

Nama : Tirtayuda Munggarana

NIM : 1103202108

Kelas : TK-44-G04

Rangkuman Random Forest

Pohon keputusan adalah model yang mudah digunakan dan diinterpretasikan. Sementara itu, random forest memerlukan pembuatan dataset seed dengan ukuran yang sama dengan data asli, dengan langkah selanjutnya adalah membuat pohon keputusan menggunakan dataset bootstrap seed.

Data bootstrap adalah pengambilan sampel dengan teknik penggantian yang digunakan untuk membuat banyak sampel baru dari kumpulan data. Dalam konteks pembelajaran mesin, bootstrapping sering digunakan untuk membuat kumpulan data baru yang ukurannya sama dengan data asli.

Proses ini dilakukan dengan memilih sampel secara acak dengan penggantian data asli sesuai ukuran yang diinginkan. Hutan acak menggunakan subset variabel acak pada setiap langkah untuk membangun pohon. Hanya dua variabel yang dipertimbangkan pada setiap langkah, dan pemilihan variabel yang dipertimbangkan bersifat acak.

Dengan menggunakan sampel bootstrap dan mempertimbangkan subset variabel pada setiap langkah, berbagai jenis pohon dapat dihasilkan. Saat mengevaluasi kinerja hutan acak, data yang tidak disertakan dalam kumpulan data bootstrap disebut data "out-of-bag" (OOB).

Setiap pohon individu di hutan acak dievaluasi menggunakan data OOB, di mana pohon tersebut dapat memberi label sampel OOB dengan benar atau salah. Akurasi hutan acak dapat diukur sebagai proporsi sampel OOB yang diklasifikasikan dengan benar.

Proporsi sampel OOB yang diklasifikasikan dengan benar berhubungan dengan keakuratan hutan acak, sedangkan proporsi sampel OOB yang salah diklasifikasikan berhubungan dengan kesalahan OOB. Untuk memilih konfigurasi hutan acak yang paling akurat, percobaan dilakukan menggunakan berbagai konfigurasi dan dipilih hutan acak dengan akurasi tertinggi.