

Nama: Danny Hamtar Pangestu
Nim: 1103210037

Random Forest - STATQUEST

The screenshot shows a YouTube video player with the title 'statquest random forest'. The video content displays a decision tree on the left and a data table on the right. The decision tree has a root node (blue box) with two branches leading to child nodes (blue boxes), which further branch into leaf nodes (green boxes). A red arrow points from the root node to the data table. The data table has columns: Chest Pain, Good Blood Circ., Blocked Arteries, Weight, and Heart Disease. The first row of data is circled in red: Chest Pain: Yes, Good Blood Circ.: No, Blocked Arteries: No, Weight: 168, Heart Disease: (blank). Below the table is a green box labeled 'Heart Disease' with a table showing 'Yes' with a count of 1 and 'No' with a count of 0. Text below the tree says 'The second tree also says "Yes"...'. The video player interface at the bottom shows a progress bar at 5:15 / 9:53 and the title 'Classifying a new sample with a Random Forest'.

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	No	No	168	

Heart Disease	
Yes	No
1	0

Dalam video "StatQuest" tentang Random Forest oleh Josh Darmer

Random Forest dibangun dari pohon keputusan, yang dikenal karena kesederhanaannya tetapi kurang fleksibilitasnya dalam mengklasifikasikan sampel baru. Darmer menjelaskan bahwa hutan acak mengatasi masalah ini dengan menggabungkan sederhanaanya pohon keputusan dengan fleksibilitas, sehingga meningkatkan akurasi prediksi. Proses pembangunan hutan acak dimulai dengan membuat kumpulan data bootstrap dari sampel acak dari data asli.

Kemudian, pohon keputusan dibangun menggunakan dataset bootstrap, tetapi hanya menggunakan subset variabel acak pada setiap langkah. Langkah selanjutnya adalah menjalankan data baru melalui semua pohon dalam hutan acak dan memilih label yang paling banyak muncul sebagai prediksi akhir. Untuk mengevaluasi kinerja hutan acak, digunakan metode kumpulan data out-of-bag, di mana sampel yang tidak dimasukkan dalam dataset bootstrap digunakan untuk mengukur akurasi model. Terakhir, untuk meningkatkan keakuratan hutan acak, variasi jumlah variabel yang digunakan per langkah dapat dieksplorasi, dan pengaturan yang paling akurat dipilih. Dengan cara ini, video ini memberikan pemahaman yang jelas tentang konsep dan penggunaan hutan acak dalam analisis data dan pembelajaran mesin.