Mushroom Classification

-dead or alive-

Team Cerberos

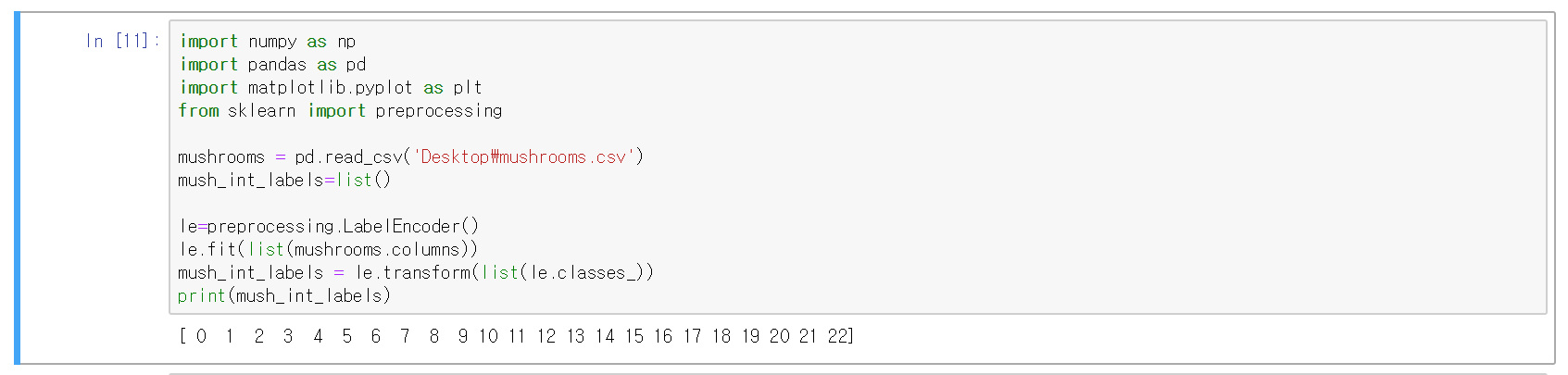
2018-07-02

**1. Introduction**

머신 러닝에 적합한 data set을 설정하여 예측모형을 만들고, ROC, AUC 곡선 등을 통해 분석해본다. 또한 결과를 github repo에 commit한다.

**2. Use sklearn package – ROC curve, AUC**

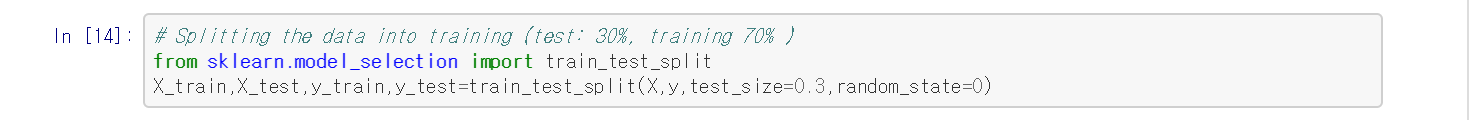
(1) Sklearn package를 통해 텍스트 데이터인 feature들을 int 형태로 바꿔준다.



(2) Models for prediction problems

- 70%의 데이터를 training에 사용하고 30%의 데이터를 test에 사용한다.

- 사용한 모델 : Logistic Regression, Decision Tree Regression, Random Forest Regression, Multi-Layer Perceptron (Machine Learning)







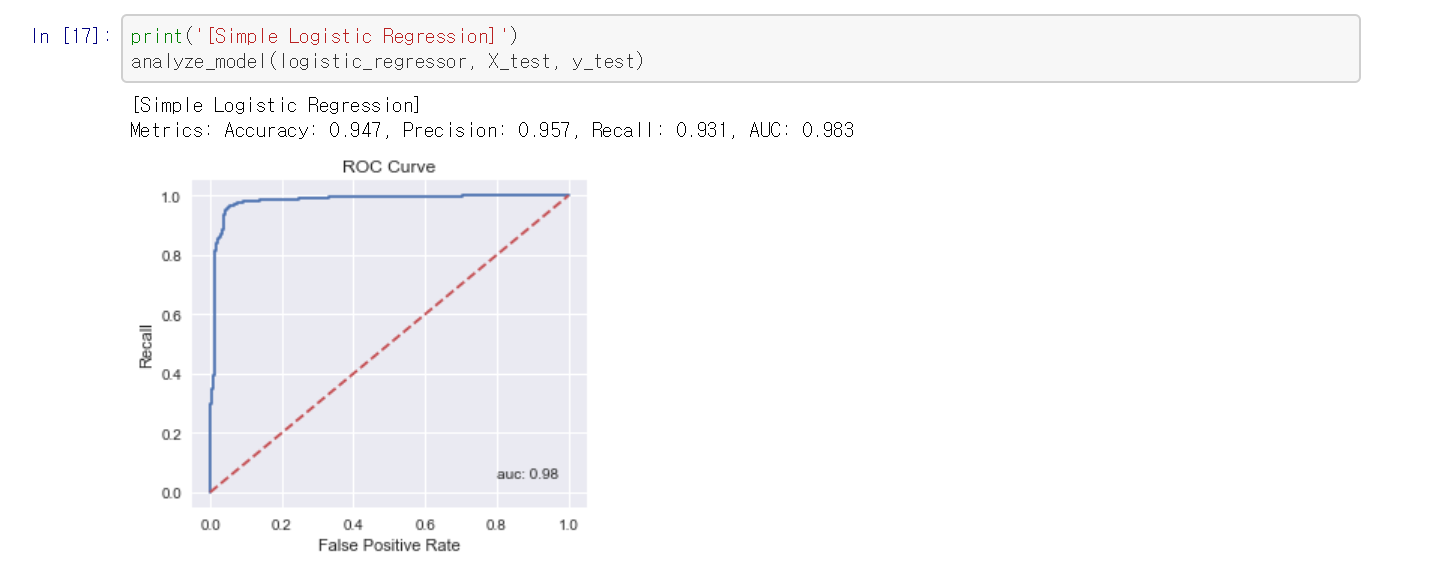
(3) Sklearn package를 통해 ROC 곡선을 그리고 AUC를 구해본다

