

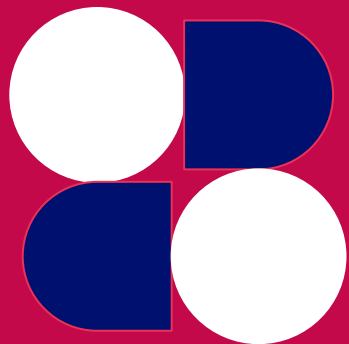


Programa PUCP DATA LITERACY

Formación tecnológica

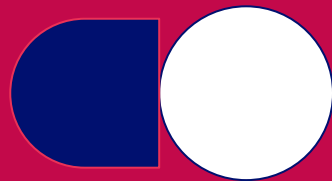
Dirección Académica de Planeamiento y Evaluación
Dirección de Gestión del Talento Humano





SESIÓN 8:

Aprendizaje supervisado





Contenido:



1

Introducción al Aprendizaje Automático y tipos de aprendizaje

2

Regresión vs Clasificación

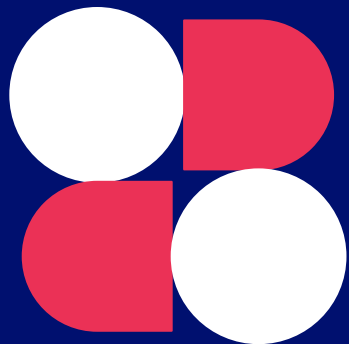
3

Entrenamiento y evaluación de modelos

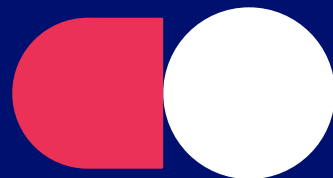


Al finalizar la sesión, lograrás:

- Comprender el concepto de aprendizaje supervisado
- Aprender sobre los modelos de regresión y de clasificación, similitudes y diferencias
- Entender el proceso mediante el cual se evalúan los modelos de regresión y de clasificación



Introducción al Aprendizaje Automático y tipos de aprendizaje





¿Qué es aprendizaje automático o machine learning?



Aprendizaje automático o ML



Machine Learning (ML) es una rama de la **inteligencia artificial** que permite a las máquinas aprender de los datos sin ser programadas explícitamente. La idea es que las **máquinas** sean capaces de **aprender patrones** y hacer predicciones a partir de los datos.





Tipos de Aprendizaje



Aprendizaje
Supervisado

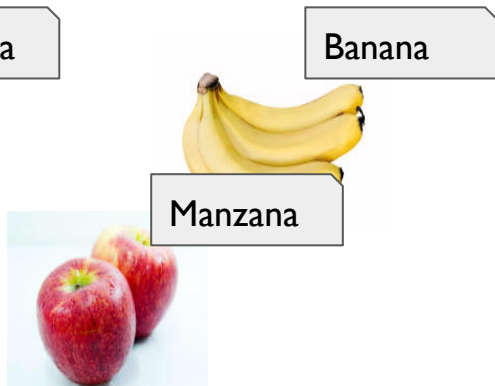
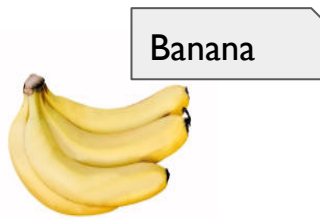
Aprendizaje no
supervisado

Aprendizaje por
refuerzo

Clase de hoy

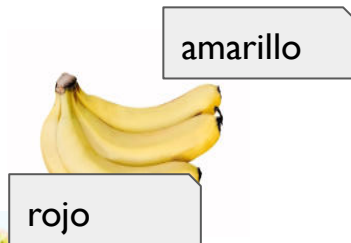
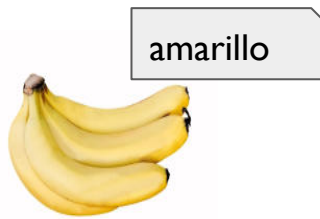


Aprendizaje supervisado



El modelo **aprende de ejemplos etiquetados**.

En otras palabras podemos evaluar si el modelo atinó o no a la respuesta correcta.



El modelo **aprende de ejemplos etiquetados**.

En otras palabras podemos evaluar si el modelo atinó o no a la respuesta correcta.



Aprendizaje no supervisado

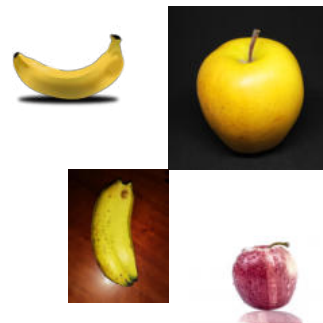


El modelo **explora los datos sin etiquetas** y busca patrones o agrupaciones.

No tiene una respuesta correcta desde el principio y nosotros tampoco.

