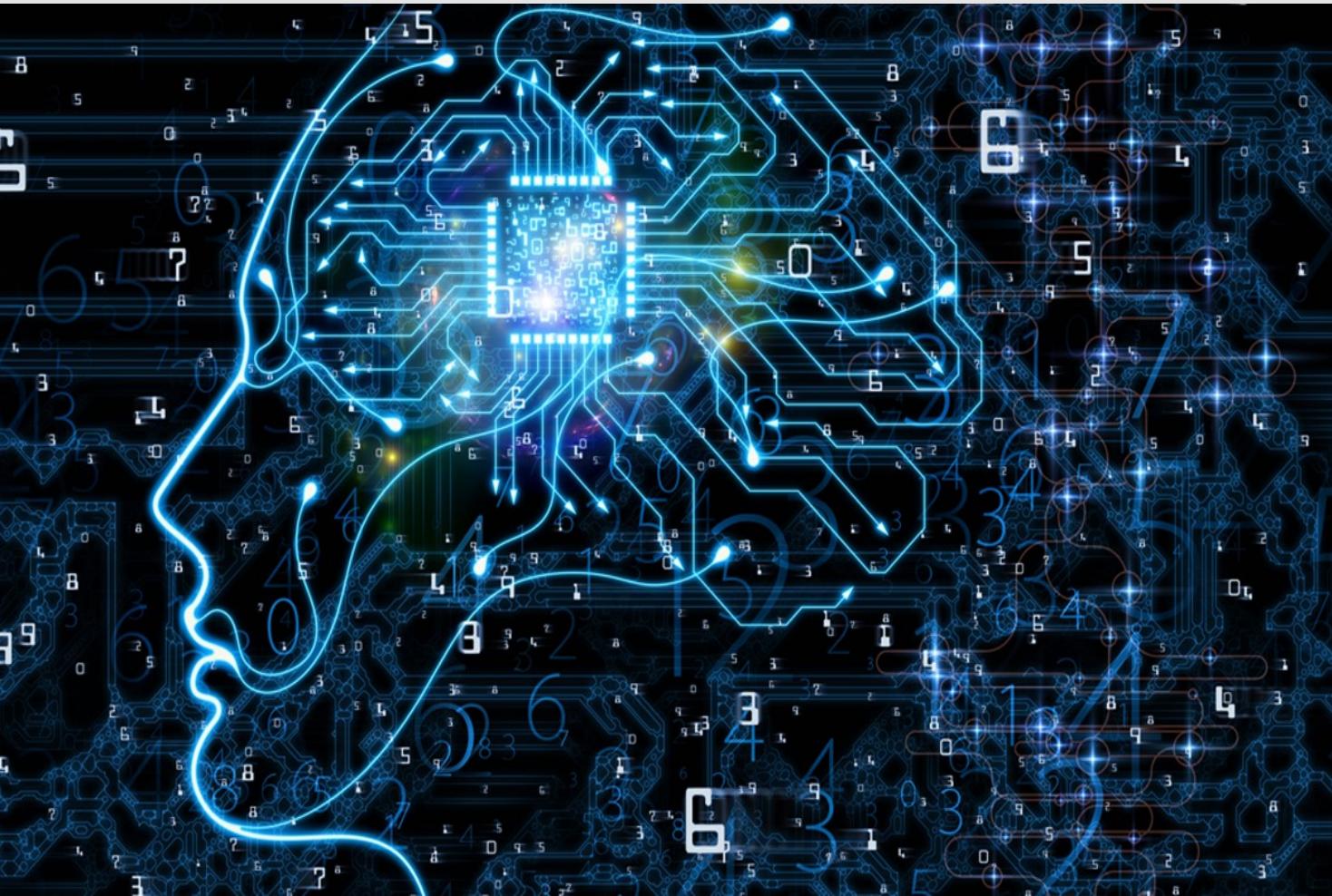


Introducción a Machine Learning Sesión 1

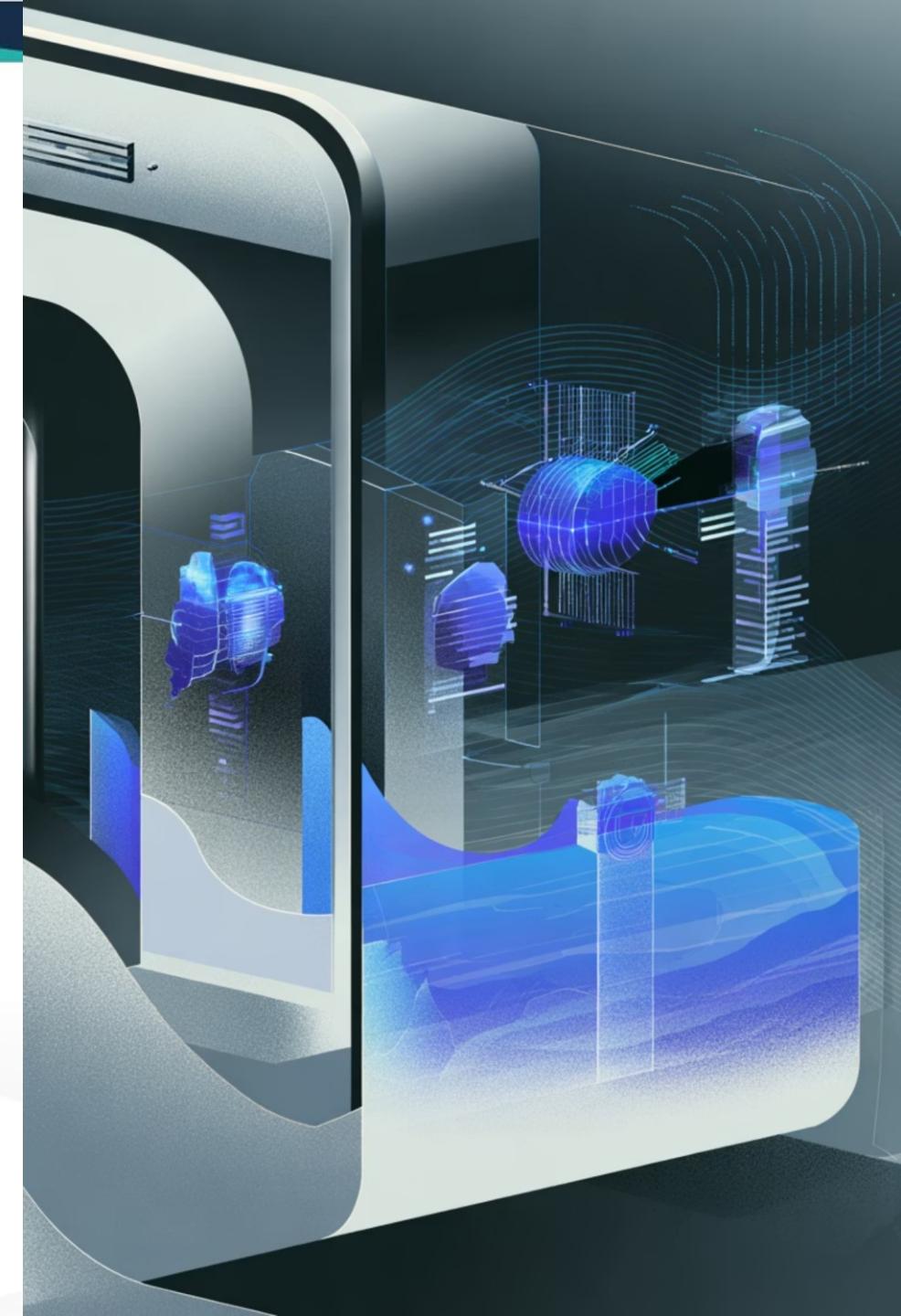


**CORPORACIÓN UNIVERSITARIA
DEL HUILA - CORHUILA**
"Diseño y prestación de servicios de
docencia, investigación y extensión
de programas de pregrado"



Introducción a ML

- Machine Learning
- Aprendizaje supervisado.
- Algoritmos de clasificación
- EDA
- Preprocesamiento
- Naive Bayes
- Árbol de Decisión
- Gradient Boosting Classifier.



