

Muhammad Harits

1103223159

Bab 7: Ensemble Learning dan Random Forests

Ringkasan Bab

Bab ini membahas ide ensemble learning: menggabungkan banyak model agar performanya lebih baik daripada hanya satu model terbaik.

Konsep Utama

- Voting Classifier
 - Hard Voting: mayoritas suara dari seluruh model.
 - Soft Voting: rata-rata probabilitas kelas — biasanya lebih akurat.
- Bagging (Bootstrap Aggregating)
 - Melatih model yang sama pada subset data yang berbeda.
 - Sampling dengan penggantian (with replacement).
- Pasting
 - Sama seperti bagging, tetapi tanpa penggantian.
- Out-of-Bag (OOB) Evaluation
 - Data yang tidak terambil oleh sampling menjadi validation otomatis.
- Random Forest
 - Ensemble dari banyak Decision Tree + bagging.
 - Penambahan randomness: setiap node hanya mempertimbangkan subset fitur.
 - Lebih stabil dan akurat dibanding Decision Tree tunggal.
- Boosting
 - Melatih model secara berurutan, setiap model fokus memperbaiki kesalahan model sebelumnya:
 - AdaBoost
 - Gradient Boosting (GBRT)
- Stacking (Stacked Generalization)
 - Alih-alih voting, prediksi dari beberapa model digunakan sebagai input model lain (meta-learner).