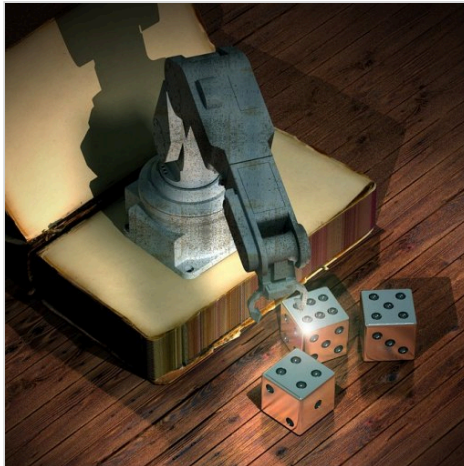


L U M B R E



# Bagging y Random Forest

**Grado:** 5

**Materia:** Aprendizaje Supervisado · Machine Learning

11/10/2026

## Antes de empezar

Un árbol es inestable y se sobreajusta; cien árboles decorrelados y promediados forman uno de los clasificadores más robustos que existen. El Random Forest domesticó la voracidad del árbol sin sacrificar su potencia, y reinó en competitions durante años.

### QUÉ APRENDERÁS

- Entender cómo el bagging reduce la varianza promediando muchos árboles entrenados con muestreo bootstrap, y qué añade el Random Forest al seleccionar características al azar.

## Bagging y Random Forest

### Muchos árboles débiles, un bosque fuerte

La sabiduría de los modelos: ninguno individual acierta siempre, pero el promedio de muchos árboles distintos, entrenados sobre datos y características diferentes, cancela errores aleatorios y revela la señal estable que subyace.

#### VIDEO

### Random Forests: cómo funcionan

Vídeo en español que explica cómo el promedio de muchos árboles remuestreados reduce la varianza.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=jLJ4uKS9D5A>

## El problema del árbol único

Ya vimos que el árbol es inestable: pequeños cambios en los datos reorganizan por completo sus cortes. Esa alta varianza significa que su predicción depende demasiado de qué ejemplos concretos vio. El bagging ataca justo ese defecto.

