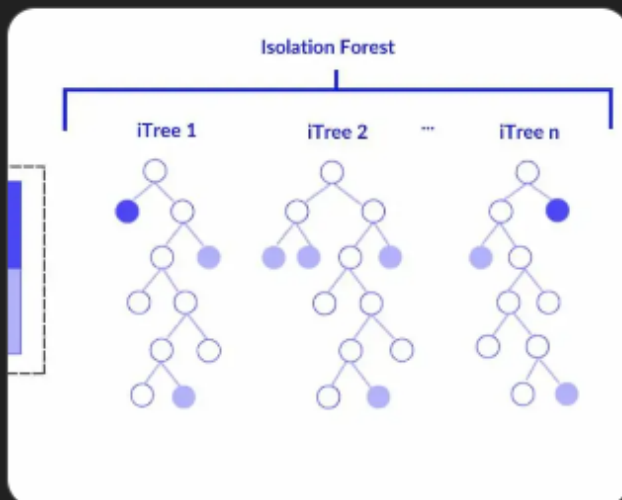
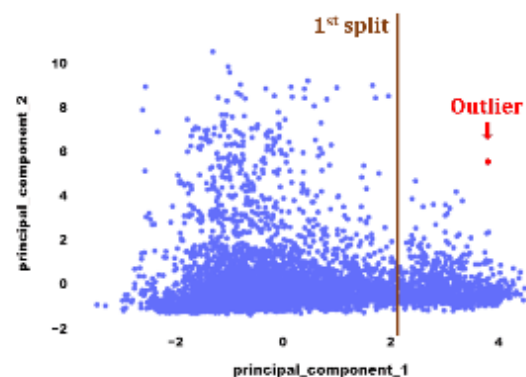


# ✓ What is Isolation Forest?



Isolation Forest randomly splits the sample until a po



Isolation Forest (iForest) is an anomaly detection machine learning algorithm.

It works by:

- Randomly isolating points.
- Normal points require many cuts to isolate.
- Anomalies require very few cuts.

## ★ Key Idea

Anomalies are:

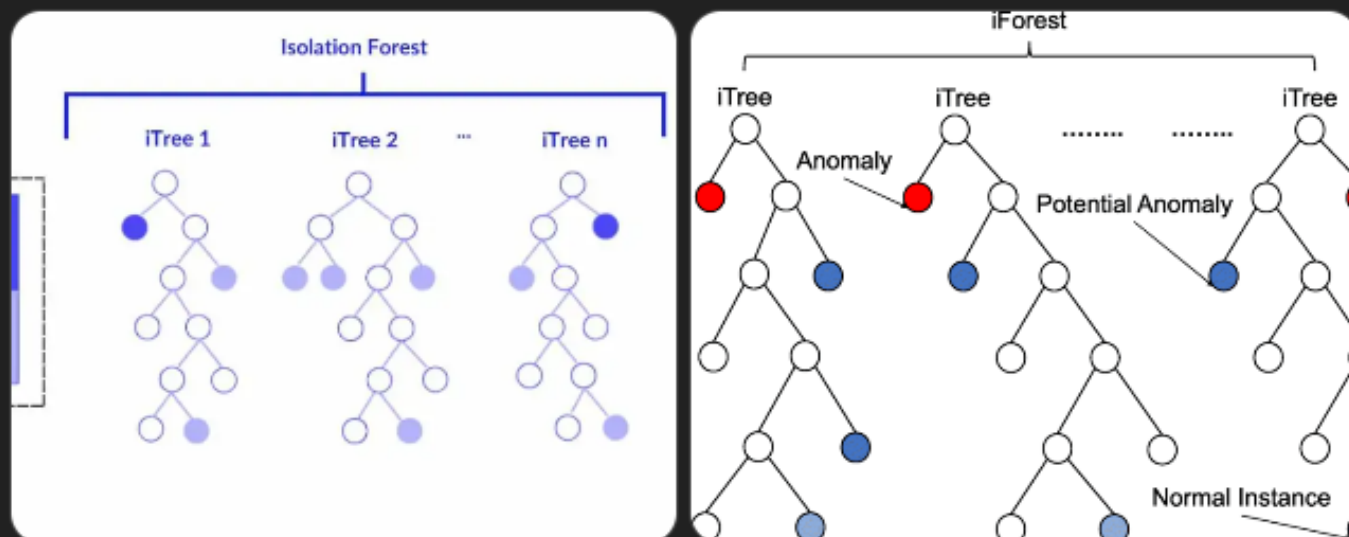
- far away from others
- easy to separate
- isolated with **few random splits**

Normal points are:

- buried deep inside clusters
- take **many splits** to isolate



# ? What is an Isolation Tree? (iTree)



An Isolation Tree is the **basic building block** of Isolation Forest.

It is similar to a decision tree BUT:

✓ **Splits are random, NOT based on entropy or gini**

- Choose a random feature
- Choose a random split value

✓ **The goal is NOT accuracy**

The goal is to **isolate** each point.

✓ **Depth = number of splits required to isolate a point**

- Small depth → **likely anomaly**
- Large depth → **likely normal**

## How Isolation Forest Works (Step-by-Step)

1. Build **many** isolation trees (10–100 trees)
2. Each tree:
  - randomly selects rows
  - randomly selects a feature
  - randomly splits it
3. Measure **path length** of each data point.
4. Anomalies have **short average path length**.

## Now Let's Take a Simple Example (Dry Run)

We take this small dataset:

Point	Value
A	10
B	12
C	11
D	13
E	50 (ANOMALY)

Clearly,  $E = 50$  is far from others.



# Build an Isolation Tree (Dry Run)

We repeatedly make random cuts until each point is isolated.

## Step 1 — Random Split 1

Range of values = [10, 50]

Let's say random split chosen:

```
java
```

```
Split at value = 20
```

**Left side ( $\leq 20$ ):**

A=10, B=12, C=11, D=13

**Right side ( $> 20$ ):**

E=50

🔗 E is isolated in just 1 split → very suspicious

## Left Subtree (Step 2)

Values now: 10, 12, 11, 13

Random split chosen:

```
mathematica
```

```
Split at 12.5
```

Left: A=10, C=11

Right: B=12, D=13



## Step 3 — For A=10 and C=11

Random split:

mathematica

Split at 10.8

Left: A=10

Right: C=11

Both isolated in 3 steps.

---

## Step 3 — For B=12 and D=13

Random split:

mathematica

Split at 12.4

Left: B=12

Right: D=13

Both isolated in 3 steps.

---

## ★ Path Length Summary

Point	Splits needed
E = 50	1 step (very LOW → anomaly)
A = 10	3
B = 12	3
C = 11	3
D = 13	3

### Interpretation:

- E is isolated extremely fast → **anomaly**
- Others need more splits → **normal**

## ★ Why E is anomaly?

Because it is far from the cluster

→ easy to isolate

→ requires 1 cut.