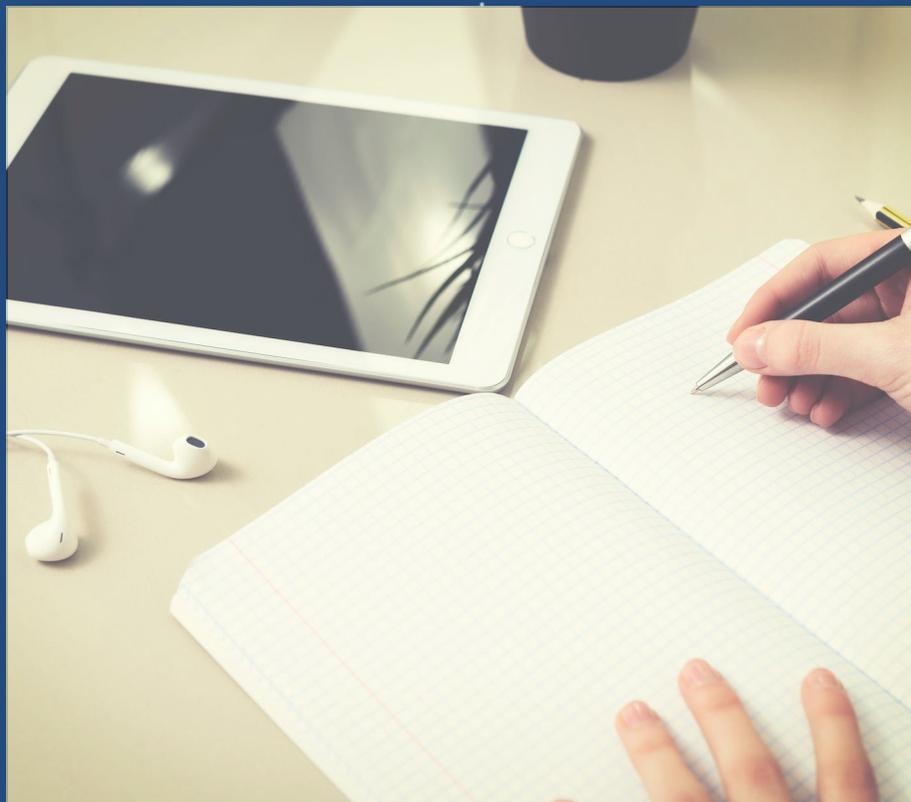


# Actividad 2

## Aprendizaje Supervisado

# Aprendizaje Supervisado



El aprendizaje supervisado es un paradigma dentro del campo del aprendizaje automático en el cual un modelo se entrena utilizando un conjunto de datos etiquetado. En este enfoque, cada ejemplo en el conjunto de datos de entrenamiento consiste en una entrada y su correspondiente salida deseada o etiqueta. El objetivo del modelo es aprender la relación entre las entradas y las salidas, de manera que pueda hacer predicciones precisas sobre nuevas instancias no vistas. En términos más simples, el aprendizaje supervisado implica enseñar al modelo a partir de ejemplos previamente etiquetados, permitiéndole generalizar patrones y hacer predicciones o clasificaciones en datos desconocidos.

# Elementos Clave del Aprendizaje Supervisado



En el Aprendizaje Supervisado, se destacan varios elementos fundamentales que son esenciales para comprender cómo funciona este enfoque y cómo se aplica en la práctica.



**Conjunto de  
Datos  
Etiquetado**



**Modelo de  
Aprendizaje**



**Proceso de  
Entrenamiento**



**Predicción en  
Datos No Vistos**

Algunos modelo de aprendizaje supervisado:

