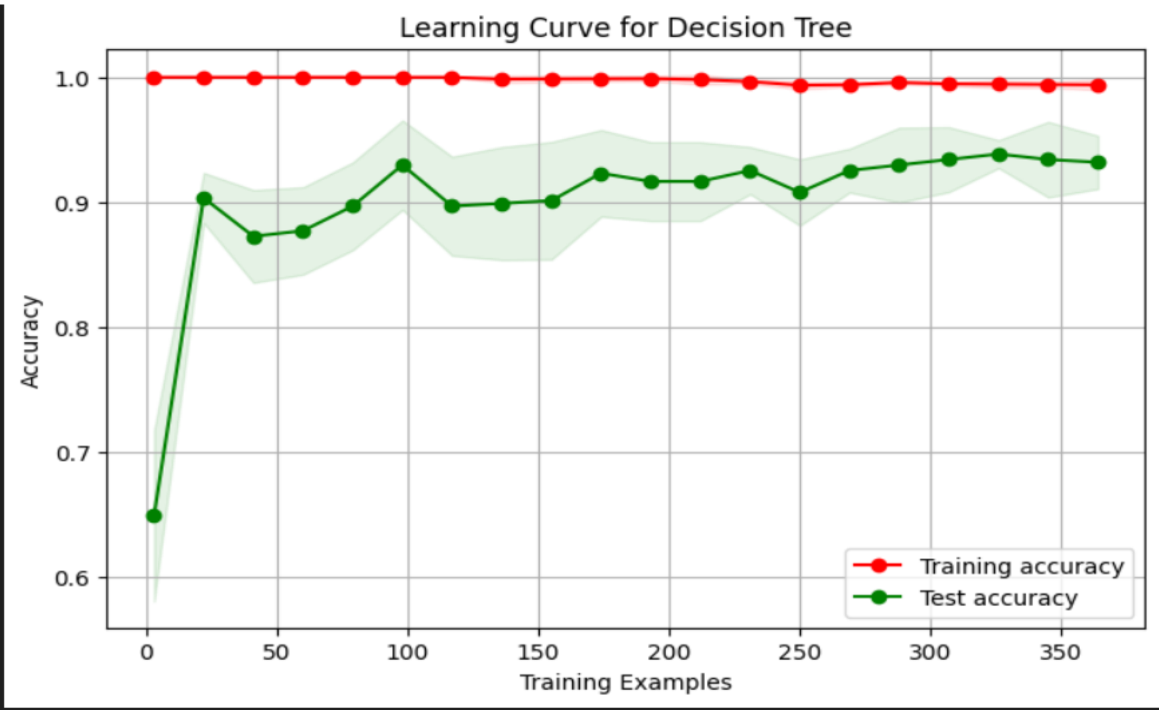
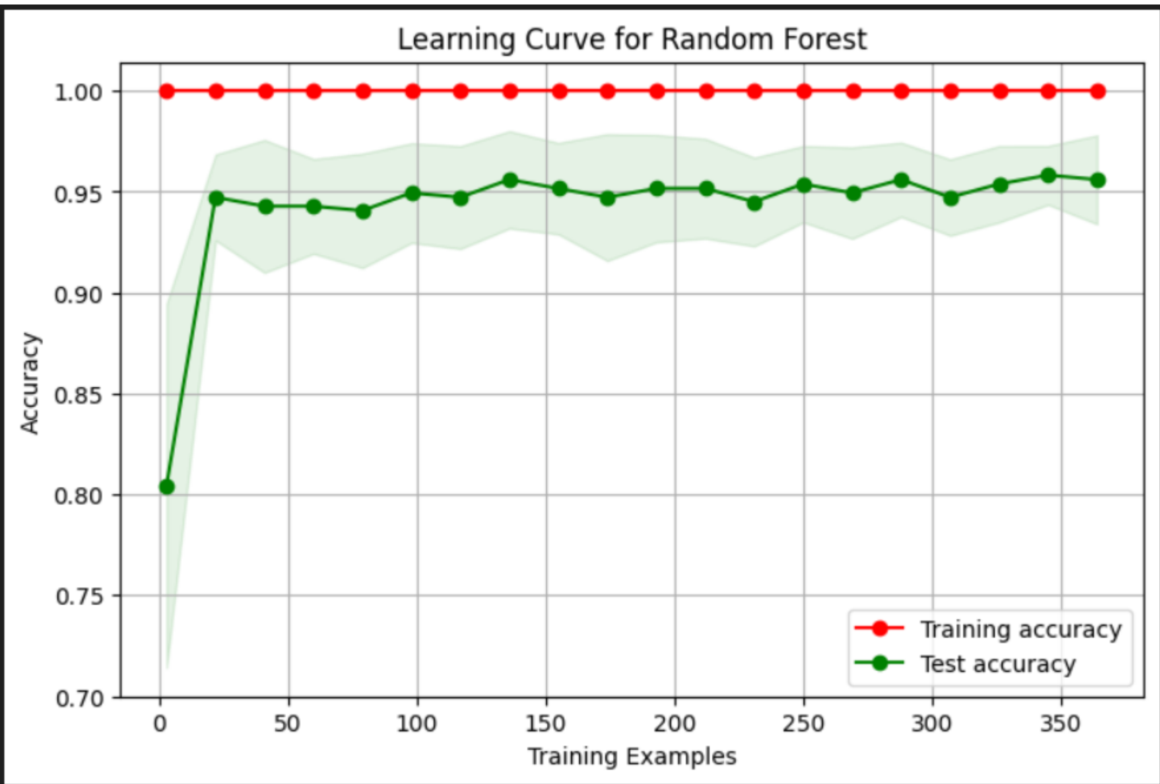
1.1)




```
Decision Tree - Training Accuracy: 0.9934, Test Accuracy: 0.9211
```




```
Random Forest - Training Accuracy: 1.0000, Test Accuracy: 0.9561
```

