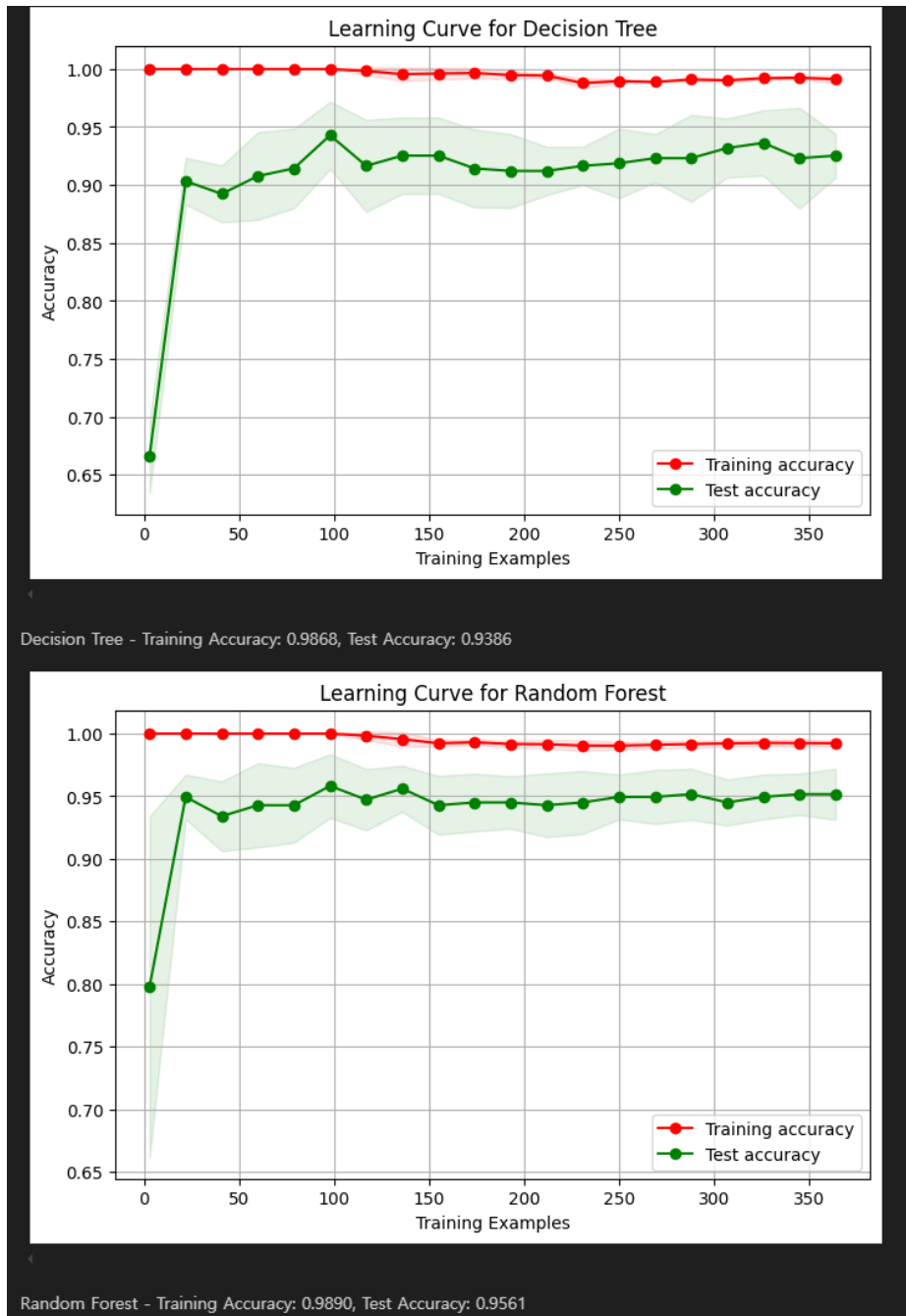


## ML Assignment Report

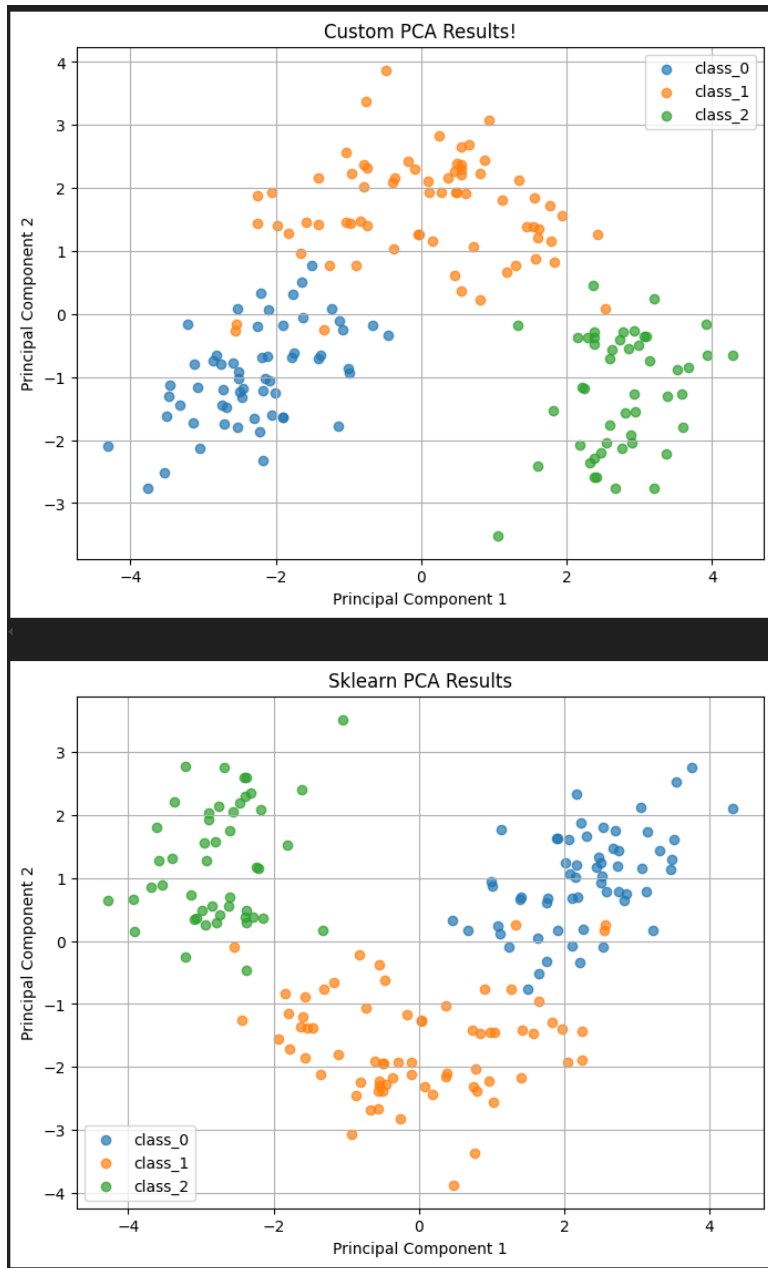
성우제

1.1)



1.2) Random Forest가 Decision Tree에 비해 accuracy가 높았다. 이는 앙상블 기법을 적용한 random forest가 다수의 decision tree 모델을 활용해서 분류를 진행하기 때문인 것으로 보이며, 따라서 random forest가 더 좋은 모델이다.

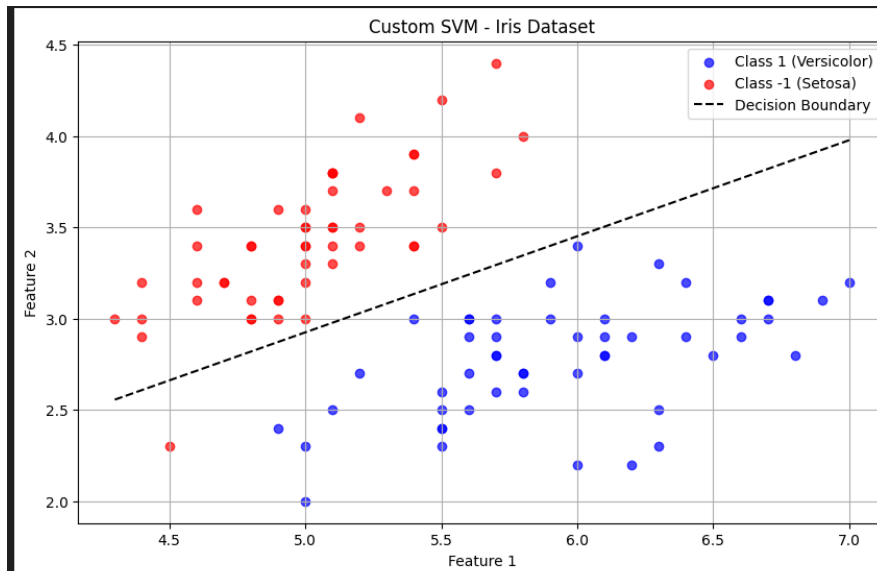
2.1)



2.2) PCA는 차원을 축소하여 차원의 저주를 해소해줄 수 있다. 또한 새롭게 추출한 특성은 서로 간의 공분산이 0이므로 다중공선성 또한 없다. 하지만 PCA로 만든 특성은 종속 변수에 대한 설명력을 보장해주지는 않는다. 단순히 분산을 크게 만들어준 것이기 때문이다.

3.1) Soft margin SVM은 과적합을 방지하기 위해 약간의 오분류를 허용하고, Hard margin SVM은 오분류를 허용하지 않는다. Soft margin은 bias가 hard margin에 비해 높을 순 있지만 variability가 작아 과적합에 강건하다.

3.2)



our result should look like the following!