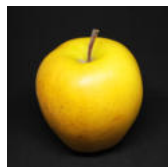


Aprendizaje no supervisado



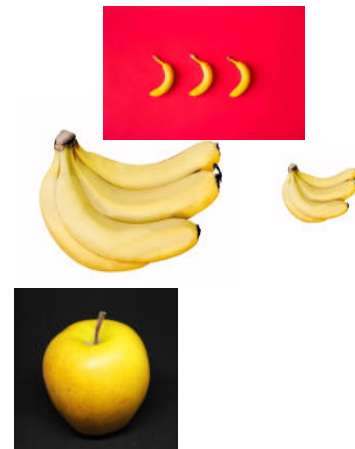


Aprendizaje no supervisado





Aprendizaje no supervisado



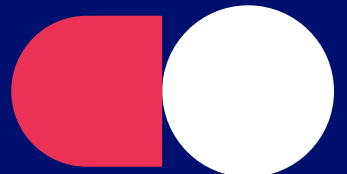


El modelo aprende a **tomar decisiones** en un entorno interactivo, recibiendo **recompensas o penalizaciones** según sus acciones.

<https://www.youtube.com/watch?v=pT2Yzr1RqBo>



Regresión vs clasificación





APRENDIZAJE SUPERVISADO

REGRESIÓN

CLASIFICACIÓN



Regresión

Definición: el modelo predice un valor **continuo** (números).

Ejemplo: **Predecir el puntaje final** de un estudiante en una materia basado en su rendimiento en el primer examen y participación en clase (alta, media o baja).

Clasificación

Definición: el modelo predice **categorías o clases** (por ejemplo, aprobado o reprobado).

Ejemplo: **Predecir la clasificación de aprobado o no aprobado** de un estudiante en una materia basado en su rendimiento en el primer examen y participación en clase (alta, media o baja).



Regresión

Predecir el puntaje final

Esto es la **VARIABLE
DEPENDIENTE**

En regresión la variable dependiente es una variable cuantitativa nivel de medición intervalo o de razón

Clasificación

**Predecir la clasificación
aprobado o no aprobado**

Esto es la **VARIABLE
DEPENDIENTE**

En clasificación la variable dependiente es una variable cualitativa nivel de medición nominal o de ordinal



Variable independentess

Variables independientes o explicativas:

Nota primer examen

Clasificación de la participación del estudiante

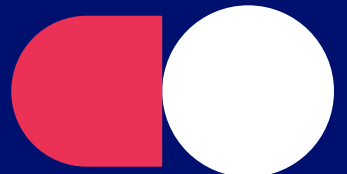
Pueden ser cualitativas o cuantitativas



id estudiante	Nota primer examen	Clasificación de la participación	Puntaje
1	5	alta	4.9
2	4	baja	3.8
3	3	media	2
4	2	alta	3.5



Actividad # I





Situación # 1



Queremos predecir la cantidad de estudiantes que se inscribirán en un curso basado en datos históricos como el número de estudiantes interesados, el % de estudiantes cuya nota ha terminado por encima de 4, cuántos estudiantes han finalizado con éxito este curso previamente

1. ¿Regresión o clasificación?
2. Variable dependiente
3. Variable independiente



Situación # 1



Queremos predecir la cantidad de estudiantes que se inscribirán en un curso basado en datos históricos como el número de estudiantes interesados, el % de estudiantes cuya nota ha terminado por encima de 4, cuántos estudiantes han finalizado con éxito este curso previamente

1. Regresión
2. Variable dependiente:
Número de estudiantes que se inscribirán en un curso
1. Variable independiente
 - a. % de estudiantes que han culminado la materia con una nota por encima de 4
 - b. Número de estudiantes interesados
 - c. Número de estudiantes que han aprobado la nota



Situación # 2



Queremos predecir el índice de satisfacción de los estudiantes con el profesor, basado en datos históricos como el número de horas de asistencia a clase, calificación final obtenida en el curso, y opiniones de los estudiantes sobre el profesor en redes sociales.

1. ¿Regresión o clasificación?
2. Variable dependiente
3. Variable independiente



Situación # 2



Queremos predecir el índice de satisfacción de los estudiantes con el profesor, basado en datos históricos como el número de horas de asistencia a clase, calificación final obtenida en el curso, y opiniones de los estudiantes sobre el profesor en redes sociales.

1. Regresión
2. Variable dependiente:
Índice de satisfacción de los estudiantes
 1. Variable independiente
 - a. número de horas de asistencia a clase
 - b. calificación final obtenida en el curso
 - c. opiniones de los estudiantes sobre el profesor en redes sociales.



Situación # 3



Quiere predecir si un **proyecto de investigación** será **aprobado** o **rechazado** para financiamiento en base a las características de los proyectos anteriores, como el **tema de investigación**, **presupuesto solicitado**, **experiencia del investigador** y **palabras claves del abstract del proyecto**.

