

Nama : Alif Firmansyah
NIM : 1103204105
Kelas : TK44G4

Rangkuman Decision Tree StatQuest

Dalam video ini, StatQuest menjelaskan tentang Decision Tree dan klasifikasi, yang membuat pernyataan dan kemudian mengambil keputusan berdasarkan apakah pernyataan tersebut true atau false. Ketika mengklasifikasikan sesuatu ke dalam kategori, ini disebut pohon klasifikasi, dan ketika memprediksi nilai numerik, ini disebut pohon regresi. Classification Tree mudah digunakan, dan prosesnya dimulai dari atas lalu turun ke bawah hingga mencapai titik klasifikasi. Terminologi yang digunakan untuk pohon klasifikasi terdiri dari simpul akar, simpul atau cabang internal, dan simpul daun. Untuk membangun Classification Tree dari data mentah, pertanyaan di bagian atas pohon dipilih dengan melihat seberapa baik suatu variabel memprediksi apakah seseorang menyukai es dingin, misalnya.

Selanjutnya, pembicara menjelaskan cara menghitung 'gini impurity' pada daun Decision Tree untuk mengukur impurity pada prediksi. Dengan menggunakan contoh prediksi siapa yang menyukai film "Cool as Ice" berdasarkan apakah mereka menyukai soda atau popcorn, pembicara menghitung gini impurity untuk setiap daun yang berisi campuran orang yang menyukai dan tidak menyukai film tersebut. Bobot setiap daun dihitung berdasarkan jumlah total orang yang diwakilinya, dan rata-rata tertimbang diambil untuk menentukan total gini impurity untuk bagian Decision Tree tersebut. Pembicara juga menyebutkan metode lain untuk mengukur impurity, seperti entropi dan perolehan informasi, namun berfokus pada gini impurity karena mudah dan populer.

Lalu video menjelaskan cara menghitung gini impurity dengan mengurutkan baris berdasarkan umur dari nilai terendah hingga nilai tertinggi kemudian menghitung rata-rata umur semua orang yang berdekatan dan nilai gini impurity untuk setiap rata-rata umur. Lebih lanjut, mereka menjelaskan bagaimana nilai gini impurity dihitung untuk setiap fitur untuk menentukan fitur mana yang harus berada di puncak pohon. Berdasarkan nilai gini impurity, fitur disusun, dan pohon dibangun untuk menetapkan nilai keluaran untuk setiap daun berdasarkan nilai mayoritas untuk menentukan apakah orang tersebut menyukai film 'Cool as Ice' atau tidak.

Konsep overfitting pada pohon keputusan dibahas dan dua metode untuk mengatasinya diperkenalkan. Salah satu metodenya disebut pemangkasan, yang melibatkan pemindahan beberapa cabang pada pohon. Metode lainnya melibatkan pembatasan pertumbuhan pohon, misalnya dengan menetapkan jumlah minimum orang untuk setiap daun. Validasi silang juga disebutkan sebagai cara untuk menguji efektivitas berbagai metode dalam menangani overfitting.