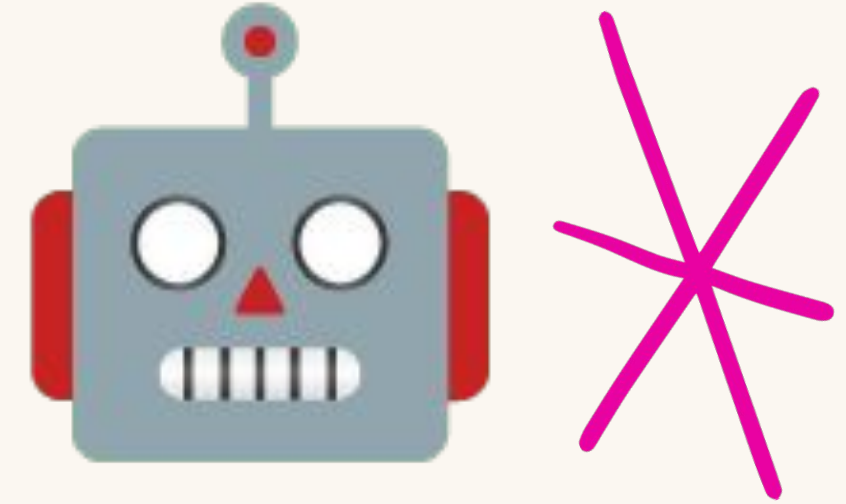


Desbloqueando el Poder de los Datos



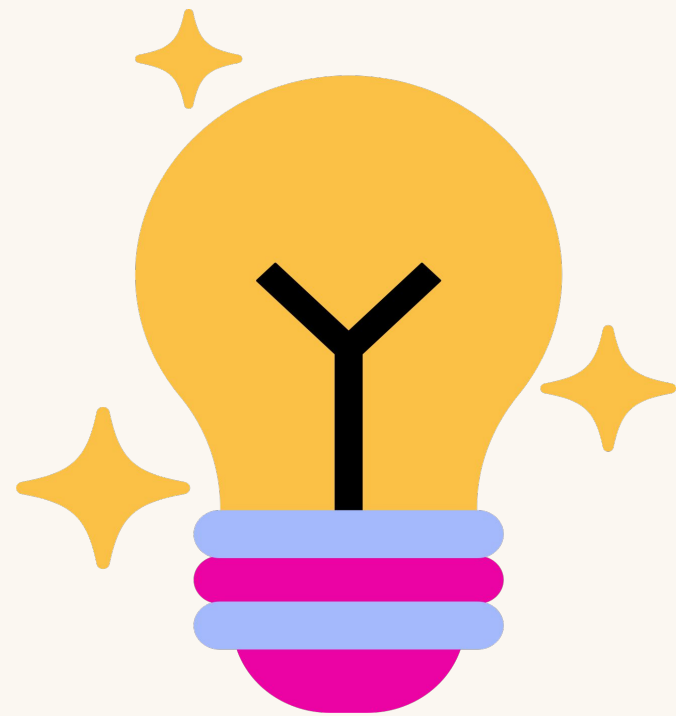
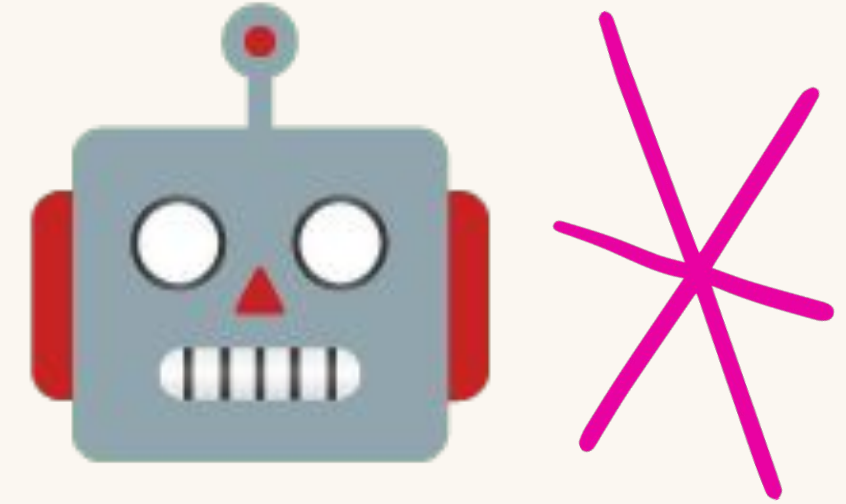
Inteligencia Artificial & Ciencia de Datos para todos

Descanzo. Regresamos a las: 8:05 a.m.

¿Te gustaría comenzar el día con alguna canción en específico?

Coméntala en el chat 🎵💬

Desbloqueando el Poder de los Datos



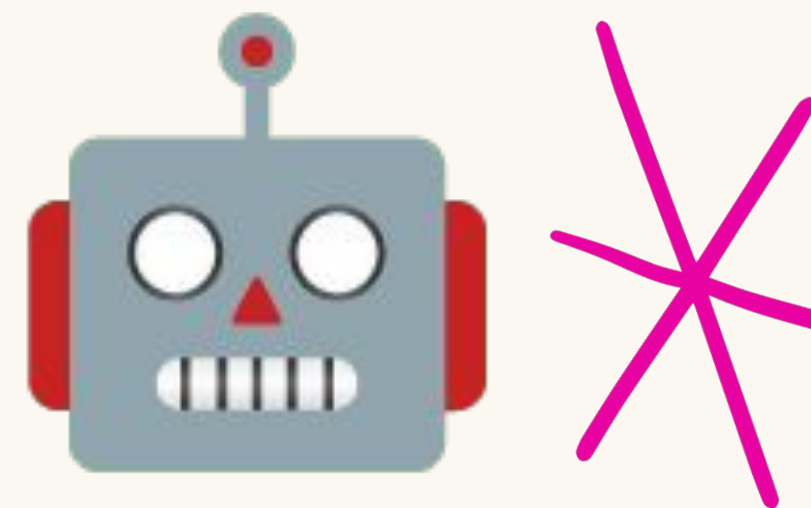
Inteligencia Artificial & Ciencia de Datos para todos

Comenzamos a las 7:05 a.m. en punto.

¿Te gustaría comenzar el día con alguna canción en específico?