'1.2)

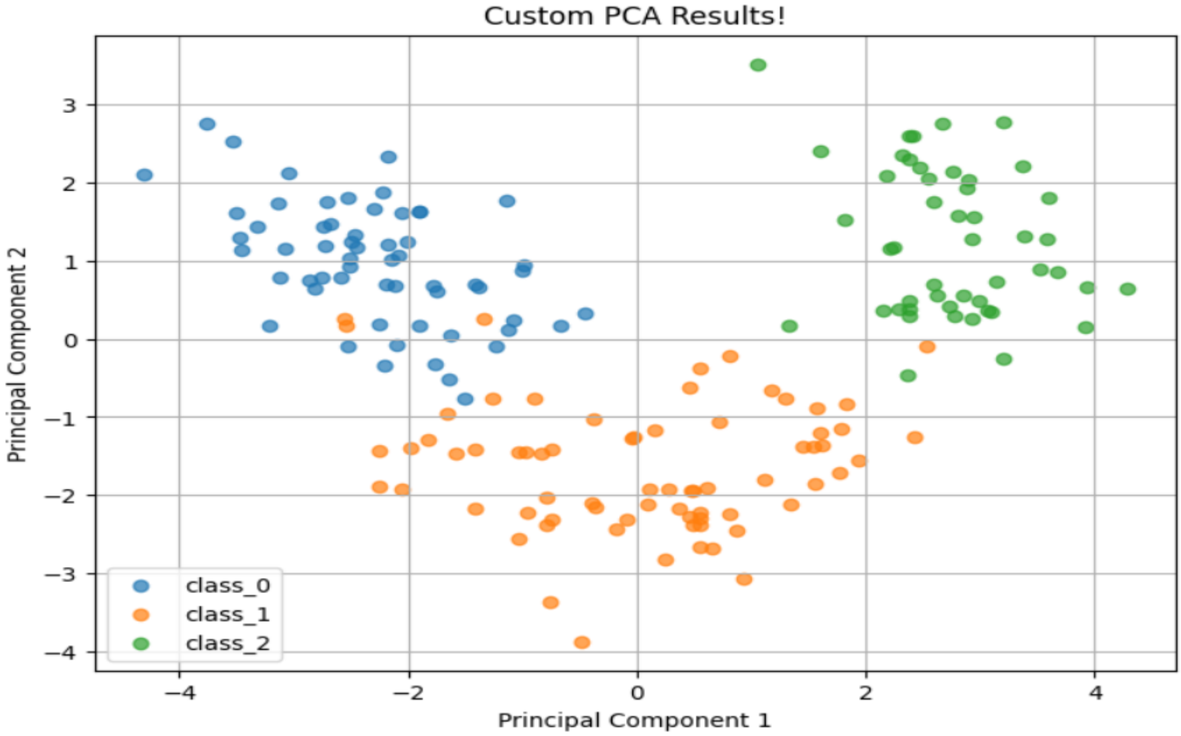
Based on the provided learning curves and accuracy metrics for the **Decision Tree** and **Random Forest** models, here is the analysis.

