**Comparative Analysis of Decision Tree and Random Forest Algorithms for Predicting Startup Success**

**Team 6**

**Collaborators:**

1. Gouri Dumale (G49564205) [gouri.dumale@gwmail.gwu.edu](mailto:gouri.dumale@gwmail.gwu.edu)

2. Nemi Makadia (G29362869) [nemi.makadia@gwmail.gwu.edu](mailto:nemi.makadia@gwmail.gwu.edu)

3. Bhanu Sai Praneeth Sarva (G46159306) [bhanusaipraneeth.sarva@gwmail.gwu.edu](mailto:bhanusaipraneeth.sarva@gwmail.gwu.edu)

4. Smit Pancholi (G31443926) [smitsnehal.pancholi@gwmail.gwu.edu](mailto:smitsnehal.pancholi@gwmail.gwu.edu)

**Description:**

This dataset comprises approximately 54,000 rows and 39 columns, offering a diverse range of company-related information. It includes details such as company names, URLs, market sectors, geographical data (country, state, region, city), founding dates, and the timing of initial and final funding rounds. Additionally, the dataset encompasses various investment types, such as seed funding, venture capital, equity crowdfunding, undisclosed investments, convertible notes, debt financing, angel investments, grants, private equity, post-IPO equity, post-IPO debt, secondary market transactions, product crowdfunding, and rounds labeled A to H. The status of companies is classified into three categories: acquired, operating, or closed.

**SMART Questions:**

1. What are the key indicators or features that the decision tree models identified as significant in determining the fate of a company (closed, operating, acquired)?

2. What are the top 5 global markets based on the count of startups, and how does this distribution vary across different industries or sectors?

3. Are there specific thresholds or values in total investment that significantly impact a company's likelihood of being 'acquired'?

**Modeling Methods:**

We aim to conduct a classification analysis using Random Forest and Decision Tree methods to predict the accuracy of our model in determining the status of startups. Specifically, we seek to leverage these methods to assess how features such as funding rounds, total investment, and other relevant factors contribute to predicting whether a startup will be classified as 'closed,' 'operating,' or 'acquired.

**GitHub repo:** <https://github.com/smitpancholi313/DataMiningProject.git>

**Dataset link:** [StartUp Investments (Crunchbase) (kaggle.com)](https://www.kaggle.com/datasets/arindam235/startup-investments-crunchbase/data)