed investor in 2017 (excl. Accelerators)
- Established strategic partnership with AngelList

Technology

- ML model feeds on database of more than 30,000 deals
- Explored over 400 distinct characteristics
- Further filtered to 20 indicators most predictive of future success

Key Insights

- Average seed investment
\$1.5 million for startups that advance to Series A
\$0.5 million for startups that fail to advance
- Background of founders
Deal with two founders from different universities is twice as likely to succeed as those with founders from the same university

Result

- Increased deal flow by up to 20 deals per week
- Significant improvement in diversity and quality of deals sourced
- 50% of seed stage deals have led to follow-on investments (above industry average)
- Unrealized returns among top 20% of investors

Data Science and Analytics for Deal Scoring

Data Science for Deal Scoring

Idea: Build a machine learning algorithm to identify the probability of a new deal being "good" or "bad"

Human tagging



Supervised Classification

Advantages

- Gives the probability/likelihood of a deal being "good" or "bad"

Disadvantages

- Requires a lot of manpower and time to do tagging
- Requires a lot of training data to achieve high accuracy, which may be costly
- Difficult to identify the best features

No tags



Unsupervised Classification

Advantages

- Does not require manual data tagging
- Useful for very large data sets

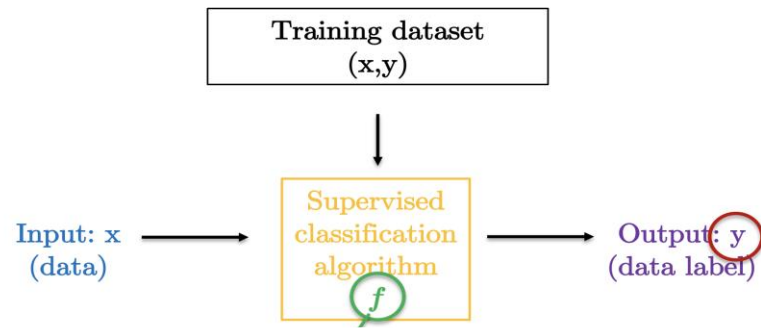
Disadvantages

- We don't know how deals are clustered together or what is it based on
- Requires even more data than supervised classification to be accurate

