

# **Rangkuman Video Materi Random Forest**

Axel David, 1103210017, TK4504

## **A. Pengenalan Random Forest**

Random Forest merupakan sebuah algoritma yang dikembangkan sebagai peningkatan dari pohon keputusan (decision trees) dengan menggabungkan sederhana dengan fleksibilitas, yang menghasilkan peningkatan akurasi dalam pembelajaran prediktif. Algoritma ini diperkenalkan sebagai solusi untuk meningkatkan performa pohon keputusan yang cenderung overfitting terhadap data latih.

Dalam konteks Random Forest, sejumlah besar pohon keputusan dibangun dan dikombinasikan untuk meningkatkan akurasi prediksi. Setiap pohon keputusan dibangun dengan mempertimbangkan subset acak dari fitur-fitur yang tersedia dalam dataset. Ketika datang saatnya untuk melakukan prediksi, hasil dari setiap pohon digabungkan untuk memutuskan kelas mayoritas. Hal ini membuat Random Forest menjadi algoritma yang kuat dalam klasifikasi data dan prediksi, terutama saat berurusan dengan dataset yang kompleks dan tidak terstruktur. Dengan pendekatan yang inovatif ini, Random Forest telah menjadi salah satu algoritma yang paling populer dan efektif dalam analisis data dan pembelajaran mesin.

## **B. Tujuan Penggunaan Random Forest**

Motivasi utama untuk menggunakan Random Forest adalah meningkatkan performa dan stabilitas prediksi dibandingkan dengan pohon keputusan tunggal. Meskipun pohon keputusan memiliki kemudahan interpretasi dan dapat memberikan hasil yang baik pada data latih, namun kecenderungan untuk overfitting dapat mengurangi kemampuan generalisasi model pada data uji.

Dengan menggunakan Random Forest, banyak pohon keputusan dibangun secara independen dan hasil dari masing-masing pohon dikombinasikan untuk membuat prediksi akhir. Hal ini mengurangi risiko overfitting karena agregasi hasil dari banyak pohon keputusan yang berbeda. Selain itu, Random Forest juga mampu menangani dataset yang besar dan kompleks dengan baik, serta toleran terhadap data yang berisik. Dengan kemampuannya untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting, Random Forest telah menjadi salah satu pilihan utama dalam analisis data dan pembelajaran mesin, terutama dalam kasus-kasus di mana keakuratan prediksi sangat penting.

## **C. Membuat Bootstrapped Dataset**

Langkah pertama dalam membangun Random Forest adalah menciptakan dataset bootstrapped. Ini dilakukan dengan mengambil sampel secara acak dari dataset asli dengan penggantian, yang berarti beberapa sampel dapat dipilih lebih dari sekali, sementara yang lain mungkin tidak dipilih sama sekali. Pengambilan sampel bootstrap ini memungkinkan pembentukan banyak dataset yang berbeda untuk digunakan dalam

pembuatan pohon keputusan di dalam Random Forest. Dengan menggunakan teknik ini, setiap pohon keputusan yang dibuat akan memiliki dataset yang unik, yang membantu meningkatkan variasi dan keakuratan model keseluruhan. Selain itu, sampel bootstrap juga membantu mengurangi varians dan overfitting dalam model, karena setiap pohon dibangun dengan menggunakan subset acak dari data, yang mewakili variasi yang lebih luas dalam data. Dengan demikian, langkah pertama ini menjadi fondasi penting dalam pembangunan Random Forest yang efektif.

#### **D. Membuat Decision Tree Dengan Subset Variabel Acak Pada Setiap Langkah**

Langkah kedua dalam pembangunan Random Forest adalah membuat pohon keputusan dengan menggunakan subset acak dari variabel-variabel yang tersedia pada setiap langkah. Dalam setiap simpul atau percabangan dalam pohon, hanya sebagian kecil dari variabel yang dipertimbangkan untuk membuat keputusan. Penggunaan subset acak ini membantu mencegah ketergantungan antar pohon keputusan yang dibangun, sehingga mendorong variasi dan keragaman antara pohon-pohon tersebut. Dengan cara ini, setiap pohon keputusan cenderung lebih independen dan tidak terlalu dipengaruhi oleh satu sama lain, yang menghasilkan hasil yang lebih stabil dan akurat saat digabungkan dalam Random Forest.

Selanjutnya, dalam pembangunan Random Forest adalah mengulangi langkah-langkah pembuatan pohon keputusan dan penciptaan dataset bootstrapped untuk sejumlah kali tertentu. Dengan melakukan iterasi ini, sejumlah besar pohon keputusan dapat dibangun, masing-masing dengan dataset dan subset variabel yang berbeda. Proses ini menghasilkan ensemble pohon keputusan yang kuat dan variatif, yang meningkatkan kemampuan keseluruhan Random Forest dalam melakukan klasifikasi dan prediksi. Semakin banyak pohon keputusan yang dibangun, semakin baik juga kemampuan generalisasi dan akurasi prediksi yang dihasilkan oleh Random Forest. Oleh karena itu, langkah ini merupakan kunci dalam memastikan performa optimal dari algoritma Random Forest.

#### **E. Mengklasifikasikan Sampel Baru Dengan Random Forest**

Random Forest mengklasifikasikan sampel baru dengan menjalankan data melalui setiap pohon keputusan dan mengagregasi hasil untuk menentukan mayoritas suara. Dengan menggunakan ensemble hasil dari berbagai pohon keputusan, Random Forest dapat memberikan prediksi yang lebih stabil dan akurat daripada pohon keputusan tunggal. Proses agregasi ini juga membantu mengurangi varians dalam prediksi dan meningkatkan keandalan hasil.

#### **F. Definisi Bagging**

Bagging merupakan singkatan dari Bootstrap Aggregating, merupakan teknik pengambilan sampel yang digunakan dalam pembangunan Random Forest. Teknik ini melibatkan pengambilan sampel secara acak dari dataset asli dengan penggantian, yang menghasilkan banyak dataset yang berbeda untuk digunakan dalam pembuatan pohon

keputusan di dalam Random Forest. Dengan menggunakan bagging, variasi dan overfitting dalam model dapat dikurangi, sehingga meningkatkan akurasi dan stabilitas prediksi.

#### **G. Evaluasi Random Forest**

Evaluasi Random Forest dilakukan untuk mengukur kinerja dan akurasi model. Salah satu metode evaluasi yang umum digunakan adalah menggunakan sampel out-of-bag, yaitu sampel-sampel yang tidak termasuk dalam dataset bootstrapped yang digunakan untuk melatih pohon keputusan tertentu. Dengan menggunakan sampel out-of-bag, akurasi model dapat diestimasi dengan lebih akurat tanpa memerlukan data validasi tambahan.

#### **H. Optimasi Random Forest**

Optimasi Random Forest dapat dilakukan dengan menyesuaikan beberapa parameter kunci, seperti jumlah pohon keputusan, jumlah variabel yang digunakan per langkah, dan kedalaman maksimum pohon. Penyesuaian parameter ini dapat dilakukan dengan menggunakan teknik seperti cross-validation untuk menemukan kombinasi parameter yang memberikan hasil terbaik dalam hal akurasi dan stabilitas prediksi. Dengan mengoptimalkan Random Forest, performa dan efisiensi model dapat ditingkatkan, sehingga memastikan hasil yang lebih baik dalam klasifikasi dan prediksi data baru.