

Radom Forest

Decision Tree mudah dibuat, mudah digunakan, dan mudah diinterpretasikan decision tree memiliki satu aspek yang menghalanginya untuk menjadi alat yang ideal untuk pembelajaran prediktif yakni dalam akurasi. Dengan kata lain, yang berfungsi dengan baik pada data yang digunakan untuk membuatnya tetapi tidak fleksibel dalam mengklasifikasikan sampel baru. random forest menggabungkan kesederhanaan decision tree dengan fleksibilitas dengan menghasilkan peningkatan besar dalam akurasi.

Cara membuat random forest

- Menyiapkan bootstrapped dataset, dibuat dengan cara mengambil data acak dengan ukuran yang sama dari dataset aslinya. Dalam pengambilan acak ini kita dapat mengambil data yang sama lebih dari sekali.
- Membuat decision tree menggunakan bootstrapped dataset namun hanya gunakan bagian yang acak dari variabel atau kolom pada tiap tahap. Dua kolom dipilih secara acak dan yang terbaik untuk pemisahan cabang akan dipilih sebagai root. Dilanjutkan dengan memilih dua kolom acak lagi sebagai cabang dan seterusnya hingga mencapai leaf.
- Setelah decision tree selesai dibuat, buat decision tree lagi menggunakan bootstrapped dataset yang sama menggunakan semua kemungkinan decision tree yang dapat dibuat (idealnya dilakukan sebanyak 100 kali). Semua decision tree yang terbuat akan menghasilkan output dan mayoritas dari semua output tersebut yang akan menjadi hasil akhir dari program.

Mengoptimalkan random forest, ketika membuat pohon langkah pertama yang dilakukan hanya menggunakan dua variabel yaitu kolom data untuk membuat keputusan pada setiap langkah. Kemudian membandingkan *error out-of-bag* untuk random forest yang dibangun hanya dengan menggunakan dua variabel per-langkah ke random forest yang dibangun dengan menggunakan tiga variabel per langkah lalu menguji banyak pengaturan berbeda dan memilih random forest yang paling akurat. Pertama membangun random forest dan kedua memperkirakan keakuratan random forest lalu mengubah jumlah variabel yang digunakan per langkah kemudian melakukannya berkali-kali dan memilih salah satu yang paling akurat.