

NAMA : ILYASIN BINTANG ANUGRAH

NIM : 1103228242

Random forest merupakan salah satu algoritma machine learning yang umum digunakan oleh data scientist, khususnya dalam permasalahan klasifikasi dan regresi. Random forest membutuhkan gabungan beberapa decision tree untuk memprediksi hasil yang akurat.

Konsep sederhana dari random forest adalah beberapa decision tree yang tidak berkorelasi akan bekerja lebih baik sebagai kelompok dibandingkan individu.

Saat menggunakan random forest sebagai pengklasifikasi, satu decision tree menyumbang satu suara. Setiap decision tree bisa menghasilkan jawaban yang sama atau berbeda satu sama lain. Misalnya decision tree A, B, E dan F memprediksi hasil 1. Sementara decision tree C dan D memprediksi hasil 0. Karena ada banyaknya alternatif jawaban dalam decision tree dan kemungkinan bias yang tinggi, random forest mengambil prediksi hasil dari beberapa decision tree berdasarkan suara mayoritas dan memprediksi hasil yang lebih akurat.

Semakin banyak hasil decision tree yang diambil, semakin tinggi akurasi terutama ketika masing-masing pohon tidak berkorelasi satu sama lain.

Namun bukan berarti hasil dari random forest selalu tepat. Karena itu ada dua asumsi untuk mendapatkan hasil yang lebih baik:

Harus ada beberapa nilai aktual dalam variabel fitur dari kumpulan data, sehingga pengklasifikasi dapat memprediksi hasil akurat.

Prediksi dari setiap pohon harus memiliki korelasi yang sangat rendah.