Minggu ke-6

Praktikum Decision Tree

Ali Ridho Barakbah

Knowledge Engineering Research Group
Department of Information and Computer Engineering
Electronic Engineering Polytechnic Institute of Surabaya

Klasifikasi dengan Decision Tree

```
from sklearn.tree import DecisionTreeClassifier

…

dtc=DecisionTreeClassifier()
dtc.fit(train_data, train_label)
class_result=dtc.predict(test_data)

…

acc=dtc.score(train_data, train_label)
```

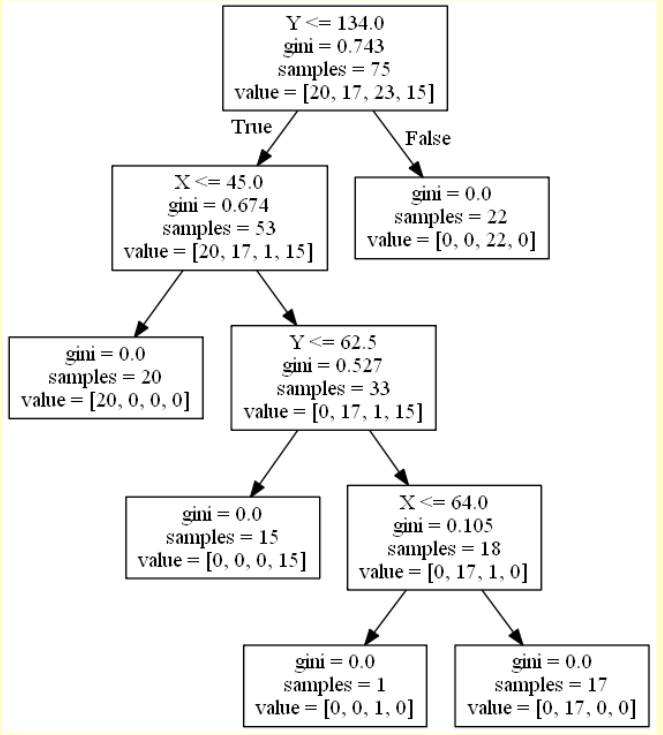
Menampilkan Gambar Decision Tree

```
from sklearn import tree
import graphviz

…

dot_data = tree.export_graphviz(dtc, out_file=None,
                                feature_names=train_data.columns.values)

graph = graphviz.Source(dot_data, format="png")
graph.render(view=True)
```

Contoh Klasifikasi pada Ruspini Dataset

```
import pandas as pd
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
import graphviz

dataset = pd.read_csv('ruspini.csv')

train_data=dataset[['X', 'Y']]
train_label=dataset[['CLASS']]

dtc=DecisionTreeClassifier()
dtc.fit(train_data, train_label)
test_data=[[130, 45]]
class_result=dtc.predict(test_data)
print('Class = ', class_result)

acc=dtc.score(train_data, train_label)
err=round((1-acc)*100, 2)
print('\n\nError ratio = ', err, '%')

dot_data = tree.export_graphviz(dtc, out_file=None,
feature_names=train_data.columns.values)


graph = graphviz.Source(dot_data, format="png")
graph.render( view=True)
```

```
Class =  [4]

Error ratio =  0.0 %
```



Latihan 6 - #Assignment
(Decision Tree dengan nilai numerik)

1. dataset ← titanic.csv
2. test_dataset ← titanic_test.csv
3. train_data ← ambil dataset kolom fitur (Sex, Age, Pclass, Fare). Lakukan pengisian missing value pada fitur Age dengan nilai mean dari masing-masing class
4. test_data ← ambil test_dataset kolom fitur (Sex, Age, Pclass, Fare).
5. train_label ← ambil dataset kolom kelas (Survived)
6. test_label ← titanic_testlabel.csv
7. Lakukan klasifikasi test_data terhadap train_data dengan Decision Tree, dan berapakah error rationya?
8. Tampilkan hirarki dari Decision Tree

Pengumpulan Tugas

- Buatlah coding dengan Bahasa pemrograman/tools apapun untuk semua assignment
- Buatlah laporan dalam slide ppt. Laporan terdiri dari screenshot coding dan hasil running untuk setiap assignment.
- Simpan laporan dalam file pdf dengan format penamaan:
  DM_M6_NRP_namadepan.pdf
- Upload file tersebut di alamat ini:
  http://ridho.lecturer.pens.ac.id/lecture.php
  (Pilih Mata Kuliah : Data Mining)
- Deadline upload: Minggu, 9 Mei 2021