

04 – Modelamiento



Centro de Servicios y Gestión Empresarial
SENA Regional Antioquia



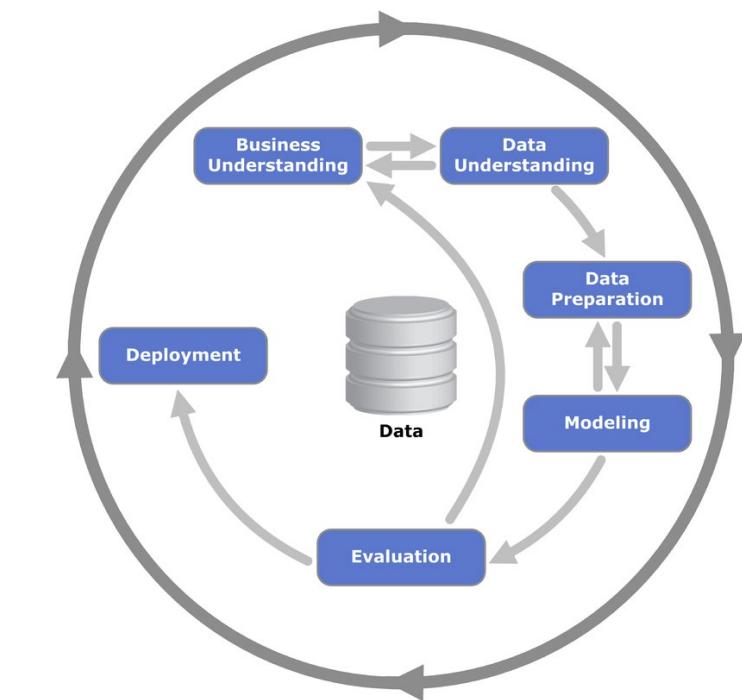
www.sena.edu.co

Modelamiento

Modelamiento

En esta fase se prueban diferentes algoritmos y modelos que podrían resolver el problema planteado. Los modelos son entrenados utilizando los datos preparados. Los parámetros de los modelos se ajustan para maximizar su rendimiento.

Objetivo: Aplicar técnicas de modelado matemático o de machine learning.



Inteligencia Artificial



Inteligencia Artificial



La inteligencia artificial (IA) se ha convertido en un campo de rápido desarrollo, con un impacto profundo en nuestras vidas.

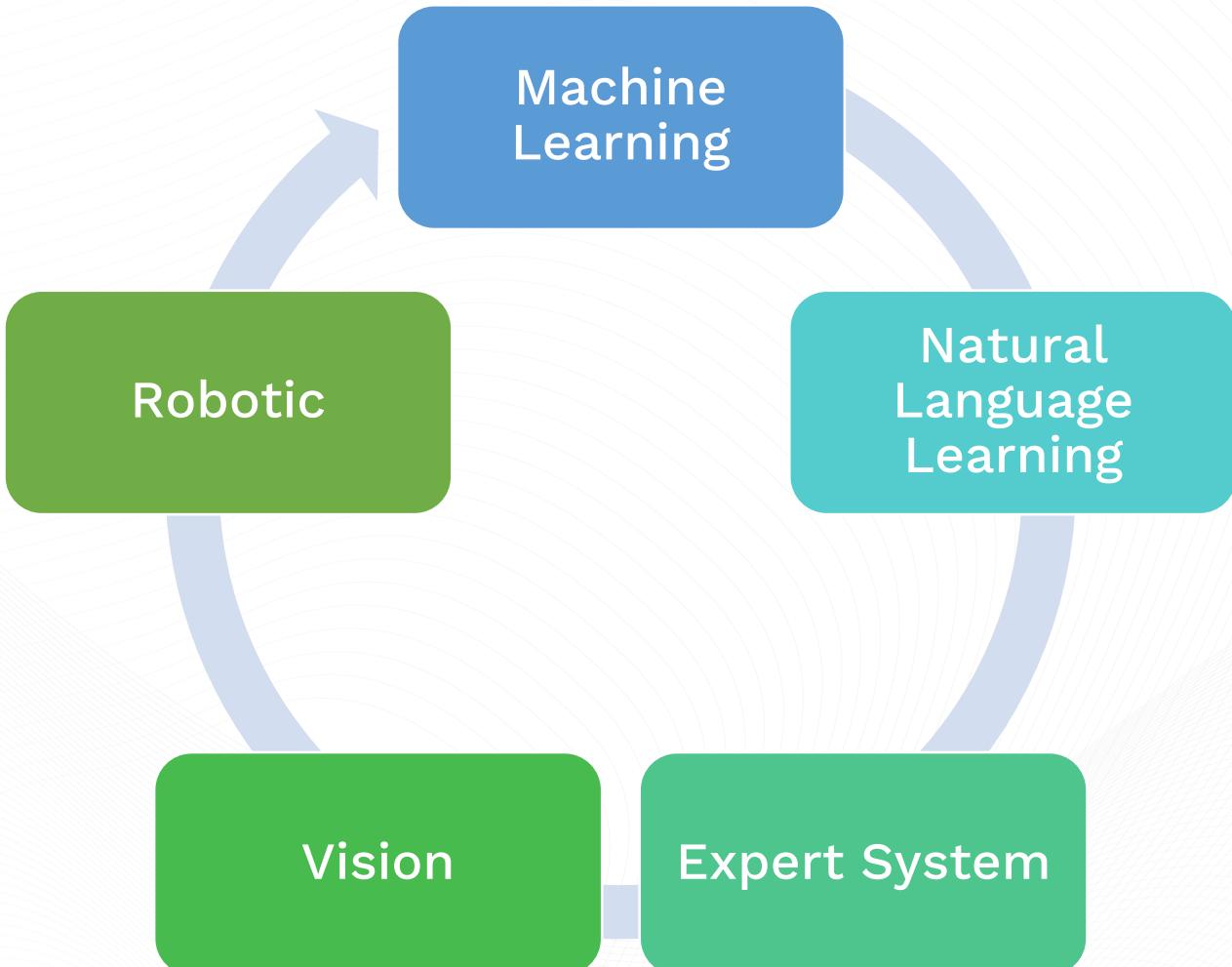


Se refiere a la capacidad de las máquinas para realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la resolución de problemas y la toma de decisiones.





Tipos de Inteligencia Artificial



Aprendizaje Automático Supervisado

En el aprendizaje supervisado, el modelo se entrena con datos etiquetados. Cada ejemplo de entrenamiento contiene una entrada y una salida deseada. El modelo aprende a asociar las entradas con las salidas deseadas, y luego puede predecir las salidas para nuevas entradas.



Tipo de Problema

Clasificación

Ejemplos

Identificar el tipo de correo electrónico (spam o no spam), clasificar imágenes de animales, predecir si un cliente comprará un producto.

Regresión

Predecir el precio de una casa, la temperatura del aire, el valor de una acción.

Aprendizaje Automático No supervisado

En el aprendizaje no supervisado, el modelo se entrena con datos sin etiquetar. El objetivo es encontrar patrones, estructuras o agrupaciones en los datos. El modelo debe aprender a agrupar los datos de forma significativa sin recibir información sobre la salida deseada.



1 Agrupamiento

Los algoritmos de agrupamiento buscan identificar grupos o clústeres en los datos. Por ejemplo, se puede utilizar para segmentar clientes en diferentes grupos según sus preferencias de compra.



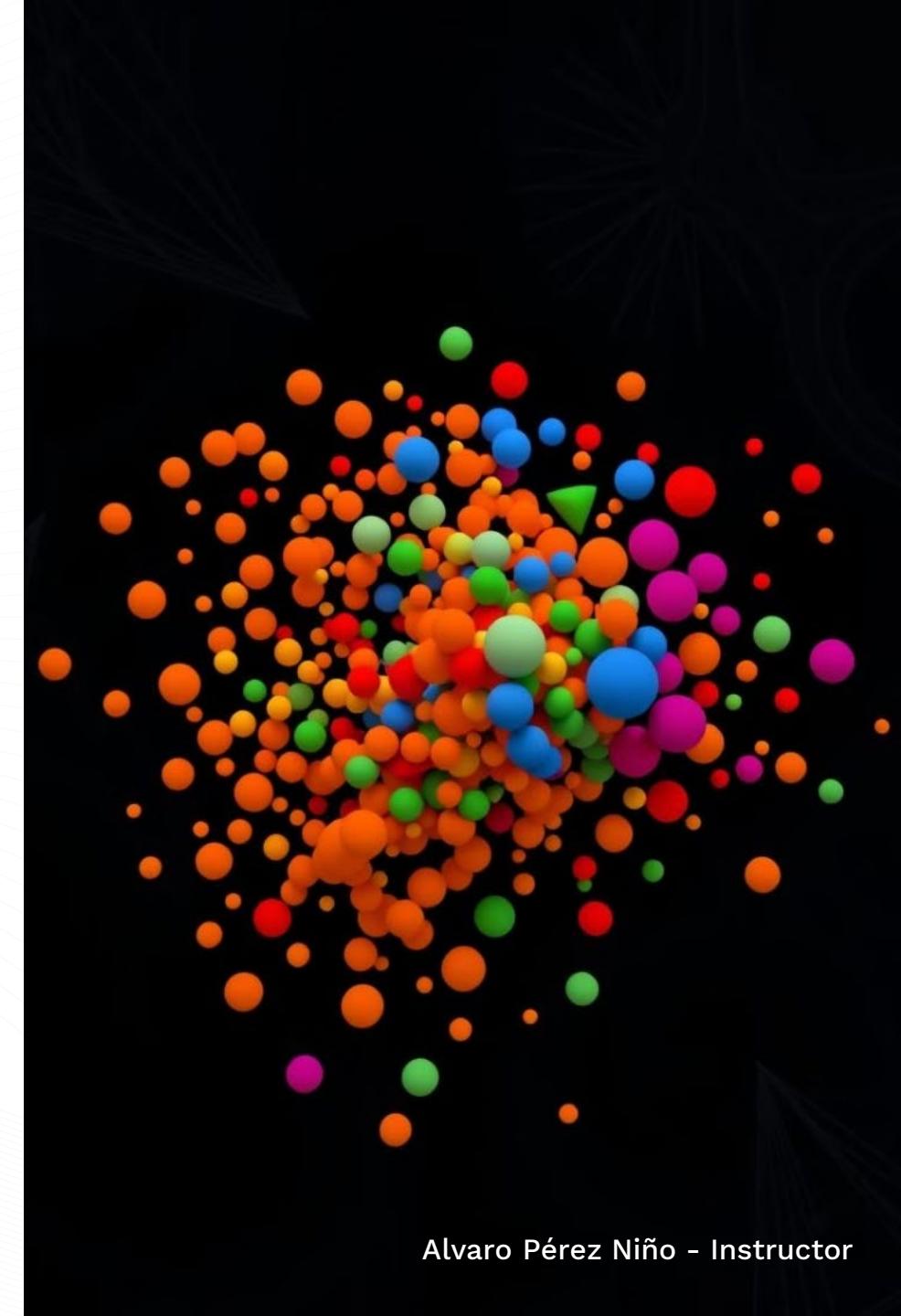
2 Reducción de Dimensionalidad

Estas técnicas reducen el número de variables en un conjunto de datos manteniendo la información más relevante. Se utiliza para simplificar los datos y mejorar la eficiencia de los algoritmos de ML.



