

## Data Science : Assignment 2 Decision Tree

2018008504 손승일

환경 :

Vscode, python으로 구현

실행 방법:

Vscode terminal에서 `python dt.py dt_train.txt dt_test.txt dt_result.txt`를 하면 `dt_result.txt`가 생성되고 `python dt.py dt_train1.txt dt_test1.txt dt_result1.txt` 을 입력하면 `dt_result1.txt`가 생성됨.

이후 test program파일에 `dt_result.txt`와 `dt_result1.txt`를 넣고 cmd창에서 `dt_test.exe dt_answer.txt dt_result.txt` 와 `dt_test.exe dt_answer1.txt dt_result1.txt`를 입력해서 실행하게 해서 점수를 받게 함.

구현 함수

1. `get_entropy()`

data들의 entropy 계산

2. `get_info_gain()`

Attribute들의 값의 비율을 구한다.

3. `get_next_attr()`

Attribute에 대해 information gain을 구하고 최대 gain을 갖는 attribute를 구하고 그걸 반환한다.

4. `decision_tree()`

해당 함수로 train하고 tree 모델을 생성한다.

5. `test_decision_tree()`

해당 함수로 test.txt 받아서 그걸로 결과값을 생성하게 한다.

6. `main()`

Train file로 tree 모델을 만들고 `test_decision_tree()` 함수로 `result.txt` 를 받게 한다

결과값

그냥 train과 test로 만든 `result.txt`

```
C:\Users\손승일\Desktop\학 교\데 사\Assignment2\test program>dt_test.exe dt_answer.txt dt_result.txt
5 / 5
```

Train1과 test1로 만든 `result1.txt`

```
C:\Users\손승일\Desktop\학 교\데 사\Assignment2\test program>dt_test.exe dt_answer1.txt dt_result1.txt
320 / 346
```