Objetivo de Aprendizaje



Meta Principal

Identificar los conceptos de aprendizaje supervisado para el desarrollo de algoritmos de clasificación binaria



Conocimientos Clave

Dominar técnicas de clasificación y preprocessamiento de datos



Aplicación Práctica

Implementar algoritmos de Machine Learning en problemas reales



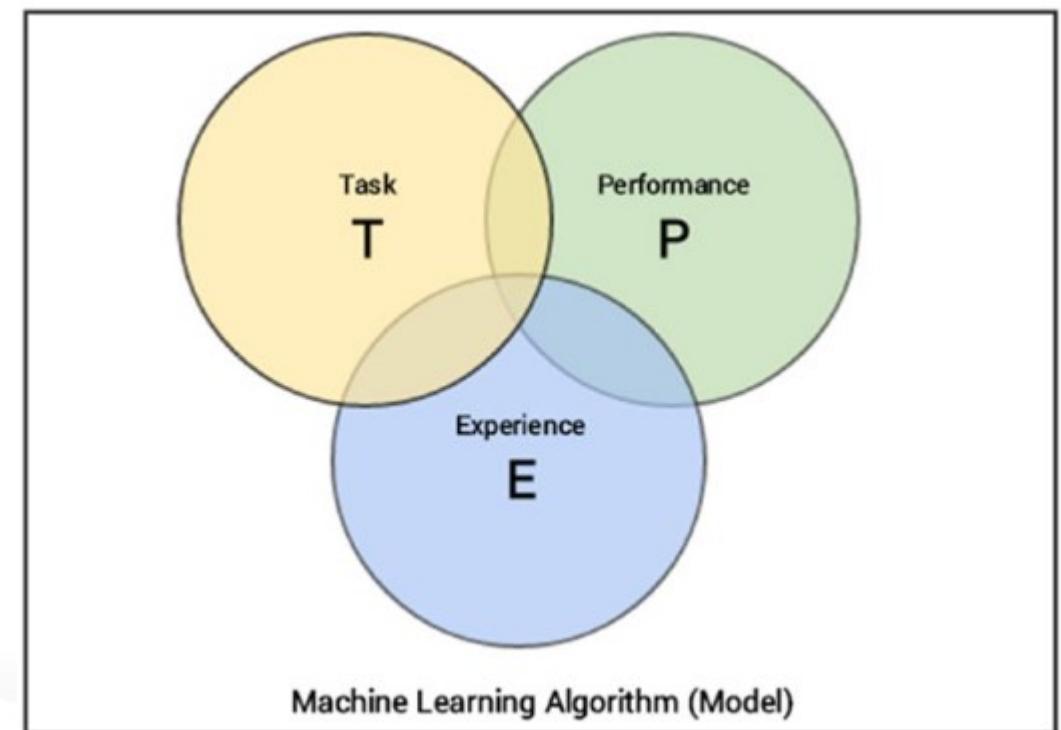
**CORPORACIÓN UNIVERSITARIA
DEL HUILA - CORHUILA**
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado"

