

Chapter 7 : Summary

Diedrick Darrell Darmadi - 1103223031

Bab 7 menjelaskan konsep dasar ensemble learning, yaitu teknik menggabungkan banyak model lemah untuk membentuk model keseluruhan yang lebih kuat. Konsep ini diperkenalkan melalui contoh voting classifier yang menggabungkan beberapa model dasar dan menghasilkan performa lebih baik dibanding model tunggal.

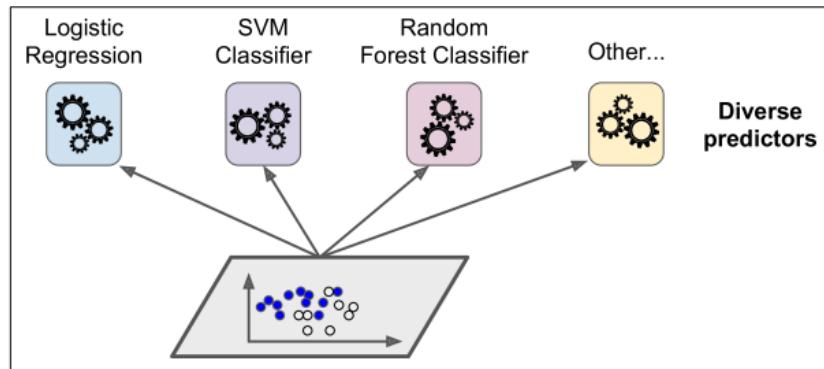


Figure 7-1. Training diverse classifiers

Bab ini kemudian membahas dua kategori utama ensemble: *bagging* (bootstrap aggregating) dan *pasting*. Bagging menggunakan sampling dengan pengembalian untuk membuat subset data bagi tiap model, sementara pasting menggunakan sampling tanpa pengembalian. Efek bagging terhadap pengurangan varians ditunjukkan melalui contoh visual pembagian sampel dan hasil agregasi.

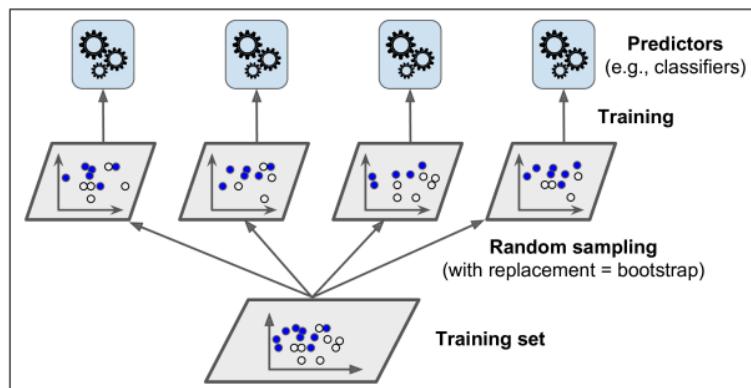


Figure 7-4. Pasting/bagging training set sampling and training

Setelah itu dijelaskan *Random Forest*, yaitu ensemble dari banyak decision tree yang dilatih pada subset data dan subset fitur acak. Pengacakan fitur membuat tiap pohon berbeda sehingga prediksi keseluruhan lebih stabil. Contoh bagaimana pohon-pohon acak menghasilkan decision boundary yang lebih halus dan general ditampilkan dalam bab ini.

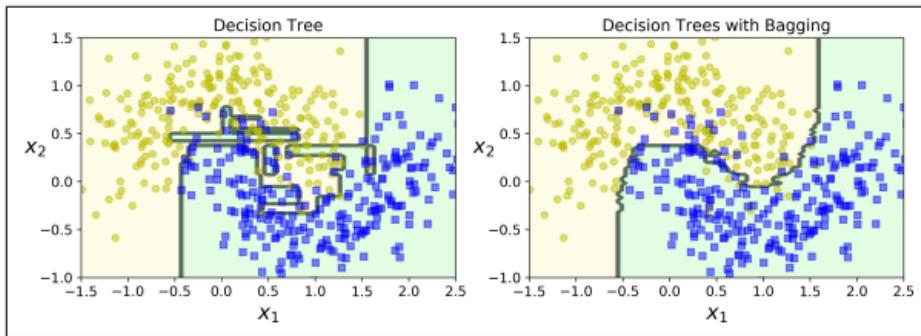


Figure 7-5. A single Decision Tree versus a bagging ensemble of 500 trees

Bab ini berlanjut ke pembahasan *Out-of-Bag (OOB) evaluation*, yaitu teknik validasi otomatis pada data yang tidak ikut terpilih saat bootstrap. OOB score memberikan estimasi performa tanpa memerlukan validation set tambahan. Bagian berikutnya membahas *feature importance* pada Random Forest, yaitu penilaian pentingnya suatu fitur berdasarkan kontribusinya dalam mengurangi impurity.

Ensemble boosting kemudian diperkenalkan sebagai teknik yang membangun model secara bertahap, di mana setiap model baru fokus pada kesalahan model sebelumnya. Dua metode utama dibahas: AdaBoost dan Gradient Boosting. Contoh pengaruh boosting terhadap decision boundary untuk beberapa iterasi ditampilkan secara bertahap

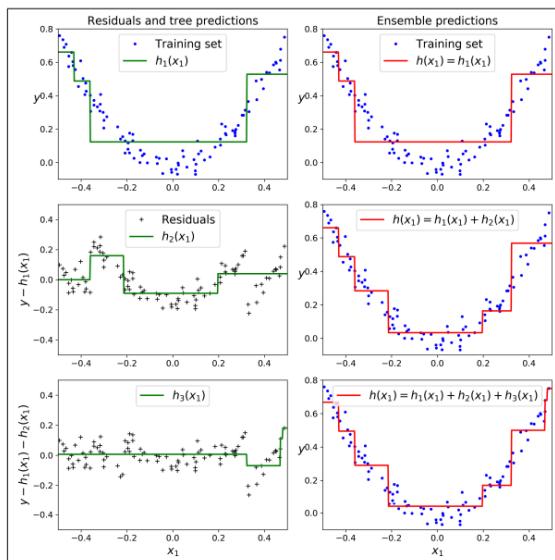


Figure 7-9. Gradient Boosting

Bab ini ditutup dengan penjelasan mengenai *Stacking*, yaitu teknik di mana model-model level pertama digabungkan menggunakan model meta-learner di level kedua.