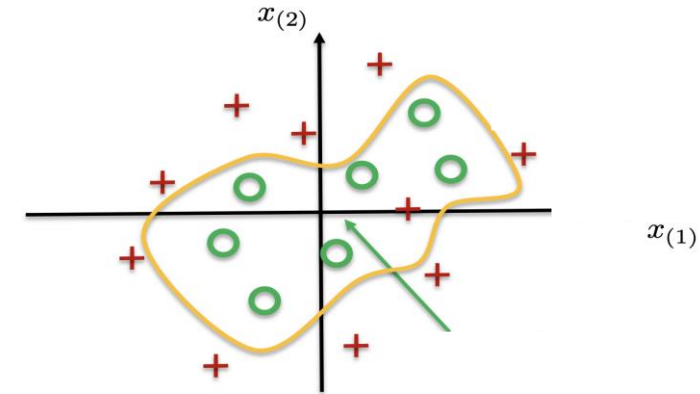
Supervised Classification

Formalization

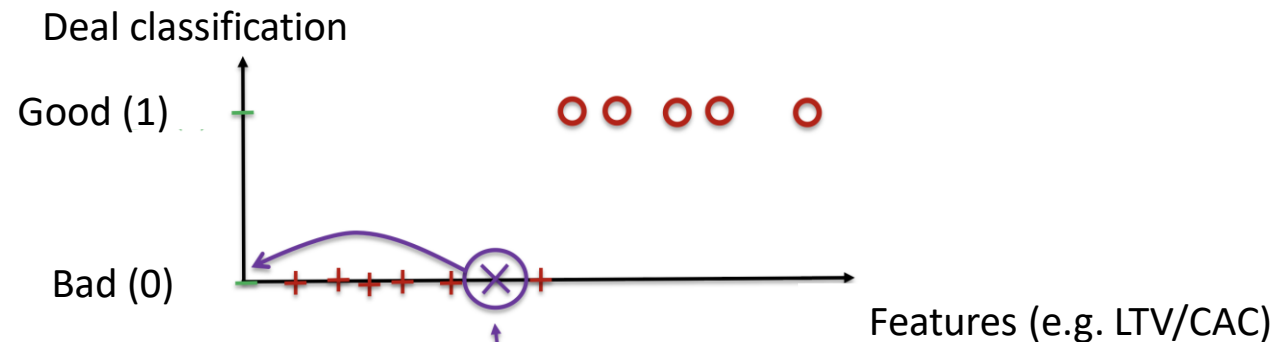
How Supervised Classification works



How Supervised Classification looks



To predict whether a deal is "good" or "bad", given existing features

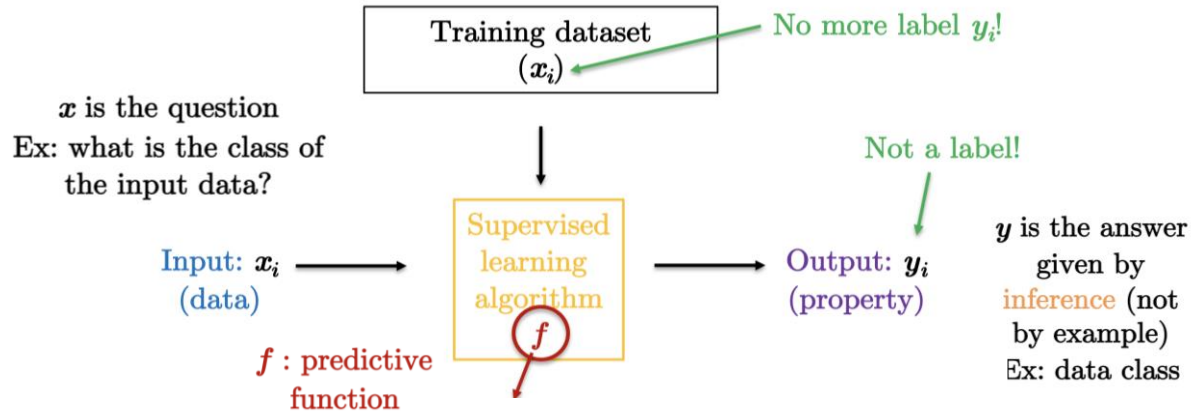


Is the deal good or bad? Supervised classification predicts it is bad.

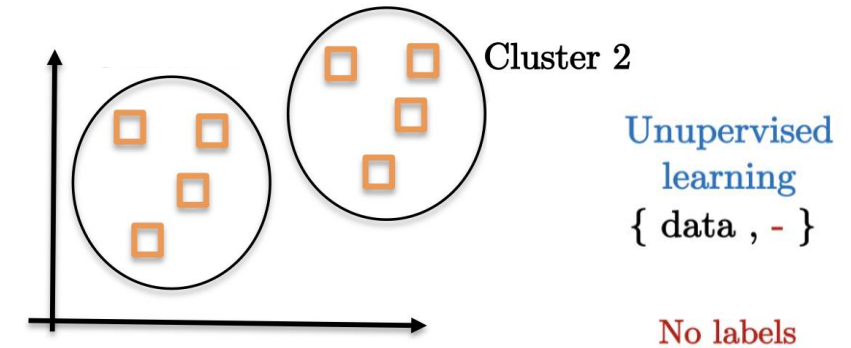
Unsupervised Classification

Design a predictive function given a training set which has no label to find structures in data

