



# Modelos de ensamble (parte I)

Tutoría

¿En qué se diferencian  
los problemas de  
regresión v/s aquellos de  
clasificación?



# Autoaprendizaje

## *Recursos asincrónicos*

- ¿Revisaste los recursos de la semana 1 (Guía y desafío)?
- ¿Tienes dudas sobre alguno de ellos?



# Ideas fuerza



Los **Árboles de decisión** poseen **alta interpretabilidad** pero sufren de un ajuste excesivo de los datos de entrenamiento **overfitting**.



Los ensambles permiten construir modelos con mejor rendimiento por medio de la **combinación de modelos**. **Bagging** puede ser usado con modelos **homogéneos o heterogéneos**



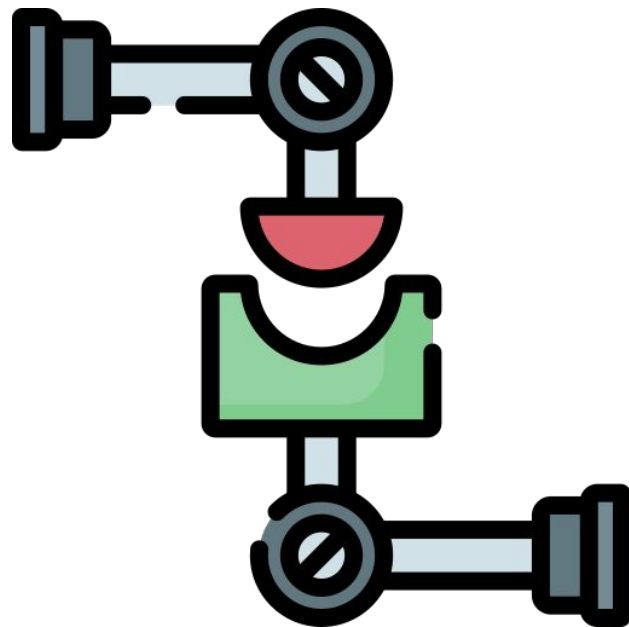
**Random Forest** posee un **alto rendimiento**, corrigiendo las **altas correlaciones** de los estimadores de Bagging. **Su rendimiento no es favorable para problemas lineales.**