1. ¿Regresión o clasificación?
2. Variable dependiente
3. Variable independiente

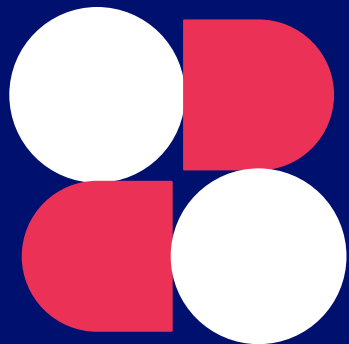


Situación # 3

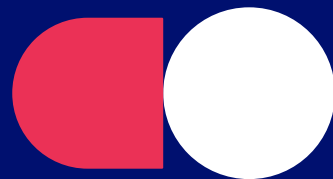


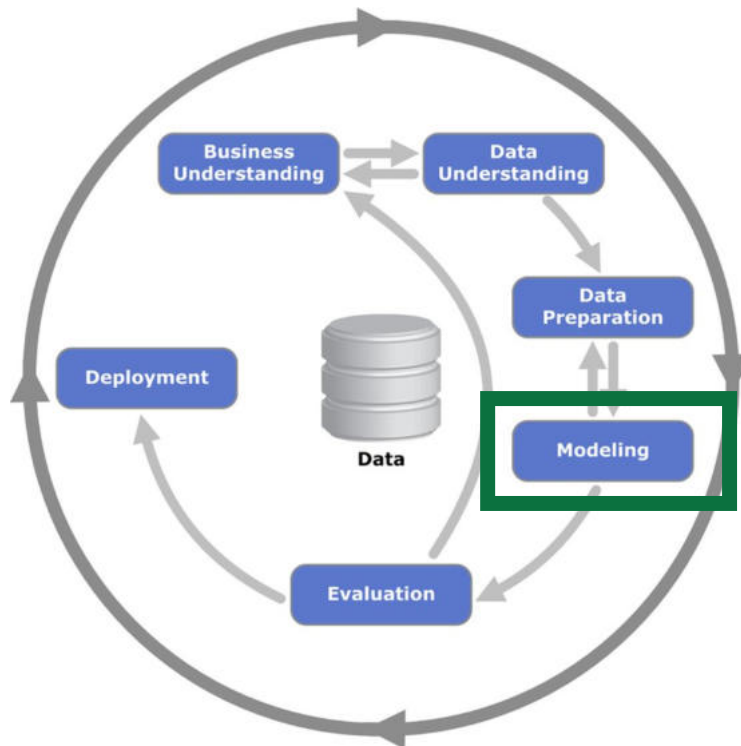
Quiere predecir si un **proyecto de investigación** será **aprobado** o **rechazado** para financiamiento en base a las características de los proyectos anteriores, como el **tema de investigación**, **presupuesto solicitado**, **experiencia del investigador** y **palabras claves del abstract del proyecto**.

1. Clasificación
2. Variable dependiente:
Clasificar al proyecto de investigación en aprobado o no
 1. Variable independiente
 - a. tema de investigación
 - b. presupuesto solicitado
 - c. experiencia del investigador
 - d. palabras claves del abstract del proyecto.



Entrenamiento y evaluación de modelos







Regresión

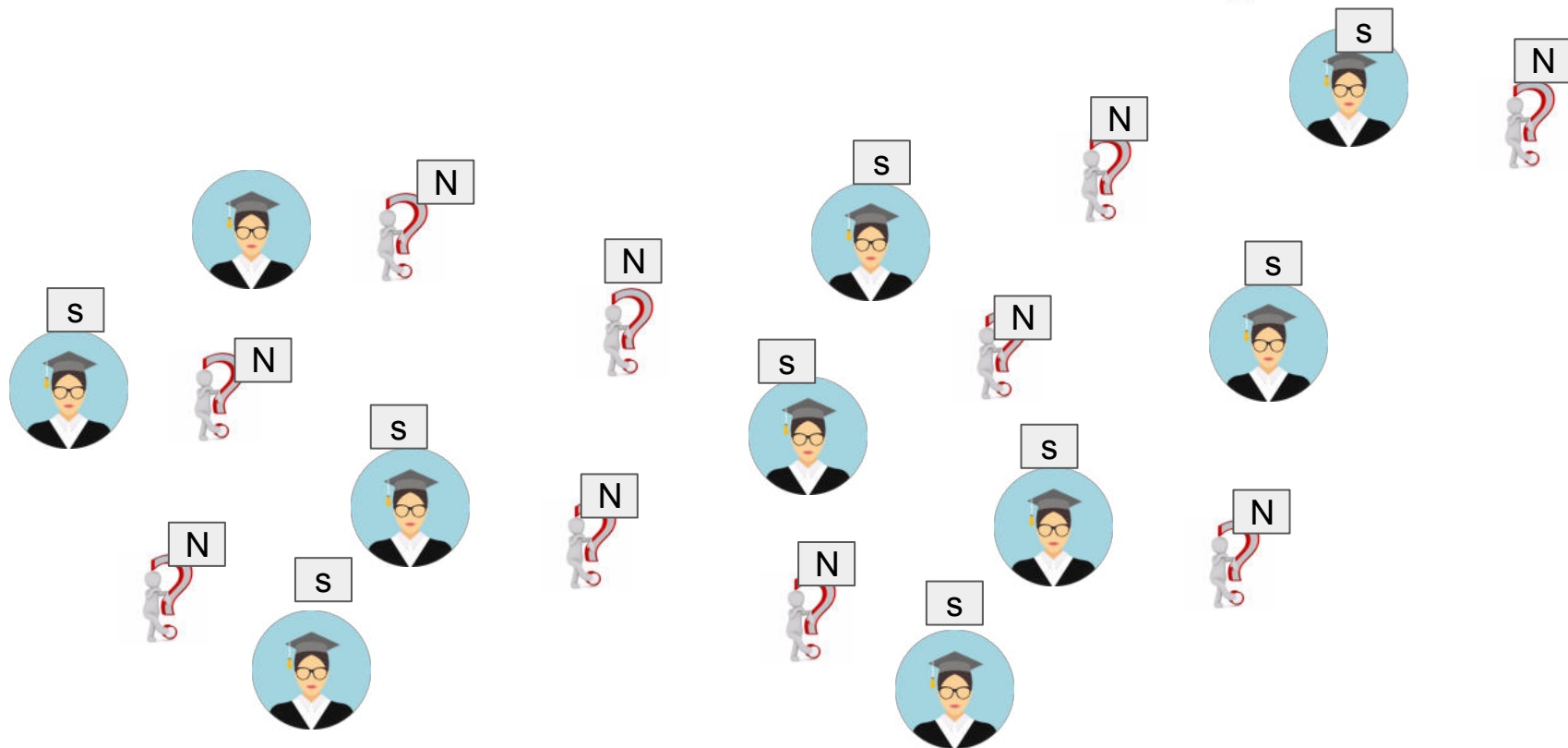
- Regresión lineal
- Árbol de decisión
- Máquinas de Soporte Vectorial (SVR)
- K-Vecinos más Cercanos (K-NN)
- Random forest
- Redes neuronales