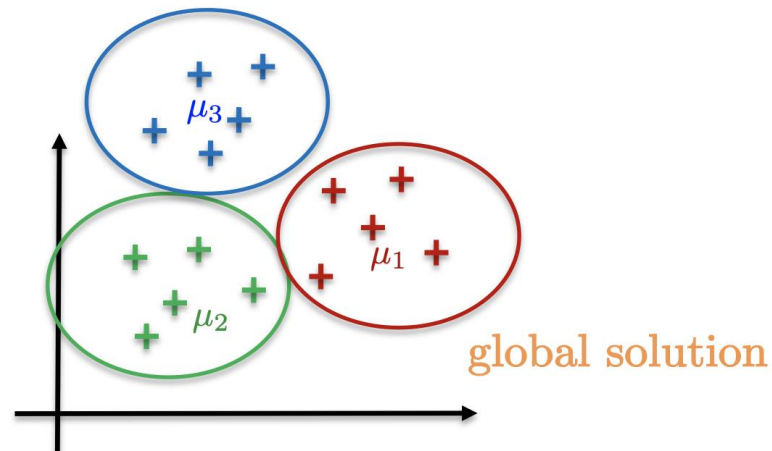
How Unsupervised Classification works



How Unsupervised Classification looks



To cluster deals together based on certain features



In this case, there are three different "clusters" of deals

Evaluation of Data Science and Analytics

Data Science and Analytics can be used in VCs to help analyze deals



However, it requires large sets of data to be effective



Less Applicable for:

Smaller VC firms

Less data available to accurately build models



More Applicable for/on:

Larger VC firms

Growth / Late stage VCs

Late stage start-ups

Greater economies of scale and number of data points of success metrics for building models, i.e. financial data, MAU, DAU

Thesis:

Unsupervised Classification is more Suitable

- Deal analysis is complex and requires large datasets to train
- Too labour intensive for supervised classification

Can be an effective tool given limited areas to analyse

- For example, one can analyse social media statistics to determine a start-up's social media presence and growth

An assistive tool for human analysts rather than replacing

- Data science can be used as an assistive tool to highlight deals with certain success metrics

Application for VCs

Case Studies: Vertex Ventures and (VC Firm)



SEA/Worldwide

Early and Growth Stage

Makes use of data science technologies like:

- Web crawling and information aggregation
- Natural Language Processing, including text categorization, for extracting keywords
 - Classification models using NMF
- Graph database and analysis and etc



North America

Early Stage VC

Built a due diligence bot using:

- Binary classification of 1 (Good) or 0 (Bad) based on success factors to train bot
 - More accurate in predicting failures as time to failure is much shorter than success
- Bot rates 1 – 5 based on investment potential

Key Technologies Identified:

Network Analysis

Web Crawling

Classification Models

Natural Language Processing

Data Analytics Solutions

Data Science Solutions

Introduction

Deal Sourcing

Deal Scoring

Portfolio Support

Fund Operations

Recommendations

Data Science and Analytics for Portfolio Support

Advising and Assisting Portfolio Companies

Case Studies: Georgian Partners and Openspace Ventures

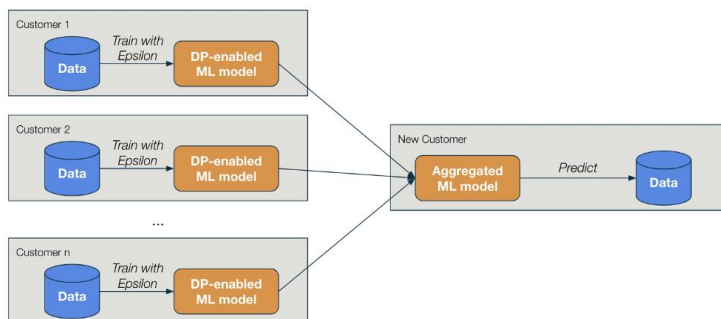


Georgian Partners

Description

- Epsilon enables its portfolio companies to quickly adopt differential privacy methods.

