

<b>ARTIFICIAL INTELLIGENCE .....</b>	<b>3</b>
<b>Major schools: .....</b>	<b>3</b>
Symbolism .....	3
Connectionism .....	3
Actionism .....	3
<b>Types of AI: .....</b>	<b>4</b>
Strong .....	4
Weak .....	4
<b>Machine Learning .....</b>	<b>4</b>
Main problems: .....	4
Classification .....	4
Regression .....	4
Clustering .....	5
Types of ML: .....	5
Supervised .....	5
Unsupervised .....	5
Semi-supervised .....	5
Reinforcement .....	6
ML process: .....	6
Data preparation .....	7
Data cleansing .....	7
Data filtering .....	7
Data loss Handling .....	7
Handling of possible error or abnormal values .....	7
Merging of data .....	7
Data consolidation .....	7
Model training .....	7
Gradient descent .....	7
Parameters .....	8
Hyperparameters .....	8
Grid search .....	8
Random search .....	8
Overfitting .....	8
Underfitting .....	9
Model evaluation .....	9
Precision .....	9
Recall .....	9
Accuracy .....	10
F1 Score .....	10
Confusion matrix .....	10
Cross-Validation .....	10
Model deployment and integration .....	10
Common ML algorithms: .....	11
Supervised .....	11
Classification .....	11

Logistic regression .....	11
SVM .....	11
NN .....	12
Decision tree .....	12
Random forest.....	12
Gradient boosted decision tree.....	13
K-NN.....	13
Naïve Bayes.....	14
Regression .....	14
Linear regression .....	14
SVM .....	14
NN .....	15
Decision tree .....	15
Random forest.....	15
Gradient boosted decision tree.....	16
K-NN.....	16
Unsupervised.....	17
Clustering .....	17
K-means clustering.....	17
Hierarchical clustering.....	17

# Artificial Intelligence

La IA es una disciplina científica que estudia y desarrolla teorías, técnicas y sistemas de aplicación utilizados para simular y ampliar la inteligencia humana. El término fue acuñado por primera vez por John McCarthy en 1956, quien lo definió como "la ciencia y la ingeniería de fabricar máquinas inteligentes, especialmente programas informáticos inteligentes". La premisa misma de la tecnología de IA es permitir que las máquinas aprendan de los datos recopilados y tomen decisiones similares a las humanas.

## Major schools:

### Symbolism

Sugiere que las unidades cognitivas básicas de los humanos son los símbolos, y que la cognición humana es un proceso de razonamiento basado en varios símbolos. Como tanto los humanos como las computadoras son sistemas de símbolos físicos, las computadoras pueden usarse para simular comportamientos humanos inteligentes.

### Connectionism

Los investigadores creen que la neurona, y no los procesos simbólicos, es la unidad básica del pensamiento. El conexionismo comienza con las neuronas y estudia las redes neuronales y los modelos cerebrales para crear un nuevo camino de desarrollo para la IA.

### Actionism

Sugiere que la inteligencia depende de la percepción y las acciones, y que la inteligencia no requiere conocimiento, representación o razonamiento. La IA puede evolucionar como la inteligencia humana y el comportamiento inteligente sólo puede manifestarse en el mundo real interactuando con el entorno.

## Types of AI:

### Strong

Esta hipótesis pretende crear máquinas inteligentes que repliquen funciones humanas, como el razonamiento y la resolución de problemas, y que sean perceptivas y conscientes de sí mismas.

### Weak

La IA débil tiene como objetivo construir máquinas inteligentes que puedan realizar tareas específicas pero que dependan en gran medida de la interferencia humana. Estas máquinas pueden parecer inteligentes, pero no son conscientes de sí mismas

## Machine Learning

El aprendizaje automático (ML) se refiere a la capacidad de las computadoras para aprender, simular o implementar el comportamiento humano para adquirir nuevos conocimientos o habilidades y actualizar continuamente las estructuras de conocimiento existentes para mejorar el rendimiento.

