InvestRight

ABHISHEK YADAV
19B090001
Dept. of Mathematics
IIT BOMBAY

Project Objective:
Startup Success Prediction

Startup Success Prediction

The final objective is to build a tool that given the general information about a Startup can rate it as whether it is going to be successful* or not.

The tool will not only give the binary classification but will also let you compare it with top 5 already existing similar startups and how they've performed in the past.

New Business Failure Rate

42	46	63	55
Finance	**Wholesale**	**Information**	**Transport**
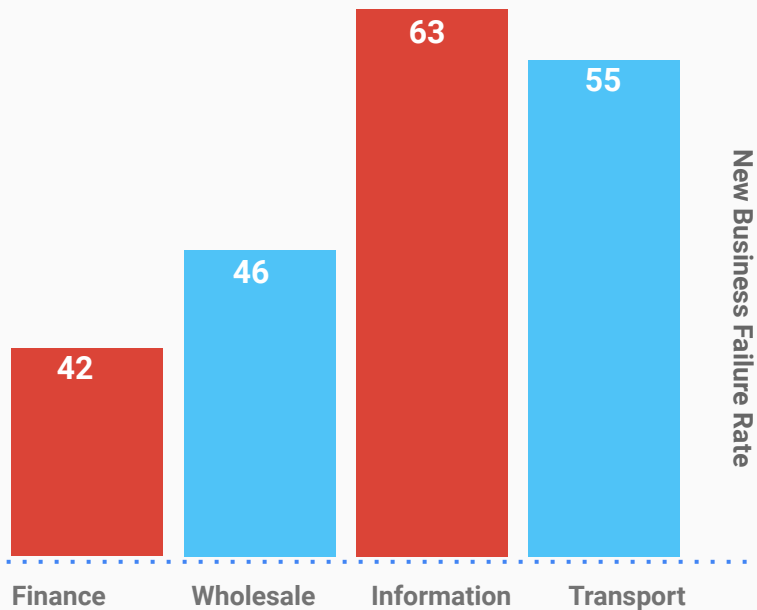
*criteria for success is defined as a Startup either going to an IPO or being acquired by a Company/Conglomerate

Project Definition: The Problem

- In the current situation of technological advancement where we have thousands and millions of startups coming into the market and playing a very important role in the economy. Which creates a new problem for Investors and CEOs of startups as well.

- Investors are eager to invest in new ideas and technological startups but because of such a large number of new startups it's next to impossible for a human to judge all at a same time. Hence there is a burning need of some tools and techniques that can evaluate and tell which startup is going to be a successful one.

- Many business groups don't know which field to choose for a successful business and hence they need something which can tell them which field is doing good in current market situation.

Project Definition: The Solution

- We leverage technological advancement in Artificial Intelligence and Machine Learning with available data regarding past startups to define how different factors drive the success of a startup.

- Our final model makes use of startup features such as total funding, operating status, number of funding rounds etc. to determine whether a startup would achieve competitive advantage in the market

- Our solution would help investors and companies in making the right investment choice

The Technology:
Data Mining (ML+DL)

Classification Models

In the present work, we have a binary classification task - the target feature is either a successful or an unsuccessful company.

It is a type of supervised learning, a method of machine learning where the output categories are predefined.

It is important to choose not only the algorithm that better fits the problem but also which adapts well to characteristics of the dataset.

Logistic Regression

LR is one of the fastest algorithms to train. Since it has low variance it is less prone to over fit.
It also does not make assumptions about distributions of classes in feature space.

Random Forest / XGBoost

Random Forest:RF is an ensemble of decision trees. For prediction, RF takes inputs from all the trees and then predicts the outcome.

XGBoost:It is a highly preferred algorithm because of less feature engineering required.

MLP

A multilayer perceptron is a class of feedforward artificial neural network which utilizes a supervised learning technique called backpropagation for training

Strategy
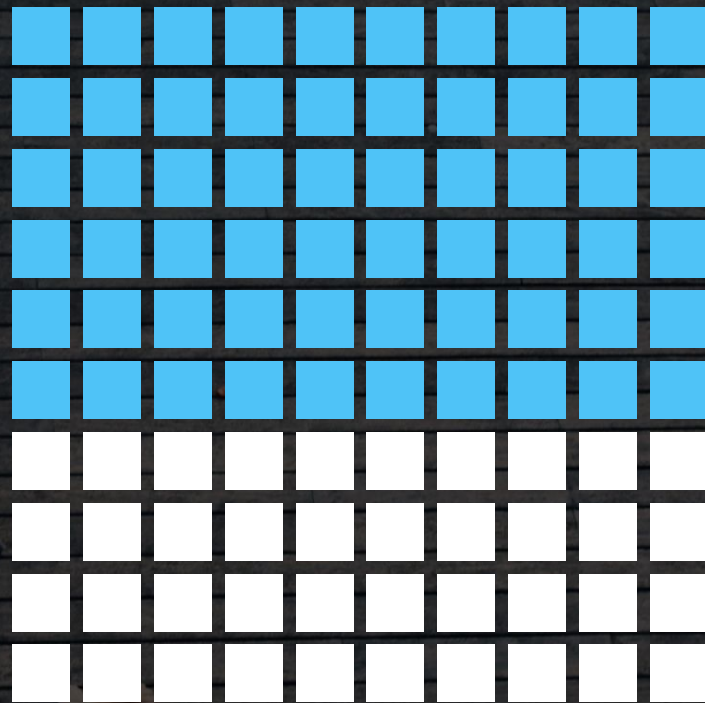
Strategy

Step 1

Data Collection,
Data Cleaning & Filtering

Step 2

Data Transformation,
Data Visualization,
Features Selection

Step 3

Modelling,
Evaluation & Optimization,
Best Model Selection,
Results Post Processing