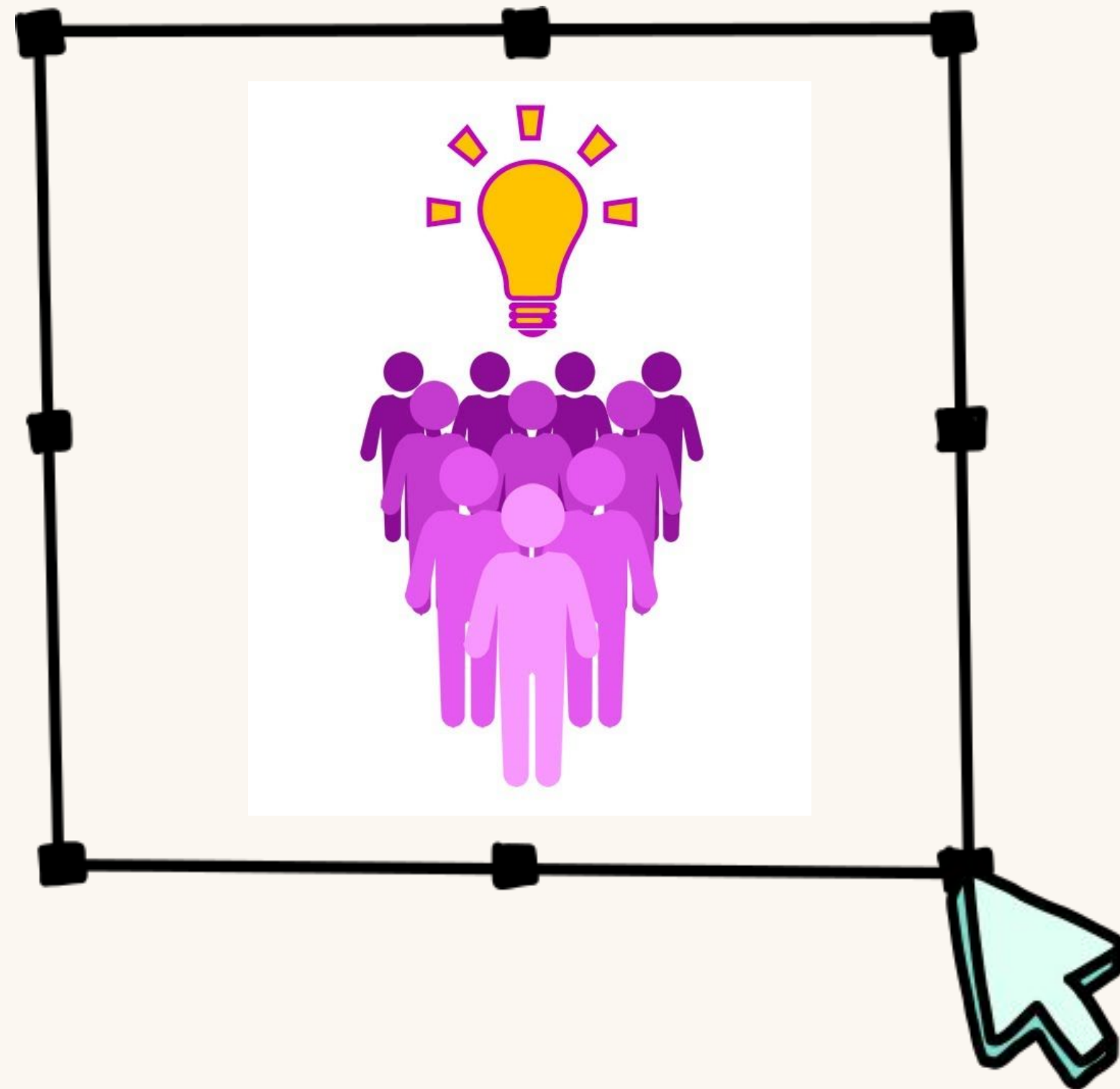
Coméntala en el chat 🎵💬

Desbloqueando el Poder de los Datos



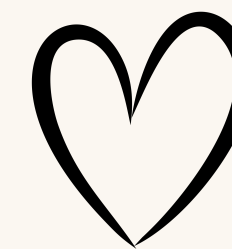
Inteligencia Artificial
& Ciencia de Datos para todos





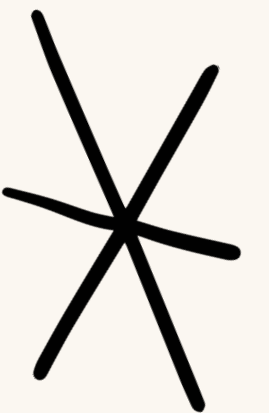
Modelos de Ensemble

Octubre 22, 2024





1. Solución del taller
2. Qué se viene para el tercer corte
3. Tema de hoy:
 - Modelos de Ensamble
 - La sabiduría de multitudes
 - Técnicas de ensamble
 - Modelos más populares
 - Random Forest
 - Gradient Boosting



Temas tercer corte:

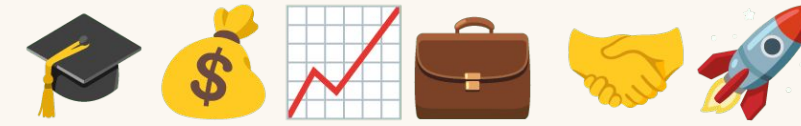
Clase #	Fecha	Tema
Clase #11	Octubre 15, 2024	Modelos NO supervisados: K-Medias
Clase #12	Octubre 22, 2024	Modelos de Ensemble: Random Forest y GB
Clase #13	Octubre 29, 2024	Comunicación de negocios y desarrollo profesional
Clase #14	Noviembre 5, 2024	Introducción a redes neuronales
Clase #15	Noviembre 12, 2024	Presentaciones finales
Clase #16	Noviembre 19, 2024	Presentaciones finales

Página web del curso con videos, diapositivas, código y más:
<https://github.com/vivianamarquez/unicomfauca-ai-2024>

Talleres tercer corte:

Taller #	Descripción	Mandatorio	Se aceptan entregas tardes	Enlace	Fecha de entrega	Porcentaje en el tercer corte	Porcentaje en el curso
Taller # 11	K-Medias	Sí	Sí	Diapositiva 59	Octubre 21, 2024	10%	4%
Taller # 12	Random Forest	Sí	Sí	Diapositiva 31	Octubre 28, 2024	10%	4%
Encuesta	Responder a la pregunta: ¿Qué te gustaría saber/aprender en la clase de desarrollo profesional?	Sí	NO	Enlace	Octubre 28, 2024	5%	2%
Excel Grupos	Decidir si van a trabajar individualmente o en parejas para el proyecto final, escoger proyecto y seleccionar datos	Sí	Preferiblemente no, pero se acepta hasta Nov 4	Enlace	Octubre 28, 2024	5%	2%
Primera Entrega Proyecto Final	Entrega de diapositivas y código del proyecto final	Sí	NO		Noviembre 4, 2024	15%	6%
Segunda Entrega Proyecto Final	Entrega de diapositivas y código del proyecto final incorporando la retroalimentación	Sí	NO		Noviembre 11, 2024	25%	10%
Presentación	Presentación final	Sí	NO (Estar presente en las dos sesiones)		Noviembre 12 y 19	30%	12%

Sobre el proyecto final



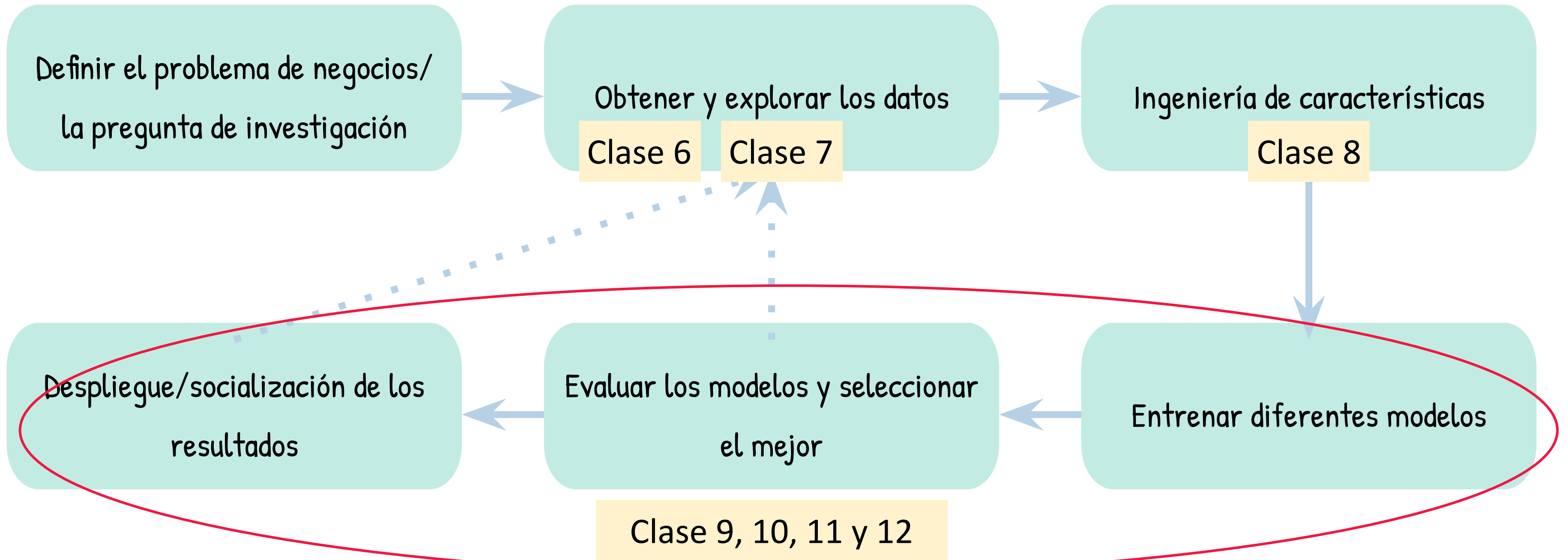
- El objetivo del proyecto final es que ustedes tengan un proyecto que puedan incluir en su portafolio profesional para demostrar sus habilidades técnicas en inteligencia artificial y comunicación de negocios.
- Respecto al nivel de dificultad del código, debe sentirse parecido a los talleres que hemos ido entregando durante el semestre. La diferencia es que ahora van a tener la oportunidad de escoger un tema que les guste y sea relevante para su vida profesional.

