Vrushank Shah

Predicting the success of a startup company.

Abstract:

More than 50% of startup companies fail in the initial four years. Further, three out of every four venture-backed firms fail. The algorithm proposed in this paper will help to predict the success of a startup company based on financial and managerial variables. This prediction will help investors to get an idea whether the investing in a startup will be successful or not? Apart from implementing model consisting of all the factors mentioned below and predicting the success of a startup company, various other models will be created representing various milestones achieved by the company. This paper will help the startup companies to know which factors are essential for getting an investment. The algorithm will be based on more than 15,000 companies' data collected from crunchbase.com. The financial variables include: Investments in each funding rounds, valuation after each round of funding, Current market value, Total funds, Investments and acquisitions by the company, financial background of key people and the managerial variables includes: Number of employees, Competitors, Location, Age of the company, Founders background, Burn Rate and various news articles on the company scrapped from internet. A variety of methods will be used to determine the best model such as random forest, text parsing, logistic regression, decision tree and Survival analysis.

Introduction:

According to Adora Cheung, co-founder and CEO of Homejoy," Startup is a state of mind, It's when people join your company and are still making the explicit decision to forgo stability in exchange for the promise of tremendous growth and the excitement of making an immediate impact." More than 100 million startups are launched per year, which is about 3 startups per second. But more than 50% of startups fail in the initial four years. There are various reasons for a startup to fail for example lack of focus, Raising too much money too soon, lack of general and domain-specific business knowledge, etc. There are very few studies which are performed to understand the reasons for the success of a startup company. There are various articles which describe the reason for failure for the startup companies but without the backing of data. This paper tries to create an accurate predictive model to predict whether a startup firm will succeed or fail.

The data for this paper was taken from crunchbase.com. More than 15,000 companies data were analyzed and used to build a model. All the companies that started between 2000-2014 were used in this paper. The key factors included the amount of seed funding, the time taken for seed funding, the company's valuation and other managerial variables. The results were explained using the survival model and Logistic regression.

Section 1-

Key factors involved:

1. **Seed funding:** It is a form of initial investment by an investor in a company in exchange for an equity stake in the company. This is the most important stage for a startup firm.
2. **Series funding:** After seed funding is achieved the startup companies start getting series funding. There could be series A,B,C,D,E,F,G fundings. Here the alphabets correspond with the development stage of the companies that are raising the capital.
3. **Rounds of funding:** Total rounds of funding received by a company.
4. **Time to get seed funding:** Getting seed funding is very essential for a startup firm but no company will like to wait for months to get its initial funding. More the time it takes for a company to get the initial funding the value of its product decreases. This parameter measures the number of months required for a startup company to get seed funding.
5. **Valuation after each round of funding:** Valuation after seed funding is calculated using the formula: (100*(Seed amount)/15). Valuation after series A funding is calculated using the (100*(Series A amount)/8). Valuation after series B,C,D,E,F,G is calculated using the formula: (100*(Amount)/5).
6. **Number of Milestones:** Number of achievements by the company