- ROC 곡선 : FPR과 TPR을 각각 x, y축으로 놓은 그래프, 곡선이 왼쪽 위로 갈수록 FPR이 낮고 TPR이 높으므로 성능이 좋다.

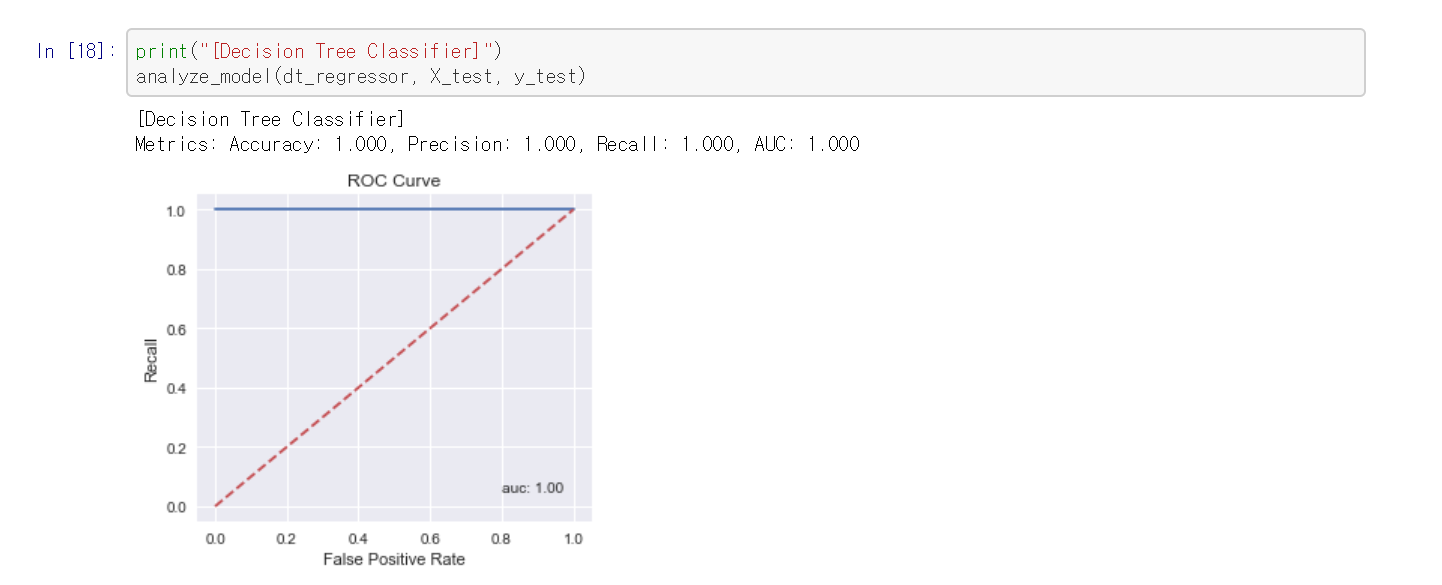
(TPR : 1인 케이스에 대하여 1이라고 예측한 비율 / FPR : 0인 케이스에 대하여 1이라고 예측한 비율)

- AUC : ROC 커브의 밑면적을 구한 값, 1에 가까울수록 성능이 좋다.

