

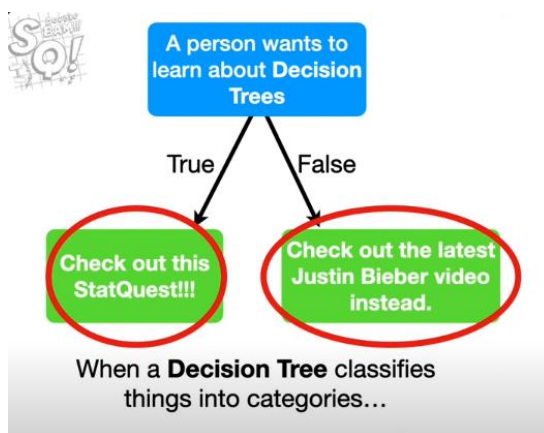
Nama : Mochamad Phillia Wibowo

NIM : 1103204191

Kelas : TK-44-01

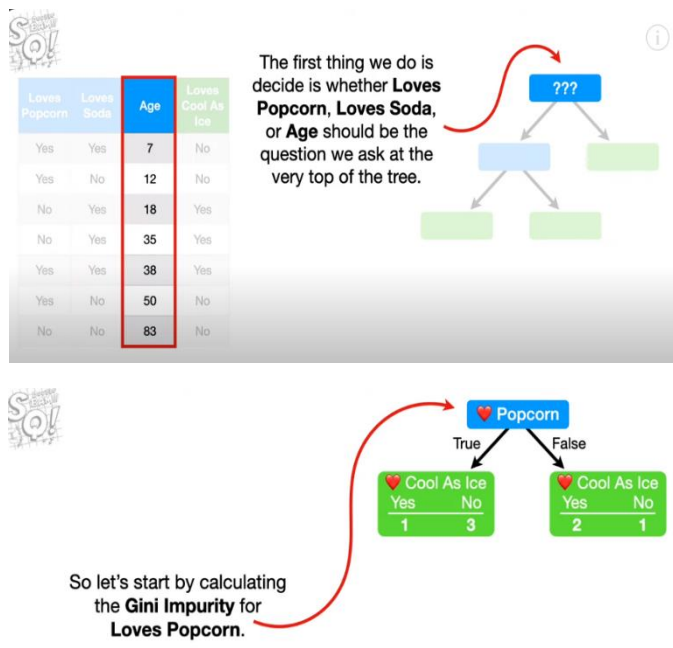
## Lecture 1 : Rangkuman *Decision Tree From StatQuest*

- **Decision Tree dan Klasifikasi**



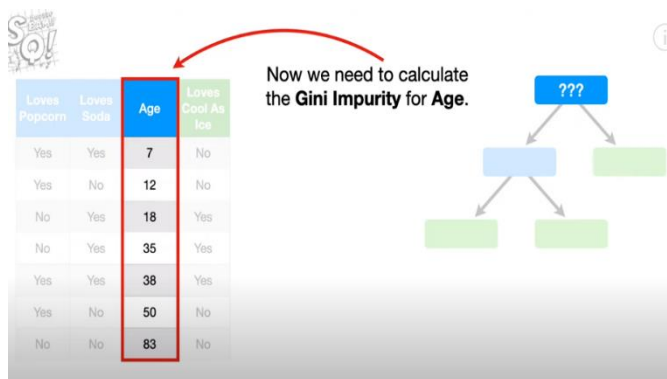
*Decision Tree* adalah algoritma yang memetakan keputusan berdasarkan serangkaian pernyataan, yang dapat digunakan untuk klasifikasi data atau memprediksi nilai. Proses pembangunan pohon dimulai dengan data mentah dan langkah demi langkah memilih pertanyaan terbaik untuk membagi data menjadi kelompok yang lebih *homogen*.

- **Pengukuran dengan Ketidakmurnian Menggunakan Impuritas Gini**



① Dalam kasus memprediksi preferensi *popcorn* dan film "*Cool as Ice*", impuritas Gini digunakan untuk mengukur ketidakmurnian pada setiap langkah pohon keputusan. Misalnya, ditemukan bahwa kecenderungan terhadap minuman soda lebih baik dalam memprediksi preferensi film daripada faktor lainnya.

