Project Outline:

Client: David Lu (Golden Gate Ventures, Singapore)

Industry: Venture Capital

Executive Summary:

The project aims to showcase the potential value that automation can bring to the VC industry. We aim to utilise machine learning to assist in sourcing deals, an important aspect of growing the firm's portfolio.

By drawing data to complement our understanding of the startup ecosystem and what makes them thrive, investors can locate potential deals faster than other competitors and understand what startups require financially and strategically based on profiling similar ventures in other regions.

After some iterations we propose two solutions in this preliminary scope; the first being a visualisation of startup movements from the US to other regions, changing industry trends of startup success as well as the breakdown of founders and how that also contributes to a successful startup. The second solution is a data science solution that will be able to identify investment opportunities by feeding data extracted from news and websites into a model to be evaluated.

Completing the project requires a strong team with diverse skills. We would need data analysts proficient in extracting and processing data, software engineers with the ability to build out the product and project managers to ensure that the team works cooperatively towards the goal and within deadlines.

With the right guidance and a strong team, the ANOVA team believes that we will be able to prototype these solutions and demonstrate the potential of machine learning in supplementing the existing processes in the deal sourcing pipeline.

Automation in Venture Capital:

Context:

Sourcing is an essential and demanding responsibility of associates at VC firms. As the number of VC firms grow the available capital across the industry explodes, slowly turning capital into a commodity. Therefore, it is essential the VC firms act fast and accurately in order to secure the best deals with their funds to maximising their future returns.

In the age of machine learning and automation, there is potential for a data science solution is that is able to detect and flag interesting investment opportunities to the VC firm, thereby speeding the sourcing process and capturing any gems that traditional approaches would otherwise miss.

We as students are not expected to build a full-fledged data science solution that automates the sourcing process, but we are expected to compile and utilise relevant data sets, as well as web scraping and web crawling techniques to obtain the essential data that will allow us to build and test a prototype model.

Key Ideas:

- Microsegmentation of startups. The segmentation of the industry, activity and revenue model should be as detailed as possible. For example, on CrunchBase, mobike and Uber are classified under the same category, but they are clearly in different industries with different target markets.

- VC firms invest in good industries, thus modelling trends of investment and their industries. If micro-segmentation of startups is possible, even better. Hence, understanding the market movement of startups is imperative.

- Understand the core relations between cloned startups in different geographical locations (e.g. Didi, Uber and Grab) and the shift of those ideas originating from the US to other locations such as South East Asia and Australia.

- One end solution could be the optimal allocation of funds by investing in startups right before their critical evaluation spike.

- The outcomes of portfolio investments can be either: unicorns, acquired (for unfavourable evaluation), raised money but failed to execute, or still in progress.

Preliminary Scope:

The entirety of the project can be broken down into 4 major steps. Based on our discussions with the client, the desired end solution has 2 key parts.

The descriptive solution involves gathering insights on successful investment decisions by analysing industry trends. With a solid high-level understanding of the industry, this will strengthen the model built in the 2nd solution.

The predictive solution is a model used to evaluate potential startup data gathered from the internet, and functions to speed up the sourcing pipeline and identify opportunities that might otherwise be missed.

