



**Indonesia AI**  
AI for Everyone, AI for Indonesia

**# 2023JAGO**  
**TEKNOLOGIAI**

**AI Career Bootcamp**

# **Random Forest & Ensemble Learning**

## Pembukaan

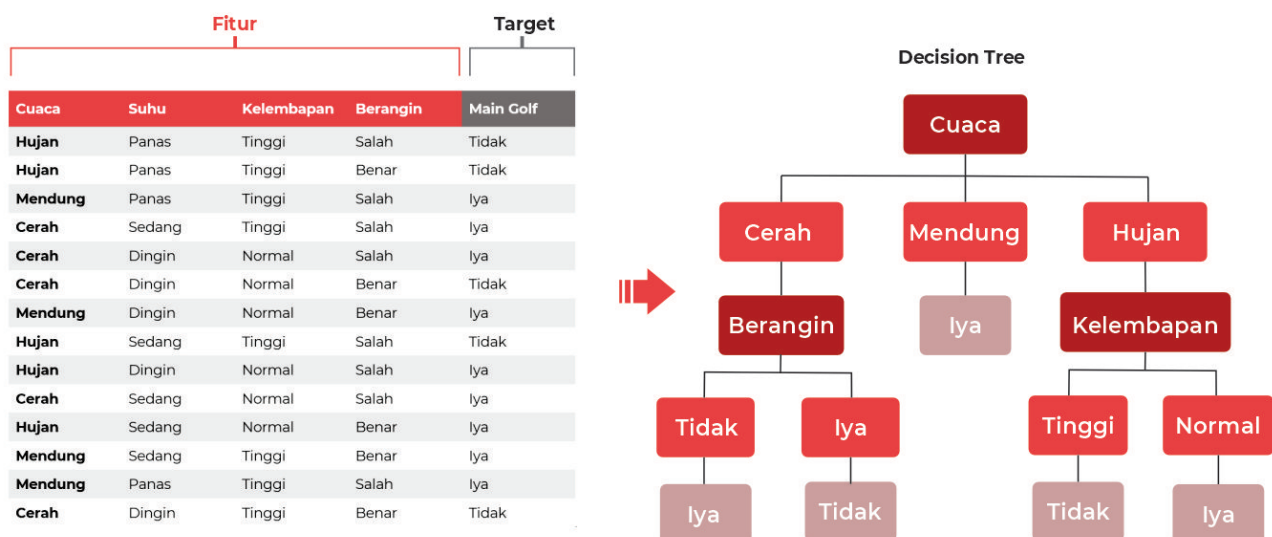
Guide Book ini memuat beberapa informasi-informasi utama yang akan disampaikan oleh mentor di program AI Career Bootcamp Kelas Computer Vision yang bisa dijadikan pegangan para students untuk mempersiapkan diri sebelum sesi Live Class berlangsung.

## Deskripsi

Random Forest adalah salah satu jenis algoritma Machine Learning yang digunakan untuk memprediksi output berdasarkan input. Algoritma ini merupakan bagian dari kelas algoritma Ensemble Learning (tipe Bagging), yang merupakan teknik yang memanfaatkan banyak model Machine Learning untuk meningkatkan akurasi prediksi. Dalam algoritma Random Forest, digunakan beberapa pohon keputusan (decision tree) yang dihasilkan secara acak untuk membuat prediksi.

Setiap pohon keputusan pada Random Forest dibangun dengan menggunakan subset acak dari data training dan subset acak dari fitur yang tersedia. Setiap pohon diukur dengan nilai kebaikan (goodness) yang digunakan untuk memilih fitur terbaik untuk membagi data. Kriteria pemilihan fitur ini bisa berbeda-beda tergantung pada masalah yang ingin dipecahkan. Misalnya, dalam masalah klasifikasi biner, Gini index dapat digunakan sebagai kriteria pemilihan fitur, sementara pada masalah regresi, Mean Square Error (MSE) lebih tepat digunakan.

Ilustrasi Decision Tree:



Setelah sejumlah pohon keputusan terbentuk, hasil prediksi dari setiap pohon digabungkan untuk menghasilkan output akhir. Proses ini dilakukan dengan mekanisme Max Voting atau Average Voting dari setiap pohon tersebut. Dengan teknik Ensemble Learning ini, algoritma Random Forest dapat menghasilkan prediksi yang lebih akurat dan stabil dibandingkan dengan menggunakan satu pohon keputusan saja.

