



Modelamiento

Clase Sincrónica

Implementar modelos de aprendizaje automático por medio de técnicas estadísticas, adecuando los diferentes algoritmos debidamente a la situación y requerimientos necesarios

- Unidad 1: Introducción al Machine Learning
- Unidad 2: Aprendizaje Supervisado y No Supervisado
(Parte I: No supervisado)
(Parte II: Clasificación)
(Parte III: Clasificación)
(Parte IV: Regresión)
(Parte V: Series de tiempo)
- Unidad 3: Aplicando lo aprendido
(Parte I: Preprocesamiento de datos)
(Parte II: Modelamiento)



Te encuentras aquí



¿Qué aprenderás en esta sesión?

En esta sesión aprenderás se finalizará el repaso del proceso de un proyecto de data science, poniendo en especial foco en la parte de modelamiento y entrega del proyecto.

¿Para qué se utilizan los frameworks? ¿Conoces alguno para proyectos de data science?



¿Cómo podemos identificar Outliers en datos?



¿Por qué es importante
detectar correlaciones?
¿Qué métodos conoces
para hacerlo?



¿Que es la maldición de
la dimensionalidad?
¿Cómo podemos
trabajarla?

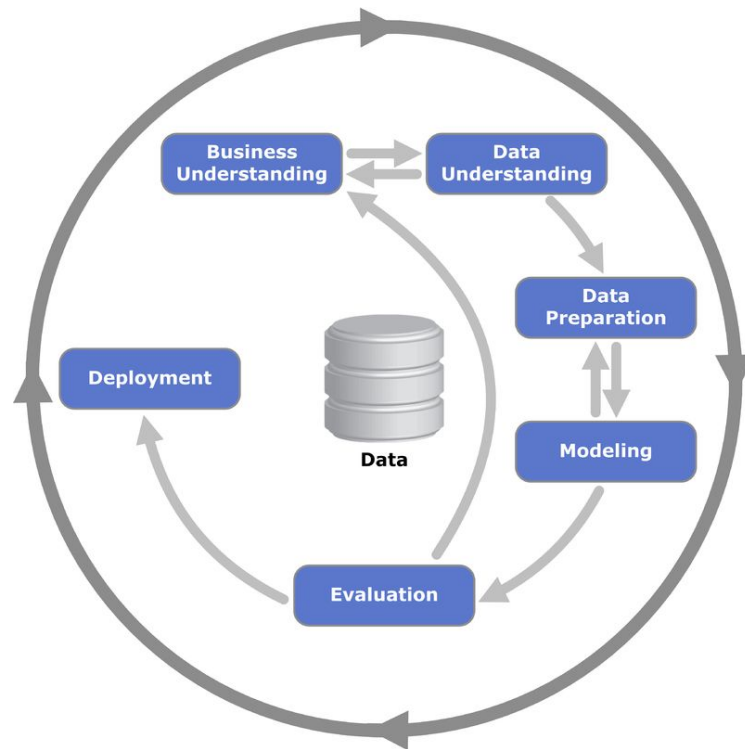


/* Retomando CRISP-DM */

CRISP DM

Metodologías para Data Science

1. Entendimiento del Negocio
2. Entendimiento de la Data
3. Preparación de la Data
4. Modelamiento
5. Evaluación
6. Despliegue



Modelamiento

CRISP DM

Identificar Solución

A partir de las etapas previas, principalmente el entendimiento del negocio. Identificar que tipo de problema se está resolviendo (regresión, clasificación, clustering, etc)

Estudio Bibliográfico

Buscar problemas similares y soluciones propuestas e implementadas en otros lugares.

Propuesta de Soluciones a Probar

En base a las etapas previas y al estudio se definen al menos 3 formas posibles de abarcar el problema. Balanceo, modelos a utilizar, tipo de modelamiento, Diferentes capas, etc.

Modelar

Implementar las diferentes propuestas de soluciones descritas anteriormente y guardar tanto los modelos o soluciones.

Optimización del modelo

Refinar el modelo y optimización de los hiperparametros.

¿Cómo se realiza un estudio bibliográfico?