*El **aprendizaje automático (ML)** es una rama de la IA donde las máquinas aprenden a resolver problemas mirando datos, sin necesidad de ser programadas explícitamente para cada tarea.*

## Main problems:

### Classification

La clasificación es un concepto utilizado para producir resultados discretos y clasificar datos en regiones específicas.

*Clasificar es como poner cosas en cajas: el modelo decide a qué categoría pertenece algo (por ejemplo, si un correo es spam o no).*

### Regression

Por otro lado, la regresión se utiliza para evaluar la relación entre variables independientes y variables dependientes.

*La regresión encuentra relaciones entre números, como predecir el precio de una casa basado en su tamaño y ubicación.*

## Clustering

La agrupación (Clustering) es un proceso de agrupar conjuntos de elementos en varios grupos. Los elementos u objetos deben ser similares dentro del grupo y diferentes de otros objetos en otros grupos. El objetivo de la agrupación es identificar patrones y similitudes en los datos que pueden usarse para obtener información y hacer predicciones. Los diferentes algoritmos de agrupación utilizan diferentes métodos para agrupar puntos de datos según sus características y medidas de similitud, como la distancia o la densidad. La agrupación en clústeres se utiliza comúnmente en diversas aplicaciones, como segmentación de clientes, clasificación de imágenes y texto, detección de anomalías y sistemas de recomendación.

*El agrupamiento consiste en encontrar grupos en los datos que son similares entre sí, como clasificar animales por su tamaño y dieta.*

## Types of ML:

### Supervised

En el aprendizaje automático supervisado, un modelo hace predicciones o decisiones basadas en datos pasados o etiquetados. Los datos etiquetados se refieren a conjuntos de datos a los que se les asignan etiquetas o etiquetas y, por lo tanto, se vuelven más significativos.

*El aprendizaje supervisado es como enseñar a un niño usando ejemplos: le mostramos entradas y salidas correctas para que aprenda a relacionarlas.*

### Unsupervised

En el aprendizaje no supervisado, no tenemos datos etiquetados. Un modelo puede identificar patrones, anomalías y relaciones en los datos de entrada.

*En el aprendizaje no supervisado, la máquina aprende por sí sola buscando patrones en los datos sin saber previamente qué buscar.*

### Semi-supervised

En el caso del aprendizaje semisupervisado, los datos de entrenamiento contienen una pequeña cantidad de datos etiquetados y una gran cantidad de datos sin etiquetar. Algunas

áreas en las que se aplica incluyen etiquetado de datos, detección de fraude y traducción automática.

*Es una mezcla: la máquina usa unos pocos ejemplos etiquetados y muchos datos sin etiquetar para aprender.*

## Reinforcement

Al utilizar el aprendizaje por refuerzo, el modelo puede aprender en función de las recompensas que recibió por su acción anterior.

El aprendizaje por refuerzo tiene un entorno y un agente. El agente realiza algunas acciones para lograr un objetivo específico. Cada vez que el agente realiza una tarea que lo lleva hacia la meta, es recompensado. Y cada vez que da un paso que va en contra de ese objetivo o en sentido contrario, es penalizado.

*El aprendizaje por refuerzo es como entrenar a un perro: se recompensa al modelo por hacer las cosas bien, lo que lo motiva a mejorar.*

## ML process:

Las tres etapas de la construcción de un modelo de aprendizaje automático son:

- Construcción de modelos: Elija un algoritmo adecuado para el modelo y entrénelo según los requisitos.
- Pruebas de modelos: Verifique la precisión del modelo a través de los datos de prueba.
- Aplicar el modelo: Realice los cambios necesarios después de las pruebas y utilice el modelo final para proyectos en tiempo real.

### Flujo de trabajo de ejemplo:

1. **Defina el problema:** determine si se trata de una clasificación, regresión, agrupamiento, etc.
2. **Examine el conjunto de datos:** verifique el tamaño, las características, el equilibrio, la linealidad, etc.
3. **Elija modelos iniciales:** según el tipo de problema, pruebe primero con modelos más simples (por ejemplo, regresión logística, árboles de decisión).

