Analysis of Business Sustainability of Startups in India

Dr. C Komalavalli¹, Dr. Chetna Laroiya² and Dr. Disha Grover³

IT Department, Jagan Institute of Management Studies, India
IT Department, Jagan Institute of Management Studies, India
IT Department, Jagan Institute of Management Studies, India

Abstract

The startup ecosystem in India has transformed the country in different aspects recently. Startup ecosystems are driven by several factors such as funding, innovative ideas, services, platform etc. Startups are growing in almost all parts of India, driving a new job opportunities. Startups are now a days receiving much attention from the research scholars and corporate specialists. However, startups are shutting down within a short span of time. In this paper, exploratory data analysis has been performed to find the factors affecting the startup stability in Delhi, NCR (India). Experimental results have been shown with the help of tables and figures showing the impact of different influential factors for startup stability, thus helps in understanding of startup ecosystem.

Keywords:
Startup, Uni-variate Analysis, Bivariate Analysis, Eco System

1. INTRODUCTION

Startups are considered as nation builders globally. This statement is true for India also, Since India is a fastest growing countries in terms of entrepreneurship[1]. With their success, country’s economy growth also increases rapidly. Start up is defined as a new business that is in the initial stages of operation and financed by an individual or small group of individuals. This term was coined a long way back, but Indian people are not aware of that term. Start-ups are nothing but an Idea that manifests into a commercial undertaking. Government of India has launched Start up India initiative to encourage the youngsters to take up the initiatives. Indian government is serious in promoting entrepreneurship at the startup level and has taken a number of initiatives to ensure appropriate support. In this aspect it is relevant to mention ‘Make in India’ campaign introduced in September’14 to attract foreign investments[2] and encourage domestic companies to participate in the manufacturing sector. The government increased the foreign direct investment (FDI) limits for most of the sectors and strengthened intellectual property rights (IPRs) protection to instill confidence in the startups.In order to make the country as number one destination for startups, Government of India (GoI) has introduced a new campaign called ‘Standup India’ in 2015[2] aimed at promoting entrepreneurship among women and to help startups with bank funding. Another commendable and far reaching initiative is ‘Digital India’ introduced in 2015[3] to ensure government services are made available to every citizen through online platform that aims to connect rural areas by developing their digital infrastructure which translates into a huge business opportunity for startups[4]. There are different factors influencing. Endurance of the startup and researchers are designing various models to predict the success of the startup.

2. PURPOSE OF THE STUDY

The purpose of the research is to study the impact of various factors influencing the success and sustainability of a startup. Government policies also play an important role in the success of a startup. This study is exploratory on secondary data in nature. Data is analysed in Time series and Cross sectional or Longitudinal (same sample at different points in time) to find out the most promising feature, that impacts the stability of the start up.

3. RESEARCH METHODOLOGY

Basic process of data analysis includes data acquisition, data preprocessing, data analysis and result interpretation.[5]
3.1. Data Acquisition

Startup ecosystems are influenced by internal as well as external factors also[6]. Impact of these factors depend on the location of the startup too. In this paper, only few external factors affecting startups are only taken into consideration. Data used in study are secondary data and source of the data is data.gov.in. The dataset contains the details of the startups spanning for last 30 years starting from 1900 to 2019. Our focus of study is limited to Delhi and NCR region. Dataset contains 6 features and 315,450 rows in the dataset. Details of the dataset are described in next section.

3.2. Data Preparation

Data needs to be cleaned for further processing[7]. We have removed null values from the dataset in order to avoid ambiguity.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company_Status</td>
<td>Current Status of the company.</td>
</tr>
<tr>
<td>Company_Class</td>
<td>Type of Company</td>
</tr>
<tr>
<td>Company Category</td>
<td>Domain of the company</td>
</tr>
<tr>
<td>Company SubCategory</td>
<td>Sub Domains of the company</td>
</tr>
<tr>
<td>Data_Of_Registraion</td>
<td>Registration Date</td>
</tr>
<tr>
<td>Paidup_Capital</td>
<td>Amount paid for the startup</td>
</tr>
</tbody>
</table>

From the dataset, Company_Status is considered as dependent variable (Y) for the analysis and contains 13 different categories. For better understanding of the analysis, categories have been reduced to three i.e: Active, Strike Off and Dormant. Company_Class, Company_Category consists of three values. Company subcategory has domain of five values. Data of registration has been divided into decades and twelve decades are considered for the analysis. Company_Class(X1), Company_Category(X2), Company_SubCategory(X3) are categorical variables but Data_Of_Registraion(X4) and Paidup_Capital(X5) are continuous variable.

3.3. Analysis of Data

Data is analysed using Excel Analysis Toolpak and Data interpretation is based on statistical data returned by Regression analysis.

3.4. Descriptive Statistics

Univariate analysis on the dependent variable, company status (Y) is performed and visualized with the help of Pie chart.

