



TEMA 70

IA: Finalidad y clasificación: ML, DL, NLP, visión artificial, sistemas expertos, robótica, y agentes inteligentes. Aspectos éticos.

Versión	30.1
Fecha de actualización	17/09/2024



ÍNDICE

ÍNDICE	2
1. PROBLEMAS DE BÚSQUEDA Y OPTIMIZACIÓN. RESOLUCIÓN DE PROBLEMAS MEDIANTE ESPACIO DE ESTADOS.....	3
1.1 ESTRATEGIAS DE BÚSQUEDA	3
1.2 OPTIMIZACIÓN	4
2. INTELIGENCIA ARTIFICIAL: FINALIDAD Y CLASIFICACIÓN.	4
3. APRENDIZAJE AUTOMÁTICO O MACHINE LEARNING. DEEP LEARNING.....	5
3.1 MODELOS.....	5
3.2 TIPOS DE APRENDIZAJE.....	5
3.3 PREPROCESADO.....	6
3.4 PRINCIPALES TÉCNICAS DE CLASIFICACIÓN	6
3.5 PRINCIPALES TÉCNICAS DE REGRESIÓN	7
3.6 PRINCIPALES TÉCNICAS DE AGRUPAMIENTO	8
4. PROCESAMIENTO DE LENGUAJE NATURAL (NLP)	9
4.1 COMPONENTES.....	10
5. VISIÓN ARTIFICIAL.....	10
6. SISTEMAS EXPERTOS.....	10
7. ROBÓTICA.....	11
8. INTELIGENCIA ARTIFICIAL DISTRIBUIDA. AGENTES INTELIGENTES.	13
8.1 TIPOS BÁSICOS DE GENTE INTELIGENTE.....	13
8.2 CARACTERÍSTICAS.....	13
8.3 TIPOS DE SISTEMA SEGÚN LA COORDINACIÓN ENTRE AGENTES	14
8.4 SISTEMAS DE COMUNICACIÓN Y COORDINACIÓN	14
8.5 LENGUAJES DE COMUNICACIÓN ENTRE AGENTES	14
8.6 APLICACIONES.....	14
9. IA GENERATIVA (GENAI)	15
10. ASPECTOS ÉTICOS.....	16
TEMAS RELACIONADOS	17
LISTA DE REPRODUCCIÓN: REDES NEURONALES.....	17



1. Problemas de Búsqueda y Optimización.

Resolución de Problemas mediante Espacio de Estados

Consiste en representar el problema como un espacio de estados, buscando un conjunto de transiciones que permitan pasar del estado inicial al estado objetivo. Las búsquedas pueden ser **completas** (siempre encuentran la solución, si existe) y/u óptimas (encuentran la mejor solución posible).

1.1 Estrategias de búsqueda

- **Métodos ciegos o no informados:** Se evalúan los estados sistemáticamente, terminando cuando se alcanza el estado objetivo o no queden más estados por evaluar. Son ineficientes.
 - **Primero en anchura:** se evalúan todos los nodos de un nivel antes de pasar al siguiente. Es completa. Sólo es óptima si el coste del camino hasta la solución es creciente con la profundidad.
 - **Primero en profundidad:** se van probando caminos, siguiendo cada uno de ellos hasta el final. No es completa ni óptima.
 - **Coste uniforme:** se elige el nodo cuyo coste acumulado sea menor. Es completa (si no existen caminos infinitos de coste infinito) y óptima.
 - **Profundidad limitada:** se realiza la búsqueda estableciendo un límite de profundidad. No es completa si el límite es inferior a la profundidad de la solución. No es óptima.
 - **Profundidad iterativa:** similar al anterior, pero el límite se va aumentando si no se encuentra el estado objetivo. Es completa. Es óptima si el coste crece con la profundidad.
 - **Bidireccional:** se busca desde los estados inicial y objetivo simultáneamente. Es completa y óptima.
- **Métodos heurísticos o informados:** Se busca primero en los estados más prometedores, dado que se tiene alguna información sobre la cercanía al estado objetivo. Permiten encontrar una buena solución rápidamente, pero puede no ser la óptima.
 - **Primero el mejor:** utiliza una función de evaluación $F(X)$ que considera el coste mínimo para llegar al estado objetivo pasando por el estado X . Se elige como nodo a evaluar el que minimice $F(X)$.
 - **Búsqueda Greedy:** solo considera el coste mínimo desde el nodo X hasta la solución. No es completo ni óptimo.
 - **Algoritmo A*:** considera además el camino hasta X . Es completo y óptimo.
 - **Mejora iterativa:** utiliza una función de calidad que mide lo buena que es una solución y guía la búsqueda. No son completos ni óptimos.
 - **Escalada simple:** se evalúan los estados siguientes de uno en uno. El primero que sea mejor que el estado actual pasa a ser el nuevo estado.
 - **Escalada por máxima pendiente:** se evalúan todos los estados siguientes al estado actual y se escoge el mejor.
- **Búsqueda con adversario:** Existe un elemento externo que actúa como oponente. Se representa con árboles que muestran todas las secuencias de jugadas posibles. Técnicas:
 - **Minimax:** se aplica la función de evaluación en los nodos hijo y se propagan los valores hasta el nodo raíz. A cada nodo MAX se le asigna el máximo de los valores de sus hijos y a cada nodo MIN el mínimo. El valor propagado a la raíz permite decidir la jugada.



- **Poda α - β :** mejora del anterior eliminando ramas del árbol de búsqueda que se sabe no van a mejorar las obtenidas.

1.2 Optimización

En el caso más simple, un problema de optimización consiste en maximizar o minimizar una función eligiendo sistemáticamente valores y computando el valor de la función. La generalización de la teoría de la optimización y técnicas asociadas comprende un área grande de las matemáticas aplicadas.

