

# IA

## COMPRENDER LA INTELIGENCIA ARTIFICIAL

DANIEL PEÑA



# IA

## COMPRENDER LA INTELIGENCIA ARTIFICIAL

DANIEL PEÑA

Funcas

**PATRONATO**

ISIDRO FAINÉ CASAS  
ANTONIO JESÚS ROMERO MORA  
FERNANDO CONLLEDO LANTERO  
ANTÓN JOSEBA ARRIOLA BONETA  
MANUEL AZUAGA MORENO  
CARLOS EGEA KRAUEL  
MIGUEL ÁNGEL ESCOTET ÁLVAREZ  
AMADO FRANCO LAHOZ  
JOSÉ MARÍA MÉNDEZ ÁLVAREZ-CEDRÓN  
PEDRO ANTONIO MERINO GARCÍA  
ANTONIO PULIDO GUTIÉRREZ

**DIRECTOR GENERAL**

CARLOS OCAÑA PÉREZ DE TUDELA

Impreso en España  
Edita: Funcas  
Caballero de Gracia, 28, 28013 - Madrid  
© Funcas

Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta publicación, así como la edición de su contenido por medio de cualquier proceso reprográfico o fónico, electrónico o mecánico, especialmente imprenta, fotocopia, microfilm, *offset* o mimeógrafo, sin la previa autorización escrita del editor.

ISBN impreso: 979-13-87770-20-4  
ISBN digital: 979-13-87770-21-1  
Depósito legal: M-10624-2026  
Maquetación: Funcas

## Contenido

---

Prefacio	5
Capítulo I. Introducción	7
Capítulo II. Los pilares de la inteligencia artificial	11
Capítulo III. Datos e información	17
Capítulo IV. Reglas de predicción y redes neuronales	23
Capítulo V. Clasificación y discriminación	35
Capítulo VI. Construcción de grupos o <i>clusters</i>	43
Capítulo VII. El reconocimiento de voz	51
Capítulo VIII. Generación de lenguaje natural	57
Capítulo IX. Reconocimiento de imágenes	67
Capítulo X. Sistemas de recomendación	77
Capítulo XI. La inteligencia artificial en la educación	83
Capítulo XII. Inteligencia artificial en finanzas	91
Capítulo XIII. Inteligencia artificial en la banca	99
Capítulo XIV. Conclusiones	105
Sobre el autor	113



## Prefacio

Este libro presenta una introducción a los conceptos de inteligencia artificial (IA) para lectores que quieran iniciarse en este importante tema. Apoya, complementa y amplía, el contenido de una colección de doce videos realizados por el autor para Funcas, con el mismo objetivo y título, dirigidos a presentar los fundamentos de la inteligencia artificial. Los videos pueden verse en la web de Funcas (<https://www.funcas.es/articulos/big-data-aprendizaje-estadistico-automatico-e-inteligencia-artificial/>), y también en Youtube (<https://www.youtube.com/playlist?list=PLQR2nth3aeQRdcUKsluMzKlbfEeeSgzEC>).

Los capítulos del libro están organizados siguiendo la misma secuencia de la serie de videos, pero ampliando el material presentado en ellos con explicaciones más detalladas de algunos aspectos importantes. Se han evitado los detalles técnicos y el texto principal se centra en proporcionar explicaciones intuitivas de los métodos de IA, sin utilizar fórmulas. Sin embargo, en algunos capítulos se ha añadido un apéndice con algunos detalles estadísticos, cuyo seguimiento solo requiere ideas básicas de esta disciplina, para proporcionar explicaciones algo más precisas al lector interesado en profundizar en estos temas.

Comprender las bases de la inteligencia artificial es hoy esencial para poder utilizarla de forma efectiva, interpretar correctamente sus resultados y prepararse para el gran impacto que puede suponer en el futuro para nuestra sociedad. Solo a partir de este conocimiento es posible convertirla en una valiosa herramienta en nuestro trabajo y en nuestro ocio, evaluar sus beneficios, reconocer sus límites y abordar de manera responsable los desafíos éticos y sociales que plantea.

El libro contiene tres partes: los capítulos 2 a 6 incluyen los fundamentos de la IA y las ideas básicas de análisis de datos masivos que se utilizan en sus aplicaciones. Los capítulos 7 a 9 analizan tres ámbitos distintos donde se ha desarrollado la IA para conseguir los objetivos de escuchar, mediante el análisis de sonidos y el reconocimiento de voz; leer y comprender, con el análisis del lenguaje escrito y la creación de sistemas de traducción automática y generación de textos; y ver, mediante el análisis de imágenes o visión artificial, reconociendo personas y objetos y generando bajo demanda nuevas imágenes o videos. Finalmente, los capítulos 10 a 13 analizan algunos ejemplos ele-

gidos para ilustrar las muchas aplicaciones de la inteligencia artificial que afectan ya a nuestras vidas: los sistemas de recomendación, la mejora de la enseñanza y el incremento de eficiencia en procesos económicos y financieros.

Quiero agradecer al director general de Funcas, Carlos Ocaña, su iniciativa y apoyo tanto para la grabación de los videos como para la publicación de este libro, así como su continuo impulso a mis iniciativas en Funcas desde que comenzamos nuestra colaboración hace varios años. En la realización de los videos he tenido la fortuna de contar con el apoyo fundamental de Daniel Conejero en todas las etapas de su elaboración y grabación, y con el excelente trabajo de diseño visual, selección y animación de imágenes y edición de Antonio Caldeiro y José Macías. Además, Daniel ha mejorado el libro con sus comentarios. También lo ha hecho con las sugerencias de Pilar Poncela, mi colaboradora principal en todas las actividades realizadas en Funcas y coautora de los varios libros editados sobre *Big Data* e IA basados en jornadas científicas propiciadas por Funcas. Agradezco mucho a mi hijo Álvaro Peña sus comentarios y el descubrimiento de NotebookLM, con el que he realizado algunas de las imágenes del libro. Sin el apoyo, cariño y generosidad de mi mujer, Jette Bohsen, en todas las etapas de este proyecto, no hubiera podido concluirlo. Finalmente, agradezco por adelantado a los lectores sus comentarios o sugerencias para mejorar tanto este texto como sus videos asociados, y les ruego los envíen a mi dirección de correo electrónico, [Daniel.pena@uc3m.es](mailto:Daniel.pena@uc3m.es).

# CAPÍTULO I

## Introducción

La expresión *inteligencia artificial* ha pasado de ser un término propio de la ciencia ficción, o de ámbitos académicos muy especializados, a formar parte del lenguaje cotidiano. Aparece en conversaciones informales, en titulares de prensa, en campañas publicitarias y en debates políticos. Sin embargo, pese a su uso cada vez más frecuente, no siempre está claro a qué nos referimos exactamente cuando hablamos de inteligencia artificial. En este libro trataremos de explicar, de manera sencilla pero rigurosa, las ideas que se esconden detrás de estos nuevos métodos que van a aparecer, cada vez con más frecuencia, en nuestra vida.

La inteligencia artificial (IA) nace con los ordenadores, en la segunda mitad del siglo XX. El nombre de IA se utilizó por primera vez en una reunión de informáticos en Dartmouth (EE. UU.) en 1955 organizada por McCarthy. Su objetivo era conseguir que los recién llegados ordenadores pudiesen emular las capacidades de los seres humanos, como hablar y entender el lenguaje, ver y comprender imágenes, o escribir y responder preguntas por escrito.

Un avance importante en la década inicial de la IA fue el programa ELIZA, uno de los primeros programas de procesamiento del lenguaje natural (NLP), desarrollado en 1966 por Joseph Weizenbaum, un científico informático del MIT en EE. UU. El programa simulaba una conversación con un humano (chatbox) utilizando patrones de texto para generar respuestas. Su módulo más famoso fue DOCTOR, que imitaba a un psicoterapeuta que escuchaba y hacía preguntas escritas por la consola a su paciente. El sistema era tan convincente que muchos usuarios creyeron que ELIZA realmente “entendía” la conversación y “comprendía” el estado de ánimo del interlocutor, lo que sorprendió incluso al propio autor del programa, que fue de los primeros en alertar sobre los riesgos asociados a este tipo de aplicaciones de la IA.

Muchos de los esfuerzos posteriores de la IA se concentraron en programar máquinas capaces de jugar al ajedrez y otros juegos, donde existían, o podían generarse fácilmente, muchos datos de entrenamiento. Un éxito importante de la IA se produjo en

1997, cuando el programa Deep Blue, desarrollado por IBM, derrotó a Kasparov, campeón del mundo de ajedrez. Este programa se basaba en calcular un número gigantesco de variantes que evaluaba con criterios diseñados por expertos. Podemos concluir que a finales del siglo pasado la inteligencia artificial era una herramienta prometedora, con resultados sorprendentes en juegos complejos, pero de escasa utilidad práctica.

El crecimiento de la IA se produce principalmente en nuestro siglo debido a tres factores: el primero, la aparición de los datos masivos con la digitalización de imágenes, textos, audios y videos, que crea grandes bancos de datos digitales masivos, que se denominan *big data*. El segundo, el enorme crecimiento de la capacidad de cálculo y de transmisión de la información. El tercero, la creación de sistemas de aprendizaje estadístico basados en grandes cantidades de datos que sustituyen a los programas informáticos basados en reglas diseñadas por expertos de razonamiento lógico. La revolución práctica y tangible para la mayoría de los ciudadanos ha sido la llegada de los métodos de la inteligencia artificial generativa (IAG), como ChatGPT, Gemini, Claude o Deepseek. Estos sistemas capaces de generar textos, imágenes o sonidos, a solicitud del usuario, se han convertido en herramientas de uso diario para muchas personas, tanto en su trabajo como en su ocio, provocando un interés general sobre la IA y sus aplicaciones presentes y futuras. Este avance ha situado la inteligencia artificial en el centro del debate público.

Conviene aclarar desde el principio que, pese a su nombre, la inteligencia artificial no implica que las máquinas “piensen” o “entiendan” el mundo del mismo modo que lo hacen los seres humanos, porque los ordenadores no poseen conciencia, emociones ni intención propia. Lo que sí hacen —y cada vez con mayor eficacia— es ejecutar muy rápidamente procedimientos estadísticos y matemáticos extremadamente complejos para detectar patrones en grandes cantidades de información y actuar siguiendo objetivos prefijados en el programa. La potencia de estas técnicas no reside en una supuesta mente artificial, sino en la combinación de datos, capacidad de cálculo y modelos matemáticos avanzados.

Como hemos visto, la revolución práctica y tangible de la inteligencia artificial en nuestro siglo no es fruto de la casualidad, sino de la convergencia de cambios profundos en la tecnología y la metodología. Una vez aclarado que no estamos ante máquinas que “piensan” en el sentido humano, sino ante sistemas que ejecutan procedimientos matemáticos para detectar patrones, los capítulos siguientes profundizan en cómo funciona este engranaje.

El capítulo II explica en qué consisten los métodos de inteligencia artificial (IA en adelante) y desarrolla sus tres pilares ya comentados: el *big data* o datos masivos, la gran capacidad de cálculo de los ordenadores y la extracción, mediante métodos estadísticos y matemáticos, de la información que contienen.

El capítulo III expone los métodos de análisis de los datos digitalizados, así como el uso de indicadores para resumirlos. Se introducen en este capítulo tres objetivos principales del análisis de datos: la predicción, la clasificación supervisada y la clasificación no supervisada.