Inteligencia Artificial para todos

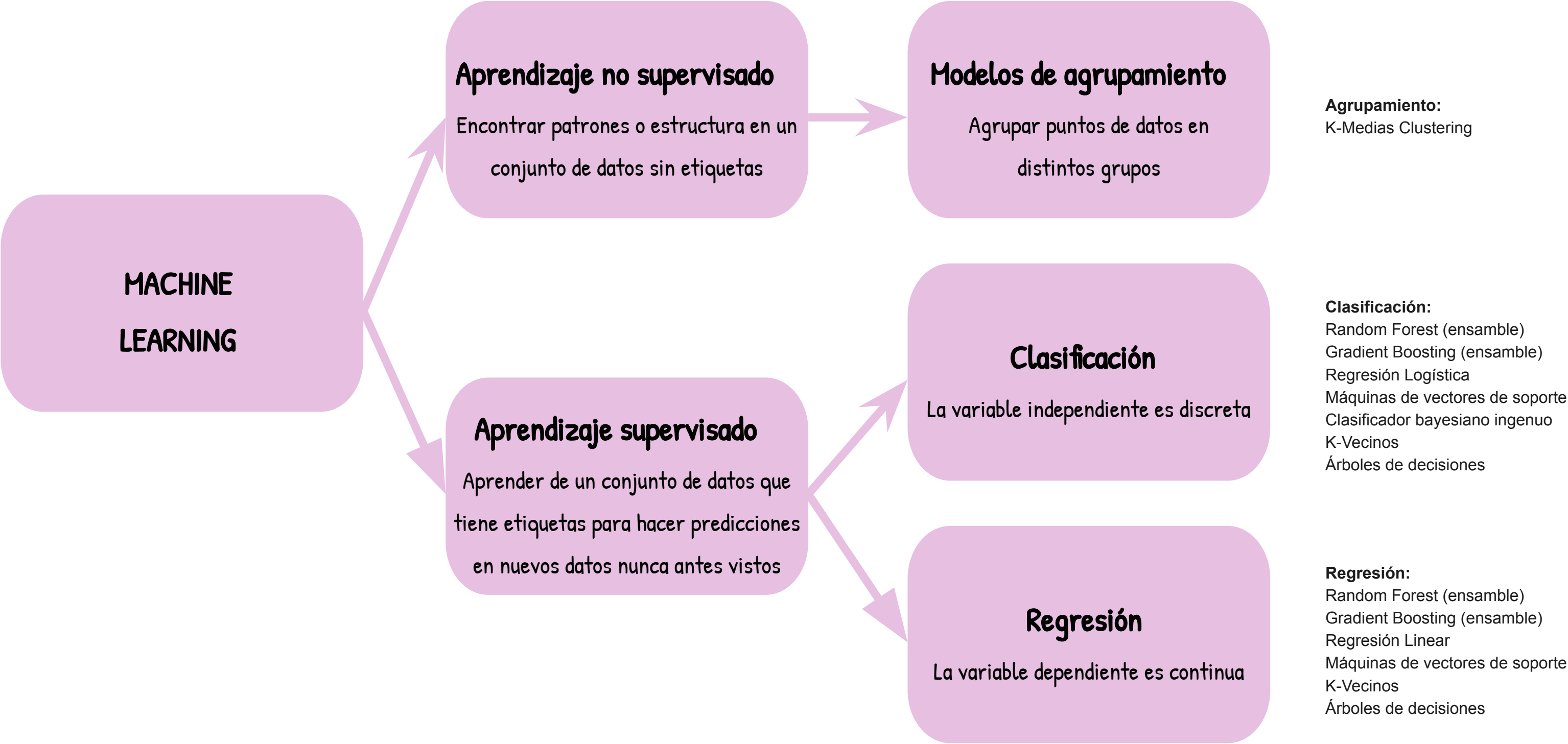
El proyecto es de su libre elección pero debe utilizar técnicas de modelado que aprendimos en el curso. Aquí hay algunas ideas:

- Para los **contadores públicos** podrían desarrollar un modelo para clasificar transacciones fraudulentas. Otro ejemplo sería un modelo para predecir si una empresa entrará en bancarrota. O un modelo que clasifique las transacciones financieras en categorías como ingresos, gastos, impuestos, etc.
- Para los **administradores de empresa** podrían desarrollar un modelo que prediga la probabilidad de que un empleado renuncie o sea despedido. Otro ejemplo sería usar un modelo de agrupamiento para dividir a los clientes con características similares para diseñar campañas de mercadeo. O un modelo que determine cuál debe ser el precio de un producto.
- Para los **ingenieros mecatrónicos** podrían desarrollar un modelo para predecir en cuánto tiempo va a fallar una maquinaria. Otro ejemplo sería un modelo que reconozca las emociones de un texto.
- Para los **ingenieros de sistemas** podrían desarrollar un modelo para clasificar accesos no autorizados a una red. Otro ejemplo sería un modelo que clasifique los logs de los errores generados por sistemas informáticos.

Pasos en un proyecto de Machine Learning



- 1. ¿Tenemos etiquetas?
- 2. ¿De qué tipo son nuestras etiquetas?



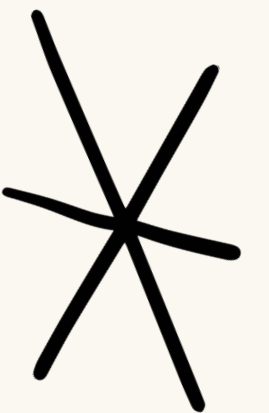
Entregas

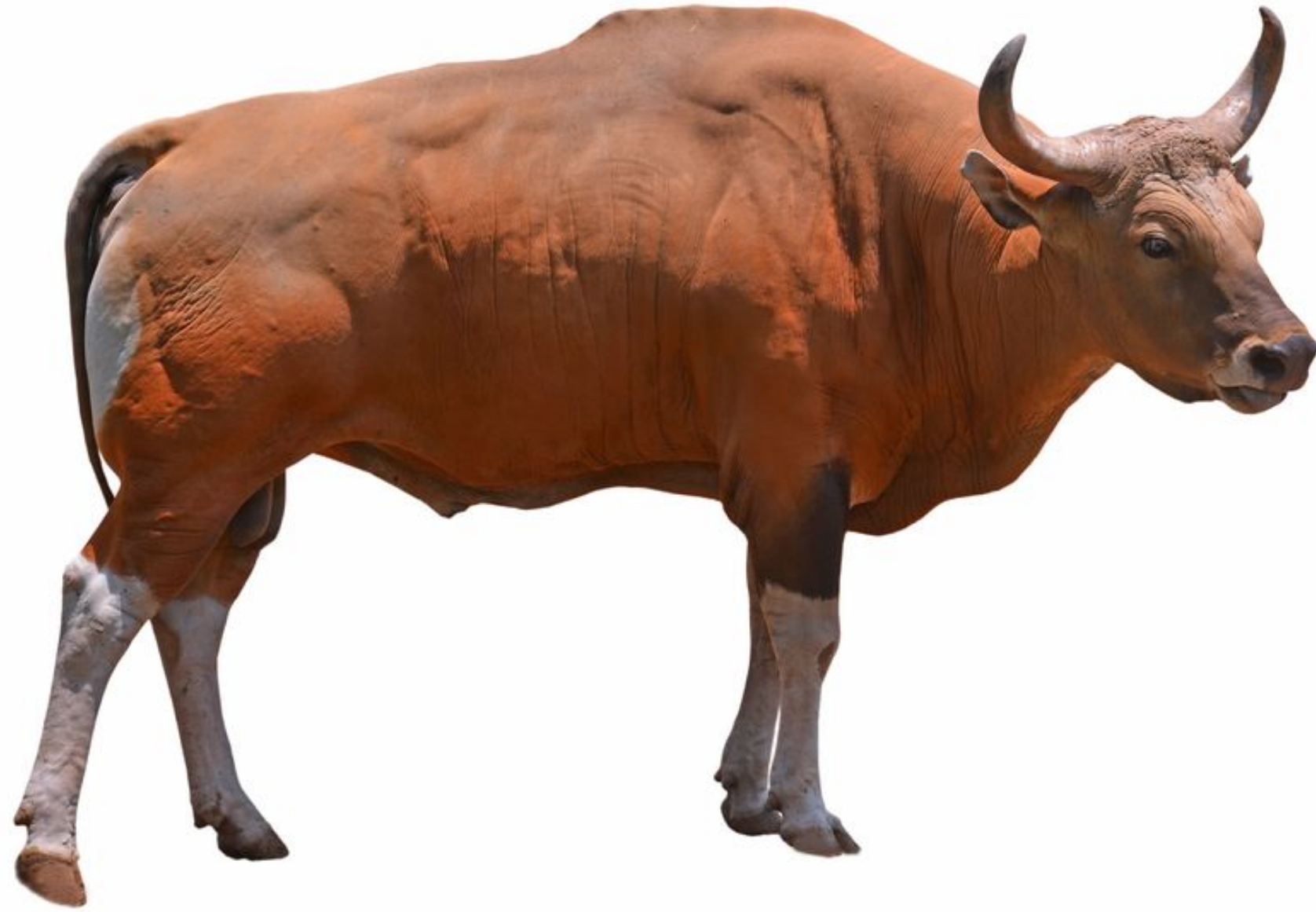
1. Paso #1 - Octubre 28 - [Enlace](#)
Escoger grupo, tema, tipo de modelo y conjunto de datos
2. Paso #2 - Noviembre 4
Primera entrega del proyecto final.
Este día deben entregar código y diapositivas.
Les haré retroalimentación que deben incorporar en la entrega final.
3. Paso #3 - Noviembre 11
Entrega final del proyecto final.
Van a entregar el mismo código y diapositivas, pero incluyendo la retroalimentación dada.
Estas serán las mismas diapositivas y código que se usará en la presentación final.
4. Paso #4 - Noviembre 12/19.
Presentación de su proyecto final.
Hay que asistir las dos fechas.

Las pautas exactas para la presentación, diapositivas y el código serán dadas la próxima clase.