Se pueden usar algoritmos que terminen en un **número finito de pasos, métodos iterativos que convergen a una solución o heurísticas** que provean soluciones aproximadas al problema. Ejemplos son:

- Algoritmos de **exploración**:
 - Método **Simplex**: evalúa la función objetivo en una serie de puntos que acotan la solución.
 - Optimización combinatoria: reducen el espacio de búsqueda eficientemente.
- Métodos iterativos basados en **gradiente**:
 - Método de **Newton**: evaluando la matriz Hessiana y el gradiente.
 - **Gradiente descendente**/ascendente: siguiendo la dirección de máxima variación.
 - **Gradiente conjugado**: se varía la dirección del gradiente para evitar mínimos locales.
 - Método de punto interior.
- Métodos **heurísticos**:
 - Evolución diferencial: usando soluciones candidatas, las cuales se recombinan y mutan.
 - Algoritmo de búsqueda diferencial.
 - **Algoritmos genéticos**: sometiendo a una serie de puntos a acciones aleatorias.
 - **Nelder-Mead**: sin llamadas a gradientes.
 - Optimización por enjambre de partículas: a partir de soluciones candidatas aproximadas.

2. Inteligencia artificial: finalidad y clasificación.

En el ámbito de la informática, la Inteligencia Artificial (IA) es una disciplina y un conjunto de capacidades cognitivas e intelectuales expresadas mediante modelos estadísticos, o combinaciones de algoritmos, que se ejecutan en sistemas de computación. La finalidad de la IA es la creación de sistemas que imiten la inteligencia humana para realizar tareas, mejorando significativamente el rendimiento, las capacidades y las contribuciones de los seres humanos, y que aprendan a medida que recopilan información.

Existen diversas maneras de clasificar los sistemas de IA. Una de ellas es la propuesta por Peter Norvig y Stuart Russell en su libro “Artificial Intelligence: A Modern Approach”, según la cual existirían los siguientes tipos de IA:

- **Sistemas reactivos**, que son capaces de responder directamente a situaciones y estímulos presentes en el entorno sin tener en cuenta el pasado ni considerar el futuro.
- **Sistemas basados en memoria**, que conservan una representación parcial del mundo y pueden utilizarla para evaluar situaciones actuales y pasadas.
- **Sistemas basados en metas**, que son capaces de establecer metas y formular planes para alcanzarlas. Utilizan información sobre el estado actual del mundo y desarrollan estrategias para lograr un objetivo deseado.



- **Sistemas basados en modelos**, que tienen una representación interna detallada del mundo y utilizan modelos para simular situaciones y escenarios. Pueden razonar y tomar decisiones basadas en estas simulaciones. Estos sistemas son capaces de predecir los efectos de sus acciones y planificar en consecuencia.

3. Aprendizaje Automático o Machine Learning. Deep Learning.

El aprendizaje automático es la rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender de los datos sin ser programadas explícitamente. El aprendizaje automático utiliza algoritmos y modelos estadísticos para analizar muestras de datos y extraer patrones y utiliza estos patrones para hacer predicciones y tomar decisiones en función de nuevas entradas de datos. Los modelos resultantes deben ser capaces de generalizar comportamientos e inferencias para un conjunto más amplio (potencialmente infinito) de datos.

El aprendizaje automático puede ser visto como un intento de automatizar algunas partes del método científico mediante métodos matemáticos. Por lo tanto, es un proceso de inducción del conocimiento.

El aprendizaje automático tiene una amplia gama de aplicaciones, incluyendo motores de búsqueda, reconocimiento de patrones, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, análisis del mercado de valores, clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, juegos, robótica, vehículos autónomos o el uso de modelos del lenguaje para aplicaciones de asistencia o consulta.

3.1 Modelos

El aprendizaje automático tiene como resultado un modelo para resolver una tarea dada. Una manera de clasificar los modelos sería la siguiente:

- Los **modelos geométricos**, que representan los datos en términos de vectores en un espacio geométrico de n dimensiones. Cada dimensión del espacio representa una característica o atributo de los datos. Si hay un borde de decisión lineal entre las clases, se dice que los datos son linealmente separables.
- Los **modelos probabilísticos**, que intentan determinar la distribución de probabilidades descriptora de la función que enlaza a los valores de las características con valores determinados. Uno de los conceptos claves para desarrollar modelos probabilísticos es la estadística bayesiana.
- Los **modelos lógicos**, que representan el conocimiento y la relación entre los diferentes elementos de un conjunto de datos mediante un conjunto de reglas lógicas. A través de la lógica, los modelos lógicos pueden representar relaciones complejas entre los datos y hacer inferencias y predicciones. Un ejemplo de modelo lógico son los árboles de decisión, en los que se emplea un árbol para representar las reglas lógicas.

Los modelos también se pueden clasificar como modelos de agrupamiento o modelos de potenciación del gradiente. Los modelos de agrupamiento, o clustering, dividen el espacio de datos en grupos o clústers, donde los objetos de un mismo grupo son más similares entre sí que con los objetos de los otros grupos.

Los modelos de potenciación del gradiente, o gradient boosting, producen un modelo predictivo en forma de un conjunto de modelos de predicción débiles, típicamente árboles de decisión. El proceso de entrenamiento se basa en la optimización de una función de costo, que mide la diferencia entre la predicción del modelo y el valor real.

3.2 Tipos de Aprendizaje

- **Aprendizaje supervisado**. Los modelos se entrenan a partir de datos etiquetados previamente. Estos datos etiquetados consisten en ejemplos de entrada y su correspondiente etiqueta, que es la salida esperada, lo que permite al modelo aprender a realizar predicciones precisas sobre nuevas entradas. Durante el



proceso de entrenamiento, el modelo ajusta sus parámetros a partir de la comparación entre las predicciones y las salidas esperadas, con el objetivo de minimizar la diferencia entre ellas. El aprendizaje supervisado se utiliza en una amplia variedad de aplicaciones, como la detección de spam, la clasificación de imágenes o la predicción de precios de acciones.

