XGBoost

# 라이브러리 링크

**install :**

https://xgboost.readthedocs.io/en/stable/install.html

**github :**

[https://github.com/dmlc/xgboost/tree/master/python-package](https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/tree/_classes.py)

# 기초 설명

****

XGBoost는 매우 효율적이고 유연하게 최적화된 gradient boosting 라이브러리 이며.

Gradient Boosting 프레임워크를 통해 머신러닝 알고리즘을 구현한다.

또한 XGBoost는 많은 데이터 과학 문제를 빠르고 정확하게 해결할수 있는 병렬 트리 부스팅을 제공합니다.

# 버전 정보

* xgboost == 1.5.2 (pip install xgboost)

# 데이터셋 설명 및 출처

* Sklearn 내부의 load\_breast\_cancer 함수를 통해 UCI ML Breast Cancer Wisconsin (Diagnostic) datasets. 을 불러옴.
* Sources : https://scikit-learn.org/stable/datasets/toy\_dataset.html#breast-cancer-dataset

# 코드 설명

* sklearn에서 불러온 데이터셋을 XGBoost 에서 사용 가능한 DMatrix 형식으로 변환한뒤.이를 XGBoost 모델에 학습 하여 학습시킨 이후 이 학습된 모델을 사용하여 테스트용 데이터셋의 결과를 예측한뒤 앞부분 10개 항목에 대한 결과를 출력한다.
* 이후 acuity\_score,precision\_score,recall\_score,f1\_score,roc\_auc\_score 와 같은 검증용 함수들을 사용하여 XGBoost 모델의 예측 결과와 실제 값 사이의 정확도를 측정하여 이를 출력 및 모델 내부에서 각 Feature 들의 중요도를 그래프를 통해 나타낸다.

# 검증 방법

* 코드 내부에서 데이터셋을 학습용 데이터셋과 검증용 데이터셋으로 나누어 이를 검증함.

(test\_size = 0.2, random\_state = 30)

* 추가적으로 sklearn의 acuity\_score,precision\_score,recall\_score,f1\_score,roc\_auc\_score 함수를 사용하여 실제값과 예측값 사이의 적합도를 평가.