

Nama : Muhammad Alfi Syahrin
NIM : 1103213144
Kelas : TK-45-05

Random Forests

Decision tree bagus untuk data yang digunakan untuknya belajar, namun tidak fleksibel ketika melakukan klasifikasi terhadap sampel terbaru. Karena hal tersebut, dibuatlah random forest yang menggabungkan kesederhanaan decision tree dengan fleksibilitas yang menghasilkan peningkatan besar dalam akurasi. Cara membuat random forest, yang pertama adalah 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. Kedua, 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. Cara mengetahui akurasi dari random forest, dapat dilakukan pengetesan menggunakan data yang tidak masuk ke dalam bootstrapped dataset (disebut "Out-of-Bag dataset"). Akurasi dapat dilihat dari banyaknya klasifikasi yang benar berdasarkan dataset asli. Optimasi random forest dapat dilakukan dengan cara mengubah banyaknya pengambilan variabel acak dalam tiap step dan digunakan yang memiliki akurat paling tinggi.