- **Aprendizaje no supervisado.** Se lleva a cabo sobre un conjunto de ejemplos formado tan sólo por entradas al sistema. No se dispone de información sobre las categorías o clases de los datos. El modelo tiene que ser capaz de explorar los datos y encontrar patrones o agrupaciones ocultas que se emplearán para etiquetar las nuevas entradas.
- **Aprendizaje semisupervisado.** Se combinan datos etiquetados y no etiquetados para entrenar a los modelos. Se emplea en aplicaciones donde la obtención de datos etiquetados es difícil o costosa.
- **Aprendizaje por refuerzo.** Se basa en la interacción de un agente con un entorno para aprender a realizar acciones que maximicen una recompensa. Durante el entrenamiento, el agente realiza una secuencia de acciones en el entorno y recibe una recompensa en función de la calidad de la acción tomada. El objetivo del agente es aprender a tomar acciones que maximicen la recompensa. Se emplea en la conducción autónoma, la robótica, los juegos o la optimización de procesos industriales.
- **Transducción.** Se centra en hacer predicciones directamente sobre una muestra de datos específicos, en lugar de aprender una regla/función general dado todo un conjunto de datos de entrenamiento. Por ejemplo, se puede emplear para predecir el sentimiento asociado a un texto sin necesidad de un conjunto de datos etiquetados previamente. Una de las principales ventajas de la transducción es que se puede utilizar para hacer predicciones en tiempo real.
- **Aprendizaje multi-tarea.** En lugar de entrenar un modelo para una tarea específica, se le entrena para realizar múltiples tareas. De esta manera, se emplea el conocimiento previamente aprendido por el sistema para enfrentarse a nuevos problemas parecidos a los ya vistos.

3.3 Preprocesado

- **Limpieza.** Generalmente, en un conjunto de datos, además de encontrar valores numéricos, en la mayoría de las ocasiones existirán datos no estructurados, inconsistentes, valores anormales, duplicados o directamente inexistentes. Las herramientas de transformación de datos nos permitirán adecuar toda la información de la que disponemos para que sea válida, siendo las técnicas más utilizadas la interpolación, la regresión lineal, el suavizado gaussiano o simplemente la eliminación de datos.
- **Filtrado / Transformación.** Conjunto de técnicas que se aplican al conjunto de datos con el objetivo de mejorar la calidad, eliminar ruido o facilitar la búsqueda de información. Aunque su aplicación se da sobre todo en imagen, existen numerosas técnicas basadas en histogramas (Ventana de Parzen).
- **Normalización.** Significa, en este caso, comprimir o extender los valores de las variables para que estén en un rango definido. En la normalización estándar, a cada dato se le resta del valor de la media del conjunto de datos y el resultado se divide entre la varianza.
- **Reducción de la dimensionalidad.** Estas técnicas se orientan a obtener una representación reducida de los datos originales, manteniendo en la mayor medida posible la integridad y la información existente en los datos. La transformación de datos puede ser lineal, como en análisis de componentes principales (PCA) o la descomposición de valores singulares (SVD), pero también existen otras técnicas de reducción de dimensionalidad no lineal como el análisis de Dirichlet latente (LDA).

3.4 Principales Técnicas de Clasificación

La clasificación consiste en asignar una etiqueta de clase a una muestra de datos de entrada en función de las características de un conjunto de entrenamiento etiquetado. La clasificación se utiliza comúnmente en una variedad de aplicaciones, como el reconocimiento de imágenes, la detección de spam, el análisis de sentimientos o el diagnóstico de enfermedades. A continuación, se presentan algunas de las técnicas de clasificación más empleadas:



- **Naive Bayes.** Utiliza el teorema de Bayes para predecir la clase de una muestra desconocida en función de la probabilidad condicional de cada atributo. El modelo supone que las características son independientes entre sí dada una clase, es decir, no se tienen en cuenta correlaciones entre variables.
- **Árboles de decisión.** Se construyen a través de una serie de decisiones binarias que dividen el conjunto de datos en subconjuntos más pequeños, y asignan una etiqueta de clase a cada hoja del árbol. Una hoja es un nodo sin descendientes (nodo terminal).
- **Regresión logística.** Se utiliza una función logística para modelar la probabilidad de pertenencia a una clase y se ajusta mediante la maximización de la verosimilitud. La variable dependiente es una variable categórica con dos o más categorías posibles. Las categorías son las clases. El modelo estima las probabilidades de pertenencia a cada categoría en función de una o más variables predictoras. Las variables predictoras son las características o atributos del dato de entrada.
- **Random Forest.** Consiste en la creación de múltiples árboles de decisión y la combinación de sus resultados para obtener una predicción final. Cada árbol se entrena con una muestra aleatoria de los datos y un subconjunto aleatorio de las características. Para decidir la clase final se utiliza el voto por mayoría de los árboles.
- **K-vecinos más cercanos (K-Nearest Neighbours, K-NN).** Clasifica una instancia según la clase predominante entre sus K vecinos más cercanos.
- **Máquinas de vectores de soporte (Support Vector Machine, SVM).** Busca un hiperplano que maximice la separación entre dos clases. Si los datos no son linealmente separables, SVM puede transformar los datos de entrada en un espacio de características de mayor dimensión para separar las clases. Los vectores de soporte son los puntos más cercanos al hiperplano de separación y son cruciales para definir la frontera de decisión.
- **Redes neuronales.** Son un paradigma de aprendizaje automático inspirado en las neuronas de los sistemas nerviosos de los animales. Las redes neuronales consisten en una estructura de múltiples capas, donde cada capa contiene una serie de nodos (neuronas) interconectados. Cada nodo realiza una operación matemática en la entrada y emite una salida. El proceso de entrenamiento consiste en ajustar los pesos de los enlaces entre los nodos para minimizar una función de pérdida. Cada nodo de salida representa una clase o categoría posible. La importancia de las redes neuronales decayó durante un tiempo con el desarrollo de los vectores de soporte y los clasificadores lineales, pero volvió a surgir a finales de la década del 2000 con la llegada del aprendizaje profundo (Deep Learning). El Deep Learning se enfoca en la construcción de modelos de redes neuronales artificiales con múltiples capas, que pueden aprender representaciones de datos altamente abstractas y complejas.
- **Gradient Boosting Tree (GBT) o Gradient Boosted Machines (GBM).** Construye una secuencia de árboles de decisión, donde cada árbol se ajusta a los residuos del modelo anterior. Los residuos se refieren a las diferencias entre los valores etiquetados y los valores predichos por el modelo en cada iteración. Este proceso se repite iterativamente hasta que se alcanza el número de árboles de decisión especificado o hasta que se alcanza un criterio de parada predefinido.
- **Reglas de asociación.** Los algoritmos de reglas de asociación procuran descubrir relaciones interesantes entre variables. Entre los métodos más conocidos se hallan el algoritmo a priori, el algoritmo Eclat y el algoritmo de Patrón Frecuente.
- **Algoritmos genéticos.** Los algoritmos genéticos son procesos de búsqueda heurística que simulan la selección natural. Usan métodos tales como la mutación y el cruzamiento para generar nuevas clases que puedan ofrecer una buena solución a un problema dado.
- **Redes bayesianas.** Una red bayesiana, red de creencia o modelo acíclico dirigido es un modelo probabilístico que representa una serie de variables de azar y sus independencias condicionales a través de un grafo acíclico dirigido.