Technology



Description

- Working with portfolio companies on recruitment and projects surrounding Data Science & Analytics
- Advising them on the different models and the approaches

Example of Portfolio Companies



Dr Zhao (OSV): “Data science serves as a differentiating factor for OSV as compared to other funds”

Case Studies: DormRoomFund



DormRoomFund

Aims to make fundraising a more equitable and straightforward process

1

Not having any prior relationships with VCs

2

Lack of access to relevant VCs

3

Challenging to stay abreast of conversations with VCs



Solutions

Analyses VC's social graphs to find mutual connections

Acts as a directory to discover relevant VCs for startups

Central repository to manage the process of reaching out to VCs

Competitor Analysis and Benchmarking

Case Studies: Signalfire and Social Capital



Beacon Software Suite

Actively tracks more than 2 million data sources and half a trillion data points

Portfolio companies can compare their performance against a collection of comparable companies to get a sense of how well they are performing relative to the market



Capital-as-a-Service Model

Founders submit a questionnaire, revenue figures and other data about their company

If Social Capital likes what it sees, it will invest up to \$250,000. If it doesn't, it will at least deliver feedback to the startup

A set of proprietary algorithms quantitatively assess if the company has achieved product-market fit, and how well it stacks up on core metrics against industry benchmarks

Chris Farmer (Signalfire): “Being massively differentiated from other venture firms and having a totally novel story is very helpful”

Case Studies: Signalfire



Talent Platform

- Executive Recruiting
- Technical Recruiting
- Talent Research
- Recruitment Process Optimization
- Leadership Development & Mentorship
- Advisor Network
- Communities

Beacon Talent Software

- Looks at the entire tech industry: over 10M data sources from 6M companies
- Ranks each employee with quality dimensions and provides real-time predictions on how likely they are to switch jobs, helps portfolio companies with identifying the right recruitment fit

In just 6 months from launching Beacon Talent, it was able to place 55 candidates, a quarter of whom on executive level, in their portfolio companies.

Data Science and Analytics for Fundraising

Fundraising Considerations

Currently, there are two major considerations in fundraising for VCs

Sourcing for Investors

Types of LPs

- Friends and Family
- Rich Individuals
- Family Offices
- Wealth Managers
- Funds of Funds
- Endowments
- Pension Plans
- Sovereign Funds

Where to find LPs?

- Websites and Desktop Research
- Proprietary Databases (Crunchbase, etc)
- Founders, especially old-school founders
- Industry and Networking Events

Opportunity for Technology here!

Due Diligence Process

Collect several information into data room

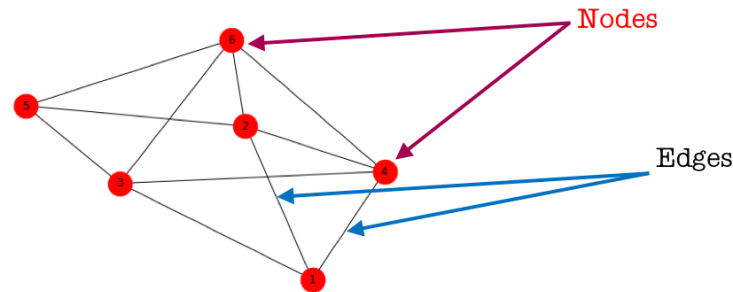
- Fund Information
- Organisation Information
- Team Information
- Market Reports
- Investment Strategy / Process
- Track Record and Case Studies
- Legals and Reference Checks

Several commercial solutions are available!



Sourcing for Investors

Network Analysis can be used to reveal new potential connections or LPs that were unknown



How Network Analysis Works

- Investigates relationships between entities, i.e. between different VCs and startups, and investments between them
- Packages such as networkx for python help identify groups



How Network Analysis can be used here: (Just 1 Suggestion)

