Decision\_Tree (Gini,Entropy)

# library link

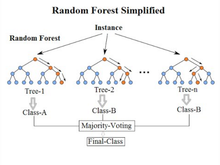
**install :**

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

**github :**

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

# basic description

****

Random forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

# 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)
* pandas (pip install pandas)

# dataset

* Housing.scv: the Data set of housing price, based on certain factors like house area, bedrooms, furnished, nearness to mainroad, etc.
* Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.  
  Belsley D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.
* Sources : <https://www.kaggle.com/yasserh/housing-prices-dataset>

# code description

* A model that learns it through Random Forest learning method based on data from datasets and predicts the number of bedrooms based on information such as house price, toilet number, warehouse number, etc.
* In contrast, the accuracy of the model changes depending on the difference between the geni and entropy methods in the Random Forest learning process and the difference in the number of trees used inside the Random Forest.

# validation

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

(test\_size = 0.2, random\_state = 42)

* Additionally, the fitness between the actual value and the predicted value is evaluated using Sklearn's acuity\_score function.