# RANDOM FOREST IN MACHINE LEARNING

# What is Random Forest?

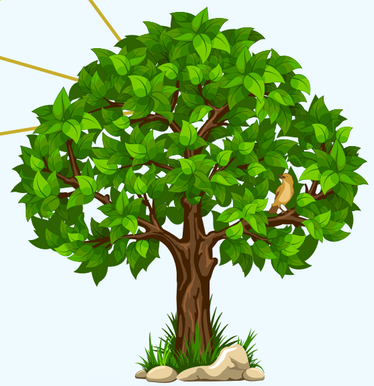
**Random Forest** is a Machine Learning algorithm based on **Supervised Learning**.

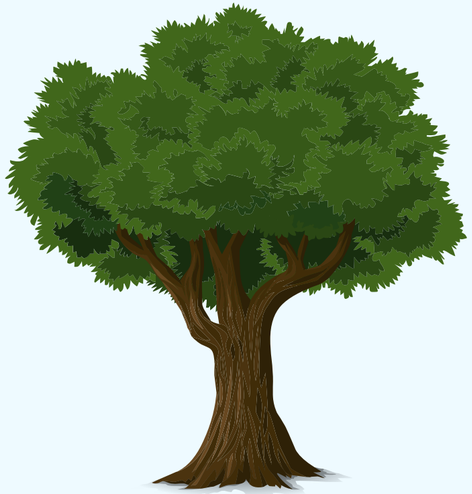
It can be used for both **Regression** and **Classification** problems.

It is a collection of multiple **Decision Trees** which is called the **Forest**.

