Light gbm

# 라이브러리 링크

**install :**

https://lightgbm.readthedocs.io/en/latest/Installation-Guide.html

**Sources :**

[https://lightgbm.readthedocs.io/en/latest/\_modules/lightgbm/sklearn.html#LGBMClassifier](https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/tree/_classes.py)

# 기초 설명

****

Light\_GBM 은 gradient boosting framework로써 tree 기반의 학습 알고리즘을 사용한다.

Light\_GBM은 더욱 빠르고 정확하면서도 메모리 소비가 적으며 분산, 병렬처리 및 GPU 연산기능을 사용할수 있도록 설계되었다.

# 버전 정보

* Python (64-bit as possible)
* wheel (pip install wheel)

For windows:

* VC runtime is needed if Visual Studio (2015 or newer) is not installed.

For Linux :

* glibc >= 2.14 (pip install glibc)
* pandas (pip install pandas)
* matplotlib (pip install matplotlib)

# 데이터셋 설명 및 출처

* Using Sklearn.datasets.load\_breast\_cancer.
* Sources : https://scikit-learn.org/stable/datasets/toy\_dataset.html#breast-cancer-dataset

# 코드 설명

* sklearn에서 불러온 데이터셋을 학습용, 데스트용 데이터셋으로 분할하여 이를 Light\_GBM모델에 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 Light\_GBM 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.