1. Solución del taller ✓
2. Qué se viene para el tercer corte ✓
3. Tema de hoy:
 - Modelos de Ensamble
 - La sabiduría de multitudes
 - Técnicas de ensemble
 - Modelos más populares
 - Random Forest
 - Gradient Boosting





<https://www.ft.com/content/bfb7e6b8-d57b-11e1-af40-00144feabdc0>

Modelos de Ensamble

Imaginemos un agente inmobiliario en Popayán



- Un cliente se enamoró de un apartamento que el agente no tiene en su lista
- El agente puede hacer una estimación del precio del apartamento al usar otros apartamentos similares como referencia.
- Desventaja: Este método no es muy preciso para hacer predicciones
- Hmm... ¿podemos pensar en otra forma de estimar el precio del apartamento? 🤔

La sabiduría de la multitud



- Reclutemos varios agentes y cada uno va hacer una predicción del precio usando ciertos apartamentos como referencia y luego tomamos el promedio de estos valores.
- La predicción de un agente en específico puede que no sea muy buena: su predicción puede ser muy baja o muy alta
- Pero al tomar el promedio de las predicciones de todos los agentes, el resultado va a ser mejor que la predicción de un sólo agente en particular

La sabiduría de la multitud



- Reclutemos varios agentes y cada uno va hacer una predicción del precio usando ciertos apartamentos como referencia y luego tomamos el promedio de estos valores.
- La predicción de un agente en específico puede que no sea muy buena: su predicción puede ser muy baja o muy alta
- Pero al tomar el promedio de las predicciones de todos los agentes, el resultado va a ser mejor que la predicción de un sólo agente en particular

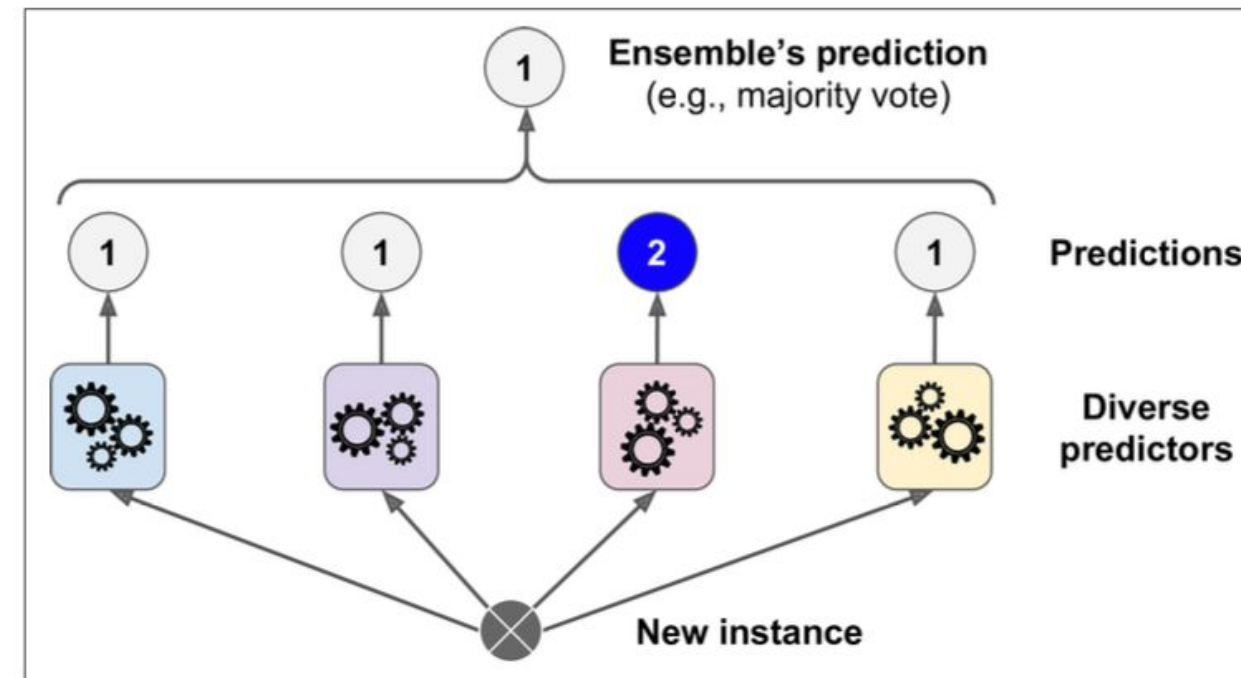
Modelos de Ensemble!



Modelos de Ensemble

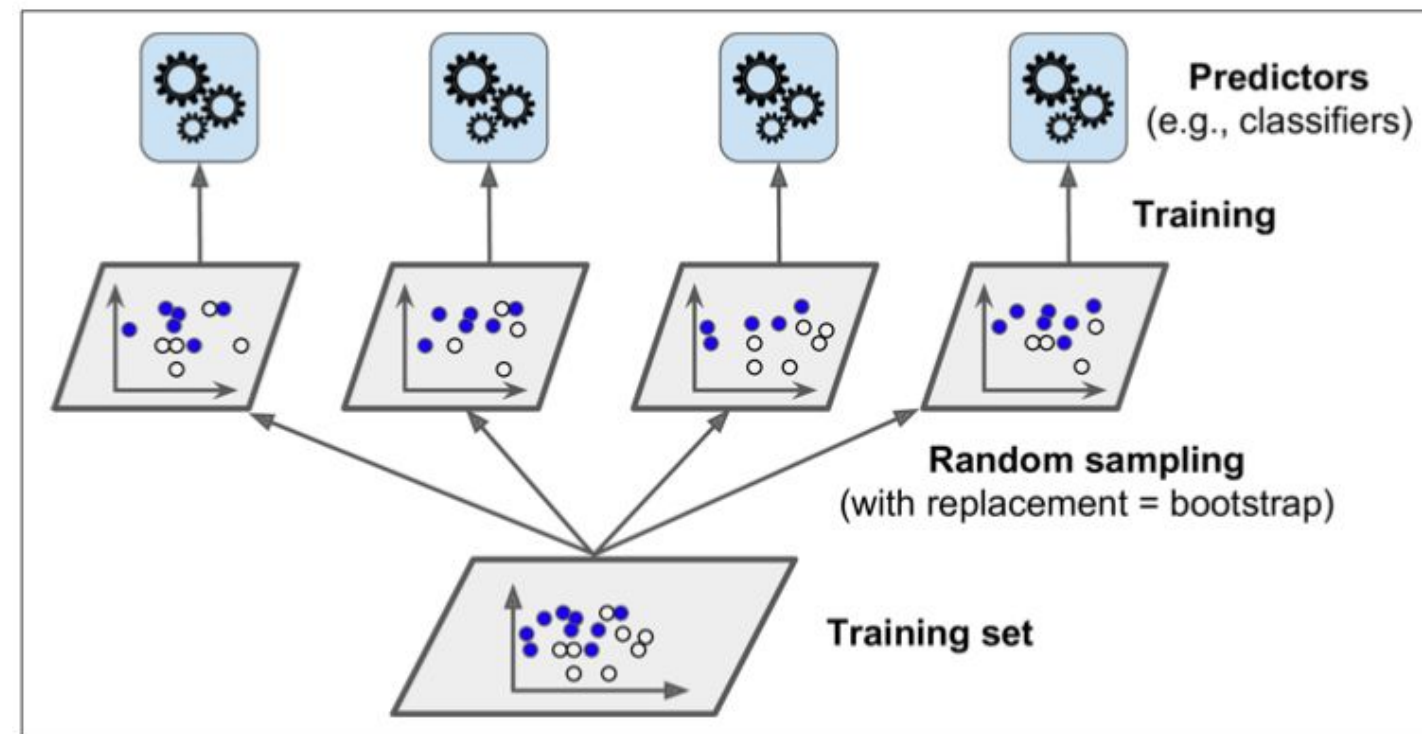
- Al juntar las predicciones de un grupo de predictores (como clasificadores o regresores), a menudo obtendrás mejores predicciones que con el mejor modelo individual.
- Los modelos de ensemble **nunca** aumentarán el sesgo o la varianza, y la mayoría de las veces lo reducirán.
- Tipos de modelos de ensemble:
 - Clasificadores de votación
 - Bagging y Pasting
 - Boosting
 - Stacking