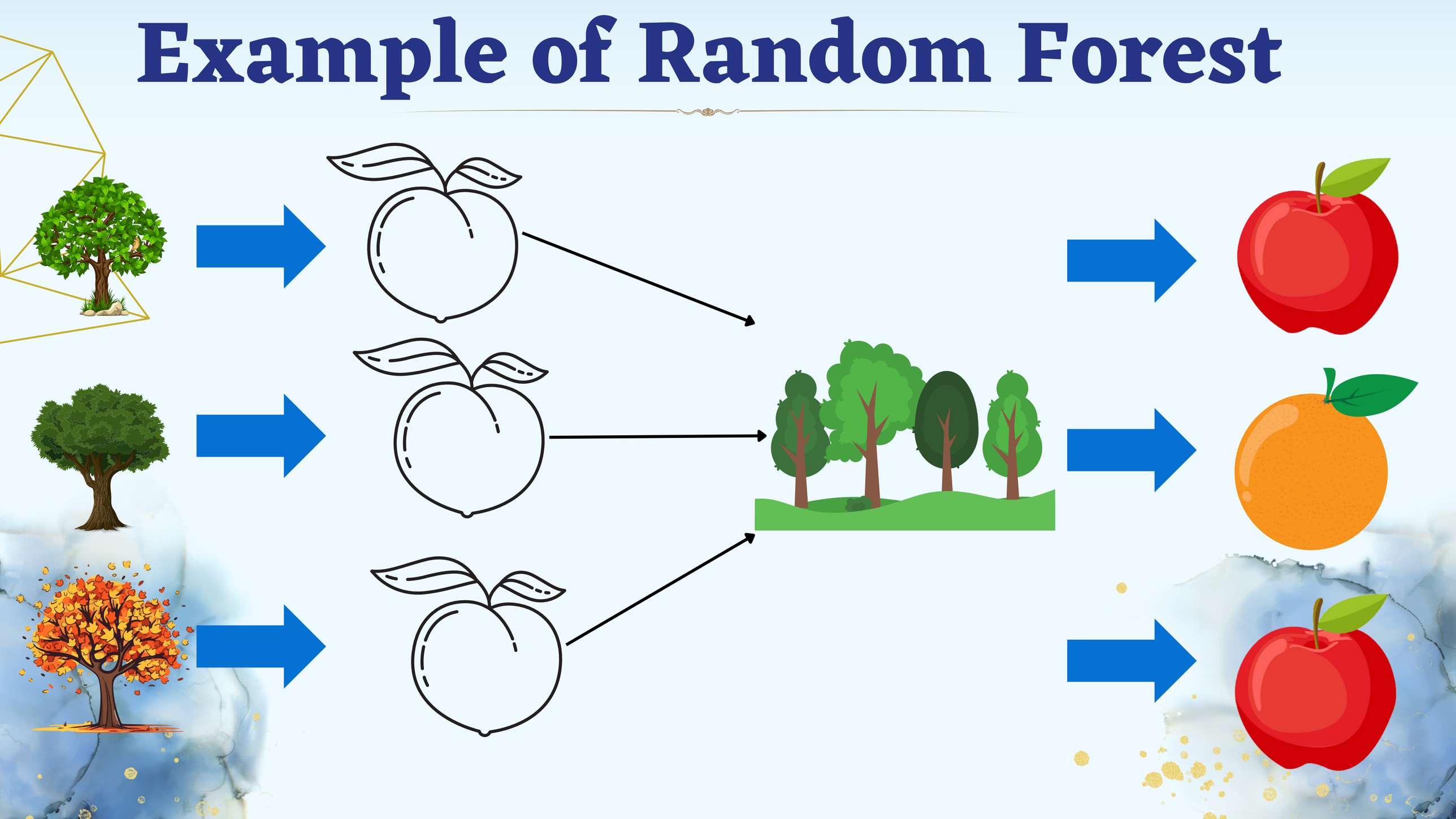
# Definition of Random Forest

**Random Forest** is constructed using multiple Decision Trees and the final decision is obtained by majority votes of the Decision Trees.