3 Detección de Anomalías

Los algoritmos de detección de anomalías identifican valores atípicos o datos que no se ajustan al patrón normal. Se utiliza para detectar fraudes, errores de sensores y otros eventos inusuales.



Procesamiento de Lenguaje Natural



Análisis de Sentimientos

Identifica la emoción o el sentimiento expresado en un texto, como positivo, negativo o neutral. Se utiliza para analizar comentarios de clientes, publicaciones en redes sociales y opiniones públicas.



Traducción Automática

Traduce texto de un idioma a otro. Se utiliza para facilitar la comunicación entre personas que hablan diferentes idiomas, para fines comerciales y educativos.



Chatbots

Son programas informáticos que simulan conversaciones con personas. Se utilizan para brindar atención al cliente, responder preguntas frecuentes y proporcionar información personalizada.



Resumen Automático

Resume textos largos de forma concisa y precisa. Se utiliza para generar resúmenes de artículos, informes y documentos.

Visión Artificial

La visión artificial permite a las máquinas "ver" y comprender imágenes y vídeos. Se utiliza para desarrollar tecnologías como reconocimiento facial, detección de objetos, análisis de imágenes médicas y vehículos autónomos.

Reconocimiento de Objetos

Identifica objetos específicos en imágenes o vídeos, por ejemplo, reconocer coches, personas, animales o productos.

Detección de Objetos

Determina la ubicación y el tamaño de los objetos en una imagen o vídeo, como la detección de obstáculos en la conducción autónoma.

Análisis de Imágenes Médicas

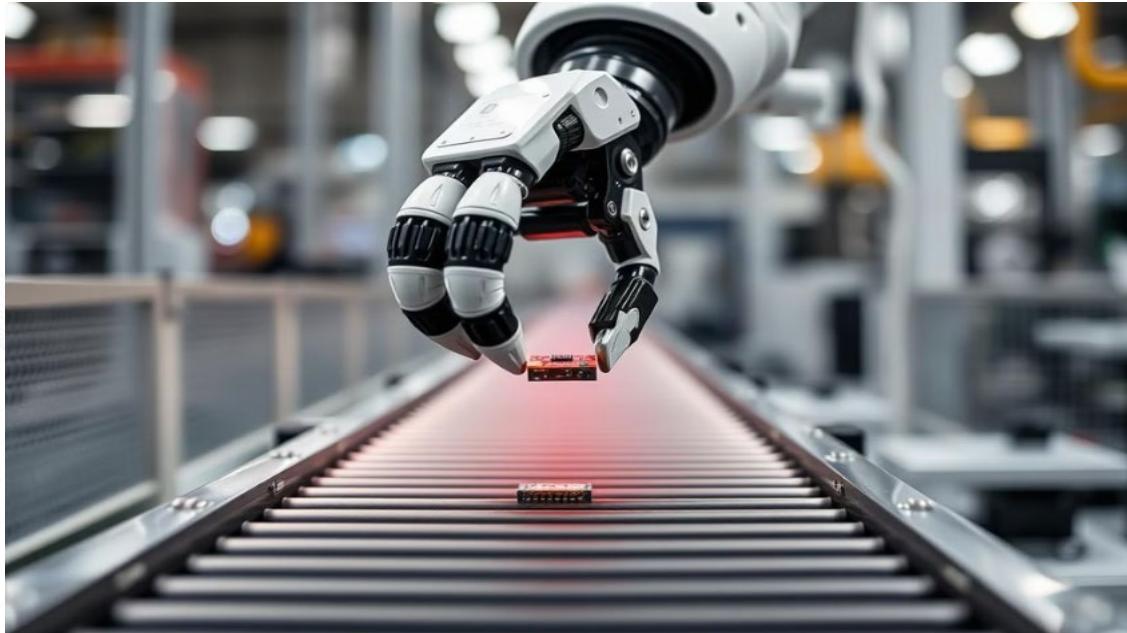
Se utiliza para analizar imágenes médicas, como radiografías, tomografías y resonancias magnéticas, para detectar enfermedades y realizar diagnósticos.



Robótica y Automatización



La robótica y la automatización implican el uso de robots y sistemas automatizados para realizar tareas repetitivas, peligrosas o complejas. Se utiliza en diversos sectores como la manufactura, la logística, la agricultura y el cuidado de la salud.



Automatización Industrial

Los robots industriales se utilizan en tareas como soldadura, pintura, ensamblaje y manipulación de materiales, mejorando la eficiencia, la calidad y la seguridad en la producción.



Robótica de Servicios

Se utilizan robots para realizar tareas de servicio, como entrega de paquetes, asistencia en el hogar, atención al cliente y limpieza.

Ética, Responsabilidad y Futuro en la IA



A medida que la IA avanza, es esencial abordar las implicaciones éticas y sociales. Se deben desarrollar mecanismos para garantizar que la IA se utilice de forma responsable, ética y que no amplíe las desigualdades existentes.

Sesgo y Discriminación

Los datos utilizados para entrenar modelos de IA pueden reflejar sesgos sociales existentes, lo que puede resultar en discriminación en las decisiones tomadas por la IA. Es importante abordar estos sesgos y desarrollar mecanismos para garantizar la equidad en los sistemas de IA.

Impacto en el Empleo

La automatización impulsada por la IA podría desplazar a algunos trabajadores. Es importante crear mecanismos para recapacitar y reconvertir a los trabajadores afectados, y desarrollar políticas para garantizar una transición justa.

Futuro de la IA

El futuro de la IA se perfila como prometedor, con nuevas tecnologías y aplicaciones que podrían revolucionar la forma en que vivimos y trabajamos. Se espera que la IA desempeñe un papel cada vez más importante en diversos ámbitos, desde la medicina hasta la educación y la investigación científica.

1

2

3

4

5

Privacidad y Seguridad

La IA maneja grandes cantidades de datos personales, por lo que es crucial proteger la privacidad y seguridad de estos datos. Se necesitan marcos legales y tecnológicos para regular el uso y la protección de datos personales en el contexto de la IA.

Gobierno y Regulación

Se necesitan marcos regulatorios para guiar el desarrollo y la aplicación de la IA, asegurando que se utilice de forma responsable y que se aborden los riesgos potenciales. Esto incluye la creación de estándares éticos, la regulación del uso de datos y la protección de los derechos humanos.

Aprendizaje Automático Machine Learning

Aprendizaje Automático

El aprendizaje automático (ML) es un subcampo de la IA que permite a las máquinas aprender de los datos sin ser programadas explícitamente. Los algoritmos de ML pueden identificar patrones, hacer predicciones y mejorar su precisión con la experiencia.