¿Qué es Machine Learning?

La idea de Machine Learning es que habrá algún algoritmo de aprendizaje que ayudará a la máquina aprender de los datos. El profesor Mitchell lo definió de la siguiente manera.

"Se dice que un programa informático aprende de la experiencia E con respecto a alguna clase de tareas T y una medida de rendimiento P, si su rendimiento en las tareas de T, medido por P, mejora con la experiencia E".

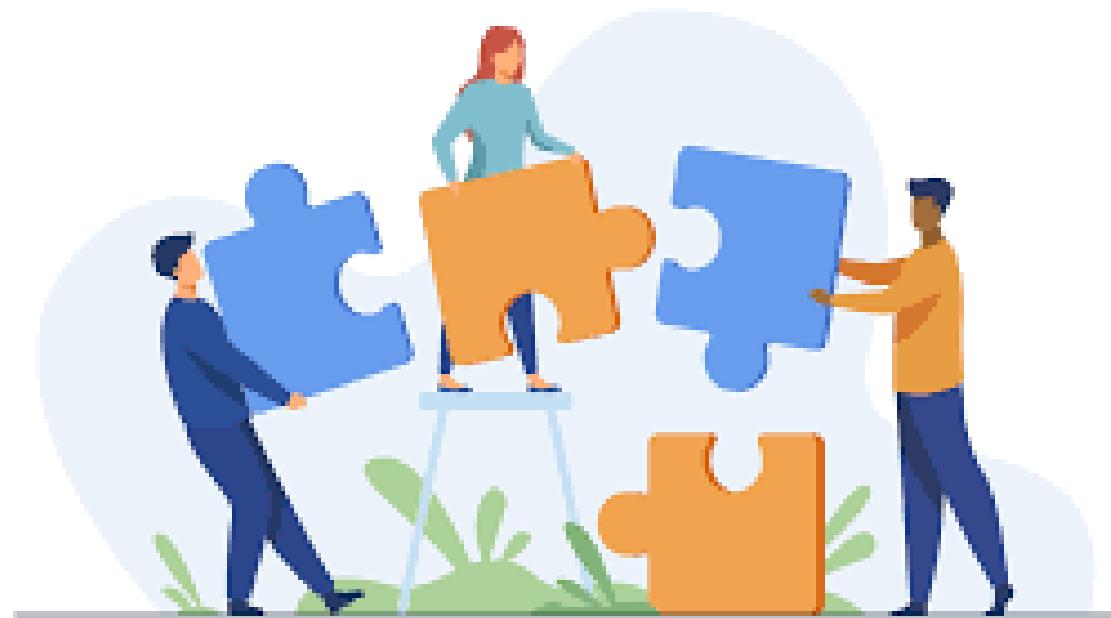
Aunque esta definición puede parecer desalentadora al principio, le pido que la lea un par de veces lentamente centrándose en los tres parámetros -T, P y E- que son los principales componentes de cualquier algoritmo de aprendizaje.



Definir la Tarea Específica

La tarea, T , que puede definirse con un doble enfoque. Desde el punto de vista del problema, la tarea, T , es básicamente el problema del mundo real que hay que resolver, que puede ser cualquier cosa, desde encontrar la mejor combinación de marketing o de productos hasta predecir los fallos de la infraestructura.

En el mundo del aprendizaje automático, lo mejor es definir la tarea de la forma más concreta posible, de manera que problema exacto que se pretende resolver y cómo se podría definir o formular el problema en una tarea específica de aprendizaje automático.