Clasificación

- Regresión logística
- Árbol de decisión
- Máquinas de Soporte Vectorial (SVR)
- K-Vecinos más Cercanos (K-NN)
- Random forest
- Redes neuronales

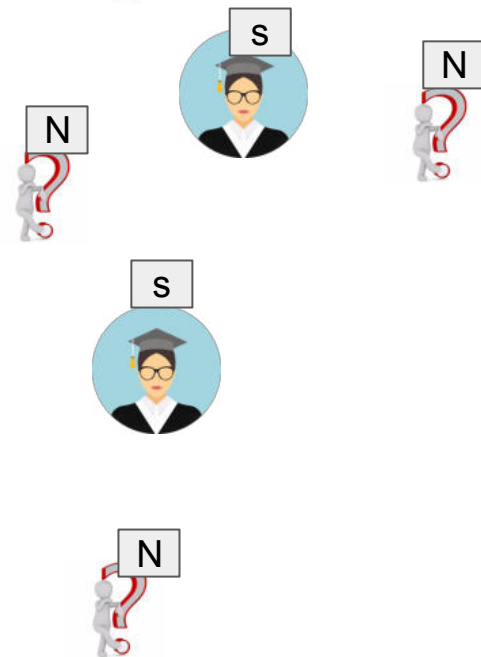
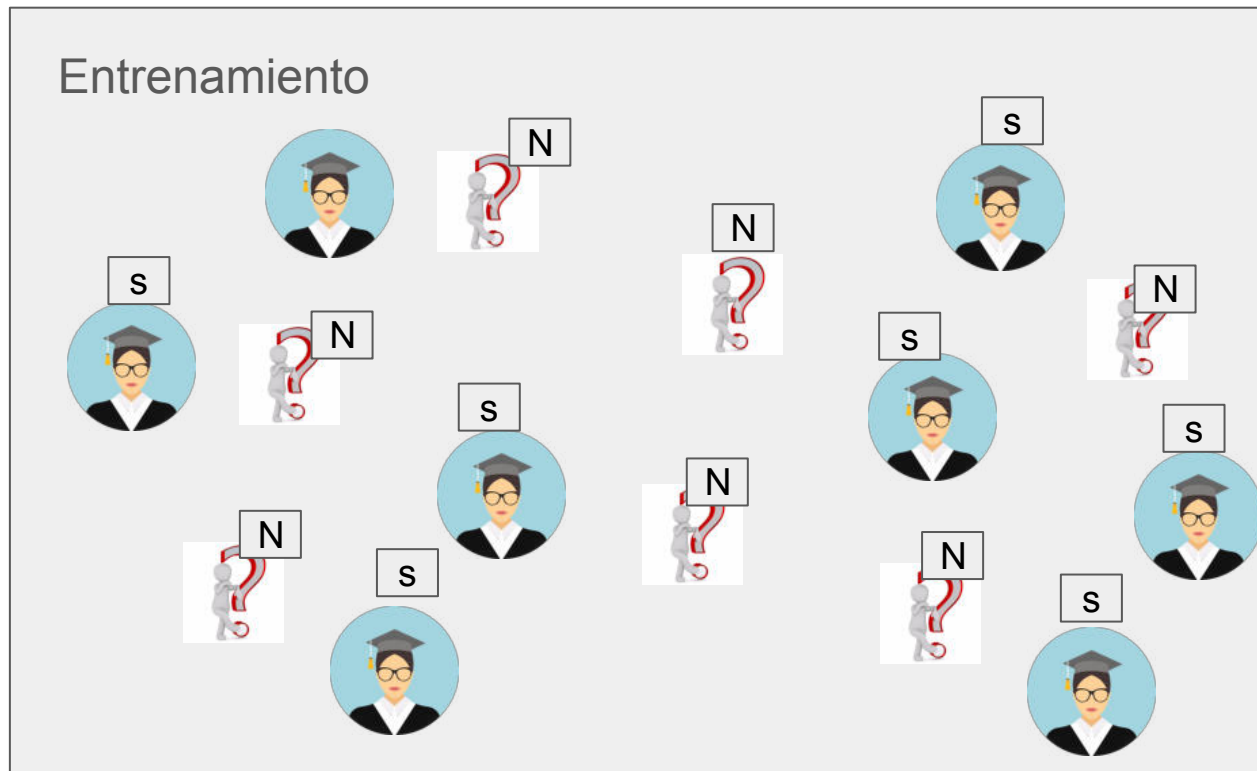