Estudio bibliográfico

¿En qué consiste?

Proceso de investigación que implica la revisión crítica y exhaustiva de las publicaciones académicas, libros, artículos técnicos y recursos en línea relacionados con el aprendizaje automático y la inteligencia artificial. Algunas fuentes útiles para realizar esto son las siguientes:

- <https://scholar.google.com/>
- <https://arxiv.org/>
- <https://towardsdatascience.com/>
- <https://medium.com/>
- <https://github.com/>
- Otros varios

¿Cómo realizar la
optimización de hiper
parámetros de los modelos?



Optimización de hiper parámetros

Algunos métodos

Durante el módulo hemos visto un método de optimización de hiperparametros ampliamente utilizado llamado **GRID SEARCH**.

Existen varios métodos y librerías que se pueden utilizar acá te mencionamos algunas otras:

1. Random Search
2. Algoritmos evolutivos
3. Otros varios

Evaluación

CRISP DM

Definir un proceso de evaluación

Decidir cómo dividir el dataset o de qué forma se va a evaluar el modelo y contra qué se va a contrastar.

Elegir métricas de evaluación.

Cálculo de métricas

Dado el proceso de evaluación calcular las métricas escogidas anteriormente.

Explicabilidad del modelo.

Encontrar la relación entre las variables de entrada y la salida del modelo, como se puede interpretar lo que dice el modelo.

Validación

A partir de las métricas y la interpretación del modelo encontrar insights y validar con el negocio.

Entrega

Una vez definido se puede devolver a las etapas anteriores o entregar la solución para ser desplegada posteriormente.

¿Cuál es la diferencia entre
interpretabilidad y
explicabilidad?



Explicabilidad vs Interpretabilidad

Diferencias

Interpretabilidad

Es una característica inherente de los modelos utilizados. Por ejemplo, los árboles de decisión o la regresión lineal son modelos interpretables.

Explicabilidad

Es un proceso que se realiza posterior al modelamiento y sirve para explicar de qué forma toman las decisiones los algoritmos.

Se utilizan comúnmente para modelos más complejos. Algunos métodos para investigar son las dependencias parciales, lime, shap, etc.

/* Random Forest */

Random Forest

Modelos de ensamblaje

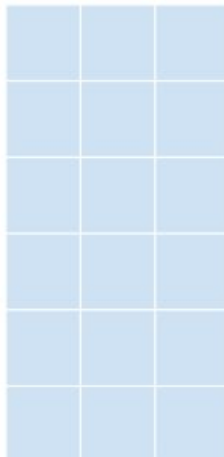
Técnica que combina múltiples modelos de aprendizaje automático (llamados "modelos base") para mejorar la capacidad de predicción y generalización en comparación con un solo modelo.

Existen varios tipos de modelos y de formas de ensamblaje. **Random Forest** es un algoritmo de ensamblaje en paralelo.