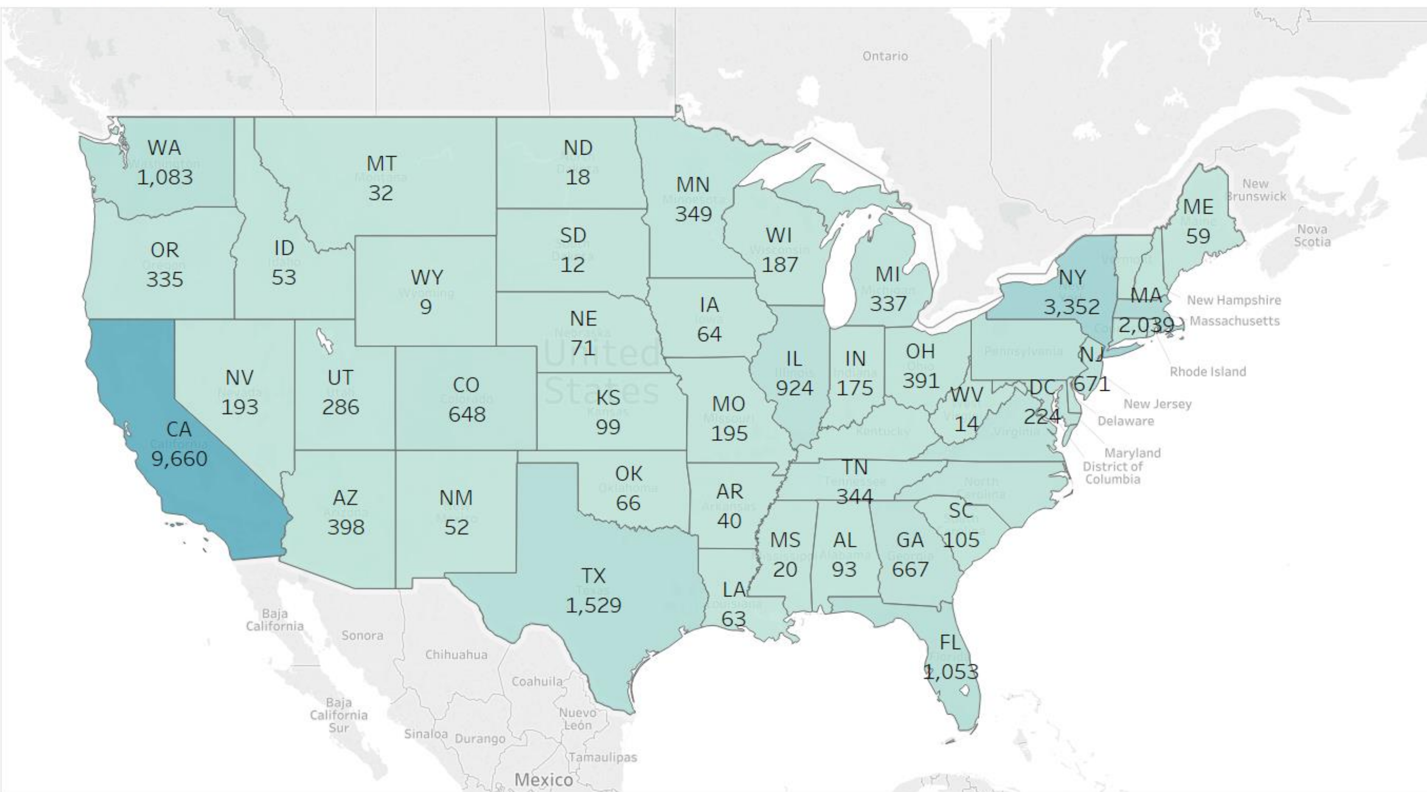
7. **Average time taken to achieve each milestone:** This parameter gives the average months taken to achieve each milestone. If the number is huge then it is a bad thing as it indicates that a company is taking the longer time to achieve milestones.
8. **Average time taken to achieve funding:** Number of months taken to receive each round of funding.
9. **Region:** The city where the company is located.
10. **Degree:** The highest education completed by the core-committee of the company.
11. **University:** The university from which the highest education was completed by the core-committee members.
12. **BurnRate:** Amount of time taken by the company to burn all its funds. It is calculated using the formula: Total fund/Number of months company was active
13. **Total funding:** The total amount of money received by the company.
14. **Category_code:** The domain of the startup company.

Section 2-

Data exploration:

a) Number of startup firms across USA between 2000-2014. The highest startup firms were from California state followed by New York.

Startup companies across USA

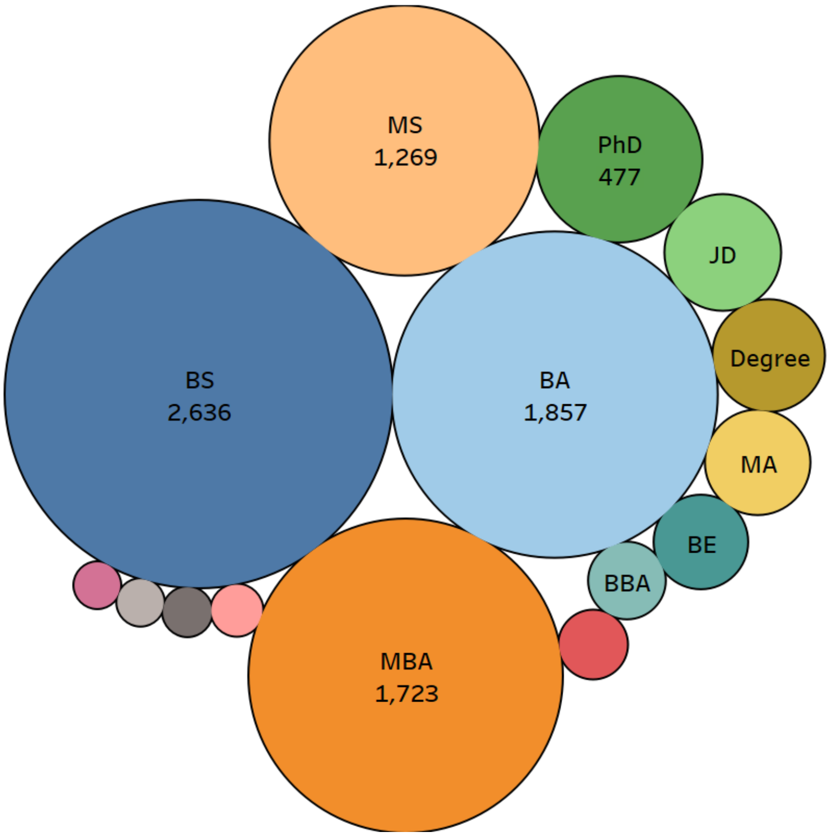


b) University and degrees of core-committee members of the startup firms (top 15):