El capítulo IV presenta los principios básicos para seleccionar reglas de predicción. Se describe una regla básica muy utilizada en la práctica, la regresión lineal, que es el fundamento de las nuevas reglas actuales más potentes basadas en las redes neuronales. También se explica el uso combinado de distintos modelos para mejorar las predicciones. En el apéndice de este capítulo se incluyen algunos de los detalles de estos métodos para facilitar su comprensión al lector interesado, así como una selección de referencias.

El capítulo V introduce el problema de clasificación partiendo de ejemplos sencillos de la vida diaria. Se explican a continuación algunos métodos de clasificación supervisada —a partir de clases prefijadas—, que es una de las técnicas más empleadas en la IA para dar respuesta a todo tipo de necesidades, por su precisión y facilidad de implementación.

El capítulo VI aborda el problema de la clasificación no supervisada, es decir, la división de grupos de datos que no tienen asociada una etiqueta de clasificación predefinida en categorías o clases. Se presentan las ideas de los métodos más utilizados y se analiza, también, cómo la combinación de métodos mixtos, basados en datos supervisados y no supervisados, permite crear métodos muy eficaces de clasificación.

El capítulo VII describe los sistemas de reconocimiento de voz, que tienen numerosas aplicaciones, desde la seguridad bancaria a la atención telefónica. Se analiza cómo los sistemas de IA “entienden” e “identifican” las palabras y cómo el uso de los datos masivos y de métodos potentes de entrenamiento que aprovechan la gran potencia de cálculo de los ordenadores actuales ha llevado a generar sistemas eficientes y flexibles.

El capítulo VIII está dedicado a los modelos de generación de lenguaje, como ChatGPT, Gemini, DeepSeek y otros, que proporcionan respuestas y contenidos a petición de los usuarios, tanto de forma escrita como hablada. La clave de este proceso es transformar las palabras en vectores numéricos que pueden manejarse con un ordenador. A continuación, el sistema puede traducir un texto a otro idioma, interpretando las palabras por el contexto, o generar un nuevo texto a solicitud del usuario.

El capítulo IX explica cómo los sistemas de IA son capaces de reconocer las imágenes, convirtiendo la información numérica, resultado de su digitalización, en patrones identificables para reconocerlas y clasificarlas. Se describen las redes neuronales de con-

volución, capaces de identificar patrones en imágenes y los sistemas generativos de imágenes, que utilizan ideas similares a los que generan textos.

El capítulo X describe una aplicación muy utilizada de la IA, los sistemas de recomendación. Estos métodos promueven la demanda de productos y servicios utilizando datos recogidos de los clientes. Se introducen las técnicas basadas en la historia de compras, de las analogías entre los productos y las analogías entre los clientes, que dan lugar a distintos sistemas de recomendación. También se plantean las críticas que han recibido algunos de estos métodos por el posible uso de información confidencial.

El capítulo XI analiza tres nuevas herramientas con las que la inteligencia artificial puede contribuir a mejorar la enseñanza. El uso de tutores virtuales, el seguimiento por el profesor sobre el proceso digital seguido por cada estudiante en su aprendizaje y el diseño de nuevos sistemas de evaluación adaptados a los medios disponibles hoy para el aprendizaje.

El capítulo XII explora cómo la inteligencia artificial está revolucionando el mundo financiero en dos grandes frentes: mejorando la eficiencia operativa y fortaleciendo la seguridad de sus operaciones. Al mismo tiempo, se analiza que, junto a sus enormes ventajas, el empleo de la IA también supone importantes riesgos y desafíos.

El capítulo XIII ilustra cómo la IA puede mejorar la eficiencia y la seguridad en las actividades de un banco con dos ejemplos concretos. En el primero, se utiliza el análisis de las redes de clientes para mejorar las decisiones financieras. En el segundo, se utilizan estas redes de clientes para mejorar la capacidad predictiva de los modelos y se ilustra con un ejemplo de estimación de la probabilidad de que un cliente entre en mora y no pueda pagar un crédito.

Finalmente, el capítulo XIV incluye algunas conclusiones sobre el futuro de la IA, sus posibles contribuciones positivas al avance de la ciencia y a nuestro bienestar, pero también sus amenazas contra las libertades, la igualdad y nuestros derechos. Se resalta la importancia de controlar su implantación con una adecuada supervisión de los poderes públicos democráticos.

# CAPÍTULO II

## Los pilares de la inteligencia artificial

### 2.1. INTRODUCCIÓN

Para comprender por qué la inteligencia artificial ha experimentado un desarrollo tan acelerado en las últimas décadas, es fundamental analizar los tres pilares sobre los que se sustenta.

El avance reciente de la inteligencia artificial no es fruto de un único descubrimiento revolucionario ni de un momento concreto de genialidad tecnológica. Por el contrario, es el resultado de la convergencia progresiva de varios factores que han evolucionado de forma paralela durante las últimas décadas. Esta coincidencia temporal y tecnológica ha sido decisiva para que la inteligencia artificial pase de ser un campo principalmente teórico o experimental a convertirse en una tecnología central de nuestra sociedad.

Estos factores pueden agruparse en tres grandes pilares fundamentales:

1. La disponibilidad de datos masivos.
2. El desarrollo de tecnologías de almacenamiento y procesamiento de la información.
3. La extracción matemática y estadística del conocimiento contenido en los datos.

La combinación de estos tres elementos ha creado un ecosistema en el que las máquinas no solo almacenan información, sino que pueden aprender de ella, detectar patrones y tomar decisiones de manera automática. La [figura 2.1](#) ilustra gráficamente estos tres pilares.

FIGURA 2.1

LOS PILARES DE LA INTELIGENCIA ARTIFICIAL



Fuente: Elaboración propia con ayuda de Chatgtp, Gemini y DeepSeek.

## 2.2. LOS DATOS MASIVOS

El primer pilar de la inteligencia artificial moderna es la existencia de datos masivos, comúnmente conocidos como *big data*. Por primera vez en la historia de la humanidad, disponemos de cantidades inmensas de datos sobre prácticamente cualquier aspecto de la actividad humana y del entorno que nos rodea.

Estos datos no se generan de forma puntual ni excepcional, sino de manera continua y, en gran medida, automática. Cada vez que utilizamos un teléfono móvil, efectuamos una búsqueda en internet, interactuamos en redes sociales, pagamos con una tarjeta de crédito o nos desplazamos con un dispositivo GPS, estamos produciendo información. A este flujo constante se suman los datos recogidos por sensores de todo tipo: cámaras, micrófonos, teléfonos móviles, satélites, dispositivos médicos, sensores industriales, relojes inteligentes o estaciones meteorológicas.

A diferencia de épocas anteriores, en las que la información era escasa, difícil de obtener y costosa de almacenar, hoy vivimos en un contexto de sobreabundancia informativa. En la actualidad, la población mundial ronda los 8.000 millones de personas y cerca del 70 % utiliza internet y teléfonos móviles de manera habitual. Esta conectividad global genera un volumen de información sin precedentes, tanto en cantidad como en diversidad.

Los datos disponibles comienzan a aumentar de forma vertiginosa a finales del siglo XX con la aparición de Internet y los primeros sensores capaces de mediciones automáticas. Timothy Berners-Lee a finales del siglo XX inventó la World Wide Web, haciendo posible el uso de Internet para cualquiera con acceso a un ordenador. La aparición de Google,

en 1998, que se ha convertido en el buscador más utilizado en nuestro siglo, contribuyó a hacer disponibles en internet una cantidad creciente de datos, estimulando a nuevos usuarios de la red que fueron dejando sus propios datos e incrementando el tráfico en internet. Los sensores digitales conectados a los ordenadores se introdujeron en los años 70 del siglo pasado, pero el desarrollo de internet y de las redes de comunicación ha llevado en este siglo al llamado *Internet de las cosas*, con miles de millones de sensores activos recogiendo información en todos los campos.

El crecimiento del volumen de datos generados ha sido asombroso. En la actualidad, la cantidad diaria de datos generada hoy equivale a que cada ser humano produjera cada día una información igual a dos veces y media todos los libros de la biblioteca del Congreso de EE. UU. (unos 20TB). Hoy, más del 90 % de la información mundial existe exclusivamente en formato digital, y una gran parte de ella es generada automáticamente por máquinas. A este ritmo, en un futuro cercano, la cantidad de información disponible en la red superará con creces la información contenida en todos los cerebros humanos juntos.

El verdadero desafío en la actualidad ya no consiste en recopilar datos, como ocurría en el siglo pasado, sino en gestionarlos, organizarlos y extraer valor de ellos. Los grandes volúmenes de datos sin un análisis adecuado no solo carecen de utilidad, sino que pueden convertirse en una carga difícil de manejar.

El impacto de los datos masivos se observa con especial claridad en el ámbito empresarial. Cuatro de las cinco mayores empresas tecnológicas del mundo —Apple, Microsoft, Alphabet (Google) y Amazon— deben gran parte de su crecimiento y su poder económico a la capacidad de recopilar, analizar y explotar datos para orientar sus estrategias, productos y servicios.

En particular, Alphabet y Amazon nacieron ya en el siglo XXI con modelos de negocio basados desde su origen en el aprovechamiento sistemático de la información. Los datos permiten optimizar procesos, personalizar servicios, anticipar comportamientos de los usuarios y tomar decisiones basadas en evidencia cuantitativa en lugar de intuiciones.

Sin llegar al tamaño de los gigantes tecnológicos anteriores, otras empresas centradas en utilizar los datos de sus usuarios para hacer recomendaciones y mejorar sus políticas, como Netflix en videos o Spotify en música, se han convertido en líderes de su sector y acumulan una audiencia mundial de cientos de millones de suscriptores.

En este contexto, los datos se han convertido en un recurso estratégico comparable—y en muchos casos superior— a las materias primas tradicionales. Sin datos suficientes, los sistemas de inteligencia artificial carecen de la base necesaria para aprender, mejorar y

adaptarse. La calidad, diversidad y representatividad de los datos condicionan directamente el rendimiento de cualquier sistema inteligente.

### 2.3. ALMACENAMIENTO Y PROCESAMIENTO DE LOS DATOS

El segundo pilar de la inteligencia artificial es el desarrollo de tecnologías capaces de almacenar y procesar estas enormes masas de datos. Este avance ha sido posible gracias a la evolución simultánea de los ordenadores, las redes de comunicación y los sistemas distribuidos.

Los lectores más veteranos recordarán que el almacenamiento y la capacidad de procesamiento de los ordenadores personales se medían inicialmente en kilobytes (KB), miles de bytes, donde un byte representa un carácter, como una letra o un número. Poco después se pasó a los megabytes (MB), que permitían guardar textos extensos y libros completos, y más tarde a los gigabytes (GB), necesarios para manejar imágenes y vídeos.

Desde hace algunos años, es habitual adquirir dispositivos de almacenamiento de varios terabytes (TB), equivalentes a miles de gigabytes, capaces de guardar no solo documentos personales, sino bibliotecas enteras, colecciones audiovisuales completas y grandes volúmenes de información histórica. A modo de comparación, se estima que la memoria humana puede almacenar entre 10 y 70 terabytes de información.

Este crecimiento no habría sido posible sin la aparición de grandes redes de ordenadores interconectados, capaces de trabajar de forma coordinada. La computación distribuida permite dividir tareas complejas en miles o millones de operaciones más pequeñas que se ejecutan en paralelo.