Random Forest

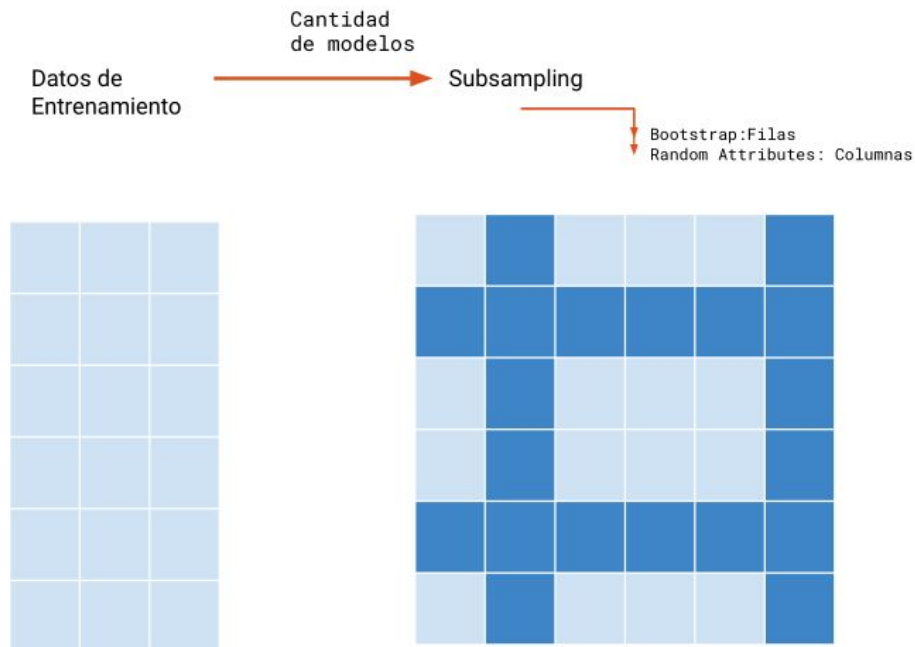
Mecanismo de acción

Datos de
Entrenamiento



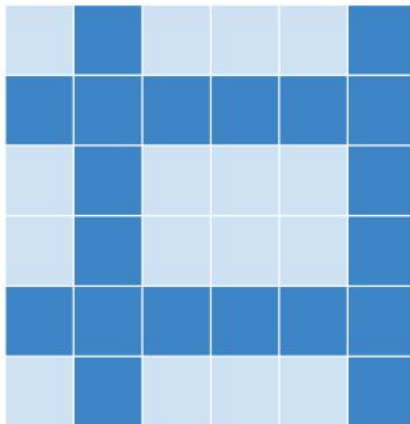
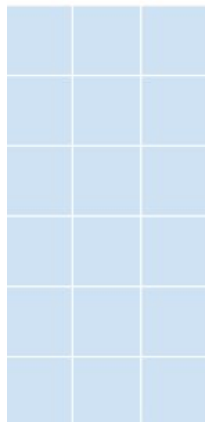
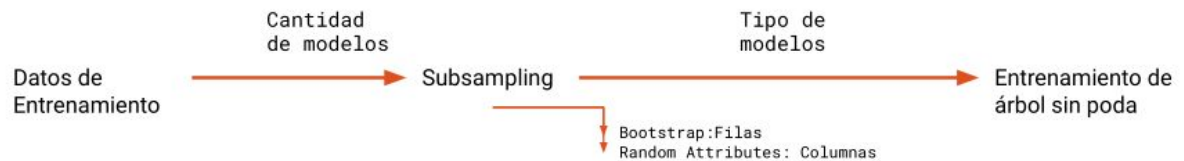
Random Forest