3.5 Principales Técnicas de Regresión



La regresión consiste en predecir un valor numérico continuo en función de las características o atributos de la entrada. La regresión se utiliza comúnmente en una variedad de aplicaciones, como la predicción de precios de bienes inmuebles, la estimación de ingresos o la predicción del tiempo. Algunas de las técnicas de regresión más empleadas son las siguientes:

- **Árboles de decisión.** Cuando se emplean para la regresión también se les denominan árboles de regresión. Cada hoja es un valor numérico que representa la estimación para el valor de la variable objetivo en esa región de características.
- **Regresión lineal.** Modela matemáticamente la variable desconocida o dependiente y la variable conocida o independiente como una ecuación lineal. En su forma más simple, la regresión lineal se expresa mediante la ecuación $y = mx + b$, donde y es la variable dependiente, x es la variable independiente, m es la pendiente de la línea y b es la intersección en el eje y . El objetivo es encontrar los valores de m y b que minimicen la distancia entre los puntos de datos reales y la línea ajustada.
- **Random Forest.** Para la regresión se utiliza el promedio de las predicciones de los árboles.
- **Redes neuronales.** Cada nodo de salida produce una salida numérica única con la predicción del valor de la variable objetivo correspondiente. Por ejemplo, en una red neuronal para predecir el precio de una casa, puede haber un nodo de salida para la predicción del precio de venta, otro nodo de salida para la predicción del precio de alquiler y así sucesivamente.
- **Gradient Boosting Tree (GBT) o Gradient Boosted Machines (GBM).**

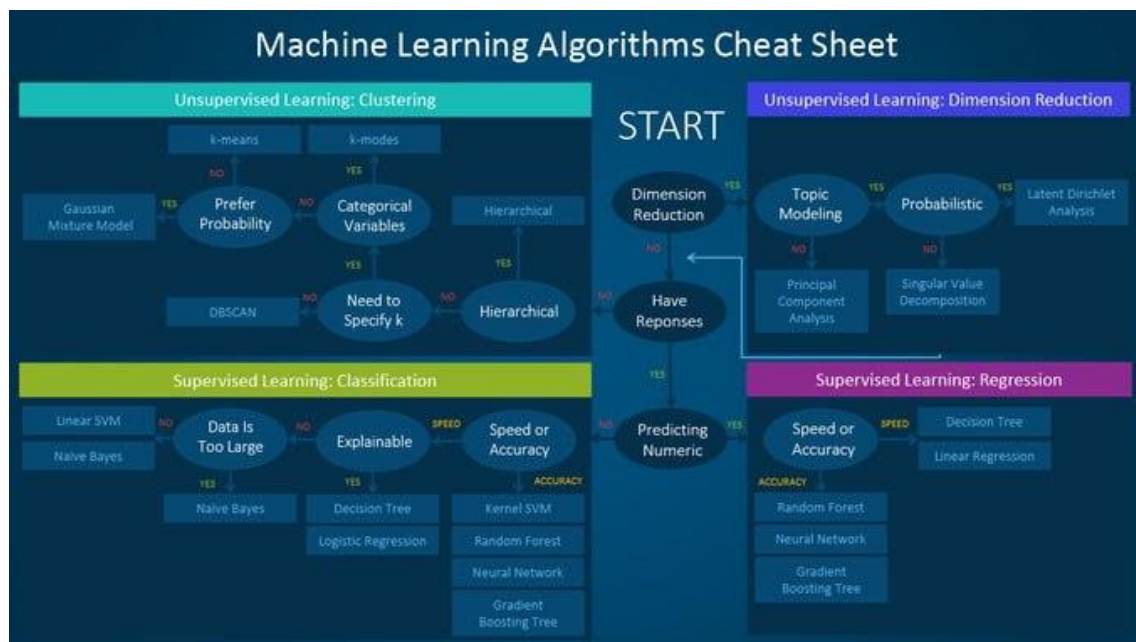


Figura 1. Flujograma del pipeline de procesamiento con técnicas de Machine Learning.

3.6 Principales Técnicas de Agrupamiento

El agrupamiento, o clustering, es una tarea de aprendizaje no supervisado que consiste en agrupar conjuntos de datos similares en subconjuntos o clústers. Algunas de las técnicas de agrupamiento más empleadas son las siguientes:

- **K-medias (K-means).** Agrupa los datos en K clústers, donde K es un número predefinido. El algoritmo funciona iterativamente, comenzando con una inicialización aleatoria de los centroides de los clústers y asignando cada punto de datos al centroide más cercano. Luego, se recalculan los centroides de cada clúster y se repite el proceso hasta que se alcanza la convergencia. El objetivo es minimizar la suma de las distancias de cada punto de datos a su centroide correspondiente. Como función de distancia se suele usar la distancia cuadrática.



- **K-mediods.** Es una versión de K-means que, en lugar de los centroides, utiliza los mediods, que son los puntos de datos reales dentro del conjunto de datos. K-mediods es menos sensible que K-means a valores atípicos y puede ser más efectivo en la agrupación de datos no esféricos.
- **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise). Agrupa los puntos de datos en función de su densidad. El algoritmo define clústers como áreas con alta densidad de puntos, separados por áreas con baja densidad.
- **Modelos de mezclas gaussianas (Gaussian Mixture Model, GMM).** Es la generalización probabilística del algoritmo de K-means. Un GMM es una combinación lineal de distribuciones gaussianas, o normales, que se ajustan a los datos. Cada componente gaussiano representa un clúster en el conjunto de datos y tiene sus propios parámetros de media y varianza.
- **Clustering jerárquico.** Puede ser aglomerativo o divisivo. El algoritmo aglomerativo comienza con cada punto de datos como su propio clúster y luego fusiona los clusters más cercanos para crear clusters más grandes, hasta que se forma un único cluster que contiene todos los datos. El algoritmo divisivo comienza con un único clúster que contiene todos los datos y luego los divide en clusters más pequeños, hasta que cada punto de datos está en su propio cluster. El resultado es una estructura de árbol, llamado dendrograma, que muestra la estructura jerárquica de los clústers. El número de clústers y qué elementos pertenecen a cada uno de ellos, dependerá de la altura a la que cortemos el dendrograma.

