

# **Machine Learning com Python**

Prof. Luciano Galdino

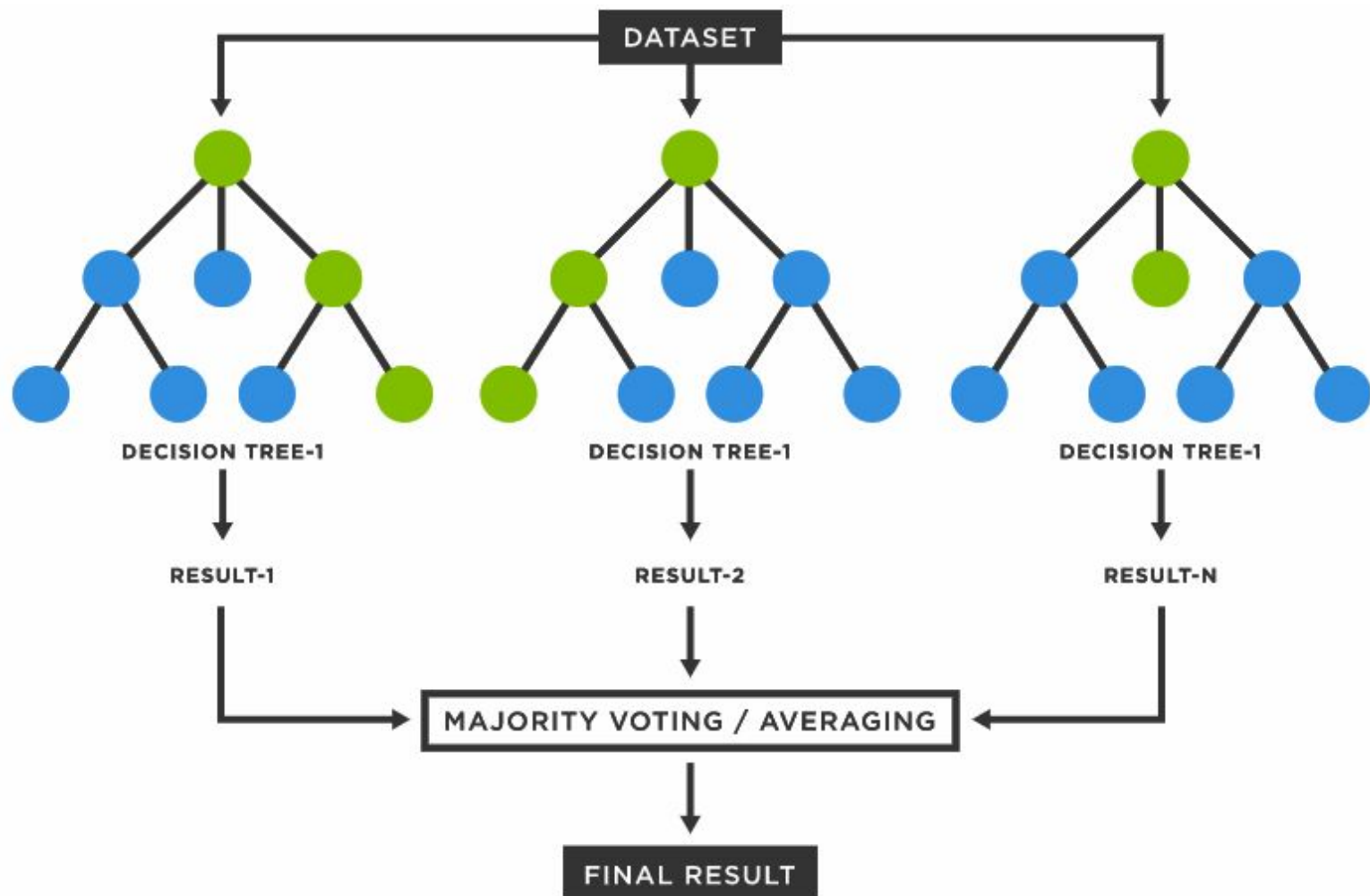
## **RANDOM FOREST (Floresta Aleatória)**

Criação aleatória de várias árvores de decisão.

Utiliza o método Ensemble (construção de vários modelos para obter um resultado único).

É mais robusto, complexo e normalmente propicia resultados melhores, mas possui maior custo computacional.

Em problemas de classificação o resultado que mais aparece será o escolhido (moda), já em regressão será a média.



# DIFERENÇA ENTRE RANDOM FOREST E DECISION TREE

## Árvore de decisão:

- apenas uma árvore.
- cria regras para seleção das melhores variáveis.
- resultado é “fruto” de uma única árvore.

## Random Forest:

- conjunto de árvores.
- seleção das variáveis aleatoriamente.
- resultado é a moda ou média de todas as árvores.

## Vantagens

- Resultados bastante precisos.
- Normalmente não necessitam de preparações sofisticadas nos dados (label Encoder e OneHot Encoder).
- Trabalha com valores faltantes, variáveis categóricas e numéricas.
- Pouca probabilidade de ocorrência de overfitting.

## **Desvantagens**

- Velocidade de processamento relativamente baixa.
- Difícil interpretação de como chegou no resultado.