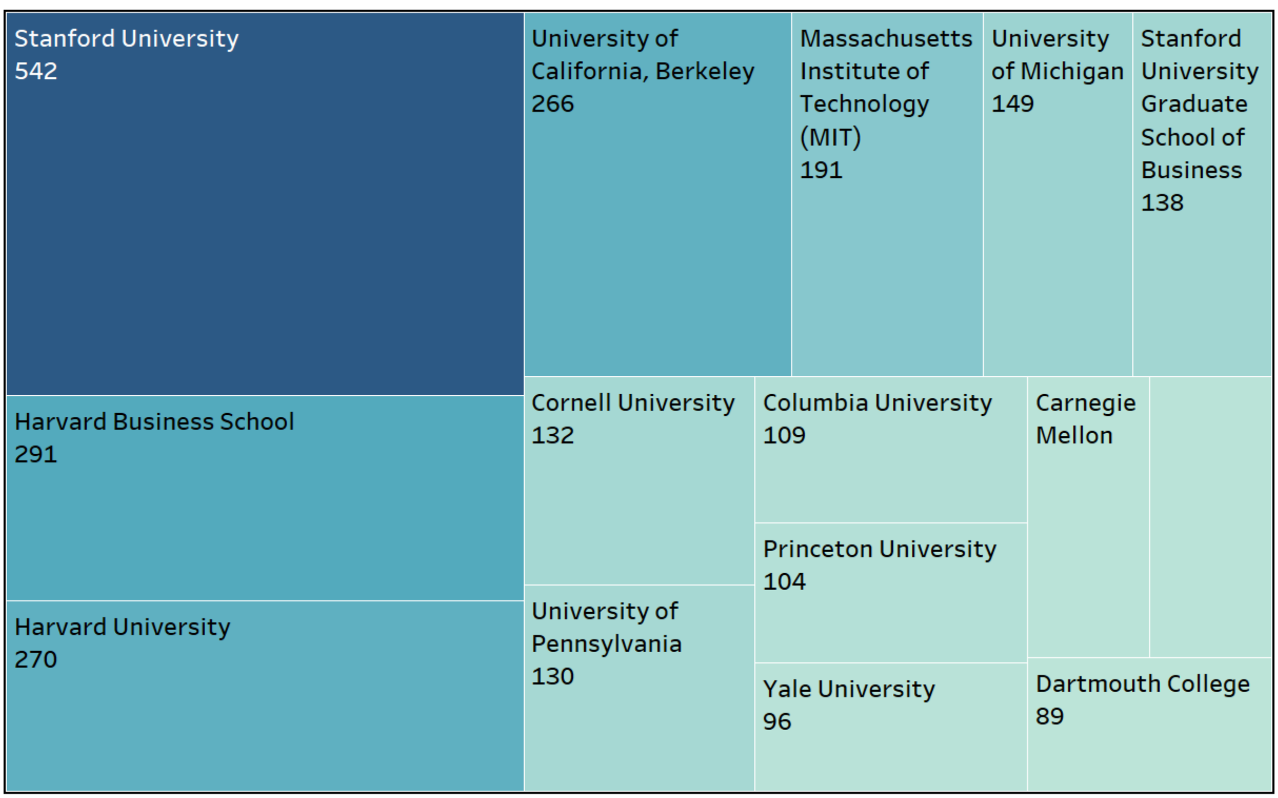
Here the bubble graph represents the highest degree achieved by the core-committee members of the startup companies. BS is the highest followed by BA and MBA. And Stanford university has highest number of CEO,CFO and founders.

