Predicting Startup Success Based On Funding

Literature Survey Report

Vibha Satyanarayana*, Sruthi V†
Department of Computer Science, PES University
Bangalore, India
Email: *vib.satya@gmail.com, †vsruthi98@gmail.com

Abstract— **In the recent years, massive growth in startup economy of India has attracted investors all around the world. But statistics [7] show that 9 out of 10 startups tend to fail. Top reasons [1] for these numbers is lack of market need, no investor interest and being outcompeted. This paves way for the requirement of the entrepreneur to know what are the factors that will affect his or her startup. Startups fail when they are not solving a market problem, and are instead tackling an interesting problem, which goes on to show the importance of market analysis location wise. We propose to tackle these problems by analysing current market trends to provide suggestions to startups as to which investors they can approach in their domain, the upcoming or established competitors in their city, the range and type of investments that their industry vertical typically receives. Apart from these factors, we are also trying to analyse founder credentials, such as skill set and educational background to name a few, and to see if these affect the startup funding in any way.**

I. Introduction

Startup entrepreneurship is crucial because it brings innovations, new jobs and competitive dynamics into the business environment. Most of present day employees who are laid off due to downsizing or are just in search of a change after dealing with the same things every day for too many years turn towards startups. But they have no insights on current market trend, need for this startup in their city, viability of the market for the service they're providing and the major investors around them. Apart from knowing their skills and aptitudes, knowledge about their competitors in the same domain can help them determine what their ideal working environment is moving forward.

Almost half of startups that fail do so because in the end, not enough people want the service they were producing. Even if the idea is worth a lot and is an interesting problem to tackle, the entrepreneur has to know if idea will work given the current place and time, was their work being done in this field, is it widely appreciated or entirely new and what peers have already faced with respect to funding. A budding businessman will be completely oblivious to the investors that were currently popular and the type of investment they provided. Knowledge about these seemingly trivial things will enhance the startup majorly and will not let ignorance be the cause of fail.

Studies [11] show that 42% of the startups failed for lack of a robust market for what the company provided and running out of funds accounted for another 29%.These numbers only add to reasons why building a model to examine startup funding is required right now. Not only will this analysis help individuals, but we can also take a closer look at the startup economy of a city and the scope of a field at that point in time.

The layout of this paper is as follows. In the next section we look at related work being done for a similar problem layout. Startup failure has been the center of study for various researchers, and to build a model to predict the outcome of investment of time and money will be beneficial to both the startups and the investors. We then proceed to our problem statement in the later sections and show how our approach takes a different route.

II. Related Work

A. Seed funding amount and time

One of the work done in this area [6] discusses predicting the outcome of a startup based on factors such as seed funding amount and seed funding time. The data was extracted from CrunchBase and preprocessed. The time period of the data is from 1999 to 2014.The authors have taken into account 20+ factors, out of which a unique one was a ranked list of positive and negative severity factors such as determination, execution, social skills that fall into the former and no flexibility, no roadmap that fall into the latter. Though the ranking order with a numerical value for these severity factors is provided, there is no clear mention as to how these are calculated, the authenticity of these numbers and the approach taken to gather such data (example discipline, bad luck, etc.).

The authors have experimented with many classifiers and have used the ROC curve to evaluate the best model. One of the atypical classifiers used was the ADTree (Alternating decision tree), where knowledge contained in the tree is split across multiple paths. A positive sum represents a one class setting and a negative sum represents a two class setting. Totally 9 models were built, and each model tried various classifiers including Naive Bayes, Simple Logistic, Random Forest, etc. Model 0 was built for no seed funding. Every higher model from then on included the lower model factors and one type of seed funding. Comparing both precision and accuracies has given fairly good results for the highest model, model 9 (precision of 96%). The results of their model was tested against current startups in the market for both success and failure. The examples provided by the authors were for 2 startups for which their model worked perfectly, that is, predicted accurate success probabilities.

B. *Crowdfunding*

Another work [4] focuses mainly on crowdfunding, that is, funding a project or venture by raising many small amounts of money from a large number of people. One of the assumptions made by the authors was that the factors during or after the funding process were not as important as the ones before the actual funding. Some of the notable attributes that were scraped or calculated are the number of facebook friends,twitter followers and sentiments.

The authors have used various classifiers including support vector machines with different kernel functions and logistic model trees. An interesting component used is the AdaBoost [12], short for ”Adaptive Boosting”, is a machine learning meta-algorithm that can be used in conjunction with many other types of learning algorithms to improve their performance.The result of their analysis showed that simple decision trees was the best classifier. But a maximum accuracy of 67% has led the authors to believe that there might be a ”hidden variable” that has not been taken into account. The timed results of this work show that complex models like logistic model trees taken a large amount of time but do not show a drastic change in accuracy as compared to simpler models such as logistic regression.

C. *Pre startup success*

This paper [10] focuses on estimating the relative importance of a variety of approaches and variables in explaining pre-startup success.The authors organized a variety of approaches and variables that are possible influences in terms of Gartners (1985) framework [3] of new ven-ture creation.This framework suggests that start-up efforts differ in terms of the characteristics of the individual(s) who start the venture, the organization which they create, the environment surrounding the new venture, and the process by which the new venture is started. The authors derived possible success and risk factors from each of these dimensions. For this paper, a random sample of 49,936 phone numbers was dialled and an interview was held at a six month, one year, two year and three year interval after initial screening.

Their results showed that few of the nascent entrepreneur’s characteristics are directly associated with success. Most of the significant findings relate to the environment: start-up capital and risk of the market are seen to be the most important features. Startups starting off with a smaller capital were more likely to succeed. None of the included individual characteristics distinguished successful nascent entrepreneurs from the unsuccessful ones. They also found that the amount of time put in a business was a direct success measure. Therefore full time startups were more likely to succeed than part-time. Another finding was that people in manufacturing more often got started than people in other sectors.

The authors only analysed the direct effects on startup success. Indirect variables such as gender of the entrepreneur were not taken into consideration although they could be influential. Secondly, not all approaches proposed in their theory section were studied,e.g.psychological approaches. Moreover, their data was far from perfect. Because it was collected at a single point in time, the number of months the entrepreneurs took to prepare for the business varies. Lastly, the study and the data is outdated and cannot be directly applicable to today's market.

D. Cognitive abilities of individuals

In this work [5], the authors focused on the cognitive abilities of the individuals receiving funding and found that funded teams were strongly correlated (R = .875; p=0.000) with higher levels of predevelopment meticulousness (i.e. attention to detail and up front planning), social influence and were less risk averse.

Data was collected by conducting interventions at 11 large companies over a time span of 12 weeks. The employees were asked to form a team of 3-6 members and come up with a business plan and attempt to obtain funding for their project. Based on the results, the following four hypotheses were made:

Hypothesis 1: Pre development meticulousness will be positively related to subsequent funding.

Hypothesis 2: The greater the teams' adeptness with respect to persuasion and other techniques for exerting social influence, the greater their financial funding.

Hypothesis 3: The higher the aversion a team has to risk, the lower the amount of funding the team will secure.

Hypothesis 4: The higher the learning motivation of the team, the greater their financial funding.

A standard multiple regression was performed between the amount of funding secured as the dependent variable and predevelopment meticulousness, social influence, risk aversion, and learning orientation. They found that three out of the four variables predevelopment meticulousness, social influence, and risk aversion explained significant amount of the variance surrounding funding and were in the predicted direction. The fourth variable, learning motivation, had a negative regression weight and was thus not significant for the hypothesis, which predicted a positive relationship. The sample size and project time limit generalization. The predictions made in this research may not be true for larger samples and higher risk projects. As the authors mentioned, they could have overlooked certain potentially important factors in their study as their intervention was restricted by the setting and scope.