**/\* Ensemble Bagging \*/**

# Ensamble

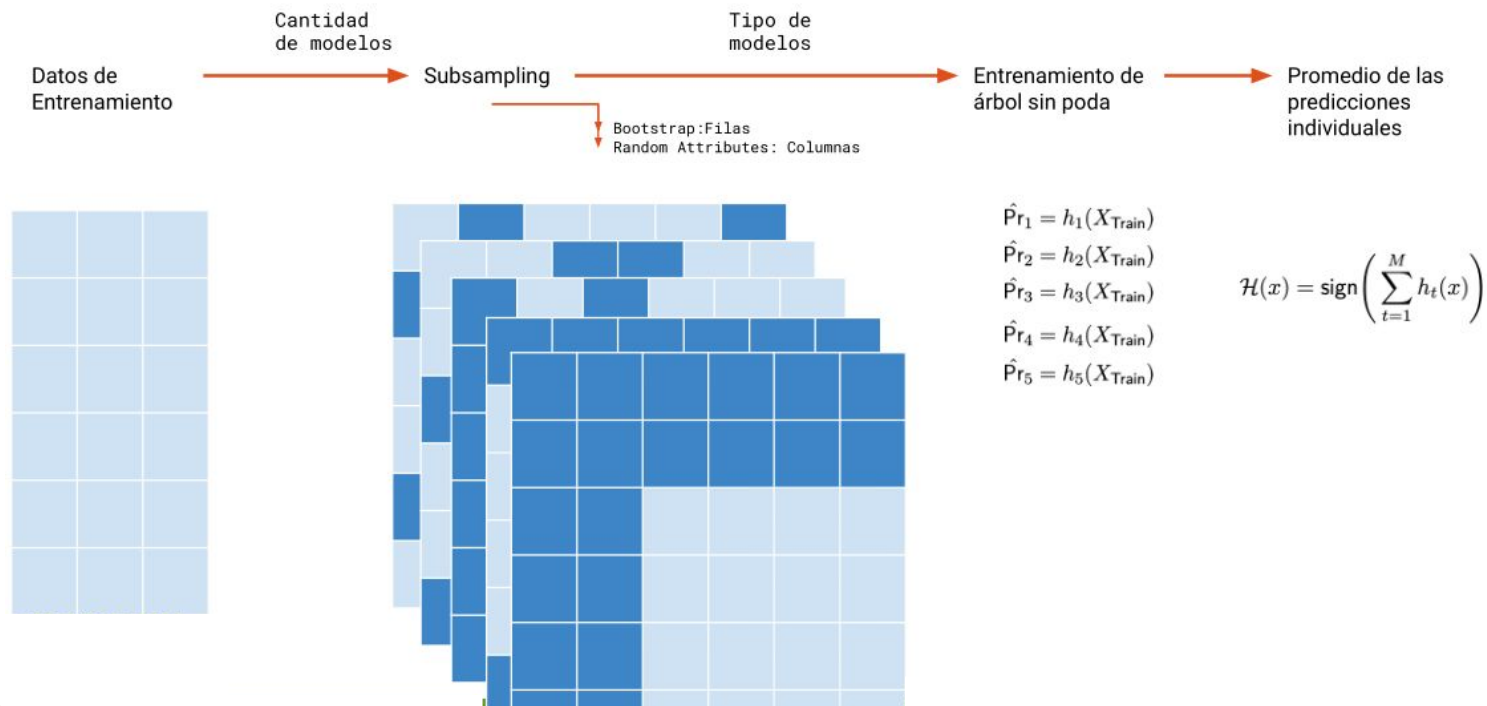
## Definición

- Los Ensamblados son algoritmos que se componen de dos o más modelos y que juntos persiguen potenciarse para lograr un mejor desempeño.
- Para decidir la clasificación (o valor) de los resultados individuales de cada clasificador, se elige democráticamente por “mayoría de votos”.



# Ensamble

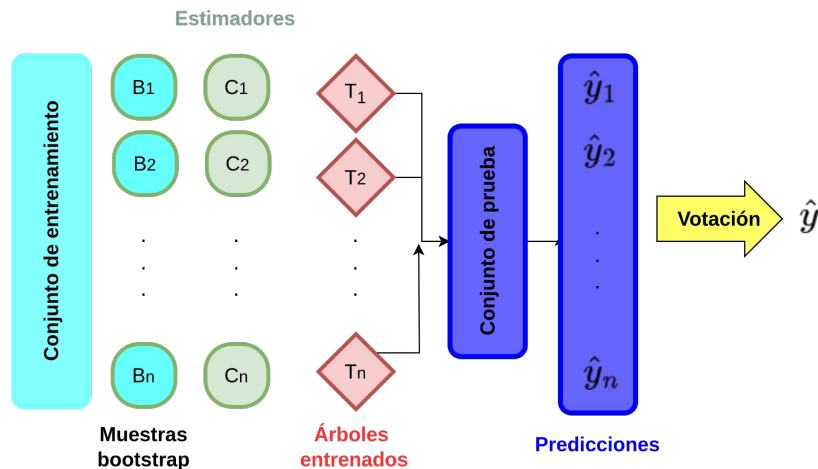
## Random Forest



# Ensamble Bagging

## Definición

- Es un algoritmo de ensamble paralelo.
- Cada estimador usa una muestra Bootstrap.
- Los estimadores (modelos) pueden ser cualquier tipo de clasificador. Por defecto en sklearn se usa predictores de árbol de decisión.
- Las estimaciones de cada modelo se combinan por medio de una votación para predecir la salida.

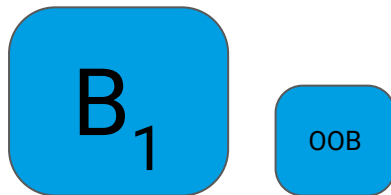




# Ensamble Bagging

## Características

- En Bagging, por defecto todos los clasificadores tienen el mismo peso.
- Cada conjunto de entrenamiento obtenido por Bootstrap dejará disponible observaciones sin usar llamadas **Out Of Bag**.



- Bagging, en general, supera el desempeño de los Árboles de decisión. Sin embargo, presenta problemas ya que los modelos que lo componen resultan ser altamente correlacionados.

**`/* Ensemble Random Forest */`**

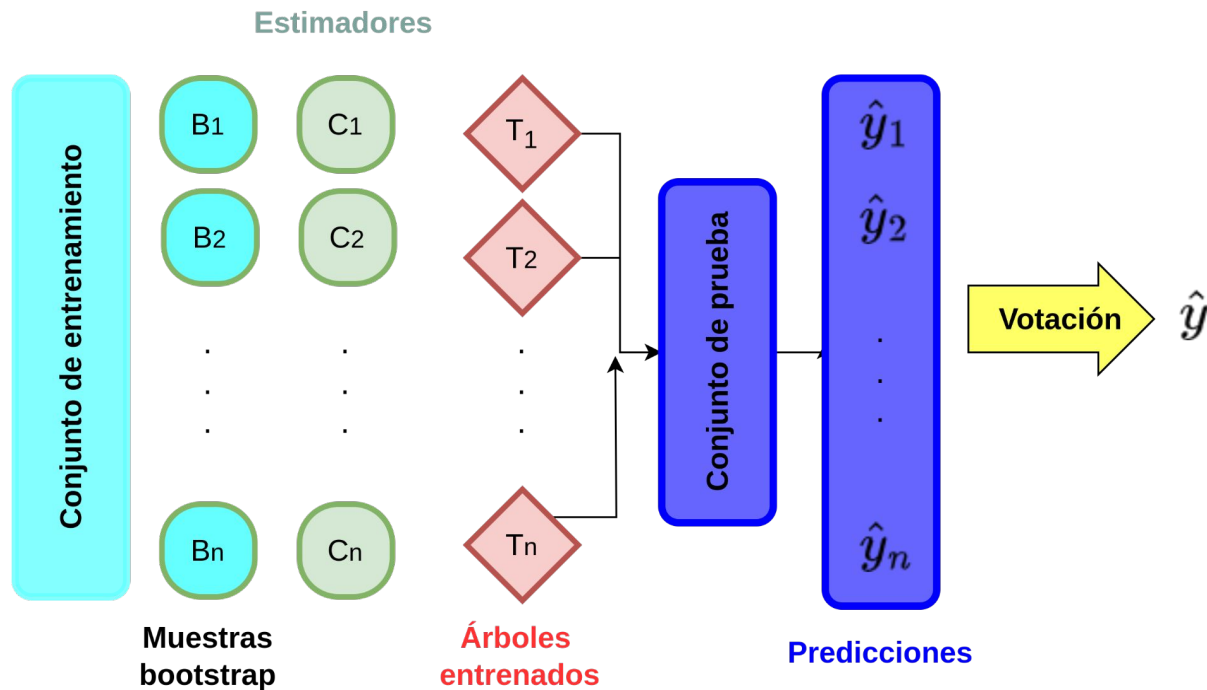
# Random Forest

## *Definición*

- Random Forest (RF) se compone exclusivamente de estimadores de árboles de decisión.
- Permite superar el problema de Bagging de estimadores altamente correlacionados.
- Incluye un componente extra de aleatoriedad, en el que para cada nodo de un árbol de decisión se selecciona en forma aleatoria un subconjunto de atributos que participarán en la selección.

# Random Forest

## Definición



# Random Forest

## *Ventajas y desventajas*

### Ventajas

- Alto desempeño (mejor que Bagging)
- Al igual que Bagging se genera durante el entrenamiento una muestra *Out Of Bag*.
- Robusto frente a outliers.
- Entrega información respecto a la importancia de los atributos.
- En general registra bajo sobreajuste

### Desventajas

- Complejo de interpretar
- Bajo rendimiento en problemas lineales

# ¡Manos a la obra! Comparando los modelos



# ¡Manos a la obra!

## *Predicción de precios de vivienda*

Veremos a continuación la implementación de estos modelos con Python, para lo que trabajaremos en Colaboratory. Abre el archivo destinado para esto y sigue las instrucciones de tu profesor.



# Desafío - Modelos de ensamble (parte I)





# Desafío

## *“Modelos de ensamble (parte I)”*

- ¿Leíste el desafío de esta semana? ¿Comprendes bien lo que se solicita en cada caso?
- ¿Hay contenidos que necesitas repasar antes de comenzar este desafío?
- ¿Necesitas algún ejemplo o indicación para alguna pregunta o requerimiento específico?



**“La suma de las partes  
entrega como resultado un  
valor mayor que el todo”  
\*Diversidad\***



**{desafío}**  
**latam\_**

*Academia de  
talentos digitales*