Once a MVP is completed for both products, the final phrase is moving both projects off into production, handing it off to a (full-time) team to build the complete product.
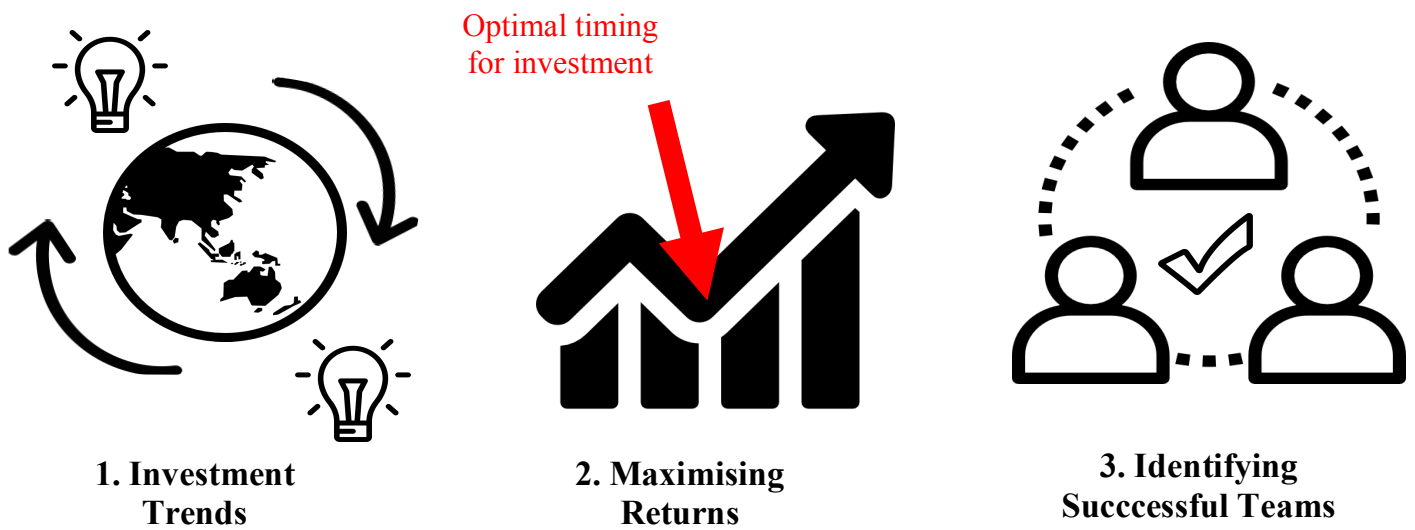
Phase 1: Data Preparation

- Extracting raw data for processing

- Microsegmentation of startups to estimate missing values (e.g. pre-money) in databases.

- Selecting features to be used in data visualisation and predictive modelling

- Building reliable databases with the desired metrics for the next stage

Phase 2: MVP Predictive Analytics – Potential success of early stage startups

Phase 3: Moving Into Industry-level Production

Preliminary End Solution

Part 1 - Descriptive Analytics

1. Investment Trends

Optimal timing for investment

2. Maximising Returns

3. Identifying Succcessful Teams

1. Identify successful investments from more established startup ecosystems in order to be financially prepared for investment into similar ideas that will migrate to SE Asia from US and/or China
    - microsegmentation of industry and ideas of startups
    - appropriate amount and stage to invest
    - time taken for migration of ideas to occur

2. Identify the % of 10x or greater returns in a particular industry and investment stage in a particular time length (e.g. 3-5 years)
    - Investment stage; when evaluations tends to exponentially grow
    - Industry; time periods where certain industries are hot investments and startups in those industries have evaluations shooting up.
    - Region; the correct timing to investment in local startups that have the environment and capital to grow

3. Identify the team composition of the founders of 10x or greater returns
    - # of founders
    - experience and track record in relevant industries
    - past successes with startups