Ejemplos de Tareas

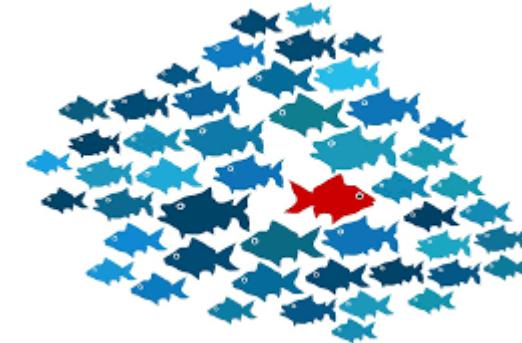
Clasificación



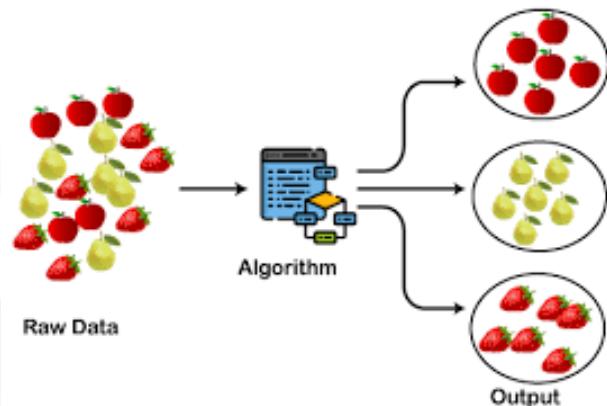
Predicción



Detección Anomalías



Clustering o agrupaciones



Procesamiento de Lenguaje Natural



Experiencia (Entrenamiento)

La idea de que un modelo o algoritmo adquiera *experiencia E* suele producirse como un proceso iterativo, también conocido como entrenamiento del modelo. Se podría pensar que el modelo es una entidad como un ser humano que adquiere conocimientos o experiencia a través de los puntos de datos, observando y aprendiendo más y más sobre los diversos atributos, relaciones y patrones presentes en los datos. Por supuesto, hay varias formas y maneras de aprender y adquirir experiencia, como el aprendizaje supervisado, el no supervisado y el de refuerzo.



Rendimiento (Performance)

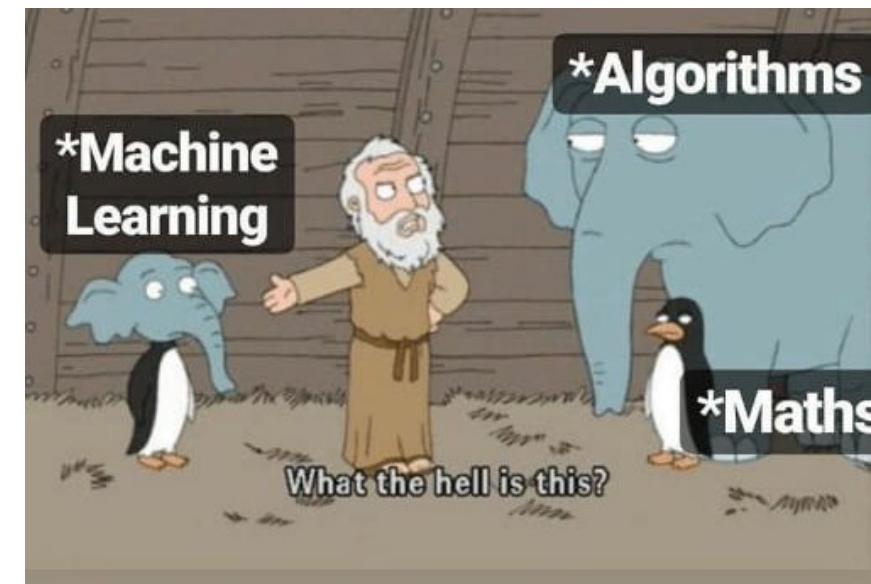
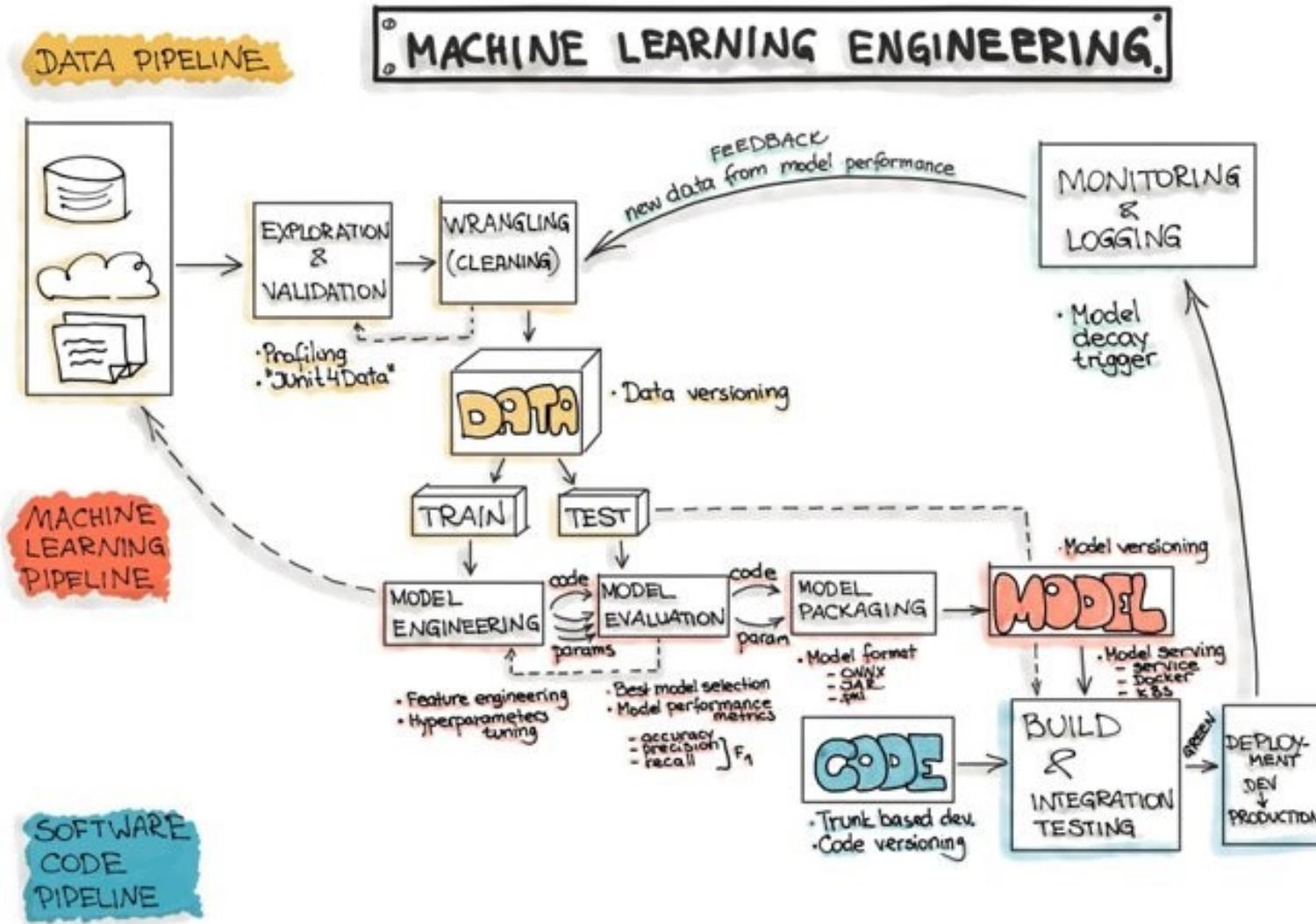
Supongamos que tenemos un algoritmo de aprendizaje automático que debe realizar una tarea, T, y que va ganando experiencia, E, con puntos de datos durante un período de tiempo. Pero, ¿cómo sabemos si está actuando bien o no?