Proceso de entrenamiento del modelo



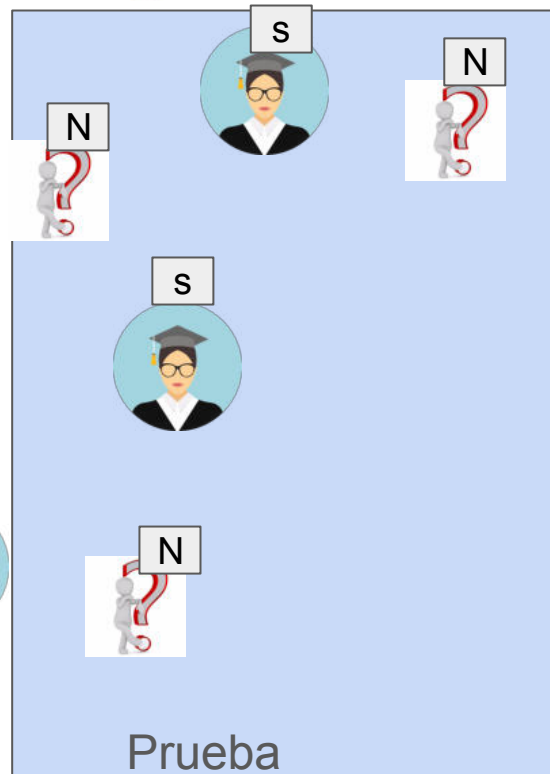
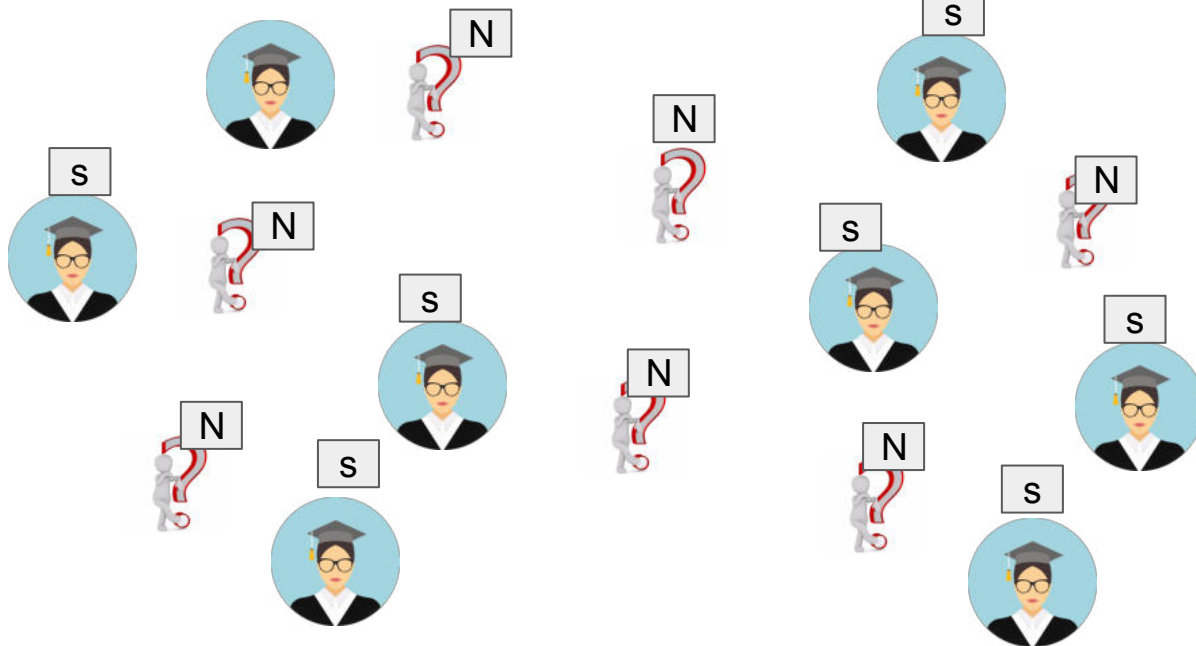


