Bagging classifier

# library link

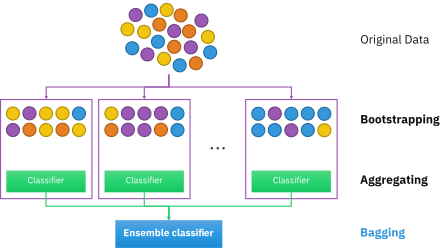
**install :**

<https://scikit-learn.org/stable/install.html>

**github :**

https://github.com/scikit-learn/scikit-learn/blob/82df48934eba1df9a1ed3be98aaace8eada59e6e/sklearn/ensemble/\_bagging.py

# basic description



bagging (from bootstrap aggregating) is a machine learning ensemble meta-algorrithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression.

# version

* NumPy >= 1.14.6 (pip install numpy)
* Scipy >= 1.1.0 (pip install scipy)
* Joblib >= 0.11 (pip install joblib
* Threadpoolctl >= 2.0.0 (pip install threadpoolctl)
* pandas >= 1.2.4 (pip install pandas)
* matplotlib == 3.22 (pip install matplotlib)
* sklearn == 1.0.2 (pip install sklearn)

# dataset

* diabetes.scv: This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.
* Sources : https://www.kaggle.com/uciml/pima-indians-diabetes-database

# code description

* Use sklearn BaggingClassifier to make Bagging\_Classifier model with decision tree.
* Read dataset and divided into train dataset and val dataset.
* train dataset in two model : Bagging\_Classifier model and Decision tree model.
* Check and print the model sore of both model.
* Make test\_data for prediction.
* print both model prediction result of test\_data.

# validation

* Inside the code, the dataset is divided into learning datasets and verification datasets to verify this.

(test\_size = 0.2, random\_state = 0)