Clasificadores de votación



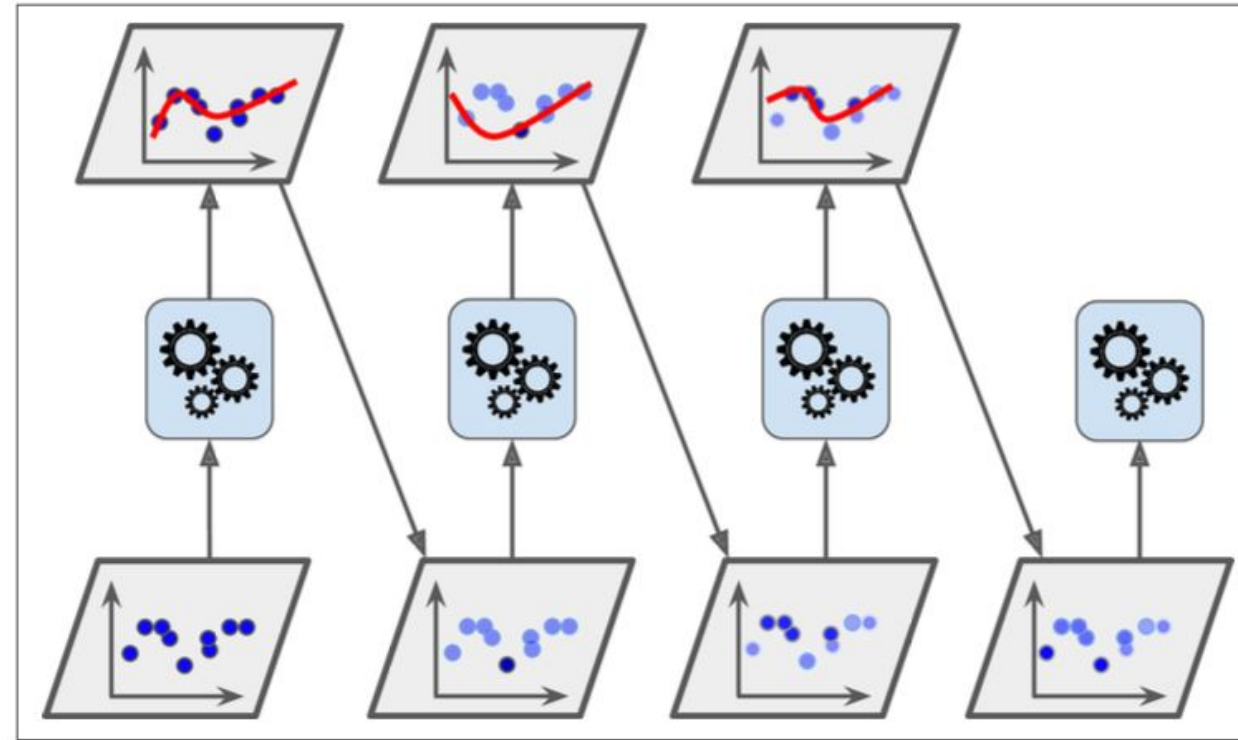
- Entrena algunos modelos y agrega sus predicciones: predice la clase que reciba más votos.
- Incluso si cada modelo es un modelo débil, el ensamble puede ser fuerte.
- Requisitos: Varios modelos débiles que sean diversos.

Bagging (agregación mediante bootstrap) y pasting



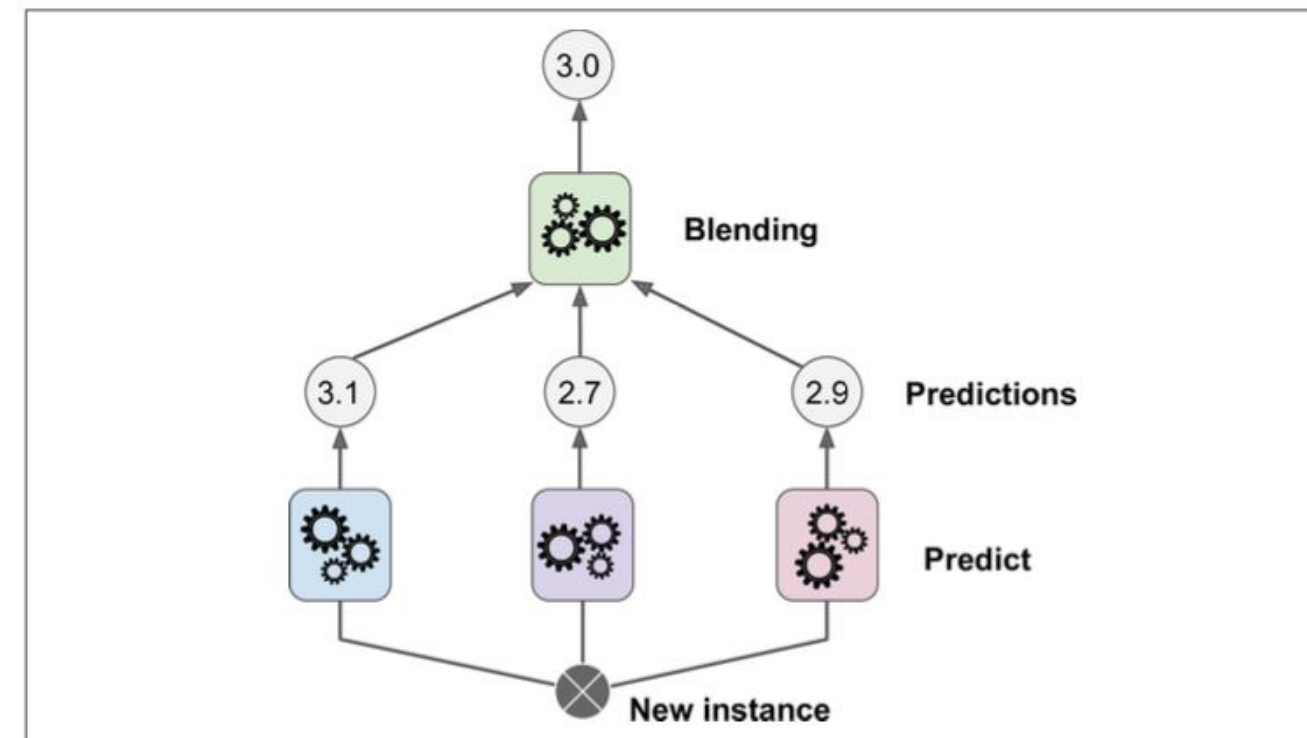
- Utiliza el mismo modelo, pero entrenándolo en diferentes subconjuntos aleatorios del conjunto de entrenamiento
- **Bagging:** El muestreo se realiza con reemplazo
- **Pasting:** El muestreo se realiza sin reemplazo

Boosting



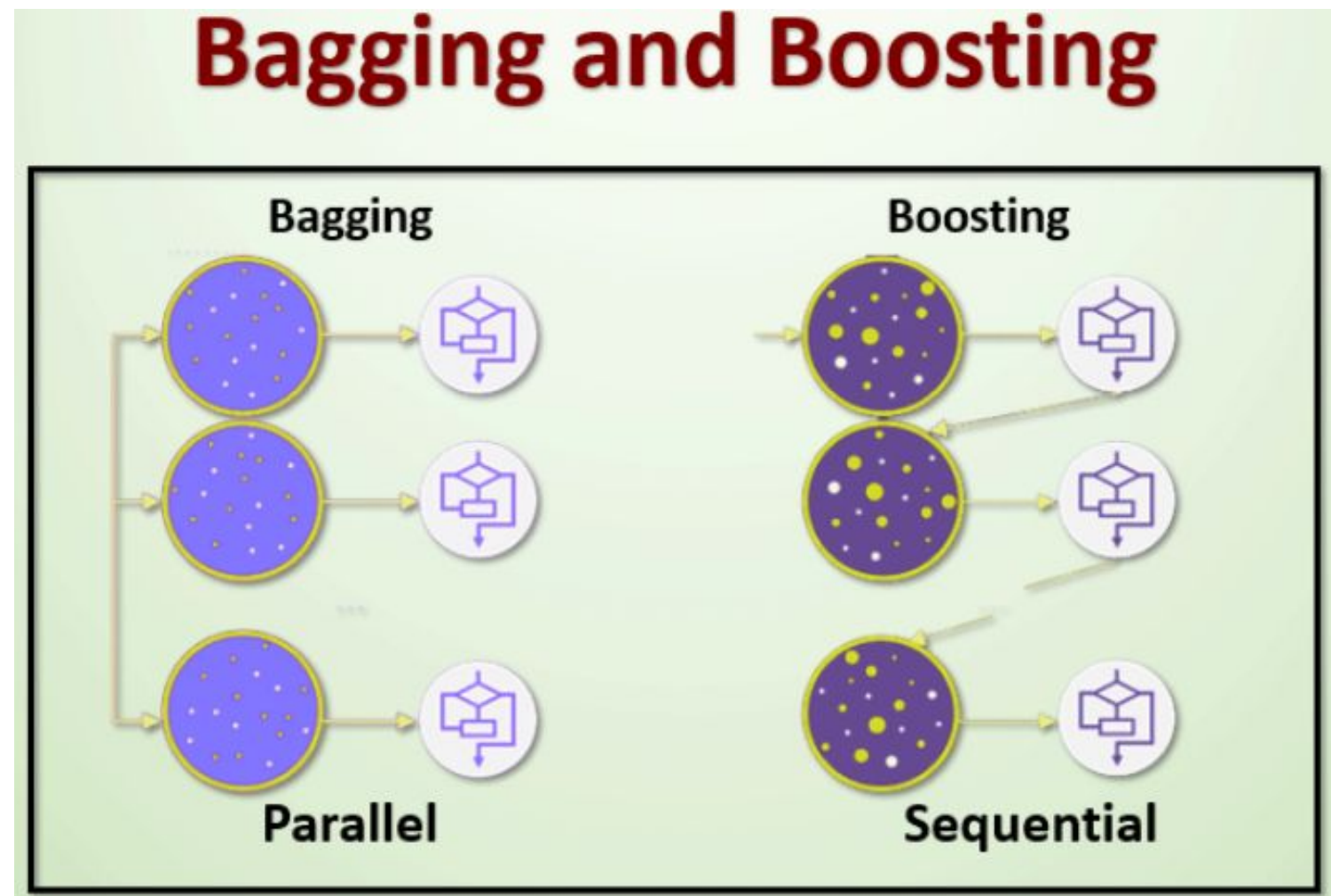
- Utiliza el mismo modelo, pero entrenandolo de manera secuencial, cada uno tratando de corregir a su predecesor
- Se da más peso a las instancias que fueron mal clasificadas en rondas anteriores