## Bootstrap: remuestrear con reposición

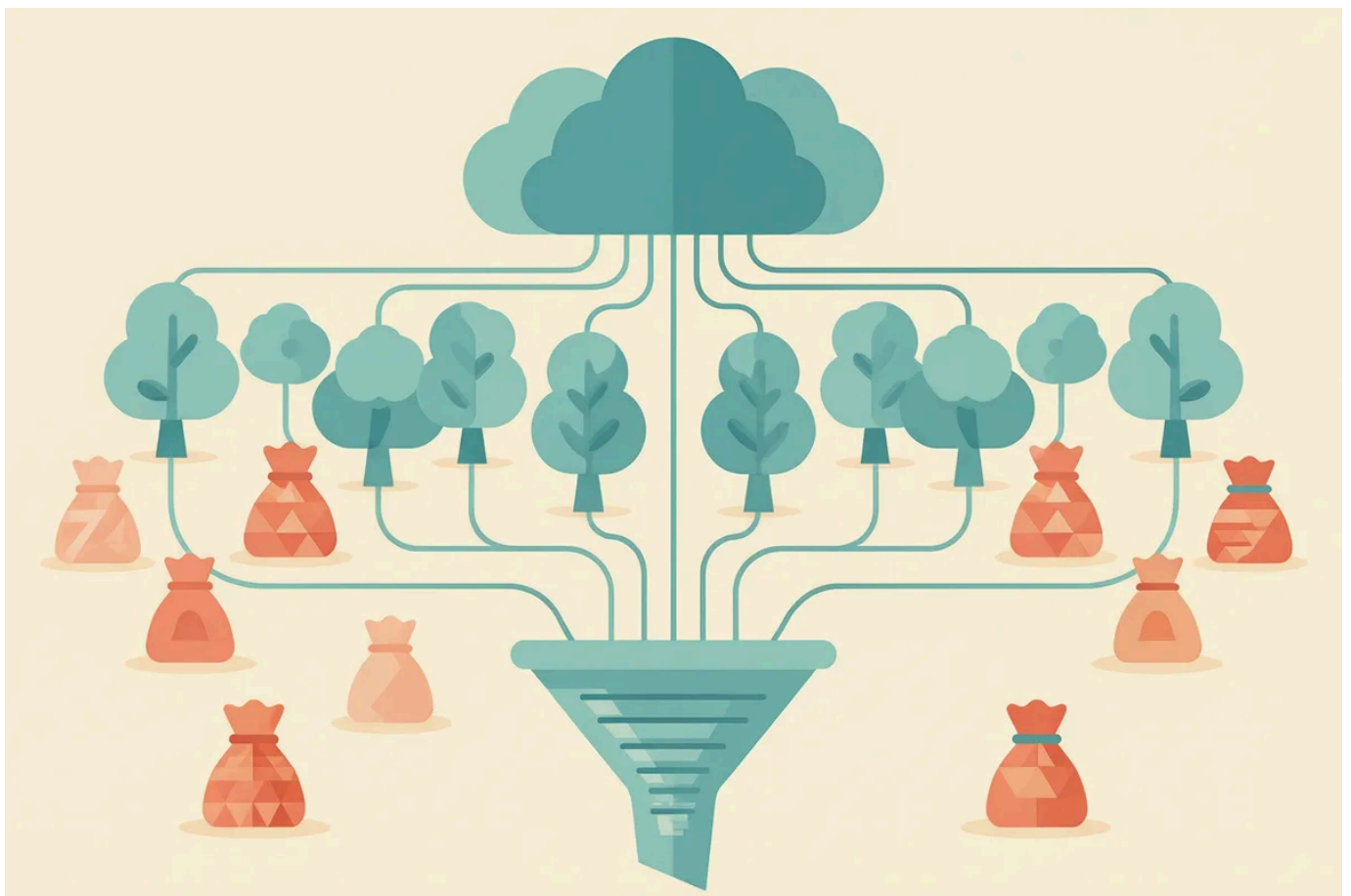
Para fabricar varios conjuntos de entrenamiento distintos, se muestrea el dataset original con reposición, extrayendo tantas filas como hay. Cada muestra omite unos ejemplos y repite otros, generando árboles que ven versiones ligeramente diferentes de la realidad.

## Promediar para reducir varianza

Se entrena un árbol por cada muestra bootstrap y se combinan sus predicciones: por mayoría en clasificación, por media en regresión. Los errores individuales, no correlacionados, tienden a cancelarse, y el promedio es mucho más estable que cualquier miembro.

## La descorrelación es la clave

El truco no es tener muchos árboles, sino árboles poco correlacionados. Si todos aprendieran lo mismo, promediar no reduciría nada. A mayor descorrelación entre miembros, mayor la caída de varianza del conjunto.



*El bagging hecho imagen: cada bolsa remuestreada cría un árbol distinto, y el embudo combina sus votos en una predicción mucho más estable que cualquier árbol suelto.*

## Qué añade el Random Forest

Además del bootstrap, en cada nodo el bosque elige el mejor corte solo entre un subconjunto aleatorio de características. Esto rompe la dominación de las variables más fuertes y descorrelaciona aún más los árboles que el simple bagging, mejorando la generalización.

## Importancia de características

Contar cuánto contribuye cada variable a reducir la impureza a lo largo de todos los árboles da una medida de importancia útil para explorar datos. Conviene tomarla como brújula, no como veredicto: sesga hacia variables de muchas categorías.

## Fuera de bolsa y robustez

Cada árbol deja de ver cerca de un tercio de los ejemplos por el bootstrap. Esos casos fuera de bolsa sirven de validación interna gratis, y el bosque, al promediar, es notablemente resistente al ruido y a la falta de escalado.

## Errores frecuentes

- Pocos árboles: el bosque necesita cientos para que el promedio estabilice la varianza.
- Leer la importancia de características como relación causal y no como correlación.
- Sobreinterpretar su precisión creyendo que no puede sobreajustar: con árboles muy profundos sí puede.

## Ejemplo resuelto: un bosque para priorizar llamadas

Un centro de atención quiere predecir qué clientes llamarán para quejarse. El montaje:

1. Partir del árbol inestable: un solo árbol cambia mucho según qué datos ve; hay que estabilizarlo.
2. Remuestrear: generar cientos de muestras bootstrap, cada una con su árbol profundo.
3. Descorrelacionar: en cada nodo elegir solo un subconjunto aleatorio de variables.
4. Votar: promediar las probabilidades de todos los árboles en la predicción final.
5. Explorar: usar la importancia de características como brújula, no como prueba causal.

## ¿Para qué sirve en la realidad?

El bosque es el por defecto honesto de la ciencia de datos tabular: apenas hay que afinar, tolera variables sin escalar, mezcla tipos y resiste el ruido. Si no le ganas con algo más complejo, no complica.

### 1. ¿Qué defecto del árbol de decisión corrige principalmente el bagging?

- ☐ Su alta varianza e inestabilidad ante cambios en los datos
- ☐ Su incapacidad para fronteras no lineales
- ☐ Su lentura al predecir
- ☐ Su necesidad de escalar características

### 2. En el Random Forest, cada nodo busca el mejor corte solo entre un subconjunto aleatorio de características.

- ☐ Verdadero
- ☐ Falso

## Cuándo conviene el bosque

Como caballo de batalla: poco que afinar, tolera escalas y tipos mixtos, maneja no linealidades y da una línea base fuerte sin preparación elaborada. Si un bosque simple no supera a tu modelo, desconfía de la complejidad añadida.

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

### 3. Bosques



#### 4. Ordena la construcción de un Random Forest:

##### Elementos:

- generar una muestra bootstrap de los datos
- entrenar un árbol profundo sobre esa muestra
- en cada nodo restringir las candidatas a un subconjunto
- repetir y combinar por votacion o media

|                  |       |
|------------------|-------|
| <b>Muestra</b>   | . . . |
| <b>Arbol</b>     | . . . |
| <b>Variables</b> | . . . |
| <b>Voto</b>      | . . . |

#### 5. Une cada término del bosque con su definición.

|    |                                                    |       |
|----|----------------------------------------------------|-------|
| 5. | remuestrear con reposición el conjunto             | ----- |
| 5. | promediar modelos entrenados en muestras distintas | ----- |
| 5. | restringir las variables candidatas en cada nodo   | ----- |
| 5. | ejemplos que un árbol no llegó a ver               | ----- |

**Opciones:** bootstrap · bagging · Random Forest · fuera de bolsa

### Muchos árboles

Random Forest

Base

árbol inestable

alta varianza

Bagging

bootstrap

promediar

Extra

variables al azar por nodo

Ventajas

poca afinación

resiste ruido

fuera de bolsa

El mapa mental interactivo es interactivo en la versión online.



¿Qué es el Random Forest? <https://www.ibm.com/es-es/think/topics/random-forest>  
Explicación de referencia en español del random forest y sus usos.

**6. Tu bosque de un solo árbol funciona regular; al subir a quinientos mejora mucho y luego se estanca. ¿Qué está pasando y por qué no explota el sobreajuste?**

---

---

---

---

---

## Referencia curricular

|                 |                                            |
|-----------------|--------------------------------------------|
| <b>Grado:</b>   | 5                                          |
| <b>Materia:</b> | Aprendizaje Supervisado · Machine Learning |

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-supla-005-bagging-y-random-forest>

Material educativo generado desde Lumbre · [lumbre.lat](https://lumbre.lat)