Project Title: Formalin Detection in Pineapple.

Dataset Description:

- Total Data Size: 250 samples
- Data Types: 2 classes

Splitting the Dataset:

Splitting a dataset into training, testing, and validation sets is a common practice in machine learning to evaluate the model's performance. Here's a general approach:

1. Training Set (typically 70-80% of the data):

- **Purpose:** Used to train the machine learning model.
- **Example Split:** 70% of 250 samples = 175 samples.

2. Testing Set (typically 10-20% of the data):

- **Purpose:** Used to evaluate the model's performance and generalization to new, unseen data.
- **Example Split:** 15% of 250 samples = 37.5 samples (rounded to the nearest whole number, so you might have 37 or 38 samples).

3. Validation Set (remaining data):

- **Purpose:** Used to fine-tune the model and avoid overfitting.
- **Example Split:** 100% - (70% + 15%) = 15% of 250 samples = 37.5 samples.

Remember, the exact percentages may vary based on the specific requirements of my project.

Class Distribution:

I mentioned there are 2 classes in the dataset. It's important to ensure that the classes are balanced across the training, testing, and validation sets. In other words, each set should have a similar distribution of samples from each class to avoid bias.