

Ensemble Learning (Random Forest)

Adinda Putri - 13523071

Ensemble learning merupakan sebuah teknik di mana beberapa model (disebut juga **base learners** atau **weak learners**) digabungkan untuk meningkatkan performa dibanding dengan menggunakan model tunggal. Salah satu teknik ensemble learning adalah **Bagging (Bootstrap Aggregating)**. Bagging melibatkan pembuatan beberapa subset dari data pelatihan dengan penggantian (bootstrap) dan melatih model pada tiap subset. Kemudian, diambil hasil prediksi mayoritas untuk classification, atau rata-rata prediksi untuk regression.

Random forest merupakan salah satu implementasi dari bagging. Random Forest menggunakan **decision tree** sebagai base learner dan secara acak memilih subset dari fitur-fitur yang tersedia untuk setiap pohon serta data pelatihan, yang menghasilkan berbagai decision tree berbeda. Pohon-pohon ini digabungkan untuk membuat prediksi akhir.

Cara Kerja:

1. Buat n estimators decision trees. Tiap decision tree tersebut menggunakan bagian acak dari data.
2. Pilih fitur secara acak untuk tiap tree.
3. Tiap tree membuat prediksi berdasarkan bagian datanya
4. Gabungkan prediksi. Untuk classification, prediksi akhir adalah prediksi mayoritas, sedangkan untuk regression prediksi akhir adalah rata-rata dari hasil prediksi semua tree

Perbandingan model dari scratch dengan dari Scikit-Learn

```
BONUS: Ensemble Methods

Bagging

K-Fold Cross-Validation Random Forest

evaluate_model_classification(rf_pipe, X_train, y_train, kf)
✓ 11m 15.9s Python

--- Evaluating Model with K-Fold Cross Validation ---
Rata-rata Akurasi: 80.15%
Rata-rata Presisi: 82.14%
Rata-rata Recall: 81.32%
Rata-rata F1-Score: 81.61%

K-Fold Cross-Validation Random Forest dari Scikit-Learn

evaluate_model_classification(rf_scikit_pipe, X_train, y_train, kf)
✓ 2.5s Python

--- Evaluating Model with K-Fold Cross Validation ---
Rata-rata Akurasi: 79.20%
Rata-rata Presisi: 81.57%
Rata-rata Recall: 79.85%
Rata-rata F1-Score: 80.50%
```

Gambar 1. K-Fold Cross-Validation Random Forest dari Scratch dan Random Forest dari Sklearn
Sumber: Penulis

```
Hold-Out Validation Random Forest

rf = RandomForest(n_estimators=50, max_depth=25, min_samples_split=2)
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
print(classification_report(y_test, y_pred))
✓ 1m 15.4s Python
```

	precision	recall	f1-score	support
0	0.70	0.80	0.75	75
1	0.85	0.76	0.80	109
accuracy			0.78	184
macro avg	0.77	0.78	0.77	184
weighted avg	0.79	0.78	0.78	184

Gambar 2. Hold-Out Validation Random Forest dari Scratch
Sumber: Penulis



Gambar 3. Hold-Out Validation Random Forest dari Sklearn

Sumber: Penulis

Dari evaluasi terlihat bahwa hasil prediksi model dari random forest from scratch mendekati model dari Scikit-Learn. Meskipun terdapat perbedaan kecil, yaitu $< 2\%$. K-Fold dan Hold-Out dari kedua model juga menunjukkan performa yang baik, tidak ada variasi ekstrim yang bisa menjadi pertanda overfitting. Perbedaan hasil evaluasi ini dapat terjadi karena perbedaan splitting algorithm, randomness dan sampling, serta implementasi internal dari Scikit-Learn yang sudah highly optimized. Perbedaan implementasi internal ini juga yang membuat waktu eksekusi untuk Random Forest dari scratch sangat lama bahkan hingga 600x lebih lambat dari model Scikit-Learn.

Ruang Improvement

- Untuk meningkatkan optimasi dan efisiensi, dapat digunakan library multiprocessing untuk eksekusi concurrent dan parallel processing.
- Memory Optimization dengan cara memilih struktur data yang lebih efisien.
- Bisa ditambahkan juga log output mengenai hasil split dan node baru tiap epoch agar konstruksi tree bisa dimonitor

Referensi:

[1] *Kenali Ensemble Learning pada Tipe Machine Learning*, DQLab, 1 September 2023. [Daring]. Tersedia: <https://dqlab.id/kenali-ensemble-learning-pada-tipe-machine-learning>. [Diakses: 4 September 2025].

[2] *Random Forest Algorithm in Machine Learning*, GeeksforGeeks, diperbarui 1 September 2025. [Daring]. Tersedia:

<https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/>.
[Diakses: 4 September 2025].