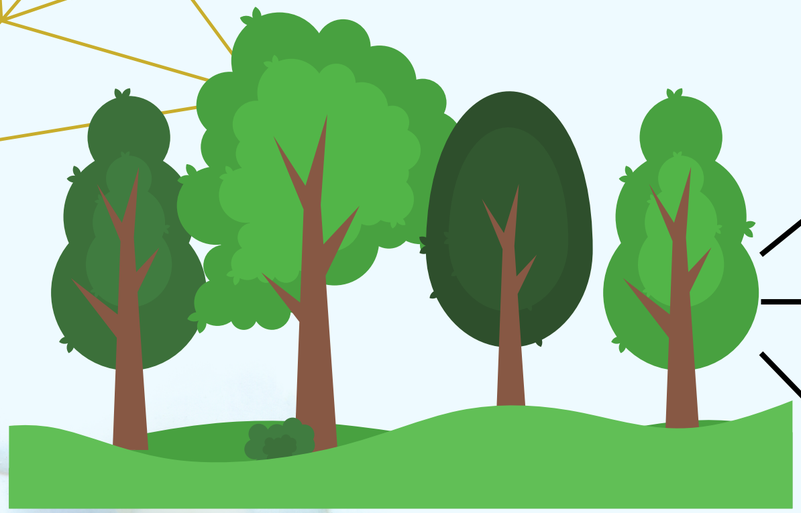


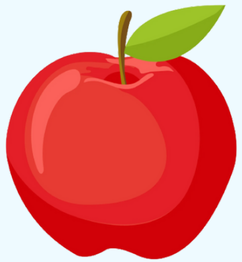


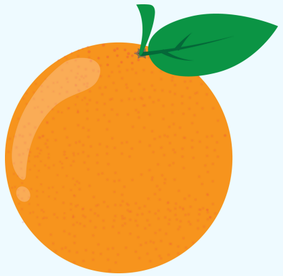
# Example of Random Forest

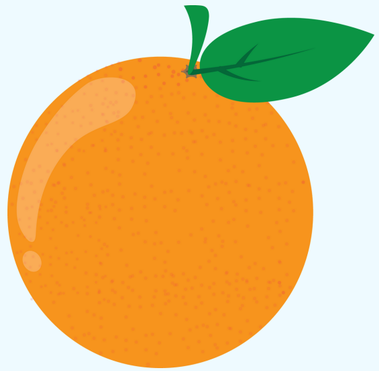


# Example of Random Forest

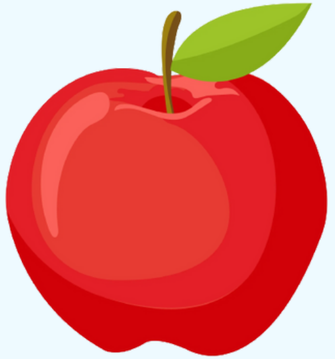




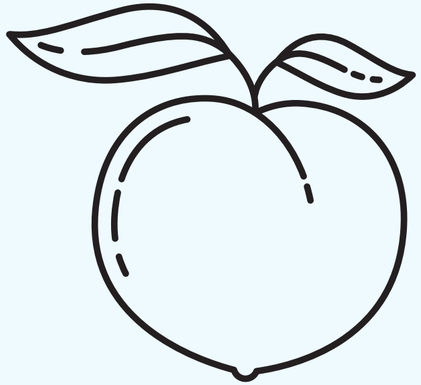


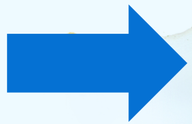


1



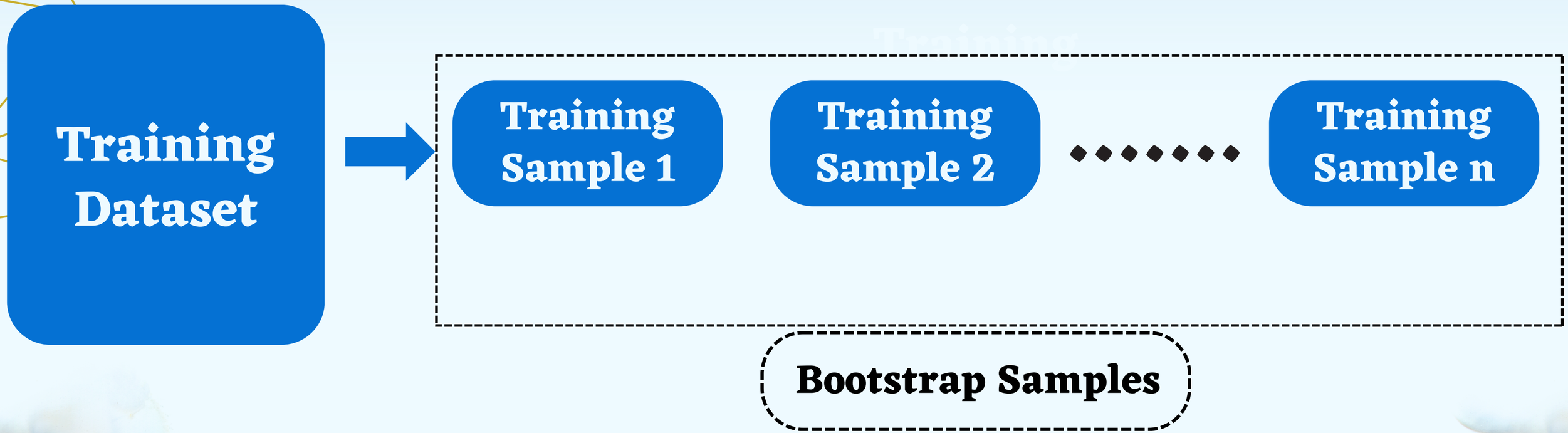
2





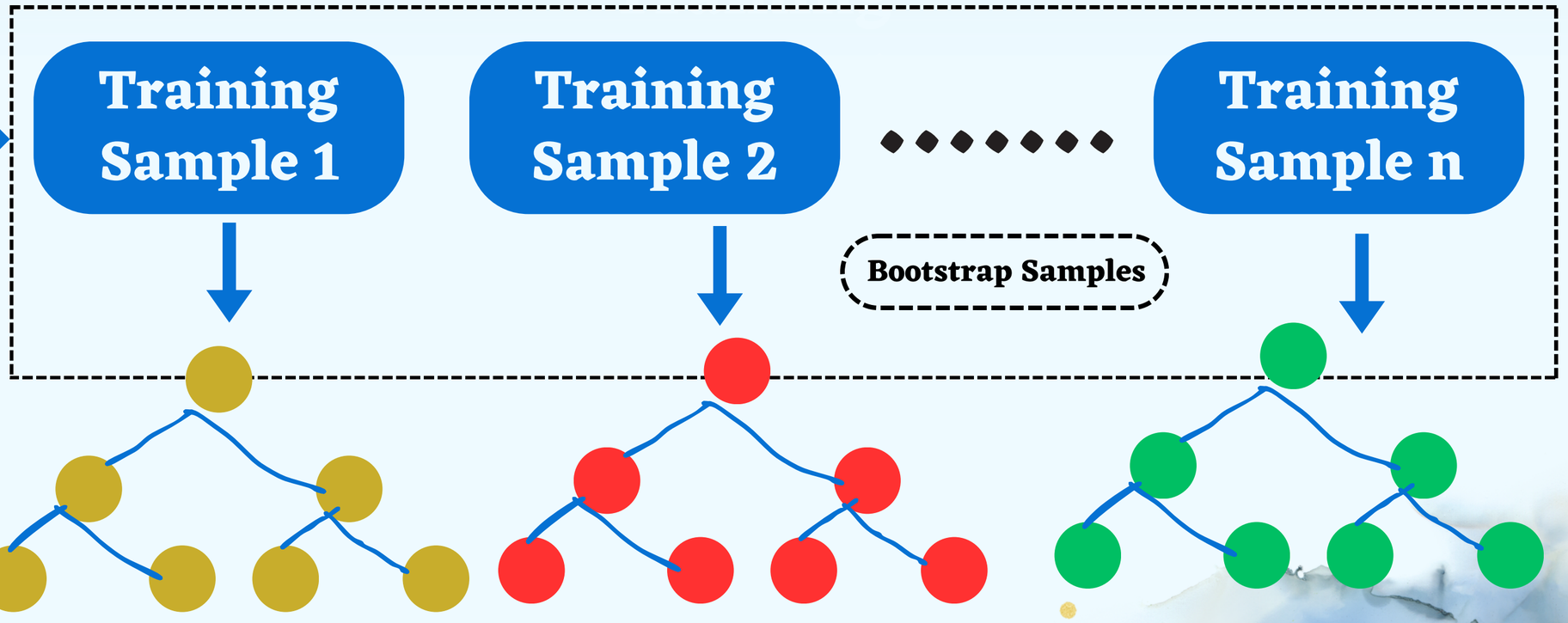
# How Random Forest works?

