

Bài tập về nhà Decision Tree, Random Forest

- Thực hiện các yêu cầu dưới đây với tập dữ liệu Australian credit
- Down dữ liệu tại đường link dưới đây và đặt vào folder /data

[https://archive.ics.uci.edu/ml/datasets/Statlog+\(Australian+Credit+Approval\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval))

```
[3]: # import packages, ...  
  
# set random_state = 0
```

1. Tìm hiểu dữ liệu

```
[ ]: '''  
- Tìm hiểu cấu trúc data tại trang web UCI  
- Down các files, đặt trong folder data/  
- Chú ý:  
  - Đọc mô tả dữ liệu trong file 'australian.doc'  
  - Kiểm tra DecisionTree/RandomForest của Sklearn có hỗ trợ với kiểu dữ liệu  
    → được cung cấp không  
'''  
data_path = './data/australian.dat'
```

2. Decision Tree

2.1. Khảo sát các giá trị khác nhau của max_depth

```
[ ]: # code
```

2.2. Tìm kiếm siêu tham số

```
[ ]: #code
```

3. Random Forest

3.1 Khảo sát các giá trị khác nhau của n_estimators

```
[ ]: #code
```

3.2 Tìm kiếm siêu tham số

```
[ ]: # code
```