Vrushank Shah

Sheet 1

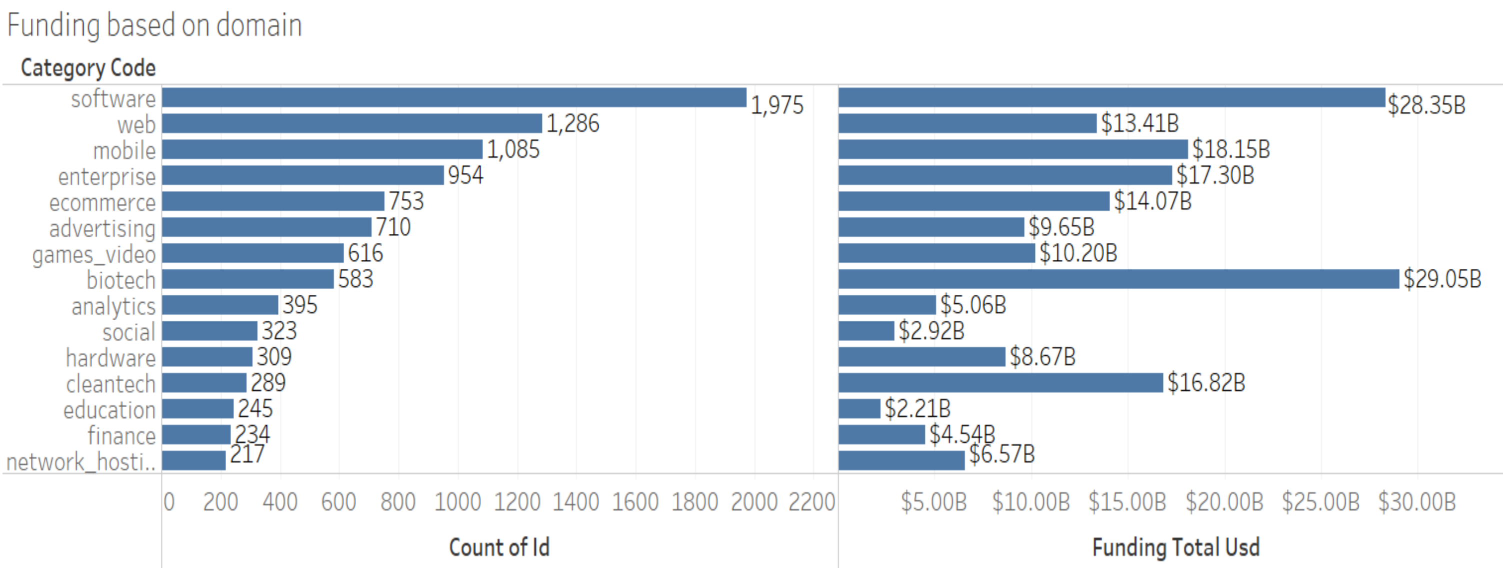


Universities attended by core-committe members



Vrushank Shah

c) Category: The below graph illustrates funding received by startup companies under each domain. The highest number of startup firms were under software domain but the highest funding was achieved by companies under biotech domain.



d) Summary for rounds of funding:

According to the table, 11608 companies were able to achieve seed funding and 7999 of companies were able to reach till series A funding. The highest amount was raised for series B funding. As the company develops the amount of funding starts increasing hence the amount for seed funding is less and from series A onwards the total raised amount is more even though the number of companies decreases.

Summary of funding rounds

Funding Round..	Startup firms	Raised Amount
seed	11,608	$9.1B
a	7,999	$54.2B
b	4,892	$56.2B
debt_round	3,409	$41.8B
angel	3,239	$2.3B
partial	3,115	$9.1B
c	2,499	$44.0B
d	1,129	$26.8B
private_equity	1,043	$26.0B
grant	776	$4.9B
e	430	$12.5B
convertible	187	$0.2B
f	145	$4.9B
crowd	111	$0.2B
post_ipo_equi..	80	$12.3B
secondary_ma..	16	$0.5B
g	13	$0.7B
post_ipo_debt	7	$2.5B
crowd_equity	3	$0.0B