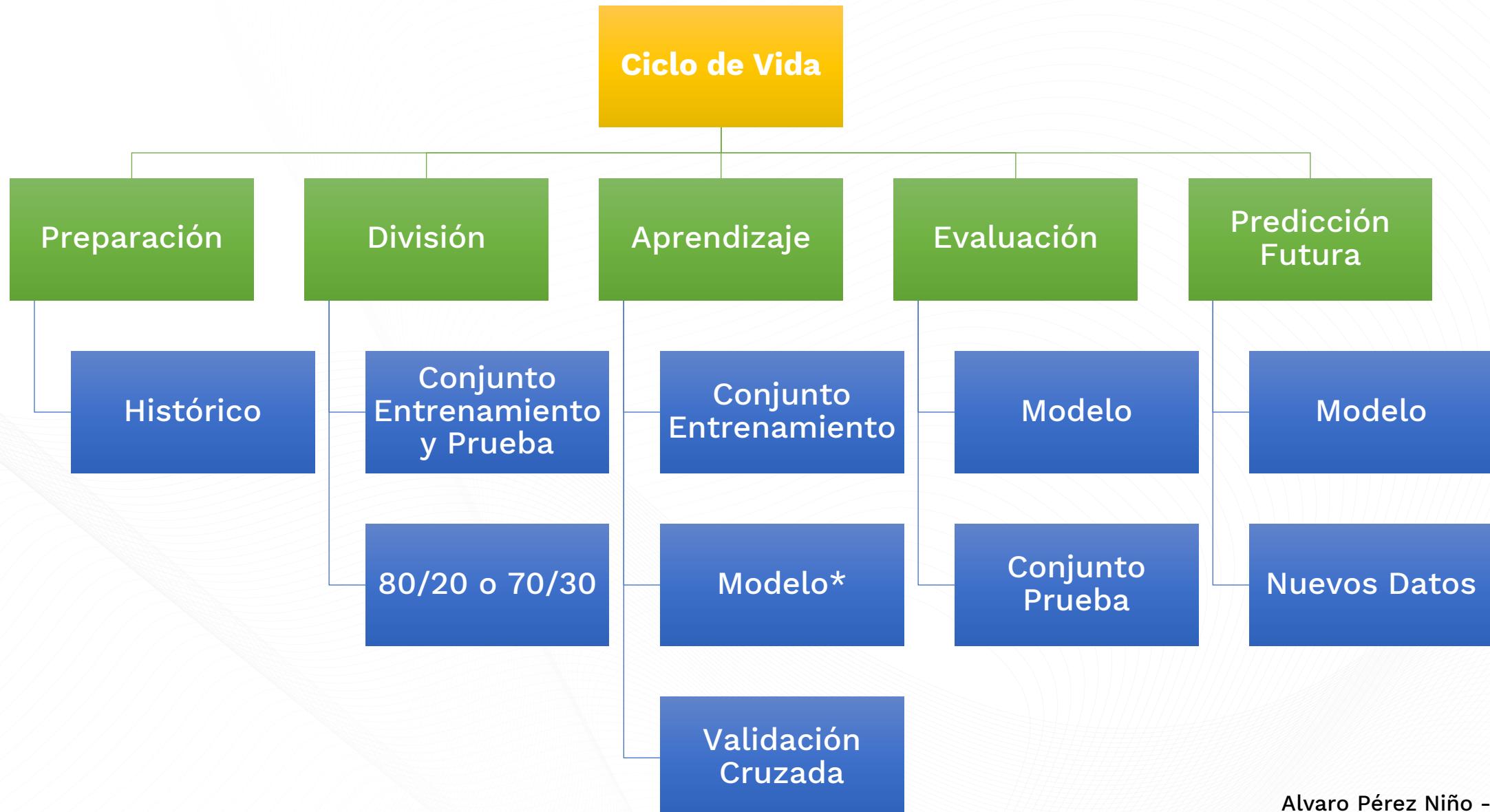


Aprendizaje Automático Supervisado

En el aprendizaje supervisado, el modelo se entrena con datos etiquetados. Cada ejemplo de entrenamiento contiene una entrada y una salida deseada. El modelo aprende a asociar las entradas con las salidas deseadas, y luego puede predecir las salidas para nuevas entradas.

A large, stylized arrow pointing to the right, divided into two sections: blue on the left and green on the right. The word "Clasificación" is written in white in the blue section, and the word "Regresión" is written in white in the green section.

Aprendizaje Supervisado

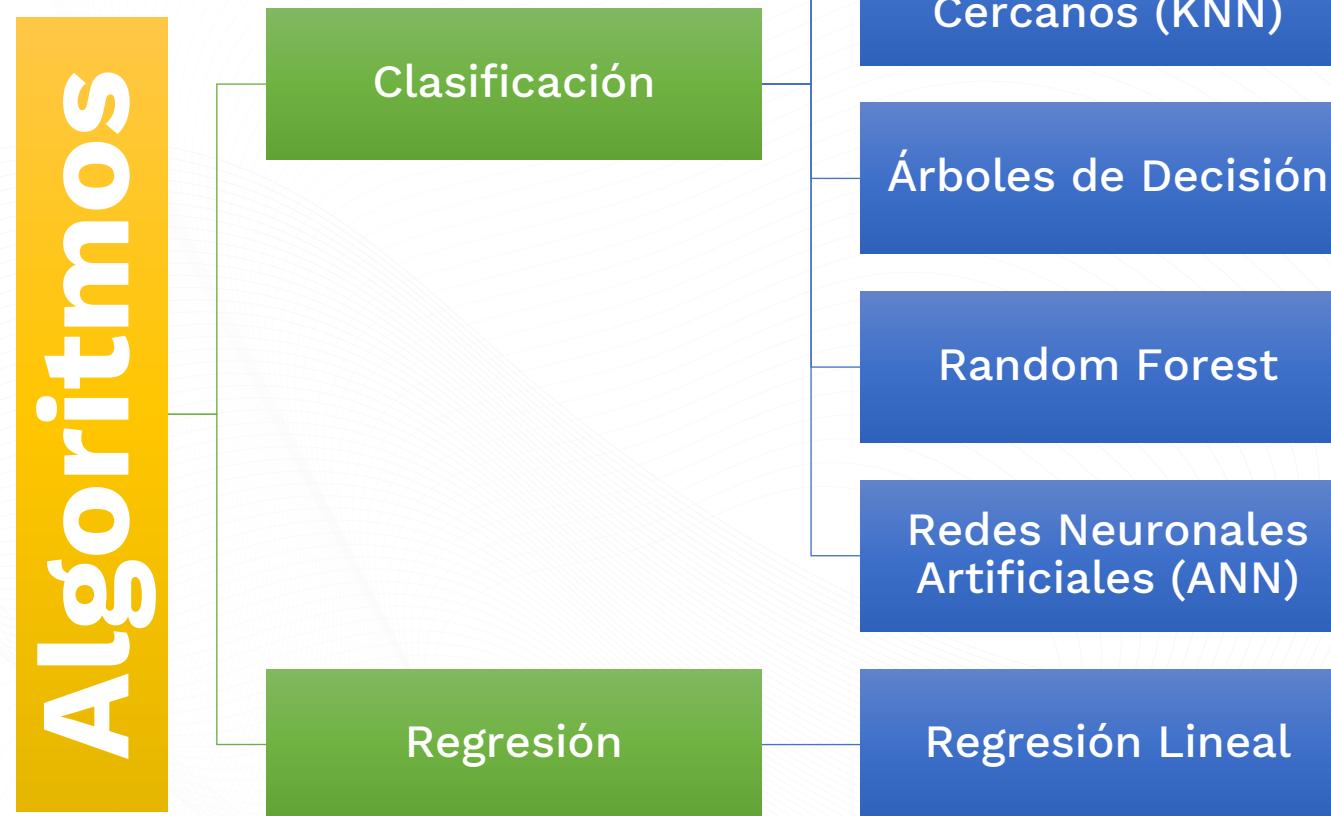


Métodos y algoritmos de Aprendizaje Automático

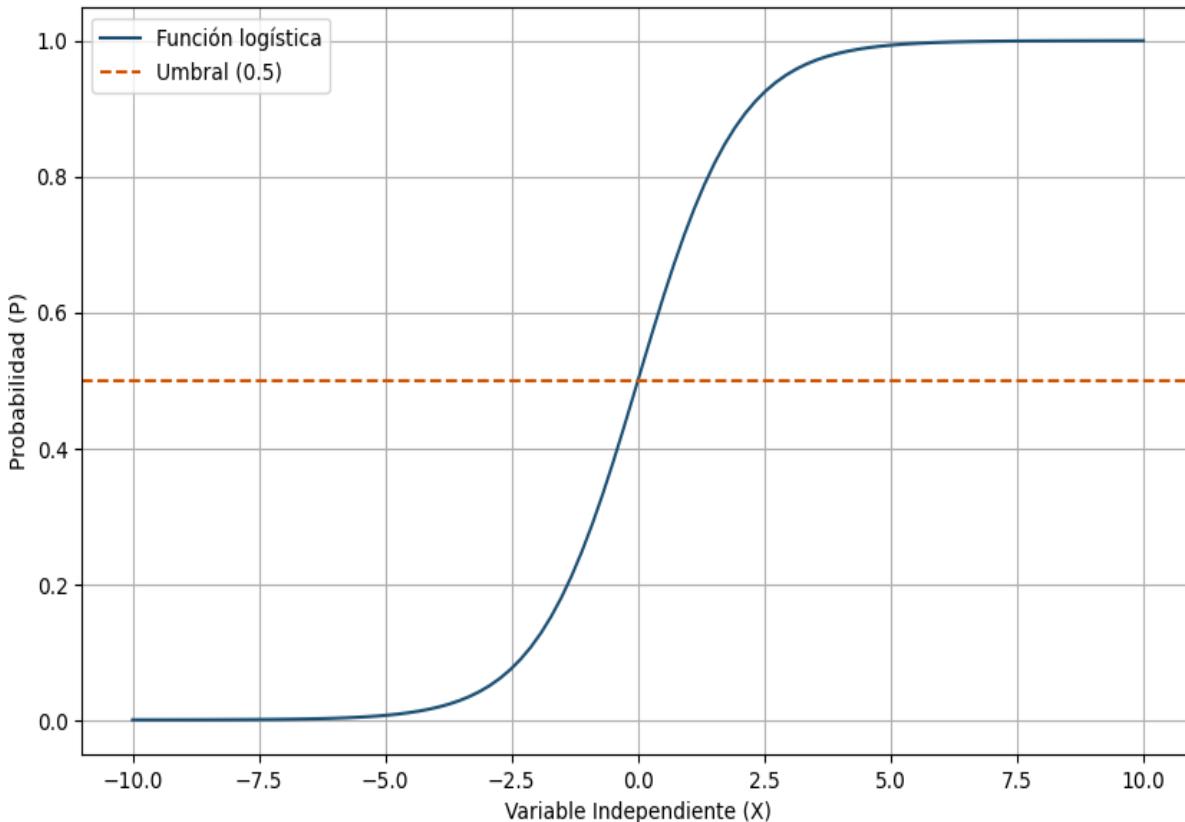
Aprendizaje Supervisado



Algoritmos



Métodos de regresión



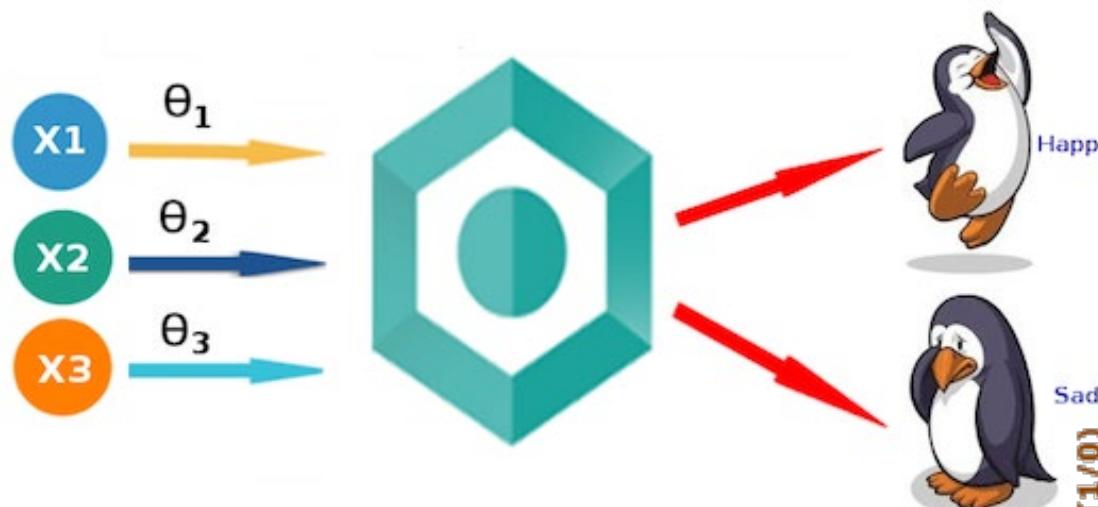
$$P(y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\dots+\beta_nX_n)}} \quad (1)$$

Es un modelo estadístico desarrollado por David Cox (1958), que se basa en la idea de predecir la probabilidad de que una variable dependiente (de tipo binaria) pertenezca a una de dos categorías en función de una o más variables independientes (predictoras).