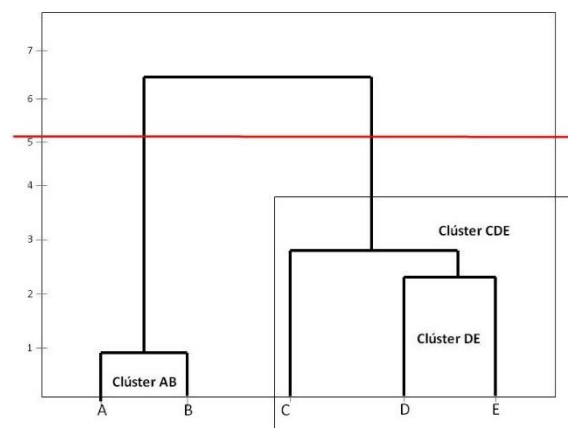


Figura 2. Dendrograma. La altura del corte determina el número y la configuración de los clúster.

4. Procesamiento de Lenguaje Natural (NLP)

El "Test de Turing", propuesto en 1950, evalúa la inteligencia de una máquina según su capacidad para generar respuestas indistinguibles de las de un ser humano y suele emplearse en NLP para medir el éxito en la generación de lenguaje natural que trata de imitar el comportamiento humano.

El NLP es una rama de la inteligencia artificial y la lingüística que estudia las interacciones entre las computadoras y el lenguaje humano. El objetivo del NLP es desarrollar tecnologías que permitan a las computadoras comprender, interpretar, manipular y generar lenguaje humano natural de manera efectiva.

El NLP no trata de la comunicación por medio de lenguas naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente y se puedan realizar por medio de programas que ejecuten o simulen la comunicación. Los modelos aplicados se enfocan no solo en la comprensión del lenguaje de por sí, sino a aspectos generales cognitivos humanos y a la organización de la memoria. El lenguaje natural sirve solo de medio para estudiar estos fenómenos.

El NLP se aplica en una amplia variedad de áreas, como el análisis de sentimientos, la traducción automática, la generación de texto automático, el reconocimiento de voz, la respuesta automática a preguntas, la extracción de información, el resumen automático de texto y la clasificación de texto, entre otras. Se destaca su uso



en asistentes virtuales como Siri o Alexa y es una técnica clave en la IA Generativa que veremos más adelante, a través de los LLM.

4.1 Componentes

- **Análisis léxico.** Identificación y análisis de palabras en un texto. Esto puede incluir la identificación de las partes del discurso, la lematización (reducción de las palabras a su forma básica) y la identificación de sinónimos y antónimos.
- **Análisis sintáctico.** Identificación y análisis de la estructura gramatical a nivel de oración.
- **Análisis semántico.** Extracción del significado de la frase y resolución de ambigüedades léxicas y estructurales.
- **Análisis pragmático.** Identificación y análisis del significado implícito de un texto, así como la inferencia de la intención del hablante o escritor. Esto puede incluir la identificación de la ironía, el sarcasmo, el humor y otros aspectos de la comunicación que no son explícitos en el texto.
- **Planificación de la frase.** Estructurar cada frase del texto con el fin de expresar el significado adecuado.
- **Generación de la frase.** La generación de la cadena lineal de palabras a partir de la estructura general de la frase, con sus correspondientes flexiones, concordancias y restantes fenómenos sintácticos y morfológicos.

5. Visión Artificial.

La visión artificial es una rama de la Inteligencia Artificial que desarrolla métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica que pueda ser tratada por sistemas de computación.

Algunas de las principales técnicas y aplicaciones de la Visión Artificial son los siguientes:

- **Extracción de características:** proceso de identificar y extraer características relevantes de una imagen, como bordes, contornos, texturas o colores.
- **Segmentación:** Dividir una imagen en regiones o segmentos más pequeños con el fin de simplificar el análisis y facilitar la extracción de características específicas.
- **Reconocimiento de objetos:** Identificar y clasificar objetos específicos en una imagen. Se pueden emplear algoritmos de Machine Learning, previamente entrenados con un conjunto de imágenes de objetos etiquetadas.
- **Seguimiento de objetos** a través de una secuencia de imágenes o en tiempo real.
- **Reconocimiento facial** para identificar a personas o su estado anímico y sus sentimientos.
- **Realidad aumentada:** Combina la información visual del mundo real con información generada por computadora para proporcionar una experiencia interactiva en tiempo real. Esto se logra al superponer objetos virtuales en una imagen o video capturado.

6. Sistemas Expertos.

Un sistema experto es un sistema informático que emula la capacidad de tomar decisiones de un humano



experto en un dominio del conocimiento. Tiene capacidad explicativa de la resolución del problema.

Tipos de sistemas expertos:

- **Razonamiento basado en reglas:** se basan en la aplicación de reglas, evaluación de resultados y aplicación de nuevas reglas según la nueva situación.
- **Razonamiento basado en casos:** permite resolver nuevos problemas basándose en soluciones similares de problemas anteriores.
- **Razonamiento basado en redes bayesianas:** fundamentados en la estadística y el teorema de Bayes.

Componentes de un sistema experto:

- **Base de conocimientos (BC):** conjunto de reglas que permiten representar los conocimientos del dominio del experto.
- **Base de hechos o memoria de trabajo:** contiene los datos de partida y se va actualizando con nuevos hechos durante la ejecución del motor de inferencia.
- **Motor de inferencia:** interpreta las reglas contenidas en la base de conocimientos y realiza procesos de inferencia que relacionan los hechos con las reglas, para obtener conclusiones.

Herramientas de desarrollo de sistemas expertos:

- Lenguajes de programación convencionales (C, Pascal...).
- Lenguajes de programación orientados a la Inteligencia Artificial (LISP, PROLOG...).
- Lenguajes de programación orientados a objetos.
- Shells. Son sistemas expertos con la BC vacía.
- Entornos de creación de sistemas expertos.

7.Robótica.

