**Capstone Project Proposal:**

**Predicting start-up success**

***Idea:*** Working in the world of M&A data aggregation, I am constantly exposed to start-ups and their evolution. What always surprises me is how some start-ups manage to get acquired at lofty valuations, sometimes in the billions of dollars, while other start-ups disappear entirely. Take Snapchat – now known as Snap Inc – for example. This messaging app is about to undergo what is predicted to be the biggest tech IPO of the year at a valuation of USD 25bn. By contrast, Barnes & Noble, probably the most well-known bookstore chain in the US, with its 130-year history and close to 640 brick-and-mortar stores around the country, has a market cap of USD 760m, 30x times less than the expected market cap of Snapchat! Snapchat’s idea was original to some degree but not unique. There had been start-ups alongside Snapchat that had proposed similar ideas: Wickr, Hash, Ansa, Secret. Wickr still exists, but the other three have disappeared. The million-dollar question is: who could have predicted Snapchat's success when it was founded, and who could have predicted Secret's failure when it was launched? This question has always been of huge interest to me, and I am hoping to answer it with the data science tools that I am acquiring in the course.

**Target audience:** This project is intended for anyone looking to start a business. Which factors should they consider when looking for financing? How many funding rounds are optimal for success? Which investors have a track record of helping companies succeed? Are add-on acquisitions conducive to growth? Most importantly, is it even possible to forecast success given observable parameters?

**Data:**

The project will be based on two data sources for the most part:

* Crunchbase data on start-ups. The data is broken down into three separate data sets: company profiles, funding rounds from investors, and acquisitions made by start-up companies.
* Potential data: web scraping using Hacker News and Techcrunch. This data set will be used to assess the PR aspect of running a start-up. Does a lot of publicity lead to success?

**Approach:**

I will attempt to forecast success or failure of a given start-up using multiple different parameters. Success is determined by either a company getting acquired or undergoing an IPO. Failure is considered to be bankruptcy or other non-bankruptcy-related closure.

The following parameters will be used to build the forecasting model:

* Investors backing the company
* Number of funding rounds prior to success/failure
* Total funding received
* Size of the first round
* Number of add-on acquisitions made overall
* Number of add-on acquisitions in the first and second year of being founded
* Number of mentions in Hacker News or Techcrunch in the first two years of being founded.
* And other potential parameters to be considered during the course of the analysis.

**Challenges:** There are multiple challenges with this project as currently foreseen:

1. The most obvious is the need to clean the data for the project. There are a lot of missing entries in the downloaded data set. A few wrong entries have been spotted as well. Finally, date manipulations will be required to extract date and interval information.
2. Web scraping is another challenge, which might prove to be too time-consuming for the project.
3. However, the most difficult part of the project will probably be dealing with the time series aspects of the study. We need to account for the fact that some start-ups might have closed due to time-related effects (Great Recession, post- dot-com bubble, etc.). The challenge here is to be able to forecast start-up success/failure while only having historical, time-dependent data available.

**Deliverables:**

* Final report with the findings
* PowerPoint presentation
* Jupyter Notebook with the underlying code

**References:**

The study question is interesting, but it is not new. Several studies have been done assessing factors influencing start-up growth. Previous studies have focused on education vs. experience of entrepreneurs, management style, cash-burn-rate in the first year, and even the name of a start-up. Most of the studies are based on surveys or on data focusing on specific geographies. However, no expansive, predictive study using Crunchbase comprehensive dataset is known to date.

Selection of previous studies:

[**http://science.sciencemag.org/content/347/6222/606**](http://science.sciencemag.org/content/347/6222/606)

[**https://hbr.org/2016/05/4-factors-that-predict-startup-success-and-one-that-doesnt**](https://hbr.org/2016/05/4-factors-that-predict-startup-success-and-one-that-doesnt)

[**https://mackinstitute.wharton.upenn.edu/2016/startup-survival-and-a-balanced-burn-rate/**](https://mackinstitute.wharton.upenn.edu/2016/startup-survival-and-a-balanced-burn-rate/)

[**http://www.ivc-online.com/Portals/0/RC/Startup%20Success/Startup%20Report%201999-2014\_PR\_Final.pdf**](http://www.ivc-online.com/Portals/0/RC/Startup%20Success/Startup%20Report%201999-2014_PR_Final.pdf)

[**http://www.microtheory.uni-jena.de/research/projects/research-project-start-ups/**](http://www.microtheory.uni-jena.de/research/projects/research-project-start-ups/)

[**http://www.aebrjournal.org/uploads/6/6/2/2/6622240/4\_makale\_critical\_success\_factors\_of\_the\_survival\_of\_start.pdf**](http://www.aebrjournal.org/uploads/6/6/2/2/6622240/4_makale_critical_success_factors_of_the_survival_of_start.pdf)

[**http://econpapers.repec.org/article/nweeajour/y\_3a2014\_3ai\_3a3\_3ap\_3a15-24.htm**](http://econpapers.repec.org/article/nweeajour/y_3a2014_3ai_3a3_3ap_3a15-24.htm)