Stacking



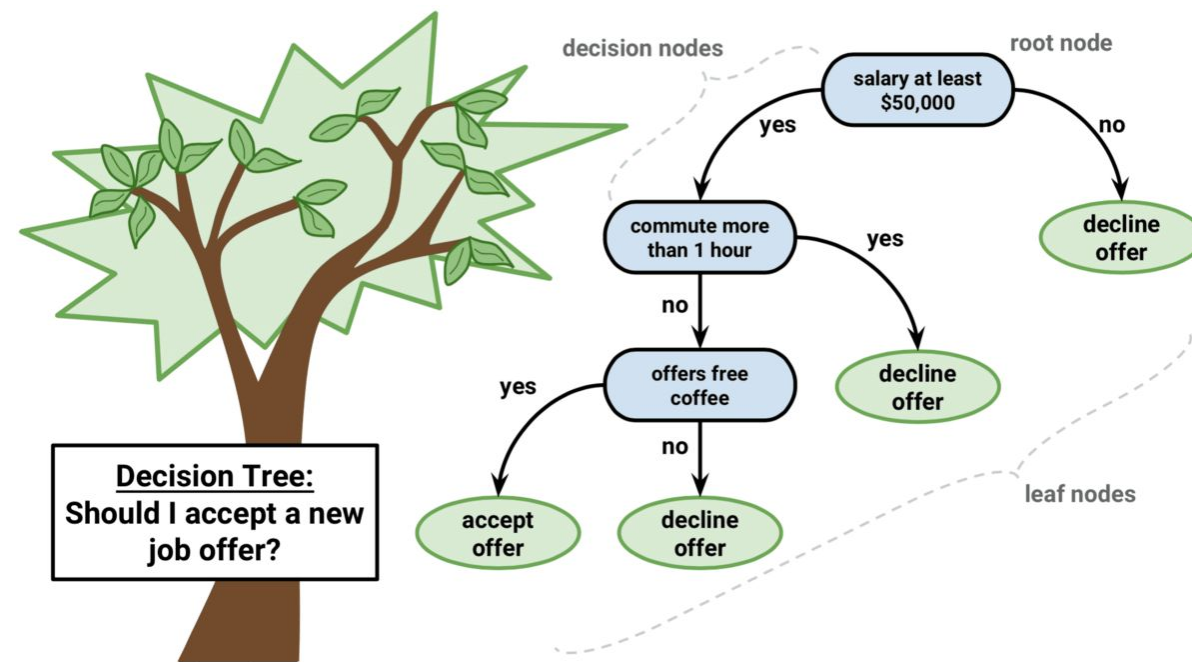
- En lugar de usar funciones triviales (como la votación mayoritaria) para juntar las predicciones de todos los modelos, ¿por qué no entrenamos un modelo para realizar esta agregación?

En resumen



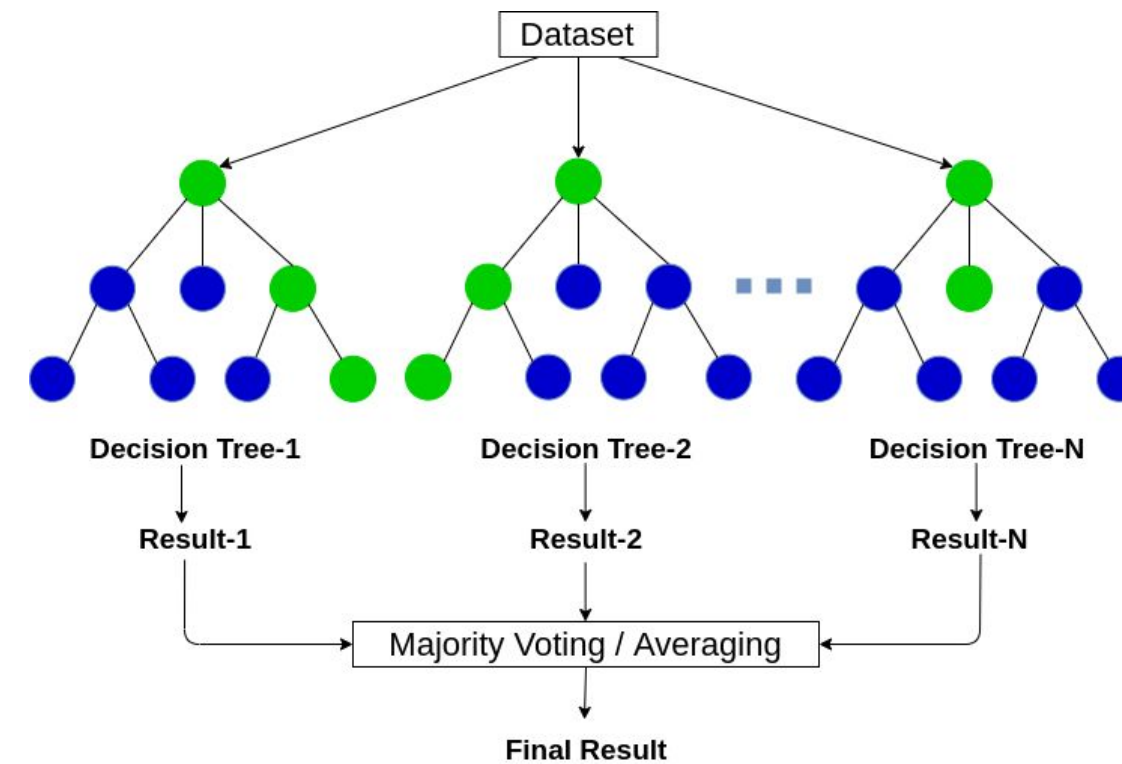
- La **votación** aprovecha modelos diversos y les permite "votar" sobre la mejor predicción.
- **Bagging** y **Pasting** utilizan el mismo modelo en diferentes muestras del conjunto de datos y agregan los resultados.
- **Boosting** construye modelos de manera iterativa para corregir los errores del modelo anterior.
- **Stacking** utiliza las predicciones de múltiples modelos como entrada para otro modelo que decide la predicción final.

Árboles de decisiones



- Los árboles de decisión son modelos de aprendizaje supervisado que se utilizan tanto para clasificación como para regresión.
- Se estructuran como un árbol con nodos de decisión y hojas que representan las predicciones finales.
- Los árboles de decisión son fáciles de interpretar y visualizar.
- Tienden a sobreajustarse si no se limitan la profundidad o el número de nodos.
- Son la base de los modelos de ensamble más populares: el Random Forest y Gradient Boosting

Random Forest (Bosque aleatorio)