Gracias a este enfoque, los sistemas de inteligencia artificial pueden entrenarse sobre conjuntos de datos gigantescos en tiempos razonables. Lo que antes habría requerido años de cálculo en un único ordenador, hoy puede resolverse en horas o días utilizando infraestructuras distribuidas.

Respecto a la transmisión de datos, hemos pasado de una velocidad promedio en el año 2000 de un megabyte por segundo a un gigabyte en la actualidad, lo que supone multiplicar por mil la velocidad media de transmisión. Pero el cambio más importante es en la velocidad de computación, que se ha multiplicado por 1.000.000. La velocidad de un PC en el año 2000 era de unos  $2 \times 10^9$  flops, es decir, podía realizar dos mil millones de operaciones de coma flotante en un segundo. Tomando de nuevo como referencia una población de la Tierra de ocho mil millones de personas, la capacidad de cálculo de un PC en el año 2000 era equivalente a tener a  $\frac{1}{4}$  de la población mundial ( $8 \times 10^9$  personas) haciendo una operación cada segundo con una calculadora manual.

Hoy podemos utilizar un ordenador PC CES (NVIDIA) con una velocidad de 300TFLOP=  $3 \times 10^{15}$  flops, equivalente a la capacidad de cálculo de todos los habitantes de 1 millón de planetas como la Tierra.

Este pilar tecnológico es, en gran medida, invisible para el usuario final. Sin embargo, constituye el músculo que sostiene a la inteligencia artificial moderna y permite que sus aplicaciones funcionen de forma fluida y casi instantánea.

## 2.4. EXTRACCIÓN ESTADÍSTICA DE LA INFORMACIÓN

El tercer pilar fundamental de la inteligencia artificial es la aplicación de métodos matemáticos —principalmente estadísticos— para extraer la información contenida en los datos. No basta con almacenar datos: es necesario transformarlos en conocimiento útil.

Como veremos en los capítulos siguientes, los métodos estadísticos permiten identificar patrones, establecer relaciones, clasificar información y realizar predicciones de forma automática. A través de estos métodos la inteligencia artificial puede estimar probabilidades, detectar anomalías y tomar decisiones basadas en datos históricos.

La estadística ha pasado a ocupar un papel central en la toma de decisiones automatizada. Gracias a ella, los sistemas inteligentes pueden aprender de la experiencia pasada y mejorar su rendimiento con el tiempo.

Es importante subrayar que esta “inteligencia” no implica comprensión en el sentido humano. Los sistemas no entienden el significado de los datos, sino que operan sobre correlaciones matemáticas. Aun así, estas correlaciones pueden ser lo suficientemente precisas como para resultar extremadamente útiles en contextos prácticos.

## 2.5. IMPORTANCIA ACTUAL DE LA IA

La combinación de datos masivos, capacidad de almacenamiento y análisis estadístico está transformando nuestra vida cotidiana de manera profunda. Estos cambios afectan a campos esenciales y diversos como el aprendizaje, la salud, el trabajo, el ocio e incluso las relaciones personales.

En el ámbito educativo, el análisis de datos permite personalizar los procesos de aprendizaje, identificar dificultades específicas y adaptar los contenidos al ritmo de cada estudiante. En la salud, el *big data* facilita diagnósticos más precisos, tratamientos personalizados y la detección temprana de enfermedades.

En el mundo laboral, los datos optimizan procesos, mejoran la productividad y conducen a mejores decisiones basadas en la evidencia empírica, en lugar de en la intuición y las corazonadas. En el ocio, los sistemas de recomendación sugieren contenidos adaptados a nuestros gustos y hábitos. Incluso en las relaciones personales, los algoritmos influyen en la forma en que nos comunicamos y conocemos a otras personas.

También la economía y las finanzas han experimentado una transformación radical. Los tradicionales parqués bursátiles, donde los agentes negociaban acciones de forma presencial, han sido sustituidos en gran medida por salas de ordenadores interconectados que realizan millones de transacciones mediante algoritmos en fracciones de segundo.

En conjunto, la IA está dando lugar a una nueva forma de entender la información y el conocimiento. No sustituye al pensamiento humano, pero sí amplifica nuestra capacidad para analizar la realidad y actuar sobre ella. Comprender sus fundamentos resulta esencial para interpretar correctamente su impacto presente y futuro en nuestra sociedad.

# CAPÍTULO III

## Datos e información

### 3.1. LOS NUEVOS DATOS DIGITALES

Durante siglos, la noción de dato estuvo estrechamente ligada a la idea de medición. Hasta nuestro siglo, los datos se entendían fundamentalmente como listas de números, obtenidos al medir una o varias variables de interés. Por ejemplo las bases de datos contenían ventas mensuales de una empresa, precios de activos financieros, temperaturas registradas a lo largo del tiempo o respuestas de una muestra de ciudadanos a una encuesta.

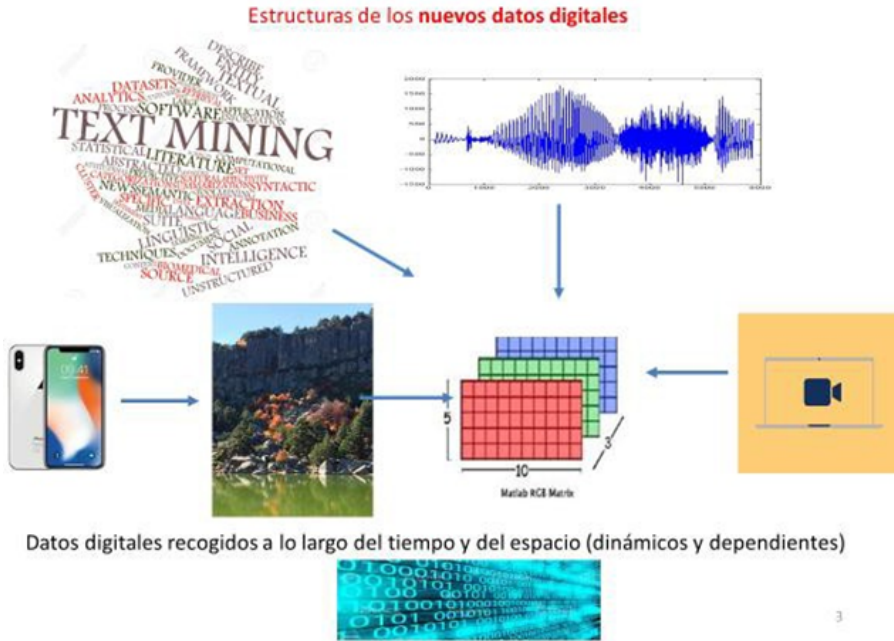
Estos datos solían organizarse en tablas o matrices, con una estructura clara y definida. Cada fila representaba una observación, cada columna una variable, y el significado de cada número estaba perfectamente establecido desde el inicio. La información se obtenía, además, de forma mayoritariamente manual: mediante la observación directa de la realidad, la realización de experimentos cuidadosamente diseñados o la elaboración de encuestas.

Este proceso implicaba un coste significativo, tanto económico como temporal. Recoger datos era caro, lento y laborioso, lo que limitaba su volumen y frecuencia. Por esta razón, los datos se obtenían con un objetivo muy concreto y estaban generalmente bajo el control exclusivo de quien los había recopilado. El acceso a la información era restringido y su reutilización, poco habitual.

Este concepto tradicional de datos ha cambiado de forma radical en las últimas décadas. Hoy en día, se consideran datos no solo las mediciones numéricas clásicas, sino cualquier elemento que contenga información sobre un fenómeno de interés y que podamos digitalizar para utilizarlo en un ordenador. Un texto escrito, los clics realizados en una página web, una imagen o un vídeo captados con un teléfono móvil, la grabación de una conversación, una trayectoria GPS o, incluso el tiempo que una persona permanece observando un contenido, son también datos. En otras palabras, cualquier rastro digital susceptible de aportar información es potencialmente analizable si lo digitalizamos. La [figura 3.1](#) presenta ejemplos de los nuevos datos.

FIGURA 3.1

EJEMPLOS DE DATOS DIGITALES. EN LA PRIMERA FILA, TEXTOS Y SONIDOS, EN LA SEGUNDA EL PROCESO DE DIGITALIZACIÓN DE UN IMAGEN REPRESENTÁNDOLA EN EL SISTEMA RGB, EN LA TERCERA LOS DATOS DIGITALES RESULTANTES



Fuente: Elaboración propia.

A diferencia de los datos clásicos, estos nuevos datos no suelen estar estructurados, y se generan de forma automática y continua. Con frecuencia estos datos se utilizan, además, con múltiples fines distintos a los originales y suelen ser accesibles a un gran número de actores, bien de forma gratuita en internet o mediante pago o suscripción en la web.

Este cambio ha multiplicado exponencialmente el volumen y la diversidad de la información disponible. La digitalización de la sociedad ha convertido actividades cotidianas en fuentes constantes de datos, muchas veces sin que los propios individuos sean plenamente conscientes de ello.

### 3.2. LA EXTRACCIÓN DE CARACTERÍSTICAS DE DATOS DIGITALES

Para que los nuevos datos puedan ser analizados por un ordenador, es necesario un paso intermedio fundamental: su transformación en datos digitales estructurados. Los

sistemas informáticos no pueden trabajar directamente con significados, imágenes o sonidos, sino con representaciones numéricas de los mismos.

El primer paso consiste, por tanto, en traducir esa información heterogénea y desordenada a un formato que pueda ser procesado algorítmicamente. Este proceso implica seleccionar, resumir y codificar la información relevante, descartando aquello que no resulta útil para el objetivo del análisis.

Aquí aparece uno de los conceptos clave de la ciencia de datos moderna: la extracción de características. Estructurar la información significa resumirla en un conjunto de variables numéricas que capturen sus aspectos más relevantes. Las características son, en esencia, descriptores cuantitativos que permiten representar información compleja de manera simplificada.

La forma concreta de extraer la información relevante depende de dos factores principales: el tipo de información disponible y el objetivo del análisis. No existe una única manera correcta de estructurar los datos; la elección depende del problema que se desea resolver. Consideremos, por ejemplo, los correos electrónicos que intercambia un banco con sus clientes. Estos correos son textos no estructurados, pero pueden transformarse en datos analizables mediante distintos procedimientos. Una opción habitual consiste en seleccionar palabras clave relevantes y contabilizar su frecuencia de aparición, construyendo tablas que indiquen cuántas veces aparece cada término en cada mensaje. Este enfoque permite convertir textos en vectores numéricos comparables entre sí. A partir de ellos, es posible detectar patrones, clasificar mensajes o identificar temas recurrentes.

Un razonamiento similar puede aplicarse a los audios generados en conversaciones telefónicas entre clientes y una entidad financiera. Estas grabaciones pueden convertirse primero en texto mediante sistemas de reconocimiento del habla y, posteriormente, analizarse del mismo modo que los correos electrónicos. Además, pueden incorporarse variables adicionales, como la duración de la llamada, la fecha, la hora o la frecuencia de contacto.

Dado el gran volumen de información disponible, suele resultar conveniente condensar los datos originales en medidas resumen. Estas medidas reciben distintos nombres según el contexto: índices, indicadores o componentes.

