

## **Rangkuman Video Materi Decisions Tree**

Axel David, 1103210017, TK4504

### **A. Pengenalan Decisions Tree**

Pohon Keputusan atau Decisions Tree merupakan salah satu algoritma yang sering digunakan dalam pembelajaran mesin untuk tujuan klasifikasi. Algoritma ini bekerja dengan membagi data menjadi kelompok-kelompok yang semakin kecil berdasarkan pada serangkaian keputusan yang diambil dari fitur-fitur data. Pada setiap simpul atau percabangan dalam pohon keputusan, suatu keputusan diambil berdasarkan nilai fitur-fitur yang ada. Proses ini terus berlanjut hingga mencapai simpul-simpul akhir yang disebut sebagai daun, di mana hasil klasifikasi diperoleh.

Salah satu keunggulan dari pohon keputusan adalah kemampuannya dalam menangani data kategoris dan numerik, serta kemudahan dalam interpretasi. Dengan adanya aturan-aturan keputusan yang terbentuk dalam pohon, interpretasi model menjadi lebih mudah dimengerti oleh pengguna. Selain itu, pohon keputusan juga mampu menangani interaksi antar fitur secara alami.

### **B. Dasar Konsep Decision Tree**

Dalam pembangunan pohon keputusan, langkah penting yang perlu diperhatikan adalah pemilihan fitur-fitur yang paling informatif untuk melakukan pemisahan data yang optimal. Untuk mengukur kualitas pemisahan tersebut, digunakan ukuran ketidakmurnian seperti Gini impurity atau entropi. Melalui proses ini, pohon keputusan dapat memilih fitur-fitur yang paling relevan untuk membuat keputusan yang akurat.

### **C. Membuat Tree dengan Gini Impurity**

Ukuran ketidakmurnian seperti Gini impurity menjadi alat penting dalam proses pembangunan pohon keputusan. Dengan menggunakan Gini impurity, kualitas pemisahan setiap simpul atau percabangan dalam pohon dapat dievaluasi dengan lebih baik. Hal ini membantu dalam menentukan fitur-fitur yang paling relevan untuk digunakan dalam memisahkan data.

### **D. Cabang Numerik dan Kontinu**

Dalam membangun pohon keputusan, penting untuk memperhatikan pengelolaan fitur-fitur numerik dan kontinu. Fitur-fitur ini memerlukan pendekatan yang berbeda dalam menentukan keputusan yang optimal. Misalnya, penggunaan ambang numerik diperlukan untuk menangani fitur-fitur yang memiliki nilai kontinu, sedangkan fitur kategoris akan memerlukan keputusan berdasarkan pada kategori-kategori yang tersedia.

### **E. Menambahkan Simpul (Leaves)**

Simpul-simpul daun merupakan simpul terakhir dalam pohon keputusan di mana nilai output atau prediksi kelas didefinisikan. Proses ini memungkinkan pohon keputusan untuk memberikan hasil klasifikasi atau prediksi yang akurat berdasarkan pada fitur-fitur data yang ada.

### **F. Mendefinisikan Output**

Simpul-simpul daun dalam pohon keputusan memegang peran penting dalam menentukan nilai output atau prediksi kelas. Dalam proses ini, definisi nilai output pada simpul-simpul daun menjadi kunci dalam memperoleh hasil klasifikasi yang akurat.

### **G. Menggunakan Decisions Tree**

Pohon keputusan digunakan untuk melakukan klasifikasi data baru dengan mengambil keputusan berdasarkan pada aturan-aturan yang terbentuk dalam struktur pohon. Ini memungkinkan pengguna untuk memprediksi kelas atau nilai output dari data yang tidak dikenal dengan menggunakan model yang sudah dibangun.

### **H. Pencegahan Overfitting**

Overfitting merupakan salah satu masalah umum dalam pembangunan pohon keputusan. Untuk mencegah overfitting, teknik-teknik seperti pemangkasan (pruning) dan pengaturan batas ukuran daun dapat diterapkan. Pemangkasan memungkinkan pohon keputusan untuk menjadi lebih sederhana dan mencegahnya dari menjadi terlalu sesuai dengan data latih.

### **I. Cross-validation**

Evaluasi kinerja model juga merupakan langkah penting dalam pembangunan pohon keputusan. Salah satu metode evaluasi yang umum digunakan adalah cross-validation, di mana data dibagi menjadi beberapa subset untuk melatih dan menguji model secara berulang. Dengan menggunakan cross-validation, dapat dipastikan bahwa model yang dibangun mampu melakukan generalisasi dengan baik pada data yang tidak terlihat sebelumnya.

### **J. StatQuest**

Dalam rangka mendukung pemahaman konsep pohon keputusan dan pembelajaran mesin secara umum, tersedia berbagai sumber daya tambahan seperti tutorial, buku, dan kursus online. StatQuest, misalnya, merupakan salah satu sumber daya yang populer dengan menyediakan video-video pendidikan yang mudah dipahami dan materi-materi belajar yang mendalam. Dengan memanfaatkan sumber daya ini, diharapkan pemahaman terhadap konsep-konsep tersebut dapat ditingkatkan dan kemampuan dalam mengimplementasikannya dalam praktik pembelajaran mesin dapat diperkuat.