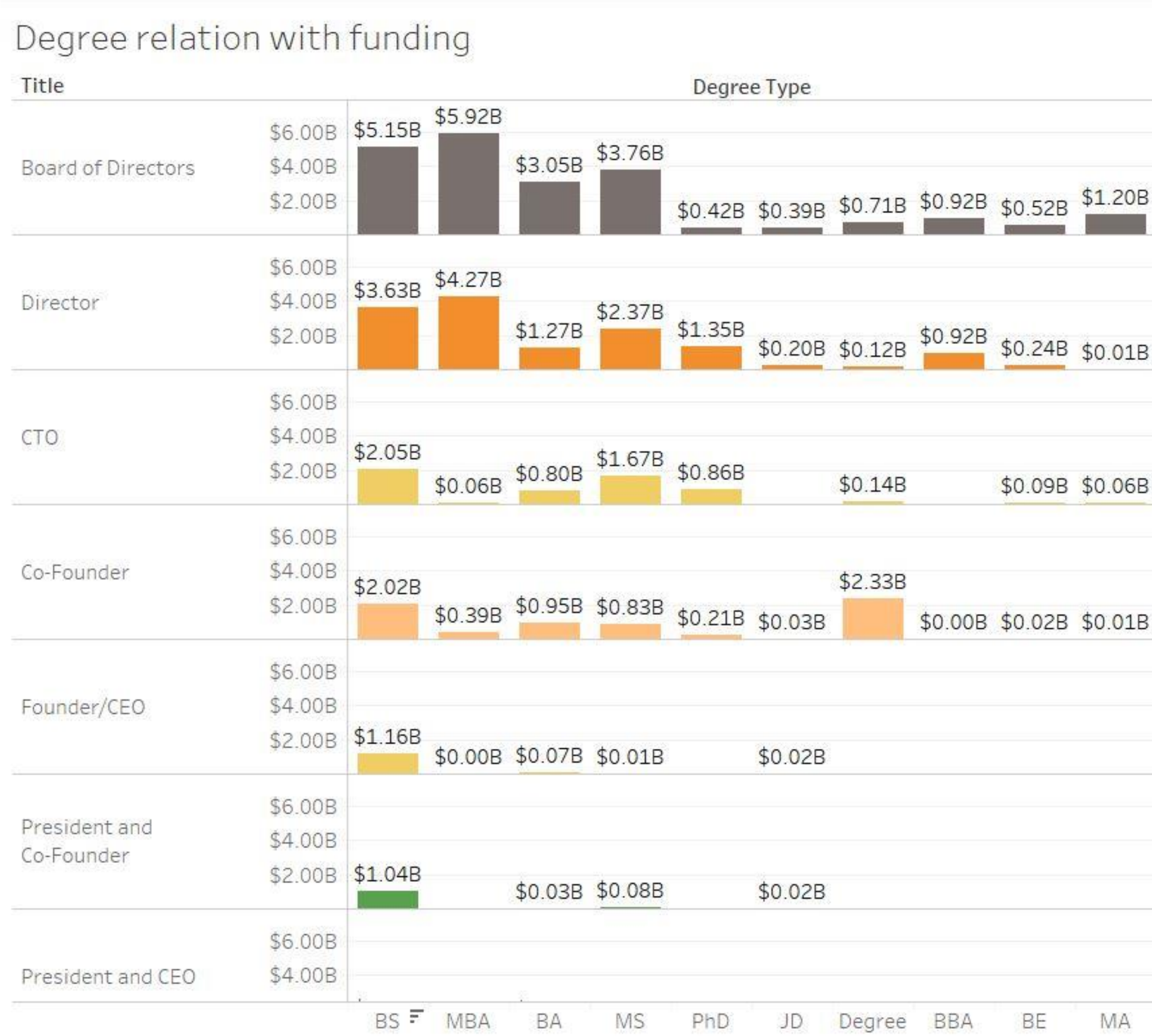
Vrushank Shah

Section 3-

Relation between success of a company and degree achieved by important people of the company:

To find the relationship between the degree and success of a company I have divided this analysis into 2 parts:

1) Descriptive analysis was done between the effect of degrees on funding. Here top 10 degrees were selected based on the total count of people who had received these degrees.



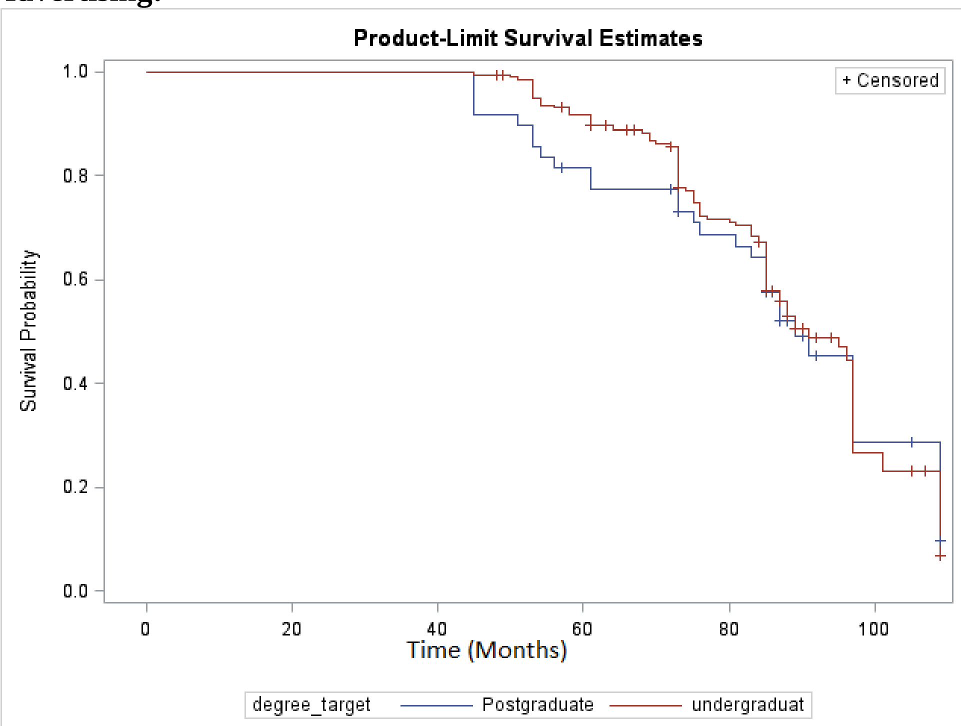
From the above bar chart, we can state that if the founders or the important people of the startup companies have a postgraduate degree like MBA, MS, Ph.D., JD then it is easy to get funding. If the company is getting good funding, then it means that it is successful. To find out whether degree effects the survival chances of a company I have built a survival model based on the postgraduate and undergraduate degree of important people in the company.