4. **Validar y comparar:** utilice validación cruzada y métricas para comparar modelos.
5. **Ajustar y refinar:** utilice el ajuste de hiperparámetros y pruebe modelos más complejos si es necesario.

## Data preparation

*Preparar los datos significa organizarlos y limpiarlos para que la máquina pueda entenderlos y usarlos de manera efectiva.*

## Data cleansing

*Limpiar los datos es como quitar errores y valores incorrectos para evitar que confundan al modelo.*

## Data filtering

*Filtrar los datos elimina la información innecesaria, dejando solo lo relevante para el análisis.*

## Data loss Handling

*Manejar la pérdida de datos implica decidir qué hacer con los valores faltantes, como rellenarlos o ignorarlos.*

## Handling of possible error or abnormal values

*Manejar la pérdida de datos implica decidir qué hacer con los valores faltantes, como rellenarlos o ignorarlos.*

## Merging of data

*Combinar datos significa juntar información de varias fuentes en un solo lugar.*

## Data consolidation

*Consolidar datos organiza la información combinada para que sea más fácil de usar.*

# Model training

## Gradient descent

El descenso de gradiente (gradient descent) es un algoritmo de optimización que se utiliza para entrenar modelos de aprendizaje automático y redes neuronales. Su objetivo es minimizar la función de costo o pérdida, es decir, encontrar los valores de los parámetros del modelo que produzcan las predicciones más precisas.

## Parameters

*Los parámetros son las reglas internas del modelo que cambian durante el entrenamiento para ajustarse mejor a los datos.*

## Hyperparameters

*Los hiperparámetros son configuraciones externas que controlan cómo se entrena el modelo, como el tamaño de los pasos en gradient descent.*

### Grid search

La búsqueda de cuadrícula realiza una búsqueda exhaustiva de todas las combinaciones posibles de hiperparámetros para formar una cuadrícula de valores de hiperparámetro. En la práctica, los rangos y pasos de los hiperparámetros se especifican manualmente. La búsqueda en cuadrícula es costosa y requiere mucho tiempo:

Este método funciona bien cuando hay relativamente pocos hiperparámetros. Por lo tanto, es factible para algoritmos generales de aprendizaje automático, pero no para redes neuronales.

*Es como probar todas las combinaciones posibles de hiperparámetros para encontrar la mejor.*

### Random search

Si el espacio de búsqueda de hiperparámetros es grande, la búsqueda aleatoria es más apropiada que la búsqueda en cuadrícula. En una búsqueda aleatoria, cada elemento de configuración se toma como muestra de posibles valores de parámetros para encontrar el subconjunto de parámetros más apropiado.

En una búsqueda aleatoria, primero se realiza una búsqueda dentro de un rango amplio y luego el rango se reduce según la ubicación del mejor resultado.

Algunos hiperparámetros son más importantes que otros y afectan las preferencias de búsqueda aleatoria.

*Es similar a grid search, pero en lugar de probar todo, selecciona combinaciones al azar.*

## Overfitting

El sobreajuste (overfitting) es una situación que ocurre cuando un modelo aprende demasiado bien el conjunto



de entrenamiento, tomando como conceptos fluctuaciones aleatorias en los datos de entrenamiento. Estos afectan la capacidad del modelo para generalizar y no se aplican a datos nuevos.

Existen múltiples formas de evitar el sobreajuste, como, por ejemplo:

- Regularización. Implica un término de costo para las características involucradas con la función objetivo.
- Realización de un modelo sencillo. Con menos variables y parámetros, la varianza se puede reducir.
- También se pueden utilizar métodos de validación cruzada como k-folds.
- Si es probable que algunos parámetros del modelo causen un sobreajuste, se pueden utilizar técnicas de regularización como LASSO que penalizan estos parámetros.

