

# DecisionTreeClassifier

February 27, 2023

## 1 Explorasi Decision Tree Classifier

Load libraries yang diperlukan untuk melakukan pembelajaran.

```
[1]: # Load libraries
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.tree import DecisionTreeClassifier, export_text
from sklearn.model_selection import cross_validate
import pickle
```

Baca datanya dan bagi menjadi data training dan test.

```
[2]: # Load the data
breast_cancer = load_breast_cancer()

# split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(breast_cancer.data,
    ↪breast_cancer.target, test_size=0.2, random_state=42)
```

Latih algoritma dengan dataset training.

```
[3]: # Make an object and fit the data
DTC = DecisionTreeClassifier()
DTC.fit(X_train, y_train)
```

```
[3]: DecisionTreeClassifier()
```

Simpan model pada suatu file, kemudian load kembali file tersebut.

```
[4]: # Save the model
with open('DecisionTreeClassifier_model.pkl', 'wb') as f:
    pickle.dump(DTC, f)
```

```
[5]: # Load the model
with open('DecisionTreeClassifier_model.pkl', 'rb') as f:
    DTC = pickle.load(f)
```

Lakukan prediksi menggunakan algoritma yang telah di-train menggunakan dataset test.

```
[6]: # Predict result
y_pred = DTC.predict(X_test)
```

Tampilkan confusion matrix dari prediksi

```
[7]: cm = confusion_matrix(y_test, y_pred)
print(cm)
```

```
[[39  4]
 [ 4 67]]
```

Tampilkan laporan perbandingan hasil prediksi dengan hasil asli prediksi.

```
[8]: report = classification_report(y_test, y_pred)
print(report)
```

	precision	recall	f1-score	support
0	0.91	0.91	0.91	43
1	0.94	0.94	0.94	71
accuracy			0.93	114
macro avg	0.93	0.93	0.93	114
weighted avg	0.93	0.93	0.93	114

Terdapat tiga metrik penilaian yaitu Precision, Recall, dan F1-score. Berikut penjelasan singkat masing-masing: - Precision: Dari semua tebakan positif, berapa banyak yang benar-benar positif? - Recall: Dari semua data yang positif, berapa banyak yang benar ditebak positif? - F1-score: Harmonic Mean dari Precision dan Recall.

Nilai tersebut merupakan nilai per kelas, sehingga untuk didapatkan nilai keseluruhan dilakukan rata-rata dengan macro average dan weighted average.

Tampilkan hasil learning kedalam bentuk tree.

```
[9]: # Hasil Tree
dtree = export_text(DTC, feature_names=list(breast_cancer["feature_names"]))
print(dtree)
```

```
|--- mean concave points <= 0.05
|   |--- worst radius <= 16.83
|   |   |--- radius error <= 0.63
|   |   |   |--- worst smoothness <= 0.18
|   |   |   |   |--- smoothness error <= 0.00
|   |   |   |   |   |--- worst concavity <= 0.19
|   |   |   |   |   |   |--- class: 1
|   |   |   |   |   |   |--- worst concavity > 0.19
|   |   |   |   |   |   |   |--- class: 0
```

```

| | | | |--- smoothness error > 0.00
| | | | | |--- worst texture <= 33.35
| | | | | | |--- class: 1
| | | | | |--- worst texture > 33.35
| | | | | | |--- worst texture <= 33.56
| | | | | | |--- class: 0
| | | | | | |--- worst texture > 33.56
| | | | | | |--- class: 1
| | | | |--- worst smoothness > 0.18
| | | | |--- class: 0
| | | |--- radius error > 0.63
| | | |--- mean smoothness <= 0.09
| | | |--- class: 1
| | | |--- mean smoothness > 0.09
| | | |--- class: 0
| |--- worst radius > 16.83
| | |--- mean texture <= 16.19
| | |--- class: 1
| | |--- mean texture > 16.19
| | |--- concave points error <= 0.01
| | |--- class: 0
| | |--- concave points error > 0.01
| | |--- class: 1
|--- mean concave points > 0.05
| |--- worst concave points <= 0.15
| | |--- worst perimeter <= 115.25
| | |--- worst texture <= 27.43
| | |--- class: 1
| | |--- worst texture > 27.43
| | |--- class: 0
| | |--- worst perimeter > 115.25
| | |--- class: 0
| |--- worst concave points > 0.15
| | |--- fractal dimension error <= 0.01
| | |--- class: 0
| | |--- fractal dimension error > 0.01
| | |--- class: 1

```

Lakukan langkah pembelajaran yang sama menggunakan cross validation.

```
[10]: cv_result = cross_validate(DecisionTreeClassifier(), X_train, y_train, cv=10,
    ↪return_estimator=True)
```

```
[11]: DTC_CV = cv_result["estimator"][cv_result["test_score"].argmax()]
```

```
[12]: y_pred_cv = DTC_CV.predict(X_test)
```

```
[13]: cm_cv = confusion_matrix(y_test, y_pred)
      print(cm_cv)
```

```
[[39  4]
 [ 4 67]]
```

```
[14]: report_cv = classification_report(y_test, y_pred_cv)
      print(report_cv)
```

	precision	recall	f1-score	support
0	0.89	0.95	0.92	43
1	0.97	0.93	0.95	71
accuracy			0.94	114
macro avg	0.93	0.94	0.94	114
weighted avg	0.94	0.94	0.94	114

Dapat dilihat dari rata-rata metrik bahwa prediksi menggunakan cross validation sedikit lebih baik daripada sebelumnya.