¿Cómo se comporta? Aquí es donde entra en juego el rendimiento, P, del modelo.

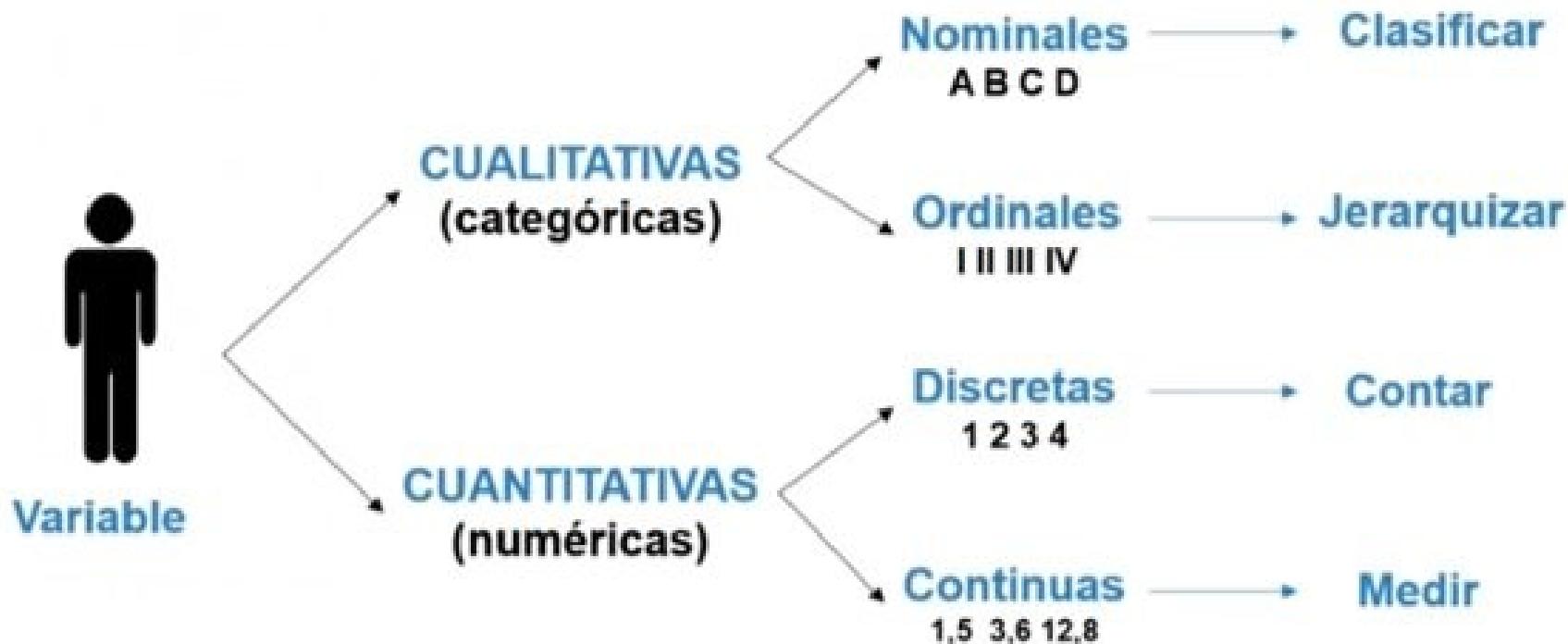
El rendimiento, P, suele ser una medida cuantitativa o métrica que se utiliza para ver lo bien que el algoritmo o modelo realiza la tarea, T, con experiencia, E. Aunque las métricas de rendimiento suelen ser métricas estándar que se han establecido tras años de investigación y desarrollo, cada métrica suele calcularse específica para la tarea, T, que se intenta resolver en un momento dado.