Proceso de entrenamiento del modelo





Proceso de entrenamiento del modelo

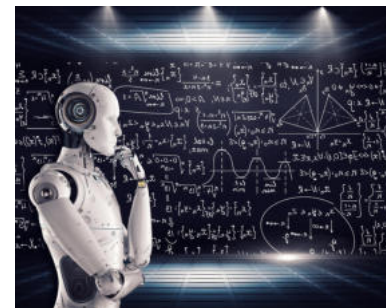
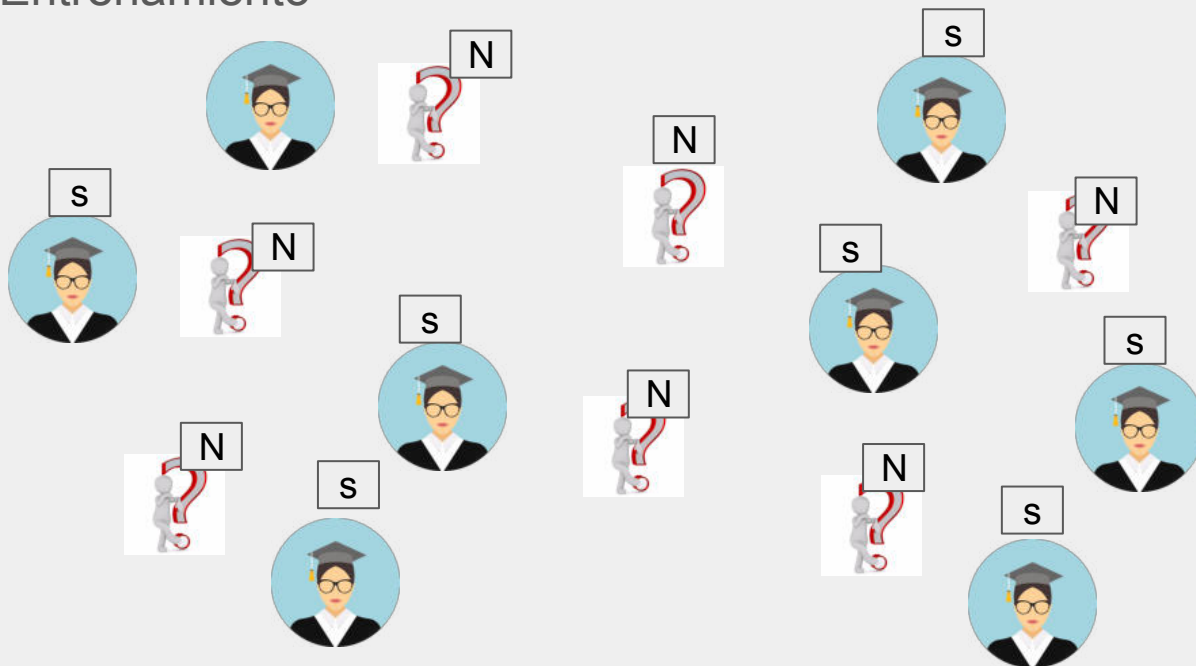




Proceso de entrenamiento del modelo



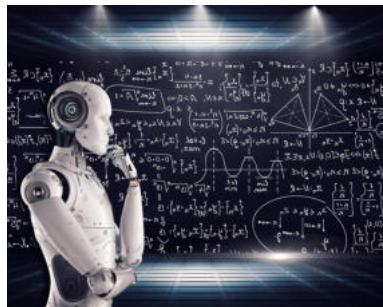
Entrenamiento



Machine learning



Proceso de entrenamiento del modelo



Machine learning



Se gradúa

No se gradúa



Existe un riesgo: sobreajuste (overfitting)



El vendedor te toma las **medidas** para hacerlo **a la perfección**, asegurándose de que cada centímetro del vestido o traje encaje perfectamente a tu cuerpo.



Existe un riesgo: sobreajuste (overfitting)



¿Qué pasa si engordas o adelgazas?
¿Qué pasa si se lo quieres prestar a
alguien para un evento similar para el
cual lo mandaste a diseñar?



Existe un riesgo: sobreajuste (overfitting)



El vestido **ya no te servirá**. No es **flexible** a los cambios, y no puedes usarlo en otras situaciones, aunque en la prueba inicial parecía perfecto.



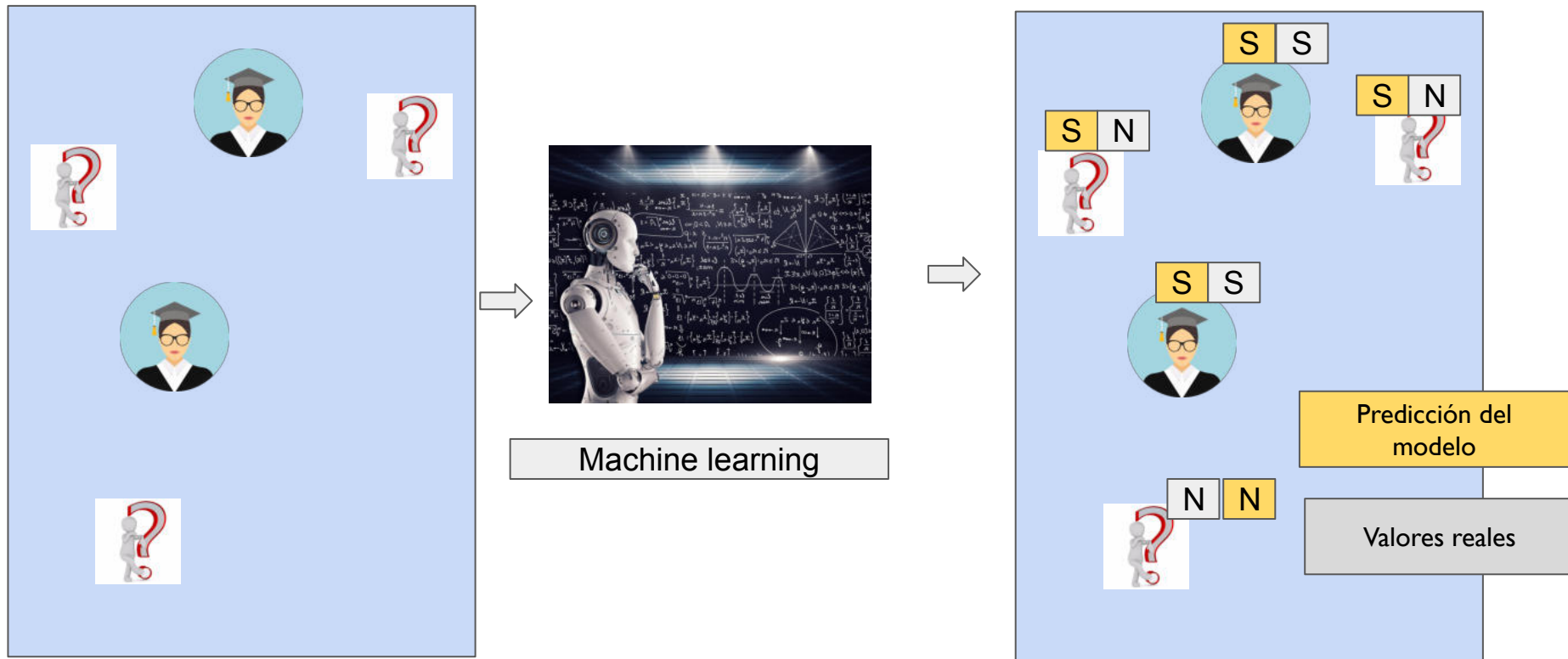
Existe un riesgo: sobreajuste (overfitting)