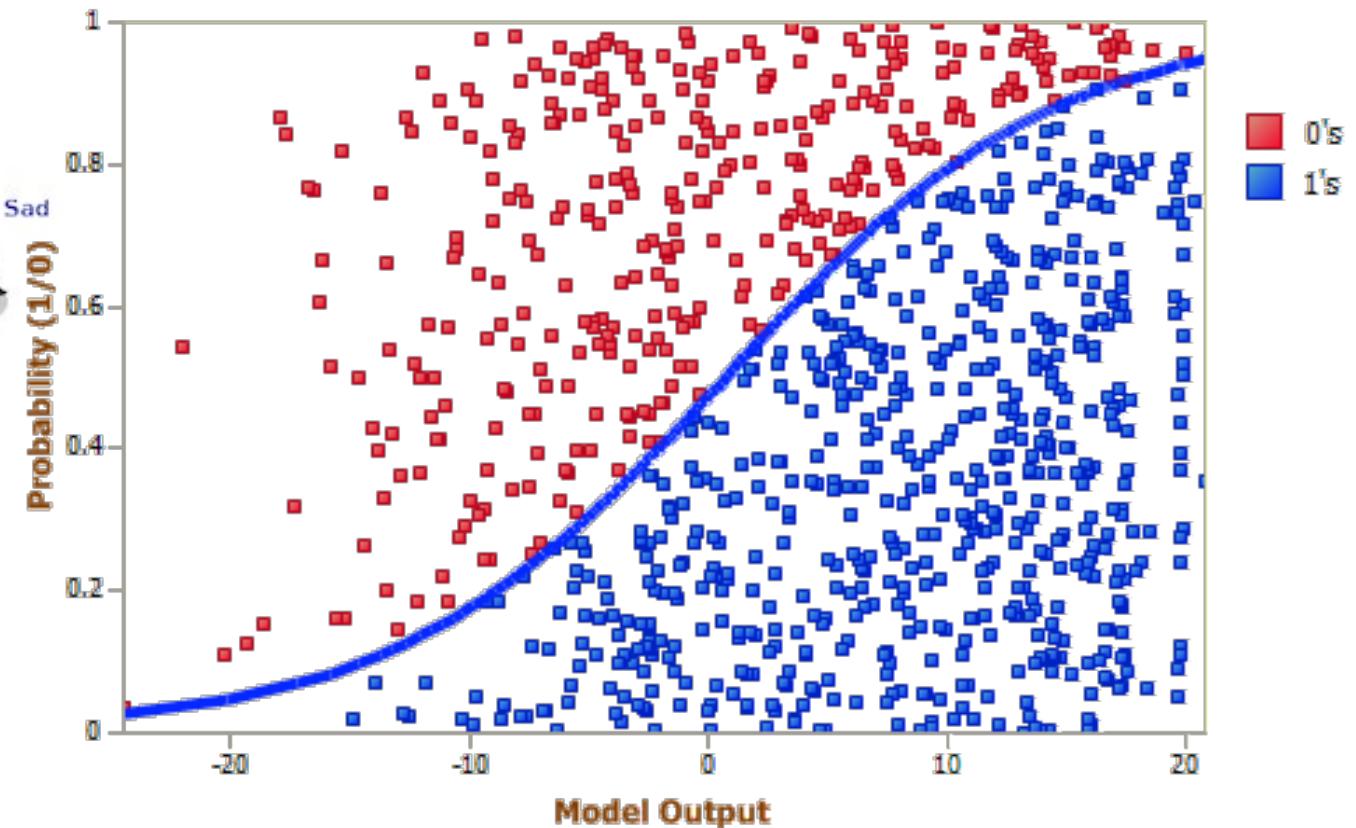
Lineal: Cuando la variable dependiente es continua, y la relación entre las variables es aproximadamente lineal.

Logística: Cuando la variable dependiente es binaria o categórica, y busca predecir la probabilidad de pertenecer a una clase mediante la función sigmoide (logit).

Métodos de regresión

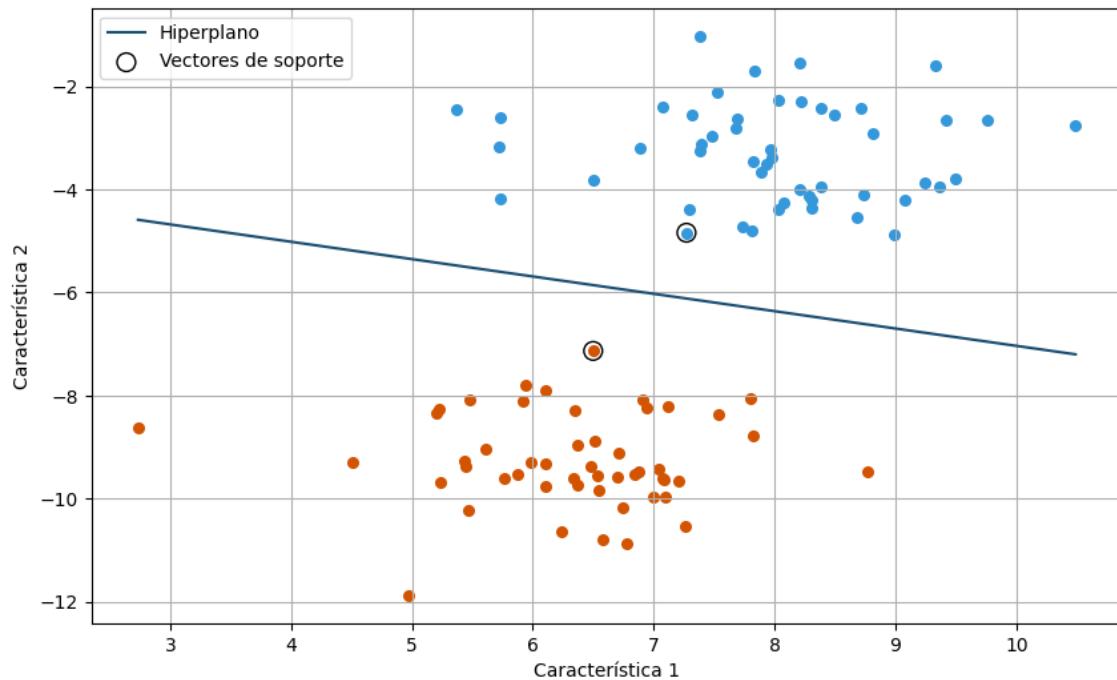


Ejemplo: Mediante las características, demográficas, culturales y educativas. Determinar si una persona será un profesional feliz o aburrido.





Máquina de Vectores de Soporte (SVM, Support Vector Machine)



$$\mathbf{w} \cdot \mathbf{x} - b = 0$$

Es un modelo desarrollado por Vladimir Vapnik (1963), utilizado tanto para clasificación como para regresión.

Su objetivo es encontrar un hiperplano de decisión óptimo que separe las clases dentro de un espacio de alta dimensionalidad, maximizando el margen entre dicho hiperplano y los puntos de datos más cercanos de cada clase, conocidos como vectores de soporte.

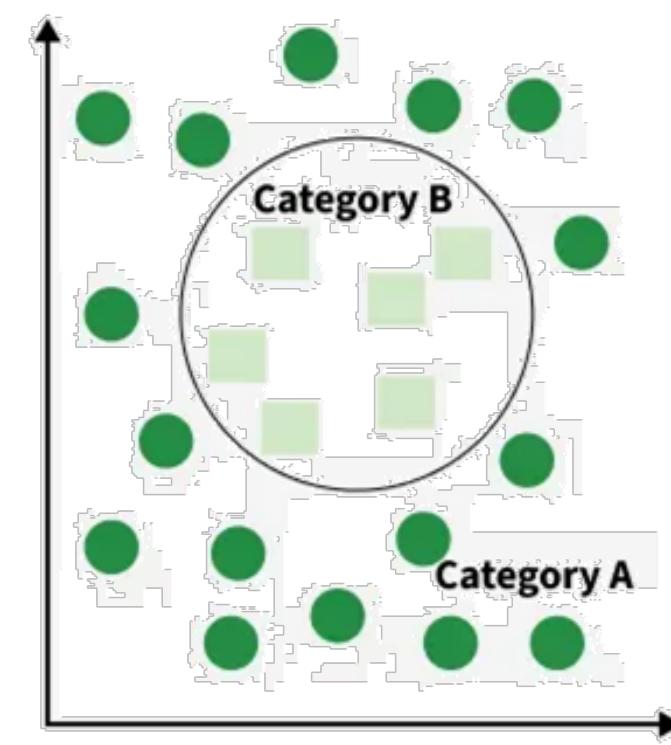
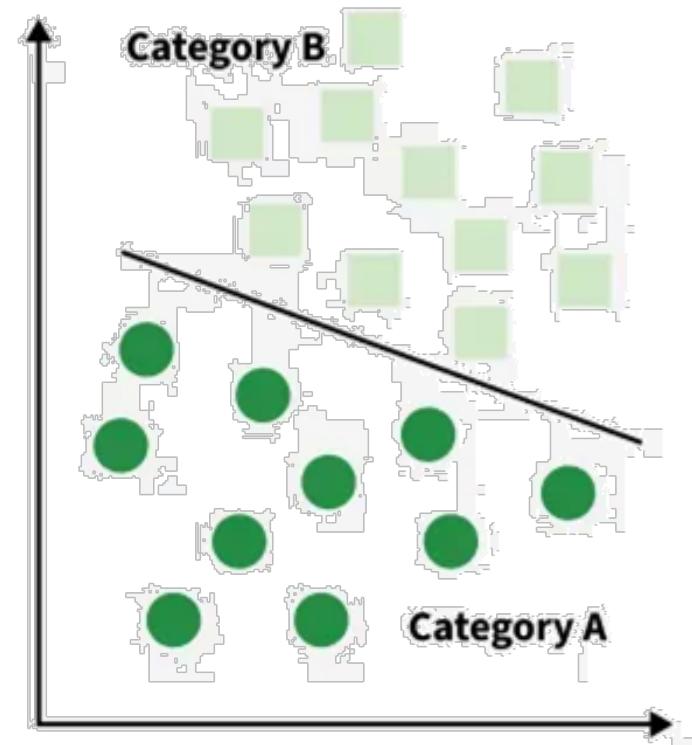
Es especialmente efectiva en problemas no lineales gracias al uso de funciones kernel, que proyectan los datos a espacios donde las clases son separables.



