

Classification Tree

Cara Kerja:

1. Pilih fitur terbaik untuk *split* data dengan menggunakan *Gini Impurity*, *Entropy*, atau kriteria lainnya.
2. Bagi data ke dalam dua kelompok, satu kelompok untuk data yang memenuhi kondisi *split* dan kelompok lainnya untuk data yang tidak memenuhi kondisi *split*.
3. Ulangi langkah 1 dan 2 hingga terbentuk pohon dengan *node* memiliki kelas yang sama atau *maximum depth* terpenuhi.
4. Gunakan pohon yang telah terbuat untuk menentukan kelas dari data yang ingin ditentukan

Perbandingan:

- *Scratch*

```
[33]: from decisionTree import DecisionTreeClassifierScratch

ct_scratch = DecisionTreeClassifierScratch()
ct_scratch.fit(X_train, y_train)
y_pred_ct_scratch = ct_scratch.predict(X_test)

validate_model(ct_scratch, method_name="Classification Tree from Scratch")
```

Hold-Out Validation (Classification Tree from Scratch):
F1 Score: 0.891566265060241

	precision	recall	f1-score	support
0	0.93	0.94	0.94	72
1	0.90	0.88	0.89	42
accuracy			0.92	114
macro avg	0.92	0.91	0.91	114
weighted avg	0.92	0.92	0.92	114

K-Fold Cross-Validation (Classification Tree from Scratch):
F1 Scores for each fold: [0.9428571428571428, 0.9014084507042254, 0.8857142857142857, 0.9166666666666666, 0.819672131147541]
Mean F1 Score: 0.8932637354179722
Standard Deviation of F1 Score: 0.04134309627852532

- *Library*

```
[34]: from sklearn.tree import DecisionTreeClassifier

ct = DecisionTreeClassifier()
ct.fit(X_train, y_train)
y_pred_ct = ct.predict(X_test)

validate_model(ct, method_name="Classification Tree from Library")
```

Hold-Out Validation (Classification Tree from Library):
F1 Score: 0.9047619047619048

	precision	recall	f1-score	support
0	0.94	0.94	0.94	72
1	0.90	0.90	0.90	42
accuracy			0.93	114
macro avg	0.92	0.92	0.92	114
weighted avg	0.93	0.93	0.93	114

K-Fold Cross-Validation (Classification Tree from Library):
F1 Scores for each fold: [0.9411764705882353, 0.9552238805970149, 0.8695652173913043, 0.9459459459459459, 0.8064516129032258]
Mean F1 Score: 0.9036726254851452
Standard Deviation of F1 Score: 0.0573870098510817

Dari implementasi secara *scratch* dan *library*, terlihat bahwa penggunaan *library* memiliki nilai F1 score yang lebih tinggi.

Improvement:

Improvement yang dapat dilakukan pada algoritma *Classification Tree* secara *scratch* dapat dilakukan dengan memperbaiki kriteria *split* yang dilakukan, yang dalam implementasi ini menggunakan *Gini Impurity*.