*Cuando un modelo aprende demasiado de los datos de entrenamiento y no generaliza bien, como memorizar respuestas en lugar de entenderlas.*

## Underfitting

*Cuando un modelo no aprende lo suficiente y falla en capturar patrones importantes.*

## Model evaluation

### Precision

*La precisión mide la proporción de predicciones positivas que realmente son correctas:*

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

### Recall

*El recall mide la proporción de verdaderos positivos que fueron identificados correctamente:*

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

## Accuracy

*La exactitud evalúa cuántas predicciones totales fueron correctas:*

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Observations}}$$

## F1 Score

*El F1 Score es la media armónica entre precisión y recall. Es útil cuando las clases están desbalanceadas:*

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

## Confusion matrix

Una matriz de confusión (o matriz de error) es una tabla específica que se utiliza para medir el rendimiento de un algoritmo. Se utiliza principalmente en aprendizaje supervisado; en el aprendizaje no supervisado, se llama matriz de correspondencia.

La matriz de confusión tiene dos parámetros:

- Actual
- Previsto

También tiene conjuntos idénticos de características en ambas dimensiones.

## Cross-Validation

*Es una técnica para evaluar un modelo dividiendo los datos en múltiples subconjuntos. El modelo se entrena en una parte y se prueba en la otra, rotando los subconjuntos, lo que asegura que se pruebe en todo el conjunto de datos. Esto evita depender de un único conjunto de entrenamiento/prueba.*

## Model deployment and integration

*Desplegar el modelo una vez esté probado y adaptado a las necesidades del cliente, considerando el flujo de trabajo necesario para operar de manera adecuada.*

## Common ML algorithms:

### Supervised

#### Classification

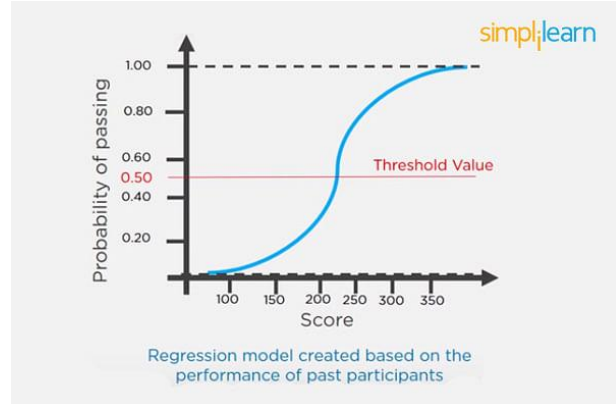
##### *Logistic regression*

La regresión logística es un algoritmo de clasificación que se utiliza para predecir un resultado binario para un conjunto determinado de variables independientes.

El resultado de la regresión logística es 0 o 1 con un valor umbral generalmente de 0,5. Cualquier valor superior a 0,5 se considera 1 y cualquier punto inferior a 0,5 se considera 0.

*A pesar de su nombre, es un algoritmo de clasificación supervisado. Se utiliza para predecir probabilidades y asignar observaciones a clases discretas utilizando una función sigmoide.*

*Ejemplo: Determinar si un correo es spam o no.*



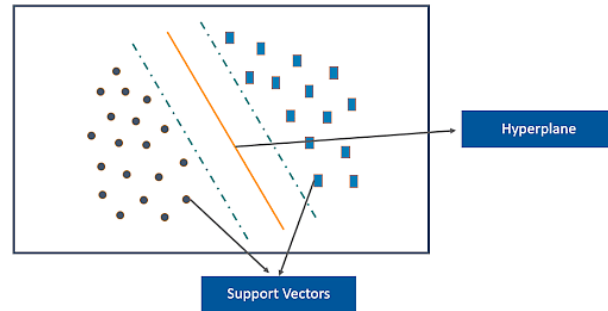
##### SVM

Los vectores de soporte son puntos de datos más cercanos al hiperplano. Influye en la posición y orientación del hiperplano. La eliminación de los vectores de soporte alterará la posición del hiperplano. Los vectores de soporte nos ayudan a construir nuestro modelo de máquina de vectores de soporte.