### 3.3. RESUMIR LOS DATOS EN INDICADORES

Los índices se construyen normalmente ponderando distintas variables en función de su importancia relativa. Cuando todas las variables tienen el mismo peso, el índice puede ser tan simple como la media aritmética o promedio de los datos. En otros casos se

calcula una media ponderada, dando ponderaciones distintas a las variables. Esto es lo que ocurre, por ejemplo, en indicadores económicos que utilizamos con frecuencia, como el índice de precios al consumo, donde los precios de los productos se ponderan por su peso en la cesta de la compra, o en los indicadores bursátiles, como el índice Ibex, que resume el precio de un conjunto de acciones ponderando cada una en la suma por una medida de su importancia relativa.

Cuando no se conoce de antemano la importancia de cada variable, la ponderación puede determinarse empíricamente, atendiendo a la capacidad predictiva de las variables originales. Es decir, se eligen los pesos de forma que el índice resultante sea capaz de reproducir lo mejor posible la información contenida en el conjunto inicial de variables. Por ejemplo, podemos hacer un indicador del precio de las viviendas en una zona ponderando las variables que condicionan su precio, como su situación, tamaño, número de habitaciones, etc., de manera que el índice obtenido sea capaz de prever lo mejor posible el precio real de la vivienda. En este caso, el índice resultante también recibe el nombre de componente principal, o componente capaz de prever. Este tipo de técnicas permite reducir la complejidad de los datos sin perder —o perdiendo lo menos posible— información relevante.

En el contexto de datos no estructurados como los correos electrónicos, un indicador resumen puede ser la contabilización de palabras clave, transformando el texto en vectores numéricos procesables. En el análisis de textos se puede utilizar una técnica similar a componentes principales para resumir matrices lingüísticas, que contienen la frecuencia relativa de aparición de palabras conjuntamente en un idioma, representando palabras a través de factores que condensan su significado o contexto, de manera que palabras similares tengan coordenadas similares indicando su proximidad.

Estos ejemplos ilustran cómo la ciencia de datos transforma la información heterogénea en descriptores cuantitativos que permiten representar realidades complejas de manera simplificada. En los capítulos siguientes veremos ejemplos concretos de la digitalización de sonidos, imágenes y textos.

Además del tipo de información, el segundo factor clave en la estructuración de los datos es el objetivo del análisis, como comentamos en la sección siguiente.

### **3.4. OBJETIVOS DEL ANÁLISIS**

De forma general, pueden distinguirse dos grandes tipos de objetivos en el análisis de un conjunto de datos. El primero es la predicción y el segundo, la clasificación, que puede ser supervisada o no supervisada.

Las reglas de predicción tienen como finalidad prever el valor futuro de una o varias variables de interés. Un ejemplo típico es estimar el volumen de negocio que puede generar un cliente bancario en el futuro. La variable que se desea prever suele depender de un conjunto amplio de otras variables, que llamaremos explicativas. Algunas están bajo el control de la entidad —como las condiciones ofrecidas al cliente—, mientras que otras dependen de factores externos o inciertos. El objetivo es construir modelos que, a partir de la información disponible, permitan realizar previsiones lo más precisas posible.

Un segundo grupo de herramientas se emplea en problemas de clasificación, que pueden dividirse en dos tipos. El primero es la clasificación supervisada, también llamada discriminación, donde los datos disponibles incluyen ejemplos previamente etiquetados o clasificados, y el objetivo es aprender a asignar correctamente nuevas observaciones a una de las categorías existentes. Ejemplos clásicos de este tipo de problemas son reconocer el lenguaje hablado, asignar un sonido a una palabra o clasificar una solicitud de crédito como buena, dudosa o de alto riesgo. El sistema aprende a partir de ejemplos conocidos y generaliza ese conocimiento a las nuevas situaciones para clasificarlas por analogía con las anteriores.

El segundo tipo de clasificación es la clasificación no supervisada o agrupación en *clusters*, y aparece cuando los datos no disponen de etiquetas previas. En este caso, partimos de un conjunto de elementos —observaciones, textos, imágenes o videos— y se desea dividirlos en grupos internamente homogéneos, pero distintos entre sí. Un ejemplo sería agrupar activos financieros teniendo en cuenta la evolución dinámica de su rendimiento, sin imponer categorías predefinidas. El objetivo es descubrir estructuras ocultas en los datos que ayuden a comprender mejor el fenómeno analizado.

En los capítulos siguientes se abordará cada uno de estos métodos con mayor detalle, analizando sus fundamentos, sus aplicaciones y sus limitaciones. Comprender cómo se transforman los nuevos datos en conocimiento es esencial para entender el verdadero alcance de la inteligencia artificial en nuestra sociedad.

### 3.5. LECTURAS ADICIONALES

Se indican a continuación algunos libros de introducción a la estadística orientados a las aplicaciones.

- Anderson, David R., Sweeney, Dennis J. y Williams, Thomas A. (2019). *Estadística para negocios y economía*. Editorial: Cengage Learning.  
Una buena introducción a la estadística para las ciencias sociales

- Cao Abad, Ricardo, Francisco Fernández, Mario, Naya Fernández, Salvador. (2001). *Introducción a la estadística y sus aplicaciones*. Editorial: Ediciones Pirámide.  
Libro orientado tanto a estudiantes como a profesionales, que combina conceptos teóricos con aplicaciones prácticas, ejercicios resueltos y propuestos, con un enfoque en la interpretación de resultados más que en la complejidad matemática.
- Knaflic, S. N. (2020). *Storytelling with Data: Let's Practice!*. Wiley.  
Aunque centrado en la comunicación de resultados, ofrece una perspectiva clara sobre la importancia de resumir datos complejos mediante indicadores y medidas sintéticas que faciliten su interpretación y toma de decisiones.
- Montgomery, Douglas C. y Runger, George C. (2015). *Probabilidad y estadística aplicadas a la ingeniería*. Editorial: Limusa Wiley.  
Un buen texto clásico orientado a las aplicaciones.
- Peña, Daniel (2002). *Análisis de datos multivariantes*. Editorial: McGraw-Hill Interamericana.  
Texto más avanzado que describe con detalle varios de los métodos de clasificación.
- Peña, Daniel. (2014). *Fundamentos de Estadística*. Editorial: Alianza Editorial.  
Texto de introducción a la estadística aplicada, cubriendo desde el análisis exploratorio de datos hasta la construcción de modelos probabilísticos, con ejemplos orientados al uso profesional y académico.
- Ugarte, M. D., Militino, A. F., & Arnholt, A. T. (2015). *Probability and Statistics with R* (2.<sup>a</sup> ed.). CRC Press.  
Buen texto de introducción a la estadística utilizando el lenguaje R, muy usado en las aplicaciones estadísticas.

Una explicación más detallada de la relación entre la estadística y la inteligencia artificial, indicando su confluencia con la aparición de los datos masivos se encuentra en el artículo:

Peña, D. (2026). Datos masivos: de la Estadística a la Inteligencia Artificial. *Revista de la Real Academia de Ciencias Serie General (RACSG)*, 114 (en prensa).

## CAPÍTULO IV

### Reglas de predicción y redes neuronales

#### 4.1. TIPOS DE PROBLEMAS DE PREDICCIÓN

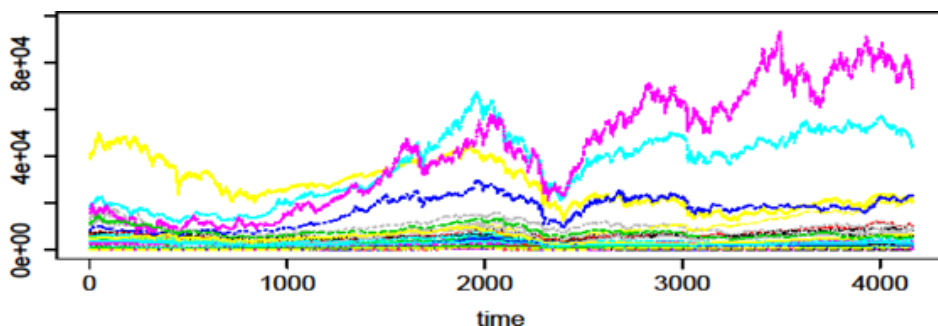
Supongamos que deseamos prever los valores futuros de una variable numérica resultado de una medición. La variable puede ser estática, entendiéndose por ello que es estable y está asociada a una situación que permanece aproximadamente constante en el tiempo, como el número de goles en un partido de fútbol, o dinámica, porque sus valores evolucionan en el tiempo, como la cantidad de lluvia que caerá mañana en una ciudad. Podemos considerar dos métodos para realizar esta predicción. La primera es la predicción univariante, que utiliza solo los valores pasados de la variable a prever, suponiendo que han sido recogidos en circunstancias similares a las de la predicción que queremos hacer. El segundo método es una predicción con variables explicativas, cuando, además, disponemos de los datos de otras variables relacionadas con las que queremos prever y que, confiamos, pueden ayudarnos a realizar predicciones más precisas. Por ejemplo, para prever la cantidad de lluvia mañana en Madrid, podemos utilizar, además de la lluvia en Madrid en los días anteriores, las cantidades caídas en otras ciudades cercanas a Madrid.

La **figura 4.1** presenta la evolución a lo largo del tiempo de los valores que han tomado un conjunto de variables, representadas por colores distintos. Estas secuencias de valores de una variable se denominan series temporales. Para prever los valores futuros de una serie, por ejemplo la azul, podemos utilizar solo los valores históricos de esta variable. Sin embargo, la figura indica que las series están relacionadas entre sí, ya que varias tienden a crecer juntamente con el tiempo, y podemos mejorar las predicciones univariantes de una serie utilizando, además de su propia historia, su relación con las demás.

Los procedimientos que vamos a estudiar a continuación consideran el caso más frecuente: disponemos de una variable a predecir, que llamaremos variable dependiente, y un conjunto de otras variables explicativas relacionadas con la anterior, que llamaremos variables explicativas o independientes. El objetivo es construir una regla de predicción tan buena como sea posible.

FIGURA 4.1

UN CONJUNTO DE SERIES TEMPORALES RELACIONADAS ENTRE SÍ



Fuente: Elaboración propia.

## 4.2. COMPARACIÓN DE REGLAS DE PREDICCIÓN

Para elegir entre posibles reglas de predicción el primer paso es establecer un criterio para evaluar la calidad de la predicción que proporcionan. Dada una regla de predicción, por ejemplo, tomar el último valor observado de la serie, o utilizar un promedio ponderado de los últimos valores de las variables explicativas, etc. (veremos después cómo obtener reglas de predicción), se define el error de predicción como la diferencia entre el valor observado y el previsto. Un procedimiento habitual para comparar las reglas de predicción es aplicarlas a los datos pasados, calcular sus errores y seleccionar la que produzca el mínimo error promedio en los datos observados.

Como los errores de predicción pueden ser positivos o negativos para que no se compensen entre sí los elevamos al cuadrado para hacerlos siempre positivos y después los sumamos, extraemos la raíz cuadrada de la suma y dividimos por el total de datos. Con esta medida promedio del error, que se denomina error cuadrático medio, podemos comparar el error promedio cometido con distintas reglas de predicción.

