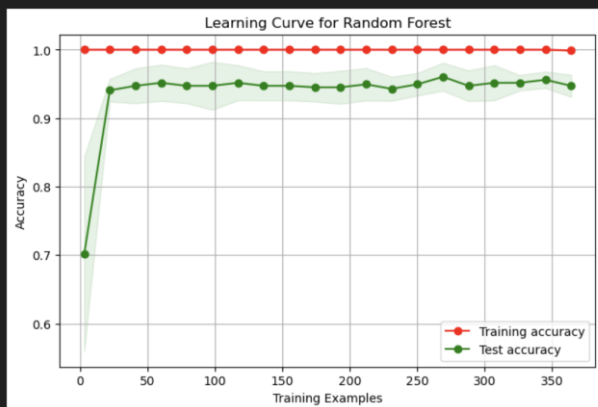
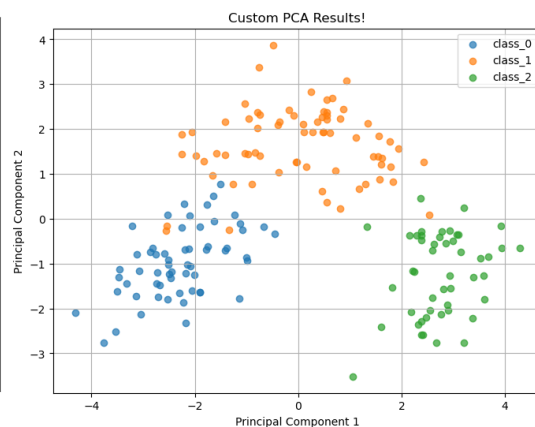
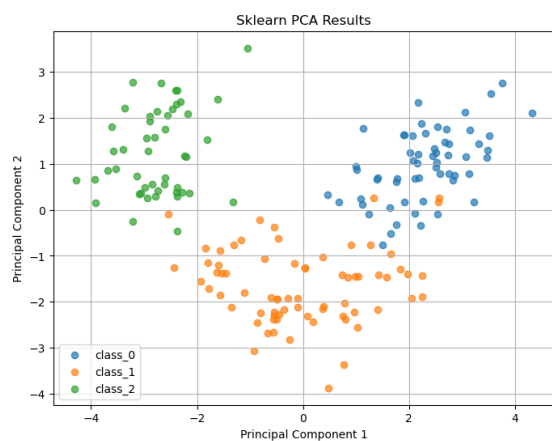


Decision Tree - Training Accuracy: 0.9934, Test Accuracy: 0.9211



Random Forest - Training Accuracy: 0.9978, Test Accuracy: 0.9649

랜덤 포레스트 쪽이 Accuracy가 더 높기 때문에 이 상황에서 더 적절한 모델이라고 할 수 있습니다.



PCA는 데이터의 주요 패턴을 캡처하면서 차원을 줄이는 기법입니다. 차원을 줄이고 불필요한 노이즈를 제거하며 시각적으로도 이해하기 쉽지만, 정보가 손실될 수 있고 선형 관계를 가정한다는 단점이 있습니다. 대체제로 비선형 기법인 UMAP가 있습니다. 데이터 구조를 잘 보존하지만 가격이 비쌉니다.

Soft Margin SVM은 일부 데이터의 오분류를 허용하나 Hard Margin은 모든 데이터의 올바른 분류를 가정합니다. Hard Margin은 완전 분리가 가능할 경우 정확도가 높으나 선형 분리가 불가능한 경우 문제가 생기며 과적합 문제도 있습니다. Soft Margin은 더 유연하나, 정확히 작동하게 하기 위해 많은 튜닝이 필요합니다.