Vrushank Shah

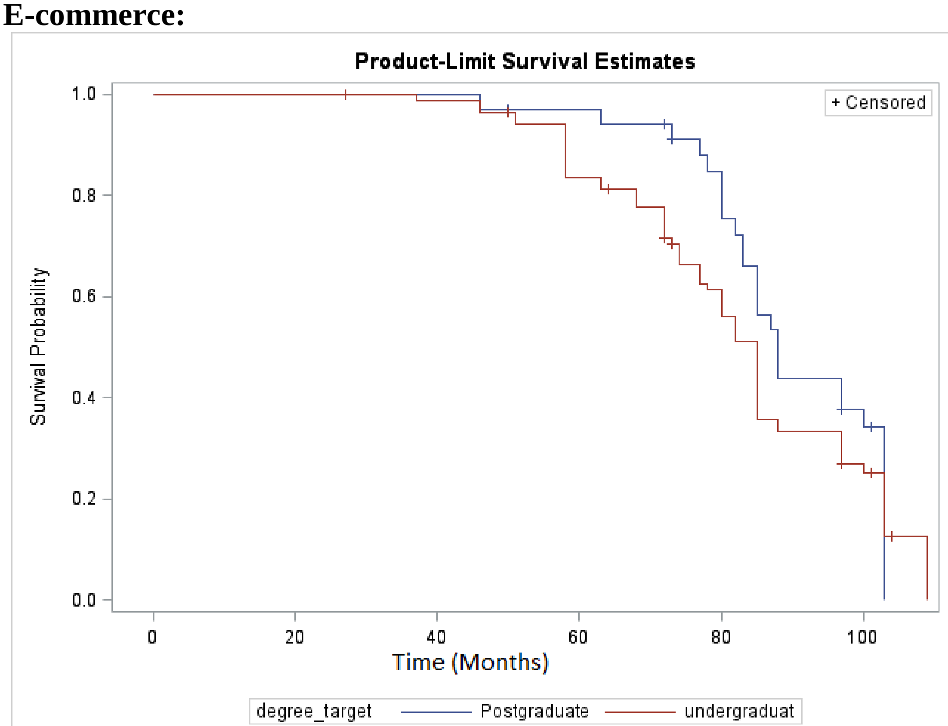
2) Survival analysis:

This analysis was done to find whether there is an association between the degree of the core-committee people and the survival of the company. Here red line means core-committee people have a postgraduate degree which includes Ph.D., MS, MBA, MD, etc and the blue line indicates core-committee people has an undergraduate degree which includes BS, BE, BA, etc. The survival model for each domain is presented in this paper. Here Advertising, E-commerce, Enterprise, Analytics, Software, Mobile, Video and web domains are selected to perform survival analysis. Since these domains have the highest number of startup companies which will help to increase the accuracy of survival model.

Advertising:

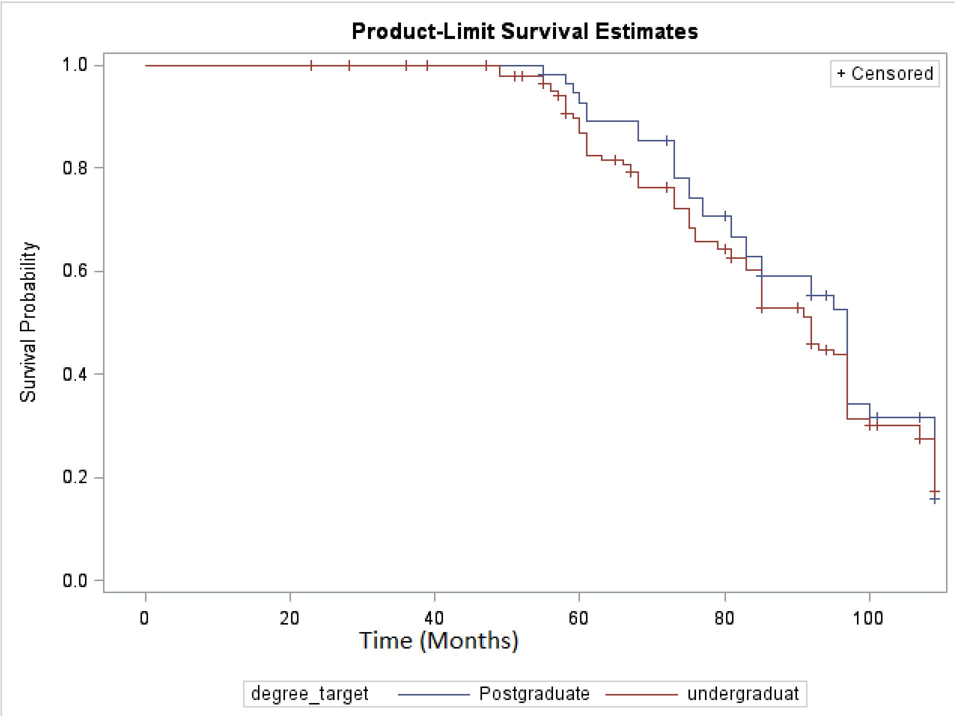


E-commerce:

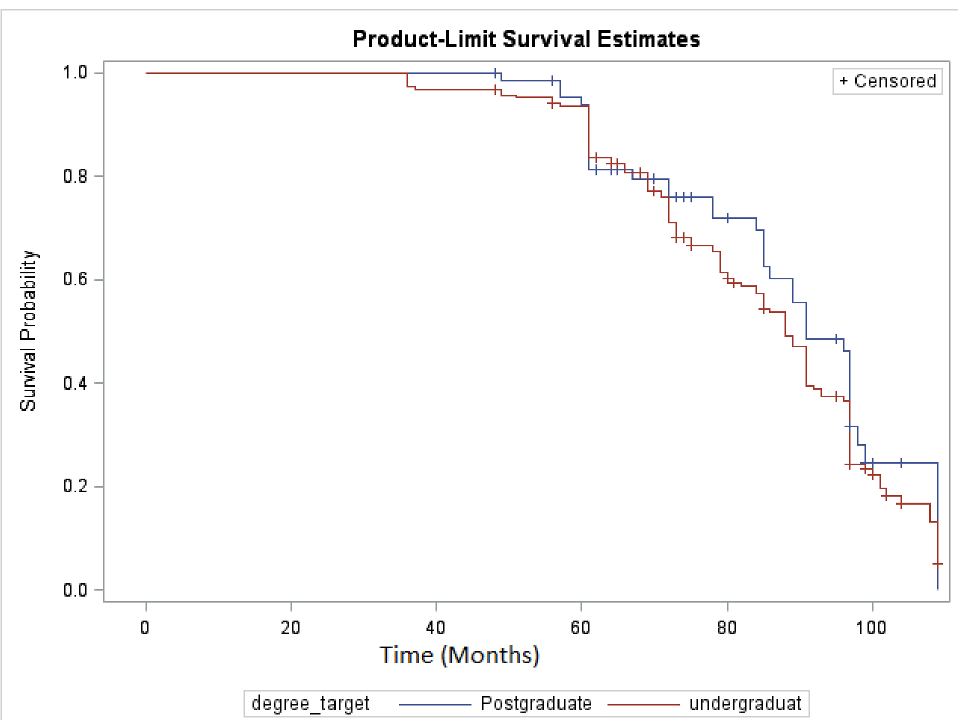


Vrushank Shah

Mobile:

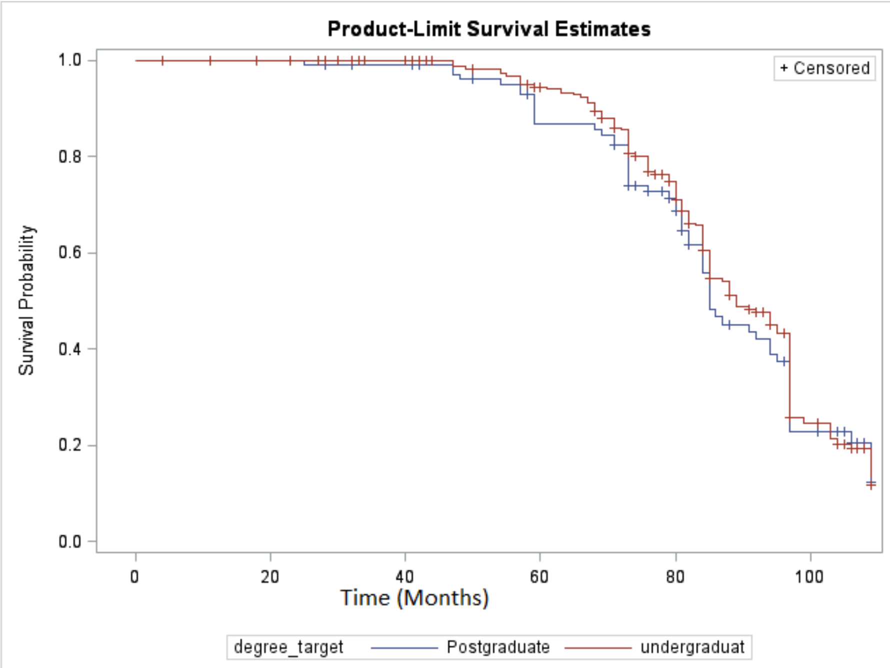


Software:



Web:

Vrushank Shah



Summary of survival analysis after **60 months** of starting a company:

Category	Survival chances for Post-graduate	Survival chances for under-graduate
Advertising	**70%** ↓	**90%** ↑
Analytics	**80%** ↑	**60%** ↓
E-commerce	**95%** ↑	**80%** ↓
Enterprise	**95%** ↑	**85%** ↓
Video	**95%** ↑	**85%** ↓
Mobile	**90%** ↑	**80%** ↓
Software	**80%**	**80%**
Web	**85%** ↓	**95%** ↑

From the above table, we can conclude that for most of the domains having post-graduate people in the core committee increases the survival chances. Thus from the above 2 comparisons, I concluded that there is a strong association between degrees of core-committee people and the success of a startup company. Also, having post-graduate core-committee people helps to get more funding and also leads to high survival chances of the company.

Section 4-

Predictive Model:

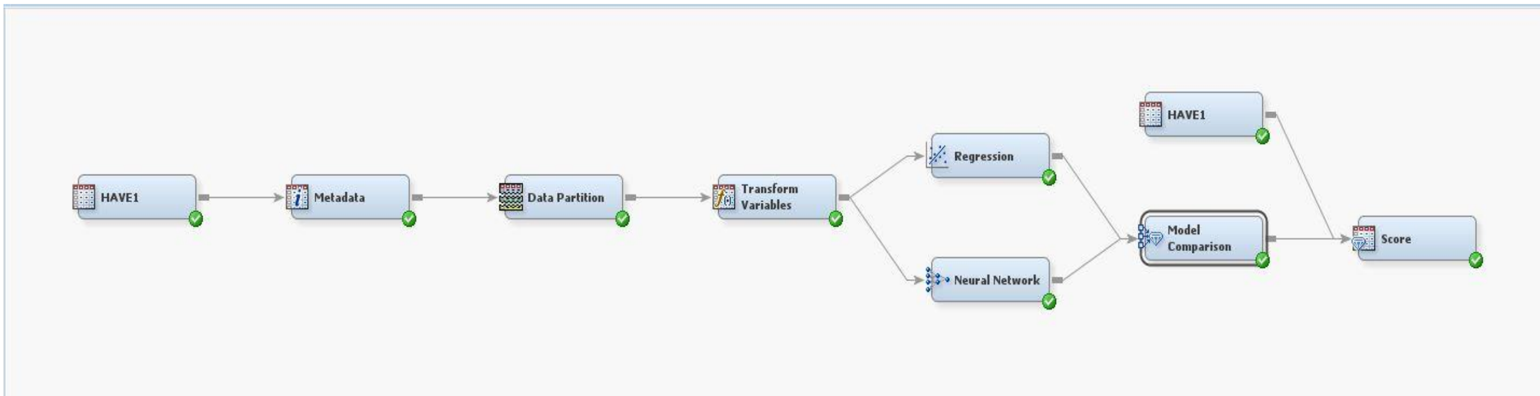
Data preparation: All the variables mentioned in section 2 were used to build the model. After performing initial data exploration variable total_funds and total_valuation had high skewness and kurtosis. A log transformation was performed on total_funds and total_valuation to decrease its skewness and kurtosis. The companies which were started between 2000-2014 were used for analysis. Also, only that companies were taken who had atleast got 1 round of funding. A target variable was created and 1 was assigned to the companies which are closed or acquired and 0 was assigned to the companies which are still operating.

Vrushank Shah

Process flow:

SAS enterprise miner was used to build the predictive model. A stratified sampling was performed on the target variable. Data was divided into 60% training,20% testing and 20% validation dataset. A transform node was used to do log transformation on total_funds and total_valuation variables. A stepwise regression model was run on the key factors variables and Neural Network was also run on the same variables. After doing a model comparison the regression model turned out to be better with 87% accuracy and 0.79 ROC index.

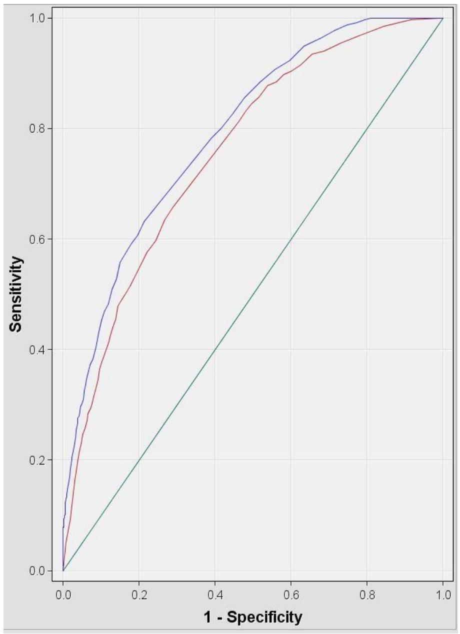
Process Flow:



Results:

ROC curve:

Here blue line indicates the regression model and red line indicates neural network model. Since the area under curve for regression model is more than neural network we will choose regression model.



Significant variables:

Vrushank Shah

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Burnrate	1	6.3866	0.0115
Fundingdays	1	31.6459	<.0001
LOG_TotalValuation	1	0.5440	0.4608
LOG_funding_total_usd	1	0.0187	0.8912
Milestonedays	1	5.6288	0.0177
Seed	1	111.0269	<.0001
category_code	40	56.3893	0.0444
funding_rounds	1	19.7872	<.0001
milestones	1	20.6500	<.0001
months	1	146.0554	<.0001

Scoring:

I tested my predictive model on a few startups which have both succeeded and failed. For example, I test my model on two companies:

1) A closed company named Minekey which satisfied all the parameters of the model such as it raised total $36M ,2 milestones achieved,2 funding rounds but had high burnrate and took around 23 months to achieve first funding and the model showed that the probability of surviving is 0.35 hence it is not successful.
2) An operating company named Rubicon project which was founded in 2007 and has raised $261M, 7 funding rounds, achieved 5 milestones and took around 4 months on an average to achieve each milestone. My model predicted that its probability of surviving is 0.87 hence it is successful.

Section 5:

Conclusion:

Based on the survival analysis we can conclude that there is a strong relationship between degree and being a successful startup company. Because getting funds based on the idea does not lead to a successful company there should be people in the core-committee that has general and business-specific knowledge. The predictive model has got an accuracy of 87% hence using the variables which are significant one can predict whether the company which is an initial stage will be successful in future or no ?

References:

1) Predicting the outcome of startups:Less failure ,More Success by Amar Krishna,Ankit Agarwal,Alok Chaoudhry