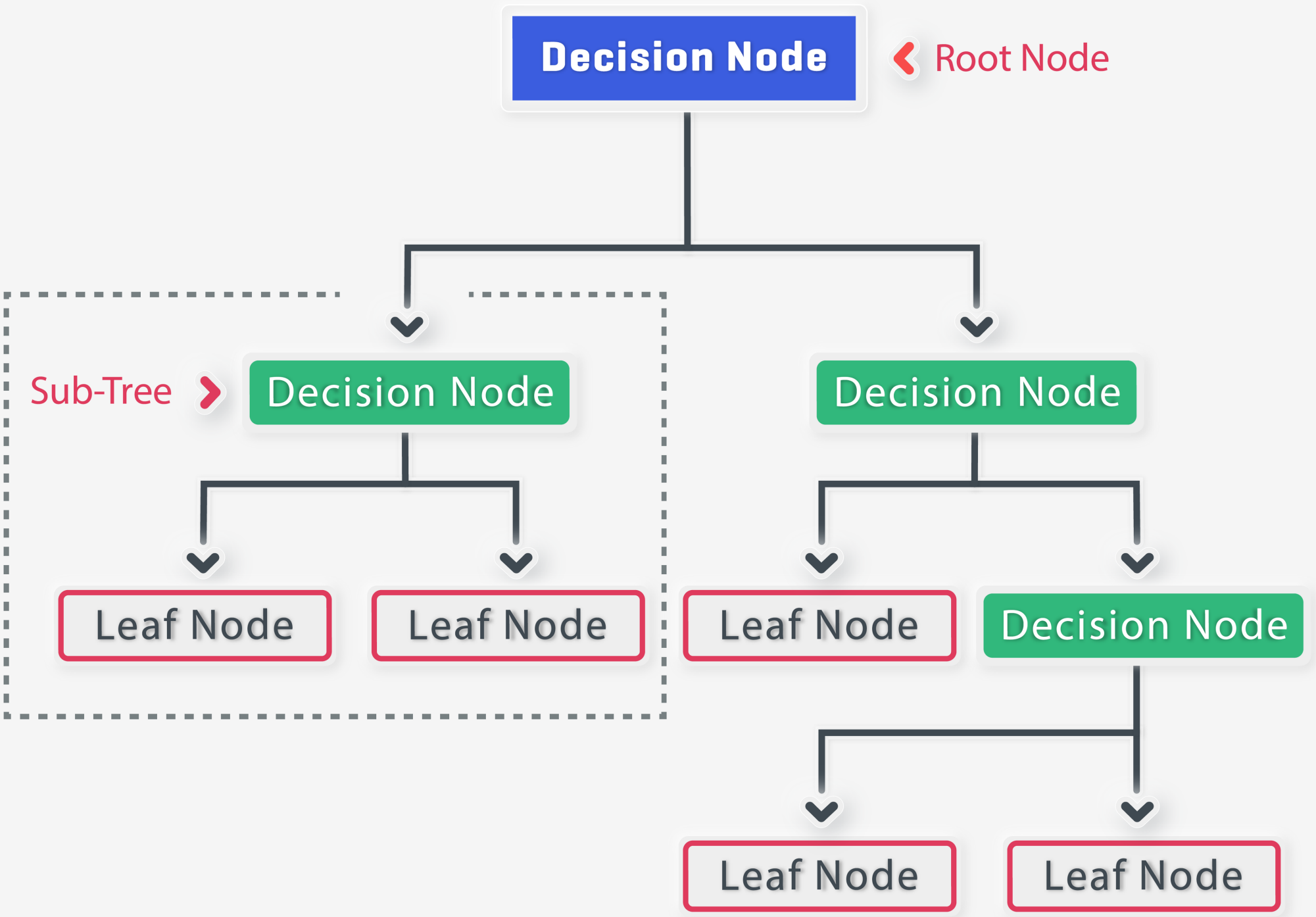
DECISION TREES



DATA-DRIVEN
SCIENCE

Decision Trees

Decision Trees (DTs) are a **non-parametric** supervised learning method used for classification and regression.

The goal is to create a model that **predicts the value** of a target variable by learning simple decision rules inferred from the data features.

A tree can be seen as a piecewise **constant approximation**.

DATA-DRIVEN
SCIENCE

Advantages of Decision Trees

Simple to understand and to interpret.
Trees can be **visualised**.

Requires **little** data preparation

The cost of using the tree (i.e., predicting data) is **logarithmic** in the number of data points used to train the tree.

Able to **handle** both numerical and categorical data

DATA-DRIVEN
SCIENCE

Advantages of Decision Trees

Able to handle **multi-output** problems.

Possible to validate a model using **statistical tests**

Uses a **white box** model. If a given situation is observable in a model, the explanation for the condition is easily explained by boolean logic.

Performs well even if its assumptions are somewhat violated by the true model from which the data were generated.

DATA-DRIVEN
SCIENCE

Disdvantages of Decision Trees

Decision-tree learners can create **over-complex trees** that do not generalise the data well

---

Predictions of decision trees are **neither smooth nor continuous**, but piecewise constant approximations.

Decision trees can be **unstable** because small variations in the data might result in a completely different tree being generated. This problem is mitigated by using decision trees within an ensemble.

DATA-DRIVEN
SCIENCE

Disdvantages of Decision Trees

There are concepts that are hard to learn because decision trees **do not express** them easily, such as XOR, parity or multiplexer problems.

Decision tree learners **create biased trees** if some classes dominate. It is therefore recommended to balance the dataset prior to fitting with the decision tree.

DATA-DRIVEN
SCIENCE

Get daily updates on
Data Science concepts

FOLLOW NOW

DATA-DRIVEN
SCIENCE