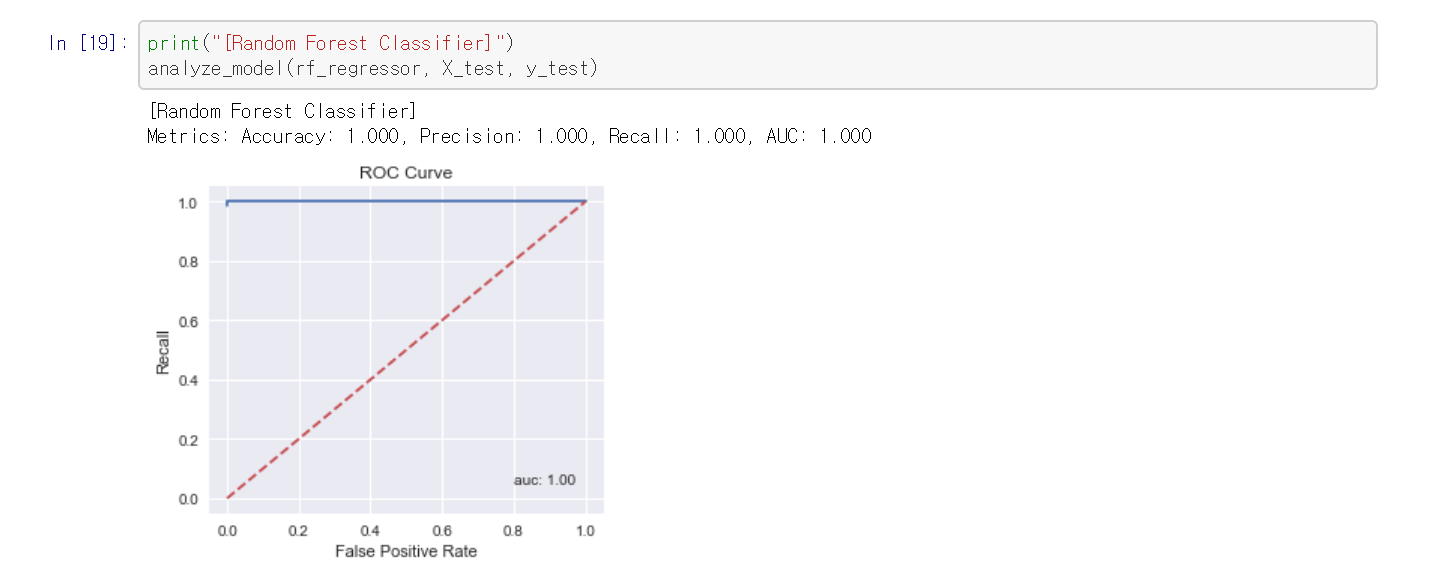
1) Logistic Regression :



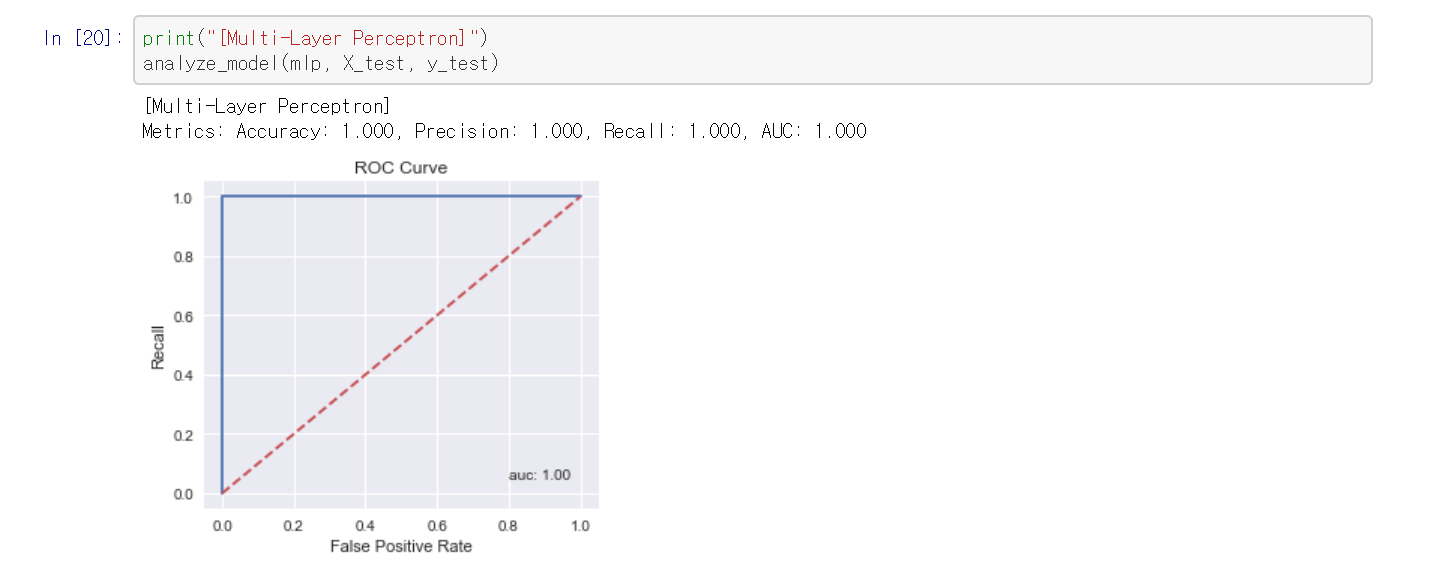
- Decision Tree Classifier :



- Random Forest Classifier :



- Multi-Layer Perceptron(MLP ; Machine Learning) :



**6. Think :**

**1) interprete the model**

**2) improve the model**