**1**

**2**

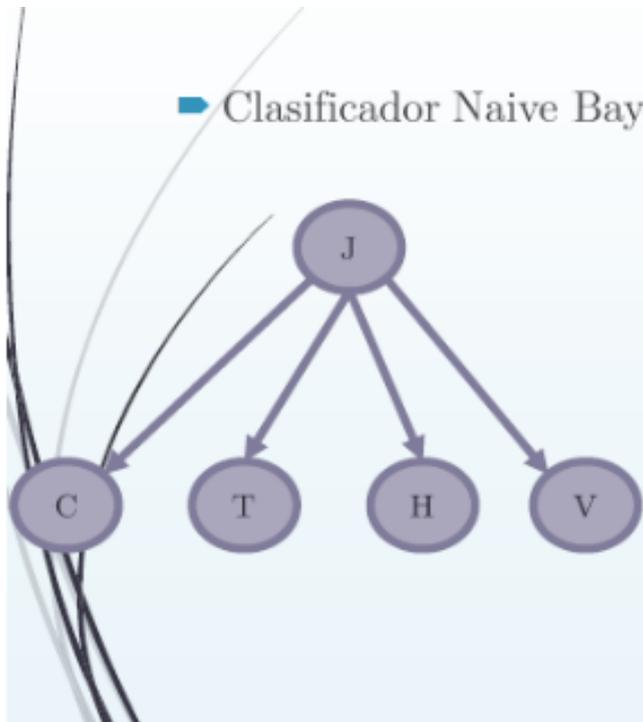
**3**

**4**

En conjunto, estos elementos son los pilares sobre los cuales se construye y desarrolla el proceso de aprendizaje supervisado. Cada uno de ellos desempeña un papel crucial en la capacidad del modelo para aprender a partir de los datos proporcionados y realizar predicciones precisas en nuevos datos no vistos.



➔ Clasificador Naive Bayes



Cielo	Temperatura	Humedad	Viento	Se Jugó
Lluvia	Calor	Alta	No	No
Lluvia	Calor	Alta	Si	No
Nublado	Calor	Alta	No	Si
Soleado	Templado	Alta	No	Si
Soleado	Frio	Normal	No	Si
Soleado	Frio	Normal	Si	No
Nublado	Frio	Normal	Si	Si
Lluvia	Templado	Alta	No	No
Lluvia	Frio	Normal	No	Si
Soleado	Templado	Normal	No	Si
Lluvia	Templado	Normal	Si	Si
Nublado	Templado	Alta	Si	Si
Nublado	Calor	Normal	No	Si
Soleado	Templado	Alta	Si	No

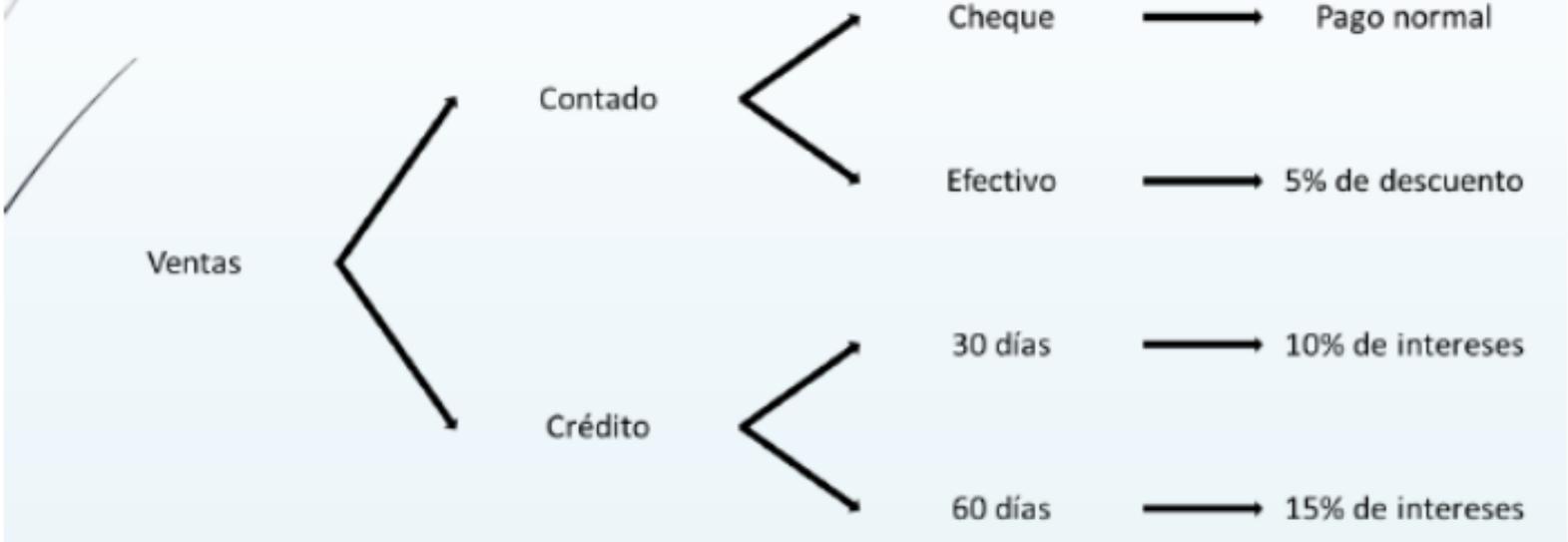


## Modelo de Aprendizaje

El modelo de aprendizaje es la representación matemática o computacional del proceso mediante el cual se realiza la tarea de predicción o clasificación. Puede ser un algoritmo de aprendizaje supervisado, como una regresión lineal, una máquina de vectores de soporte (SVM) o una red neuronal, entre otros. Este modelo se entrena utilizando el conjunto de datos etiquetado, de modo que pueda aprender la relación entre las características de entrada y las etiquetas de salida.

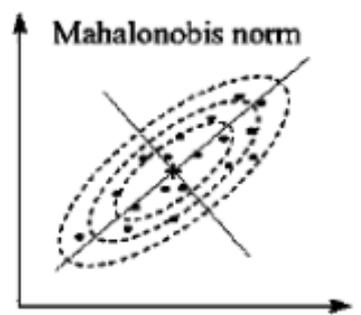
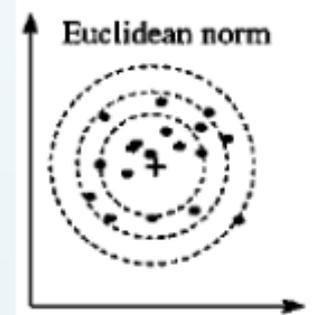
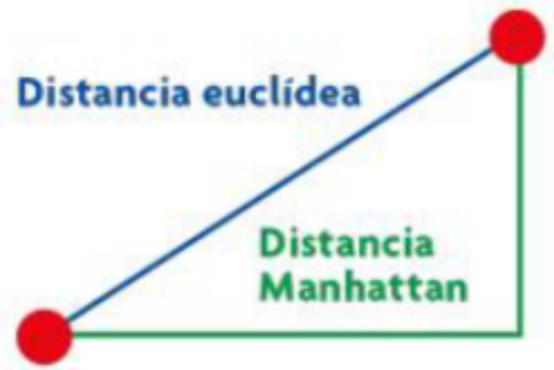
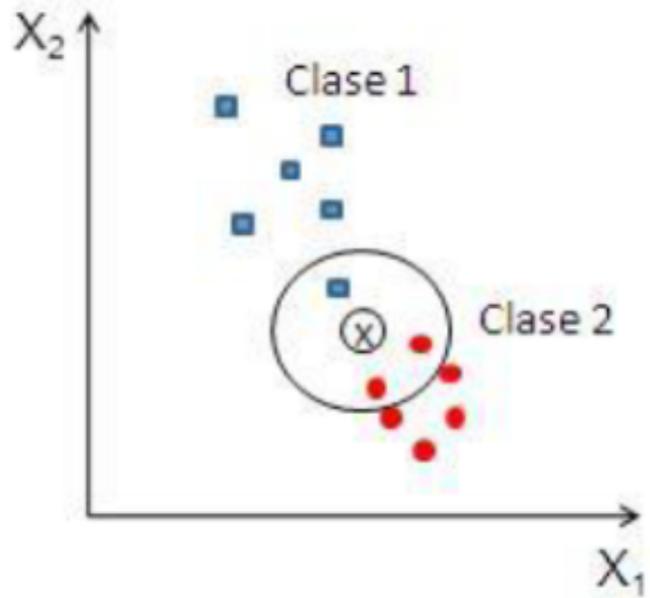


### ▶ Árbol de decisión de aprendizaje





Vecinos cercanos



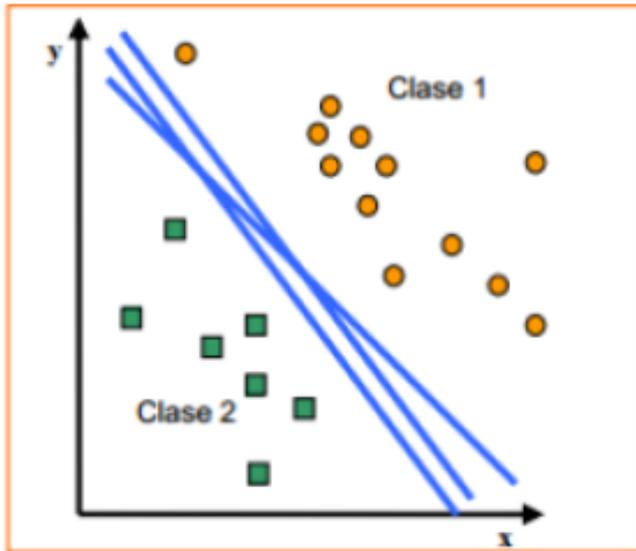


## Proceso de Entrenamiento

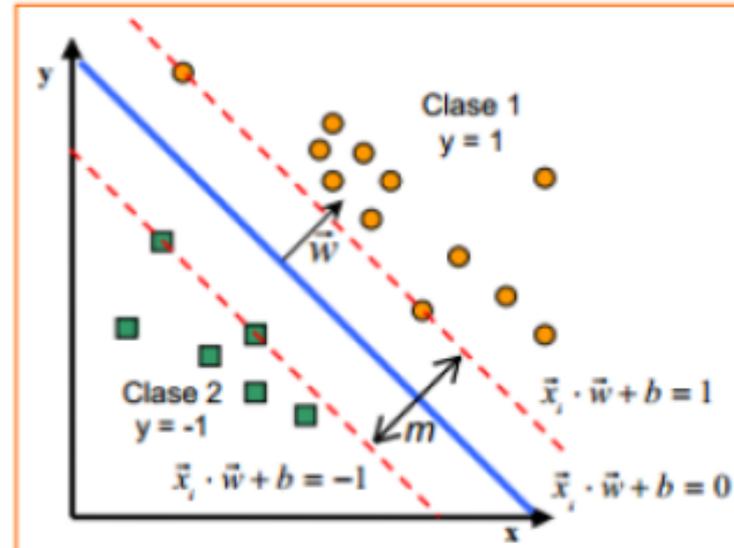
Durante el proceso de entrenamiento, el modelo ajusta sus parámetros internos para minimizar la discrepancia entre las predicciones que genera y las etiquetas reales proporcionadas en el conjunto de datos de entrenamiento. Este ajuste de parámetros se lleva a cabo utilizando técnicas de optimización que buscan encontrar los valores óptimos que permitan al modelo realizar las predicciones más precisas posibles.



■ Máquina de soporte vectorial



**Posibles hiperplanos de separación**



**Hiperplano de separación óptimo**



## Conjunto de Datos Etiquetado

Este componente constituye la base sobre la cual se construye el proceso de aprendizaje. Se trata de un conjunto de ejemplos de entrada, cada uno asociado con una salida o etiqueta correspondiente. Estas etiquetas representan la respuesta deseada que el modelo debe aprender a predecir a partir de las características de entrada. La presencia de etiquetas en los datos es lo que distingue al aprendizaje supervisado de otros enfoques de aprendizaje automático.



## Predicción en Datos No Vistos

Una vez que el modelo ha sido entrenado adecuadamente, se evalúa su desempeño haciendo predicciones en datos que no han sido utilizados durante el entrenamiento, es decir, datos que no formaban parte del conjunto de datos de entrenamiento. Estos datos no vistos proporcionan una medida de la capacidad de generalización del modelo, es decir, su habilidad para realizar predicciones precisas en situaciones nuevas y no vistas previamente.

# Ejemplo Conceptual



Supongamos que queremos entrenar un modelo para reconocer dígitos escritos a mano. El conjunto de datos de entrenamiento contendría imágenes de dígitos junto con las etiquetas que indican qué dígito representa cada imagen. Durante el entrenamiento, el modelo aprendería a asociar características visuales específicas con los dígitos correctos. Después del entrenamiento, podríamos usar el modelo para predecir el dígito de nuevas imágenes no vistas con una buena probabilidad de precisión.

Ejemplos de Aplicación en Diferentes Contextos:



**Clasificación de  
Correos  
Electrónicos**



**Diagnóstico  
Médico**



**Reconocimiento  
de Imágenes**



**Análisis de  
Sentimientos en  
Redes Sociales**



**Predicción del  
Tiempo**



**Recomendación  
de Productos**



## Diagnóstico Médico

*Predecir si un paciente tiene una enfermedad específica.*

**Ejemplo:** Utilizando datos médicos previos, un modelo puede aprender a asociar ciertos síntomas y resultados de pruebas con diagnósticos específicos para ayudar en el diagnóstico de nuevos pacientes.



## Análisis de Sentimientos en Redes Sociales

*Determinar el tono emocional de un mensaje.*

**Ejemplo:** Un modelo de aprendizaje supervisado puede entrenarse con mensajes previamente etiquetados como positivos, negativos o neutros en redes sociales. Después, puede clasificar automáticamente el tono de nuevos mensajes.



## Recomendación de Productos

*Sugerir productos a usuarios en plataformas de comercio electrónico.*

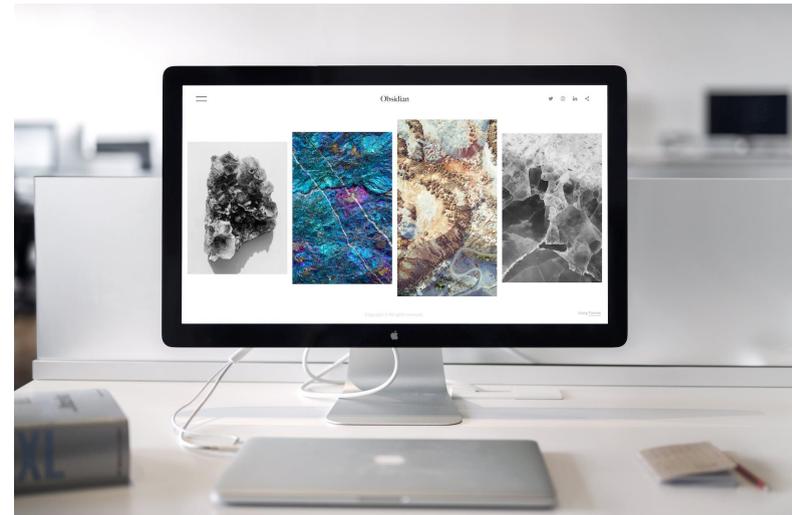
**Ejemplo:** Un modelo puede entrenarse con datos de comportamiento de usuarios y sus compras pasadas para predecir productos que podrían interesarles.



# Reconocimiento de Imágenes

*Identificar objetos en imágenes.*

**Ejemplo:** Un modelo puede entrenarse con un conjunto de imágenes etiquetadas con objetos específicos, como perros, gatos, automóviles, etc. Luego, puede identificar estos objetos en nuevas imágenes.



## Clasificación de Correos Electrónicos

*Identificar si un correo electrónico es spam o no.*

**Ejemplo:** Un modelo de aprendizaje supervisado puede entrenarse con correos electrónicos previamente etiquetados como spam o no spam.

