**Context**

A **startup** or **start-up** is a company or project begun by an entrepreneur to seek, develop, and validate a scalable economic model. While entrepreneurship refers to all new businesses, including self-employment and businesses that never intend to become registered, startups refer to new businesses that intend to grow large beyond the solo founder. Startups face high uncertainty and have high rates of failure, but a minority of them do go on to be successful and influential. Some startups become unicorns: privately held startup companies valued at over US$1 billion. [Source of information: Wikipedia]



Source: bbva

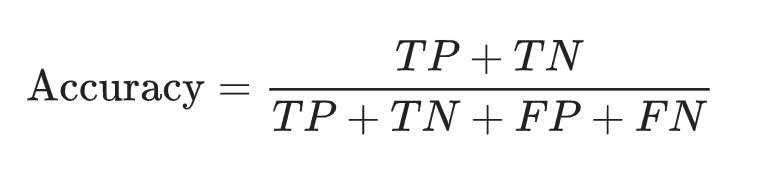
 Startups play a major role in economic growth. They bring new ideas, spur innovation, and create employment thereby moving the economy. There has been an exponential growth in startups over the past few years. Predicting the success of a startup allows investors to find companies that have the potential for rapid growth, thereby allowing them to be one step ahead of the competition.

**Objective**

The objective is to predict whether a startup which is currently operating turns into a success or a failure. The success of a company is defined as the event that gives the company's founders a large sum of money through the process of M&A (Merger and Acquisition) or an IPO (Initial Public Offering). A company would be considered as failed if it had to be shut down.

**Evaluation Criteria**

Submissions are evaluated using **Accuracy Score**.



**How do we do it?**

Once you generate and submit the target variable predictions on the test dataset, your submissions will be compared with the true values of the target variable.

##### ****About the Data****

The data contains industry trends, investment insights and individual company information. There are 48 columns/features. Some of the features are:

* age\_first\_funding\_year – quantitative
* age\_last\_funding\_year – quantitative
* relationships – quantitative
* funding\_rounds – quantitative
* funding\_total\_usd – quantitative
* milestones – quantitative
* age\_first\_milestone\_year – quantitative
* age\_last\_milestone\_year – quantitative
* state – categorical
* industry\_type – categorical
* has\_VC – categorical
* has\_angel – categorical
* has\_roundA – categorical
* has\_roundB – categorical
* has\_roundC – categorical
* has\_roundD – categorical
* avg\_participants – quantitative
* is\_top500 – categorical
* status(acquired/closed) – categorical (the target variable, if a startup is ‘acquired’ by some other organization, means the startup succeed)

**To load training dataset, use below command:**

import pandas as pd

startup\_data = pd.read\_csv("[training\_set\_label.csv](https://raw.githubusercontent.com/dphi-official/Datasets/master/startupdata/training_set_label.csv)“)

**To load the testing dataset, use below command:**

Load the test data (name it as test\_data). You can load the data using the below command.

test\_data = pd.read\_csv('[testing\_set\_label.csv](https://raw.githubusercontent.com/dphi-official/Datasets/master/startupdata/testing_set_label.csv)')

Here the target column is deliberately not there as you need to predict it.

##### ****Acknowledgement****

I would like to thank [Ramkishan Panthena](https://www.linkedin.com/in/ramkishanpanthena/), for providing us this dataset. He is a Machine Learning Engineer at GMO.