Ciclo de vida de ML



Tipos de Variables - Estadística



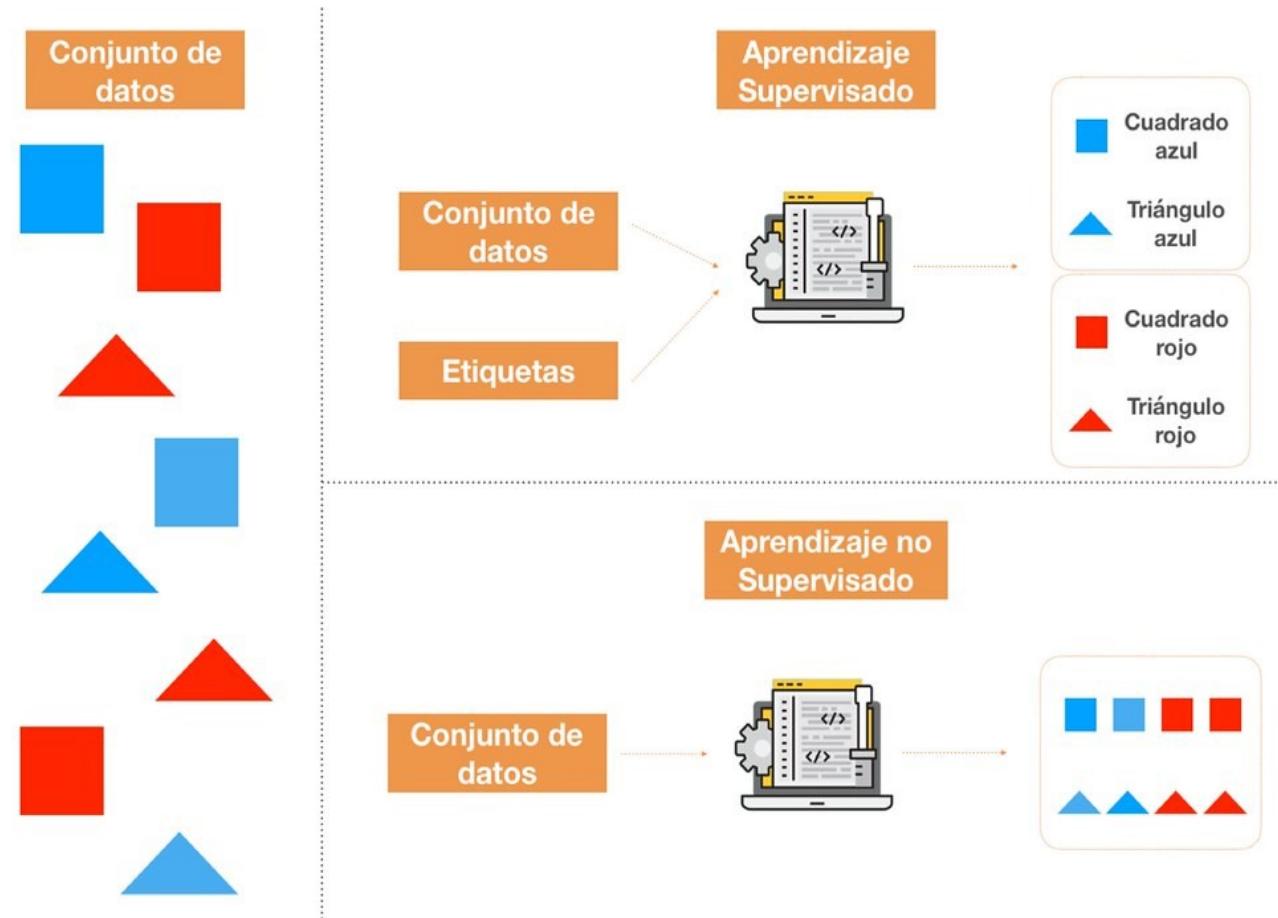
Tipos de Algoritmos de ML

Aprendizaje Supervisado

Técnica de Machine Learning donde se **entrena un modelo** utilizando un conjunto de datos etiquetado. Cada ejemplo de entrenamiento consta de una **entrada** (características o atributos) y **una etiqueta o resultado deseado**.

