

**TUGAS PENDAHULUAN PRAKTIKUM MODUL IV**  
**PEMBELAJARAN MESIN**



Oleh

**Nama : Muhammad Reza Putra Anisi    Nama : Muhammad Tijan Abdurahman**

**NIM :105220053    NIM : 105220023**

**PROGRAM STUDI ILMU KOMPUTER**  
**FAKULTAS SAINS DAN KOMPUTER**  
**UNIVERSITAS PERTAMINA**  
**2022**

Pertanyaan :

1. Jelaskan cara kerja algoritma decision tree pada library python (sertakan nama library)!
2. Jelaskan cara melakukan split pada pohon keputusan!

*Bonus: Jelaskan teknik Pre-pruning dan Post-pruning yang kamu ketahui!*

Jawaban :

1. Algoritma decision tree adalah metode pembelajaran mesin yang populer untuk tugas klasifikasi dan regresi. Dalam algoritma ini, model membangun pohon keputusan berdasarkan fitur-fitur input dan label output yang tersedia pada data pelatihan. Pohon keputusan ini dapat digunakan untuk memprediksi kelas atau nilai target untuk data baru dengan menerapkan aturan keputusan yang diambil dari cabang pohon.

Library Python yang umum digunakan untuk membangun model decision tree adalah scikit-learn. Berikut adalah langkah-langkah dasar untuk menggunakan scikit-learn dalam membangun model decision tree:

- Persiapkan data pelatihan dan data pengujian dalam bentuk numpy array -atau pandas dataframe.
- Import DecisionTreeClassifier dari scikit-learn.
- Buat objek DecisionTreeClassifier dengan parameter yang diinginkan.
- Latih model pada data pelatihan dengan memanggil method fit() pada objek DecisionTreeClassifier.
- Lakukan prediksi pada data pengujian dengan memanggil method predict() pada objek DecisionTreeClassifier. (Tulis jawaban dari pertanyaan diatas)

Referensi :

Scikit-learn documentation: <https://scikit-learn.org/stable/index.html>

Decision Trees in Scikit-learn: <https://scikit-learn.org/stable/modules/tree.html>

Iris Dataset:

[https://scikitlearn.org/stable/modules/generated/sklearn.datasets.load\\_iris.html](https://scikitlearn.org/stable/modules/generated/sklearn.datasets.load_iris.html)

2. Split atau pemisahan pada pohon keputusan adalah proses pembuatan keputusan untuk membagi data menjadi subset yang lebih kecil di setiap simpul pohon. Tujuan dari pemisahan ini adalah untuk meningkatkan kemampuan pohon keputusan dalam memisahkan kelas atau nilai target yang berbeda pada data.

Untuk melakukan split pada pohon keputusan, algoritma decision tree mencari fitur (atribut) yang paling informatif untuk memisahkan data pada simpul saat ini. Untuk menentukan atribut yang paling informatif, beberapa metrik pemilihan fitur dapat digunakan, seperti Information Gain, Gain Ratio, dan Gini Index.

Setelah fitur yang paling informatif dipilih, algoritma decision tree memisahkan data berdasarkan nilai fitur tersebut. Setiap nilai fitur akan memberikan dua cabang pada simpul saat ini, satu untuk nilai fitur yang lebih kecil atau sama dengan nilai ambang batas yang telah ditentukan, dan yang lainnya untuk nilai fitur yang lebih besar dari ambang batas. Pemisahan dilakukan untuk setiap nilai fitur yang unik pada fitur yang dipilih.

Referensi:

<https://towardsdatascience.com/how-decision-tree-algorithm-works-90d8b0a0e8f6>

<https://scikit-learn.org/stable/modules/tree.html#tree-algorithms-id3-c4-5-c5-0-and-cart>

*Bonus :*

Pre-pruning

Pre-pruning adalah teknik yang dilakukan sebelum pembentukan pohon keputusan untuk mencegah terlalu banyak cabang dan simpul pada pohon keputusan. Pre-pruning dilakukan dengan menghentikan rekursi pembentukan pohon keputusan sebelum mencapai kondisi terminasi standar atau batas kedalaman maksimum. Beberapa metode pre-pruning yang dapat digunakan, antara lain:

- Maximum Depth: Menghentikan pembentukan pohon keputusan setelah mencapai kedalaman maksimum tertentu.
- Minimum Sample Split: Menghentikan pembentukan pohon keputusan jika jumlah sampel pada simpul saat ini kurang dari ambang batas tertentu.
- Minimum Leaf Sample: Menghentikan pembentukan pohon keputusan jika jumlah sampel pada simpul daun kurang dari ambang batas tertentu.

Post-pruning

Post-pruning adalah teknik yang dilakukan setelah pembentukan pohon keputusan untuk memangkas pohon keputusan yang terlalu kompleks dan mencegah overfitting. Post-pruning dilakukan dengan menghapus simpul dan cabang yang tidak signifikan pada pohon keputusan. Beberapa metode post-pruning yang dapat digunakan, antara lain:

- Reduced Error Pruning: Menghapus simpul dan cabang yang tidak signifikan pada pohon keputusan berdasarkan perubahan error rate setelah penghapusan simpul dan cabang.
- Cost Complexity Pruning: Menghapus simpul dan cabang yang tidak signifikan pada pohon keputusan dengan meminimalkan fungsi objektif berdasarkan kompleksitas pohon keputusan dan error rate.

Referensi:

<https://towardsdatascience.com/how-to-prune-a-decision-tree-2c578130303d>

<https://scikit-learn.org/stable/modules/tree.html#tree-algorithms-id3-c4-5-c5-0-and-cart>