![Pie chart showing the status of startups](image)

The chart indicates that 70% of the startups are active state, 29.5% are strike off and 0.3% are in dormant state over the span of 29 years.

![Column chart showing the number of startups](image)

Bivariate analysis on independent variable Date_Of_Registration and dependent variable Company_Status is achieved and visualized with the help of bar chart. Bar chart illustrates that startups are active in the year range from 1911-1920, 1971-1980, 2011-2020, 2001-2010, 1931–1940 and 1921-1930 in descending order.

Correlation coefficients calculated between any two variables considering one dependent variable(X) and dependent variable(Y) and results are shown here.

a) \[
\text{COMPANY_CLASS} \rightarrow \text{COMPANY_STATUS} = -0.037381224
\]

b) \[
\text{COMPANY_CATEGORY} \rightarrow \text{COMPANY_STATUS} = 0.014151841
\]
The above result clearly indicates that all five independent variables is not determining the dependent variable company status, since the correlation coefficient value range does not lie between <=0.05 and >0.05. As correlation coefficient or univariate analysis does not identify any feature which could be held responsible for the Company_status, so we have decided to do multivariate analysis, so that we can identify pattern in the data by considering combination of more than one feature in order to justify output variable.

Multiple regression is done to find out the features which are important contributors to the status of the startups and arranged in descending order of their impact. All possible combinations of independent variables are considered and calculated the values. Since Paid_Up_Capital is a continuous variable, contribution of this variable is not considered in our study.

1. \[ Y = -0.06544 \times \text{COMPANY\_CLASS} + 0.073851 \times \text{COMPANY\_CATEGORY} + 1.355202 \]
2. \[ Y = -0.6693 \times \text{COMPANY\_CLASS} + 0.05921 \times \text{COMPANY\_SUBCATEGORY} + 1.492873 \]
3. \[ Y = 0.041765 \times \text{COMPANY\_CLASS} - 0.12264 \times \text{DECADE} + 2.56597 \]
4. \[ Y = 1.258795 \times \text{COMPANY\_CATEGORY} + 0.02872 \times \text{COMPANY\_SUBCATEGORY} + 0.6693 \times \text{COMPANY\_CLASS} + 0.073851 \times \text{COMPANY\_CATEGORY} + 1.404583 \]
5. \[ Y = 0.08228 \times \text{COMPANY\_CATEGORY} + 0.08228 \times \text{COMPANY\_CLASS} + 0.08228 \times \text{COMPANY\_CLASS} + 0.08228 \times \text{COMPANY\_STATUS} = 0.065030073 \]

The above Equations reveals that there is no one variable which is effectively contributing to the dependent variable i.e COMPANY_STATUS weather It is Active, Strike Off or in Dormant State. Still decade of registration of the Company has some positive impact on the status of the Co.

Multicollinearity occurs when input variables are not only correlated to the output variable (Y) but greatly dependent on each other too. Multicollinearity is considered to find the attributes which could be dropped from the regression analysis as it is linked to the values in the other input variable.

Multicollinearity exists between independent variables if VIF is >=5
Where as VIF = [1/(1-R2)]
R2 is a parameter to calculate how good the feature set can be to predictors of the output variable. Higher the value of R2, better are the feature set predict Y.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Variable</th>
<th>R2</th>
<th>1-R2</th>
<th>VIF (Variance Inflation Factor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>DECADE</td>
<td>0.00269</td>
<td>0.9973</td>
<td>0.05649</td>
</tr>
<tr>
<td>8</td>
<td>SUB_CATE-GORY</td>
<td>0.08209</td>
<td>0.9179</td>
<td>0.0225</td>
</tr>
<tr>
<td>9</td>
<td>COMPANY_CLASS</td>
<td>0.08185</td>
<td>1.0891</td>
<td>0.519</td>
</tr>
<tr>
<td>10</td>
<td>COMPANY_CATEGORY</td>
<td>0.08228</td>
<td>1.0896</td>
<td>0.682</td>
</tr>
</tbody>
</table>

Variance Inflation Factor clearly indicates that Company_Class and Company_Category features should be removed from the Regression model to predict Company_Status based on their p-value. Feature (Company_Class and Company_Category) if result to have higher p-value should be eliminated first.

4. CONCLUSION

This study is an effort to explore that how promising is a feature in predicting the dependent variable. Features taken for the study are based on the external factors which are driving the startups. Through exploratory data analysis on the data under consideration, some important points revealed are:

A good perccentage of startups are active (79%)
Output variable (Y) is categorical variable with 3 categories so logarithmic Regression is applicable but not the Binary Logarithmic Regression.


Although predictive modelling is out of the scope of this paper and also the feature set is too small to to create a model to predict the Company_Status but observed facts during our study are

Multiple regression identifies Company_Category as a pretty good predictor of company_Status. Company_Class and Company_Category features show multicollinearity so they might not contribute well in the Regression model if built.

Thus the work can help us to improve the understanding of the startup ecosystem. In future work, categories and subcategories can be considered for further processing.

REFERENCES