En el ámbito de las TIC, el término “robótica” puede referirse, al menos, a las siguientes dos definiciones:

- Ciencia que aglutina varias ramas tecnológicas, entre ellas, la ingeniería mecánica, la ingeniería electrónica y las ciencias de la computación, ocupándose del diseño, construcción, operación, estructura, manufactura y aplicación de los robots físicos. Estos robots se emplean en diversas tareas como la soldadura, el transporte de piezas de un punto a otro de una cadena de fabricación, los androides que simulan las capacidades humanas, como la locomoción bípeda, o los zoomórficos que imitan el sistema de locomoción de otros seres vivos.
- La Automatización Robótica de Procesos (RPA, por sus siglas en inglés). Un sistema de RPA es un paquete de software que replica las acciones de un ser humano interactuando con la interfaz de usuario de un sistema informático, liberándose de la dependencia de APIs de programación. Un RPA se basa en reglas, por lo que no es necesario que presente características propias de la IA. No obstante, la evolución y maduración de la IA, y en concreto del Machine Learning, ha hecho posible la inclusión de IA en estos sistemas. Por ejemplo, pueden incluir algoritmos de agrupamiento para clasificar documentos.

La AGE cuenta con el Servicio de Automatización de Procesos (SAI), que se corresponde con la medida 5 del Plan de Digitalización de las AAPP 2021-2025 y persigue la mejora de la calidad de los servicios y



procesos de gestión y tramitación de la AGE a través de la utilización de tecnologías de automatización inteligente (RPA). El SAI proporciona a los organismos de la Administración del Estado un servicio integral de robotización centralizada basado en tecnología RPA, cubriendo todos los aspectos de los proyectos desde el descubrimiento hasta la operación, pasando por el desarrollo y mantenimiento de las automatizaciones.



8. Inteligencia Artificial Distribuida. Agentes Inteligentes.

La Inteligencia Artificial Distribuida (DAI, por sus siglas en inglés) se enfoca en el desarrollo de sistemas de Inteligencia Artificial que se ejecutan en múltiples nodos que trabajan juntos como si fueran un solo sistema inteligente. Los sistemas DAI consisten en nodos de procesamiento de aprendizaje autónomos (agentes), que se distribuyen, a menudo a gran escala. Los nodos DAI pueden actuar de forma independiente y las soluciones parciales se integran mediante la comunicación entre nodos, a menudo de forma asincrónica.

En virtud de su escala, los sistemas DAI son robustos y elásticos y, por necesidad, están débilmente acoplados. Además, los sistemas DAI están diseñados para adaptarse a los cambios en la definición del problema o en los conjuntos de datos subyacentes debido a la escala y la dificultad de la redistribución.

Esta disciplina estudia la solución cooperativa de problemas mediante un conjunto de agentes distribuidos, donde ninguno posee toda la información para resolver el problema.

Un agente inteligente es una entidad autónoma, con conocimientos propios y capacidad para tomar decisiones e interactuar de manera flexible con su entorno y con otros agentes. En la mayoría de las ocasiones, los agentes no son desarrollados de forma independiente sino como entidades que constituyen un sistema. A este sistema se le denomina sistema multiagente.

8.1 Tipos Básicos de gente Inteligente

- Agente **reactivo**: agente de bajo nivel, capaz de responder a estímulos inmediatos del entorno en lugar de planificar o tomar en cuenta el historial de eventos pasados. No mantienen una representación interna del entorno. No son inteligentes individualmente, sino de manera conjunta.
- Agente **cognitivo**/deliberativo: posee capacidad de razonamiento sobre su base de conocimiento y utilizan la información del pasado para tomar decisiones y realizar acciones. Pueden mantener una representación interna del entorno. Están diseñados para aprender y adaptarse a su entorno a medida que interactúan con él. Tienen la capacidad de comunicarse con otros agentes y de lograr acuerdos con ellos.

8.2 Características

- **Continuidad temporal**: Un agente es un proceso que se ejecuta sin fin realizando su función.
- **Autonomía**: Un agente es autónomo si es capaz de tomar decisiones y actuar de forma autónoma sin intervención humana.
- **Sociabilidad**: Un agente puede comunicarse con otros agentes y con humanos.
- **Racionalidad**: Los agentes pueden razonar y tomar decisiones lógicas basadas en la información que recopilan y su conocimiento previo.
- **Reactividad**: Un agente actúa como resultado de cambios en su entorno.
- **Proactividad**: Los agentes tienen un conjunto de objetivos específicos que intentan lograr a través de sus acciones y decisiones.
- **Adaptabilidad**: Si el agente puede alterar su comportamiento en base a su aprendizaje.
- **Movilidad**: Capacidad de un agente de trasladarse a través de una red telemática.
- **Veracidad**: Asunción de que un agente no comunica información falsa a propósito.



- **Benevolencia:** Asunción de que un agente está dispuesto a ayudar a otros agentes si esto no entra en conflicto con sus propios objetivos.

8.3 Tipos de Sistema según la coordinación entre agentes

- **Sistemas de Resolución Distribuida de Problemas (SRDP):** Las interacciones entre agentes están prefijadas de antemano. Requiere un agente coordinador que asigne las tareas. Para la resolución distribuida de problemas, el problema principal es cómo se descompone el problema y se sintetizan las soluciones. La Resolución Distribuida de Problemas consta de 3 fases: dividir el problema en subproblemas menores, resolver dichos subproblemas y finalmente, combinar de las soluciones parciales para obtener la solución global.
- **Sistemas Multiagente (SMA):** Es un sistema compuesto por múltiples agentes autónomos que interactúan entre sí para lograr objetivos específicos, decidiendo cómo cooperan y quién realiza cada tarea. En los sistemas de multiagente, el problema principal es cómo los agentes coordinan sus conocimientos y actividades.

8.4 Sistemas de comunicación y coordinación

- **Pizarra:** Es arbitrada por un controlador, que determina el orden en que los agentes aportan sus contribuciones a la solución.
- **Mensajes:** Se intercambian directamente entre dos agentes.
- **Red de contratos:** Se basa en dos roles: contratista (gestiona la tarea) y oferente (la ejecuta).