*SVM encuentra un hiperplano que separa las clases con el mayor margen posible. Es potente*

*para problemas de clasificación y puede manejar datos no lineales usando kernels.*

*Ejemplo: Clasificar imágenes de gatos y perros.*



## NN

*Las redes neuronales, del inglés Neural Networks, simulan el cerebro humano mediante capas de nodos que aprenden patrones complejos en los datos. Se utilizan ampliamente en tareas no lineales.*

*Ejemplo: Reconocimiento de escritura a mano.*

## Decision tree

Un árbol de decisión construye modelos de clasificación (o regresión) como una estructura de árbol, con conjuntos de datos divididos en subconjuntos cada vez más pequeños mientras se desarrolla el árbol de decisión, literalmente en forma de árbol con ramas y nodos. Los árboles de decisión pueden manejar datos tanto categóricos como numéricos.

*Un árbol de decisión divide los datos en ramas basadas en reglas lógicas hasta llegar a un resultado. Es interpretable y adecuado tanto para clasificación como regresión.*

*Ejemplo: Decidir si otorgar un préstamo a un cliente basado en su historial crediticio.*

## Random forest

Un "bosque aleatorio" es un algoritmo de aprendizaje automático supervisado que generalmente se utiliza para problemas de

clasificación. Opera construyendo múltiples árboles de decisión durante la fase de entrenamiento. El bosque aleatorio elige la decisión de la mayoría de los árboles como decisión final.

*Es un ensamblado de múltiples árboles de decisión que combinan sus predicciones. Reduce el riesgo de sobreajuste y mejora la precisión.*

*Ejemplo: Predecir el diagnóstico médico basado en síntomas.*

#### *Gradient boosted decision tree*

*Es un método de ensamble que construye modelos débiles (como árboles) secuencialmente, cada uno corrigiendo los errores del anterior. Variantes populares incluyen **XGBoost** y **LightGBM**.*

*Ejemplo: Predicción del riesgo de abandono de clientes.*

#### *K-NN*

El algoritmo del vecino más cercano K es un algoritmo de clasificación que funciona de manera que un nuevo punto de datos se asigna a un grupo vecino al que es más similar.

En K vecinos más cercanos, K puede ser un número entero mayor que 1. Entonces, para cada nuevo punto de datos que queremos clasificar, calculamos a qué grupo de vecinos está más cercano.

*KNN asigna una clase a una nueva observación basada en las clases de sus vecinos más cercanos. Es intuitivo, pero computacionalmente costoso para grandes conjuntos de datos.*

*Ejemplo: Clasificar si un cliente comprará un producto según el comportamiento de clientes similares.*

### Naïve Bayes

El clasificador se llama "ingenuo" porque hace suposiciones que pueden resultar correctas o no.

El algoritmo supone que la presencia de una característica de una clase no está relacionada con la presencia de ninguna otra característica (independencia absoluta de las características), dada la variable de clase.

Por ejemplo, una fruta puede considerarse cereza si es de color rojo y de forma redonda, independientemente de otras características. Esta suposición puede ser correcta o no (ya que una manzana también coincide con la descripción).

*Este algoritmo basado en probabilidad utiliza el Teorema de Bayes asumiendo independencia entre características. Es rápido y eficaz para clasificación.*

*Ejemplo: Clasificación de texto, como análisis de sentimientos.*

### Regression

#### Linear regression

*La regresión lineal es un algoritmo de aprendizaje supervisado utilizado para predecir valores continuos. Encuentra la relación entre una característica (o múltiples) y la variable objetivo ajustando una línea recta que minimiza los errores de predicción.*