# How Random Forest works?



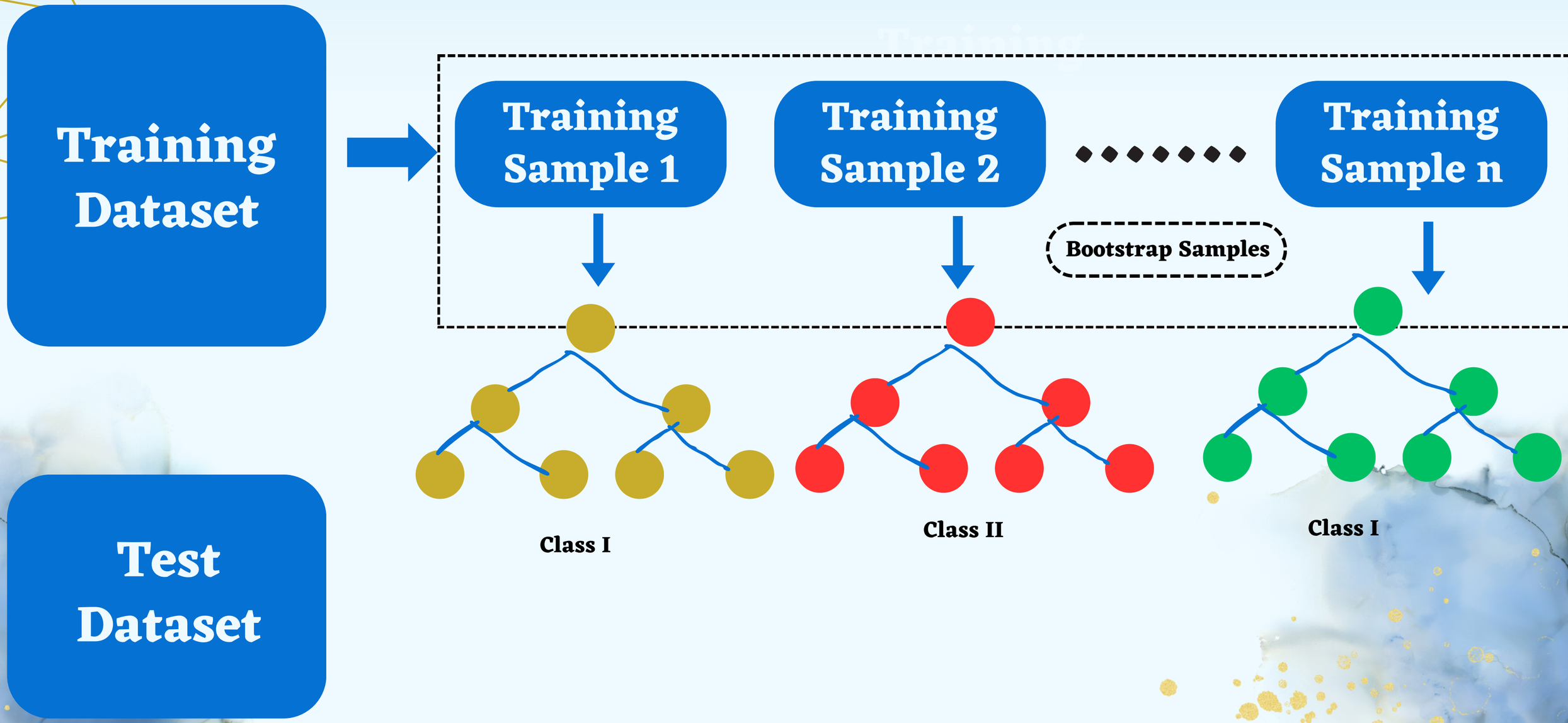
# How Random Forest works?

