

Nama : Afif Ibadurrahman Jalaluddin

Nim : 1103210221

Kelas : TK4503

Random Forest

Video tersebut membahas tentang *Random Forest*, dan ini adalah bagian pertama dari serangkaian video mengenai topik tersebut. Pembuat video, Josh Dharma, menjelaskan langkah-langkah dalam membangun dan mengevaluasi *Random Forest*. *Random Forest* sendiri dibangun dari *Decision Trees*, yang sebelumnya telah dijelaskan dalam video *StatQuest* sebelumnya.

Langkah pertama dalam membuat *Random Forest* adalah menciptakan *bootstrap dataset*. *Bootstrap* dataset dibuat dengan memilih sampel secara acak dari data asli, dengan memperbolehkan pemilihan sampel yang sama lebih dari satu kali. Setelah itu, dibangun *Decision Tree* menggunakan *bootstrap dataset* tersebut, tetapi hanya menggunakan *subset* acak dari variabel atau kolom pada setiap langkahnya.

Proses ini diulang beberapa kali, menciptakan berbagai *Decision Trees* dengan variasi. Keanekaragaman ini membuat *Random Forest* lebih efektif daripada *Decision Trees* individu, meningkatkan akurasi prediksi.

Penggunaan *Random Forest* dilakukan dengan menjalankan data baru melalui setiap *Decision Tree* yang telah dibangun. Setiap pohon memberikan *vote* untuk kategori yang sesuai, dan hasil akhir diambil dari mayoritas suara.

Untuk mengevaluasi keberhasilan *Random Forest*, digunakan *out-of-bag dataset*. Ini adalah data yang tidak termasuk dalam *bootstrap dataset* karena diizinkan adanya duplikasi. *Out-of-bag dataset* digunakan untuk mengukur akurasi model, dan *out-of-bag error* dihitung sebagai proporsi sampel yang salah diklasifikasikan oleh *Random Forest*.

Josh juga menjelaskan bahwa penentuan jumlah variabel yang digunakan pada setiap *step* dalam pembuatan *Decision Tree* dapat dioptimalkan dengan mencoba beberapa pengaturan berbeda. Pengaturan yang memberikan *Random Forest* paling akurat dipilih setelah pengujian.

Video ini memberikan gambaran tentang bagaimana membangun, menggunakan, dan mengevaluasi *Random Forest*. Bagian selanjutnya dari seri ini akan membahas cara mengatasi data yang hilang dan cara mengelompokkan sampel.