

Nama : Angelica Sharon Amelia Simanjuntak

NIM : 1103210032

RESUME

Random Forests

Proses pembuatan algoritma random forest adalah dengan membuat suatu dataset bootstrap dan subset variabel untuk decision tree. Decision Tree termasuk mudah dibuat, digunakan, serta diinterpretasikan, namun kurang fleksibel, sehingga akurasi menurun saat mengkategorikan sampel baru. Dengan bootstrapping akan memungkinkan pembentukan set data yang acak dengan penggantian serta setiap decision tree dibangun dari satu set data acak.

Langkah-langkah:

- **Step 1**
membuat kumpulan data pada bootstrap yang ukurannya sama dengan aslinya, kemudian memilih sampel secara acak dari kumpulan data asli.
- **Step 2**
membuat Decision Tree menggunakan dataset bootstrap namun hanya menggunakan sub kumpulan variabel atau kolom acak pada setiap langkah, contohnya hanya akan mempertimbangkan dua variabel atau kolom pada setiap langkah.
- **Step 3**
mengulangi step 1 dan 2 beberapa kali, membuat data bootstrap baru. Menggunakan sampel bootstrap dan hanya akan mempertimbangkan sebagian variabel pada setiap langkah dan akan menghasilkan beragam pohon, inilah yang menjadikan random forest menjadi lebih efektif dibandingkan decision tree individual.

Cara mengevaluasi random forest yaitu dengan menggunakan 2 variabel yang dimana kolom data untuk membuat keputusan pada tiap langkahnya. Lalu membandingkan error out-of-bag yang akan dibangun dengan menggunakan 2 variabel per langkahnya. Pertama dengan membangun random forest lalu yang kedua memperkirakan keakuratan random forest dan mengubah jumlah variabel yang akan digunakan pada tiap langkahnya kemudian melakukannya beberapa kali dan memilih yang paling akurat.