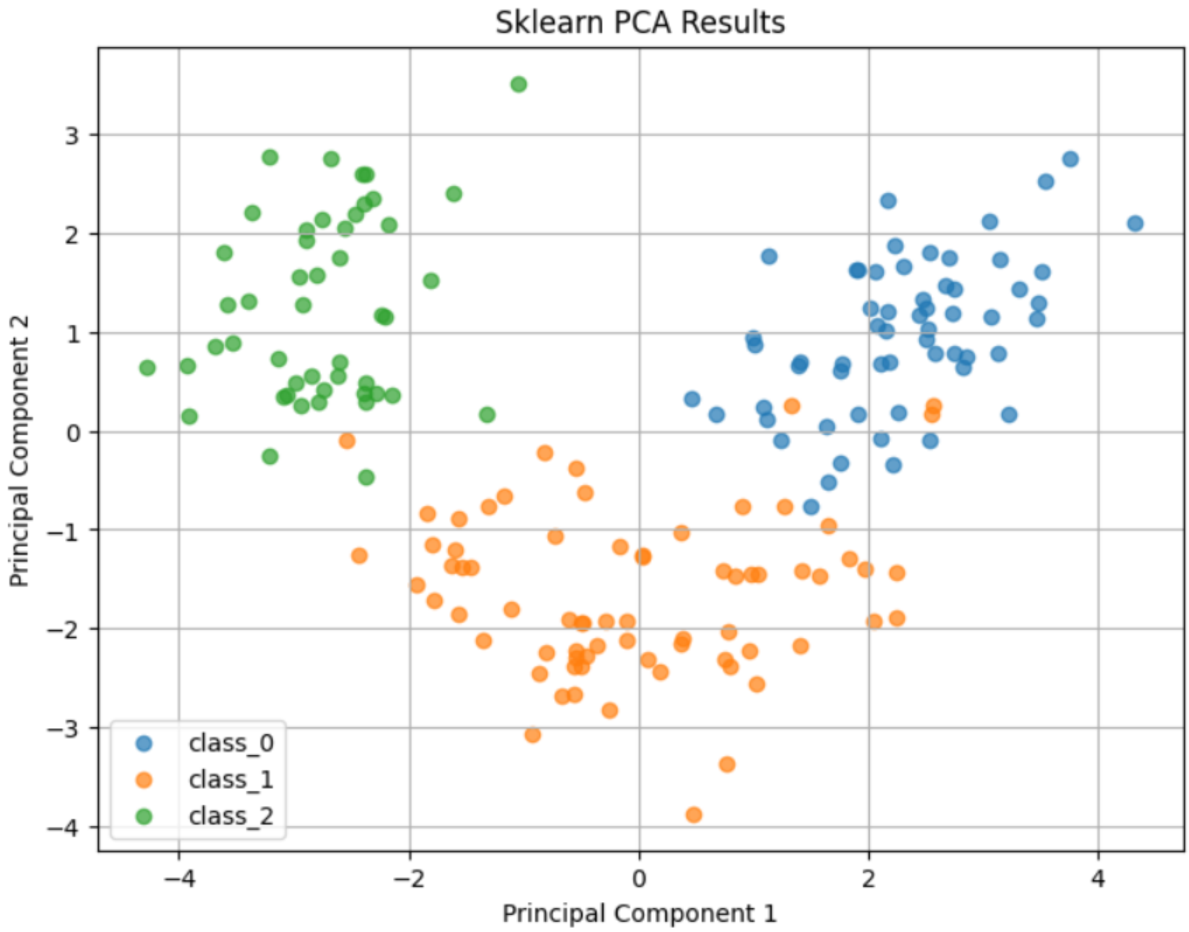
Evidence from Learning Curves

1. **Training Accuracy**:
    - Both the Decision Tree and Random Forest models achieve near-perfect training accuracy (~1.0). This is typical because these models can overfit the training data when not properly regularized.
2. **Test Accuracy**:
    - The Random Forest model consistently achieves **higher test accuracy** (~0.95) compared to the Decision Tree (~0.90).
    - The test accuracy of the Random Forest shows less variance (tighter green shaded region), indicating **better generalization** compared to the Decision Tree.
3. **Stability**:
    - The Random Forest curve is smoother and more stable across the training examples, demonstrating **robust performance across subsets of the data**.
    - In contrast, the Decision Tree shows greater fluctuations in test accuracy, suggesting it is more prone to overfitting.

The Random Forest achieves a higher test accuracy, indicating it generalizes better on unseen data.

2.1)


Custom PCA Results!
Principal Component 2
Principal Component 1
class_0
class_1
class_2
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4


Sklearn PCA Results
Principal Component 2
Principal Component 1
class_0
class_1
class_2
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

2.2)

Benefits of PCA:

1. **Dimensionality Reduction:**
   - **PCA reduces the number of features, simplifying the dataset while retaining most of the variance.**
   - **It helps improve computational efficiency and reduces the risk of overfitting in machine learning models.**
2. **Feature Decorrelation:**
   - **PCA transforms correlated features into uncorrelated components, making it easier for models to interpret the data.**
3. **Noise Reduction:**
   - **By keeping only the most significant principal components, PCA can filter out noise from the dataset.**
4. **Visualization:**
   - **PCA is often used for visualizing high-dimensional data by reducing it to 2D or 3D.**

Disadvantages of PCA:

1. **Loss of Interpretability:**
   - **The transformed principal components are linear combinations of the original features, making them hard to interpret.**
2. **Assumes Linearity:**
   - **PCA assumes that the data's variability is linear and may not work well with nonlinear relationships.**
3. **Sensitivity to Scaling:**
   - **PCA is sensitive to the scale of the features. Standardizing the data is necessary before applying PCA.**
4. **Variance-Based:**
   - **PCA prioritizes features with the highest variance, which may not**

always align with the features most relevant to the target variable in supervised learning.

3.1)

Hard Margin SVM:

- **Definition**:
  - A hard margin SVM tries to find a hyperplane that perfectly separates the data into two classes without any misclassifications.
  - It requires the data to be **linearly separable**.

Soft Margin SVM:

- **Definition**:
  - A soft margin SVM allows some misclassifications or violations of the margin to account for noise or overlap in the data.
  - It introduces a **slack variable** to penalize misclassifications.

Advantages and Disadvantages of Both

Hard Margin SVM:

1. **Advantages**:
   - Guarantees a perfectly separating hyperplane if the data is linearly separable.
   - Simpler formulation without the need for additional parameters like the penalty term (C).
2. **Disadvantages**:
   - Cannot handle datasets with overlapping classes or noise.
   - Sensitive to outliers, as even a single misclassified point makes it infeasible to find a hyperplane.

Soft Margin SVM:

1. **Advantages**:
    - Works well with noisy or overlapping data by introducing flexibility.
    - Can generalize better to unseen data by finding a balance between maximizing the margin and minimizing classification errors.
2. **Disadvantages**:
    - Requires tuning of the penalty parameter (C), which controls the trade-off between margin maximization and error minimization.
    - May overfit if C is too large or underfit if C is too small.

3.2)