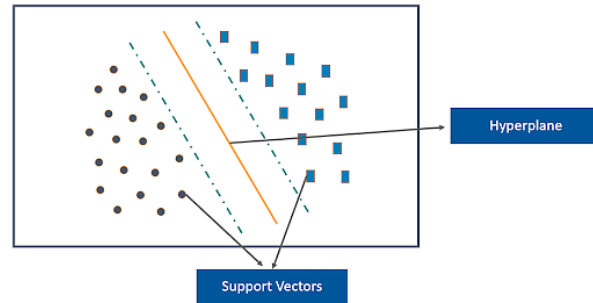
*Ejemplo: Predecir el precio de una casa basado en su tamaño.*

### SVM

Los vectores de soporte son puntos de datos más cercanos al hiperplano. Influye en la posición y orientación del hiperplano. La eliminación de los vectores de soporte alterará la posición del hiperplano. Los vectores de soporte nos ayudan a construir nuestro modelo de máquina de vectores de soporte.

*SVM encuentra un hiperplano que separa las clases con el mayor margen posible. Es potente para problemas de clasificación y puede manejar datos no lineales usando kernels.*

*Ejemplo: Clasificar imágenes de gatos y perros.*



## NN

*Las redes neuronales, del inglés Neural Networks, simulan el cerebro humano mediante capas de nodos que aprenden patrones complejos en los datos. Se utilizan ampliamente en tareas no lineales.*

*Ejemplo: Reconocimiento de escritura a mano.*

## Decision tree

Un árbol de decisión construye modelos de clasificación (o regresión) como una estructura de árbol, con conjuntos de datos divididos en subconjuntos cada vez más pequeños mientras se desarrolla el árbol de decisión, literalmente en forma de árbol con ramas y nodos. Los árboles de decisión pueden manejar datos tanto categóricos como numéricos.

*Un árbol de decisión divide los datos en ramas basadas en reglas lógicas hasta llegar a un resultado. Es interpretable y adecuado tanto para clasificación como regresión.*

*Ejemplo: Decidir si otorgar un préstamo a un cliente basado en su historial crediticio.*

## Random forest

Un "bosque aleatorio" es un algoritmo de aprendizaje automático supervisado que

generalmente se utiliza para problemas de clasificación. Opera construyendo múltiples árboles de decisión durante la fase de entrenamiento. El bosque aleatorio elige la decisión de la mayoría de los árboles como decisión final.

*Es un ensamblado de múltiples árboles de decisión que combinan sus predicciones. Reduce el riesgo de sobreajuste y mejora la precisión.*

*Ejemplo: Predecir el diagnóstico médico basado en síntomas.*

#### *Gradient boosted decision tree*

*Es un método de ensamble que construye modelos débiles (como árboles) secuencialmente, cada uno corrigiendo los errores del anterior. Variantes populares incluyen **XGBoost** y **LightGBM**.*

*Ejemplo: Predicción del riesgo de abandono de clientes.*

#### *K-NN*

El algoritmo del vecino más cercano K es un algoritmo de clasificación que funciona de manera que un nuevo punto de datos se asigna a un grupo vecino al que es más similar.

En K vecinos más cercanos, K puede ser un número entero mayor que 1. Entonces, para cada nuevo punto de datos que queremos clasificar, calculamos a qué grupo de vecinos está más cercano.

*KNN asigna una clase a una nueva observación basada en las clases de sus vecinos más cercanos. Es intuitivo pero computacionalmente costoso para grandes conjuntos de datos.*

*Ejemplo: Clasificar si un cliente comprará un producto según el comportamiento de clientes similares.*



# Unsupervised

## Clustering

### *K-means clustering*

*Es un algoritmo no supervisado que agrupa datos similares en "clusters" minimizando la distancia interna de los puntos a sus centroides.*

*Ejemplo: Agrupar clientes por comportamientos de compra.*

### *Hierarchical clustering*

La agrupación jerárquica divide un conjunto de datos en diferentes capas y forma una estructura de agrupación en forma de árbol. La división del conjunto de datos puede utilizar una política de agregación "de abajo hacia arriba" o una política de división "de arriba hacia abajo". La jerarquía de agrupación se representa en un diagrama de árbol. La raíz es el único grupo de todas las muestras y las hojas son grupos de muestras únicas.