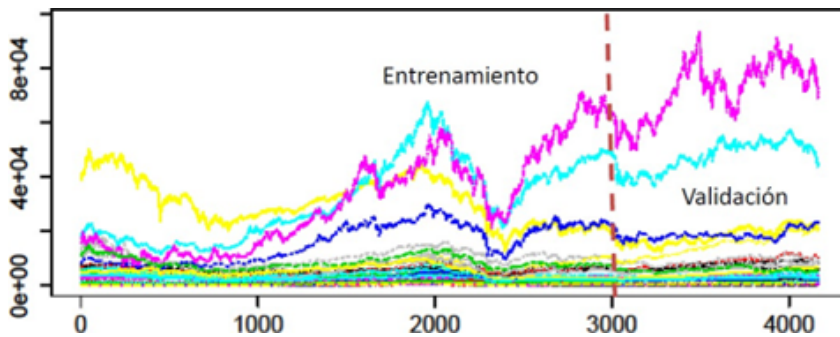
Un problema con este enfoque es que estamos utilizando los mismos datos para calcular la regla de predicción y después calcular sus errores, cuando la regla de predicción se ha construido para que estos sean lo más pequeños posibles. Por lo tanto el error cuadrático medio calculado en los datos tiende a darnos una versión optimista del error que tendremos con nuevos datos, que es lo que ocurrirá cuando apliquemos la regla elegida a valores futuros desconocidos.

Para evitar este problema y seleccionar la mejor regla de predicción simulando una predicción real, podemos dividir la muestra de datos disponible en dos partes. Con la primera, la muestra de entrenamiento, se estima la regla de predicción, obteniendo los

pesos de las variables explicativas que producen mejores menores errores en el ajuste. Con la segunda, la muestra de validación, se calcula el error de predicción de las reglas consideradas con valores nuevos no utilizados. La regla seleccionada es la de menor error de predicción en la muestra de validación. En algunos casos, la muestra de entrenamiento se subdivide además en otras dos: en la primera se estiman los parámetros de un modelo fijada su estructura, y en el segunda se elige esta estructura, seleccionando los llamados hiperparámetros que la definen. Este proceso se ilustra en la **figura 4.2**, donde los datos se han dividido en una muestra de entrenamiento y otra de validación.

FIGURA 4.2

## DIVISIÓN DE DATOS TEMPORALES ENTRE ENTRENAMIENTO Y VALIDACIÓN



Fuente: Elaboración propia.

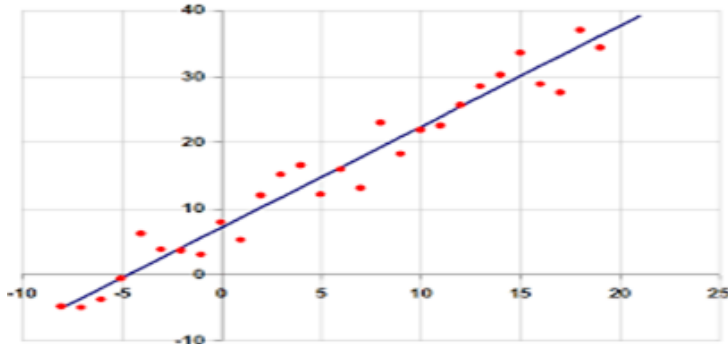
### 4.3. LA REGRESIÓN LINEAL

Una forma muy utilizada de construir una regla de predicción de una variable numérica cuando disponemos de un conjunto de variables explicativas es la regresión lineal. Con este procedimiento, introducido por Gauss y Legendre a principios del siglo XIX, se predice la variable dependiente buscando la combinación lineal, o indicador resumen de todas las variables explicativas, que produce el menor error de predicción (véase el apéndice de este capítulo para una explicación más precisa). Los pesos de cada variable en esta combinación lineal se obtienen por el método estadístico de los mínimos cuadrados, que minimiza el error cuadrático medio, es decir, el promedio de la suma de los cuadrados de los errores del ajuste del modelo a los datos observados. La **figura 4.3** presenta una recta de regresión cuando existe una única variable explicativa y una relación lineal entre la variable a predecir y la explicativa.

Cuando tenemos muchas variables explicativas, y representaremos su número con la letra  $p$ , y las variables por los símbolos  $x_1, x_2, \dots, x_p$ , donde  $p$  puede ser muy grande, pero menor que el número de datos disponible para las variables, la regresión predice el

FIGURA 4.3

RELACIÓN LINEAL ENTRE DOS VARIABLES Y LA RECTA DE REGRESIÓN ENTRE ELLAS



Fuente: Elaboración propia.

valor de la variable respuesta o variable de salida combinando las variables de entrada de forma lineal. Es decir, se crea una nueva variable resumen de todas ellas dando pesos,  $b_1, b_2, \dots, b_p$ , a cada una de las variables de entrada y sumándolos para obtener la variable prevista  $b_0 + b_1x_1 + \dots + b_px_p$ , donde el parámetro  $b_0$  se denomina la constante del modelo y los parámetros  $b_i$  determinan la importancia de cada variable de entrada  $x_i$  en la predicción.

La regresión es adecuada cuando las relaciones son lineales, lo que suele ocurrir con variables muy agregadas que son la suma de muchas observaciones, como las ventas anuales explicadas por el gasto anual en consumo, pero con variables muy desagregadas, como las ventas en la siguiente hora en función de la temperatura, las relaciones son con frecuencia de tipo no lineal, lo que supone una limitación a este método. Por ejemplo, la regresión puede funcionar bien para prever las ventas mensuales de una gasolinera, pero será menos eficiente para prever las ventas diarias, que estarán afectadas de forma no lineal por variables climáticas. Un procedimiento general para prever con una relación no lineal entre una variable y un conjunto de variables explicativas son las redes neuronales, que se explican a continuación.

#### 4.4. LAS REDES NEURONALES

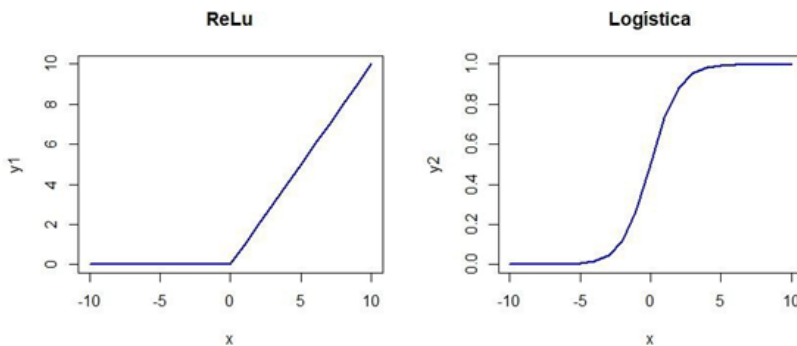
Las redes neuronales tienen una larga historia que se inicia a mediados del siglo pasado, pero se han ido generalizando y ampliando hasta convertirse en la herramienta central para resolver problemas en inteligencia artificial. Rosenblatt (1958), un psicólogo americano, concibió un modelo de aprendizaje, el perceptrón, que recibe un conjunto de  $p$  señales de entrada, que representaremos por los símbolos  $x_1, x_2, \dots, x_p$ , y construye una predicción de una variable de salida,  $y$ , en dos pasos. En el primero, las variables de

entrada se combinan de forma lineal, como en la regresión, es decir, se crea una nueva variable resumen de todas ellas  $h = b_o + b_1x_1 + \dots + x_p b_p$ , donde el parámetro  $b_o$  se denomina sesgo y los parámetros  $b_i$  determinan el peso de cada variable de entrada  $x_i$  en la variable resumen,  $h$ . El segundo paso obtiene la predicción de la variable respuesta,  $y$ , aplicando una función no lineal a la variable resumen de la entrada  $h$ ,  $y=f(h)$ . Esta función no lineal suele llamarse función de activación, y determina el tipo de salida de la neurona.

La [figura 4.4](#) presenta dos ejemplos habituales de funciones de activación. La elegida inicialmente en el perceptrón fue la función logística. Otra función muy utilizada es la función lineal truncada o ReLU, *rectified linear unit*, donde la salida toma el valor cero si la variable de entrada es negativa o cero y reproduce el valor de la entrada si esta es positiva.

**FIGURA 4.4**

**LA FUNCIÓN DE ACTIVACIÓN RELU Y LA SIGMOIDE O LOGÍSTICA, FUNCIONES DE ACTIVACIÓN HABITUALES EN REDES NEURONALES**



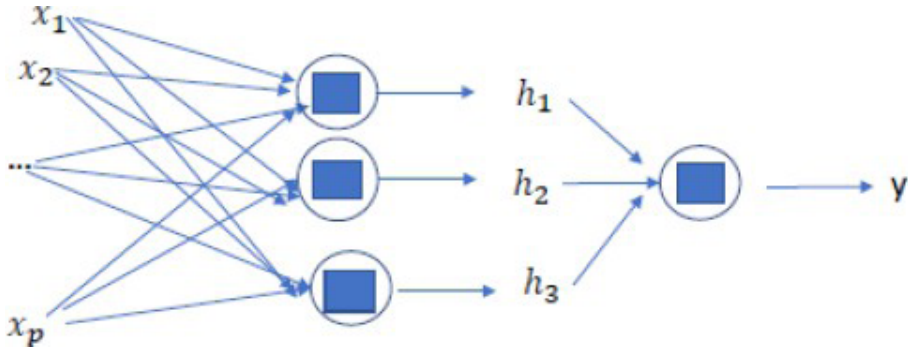
Fuente: Elaboración propia.

Los pesos de la neurona,  $b_0, b_1, b_2, \dots, b_p$ , se calculan iterativamente minimizando el error promedio cometido. Por ejemplo, si disponemos de  $n$  valores de dos variables de entrada  $x(1, i), x(2, i)$ , y de la variable de salida,  $y(i)$ , donde  $i=1, \dots, n$  y suponemos una función de activación logística, los pesos se determinan tomando un valor inicial para los pesos, calculando el error promedio cometido, viendo qué cambios en los pesos reducen el error y modificando su valor para reducirlo. Este ciclo se repite hasta que no sea posible mejorar la predicción.

Una red neuronal artificial, RNA, es un desarrollo del perceptrón combinando varias neuronas para construir una respuesta con todas ellas. La [figura 4.5](#) representa un ejemplo con una red de tres neuronas. Cada una de ellas produce una variable resumen,  $(h_i)$ , a la que se aplica la función de activación para formar la variable salida de la neu-

FIGURA 4.5

UNA RED NEURONAL CON TRES NEURONAS



Fuente: Elaboración propia.

rona. Después, las salidas de las tres neuronas se combinan linealmente para proporcionar la previsión de la variable  $y$ . Puede demostrarse que una red neuronal con varias neuronas en una única capa y función de activación tipo sigmoide puede, al aumentar el número de neuronas, aproximar cualquier función continua.

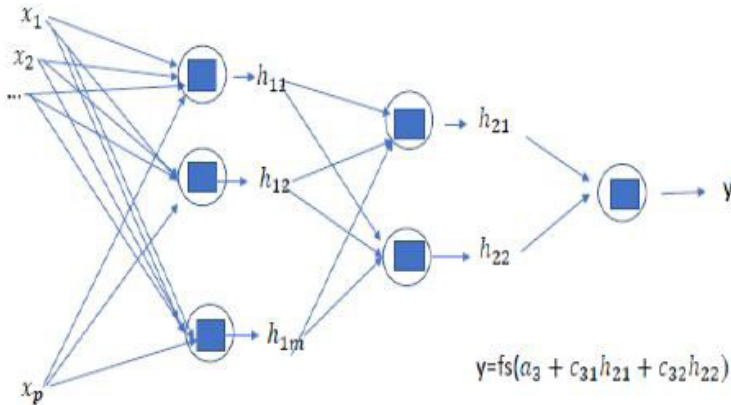
## 4.5. LAS REDES NEURONALES PROFUNDAS (DEEP LEARNING)

El siguiente paso en el avance de las redes neuronales para modelar relaciones complejas entre las entradas y las salidas es utilizar distintas capas de neuronas, como se representa en la [figura 4.6](#), donde las tres neuronas de la primera capa proporcionan respuestas,  $h_{11}$ ,  $h_{12}$ ,  $h_{13}$ , que son las entradas de la segunda capa. En ella se combinan las dos respuestas generadas por las dos neuronas de esta capa,  $h_{21}$ ,  $h_{22}$ , para crear la variable de salida,  $y$ , o respuesta final de la red. Puede demostrarse que estas redes pueden aproximar cualquier función continua y, al aumentar el número de capas, consiguen en la práctica aproximar muy bien cualquier tipo de relación compleja entre las variables. Con estas bases, durante nuestro siglo, el crecimiento de las redes con muchas capas, o *Deep learning*, aprendizaje profundo, ha sido espectacular.