Network Analysis



Databases such as
Crunchbase



Unsupervised Learning



Find New LPs and
investments

Other Uses of Network Analysis

Deal sourcing
for startups

Relationship
building
between VCs

Helping startups
fundraise too

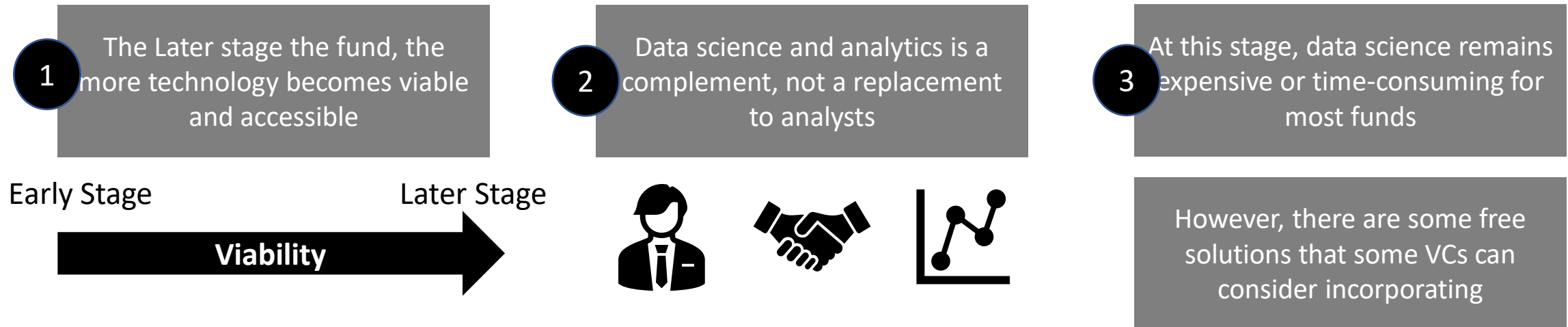
Case Study: Affinity VC



Relationship
Management
Software

Conclusions

General Thoughts and Recommendations



However, funds that can blend data science and analytics into their operations can differentiate themselves and prepare them for better success, be it deal flow generation or achieving uprounds

Recommendations for Protégé Ventures

- 1** A special projects team can work on projects such as network graphs, classification models and others
- 2** Possible to invest in an information management system, given current systems
- 3** Value add to potential and portfolio startups through data mining and providing benchmarks to startups



Thank you! Questions?

References

<https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/a-machine-learning-approach-to-venture-capital#>

<https://techcrunch.com/2019/06/26/vcs-double-down-on-data-driven-investment-models/>

<https://medium.com/@lamRavijot/rise-of-intelligent-and-data-driven-venture-investing-36edad1d4b85>

<https://medium.com/connetic-ventures/singles-or-home-runs-how-we-train-our-ai-early-stage-investing-vc-bot-69e03bb1723c>

<https://medium.com/@georgianprtnrs>

https://en.wikipedia.org/wiki/Dorm_Room_Fund

<https://medium.com/hackernoon/winning-by-eating-their-own-dogs-food-83-venture-capital-firms-using-data-ai-proprietary-da92b81b85ef>

<https://vcpreneur.com/how-vcs-lps-apply-data-science-to-make-informed-investment-decisions-527c917a3edf>

<https://www.startupreporter.eu/how-does-a-venture-capital-fund-work-in-terms-of-operations/>

<https://www.codingvc.com/behind-the-scenes-at-a-vc-fund-part-2-helping-founders-and-time-allocation/>

<https://corporatefinanceinstitute.com/resources/careers/jobs/role-of-investor-relations-ir/>

<https://towardsdatascience.com/data-science-in-venture-capital-8c13ec0c8458>

<https://towardsdatascience.com/social-network-analysis-in-venture-capital-4db03cb5d833>

<https://www.venturecapitaljournal.com/south-east-asia-still-has-goldmine-potential-for-lps/>

<https://pearonline.com/a-checklist-for-producing-a-great-fundraising-data-room/>

<https://www.docsend.com/blog/how-first-time-funds-attract-lp-investments/>

<https://digify.com/case-studies/cocoon-capital-raised-over-21-million-with-digify/>

<https://runway.is/blog/how-to-raise-a-vc-fund/>

<https://medium.com/initialized-capital/how-to-raise-a-venture-fund-735647ca90aa>

<https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374>

<https://towardsdatascience.com/network-analysis-d734cd7270f8>