- Perhitungan Impuritas Gini Untuk Fitur dan Usia



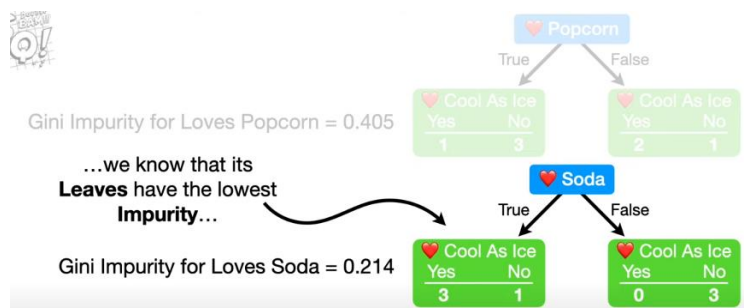
Now we need to calculate the Gini Impurity for Age.

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

Pengukuran impuritas Gini dilakukan untuk mengevaluasi tingkat ketidakmurnian pada setiap langkah pohon. Selain itu, impuritas Gini juga dihitung untuk fitur tertentu, seperti usia, untuk membantu dalam

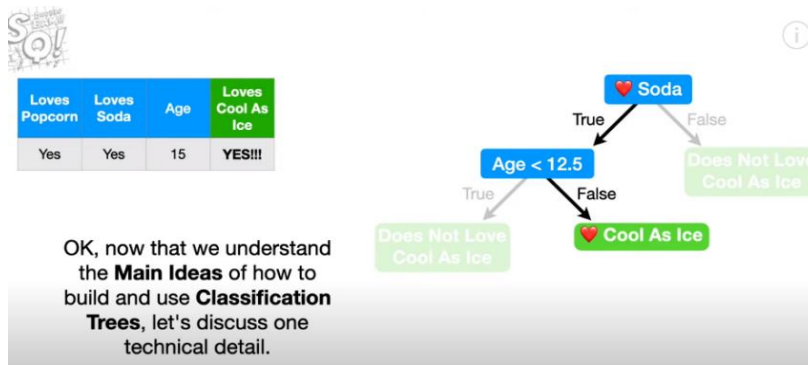
memutuskan langkah terbaik dalam pembuatan keputusan.

- Pembangunan dan Penggunaan Pohon Klasifikasi



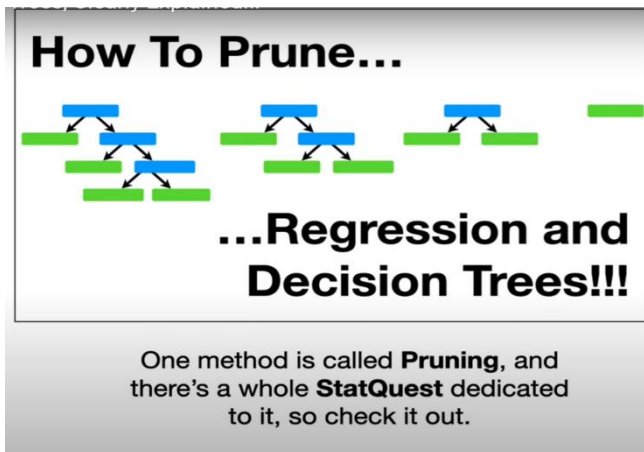
Proses pembangunan pohon klasifikasi berfokus pada mengurangi ketidakmurnian pada setiap langkahnya. Hal ini dilakukan dengan memilih

langkah yang menghasilkan impuritas Gini terendah, seperti memisahkan data berdasarkan preferensi minuman soda, *popcorn*, atau usia. Pentingnya mencegah



*overfitting* juga ditekankan untuk memastikan prediksi yang akurat.

- Pemangkasan dan Pembatasan Pohon Dalam *Machine Learning*



Untuk mengatasi *overfitting*, teknik pemangkasan dan pembatasan pertumbuhan pohon digunakan. Ini termasuk menghapus cabang yang tidak signifikan dari pohon atau menetapkan batas minimum untuk jumlah data dalam setiap daun. Langkah-langkah ini membantu

menghasilkan model yang lebih umum dan dapat diterapkan dengan baik pada data baru dalam praktik *machine learning*.