Al igual que un **vestido hecho a medida** puede ser demasiado específico, un **modelo sobreajustado no será útil** cuando se enfrente a **nuevos datos** o situaciones diferentes.



Proceso de entrenamiento del modelo





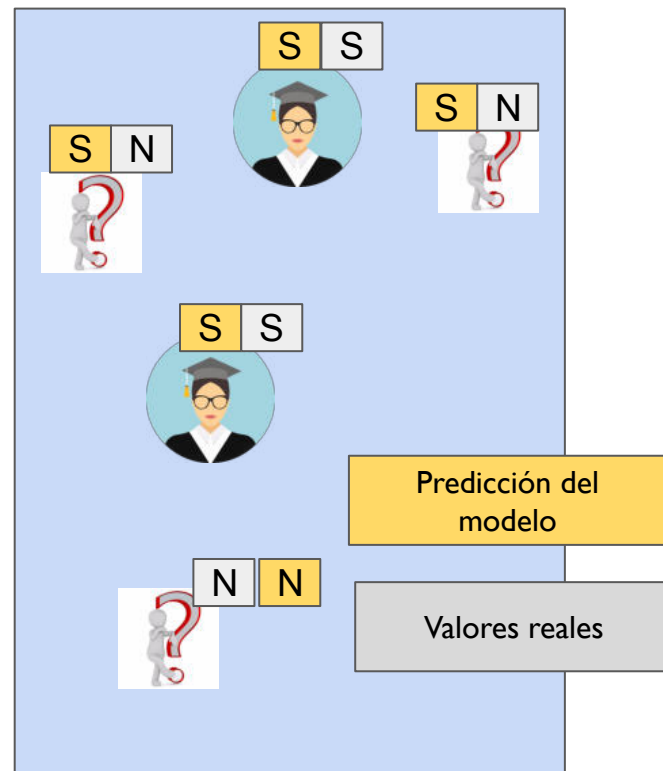
Paso 1: Separar los datos: 80% para entrenamiento, 20% para prueba.

Paso 2: El modelo aprende de los datos de entrenamiento.

Paso 3: El modelo es probado con los datos de prueba para ver cuán bien puede hacer predicciones.



**Una vez tenemos los resultados
cómo evaluamos ...**





Métricas de evaluación

Clasificación



		Valores reales de los datos	
		Se gradúo	No se gradúo
Valores que predice el modelo	Se gradúo	2	2
	No se gradúo	0	1



Métricas de evaluación

Clasificación



		Valores reales de los datos			
		Se gradúo		No se gradúo	
Valores que predice el modelo	Se gradúo	2	Verdaderos positivos	2	Falsos positivos
	No se gradúo	0	Falsos negativos	1	Verdaderos negativos



Métricas de evaluación Clasificación



Exactitud (accuracy)

Precisión (precision)

Sensibilidad (recall o sensitivity o
tasa de verdaderos positivos)

Especificidad (specificity o tasa de
verdaderos negativos)

Tasa de falsos negativos

Muchos más



Métricas de evaluación Clasificación



Exactitud
(accuracy)

Verdaderos
positivos



Verdaderos
negativos

Verdaderos
positivos



Verdaderos
negativos



Falsos positivos



Falsos negativos



Métricas de evaluación Clasificación



Precisión
(precision)

Verdaderos
positivos

Verdaderos
positivos



Falsos positivos



Ejercicio ChatGPT



Actúa como un modelo de aprendizaje supervisado en clasificación de texto. Tengo un conjunto de comentarios de estudiantes sobre un curso.