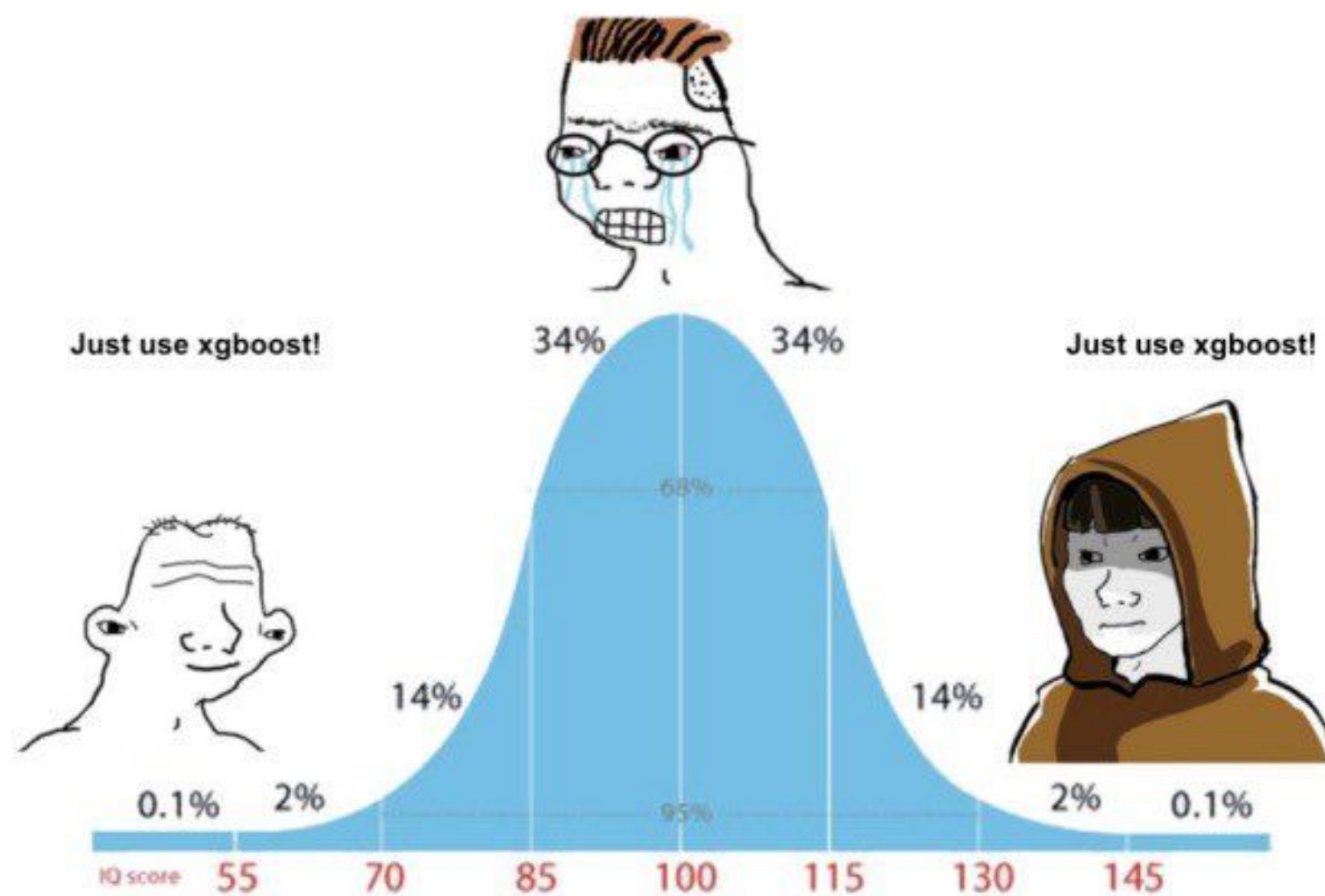


- Random Forest es modelo de ensamble que usa un conjunto de múltiples árboles de decisión con la técnica bagging
- Es un modelo muy robusto, por lo general tiene un mejor rendimiento que los modelos independientes

Gradient Boosting (potenciación por gradiente)

- Gradient Boosting es un método de ensamble que construye modelos de manera secuencial, donde cada nuevo modelo corrige los errores del modelo anterior.
- Es muy efectivo para problemas tanto de clasificación como de regresión.
- Su rendimiento es generalmente superior al de métodos más simples como Bagging o Random Forest, pero es más costoso computacionalmente.

NOOOOO you can't just throw a black box model at a problem just because you heard it does well in Kaggle. To build a robust model you must understand and visualize each of your features. You have to transform them, if appropriate. You have to have an understanding both of the statistical underpinnings of your model and of the underlying causal relationships in your data generating process. And that's just getting started, you-





Notebooks de hoy

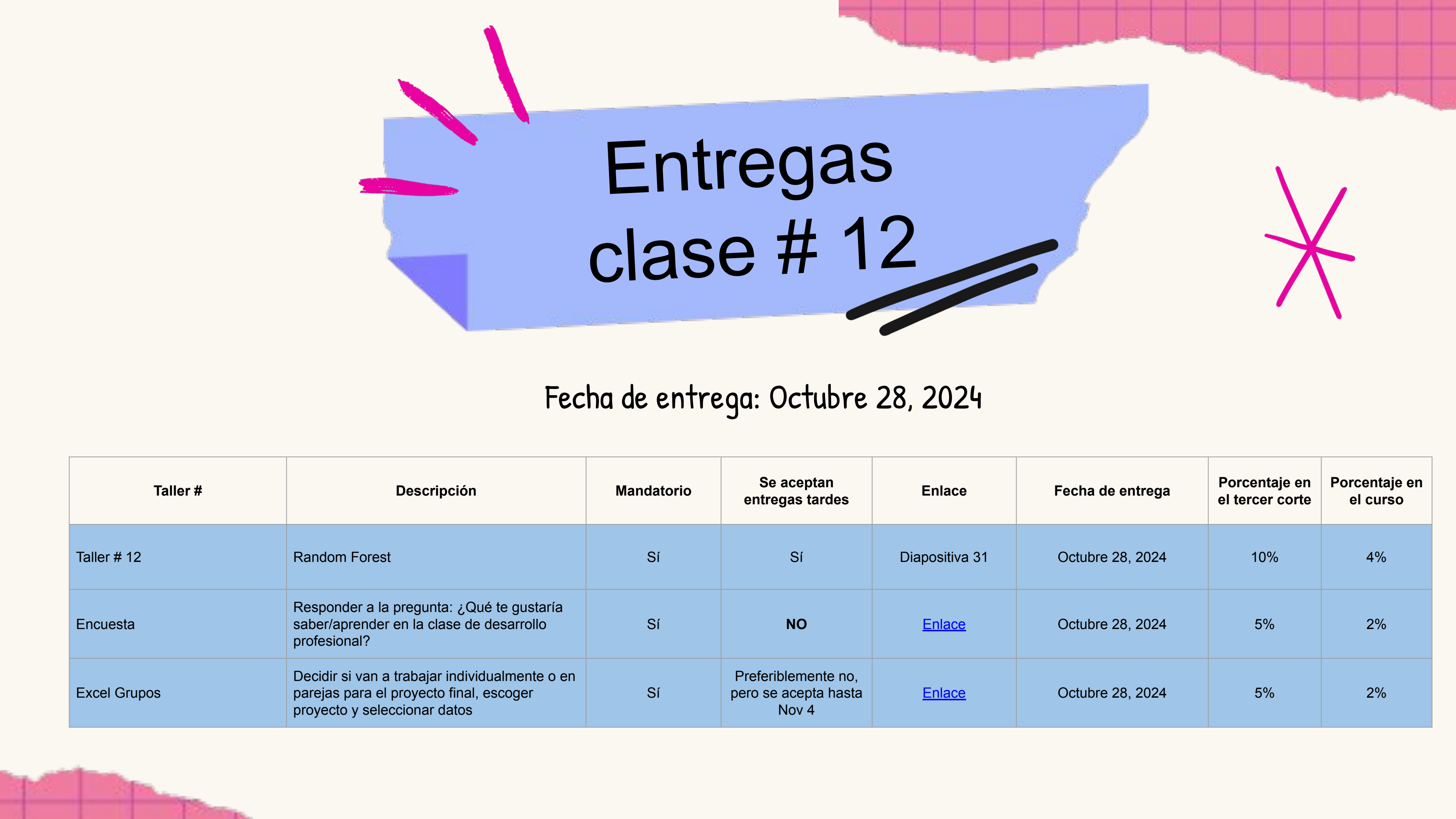
Regresión:

https://colab.research.google.com/drive/1pG-m3g_nP5-1pQeq62-y1c3d0mEdH4o1?usp=sharing

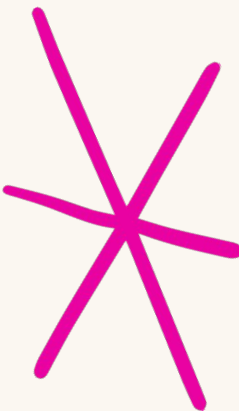
Clasificación:

<https://colab.research.google.com/drive/1lyAFXMux6VYeGuLkoXvvOKAWaermFuLc?usp=sharing>





Entregas clase # 12

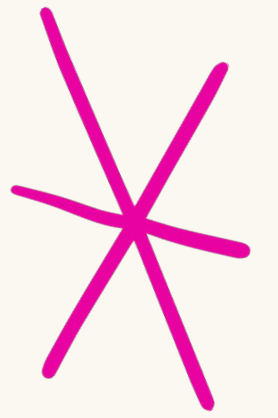


Fecha de entrega: Octubre 28, 2024

Taller #	Descripción	Mandatorio	Se aceptan entregas tardes	Enlace	Fecha de entrega	Porcentaje en el tercer corte	Porcentaje en el curso
Taller # 12	Random Forest	Sí	Sí	Diapositiva 31	Octubre 28, 2024	10%	4%
Encuesta	Responder a la pregunta: ¿Qué te gustaría saber/aprender en la clase de desarrollo profesional?	Sí	NO	Enlace	Octubre 28, 2024	5%	2%
Excel Grupos	Decidir si van a trabajar individualmente o en parejas para el proyecto final, escoger proyecto y seleccionar datos	Sí	Preferiblemente no, pero se acepta hasta Nov 4	Enlace	Octubre 28, 2024	5%	2%