Máquina de Vectores de Soporte (SVM, Support Vector Machine)

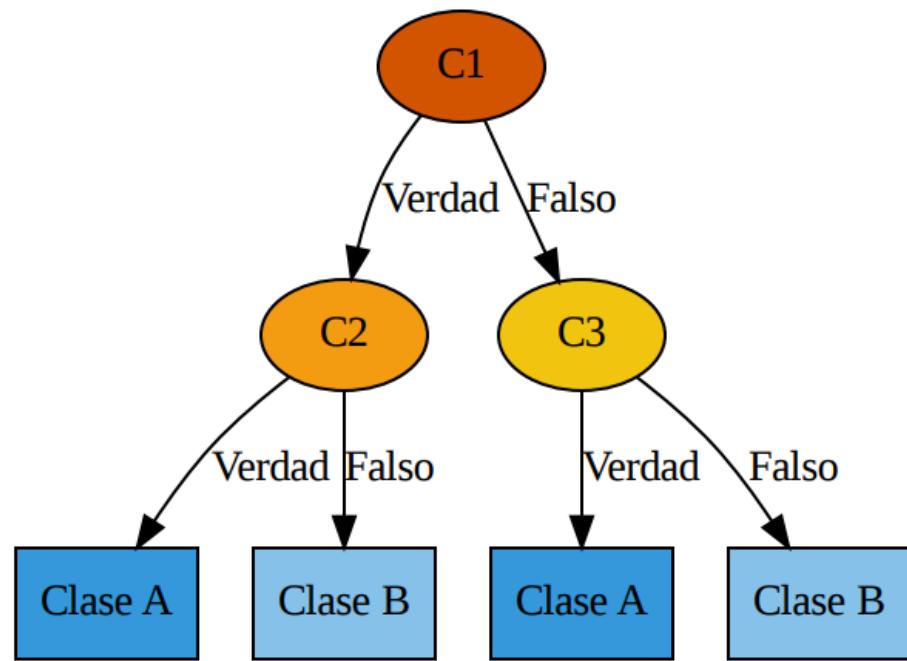


Ejemplo: Una empresa de tecnología desea implementar un sistema que clasifique automáticamente las opiniones de sus clientes en redes sociales como “positivas” o “negativas”.





Árboles de Decisión - DT



Si $X_1 \leq t_1$ entonces seguir la rama izquierda, sino la derecha
Si $X_2 \leq t_2$ entonces seguir la rama izquierda, sino la derecha, y así sucesivamente

Es un modelo desarrollado por J. Ross Quinlan (1986). Este método, divide el espacio de características en regiones más simples y homogéneas, basándose en un conjunto de reglas de decisión.

Cada nodo en el árbol representa una condición o prueba sobre una característica, y cada rama representa el resultado de la prueba.

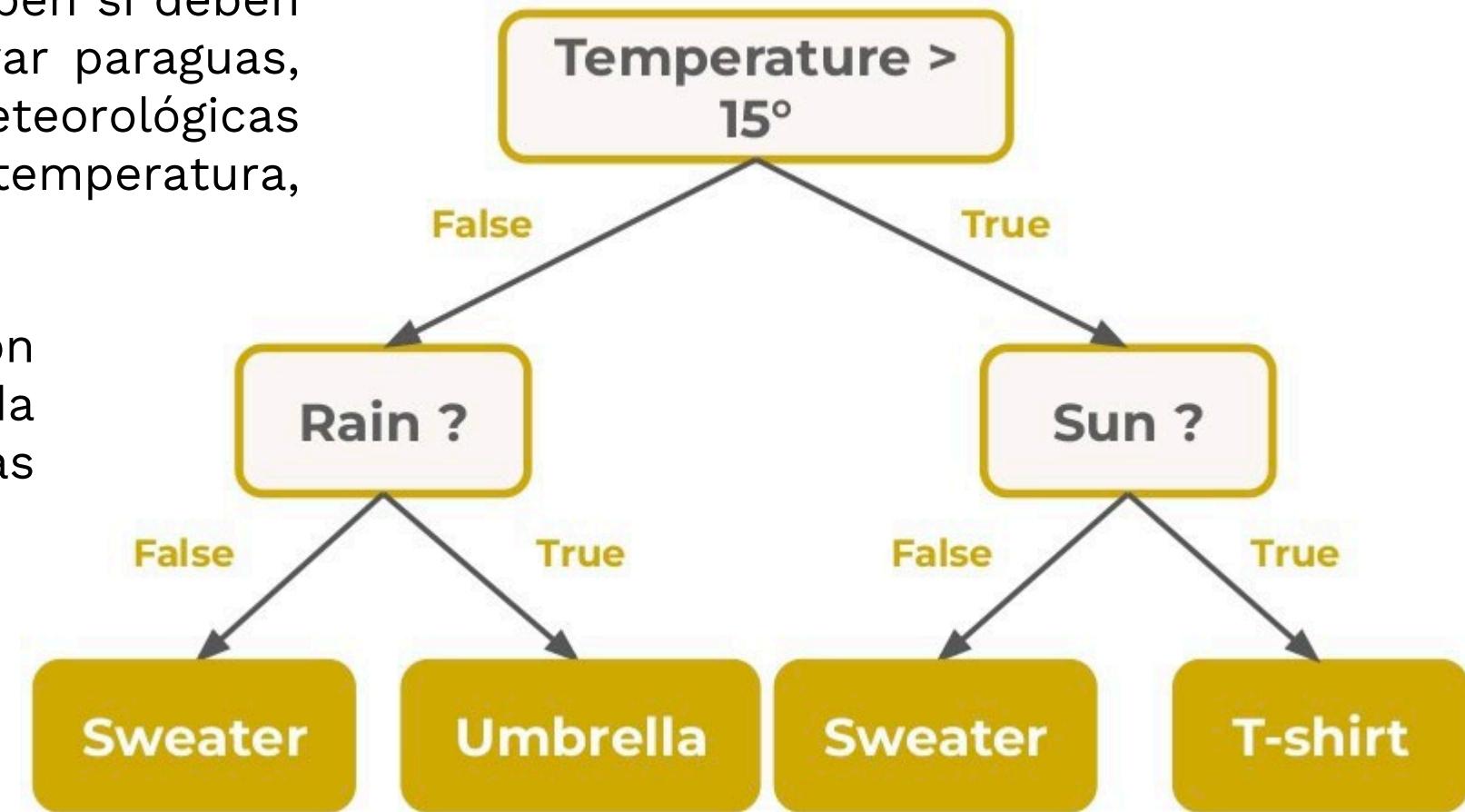
Los nodos hoja indican la clase o valor predicho.



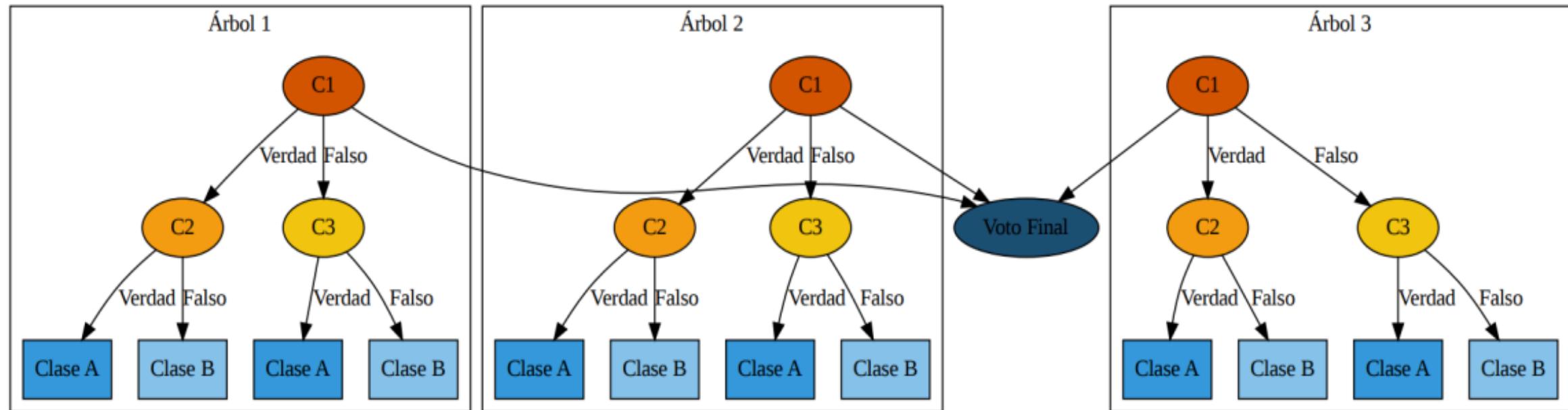
Árboles de Decisión - DT

Ejemplo: Los usuarios no saben si deben usar suéter, camiseta o llevar paraguas, ya que las condiciones meteorológicas cambian constantemente (temperatura, lluvia o sol).