Part 2 - Predictive Analytics
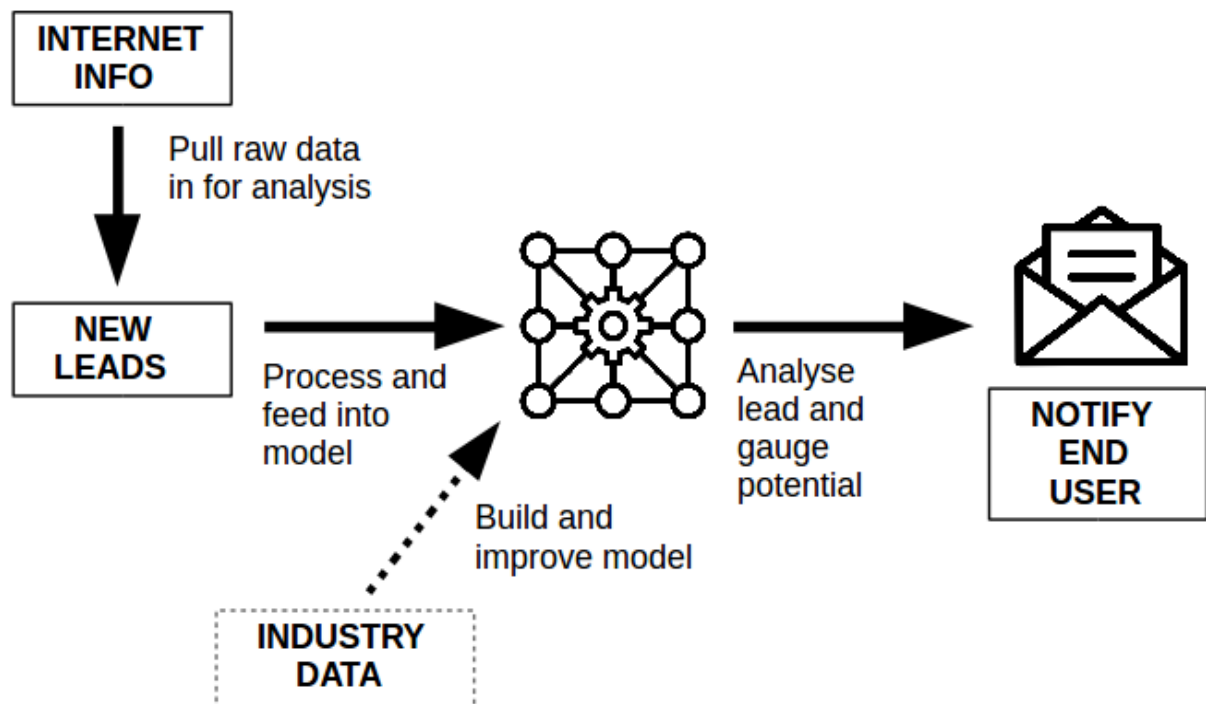
The predictive solution involves building a model to identify investment opportunities through retrieving information from the internet to evaluate. The model then determines whether the startup is worthy of attention and passes the information to the end user.

An alternative is to estimate the growth of potential startups and provide an estimate for the growth in evaluation (e.g. in 2 years).

<u>Core Functionalities:</u>

1) Machine Learning Algorithm

The model requires vast amounts of historical VC investment and evaluation data in order to build, train and test the model.

2) Web Crawling for Raw Data

Develop a web crawling software to retrieve data regarding startups on the Internet to identify opportunities and automate sourcing.

Resources:

https://en.wikipedia.org/wiki/Venture_capital (good starting point)

https://en.wikipedia.org/wiki/Seed_money (intro to seed funding)

https://hbr.org/1998/11/how-venture-capital-works (excellent next-article read)

https://www.quora.com/What-are-unconventional-ways-to-get-an-entry-level-job-at-a-VC-firm-I-dont-care-about-title-or-compensation-I-just-want-a-shot-at-helping-source-healthcare-deals-I-am-willing-to-work-for-free-for-a-trial-period-and-am-willing-to-be-an-intern (good read for entry-level students wanting to breka into the VC industry)

https://www.quora.com/profile/Crystal-Huang (broad collection of Quora responses by an associate at a VC firm called GVV capital)

https://www.linkedin.com/pulse/using-venture-capital-firms-ahmed-serougi/

https://wenku.baidu.com/view/8e94c0f3770bf78a652954bd.html (paper on VC and structure of capital markets)

https://www.ft.com/content/dd7fa798-bfcd-11e7-823b-ed31693349d3

https://techcrunch.com/2017/12/21/fly-ventures/

https://www.ft.com/content/dd7fa798-bfcd-11e7-823b-ed31693349d3

https://outsideinsight.com/insights/outside-insight-in-venture-capital/