Objetivo: función que pueda mapear las entradas a las etiquetas o resultados correctos, permitiendo hacer predicciones precisas sobre nuevas entradas.

- Datos etiquetados con relaciones conocidas
- Aprendizaje de patrones en los datos
- Evaluación mediante métricas de rendimiento

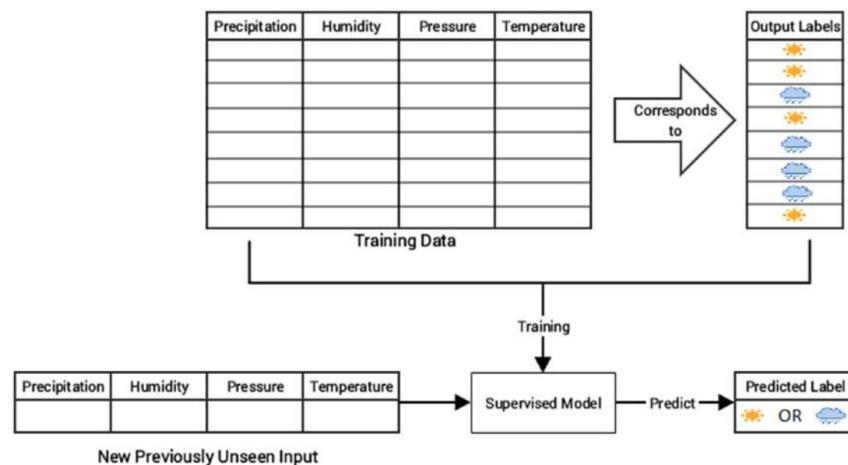


Aspectos Clave

Aplicaciones del Aprendizaje Supervisado

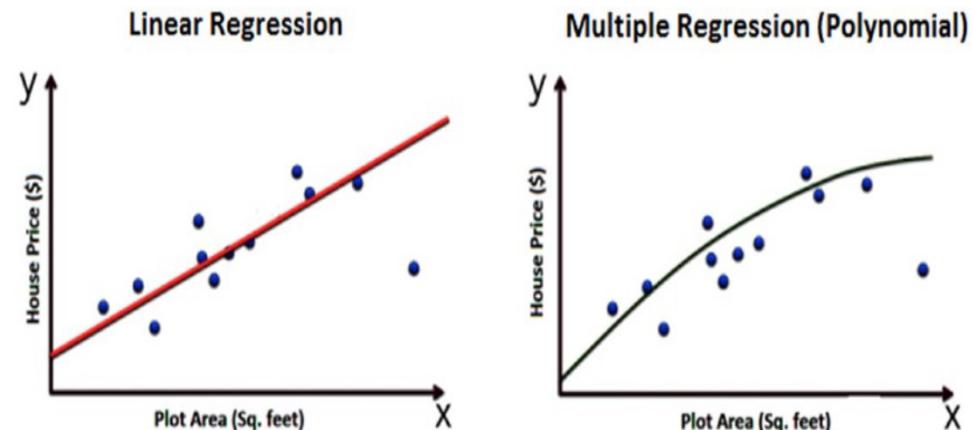
Clasificación

- Asignar etiquetas o categorías a entradas. Ejemplos: clasificar correos como spam o no spam, identificar especies de animales en imágenes.



Regresión

- Predecir valores numéricos continuos. Ejemplos: predecir el precio de una casa según tamaño, ubicación y características.



El aprendizaje supervisado es una de las técnicas más utilizadas en ML debido a su versatilidad y aplicabilidad en problemas del mundo real, desde clasificación hasta predicción y toma de decisiones.

Algoritmos de Clasificación

Los algoritmos de clasificación son técnicas de Machine Learning que asignan una etiqueta o categoría a un elemento de entrada basándose en sus características. Predicen la clase a la que pertenece el elemento usando patrones aprendidos de datos de entrenamiento etiquetados.

Correo Electrónico

Determinar si es spam o no spam

Imágenes

Identificar objetos como gatos, perros o automóviles

Documentos

Categorizar en diferentes temas

Diagnóstico Médico

Predecir enfermedades según síntomas

Recomendaciones

Sugerir productos según preferencias



Algoritmos Comunes de Clasificación

1

Máquinas de Soporte Vectorial (SVM)

Útiles para clasificación binaria y pueden manejar datos no lineales de manera efectiva.

2

Árboles de Decisión

Representan decisiones y pueden manejar datos categóricos y numéricos simultáneamente.