Luego, puede predecir la categoría de nuevos correos electrónicos.



## Predicción del Tiempo

*Prever las condiciones meteorológicas futuras.*

**Ejemplo:** Utilizando datos históricos de temperatura, presión atmosférica, etc., un modelo puede aprender a predecir el clima futuro, lo que es esencial para pronósticos precisos.





El preprocesamiento de datos desempeña un papel fundamental en el éxito de los modelos de aprendizaje supervisado.



**Manejo de  
Datos Faltantes**



**Normalización y  
Escalado**



**Codificación de  
Variables  
Categorías**



**Gestión de Datos  
Desbalanceados**

Ejemplos de Mejora de Rendimiento al aplicar técnicas como normalización, manejo de valores faltantes, codificación de variables categóricas:



**Eliminación de  
Outliers**



**Normalización  
de Texto**



**Imputación de  
Datos Faltantes**



**Selección de  
Características**



## Selección de Características

Algunas características pueden ser irrelevantes o redundantes. El preprocesamiento mediante técnicas de selección de características optimiza el conjunto de características, mejorando la eficiencia del modelo y reduciendo la posibilidad de sobreajuste.



El preprocesamiento de datos es esencial para garantizar que los conjuntos de datos sean adecuados y efectivos para la construcción de modelos de aprendizaje supervisado. Mejora la calidad de los datos, aborda problemas específicos y contribuye significativamente al rendimiento y la generalización del modelo. Considerar cuidadosamente las necesidades específicas del conjunto de datos y del modelo es clave para un preprocesamiento efectivo.



## Codificación de Variables Categóricas

Los algoritmos de aprendizaje supervisado generalmente requieren que las variables categóricas se conviertan en formatos numéricos. El preprocesamiento incluye técnicas como la codificación one-hot para manejar estas variables.



## Imputación de Datos Faltantes

En conjuntos de datos con valores faltantes, el preprocesamiento mediante técnicas de imputación (media, mediana, etc.) garantiza que el modelo no se vea afectado negativamente por la falta de información.



## Eliminación de Outliers

Un conjunto de datos puede contener valores atípicos que afectan negativamente el rendimiento del modelo. El preprocesamiento mediante la eliminación de outliers mejora la capacidad del modelo para generalizar.



## Gestión de Datos Desbalanceados

En problemas con clases desbalanceadas, donde una clase tiene muchos más ejemplos que la otra, el preprocesamiento puede incluir técnicas como el submuestreo, sobre muestreo o generación de datos sintéticos para abordar este problema.



## Manejo de Datos Faltantes

Los conjuntos de datos reales a menudo contienen valores faltantes. El preprocesamiento puede abordar este problema mediante técnicas como la imputación de datos o la eliminación de instancias incompletas.



## Normalización de Texto

En problemas de procesamiento de lenguaje natural (NLP), normalizar el texto mediante la eliminación de puntuación, la conversión a minúsculas y la lematización puede mejorar la calidad de las características y la eficacia del modelo.



## Normalización y Escalado

Diferentes características pueden tener escalas diferentes. La normalización y el escalado aseguran que todas las características contribuyan de manera equitativa al modelo, evitando la dominación de características con magnitudes mayores.



# Cual es el Problemas Específicos del Aprendizaje Supervisado

El problema específico del aprendizaje supervisado radica en la necesidad de que el modelo aprenda a generalizar adecuadamente a partir de los datos de entrenamiento para hacer predicciones precisas en datos nuevos y no vistos. Aunque los algoritmos de aprendizaje supervisado se entrenan inicialmente sobre un conjunto de datos de entrenamiento, donde se mide el error de entrenamiento, la verdadera medida del desempeño del modelo radica en su capacidad para generalizar y hacer predicciones precisas en un conjunto de datos separado, conocido como conjunto de prueba, donde se mide el error de generalización.