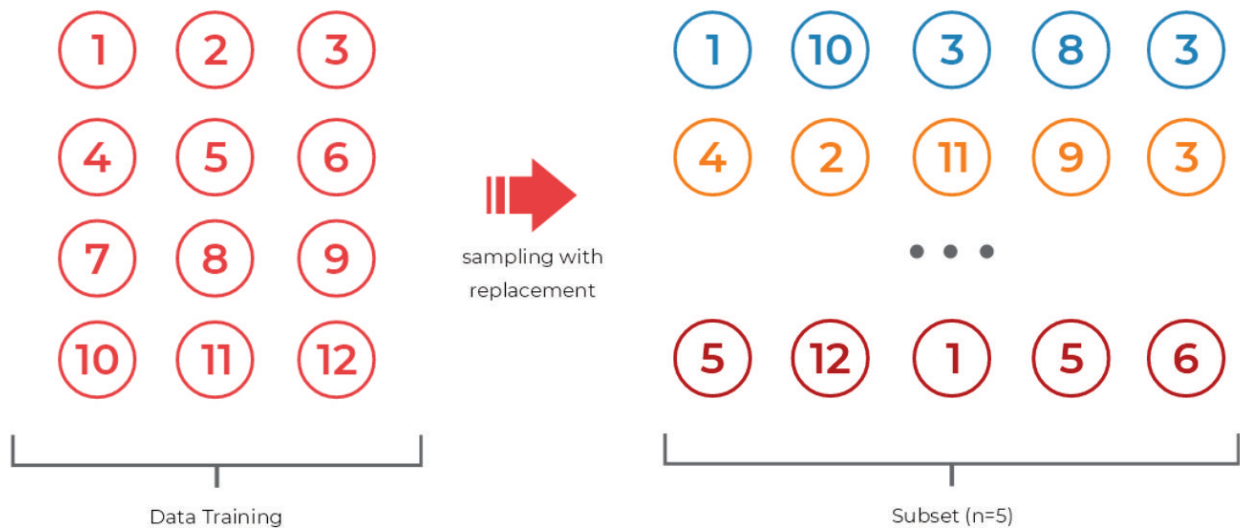
## Ensemble Learning

### ***“Unity is Strength!”***

Ensemble Learning adalah teknik dalam Machine Learning yang menggabungkan beberapa model atau algoritma untuk meningkatkan kinerja dan akurasi prediksi. Dalam Ensemble Learning, beberapa model dihasilkan dengan menggunakan data yang sama atau data yang berbeda untuk menghasilkan prediksi yang lebih akurat. Model-model tersebut kemudian digabungkan untuk menghasilkan prediksi akhir. Algoritma Random Forest mengadopsi teknik ini dengan cara menggabungkan banyak algoritma Decision Tree di dalamnya. Namun, pada praktiknya kita juga bisa menggunakan dan menggabungkan algoritma lainnya Logistic Regression dan SVM.

Terdapat beberapa jenis Ensemble Learning, di antaranya adalah Bagging, Boosting, dan Stacking.

- 1 Bagging (bootstrap aggregating) adalah teknik yang membagi data training menjadi beberapa subset yang berbeda dan membangun model pada setiap subset tersebut. Setiap model kemudian melakukan prediksi pada data testing dan hasil prediksi dari masing-masing
- 2 Boosting adalah teknik yang membangun model secara berurutan dengan memperhatikan kesalahan prediksi model sebelumnya. Model yang lebih akurat diberi bobot yang lebih besar dalam penggabungan hasil prediksi
- 3 Stacking, yaitu menggabungkan hasil prediksi beberapa model dengan model yang lebih kompleks untuk menghasilkan prediksi akhir yang lebih akurat.



Ensemble Learning merupakan teknik yang sangat berguna dalam aplikasi Machine Learning karena dapat meningkatkan kinerja dan akurasi prediksi. Dalam Ensemble Learning, beberapa model atau algoritma dihasilkan dengan menggunakan data yang sama atau data yang berbeda,

## Kelebihan & Kekurangan

Random Forest adalah salah satu algoritma machine learning yang populer dan banyak digunakan karena memiliki sejumlah kelebihan, di antaranya:

- Akurasi yang tinggi
- Mampu menangani data yang besar
- Tidak sensitive terhadap outlier dan noise

Namun, Random Forest juga memiliki beberapa kekurangan, di antaranya:

- Proses training yang membutuhkan waktu relatif lebih lama
- Membutuhkan waktu tuning parameter