Un sistema de recomendación desea sugerir la prenda adecuada según las condiciones del clima.



Random Forest (RF)



Random Forest en Clasificación

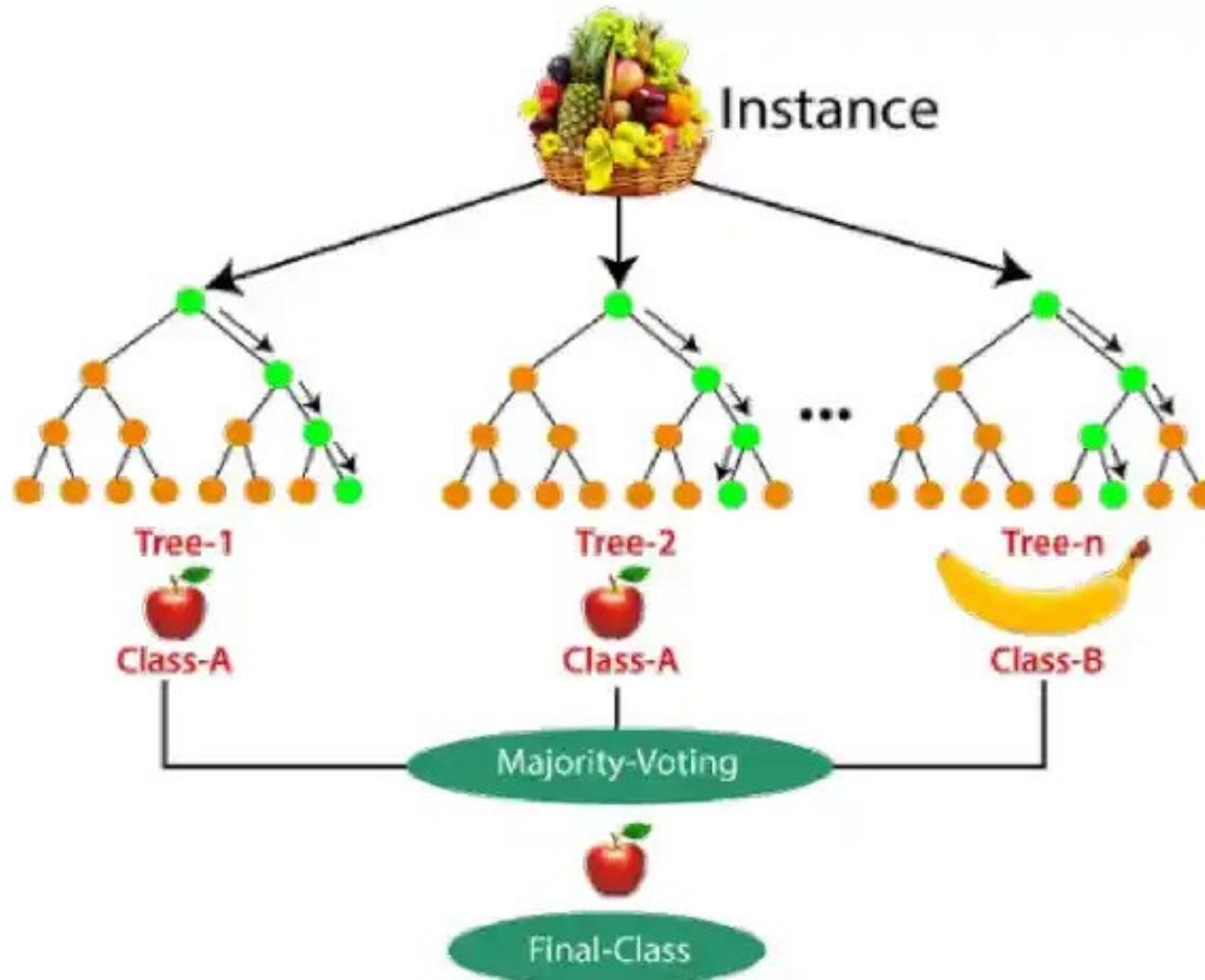
$$\hat{y} = \text{modo}(T_1(x), T_2(x), \dots, T_n(x)) \quad (3)$$

Random Forest en Regresión

$$\hat{y} = \frac{1}{n} \sum_{i=1}^n T_i(x) \quad (4)$$

Tin Kam Ho (1995) → Introduce la idea inicial de Random Decision Forests. Y luego Leo Breiman (2001) → Desarrolla y formaliza el algoritmo Random Forest

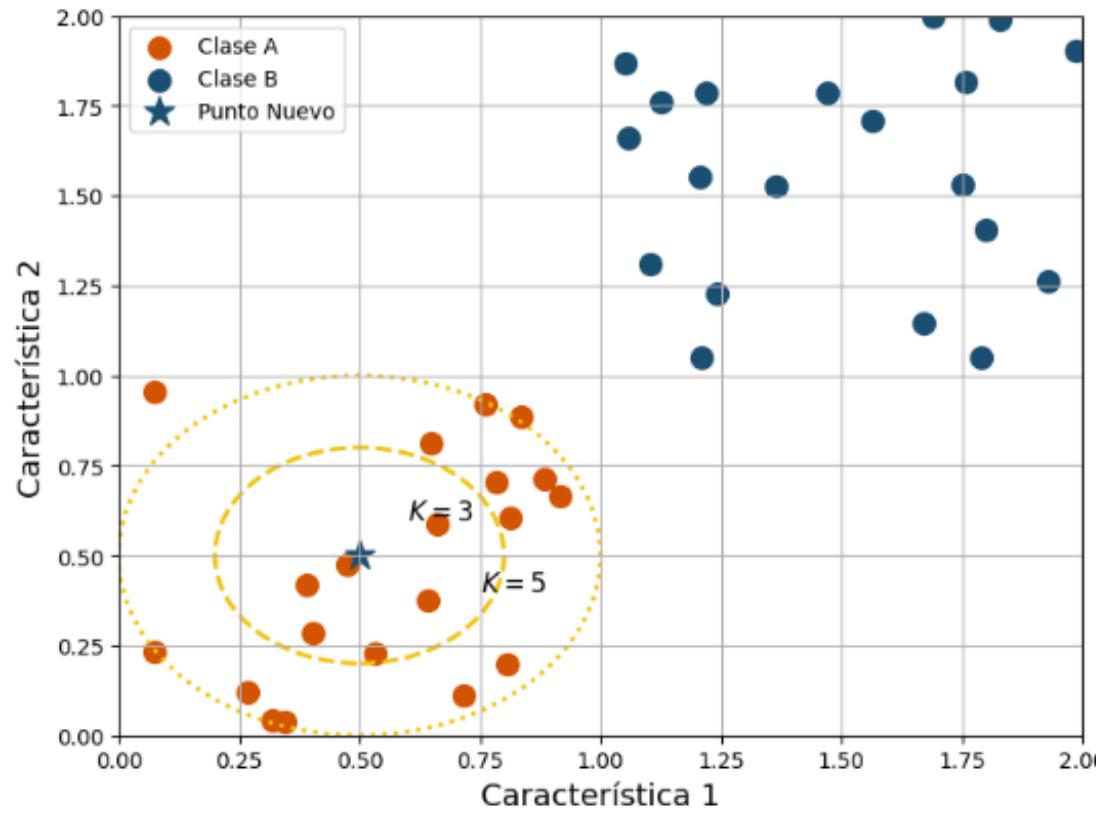
Random Forest (RF)



Ejemplo: Una empresa frutícola desea desarrollar un sistema que clasifique automáticamente frutas (como manzanas, bananos o naranjas) a partir de sus características físicas y visuales, tales como color, peso, forma y textura.

Un solo árbol de decisión puede cometer errores si los datos son ruidosos o si existen frutas con características muy similares. Esto afecta la precisión del sistema y genera clasificaciones equivocadas.

K vecino más cercano (KNN):



$$\hat{y} = \text{modo}(y_1, y_2, \dots, y_K)$$

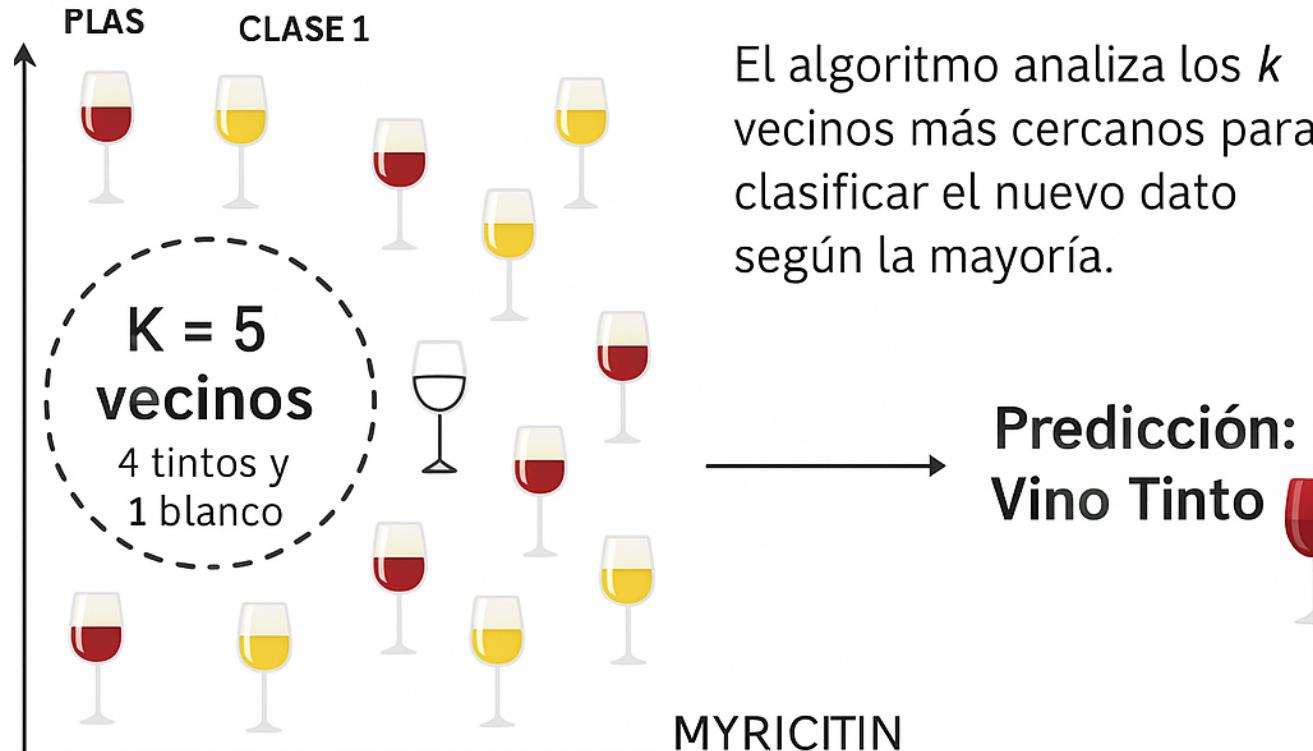
Es un modelo propuesto por Thomas Cover y Peter Hart (1967), es un método no paramétrico utilizado para clasificación y regresión.

