

Nama: Nella Aprilia

NIM: 1103210185

Kelas: TK-45-04

StatQuest: Random Forests Bagian 1 - Membangun, Menggunakan dan Mengevaluasi

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.

Langkah ke-1, membuat kumpulan data pada bootstrap. Untuk membuat kumpulan data bootstrap yang ukurannya sama dengan aslinya, kemudian hanya memilih sampel secara acak dari kumpulan data asli. Langkah Ke-2, membuat Decision Tree subset variabel acak di setiap langkah untuk membuat Random Forest adalah dengan cara membuat Decision Tree menggunakan dataset bootstrap namun hanya menggunakan sub kumpulan variabel atau kolom acak pada setiap langkah. Dalam contoh hanya akan mempertimbangkan dua variabel atau kolom pada setiap langkah. Langkah Ke-3 ulangi langkah 1 dan 2 beberapa kali. Sekarang kembali ke langkah pertama dan ulangi kemudian membuat kumpulan data bootstrap baru dan buat pohon dengan mempertimbangkan sub kumpulan variabel di setiap langkah. Menggunakan sampel bootstrap dan hanya mempertimbangkan sebagian variabel pada setiap langkah akan menghasilkan beragam pohon. Keanekaragaman inilah yang menjadikan Random Forest lebih efektif dibandingkan Decision Tree individual.

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.