Este proceso de entrenar sobre un conjunto de datos y luego evaluar sobre otro introduce una serie de desafíos y complejidades en el problema de optimización:



Generalización  
óptima



Sobreajuste  
(Overfitting)



Selección de  
Hiperparámetros



Evaluación del  
Rendimiento



## Selección de Hiperparámetros

Para mitigar el sobreajuste y optimizar el rendimiento del modelo, a menudo es necesario ajustar los hiperparámetros del algoritmo de aprendizaje, como la tasa de aprendizaje en redes neuronales o el parámetro de regularización en modelos lineales. Encontrar la combinación óptima de hiperparámetros puede ser un desafío y requiere experimentación y validación utilizando conjuntos de datos de validación.



El problema específico del aprendizaje supervisado radica en optimizar el rendimiento del modelo para generalizar de manera efectiva a partir de los datos de entrenamiento y hacer predicciones precisas en datos nuevos y no vistos. Esto implica abordar desafíos como el sobreajuste, la selección de hiperparámetros y la evaluación del rendimiento de manera rigurosa y efectiva.





## Sobreajuste (Overfitting)

El sobreajuste ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento, capturando el ruido y las características irrelevantes en lugar de las relaciones subyacentes que generalmente se encuentran en los datos. Como resultado, el modelo puede tener un rendimiento deficiente cuando se enfrenta a nuevos datos, lo que reduce su capacidad de generalización.



## Evaluación del Rendimiento

La evaluación del rendimiento del modelo en el conjunto de prueba proporciona una medida importante de su capacidad de generalización. Sin embargo, es crucial garantizar que el conjunto de prueba no se utilice durante el proceso de entrenamiento o ajuste de hiperparámetros, para evitar sesgos en la evaluación del rendimiento.





## Generalización óptima

El objetivo final del aprendizaje supervisado es que el modelo pueda generalizar de manera óptima a partir de los datos de entrenamiento para hacer predicciones precisas en datos nuevos y no vistos. Sin embargo, esto puede ser difícil de lograr, ya que el modelo debe aprender a discernir los patrones y relaciones relevantes en los datos de entrenamiento sin memorizarlos, lo que puede llevar al sobreajuste.



# Avances y Tendencias en Aprendizaje Supervisado



## Modelos Preentrenados:

Se refiere a modelos que han sido entrenados previamente en grandes conjuntos de datos y tareas específicas antes de ser ajustados para una tarea particular.

## Ventajas en Transferencia de Conocimiento:

Los modelos preentrenados poseen conocimientos generales que pueden ser transferidos a nuevas tareas, acelerando el aprendizaje y mejorando el rendimiento con conjuntos de datos más pequeños.

## Ejemplos de Aplicación en NLP (Procesamiento del Lenguaje Natural):

Modelos preentrenados como GPT-3 han revolucionado el campo del procesamiento del lenguaje natural, logrando resultados destacados en tareas como comprensión del lenguaje y generación de texto.



# Avances y Tendencias en Aprendizaje Supervisado



## Técnicas de Transfer Learning:

Es un paradigma donde un modelo entrenado en una tarea general se adapta para tareas específicas relacionadas.

## Aplicaciones en Visión por Computadora:

Modelos entrenados en grandes conjuntos de datos para reconocimiento de imágenes pueden transferir conocimientos a tareas específicas como detección de objetos o clasificación de imágenes médicas.

**Adaptación a Dominios Específicos:** Las técnicas de transfer learning permiten adaptar modelos a dominios específicos con menos datos de entrenamiento, facilitando su aplicación en contextos empresariales o industriales.



# Avances y Tendencias en Aprendizaje Supervisado



## Desafíos y Futuras Direcciones:

La investigación se centra en desarrollar modelos preentrenados y técnicas de transfer learning que puedan generalizar conocimientos a una amplia gama de tareas.

**Interpretabilidad y Ética:** A medida que los modelos preentrenados se vuelven más complejos, la interpretabilidad y las consideraciones éticas en el uso de grandes conjuntos de datos son temas clave a abordar.





**TIC**

▶ **TALENTO**  
**TECH**

**AZ** | **PROYECTOS**  
**EDUCATIVOS**