Este tipo de redes con entradas que se procesan por todas las neuronas hasta la salida se utiliza mucho para la predicción  $y$ , y con frecuencia, al hablar de redes neuronales artificiales —RNA— se piensa en ellas. Sin embargo, existen otros tipos de redes para análisis específicos que veremos en los siguientes capítulos y el nombre preciso de estas redes, como las de la [figura 4.6](#), es de redes Perceptrones multicapa, o, también, FF o *Feed-Forward*, porque la información de las variables de entrada fluye siempre hacia adelante, en dirección a la capa de salida. Existen otros tipos de redes neuronales adaptados a objetivos específicos. Por ejemplo, para el análisis y clasificación de imágenes

FIGURA 4.6

RED NEURONAL CON DOS CAPAS Y M NEURONAS EN LA PRIMERA Y DOS EN LA SEGUNDA



Fuente: Elaboración propia.

se utilizan mucho las redes de convolución y para datos temporales, o variables dinámicas, las redes recurrentes. Más recientemente, para el procesamiento de lenguaje natural se han introducido las redes de transformación, o *Transformers*, que están teniendo muchas otras aplicaciones en la actualidad y que se presentarán más adelante.

## 4.6. LA COMBINACIÓN DE VARIOS MÉTODOS DE PREDICCIÓN Y DATOS

La experiencia ha mostrado que con frecuencia se obtienen mejores predicciones mezclando distintas reglas de predicción y ponderándolas por una medida de su eficacia predictiva. Por ejemplo, podemos mezclar redes neuronales con otras de las herramientas de clasificación que veremos en los capítulos siguientes, o distintos tipos de redes.

También podemos aplicar esta idea para combinar distintos tipos de datos. Esto se aplica en los denominados sistemas de *nowcasting* —predicción inmediata— donde se mezclan datos diarios, mensuales y trimestrales para la identificación de riesgos o la predicción de precios de activos. Por ejemplo, podemos mejorar la predicción de una variable meteorológica diaria utilizando la información de la evolución de otras variables relacionadas con ella en periodos de cinco minutos a lo largo de los días precedentes o la volatilidad diaria de una acción en bolsa utilizando la evolución minuto a minuto de los rendimientos en los días previos.

## 4.7. CONCLUSIONES

Los modelos de regresión son una herramienta fundamental en estadística para modelar la relación entre una variable dependiente (objetivo) y una o más variables indepen-

dientes (predictores). Estos modelos son interpretables porque nos indican el impacto exacto de cada variable en la respuesta. Por ejemplo, si el coeficiente de una variable es 2, sabemos que al aumentar esa variable en una unidad, la respuesta aumentará en 2 unidades, *ceteris paribus*, es decir, si todas las demás variables permanecen constantes. Sin embargo, su eficacia depende de que exista una relación lineal entre las variables. La práctica ha mostrado que estos modelos son adecuados para prever datos agregados utilizando un número de variables explicativas,  $p$ , no muy grande con relación a los datos disponibles,  $n$ , de manera que  $n/p$  sea al menos 10.

Las redes neuronales son algoritmos inspirados biológicamente, diseñados para reconocer patrones complejos. A diferencia de la regresión, su enfoque es predictivo y flexible. Se consideran "aproximadores universales de funciones" y en la práctica han demostrado gran eficacia para aproximar cualquier tipo de función. Aunque cada neurona construye un indicador de la entrada, de forma similar a una regresión, como la salida de la neurona es una transformación no lineal de este indicador, combinando muchas neuronas se pueden capturar relaciones muy complejas entre las variables.

La mayor debilidad actual de las redes neuronales es la dificultad de comprender lo que está ocurriendo dentro del modelo, ya que no disponemos de una ecuación simple que explique la relación entre la entrada y la salida. A cambio, no requiere supuestos sobre los datos, y puede modelar relaciones extremadamente complejas y no lineales. No obstante, se está haciendo mucha investigación sobre redes neuronales interpretables y es previsible que este problema se resuelva con ayuda de métodos estadísticos de diseño de experimentos en el futuro.

La elección entre utilizar para la predicción la regresión o las redes neuronales no debe basarse en la "modernidad" de la técnica, sino en el problema a resolver. La regresión es más adecuada si se quiere explicar qué factores influyen en el resultado. Las redes son más potentes si lo único que importa es la precisión de la predicción, no el proceso de llegar a ella, y disponemos de grandes cantidades de datos o trabajamos con datos no estructurados (imágenes, texto, audio). En la práctica, se recomienda comenzar siempre con un modelo de regresión para establecer una línea base (*baseline*). Solo si la regresión no alcanza la precisión deseada, se justifica el salto a la complejidad de las redes neuronales. También es posible utilizar redes híbridas con una parte lineal y otra no lineal y con ciertas restricciones en su funcionamiento.

## APÉNDICE 4

### Estimación de los parámetros

En este apéndice vamos a resumir cómo se estiman los parámetros de una función de predicción.

En su forma más clásica (regresión lineal), se asume una relación lineal entre las  $p$  variables de entrada ( $x_{1i} \dots x_{pi}$ ) y la variable de salida,  $y_i$ , donde, si tenemos  $n$  datos, el índice  $i$  de las variables varía de 1 a  $n$ . La ecuación del modelo de regresión es:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + u_i,$$

donde los coeficientes  $\beta$ , que son los parámetros del modelo a estimar, ponderan el efecto de cada variable de entrada ( $x_{1i} \dots x_{pi}$ ) y  $u$  es el término de error (ruido), que normalmente se asume que sigue una distribución normal.

Los coeficientes  $\beta_i$  se estiman a partir de los datos minimizando el error promedio de predicción al cuadrado para el conjunto de datos.

$$EPP(\beta_0, \dots, \beta_p) = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_p x_{pi})^2$$

Llamaremos  $\hat{\beta}_i$  a los valores estimados que minimizan el error promedio,  $EPP$ . La predicción de la respuesta para unos valores dados de las variables explicativas ( $x_{1i} \dots x_{pi}$ ) se calcula con:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_p x_{pi}.$$

La eficacia de estos modelos depende de que exista una relación lineal entre las variables u otra forma funcional simple. La práctica ha mostrado que estos modelos son adecuados para prever datos agregados utilizando un número de variables explicativas,  $p$ , no muy grande con relación a los datos disponibles,  $n$ , de manera que el cociente  $n/p$  sea al menos 10.

Las redes neuronales profundas son modelos paramétricos compuestos por muchas capas de transformaciones no lineales. Cada capa contiene parámetros ajustables similares a los de una regresión y el objetivo de la estimación, o entrenamiento de la red, es estimar estos parámetros a partir de datos para que la red produzca predicciones lo más correctas posible de acuerdo con el  $EPP$ . Llamaremos  $\beta$  al conjunto de todos los parámetros, que pueden ser un número muy grande, de millones de parámetros distintos. Para estimarlos, los métodos analíticos clásicos no son viables y en su lugar se utilizan métodos iterativos de optimización.

Si queremos encontrar el valor  $b$  que minimiza una función  $f(b)$  de una única variable, podemos comenzar con un valor inicial cualquiera,  $b_0$ , y ver si al aumentarlo un poco la función aumenta o disminuye. Esto es equivalente a calcular la derivada de la función que se aproxima por  $(f(b_0+a)-f(b_0))/a$ . Si la función disminuye cuando  $a > 0$  nos movemos al punto  $b_1 = b_0 + a$  y comenzamos de nuevo.

La misma idea se aplica para encontrar el mínimo de una función de muchas variables,  $EPP(\beta)$ . Ahora calcularemos el gradiente de la función, que equivale a calcular la derivada en cada una de las coordenadas o direcciones de la función. Después podemos aplicar un procedimiento similar como sigue: Se parte de unos valores iniciales de los parámetros (que pueden fijarse al azar), se calcula el valor de la función a minimizar,  $EPP(\beta)$  con estos valores y el gradiente de la función, que nos indica cómo cambia la función en cada dirección o coordenada si se modifican un poco. Con ello se obtiene cómo deben modificarse los parámetros a un nuevo valor y se repite este proceso hasta que la función ya no disminuya más.

Podemos imaginar que el problema general que queremos resolver es similar a descender de una montaña con una niebla profunda que no nos permite ver el camino. En este caso existe una función de dos variables,  $y = f(x_1, x_2)$  que determina la altura y en la que estamos en función de las coordenadas del punto en la montaña,  $x_1$  y  $x_2$ . A partir de un punto inicial,  $x_1(0)$  y  $x_2(0)$  podemos explorar el terreno dando un paso hacia delante y viendo si descendemos o no, volvemos a nuestra posición inicial y damos el paso hacia atrás, comprobando cómo cambia el terreno, luego a la derecha y finalmente a la izquierda. Con estos cuatro pasos podemos decidir la dirección de movimiento donde el descenso es mayor, movernos unos pocos pasos en esa dirección y comenzar de nuevo. Lo que estamos haciendo con estos pasos es calcular el gradiente, o dirección de cambio, en el punto en que estamos y movernos en la dirección de mínimo gradiente.

Si llamamos  $\beta_0$  al valor de los parámetros en el punto inicial, el método del gradiente calcula una nueva posición,  $\beta_1$ , mediante:

$$\beta_1 = \beta_0 - \alpha \nabla f(\beta_0)$$

Donde  $\alpha$  es la longitud de paso o tasa de aprendizaje y  $\nabla f(\beta_0)$  el gradiente de la función que queremos minimizar en el punto de partida.

El gradiente de la función  $EPP(\beta)$  se calcula mediante un algoritmo llamado de retropropagación (*backpropagation*), que aplica la regla de la cadena para calcular derivadas de funciones para propagar los cambios en toda la red desde la salida hacia las capas anteriores.

Este método tiene dos inconvenientes. El primero es que puede quedar atrapado en un mínimo local. Al descender de la montaña, podemos llegar a un pequeño valle donde la pendiente sube en cualquier dirección. El segundo es que, si el conjunto de datos es muy grande (millones de ejemplos), como la función  $EPP(\beta)$  mide la suma de los errores de predicción para todos los datos y hay que calcular el gradiente para cada uno de ellos y sumarlos, utilizar todos los datos en cada iteración es extremadamente costoso.

En su lugar se utiliza lo que se denomina el gradiente estocástico, que aproxima el gradiente usando solo una pequeña parte de los datos, que se denomina *batch* o lote. Con este método, en lugar de calcular el gradiente con todos los datos, se toma al azar un pequeño subconjunto de datos, el lote o *batch*, y se itera con estos datos calculando el gradiente con ellos hasta que no sea posible mejorar. A continuación se toma otro *batch*, muestreando al azar sin reemplazamiento, y comenzando en el último valor de los parámetros del *batch* anterior, se calcula el gradiente con los datos de este *batch* y se itera hasta que no se pueda mejorar. Este proceso se repite con todos los datos de entrenamiento, de manera que, si tenemos  $N$  datos y *batches* de tamaño  $B$  se harán  $N/B$  iteraciones hasta que todos los datos han sido procesados.

Cuando se han procesado todos los datos, se ha cumplido una época o *epoch* y típicamente este proceso se repite varias épocas, hasta que el mínimo de la función se hace estable.

Este método tiene dos ventajas principales. La primera es que, al llegar a un mínimo local con un *batch*, al tomar otro y cambiar la dirección del gradiente, es más fácil salir de ese punto y llegar al mínimo global. Además, al iniciar el algoritmo en distintos puntos en distintas épocas, es mucho más fácil encontrar caminos alternativos para llegar al óptimo de la función. La segunda es que reduce enormemente el coste computacional y permite estimar los parámetros con conjuntos de datos muy grandes.

Un algoritmo del gradiente muy utilizado es el Adam, que introduce una especie de inercia (*Momentum*) en el movimiento de los parámetros, acumulando gradientes pasados. Esto ayuda a suavizar oscilaciones en el proceso de optimización. También ajusta automáticamente la tasa de aprendizaje o longitud de paso para cada parámetro según la magnitud de gradientes recientes.



## CAPÍTULO V

### Clasificación y discriminación

#### 5.1. LA CLASIFICACIÓN ESTADÍSTICA

Uno de los problemas más importantes —y más antiguos— de la inteligencia artificial consiste en clasificar objetos para reconocerlos. Dado un objeto cualquiera, el objetivo es decidir a cuál de varias categorías previamente definidas pertenece. Estas categorías pueden ser tan simples como “sí” o “no”, o tan complejas como miles de posibles significados, intenciones o identidades distintas. Este tipo de problema se conoce como *clasificación supervisada*, y su nombre refleja una idea clave: el sistema aprende a clasificar porque previamente ha sido “supervisado” mediante ejemplos correctos. Es decir, alguien —una persona o una organización— ha decidido de antemano cuáles son las clases posibles y ha proporcionado ejemplos de objetos correctamente clasificados.

Aunque pueda parecer una tarea técnica, la clasificación es una actividad profundamente humana para reconocer objetos y experiencias. Desde pequeños aprendemos a clasificar el mundo: distinguimos sonidos, rostros, emociones, peligros, oportunidades. La inteligencia artificial intenta reproducir este proceso, pero utilizando datos, modelos matemáticos y algoritmos.

La clasificación está hoy presente en multitud de aplicaciones cotidianas. Cuando dictamos un mensaje a nuestro teléfono, el sistema debe clasificar una señal sonora continua en palabras discretas. Cuando una aplicación traduce un texto, clasifica palabras y expresiones de una lengua en equivalentes de otra. Los anuncios personalizados que vemos en internet son consecuencia directa de nuestra clasificación como usuarios con determinados intereses y los sistemas de seguridad reconocen rostros clasificando imágenes.

En todos estos casos, el problema esencial es el mismo: dado un objeto descrito por un conjunto de características, decidir a qué clase pertenece.

## 5.2. DISCRIMINACIÓN LINEAL

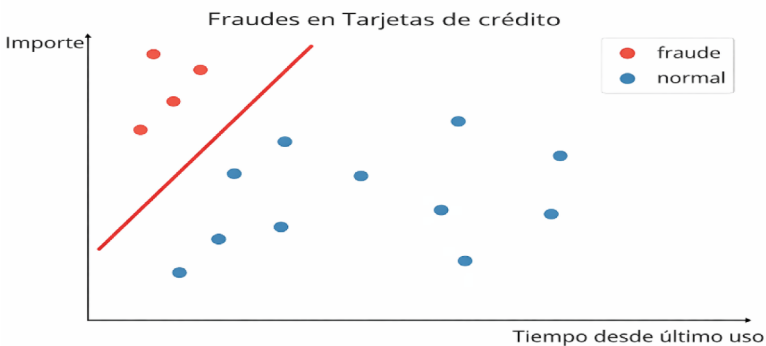
La primera herramienta diseñada para este objetivo es debida a Ronald Fisher, uno de los creadores de la Estadística, que a mediados del siglo XX propuso un método que denominó de *discriminación lineal*. Supongamos que tenemos un conjunto de objetos, cada uno de ellos caracterizado por un conjunto de variables numéricas, y que sabemos también los objetos que pertenecen a la clase A y cuáles a la B. Queremos ahora clasificar otro objeto en una de las dos clases disponiendo de los valores que toman en él las variables que le describen. Supongamos, para simplificar, que disponemos de dos variables para describir cada objeto. Entonces, podemos representar cada objeto en un plano por las coordenadas de estas dos variables y una regla de clasificación equivale a encontrar una línea en el plano que nos separe, tan bien como sea posible, las dos clases. Esta situación se ilustra en la [figura 5.1](#).

Fisher propuso encontrar la línea de separación de las clases buscando la que minimiza los errores de clasificación de los datos disponibles, que es la función lineal discriminante. Esta idea se extiende fácilmente para el caso de muchas variables de clasificación y más de dos grupos de objetos.

Vamos a ilustrar esta idea con un ejemplo. Imaginemos un banco que quiere detectar operaciones fraudulentas con tarjeta de crédito. Para simplificar, consideremos solo dos variables asociadas a cada operación: la cantidad gastada y el tiempo transcurrido desde la operación anterior. Cada operación puede representarse como un punto en un plano: una coordenada corresponde al importe y la otra al tiempo. Disponemos de datos históricos y sabemos qué operaciones fueron legítimas y cuáles fraudulentas. La [figura 5.1](#) representa estos datos y el problema que se quiere resolver es encontrar la línea roja del gráfico que es capaz de separar ambas clases.

FIGURA 5.1

### SEPARACIÓN LINEAL CON DOS VARIABLES



Fuente: Elaboración propia.

Visualmente, observamos que las operaciones fraudulentas tienden a concentrarse en ciertas zonas del gráfico, mientras que las legítimas se agrupan en otras. El problema consiste en encontrar una línea que separe ambos grupos de la mejor manera posible. Una vez encontrada, cualquier nueva operación puede clasificarse simplemente observando en qué lado de la línea cae. Bajo ciertas hipótesis sobre los datos, esta separación puede encontrarse por el método propuesto por Fisher.

Este método funciona bien cuando los datos cumplen ciertas hipótesis y cuando las clases pueden separarse razonablemente mediante una frontera lineal. Aunque hoy existen métodos mucho más sofisticados, esta idea —dibujar una frontera que separe clases— sigue siendo fundamental.

Los nuevos datos han contribuido a desarrollar métodos más potentes de clasificación que ya utilizamos: un ordenador es capaz de entender un texto y traducirlo a otra lengua, las máquinas contestan nuestros mensajes de voz clasificando los sonidos, en nuestro teléfono o en la web aparecen anuncios personalizados, fruto de nuestra clasificación como usuario de ciertos productos, los sistemas de seguridad reconocen nuestros rostros. Estos problemas son más complejos porque existen muchas clases posibles y existen distintas formas de abordarlos, que comentaremos a continuación.

Los problemas anteriores se denominan de Clasificación supervisada o Discriminación, porque las clases posibles donde clasificar se deciden *a priori*. En el capítulo siguiente hablaremos de problemas de clasificación no supervisada o métodos *cluster*, o de construcción de clases, donde se trata de identificar grupos homogéneos sin definir *a priori* cuántos grupos existen ni su composición.

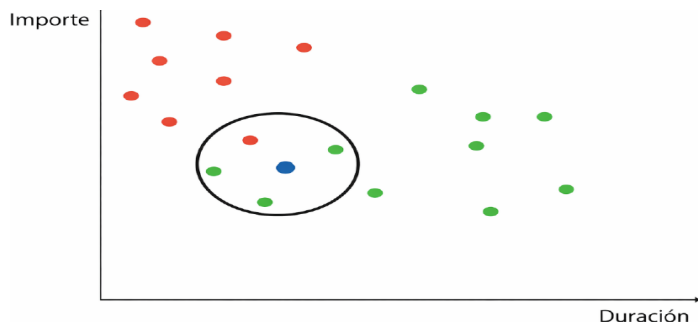
### 5.3. CLASIFICACIÓN POR VECINOS MÁS PRÓXIMOS

El método de los  $k$  vecinos más próximos se basa en una idea intuitiva: tendemos a juzgar lo desconocido por comparación con lo parecido. Si algo se parece mucho a cosas que ya conocemos, asumimos que pertenece al mismo grupo. Este método es muy utilizado, porque puede aplicarse a todo tipo de datos.

Su funcionamiento se presenta en la **figura 5.2** y lo ilustraremos con un ejemplo: supongamos que se desea clasificar una hipoteca como potencialmente problemática o como viable. Disponemos de dos variables para realizar la clasificación: el importe de la hipoteca, que se representa en el eje vertical de la **figura 5.2**, y su duración, que se indica en el eje horizontal. Los puntos rojos de la figura indican hipotecas problemáticas, y los verdes, hipotecas viables. El punto que clasificar es el azul y para ello tomamos un grupo de vecinos próximos al punto a clasificar, por ejemplo, los cuatro vecinos más próximos. Observamos que, de entre ellos, la mayoría de los puntos corresponden a hipotecas viables, ya que de los cuatro, tres son verdes y uno rojo. En consecuencia,

FIGURA 5.2

CLASIFICACIÓN POR VECINOS MÁS PRÓXIMOS



Fuente: Elaboración propia.

como la mayoría de las hipotecas similares a la considerada han sido viables, clasificaremos la nueva hipoteca, el punto azul, como una hipoteca viable.

Naturalmente, en las aplicaciones reales las variables a considerar pueden ser muchas más, lo que supondría que el punto estaría en un espacio de muchas dimensiones representadas en nuevos ejes, pero el procedimiento a aplicar sería el mismo.

Observemos que este método no construye reglas explícitas ni fronteras fijas: la decisión depende del contexto local.

#### 5.4. CLASIFICACIÓN POR MÁQUINAS DE VECTOR SOPORTE

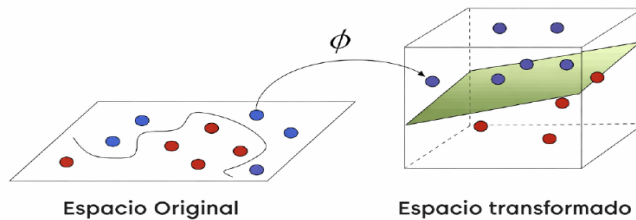
El método de *máquinas de vector soporte* generaliza el método de Fisher y busca una frontera de máxima separación entre los grupos, pero se diferencia del de Fisher en que no necesita ninguna hipótesis sobre los datos. La condición que se busca es que la frontera entre los puntos de ambos conjuntos separe lo más posible los puntos próximos a ella. La diferencia entre el enfoque de Fisher y este método es que en el primero la frontera se calcula con todos los datos observados, y tiene, por tanto, un carácter global, ya que todos los datos observados intervienen en su cálculo. En las máquinas de vector soporte, solo se tienen en cuenta los puntos que están próximos a la frontera de separación y son ellos los que definen la frontera y se utilizan en su cálculo.