8.5 Lenguajes de comunicación entre agentes

- **KIF (Knowledge Interchange Format):** Lenguaje intermedio que se utiliza para representar el conocimiento en un formato que es independiente de cualquier lenguaje de programación específico. Se emplea cuando los agentes están programados en lenguajes diferentes o tienen estructuras de datos internas diferentes.
- **KQML (Knowledge Query and Manipulation Language):** Lenguaje de comunicación de agentes basado en texto que permite su interacción y coordinación.

8.6 Aplicaciones

- Comercio electrónico, p. Ej. Para las estrategias comerciales, el sistema DAI aprende las reglas comerciales financieras a partir de muestras de datos financieros.
- Redes, p. Ej. en telecomunicaciones, el sistema DAI controla los recursos cooperativos en una red WLAN.
- Enrutamiento, p. Ej. Modelo de flujo de vehículos en las redes de transporte.
- Programación, p. Ej. Programación de flujo de trabajo donde la entidad de gestión de recursos asegura la optimización local y la cooperación para la coherencia global y local.
- Sistemas de agentes múltiples, p. Ej. vida artificial, el estudio de la vida simulada.
- Sistemas de energía eléctrica, p. Ej. COndition Monitoring Multi-Agent System (COMMAS) aplicado a la monitorización de la condición del transformador y el sistema de restauración automática.



9. IA Generativa (GenAI)

La **Inteligencia Artificial Generativa (GenAI)** es aquella que facilita la creación de contenido nuevo (**contenido sintético**), como texto, imágenes, videos o audio, a partir de grandes volúmenes de datos. Destacan los **modelos de difusión** que aprenden a través de la distribución de probabilidad de los datos y han contribuido a la generación de imágenes y videos de alta calidad (**Stable Diffusion**, ejemplo open-source). Por otro lado, los **Grandes Modelos de Lenguaje (LLM)**, generalmente hacen uso de la arquitectura **Transformer**, con la cual procesan el lenguaje mediante **tokens** (unidades básicas de texto, palabras o subpalabras). Dicha arquitectura utiliza el **mecanismo de atención** para enfocarse en distintas partes del texto dinámicamente, lo que les permite extraer el contexto y las relaciones entre las palabras.

Especialmente en contextos corporativos, aquellos entrenados con grandes conjuntos de datos y que sirven como **punto de partida** para desarrollar modelos más especializados, ahorrando tiempo y recursos se les denomina **modelos fundacionales**. Estos se adaptan mediante distintas técnicas como por ejemplo el **fine-tuning** o el RLHF (**Reinforcement Learning with Human Feedback**). Como alternativa (o mejora) al entrenamiento convencional del modelo (**dado su alto coste computacional**) se puede hacer uso de la **ingeniería de prompts** y/o la **Generación Aumentada por Recuperación (RAG)**.

La **ingeniería de prompts** se centra en el diseño cuidadoso de las instrucciones (**prompts**) que se proporcionan a los modelos de lenguaje con el objetivo de obtener respuestas más específicas, precisas y útiles. Para ello, se emplean diversas estrategias, como el enfoque **Few-shot**, donde se incluyen varios ejemplos para guiar al modelo; el enfoque **Zero-shot**, en el que no se proporcionan ejemplos directos y se confía en una descripción detallada de la tarea; o el enfoque **One-shot**, que utiliza un único ejemplo representativo.

RAG es un proceso que busca optimizar la ventana de contexto del modelo, combinando la generación con la recuperación de información sobre BBDD previamente indexadas (datos en forma de vectores, **embeddings**). Se divide en las siguientes fases:

1. **Consulta inicial:** El usuario formula una consulta que activa el sistema de recuperación (**Retriever**).
2. **Recuperación:** El **Retriever** localiza la información relevante en la **BBDD indexada**.
3. **Aumento:** La consulta se enriquece con la información recuperada, proporcionando contexto adicional.
4. **Generación:** El **LLM** procesa la consulta aumentada y es capaz de generar una respuesta más precisa con las referencias de las fuentes originales.

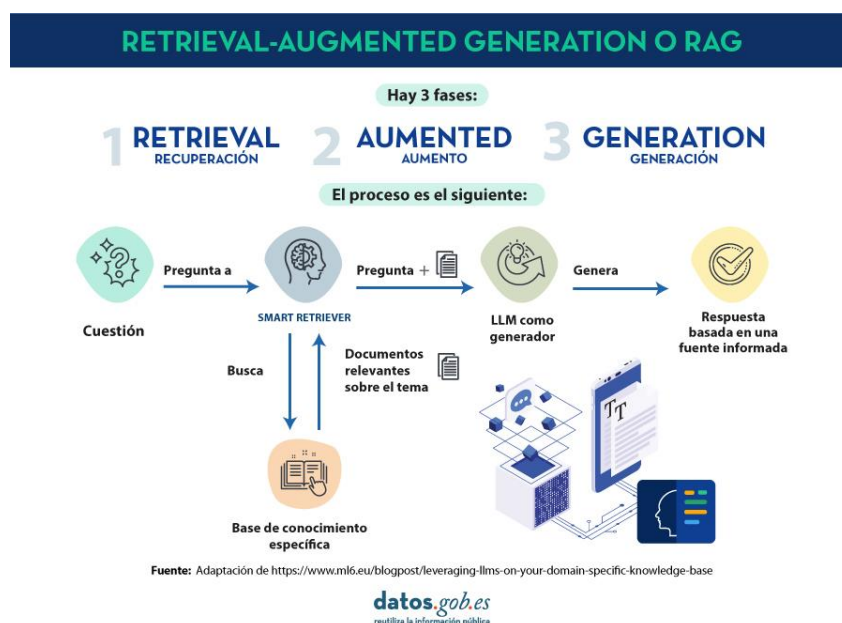


Figura 3. Esquema RAG explicativo (blog de datos.gob.es).



Por otro lado, cabe destacar la técnica de **cuantización**, que reduce su tamaño, mejora la velocidad en inferencia y disminuye el coste de procesamiento, aunque puede comprometer la calidad de las respuestas. No obstante, es una solución efectiva para desplegar modelos en dispositivos de menor capacidad, como smartphones (**on-device**, **Edge-device**), garantizando **disponibilidad** y **privacidad**. Ejemplo: IA de Apple Intelligence.

Al seleccionar un modelo de lenguaje, es crucial evaluar su rendimiento utilizando **Benchmarks** como el MMLU o el HumanEval, que miden distintas capacidades en áreas como la programación, o la ventana de contexto para incorporar datos nuevos.