Se basa en la idea de que los puntos de datos con características similares tienden a encontrarse próximos entre sí en el espacio de características.

El algoritmo calcula la distancia (euclídea) entre el punto a clasificar y los datos existentes, y asigna la categoría o valor promedio de los k vecinos más cercanos.



K vecino más cercano (KNN):

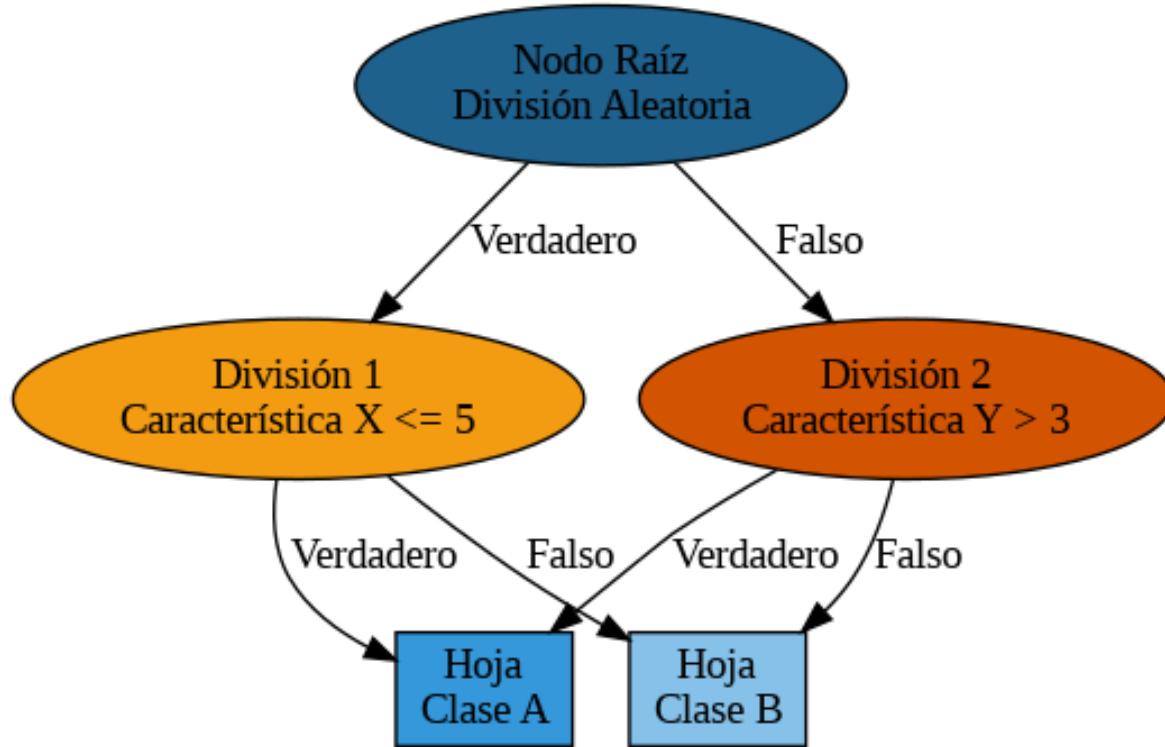


El algoritmo analiza los k vecinos más cercanos para clasificar el nuevo dato según la mayoría.

Ejemplo: El nuevo vino aún no tiene etiqueta que indique su tipo, y los valores de sus propiedades químicas se asemejan tanto a los vinos tintos como a los blancos.

Por esta razón, el laboratorio enológico busca clasificar el nuevo vino, utilizando medidas obtenidas en el análisis químico, como el contenido de Rutina y Myricetina, para determinar a qué categoría pertenece.

Extra Trees Classifier

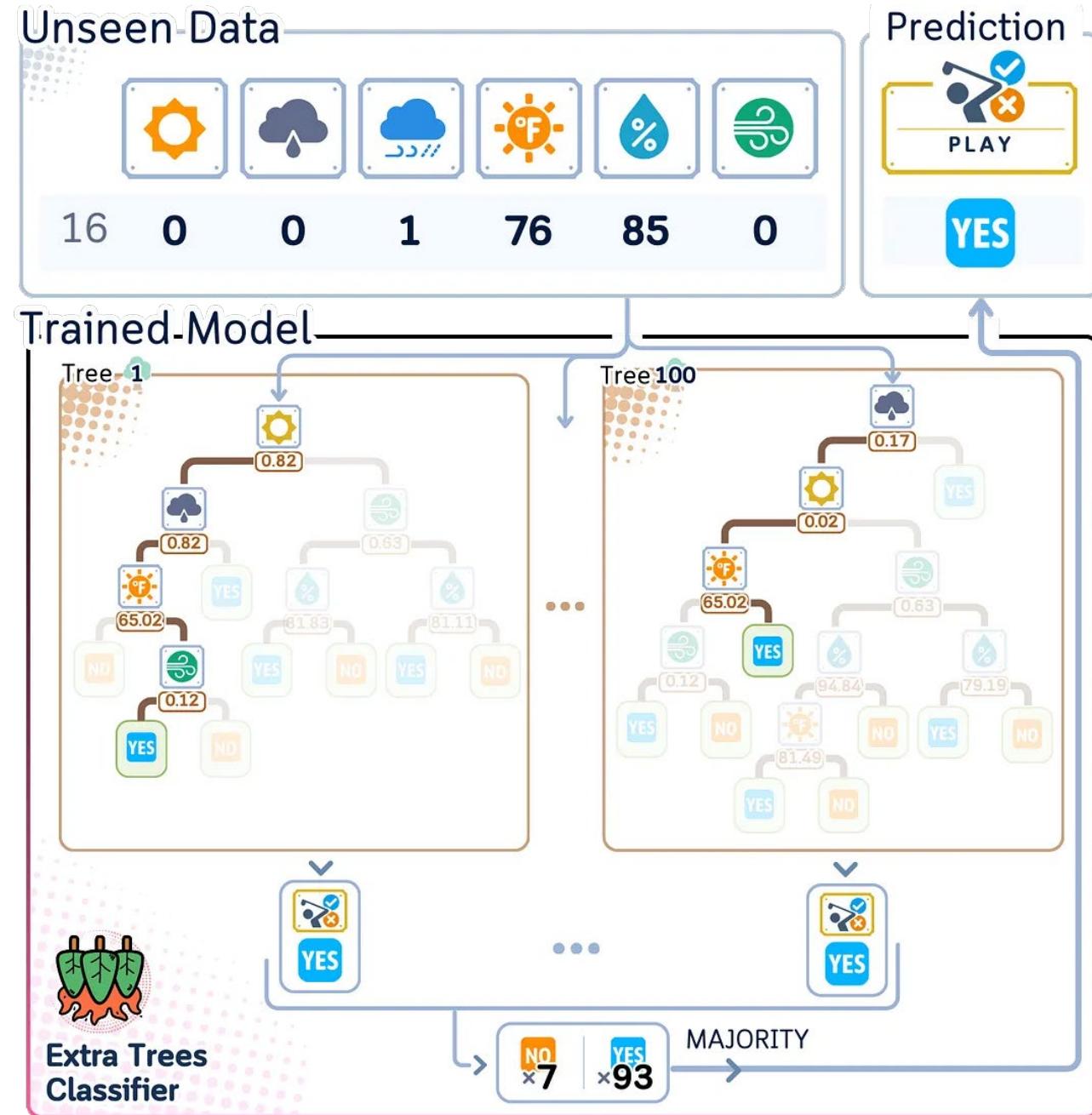
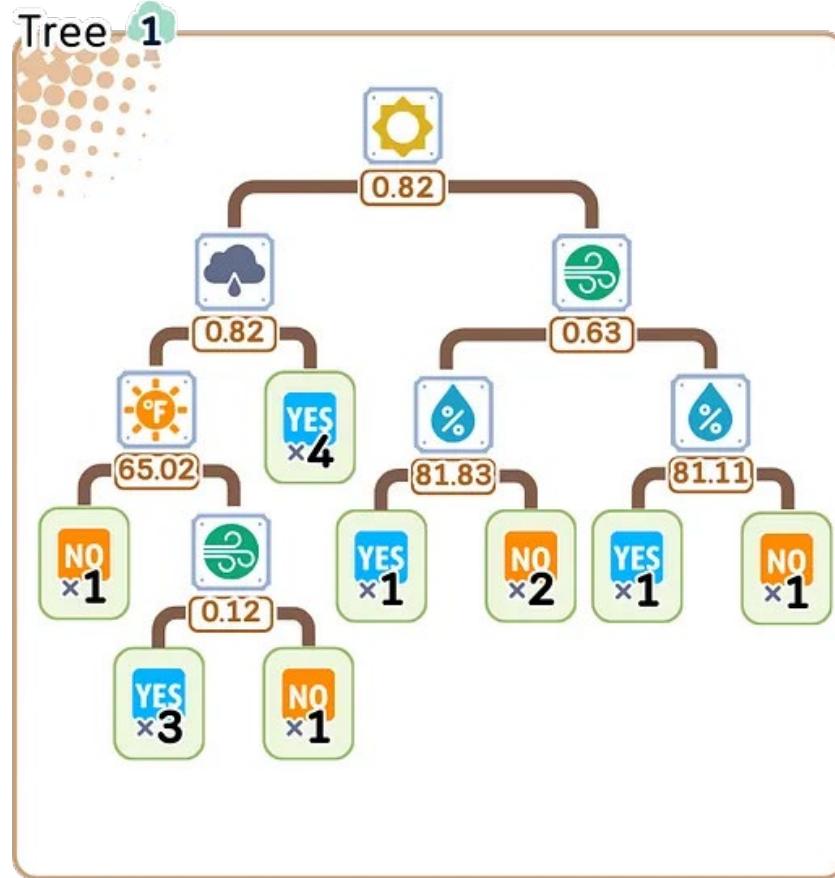


Extra Trees Classifier (Extremely Randomized Trees), propuesto por Pierre Geurts, Damien Ernst y Louis Wehenkel (2006)

Es un modelo de ensamble basado en múltiples árboles de decisión. A diferencia del Random Forest, este algoritmo introduce una mayor aleatoriedad al seleccionar tanto las muestras como los puntos de división, lo que reduce la varianza y mejora la generalización del modelo.