Tu tarea es: 1. Clasificar cada comentario como Positivo o Negativo (como si fueras un modelo de machine learning). 2. Comparar tus predicciones con las etiquetas reales. 3. Construir la matriz de confusión (VP, VN, FP, FN). 4. Calcular las métricas básicas: exactitud (accuracy), precisión (precision) y recall (sensibilidad). 5. Explicar en lenguaje simple qué significa cada resultado en el contexto universitario.

Comentarios

1. “El profesor domina el tema y responde dudas rápido.”
2. “La carga de lecturas es exagerada e inútil.”
3. “Aprendí bastante aunque las clases eran largas.”
4. “No entendí los temas y no hubo suficiente apoyo.”
5. “Los ejemplos usados fueron claros y útiles.”

Por favor entrégame: - Una tabla con columnas: Comentario | Real | Predicción | Resultado (VP, VN, FP, FN). - La matriz de confusión en formato tabla. - Los cálculos de exactitud, precisión y recall paso a paso. - Una breve interpretación de qué significan los errores para la universidad.



Ejercicio ChatGPT



Comentarios con etiqueta real:

1. “El profesor domina el tema y responde dudas rápido.” → Real: Positivo
2. “La carga de lecturas es exagerada e inútil.” → Real: Negativo
3. “Aprendí bastante aunque las clases eran largas.” → Real: Positivo
4. “No entendí los temas y no hubo suficiente apoyo.” → Real: Negativo
5. “Los ejemplos usados fueron claros y útiles.” → Real: Positivo



y para el caso de regresión?



Proceso de entrenamiento del modelo



Semestre	Inscritos
2015 1	10520
2015 2	11450
2016 1	10250
2016 2	13520
2017 1	12530
2017 2	11452
2018 1	13452
2018 2	14562
2019 1	15789



Proceso de entrenamiento del modelo



Semestre	Inscritos
2015 I	10520
2015 2	11450
2016 I	10250
2016 2	13520
2017 I	12530
2017 2	11452
2018 I	13452
2018 2	14562
2019 I	15789



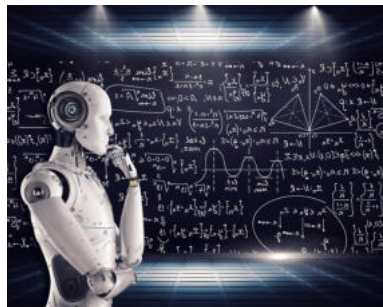
Proceso de entrenamiento del modelo



2018 1

2018 2

2019 1



Machine learning



2018 1 13125 13452

2018 2 14890 14562

2019 1 18250 15789

Predicción del modelo

Valores reales



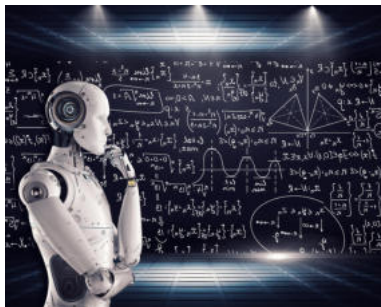
Proceso de entrenamiento del modelo



2018 1

2018 2

2019 1



Machine learning



2018 1 13125 13452

2018 2 14890 14562

2019 1 18250 15789

Predicción del modelo

Valores reales



Métricas de evaluación Regresión



RMSE - Raíz del error cuadrático
medio

MSE - Error Cuadrático Medio

MAE - Error Absoluto Medio

R^2 - Coeficiente de
Determinación

MAPE - Error Porcentual
Absoluto Medio

Muchos más



Métricas de evaluación Regresión



RMSE - Raíz del error cuadrático
medio

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{\text{real}} - y_{\text{predicho}})^2$$

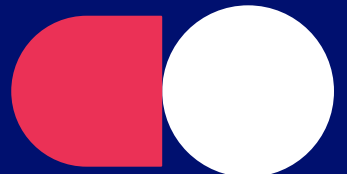
$$RMSE = \sqrt{MSE}$$



- Entrenar múltiples modelos
- Verificar la capacidad de generalización del modelo
- Asegurarse de que los resultados sean comprensibles y aplicables
- Considerar la actualización y el mantenimiento del modelo



Actividad # 2



Actúa como un experto en machine learning con más de 5 años de experiencia y que trabaja en una universidad.

Genera ideas para proyectos de aprendizaje supervisado que se puedan implementar en el área de [especificar área de la universidad]. Para cada proyecto, por favor incluye:

1. **Objetivo del Proyecto:** ¿Cuál es el propósito principal de este proyecto? ¿Qué se espera lograr?
2. **Problema que Resuelve:** ¿Qué problema específico dentro del área está resolviendo este proyecto? ¿Qué tipo de decisiones o tareas se mejoran?
3. **Tipo de Modelo:** ¿Qué tipo de modelo de aprendizaje supervisado sería adecuado para este proyecto? (Ejemplo: regresión o clasificación)
4. **Posibles Variables Dependientes:** ¿Qué variable se intenta predecir o clasificar?
5. **Posibles Variables Independientes:** ¿Qué características o factores se utilizarán para hacer la predicción?
6. **Impacto Potencial:** ¿Cómo impactará este proyecto a la organización o a los usuarios (por ejemplo, estudiantes, profesores, administradores)? ¿Qué mejoras tangibles se esperan?

Genera al menos tres ideas de proyectos, variando entre clasificación y regresión, que puedan ser útiles en este contexto.



Gracias

