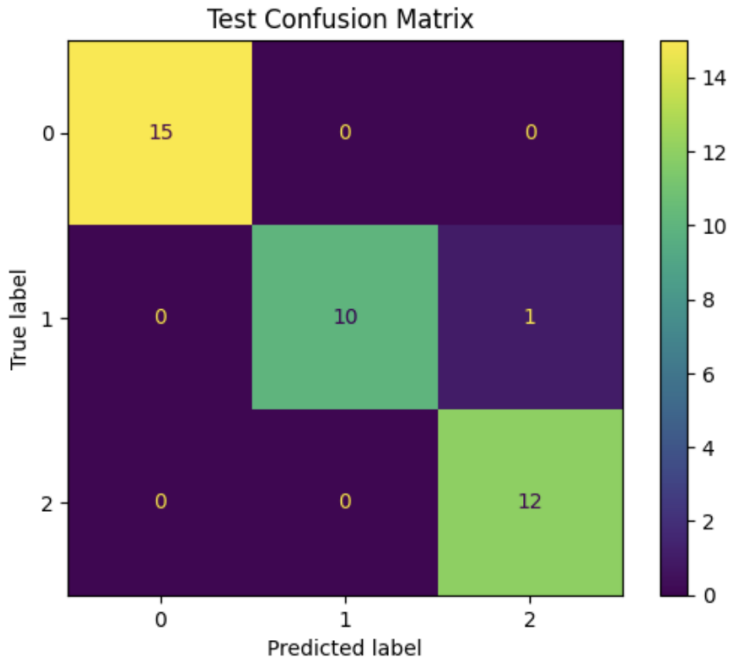
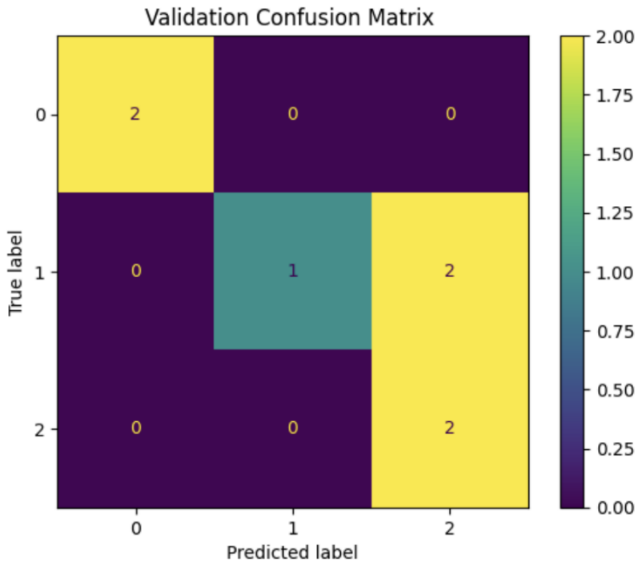
# Naive Bayes Classifier

1. We are using jupyter Notebook to complete this assignment.
   The following are the test and validation metrics of the naïve bayes classifier model.

```
Test Accuracy: 0.9736842105263158
Test Precision: 0.9743589743589745
Test Recall: 0.9696969696969697
```

```
Validation Accuracy: 0.7142857142857143
Validation Precision: 0.8333333333333334
Validation Recall: 0.7777777777777777
```



From the misclassification output, we observe that **one sample with the true label 'Iris-versicolor' was misclassified as 'Iris-virginica'**. This suggests that the model is struggling to differentiate between these two classes, likely due to similarities in their feature distributions. Here's an inference on the possible reasons and implications:

**Inference:**

1. **Feature Overlap**:
    - The attributes of **Iris-versicolor** and **Iris-virginica** may have overlapping values for certain features, making it challenging for the classifier to distinguish between the two. For instance, petal width and petal length often have close ranges for these species.
2. **Model Limitations**:
    - Since this Naive Bayes model relies on probability distributions and feature independence, it may misclassify samples when features are not fully independent or when there's significant overlap. This is a common limitation with Naive Bayes, which could lead to such classification errors, particularly if features are highly correlated.
3. **Discretization Effects**:
    - Discretizing continuous features into bins can reduce information about subtle differences, causing two distinct classes to appear similar. This may have contributed to the confusion between Iris-versicolor and Iris-virginica.