Startup Success Prediction
Project Proposal

Name: Lama Alharbi
Class : t5o03

Problem Statement:

What is the framing question of your analysis and what the purpose of the model/system you plan to build?

9 out of 10 startups fail (source: Startup Genome). This means that investing in a startup is a risky matter so to assist the investors the factors impacting the success of a startup need to be determined. The following question for this analysis is:

What are the most relevant variables that can be used to predict the success of the startup?

The main goal of building the model is to predict the status of a startup.

Who benefits from exploring this question or building this model/system?

- Investors
- Entrepreneurs

Data Description:

The dataset used is "Start Up Investments" from Crunchbase which is a platform for finding business information about private and public companies. The dataset is acquired through a secondary source which is Kaggle.

The data set contains 49438 rows, each row of the dataset represents a startup company with 39 features such as: name, category list, market, funding total in USD, country, funding rounds number, round A-H series funding and the target feature for this analysis: status.

Upon initial view, the dataset seems imbalanced as the data contains %6 acquired startups, %5 closed and above 80% operating startups.

Tools:

The tools used for this project will include:

1-Python3: as the main programming language used

2-Jupyter: as the IDE for python

3-Numpy: to manipulate the dataset

4-Pandas: to clean and preprocess the dataset

5-Matplotlib: to visualize the dataset

6-Seaborn: to visualize the dataset

7- sklearn: for feature selection and modeling

MVP goal:

a minimum viable product of this project would be a heatmap of the most important features affecting the success of a startup and a classification model for the status feature.