Training Dataset

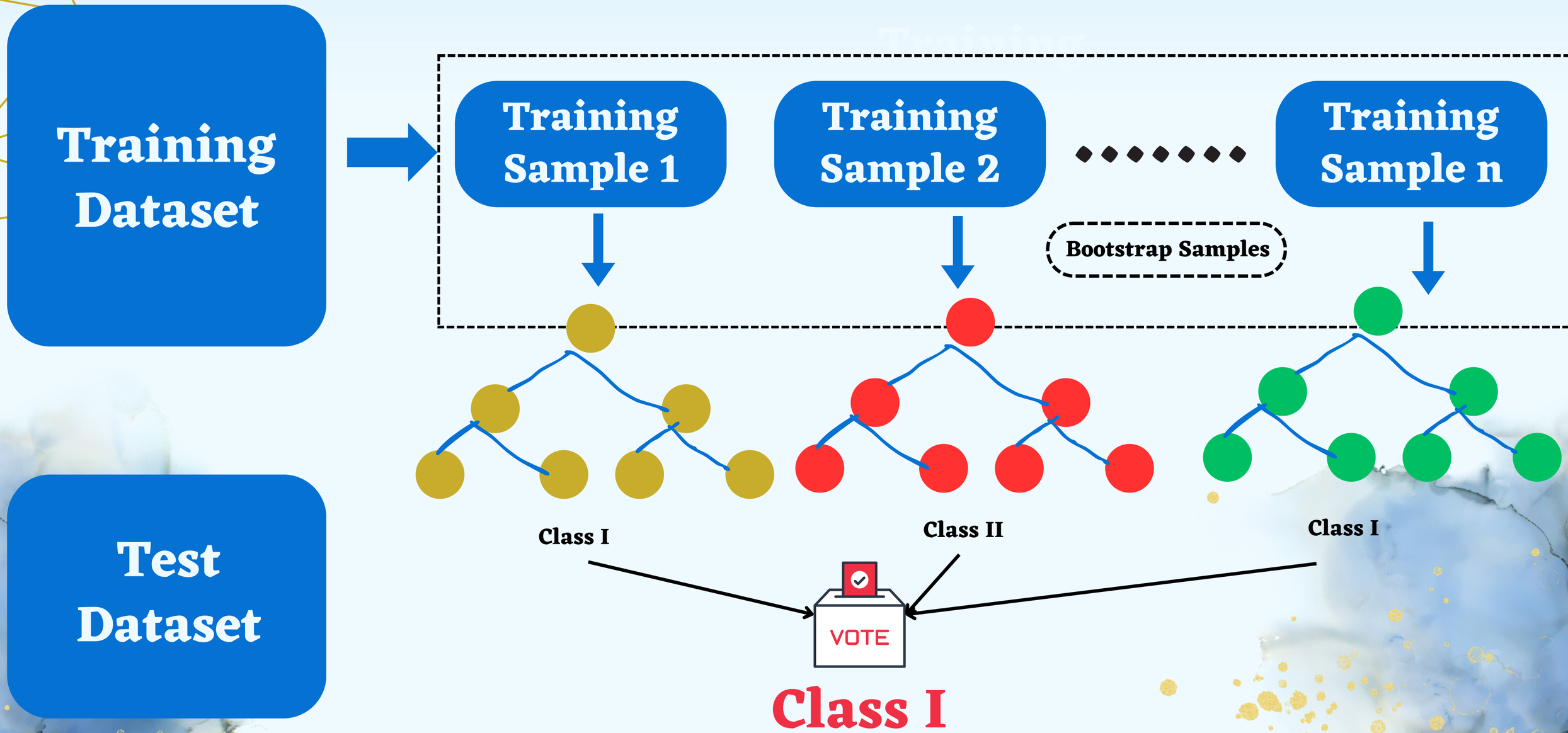



Test Dataset

# How Random Forest works?



# How Random Forest works?



# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

| id |
|---|
| 2 |
| 0 |
| 2 |
| 4 |
| 5 |
| 5 |

Dataset 1

| id |
|---|
| 2 |
| 1 |
| 3 |
| 1 |
| 4 |
| 4 |

Dataset 2

| id |
|---|
| 4 |
| 1 |
| 3 |
| 0 |
| 0 |
| 2 |

Dataset 3

| id |
|---|
| 3 |
| 3 |
| 2 |
| 5 |
| 1 |
| 2 |

Dataset 4

# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

| id |
|---|
| 2 |
| 0 |
| 2 |
| 4 |
| 5 |
| 5 |

Dataset 1

| id | x0 | x1 | y |
|---|---|---|---|
| 2 | 2.7 | 4.8 | 0 |
| 0 | 4.3 | 4.9 | 0 |
| 2 | 2.7 | 4.8 | 0 |
| 4 | 6.5 | 2.9 | 1 |
| 5 | 2.7 | 6.7 | 1 |
| 5 | 2.7 | 6.7 | 1 |

# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

| id |
|---|
| 2 |
| 1 |
| 3 |
| 1 |
| 4 |
| 4 |

| id | x2 | x3 | y |
|---|---|---|---|
| 2 | 4.1 | 5 | 0 |
| 1 | 5.9 | 5.5 | 0 |
| 3 | 4.5 | 3.9 | 1 |
| 1 | 5.9 | 5.5 | 0 |
| 4 | 4.7 | 4.6 | 1 |
| 4 | 4.7 | 4.6 | 1 |

Dataset 2

# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

| id |
|---|
| 4 |
| 1 |
| 3 |
| 0 |
| 0 |
| 2 |

| id | x2 | x4 | y |
|---|---|---|---|
| 4 | 4.7 | 6.1 | 1 |
| 1 | 5.9 | 5.9 | 0 |
| 3 | 4.5 | 5.9 | 1 |
| 0 | 4.1 | 5.5 | 0 |
| 0 | 4.1 | 5.5 | 0 |
| 2 | 4.1 | 5.6 | 0 |

Dataset 3

# Random Forest on Dataset

| id | x0 | x1 | x2 | x3 | x4 | y |
|---|---|---|---|---|---|---|
| 0 | 4.3 | 4.9 | 4.1 | 4.7 | 5.5 | 0 |
| 1 | 3.9 | 6.1 | 5.9 | 5.5 | 5.9 | 0 |
| 2 | 2.7 | 4.8 | 4.1 | 5 | 5.6 | 0 |
| 3 | 6.6 | 4.4 | 4.5 | 3.9 | 5.9 | 1 |
| 4 | 6.5 | 2.9 | 4.7 | 4.6 | 6.1 | 1 |
| 5 | 2.7 | 6.7 | 4.2 | 5.3 | 4.8 | 1 |

| id |
|---|
| 3 |
| 3 |
| 2 |
| 5 |
| 1 |
| 2 |

| id | x1 | x3 | y |
|---|---|---|---|
| 3 | 4.4 | 3.9 | 1 |
| 3 | 4.4 | 3.9 | 1 |
| 2 | 4.8 | 5 | 0 |
| 5 | 6.7 | 5.3 | 1 |
| 1 | 6.1 | 5.5 | 0 |
| 2 | 4.8 | 5 | 0 |

Dataset 4

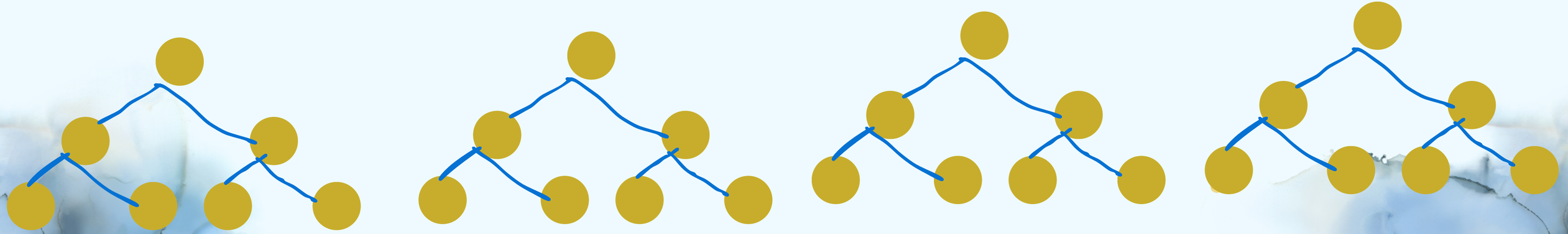
# Random Forest on Dataset

| id | x0 | x1 | y |
|---|---|---|---|
| 2 | 2.7 | 4.8 | 0 |
| 0 | 4.3 | 4.9 | 0 |
| 2 | 2.7 | 4.8 | 0 |
| 4 | 6.5 | 2.9 | 1 |
| 5 | 2.7 | 6.7 | 1 |
| 5 | 2.7 | 6.7 | 1 |

| id | x2 | x3 | y |
|---|---|---|---|
| 2 | 4.1 | 5 | 0 |
| 1 | 5.9 | 5.5 | 0 |
| 3 | 4.5 | 3.9 | 1 |
| 1 | 5.9 | 5.5 | 0 |
| 4 | 4.7 | 4.6 | 1 |
| 4 | 4.7 | 4.6 | 1 |

| id | x2 | x4 | y |
|---|---|---|---|
| 4 | 4.7 | 6.1 | 1 |
| 1 | 5.9 | 5.9 | 0 |
| 3 | 4.5 | 5.9 | 1 |
| 0 | 4.1 | 5.5 | 0 |
| 0 | 4.1 | 5.5 | 0 |
| 2 | 4.1 | 5.6 | 0 |

| id | x1 | x3 | y |
|---|---|---|---|
| 3 | 4.4 | 3.9 | 1 |
| 3 | 4.4 | 3.9 | 1 |
| 2 | 4.8 | 5 | 0 |
| 5 | 6.7 | 5.3 | 1 |
| 1 | 6.1 | 5.5 | 0 |
| 2 | 4.8 | 5 | 0 |



| x0 | x1 | x2 | x3 | x4 |
|---|---|---|---|---|
| 5.8 | 4.2 | 6.9 | 2.1 | 6.7 |

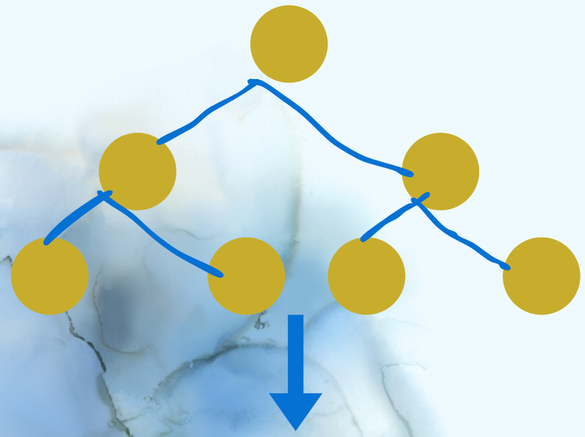
# Random Forest on Dataset

| id | x0 | x1 | y |
|---|---|---|---|
| 2 | 2.7 | 4.8 | 0 |
| 0 | 4.3 | 4.9 | 0 |
| 2 | 2.7 | 4.8 | 0 |
| 4 | 6.5 | 2.9 | 1 |
| 5 | 2.7 | 6.7 | 1 |
| 5 | 2.7 | 6.7 | 1 |

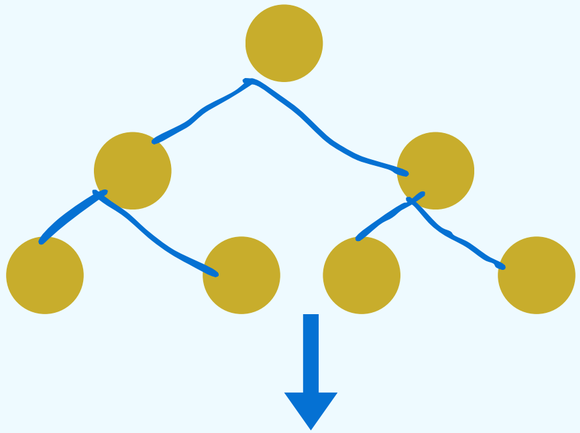
| id | x2 | x3 | y |
|---|---|---|---|
| 2 | 4.1 | 5 | 0 |
| 1 | 5.9 | 5.5 | 0 |
| 3 | 4.5 | 3.9 | 1 |
| 1 | 5.9 | 5.5 | 0 |
| 4 | 4.7 | 4.6 | 1 |
| 4 | 4.7 | 4.6 | 1 |

| id | x2 | x4 | y |
|---|---|---|---|
| 4 | 4.7 | 6.1 | 1 |
| 1 | 5.9 | 5.9 | 0 |
| 3 | 4.5 | 5.9 | 1 |
| 0 | 4.1 | 5.5 | 0 |
| 0 | 4.1 | 5.5 | 0 |
| 2 | 4.1 | 5.6 | 0 |

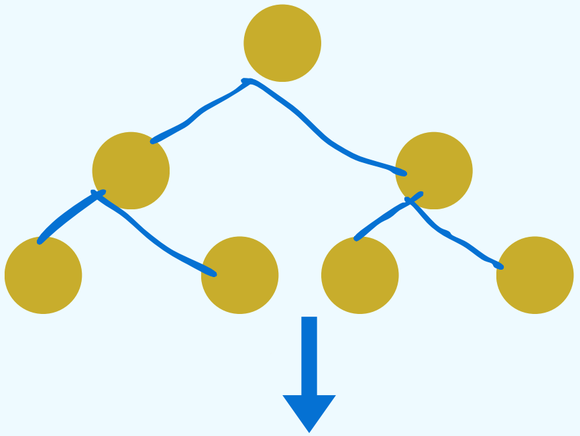
| id | x1 | x3 | y |
|---|---|---|---|
| 3 | 4.4 | 3.9 | 1 |
| 3 | 4.4 | 3.9 | 1 |
| 2 | 4.8 | 5 | 0 |
| 5 | 6.7 | 5.3 | 1 |
| 1 | 6.1 | 5.5 | 0 |
| 2 | 4.8 | 5 | 0 |



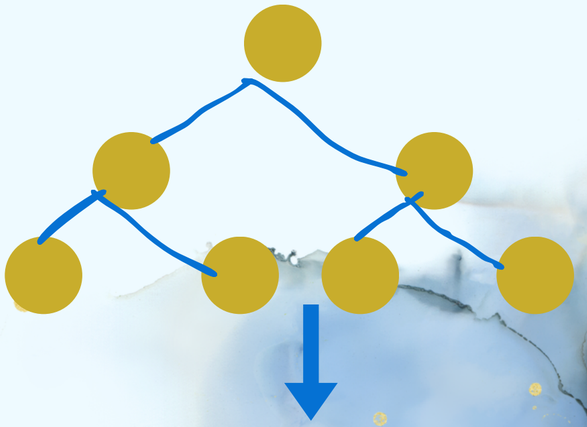
Class I



Class II



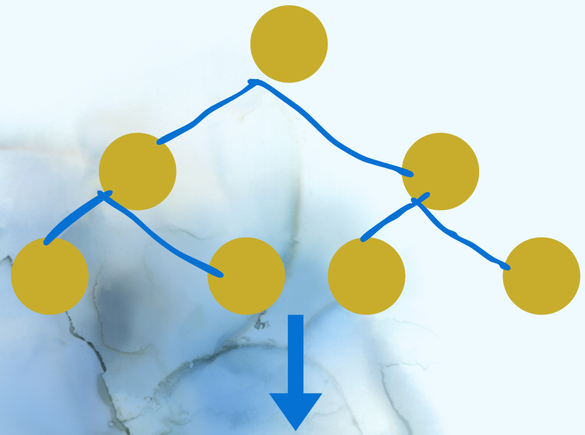
Class II



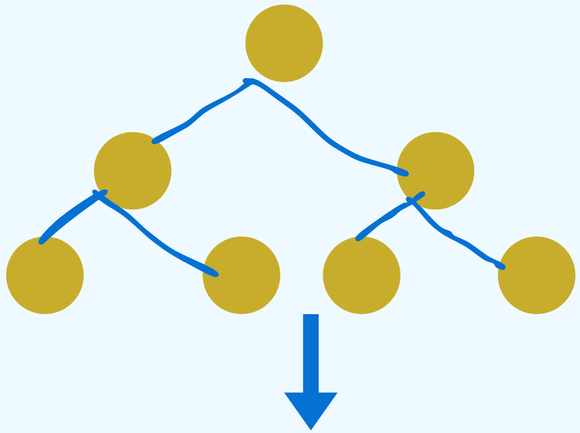
Class II

# Random Forest on Dataset

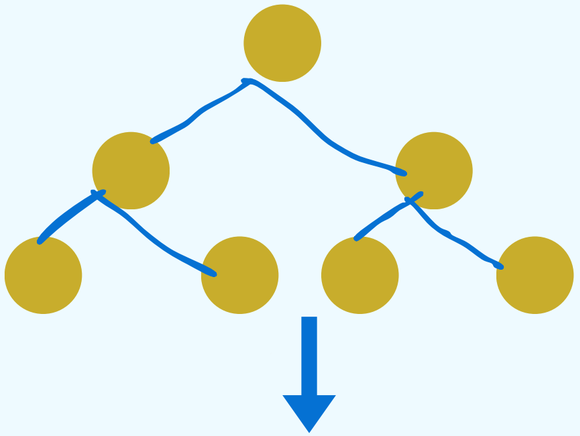
**Aggregation:** The process of combining multiple results from multiple models is called Aggregation.



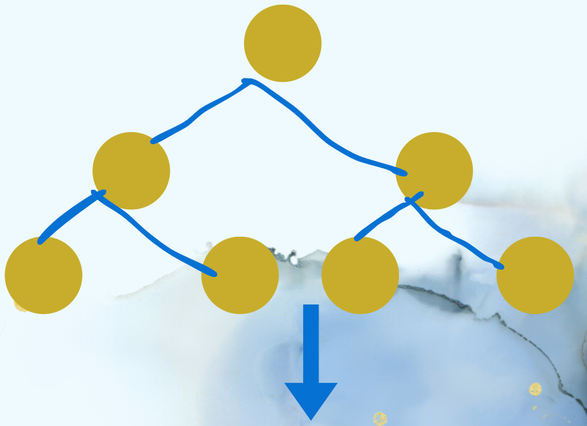
Class I



Class II



Class II



Class II

# Random Forest Important Questions

## Why is it called Random?

We have used 2 random processes: Bootstrapping and Random Feature Selection.

# Random Forest Important Questions

## How many features consider?

**Values close to log and sqrt of the total number of features will work well.**

# Random Forest Important Questions

## How to use this for Regression?

Just take the average values.