3

Naive Bayes

Basado en el teorema de Bayes, ideal para clasificación de texto y documentos.

4

Random Forest

Conjunto de árboles de decisión que combina múltiples modelos para mejorar la precisión.

La elección del algoritmo depende del problema específico, la naturaleza de los datos y los objetivos del aprendizaje automático.



**CORPORACIÓN UNIVERSITARIA
DEL HUILA - CORHUILA**
"Diseño y prestación de servicios de
docencia, investigación y extensión
de programas de pregrado"



Análisis Exploratorio de Datos (EDA)

EDA es una etapa crucial que aplica técnicas estadísticas y visuales para examinar datos, identificar patrones, tendencias, relaciones y posibles anomalías.

Propósitos Principales

- Comprender estructura y naturaleza de los datos
- Detectar anomalías y valores atípicos
- Explorar distribuciones de variables
- Identificar relaciones y correlaciones
- Seleccionar técnicas de modelado apropiadas

Beneficios Clave

- Visualización clara de datos
- Identificación de valores faltantes
- Selección de características relevantes
- Validación de suposiciones iniciales
- Resumen de hallazgos clave

Preprocesamiento de Datos

El preprocesamiento es una fase crítica que implica preparar y limpiar datos brutos antes de análisis avanzados o construcción de modelos predictivos. Es esencial para garantizar que los datos estén en condiciones óptimas y los resultados sean confiables.

01

Limpieza de Datos

Identificar y tratar valores atípicos y valores faltantes que pueden distorsionar el análisis.

04

Manejo de Desbalances

Abordar desequilibrios en las clases que afectan el aprendizaje del modelo.

02

Transformación

Normalización, estandarización y codificación de variables categóricas en formatos numéricos.

05

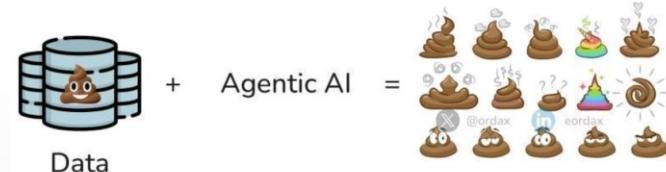
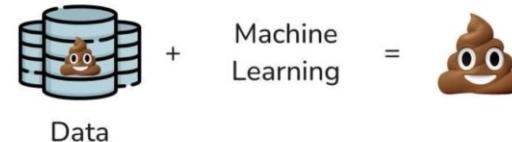
Reducción de Ruido

Eliminar información irrelevante que dificulta la precisión del análisis.

03

Selección de Características

Identificar y mantener solo las características más relevantes para reducir dimensionalidad.



Preprocesamiento de Datos

El preprocesamiento es una fase crítica que implica preparar y limpiar datos brutos antes de análisis avanzados o construcción de modelos predictivos. Es esencial para garantizar que los datos estén en condiciones óptimas y los resultados sean confiables.

01

Limpieza de Datos

Identificar y tratar valores atípicos y valores faltantes que pueden distorsionar el análisis.

02

Transformación

Normalización, estandarización y codificación de variables categóricas en formatos numéricos.

03

Selección de Características

Identificar y mantener solo las características más relevantes para reducir dimensionalidad.

04

Manejo de Desbalances

Abordar desequilibrios en las clases que afectan el aprendizaje del modelo.

05

Reducción de Ruido

Eliminar información irrelevante que dificulta la precisión del análisis.



**CORPORACIÓN UNIVERSITARIA
DEL HUILA - CORHUILA**
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado"



Naive Bayes

Naive Bayes es un algoritmo de clasificación supervisada basado en el teorema de Bayes. El nombre "Naive" (ingenuo) se refiere a la suposición simplificada de independencia entre características, que aunque no siempre es cierta, resulta efectiva en la práctica.

A pesar de esta suposición simplificada, los clasificadores Naive Bayes son rápidos de entrenar y usar, siendo ampliamente utilizados en procesamiento de lenguaje natural, clasificación de textos, sistemas de recomendación y detección de spam.

Aplicaciones Comunes

- Clasificación de texto y documentos
- Detección de spam
- Análisis de sentimientos
- Sistemas de recomendación
- Diagnóstico médico
- Detección de fraudes

Árbol de Decisión y Gradient Boosting

Árbol de Decisión

Estructura jerárquica donde cada nodo interno representa una pregunta sobre una característica, cada rama una respuesta, y cada hoja una decisión o etiqueta. Altamente interpretable y maneja datos mixtos para clasificación y regresión.

Gradient Boosting Classifier

Algoritmo de ensamble que combina múltiples árboles de decisión débiles de manera secuencial. Cada iteración mejora el rendimiento corrigiendo errores anteriores mediante ponderación, logrando mayor precisión en clasificación y regresión.

Ambos algoritmos son fundamentales en Machine Learning, ofreciendo soluciones robustas para problemas complejos de clasificación y predicción.



**CORPORACIÓN UNIVERSITARIA
DEL HUILA - CORHUILA**
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado"