Category	*Explanation*
Introduction	**Why was the project undertaken? What was the research question, the tested hypothesis or the purpose of the research?**
Selection of Data	**What is the source of the dataset? Characteristics of data? Any munging or feature engineering?**
Methods	**What materials/APIs/tools were used or who was included in answering the research question?**
Results	**What answer was found to the research question; what did the study find? Was the tested hypothesis true? Any visualizations?**
Discussion	**What might the answer imply and why does it matter? How does it fit in with what other researchers have found? What are the perspectives for future research? Survey about the tools investigated for this assignment.**
Summary	**Most important findings.**

Introduction

The dataset that we will be using is [Startup Success/Fail]

Our goal is to analyze the data from start-up companies to determine factors that can help predict the likelihood of success or failure in a startup business by analyzing factors such as business category, funding and the number of funding rounds, geographic area, and the success ratios based on those factors.

Some of the questions that we plan on exploring are: What is the rate of success and failure in start-up businesses? Do some business sectors yield a higher success rate than others? What is the funding threshold that businesses typically need to find success in their sector? Are funding rounds helpful in achieving success?

This data analysis will help entrepreneurs make informed decisions and minimize chances of failure in their business endeavors.

Selection of Data

The dataset that we used for our project is Startup Success/Fail.
This dataset contains the following features:

- **permalink** - Link to the organization
- **name** - Company name
- **homepage_url** - Company website
- **category_list** - Field of company
- **funding_total_usd** - Total funding in USD
- **status** - Operating status
- **country_code** - Country
- **state_code** - State of company location
- **region** - Region of company location
- **city** - City of company location
- **funding_rounds** - amount of rounds of funding received
- **founded_at** - the date the company was found
- **first_funding_at** - the date the company received its first funding
- **last_funding_at** - the date the company received its last funding

Munging and Data Engineering

There were features that we removed because they were irrelevant to our analysis.These features were *permalink, homepage_url, category_list, state_code, region, city, and found_at.*

After dropping the columns, we checked for NA percentage by column. We dropped all rows that had missing information because it still provided us with a decent size dataset.

For the column *status*, there were four different entries: ipo, acquired, operating, and closed. We determined that ipo and acquired were to be used as criteria for success and closed was to be used as the criteria for failure. We used OneHotEncoder to convert this column from categorical to numerical in preparation for machine learning.

We also created another dataframe called *sf_by_year*. This dataframe contains the year that the company was first funded as the index and the columns are as follows: *total_success, total_fail, total, success_ratio, fail_ratio, and funding_median.* This dataframe gives us a quick glance at how success ratio is affected year over year.

Methods

Tools:

- Google Colab
- Github
- NumPy
- Pandas

- Matplotlib
- Seaborn
- Graphviz
- Scikit
- Machine learning models
    - KNN Classification
    - Decision Tree Classification

Results

What is the rate of success and failure in start-up businesses?

55.92% succeeded while 44.08% failed.

Do some business sectors yield a higher success rate than others?

Yes, according to the data, some categories such as Android yield a 100% success rate, and some such as Machine-to-Machine (M2M) a 0% success rate.

What is the funding threshold that businesses typically need to find success in their sector?

Based on the decision tree, funding >= 1508700 yields a higher chance of success, however the decision tree is only 50.3% accurate.

Are funding rounds helpful in achieving success?

Companies that only received one funding round had a higher fail ratio than companies that received greater than one funding round.

Discussion

The accuracy scores achieved in our prediction models shows that the features that we examined are not sufficient enough to predict the outcome of a startup. Through our research we found that there are many other factors that could affect a company's chances of succeeding or failing. While funding amounts and additional funding rounds increase the likelihood of success, the increase is minimal. The category (business sector) of the startup seems to have a greater impact on the predictability of success in the data we explored.

Summary

Based on the data and predictive models used, we can conclude that predicting the success or failure of a company cannot be determined by the amount of funding or the number of rounds of funding a startup receives. The type of startup company has a larger impact on its success than the amount of funding. While funding does play a part in the success of a startup, it appears that other factors may play a more important role in determining success or failure, and funding data alone provides inconclusive results.