Los modelos actuales pueden llegar a trabajar con múltiples modalidades de entrada y salida, como texto, imágenes o audio, denominados modelos **multimodales**. Ejemplos recientes como **Gemini (Google, Deepmind)** combinan la generación de texto con capacidades visuales, posibilitando la traducción en tiempo real o la creación de resúmenes de video. Estas altas capacidades han abierto la puerta también a ser empleados como parte central de **agentes autónomos**, optimizando procesos de backend y otras tareas automatizadas. A largo plazo, de manera teórica se espera la **Inteligencia Artificial General (AGI)** que será capaz de realizar cualquier tarea intelectual humana.

Cabe destacar que, de acuerdo con el Plan de Digitalización de las AAPP 2021-2025, en su Eje 1, en línea con la medida 6 de gestión e intercambio transparente del dato se da impulso al **Datalake de la AGE** con la creación de la **Plataforma del Dato**. Esto resulta relevante, dado que en algunos casos de uso se ha utilizado dicha **Plataforma** mediante sus herramientas de analítica avanzada para el entrenamiento de modelos de lenguaje (chatbot de consulta de documentación, por ejemplo), así como la puesta en producción de algoritmos para el procesamiento de lenguaje natural (el análisis del sentimiento o categorización de contenido).

OWASP (Open Web Application Security Project) ha recopilado los principales riesgos de las aplicaciones que hacen uso de LLM. Entre los más preocupantes están las **“alucinaciones”** (información incorrecta o inventada), o los **ataques de inyección de prompts**, donde se manipulan las instrucciones para obtener respuestas inseguras. Estos riesgos pueden mitigarse mediante una adecuada gestión de privilegios y roles o un diseño cuidadoso de los datos de entrenamiento. Además, se recomienda verificar las fuentes de información y evitar la delegación excesiva de las tareas.

10. Aspectos Éticos.

Los aspectos éticos de la IA incluyen entre otros:

- La responsabilidad de los desarrolladores y los usuarios de la IA, que han de diseñar sistemas que se adhieran a principios éticos, como la transparencia, la imparcialidad.
- La privacidad y protección de datos que se emplean, por ejemplo, en el entrenamiento de los algoritmos de Machine Learning.
- La equidad y justicia para asegurar que los sistemas de IA sean imparciales y evitar sesgos discriminatorios.
- El impacto laboral para garantizar la protección de los trabajadores y fomentar la reestructuración laboral justa.
- Tomar en consideración la seguridad y el control de la IA superinteligente, que supera la capacidad intelectual humana, ya que su mal uso o falta de alineación con los valores humanos podrían tener consecuencias devastadoras.
- El impacto ambiental de la IA, especialmente en su entrenamiento, requiere grandes cantidades de datos y tiempo de cómputo, que origina un alto coste energético y contribuye al efecto invernadero.



La opacidad de muchos algoritmos puede crear incertidumbre y obstaculizar la aplicación efectiva de la legislación vigente en materia de seguridad y derechos fundamentales. Esto abarca aplicaciones como los sistemas de identificación biométrica o decisiones de IA que afectan a intereses personales importantes, por ejemplo, en los ámbitos de la contratación, la educación, la asistencia sanitaria o la aplicación de la ley. El Reglamento de Inteligencia Artificial (RIA) tiene por objeto garantizar la protección de los derechos fundamentales y al mismo tiempo promover la confianza en el desarrollo y la adopción de la IA.

A fin de aumentar la eficiencia y establecer un punto de contacto oficial con la población y otros homólogos, cada Estado miembro deberá designar una autoridad nacional de supervisión, que también representará al país en el Comité Europeo de Inteligencia Artificial. En España, será la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), que cumplirá las labores de supervisión y minimización de riesgos y busca generar un ecosistema de investigación y empresarial de Inteligencia Artificial.

Los IA generativa, ha experimentado un crecimiento exponencial en los últimos años, generando debate en diversos ámbitos profesionales, como el diseño, la música o el cine. Las principales preocupaciones giran en torno a la automatización de los diferentes procesos con su abaratamiento y la consecuente reducción de la demanda de RRHH. Además, de los problemas de copyright derivados del uso de datos protegidos por derechos de autor para entrenar estos modelos, ya que muchas de estas tecnologías se basan en información previamente registrada y no constan del consentimiento. También plantea importantes desafíos como los deepfakes. Se trata de contenido sintético generado a partir de herramientas de IA, como imágenes, videos o audios que pueden ser utilizados de manera malintencionada para aparentar que una persona ha dicho o hecho algo que nunca ha ocurrido, afectando gravemente la privacidad y reputación de la persona.

Otro concepto clave y en estudio que está vinculado con la seguridad y la confianza de la IA es la **explicabilidad (xAI)**. La IA, especialmente los basados en redes neuronales, se consideran "**cajas negras**" debido a la complejidad y la falta de transparencia en su funcionamiento. Esta característica opaca de los modelos impide que los usuarios comprendan cómo y por qué se generan ciertos resultados.

Finalmente, la distribución desigual de recursos y capacidades en torno a la IA acentúa la brecha digital a nivel global. La falta de acceso a los datos y a los Centros de Procesamiento de Datos profundiza la exclusión digital, además de crear una dependencia de modelos de lenguaje de habla inglesa, perpetuando sesgos culturales y discriminando lenguas como el gallego, euskera, valenciano o catalán. En este sentido, y en línea con la nueva [Estrategia de IA 2024](#), el gobierno ha destacado el desarrollo de un **modelo fundacional en español y lenguas cooficiales**.

TEMAS RELACIONADOS

75. Tecnologías y sistemas de explotación de datos: data lake, data warehouse, lakehouse, data fabric, data mesh, tecnologías para la protección de la confidencialidad (PET). Entornos de compartición de datos: espacios de datos, aspectos tecnológicos y organizativos.

107. Las tecnologías emergentes. Concepto. Clasificación, aspectos jurídicos y aplicaciones.

117. La transformación digital e industria 4.0: ciudades inteligentes. Internet de las Cosas (IoT).

LISTA DE REPRODUCCIÓN: REDES NEURONALES

https://youtube.com/playlist?list=PL-Ogd76BhmcC_E2RjglIJZd1DQdYHcVf0