Mecanismo de acción



Random Forest

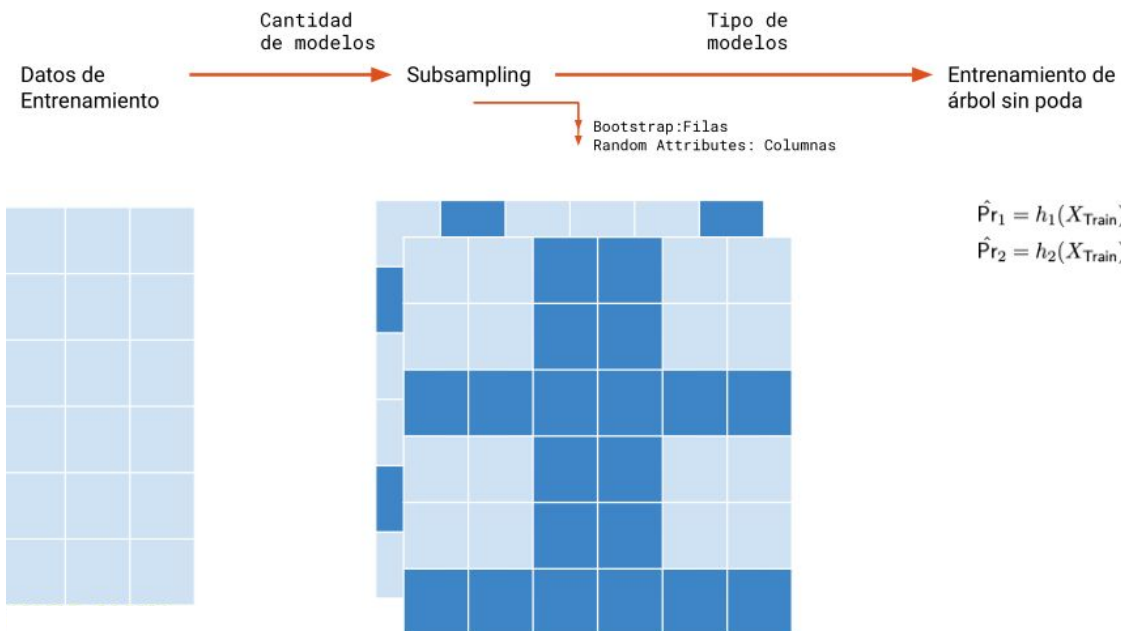
Mecanismo de acción



$$\hat{P}_{r_1} = h_1(X_{\text{Train}})$$

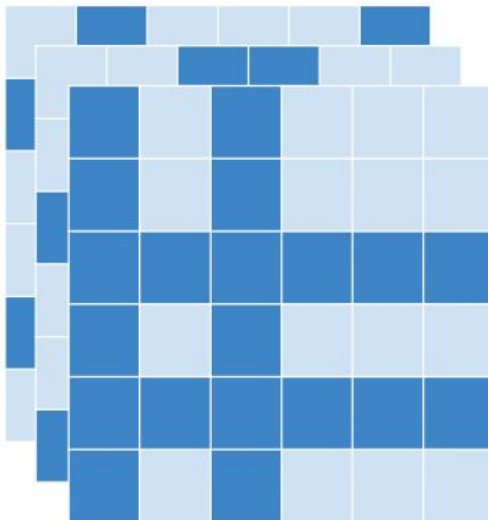
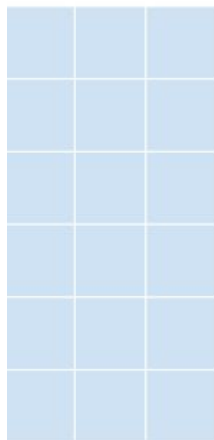
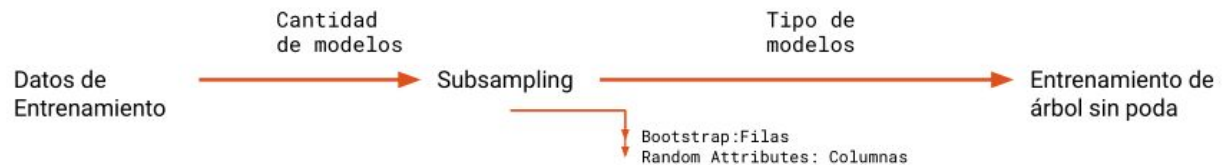
Random Forest