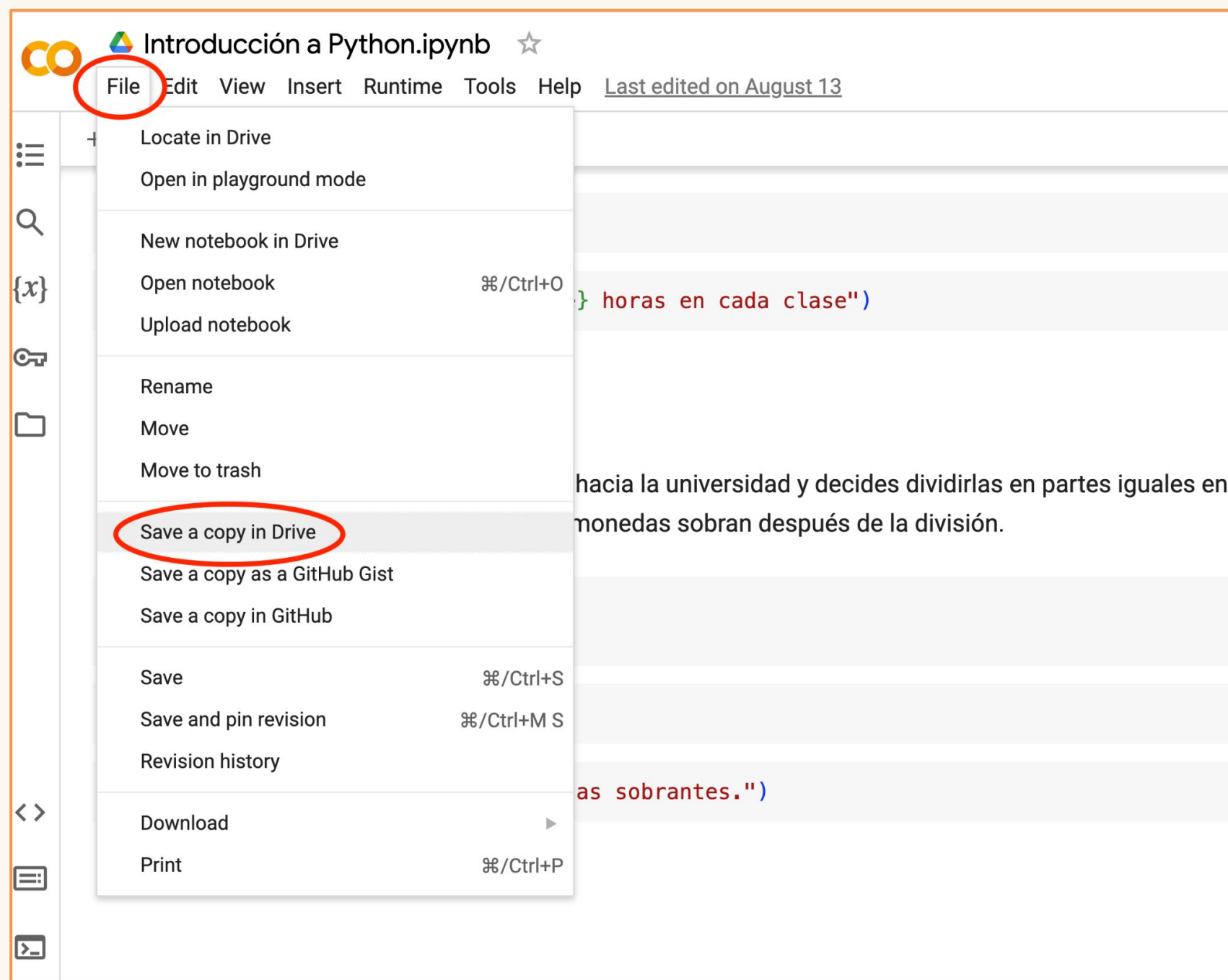
Sobre el taller # 12

- Escoger el [taller 9 \(regresión\)](#) ó el [taller 10 \(clasificación\)](#) y agregar el modelo simple "árbol de decisión" y los modelos de ensamble "Random Forest" y "Gradient Boosting".
- Obtén las métricas de rendimiento para todos los modelos y escoge el mejor modelo basadas en ellas. Una vez hayas escogido el mejor modelo, interpreta estas métricas de rendimiento y ofrece recomendaciones de negocio.
- Puedes usar como referencia los dos notebooks que usamos en clase ([modelos de regresión](#)) y ([modelos de clasificación](#)).

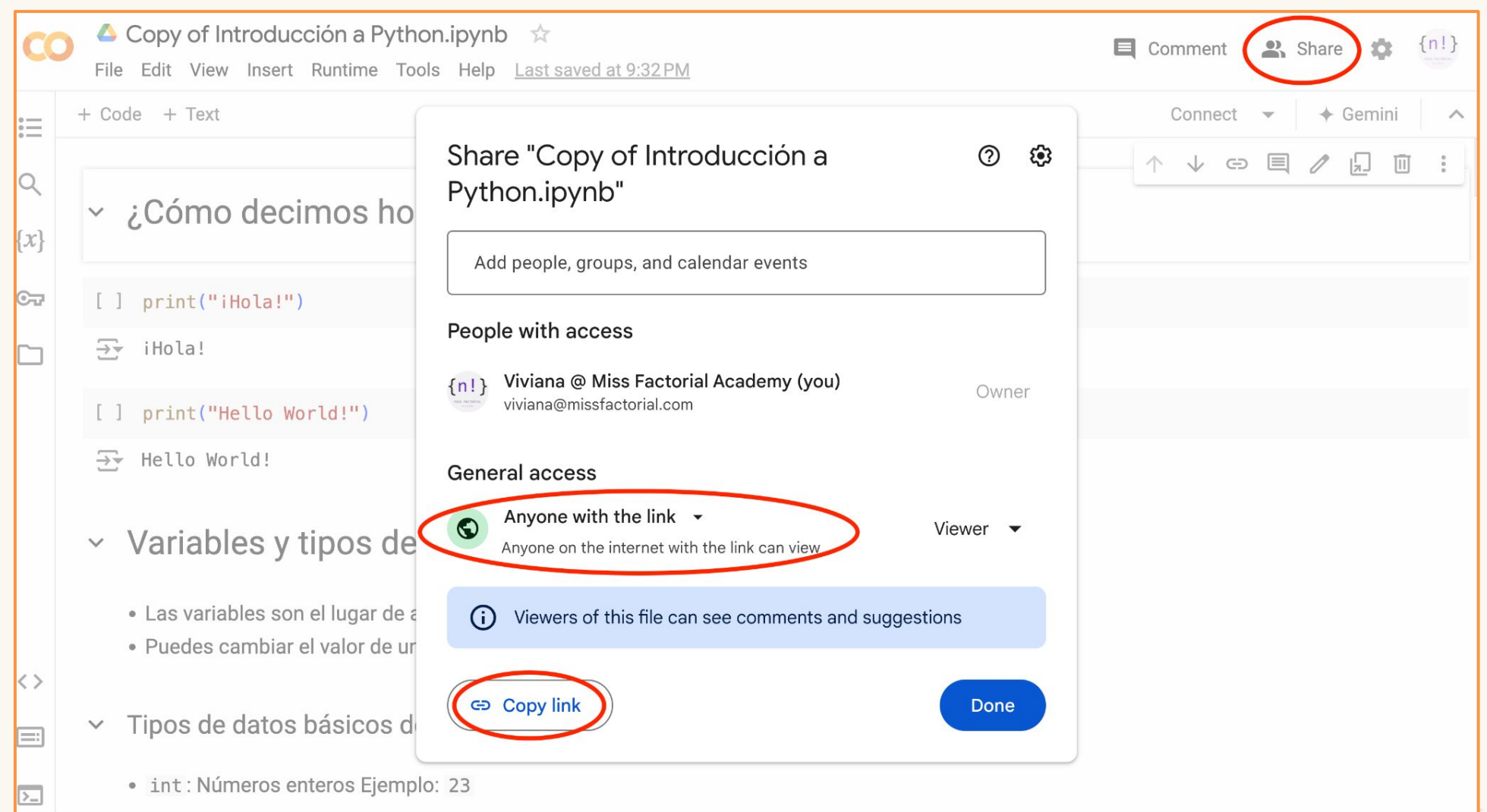


Para enviar los talleres de código

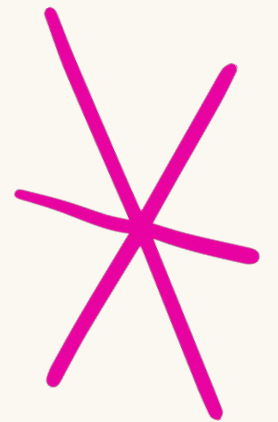
- ❑ Hacer click en **archivo** → **guardar copia en mi Drive** para que les quede una copia en su cuenta, de lo contrario, los resultados no serán guardados.
- ❑ En la copia creada, hacer click en **compartir**, asegurarse que el enlace sea visible a **cualquier persona**, copiar el enlace y enviarlo.



vroberta@unicomfaucauca.edu.co



¡Gracias!



¿Dudas? Email de la profe:

vroberta@unicomfauca.edu.co

Página web del curso con toda la info:

<https://github.com/vivianamarquez/unicomfauca-ai-2024>