Su funcionamiento se basa en combinar los resultados de varios árboles para obtener una predicción final más estable y precisa.

Extra Trees Classifier

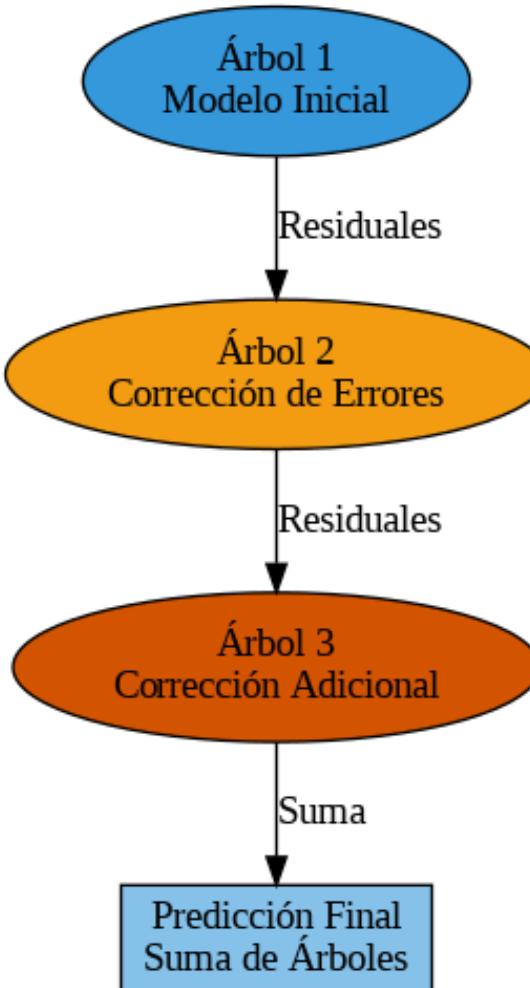


Extra Trees
Classifier

MAJORITY
NO x 7 YES x 93



Extreme Gradient Boosting

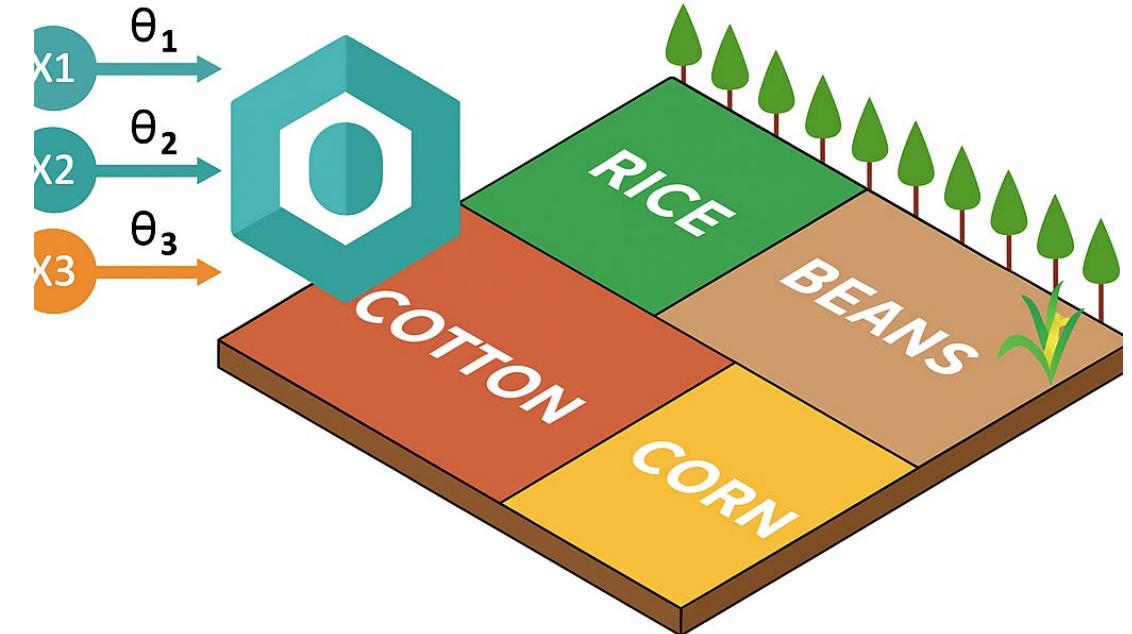
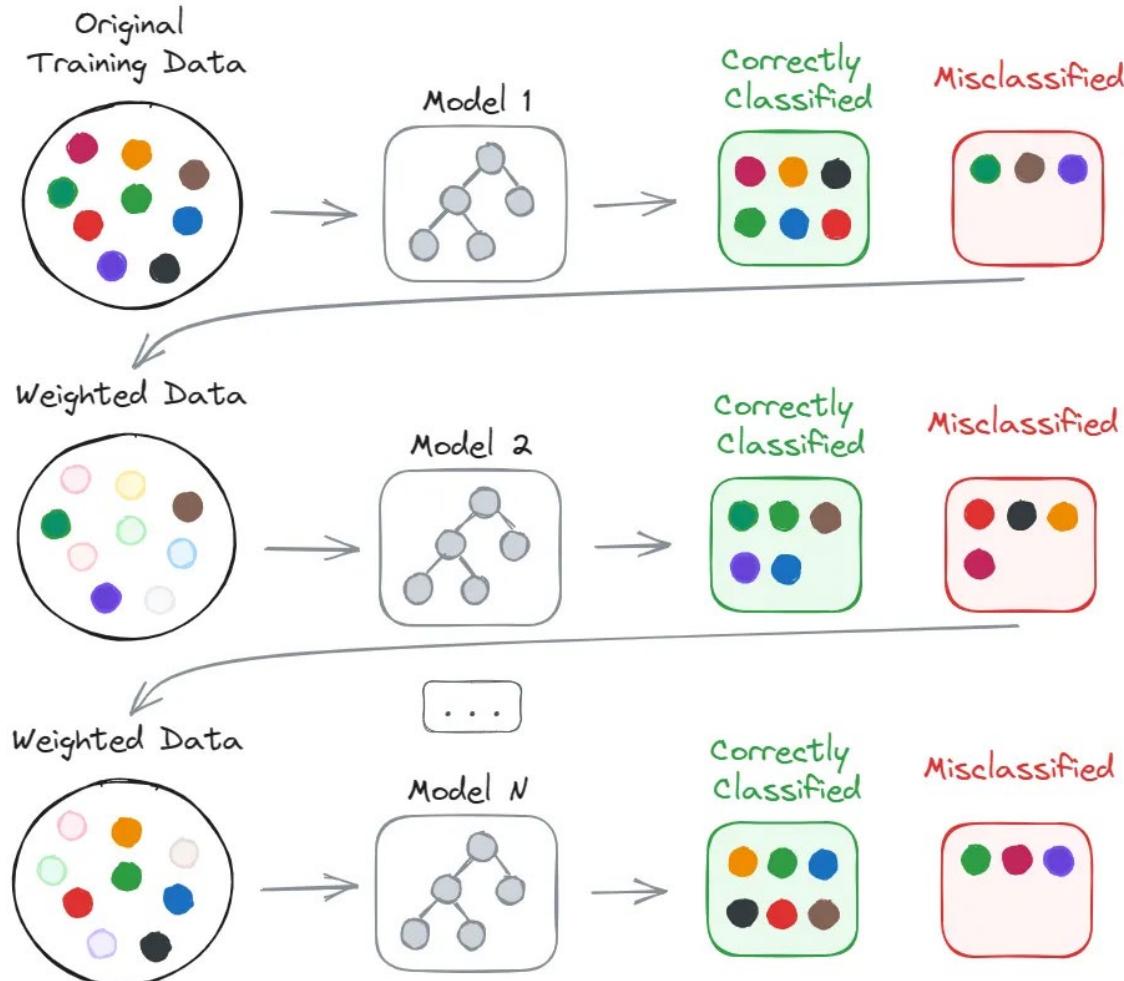


El Extreme Gradient Boosting (XGBoost), desarrollado por Tianqi Chen y Carlos Guestrin (2016), es un algoritmo de ensamble que utiliza la técnica de boosting para mejorar el rendimiento de los modelos de árboles de decisión.

Construye los árboles de forma secuencial, donde cada nuevo árbol corrige los errores cometidos por los anteriores.

Se caracteriza por su alta eficiencia, capacidad de regularización (para evitar sobreajuste) y excelente desempeño en tareas de clasificación y regresión.

Extreme Gradient Boosting



¿Cuál es el mejor cultivo para sembrar en un terren?



G R A C I A S

Presentó: Alvaro Pérez Niño

Instructor Técnico

Correo: aperezn@misena.edu.co

<http://centrodesseriviciosygestionempresarial.blogspot.com/>

Línea de atención al ciudadano: 01 8000 910270

Línea de atención al empresario: 01 8000 910682



www.sena.edu.co