E. Indian startups data

This article [9] reports findings on Indian startup data. Data was collected on every tech startup in India that ever raised more than $2M in funding since 2005. This covered 448 companies and 987 founders.The data set covered parameters such as undergrad alma mater of the founders, last company before they started their entrepreneurial journey, sectors, exits, and work experience. This data set was primarily created using the data available via CB Insights, Tracxn, Linkedin, Venture Intelligence, and Crunchbase.

The authors broke up the analysis along a few major dimensions work experience, academic background, sectors, business model choices, fundraising and exits (transitioning the ownership of one company to another). They found that in e-commerce, a majority of founders were associated with the startup ecosystem in some capacity before founding their own companies. Their data also depicted the current trend in choice of business model. This showed that B2C (Business to consumer transactions), which was generally believed to dominate the Indian Startup landscape, did not do so. Another interesting finding was that vast majority of the exits seemed to reflect companies with experienced founders.

The authors' main focus was the experience of the founder. This is an important factor, but it need not be the primary variable to predict success. There are several examples of successful startups with founders having very little work experience.

F. Work done on the same dataset

Most work done on kaggle kernels is visualization. The data was understood better and interpreted. For example, Jaghadish Rajagopalan [8] showed that the number of investments per month plot showed that mid 2015 and mid 2016 had more funding and has been marginally declining in recent days. The author also found which type of funding was more popular, which city attracted more investors, the current leading industry vertical and the top investors in the Indian ecosystem.

DougDalys work [2] concentrates on the investments made. The author found that the investment trend was steady with recent upward spikes. It was also seen that half of the investment came from the top 12 investors out of 1744 investors.

III. Problem Statement

We intend to address the issue of startup funding and suggesting investors for startups, based on their industry vertical. Budding entrepreneurs will benefit by knowledge about their peers in the same sub vertical as theirs, the type of investment that is popular for their sector and the range of investment that this particular domain receive.

We will build a model to analyse startup funding location wise, year wise, funding type wise and industry vertical wise and experiment with classifiers for best results. Apart from the existing data, we have managed to scrape data regarding the startup founders from other websites. This will help us look at founder credentials, such as educational qualification, university information, skill set and other factors. Text analysis will help us decide if good quality educational background of the entrepreneur correlates to better success in their startup in terms of funding.

A. How is our approach different form what has already been done

A significant amount of analysis on startup statistics in India has not been done. It is important to study Indian startup data because of its rapid growth rate and contribution to the economy. This is what we intend to do in our project. Most of the research work done so far seems to focus on predicting success based on the funding type or startup surroundings. Although these are important factors to consider, factors such as the founders, location and skill set greatly influence the results. In this project, analysis on the founders of the startup will be done to determine if their educational background plays any role in the success of their company. We will also compare different startups across India to determine if location plays any role. Apart from this, the type of funding, industry vertical and subvertical trends will also be studied. This information would be important for investors looking to invest in a potentially successful company.

References

[1] The top 20 reasons startups fail, cbinsights.

[2] DougDaly. Indian startup data exploration.

[3] William B. Gartner. A conceptual framework for describing the phenomenon of new venture creation. 1985.

[4] Michael D. Greenberg, Bryan Pardo, Karthic Hariharan, and Elizabeth Gerber. Crowdfunding support tools : Predicting success and failure. 2013.

[5] Peter A. Koen, Gideon D. Markman, Robert A. Baron, and Richard Reilly. Cognitive mechanisms: Which ones allow corporate entrepreneurs to obtain startup funding. 2005.

[6] Amar Krishna, Ankit Agrawal, and Alok Choudhary. Predicting the outcome of startups: Less failure, more success. 2016.

[7] Neil Patel. 90% of startups fail: Here's what you need to know about the 10%.

[8] Jaghadish Rajagopalan. Simple exploration notebook-indian startups.

[9] Bala Srinivasa and Yash Jain. Heres what the data tells us about tech startup founders. 2017.

[10] Marco van Gelderen, Roy Thurik, and Niels Bosma. Success and risk factors in the prestartup phase. 2005.

[11] Kimberly Weisul. What went wrong: 101 failed startups tell all.

[12] Wikipedia. Adaboost.