Mecanismo de acción



Random Forest

Mecanismo de acción



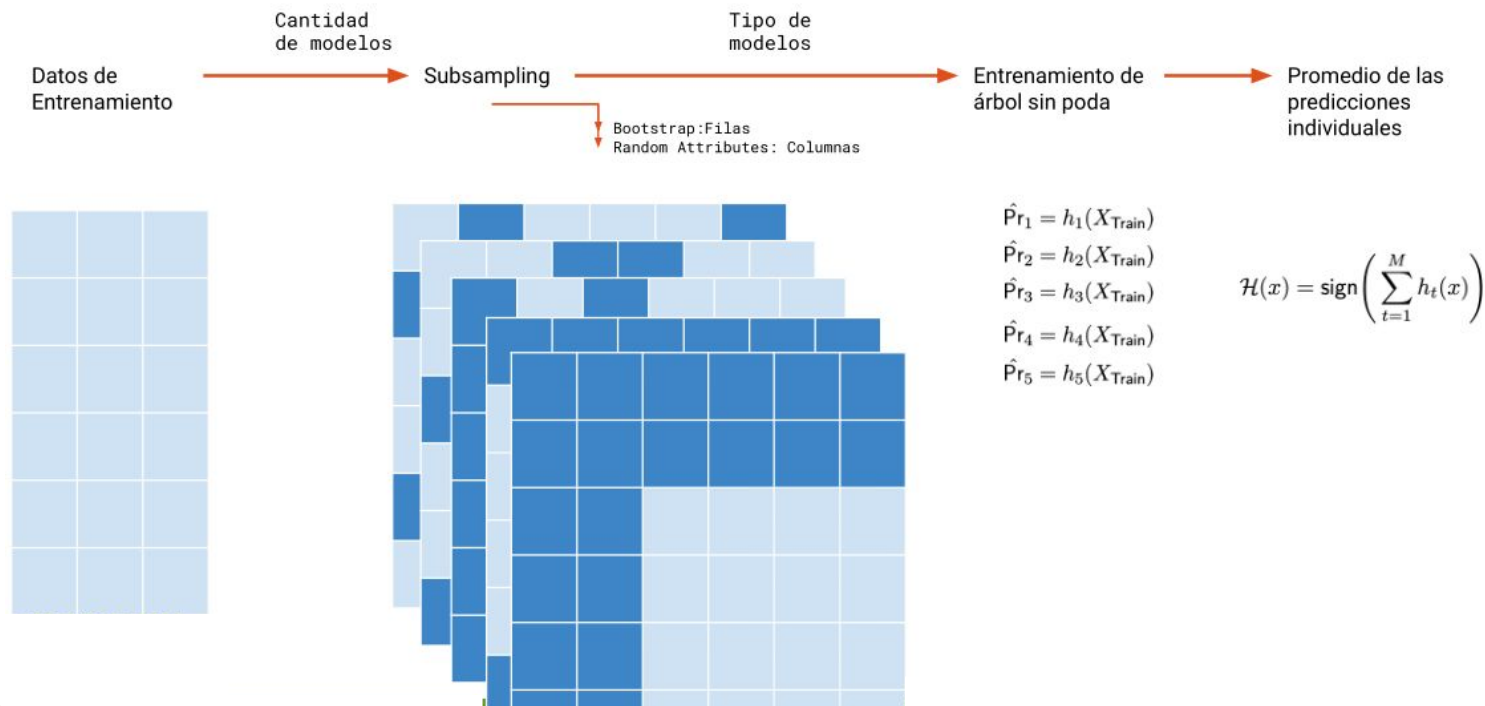
$$\hat{P}_{r_1} = h_1(X_{\text{Train}})$$

$$\hat{P}_{r_2} = h_2(X_{\text{Train}})$$

$$\hat{P}_{r_3} = h_3(X_{\text{Train}})$$

Random Forest

Mecanismo de acción



Random Forest

Hiperparámetros

Random Forest tiene como “modelo base” los árboles de decisión. Por lo que los hiperparámetros son iguales que la de los árboles de decisión.

Se agrega a ellos el número de estimadores, es decir, la cantidad de modelos base que se utilizan.

¡Manos a la obra! Apliquemos un modelo Random Forest



¡Manos a la obra!

Apliquemos un modelo Random Forest

Veremos ahora como aplicar un modelo Random Forest con Python, para lo que utilizaremos la base de datos Iris que hemos utilizando anteriormente.

Observa los pasos que te mostrará tu profesor, y puedes replicarlos también en tu propio archivo de Jupyter Notebook



Prueba - Predicción del precio de vuelos



Prueba

"Predicción del precio de vuelos"

- Descarga el archivo "Predicción del precio de vuelos".
- Tiempo de desarrollo asincrónico: desde 4 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪



Ideas fuerza



El **modelamiento** es una etapa iterativa donde se realiza el trabajo final que entrega los resultados.



La etapa de **Evaluación** es crucial para el proyecto. En la cual se tiene que entregar si los resultados son útiles para el problema o no.



Los **algoritmos de ensamblaje** son formas de utilizar algoritmos base para entregar un mayor poder predictivo. Por ejemplo **random forest**.

¿Qué conceptos no te
quedaron claros o quieres
reforzar?



Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio
- Desafío “Nombre de la puerta”



{desafío}
latam_

*Academia de
talentos digitales*

