Problem Statement: What problem will you be investigating? Why is it interesting?

Anu is an agricultural technology company that has developed controlled environment agriculture vertical growing systems powered by their proprietary Rotary Aeroponics® technology. To assess plant growth performance within the anu systems, we collect various types of data, including total plant biomass, plant emergence, plant survival, elemental analysis, water data, and whole plant images.

The challenge with assessing plant performance is that most processes are destructive. For example, in order to acquire elemental data, which can be used to understand plant health, the plants must be harvested and sent out to the lab, where they are dried down and processed. However, with imaging and machine learning techniques, plant health metrics can be extracted from images via colorimetric data. The images can be taken using RGB and NIR cameras, which allows for metrics such as NDVI to be extracted from the images (NDVI = (NIR - R) / (NIR + R)). We currently have a plant imaging pipeline that processes the images collected at harvest and extracts various colorimetric values using PlantCV, an open-source image analysis software package, specifically for plant science (analyze color PCV). This step should be able to be built into the existing pipeline which segments plant images.

This challenge presents an opportunity for an interesting opportunity to model plant tissue elemental analysis from image data, which has multiple benefits. Currently, the cost of plant tissue sampling is $13 per sample, meaning that in an experiment where we sample every other plant across 12 UTC chambers, the total cost is $4,680. By developing a pipeline that can accurately predict levels of select elements in a plant tissue sample, the company could save a significant amount of money. Secondly, with a non-destructive method of determining elemental levels within plants, we have more freedom to assess plants at various stages of growth. With timecourse data, we can assess plant health at various stages of growth non-destructively, and potentially adapt growing conditions based on what the plants need at a specific stage of growth.

Challenges: What are the challenges of this project?

1. Image quality - if image quality isn't consistently high quality, this may affect the model's ability to accurately predict elemental content
2. Some elements, such as phosphorus, don't have a standard spectral index used to target them, so it becomes more challenging to extract meaningful data without combining both plant structure data and colorimetric data
3. Determining what color or other visual metrics are associated with each plant tissue measurement

Dataset: What dataset are you using? How do you plan to collect it? You can use your own data or gather data from online data repositories.

The two datasets used for this project will be tabular data with plant colorimetric data extracted from anu plant images and the elemental analysis data from the lab.

All raw data from this project will need to remain confidential and will only be made accessible to Professor Manuel Campagnolo for review of the project.

Method or Algorithm: What method or algorithm are you proposing?
Since the data will primarily be tabular, potential options are Random Forest Regression or gradient boosting (XGBoost/LightGBM).
Naive Bayes Classifier to classify pixel colors in an image and PlantCV algorithms will be used to extract colorimetric data.

Evaluation: How will you evaluate your results? What kind of analysis will you use to evaluate and/or compare your results (e.g., performance metrics or statistical tests)?
Use regression metrics per element

- R^2 (coefficient of determination)
- MAE (Mean Absolute Error)
- RMSE (Root Mean Square Error)
- Spearman/Pearson correlation with ground truth (for ranking)

Other?