Nama : Mochamad Phillia Wibowo

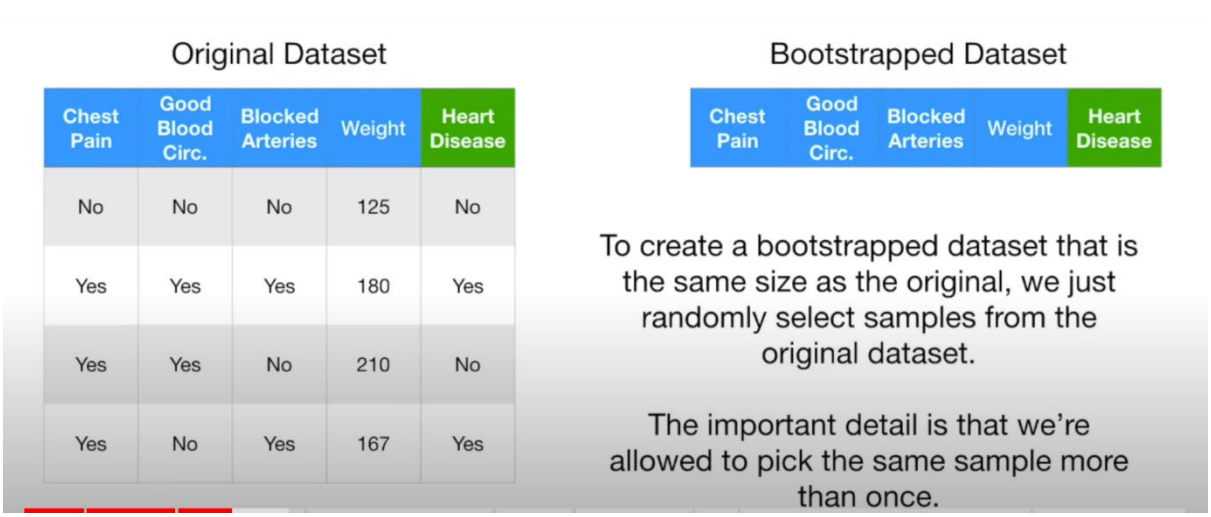
NIM : 1103204191

Kelas : TK-44-G04

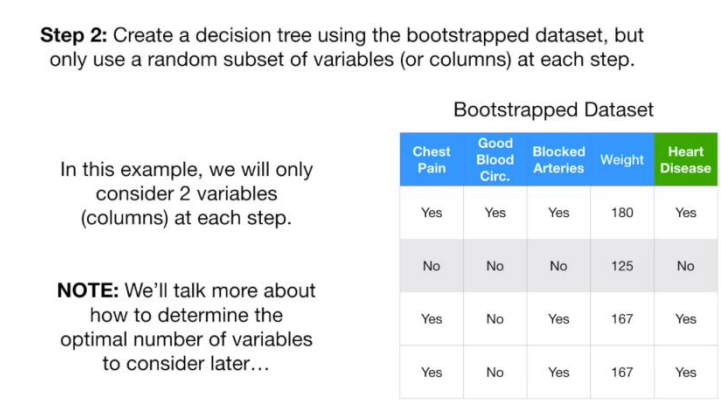
Lecture 1 : Rangkuman *Random Forests From StatQuest*

Video ini menjelaskan tentang konsep dan evaluasi dari metode pembelajaran mesin yang disebut sebagai "*random forests*" atau "hutan acak". Konsep dasar dari *random forests* adalah kombinasi antara kesederhanaan dari pohon keputusan (*decision trees*) dengan fleksibilitas, yang secara signifikan meningkatkan akurasinya.

Decision trees adalah model yang mudah dibangun, digunakan, dan diinterpretasikan. Sedangkan *random forests* membutuhkan pembuatan kumpulan data *bootstrap* dengan ukuran yang sama dengan data asli, di mana langkah berikutnya adalah membuat pohon keputusan menggunakan kumpulan data *bootstrap* tersebut.

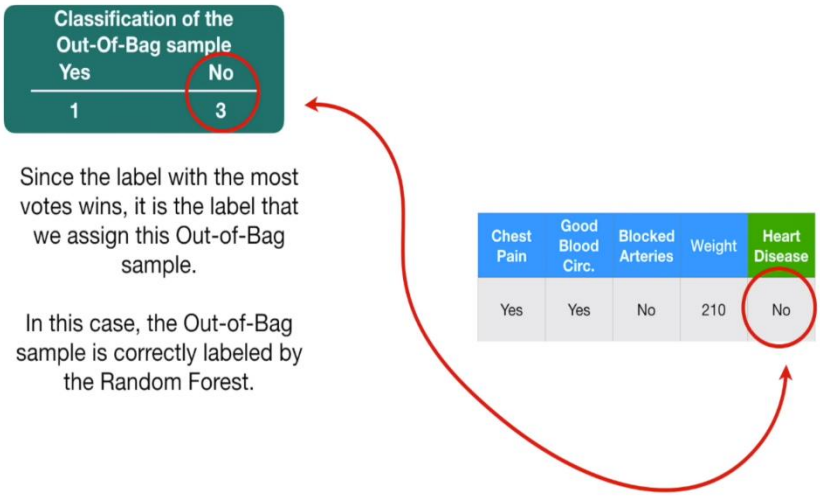


Data *bootstrap* adalah teknik pengambilan sampel dengan penggantian yang digunakan untuk menghasilkan banyak sampel baru dari satu set data yang ada. Dalam konteks pembelajaran mesin, bootstrap sering digunakan untuk membuat kumpulan data baru dengan ukuran yang sama dengan data asli. Proses ini dilakukan dengan memilih acak sampel dengan penggantian dari data asli sebanyak ukuran yang diinginkan.



Random forests menggunakan subset acak dari variabel pada setiap langkah untuk membangun pohon. Hanya dua variabel yang dipertimbangkan pada setiap langkah, dan pemilihan variabel yang akan dipertimbangkan dilakukan secara acak. Dengan menggunakan sampel *bootstrap* dan pertimbangan subset variabel pada setiap langkah, dapat diciptakan berbagai macam pohon.

Dalam evaluasi performa *random forests*, data yang tidak termasuk dalam kumpulan data bootstrap disebut sebagai "*out-of-bag*" (OOB) data. Setiap pohon individual dalam *random forests* dievaluasi menggunakan OOB data, di mana pohon dapat benar atau salah dalam menentukan label pada sampel OOB.



Akurasi dari *random forests* dapat diukur dengan proporsi sampel OOB yang diklasifikasikan dengan benar. Proporsi sampel OOB yang diklasifikasikan dengan benar adalah akurasi dari *random forests*, sedangkan proporsi sampel OOB yang diklasifikasikan dengan salah adalah kesalahan OOB. Untuk memilih konfigurasi *random forests* yang paling akurat, dilakukan pengujian dengan menggunakan berbagai konfigurasi dan memilih *random forests* yang memberikan akurasi tertinggi.