Cuando los datos no pueden separarse con una línea recta, este método recurre a una idea ingeniosa: introducir más variables transformando las originales y buscar la separación lineal en este nuevo espacio. Las nuevas variables resultan de aplicar transformaciones no lineales a las originales (por ejemplo, elevándolas al cuadrado o al cubo) y

se intenta obtener una separación lineal en un espacio ampliado que incluye todas las nuevas variables, es decir, tanto las variables originales como sus transformaciones. Esta situación se ilustra en la [figura 5.3](#).

FIGURA 5.3

CLASIFICACIÓN NO LINEAL EN UN PLANO QUE SE HACE LINEAL AL AUMENTAR EL NÚMERO DE VARIABLES A UTILIZAR



Fuente: Elaboración propia.

Aunque este espacio no sea directamente visible, por su alta dimensión, es frecuente poder encontrarlo y el resultado práctico es un clasificador muy potente, capaz de manejar relaciones complejas sin requerir suposiciones restrictivas sobre los datos.

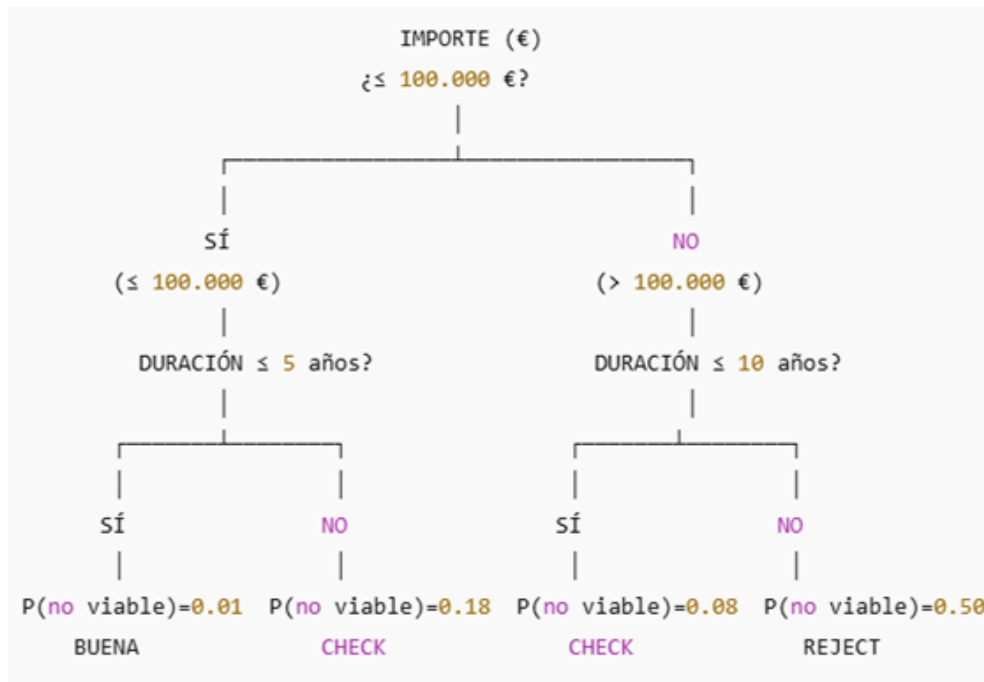
## 5.5. ÁRBOLES DE CLASIFICACIÓN

Un método muy flexible para clasificar son los *árboles de clasificación*, que se construyen dividiendo los datos por etapas de manera que, en cada una de ellas, se reduzca lo más posible el error de clasificación. Volviendo al ejemplo de clasificar hipotecas en función de su importe y duración, la [figura 5.4](#) muestra un ejemplo de un árbol de clasificación muy simple para este problema. Se comienza buscando la variable que separa más las hipotecas viables de las problemáticas y suponemos que en este caso es el importe y que, en los datos disponibles, se ha comprobado que la mejor separación se obtiene con cantidades menores o mayores de 100.000 euros. Si consideramos a continuación hipotecas que superan este límite, la siguiente mejor variable para clasificarlas es la duración del préstamo y suponemos que se obtiene que la mejor separación ocurre con plazos menores de 10 años. En el caso de hipotecas con importe debajo del límite de 100.000 euros, la mejor separación se obtiene con duraciones menores de 5 años.

El resultado de esta regla es el árbol de clasificación de la [figura 5.4](#) donde debajo de cada rama final se indica la probabilidad de que, en función de las variables de la

FIGURA 5.4

UN ÁRBOL DE CLASIFICACIÓN DE HIPOTECAS EN FUNCIÓN DE LAS VARIABLES T (DURACIÓN) E I (IMPORTE)



Fuente: Elaboración propia.

hipoteca, esta sea no viable. Esta probabilidad se calcula viendo cuántas hipotecas quedan en cada uno de los cuatro bloques finales y la proporción de hipotecas no viables en cada uno de ellos. Vemos que solo un 1 % de las hipotecas menores de 100.000 euros y menores de 5 años han dado problemas, mientras que esta proporción sube al 50 % para las hipotecas mayores de 100.000 euros y con duración mayor que diez años.

El resultado de este análisis es la siguiente regla de decisión: si, dada la cantidad y duración de la hipoteca, la probabilidad de no viabilidad es de 0.01, calificarla de “Buena” y aprobarla sin dilación. Si esta probabilidad está entre 0.08 y 0.18, investigar más sobre ella, indicado en la figura con “Check”, y si es 0,5 rechazarla, indicado con “Reject”.

Al igual que en los métodos anteriores, este método se aplica utilizando todas las variables disponibles, dando lugar a árboles muy complejos y con muchas ramas.

## 5.6. CLASIFICACIÓN COMO PREDICCIÓN

Los problemas de clasificación pueden verse como casos particulares de los problemas de predicción que analizamos en el capítulo anterior. En la clasificación, la variable respuesta, o a predecir, es la clase a la que pertenece el objeto, y como en los problemas de predicción, se utiliza un conjunto de variables explicativas. Si existen  $k$  clases posibles, la variable respuesta puede definirse como el número de grupo,  $1, 2, \dots, k$  y los métodos de predicción se aplican a esta variable respuesta para relacionarla con las explicativas de manera similar a como vimos en el capítulo anterior. La función que se optimiza para escoger los parámetros de la regla de clasificación se adapta a este problema para que se minimicen los errores de clasificación. El apéndice indica algunos de los cambios en las redes neuronales que se han presentado en el capítulo anterior para adaptarlas a problemas de clasificación.

Cuando el objeto a clasificar tiene características específicas, como una imagen o un sonido, las variables explicativas deben escogerse cuidadosamente para que sean informativas respecto al texto a clasificar. Existen redes neuronales diseñadas para, en primer lugar, extraer las propiedades básicas del objeto, y luego utilizar estas propiedades como variables para la clasificación. Las redes de convolución, que veremos en el capítulo IX dedicado al análisis de imágenes, son un buen ejemplo de estos métodos.

### APÉNDICE 5

#### Redes neuronales para clasificación

La clasificación supervisada con redes neuronales profundas parte de un conjunto de datos que incluyen una variable respuesta conocida que es un vector con  $k$  componentes, donde  $k$  es el número de clases posibles,  $y_i = (0, \dots, 1, \dots, 0)$ , donde solo uno de los componentes toma el valor uno en la posición de la clase a la que pertenece, que es la etiqueta de clase y un conjunto de  $p$  variables de entrada  $(x_{1i} \dots x_{pi})$  y donde, si tenemos  $n$  datos, el índice  $i$  de las variables varía de 1 a  $n$ .

La red define una función  $y_i = f(\beta, X)$  entre las variables de entrada, los parámetros a estimar y la función de salida que nos indica la clase. La última capa de la red neuronal tiene que proporcionarnos las probabilidades de pertenecer a cada clase y la clasificación consiste en asignar el objeto a la clase más probable.

En clasificación multiclase, la última capa suele incluir una función llamada softmax, que transforma los valores de salida de la red,  $z_j$  para  $j=1, \dots, k$ , en probabilidades. Esta función es,

$$p_j = \frac{e^{z_j}}{\sum_{j=1}^k e^{z_j}}$$

Al elevar al exponente, convertimos los valores en positivos y, al dividir por su suma, obtenemos números que suman uno y podemos interpretar como probabilidades.

En lugar de minimizar el error de predicción, que tiene poco sentido para variables respuesta cualitativas, suele utilizarse como objetivo minimizar el error de clasificación para lo que se utiliza el concepto de entropía. Consideremos un dato  $i$  que pertenece a la clase  $j$ , y llamemos  $p_{i,j}$  a la probabilidad que el modelo asigna de pertenencia de ese dato a la clase  $j$ , donde  $j=1, \dots, k$ , y  $\sum_{j=1}^k p_{i,j} = 1$ . La entropía de este modelo de clasificación para la observación  $i$  es:

$$L_i(\boldsymbol{\beta}) = - \sum_{j=1}^k p_{i,j} \text{Log}(p_{i,j}).$$

Como las probabilidades no pueden ser mayores que la unidad, sus logaritmos serán en general negativos y la entropía en situaciones de incertidumbre será positiva. Tomará el valor cero en el caso de certeza en la clasificación, es decir, cuando una de las probabilidades es uno y el resto cero, en cuyo caso la entropía para ese dato tomará su valor mínimo, cero. Observemos que las clases que tienen probabilidad cero no influyen en la entropía ya que el producto  $0\log(0)=0$ . Puede demostrarse que el valor máximo de la entropía ocurre cuando todas las probabilidades son iguales, es decir, iguales a  $1/k$ , que corresponde al caso de máxima incertidumbre en la clasificación y donde la entropía toma el valor  $\log(k)$ .

La función a minimizar será la entropía total en la clasificación de todos los  $n$  datos, que es la suma de las entropías de cada uno de ellos, y se calcula como:

$$L(\boldsymbol{\beta}) = - \sum_{i=1}^n \sum_{j=1}^k p_{i,j} \text{Log}(p_{i,j}).$$

Los parámetros de la función de predicción  $f(\boldsymbol{\beta}, X)$  se estiman mediante descenso por gradiente estocástico (SGD) o variantes como Adam.

# CAPÍTULO VI

## Construcción de grupos o *clusters*

### 6.1. INTRODUCCIÓN

Un problema de clasificación no supervisada se presenta cuando se dispone de  $n$  elementos (mediciones, textos, sonidos, imágenes, objetos, sistemas, etc.) caracterizados por  $p$  variables o atributos, pero no disponemos de etiquetas o variables de clasificación. Se desea encontrar si los elementos son homogéneos respecto a las variables, es decir, si los elementos no difieren mucho en los valores de las variables que los definen o, por el contrario, existen grupos entre ellos, de manera que los elementos de un mismo grupo tienen valores más similares de las variables que los describen y distintas de los otros grupos.

Por ejemplo, podemos disponer de un conjunto de variables económicas y sociales asociadas a empresas que queremos clasificar por su potencial de crecimiento futuro. No sabemos *a priori* cuántos grupos distintos existen y queremos utilizar los datos para construir grupos de empresas similares respecto a su futura rentabilidad. También podemos tener datos de la evolución de un conjunto de acciones a lo largo del tiempo y queremos clasificarlas en grupos por su evolución histórica. Este problema se conoce como clasificación no supervisada, o análisis de grupos, conglomerados o *clusters*, porque no sabemos cuántas clases existen ni sus características.

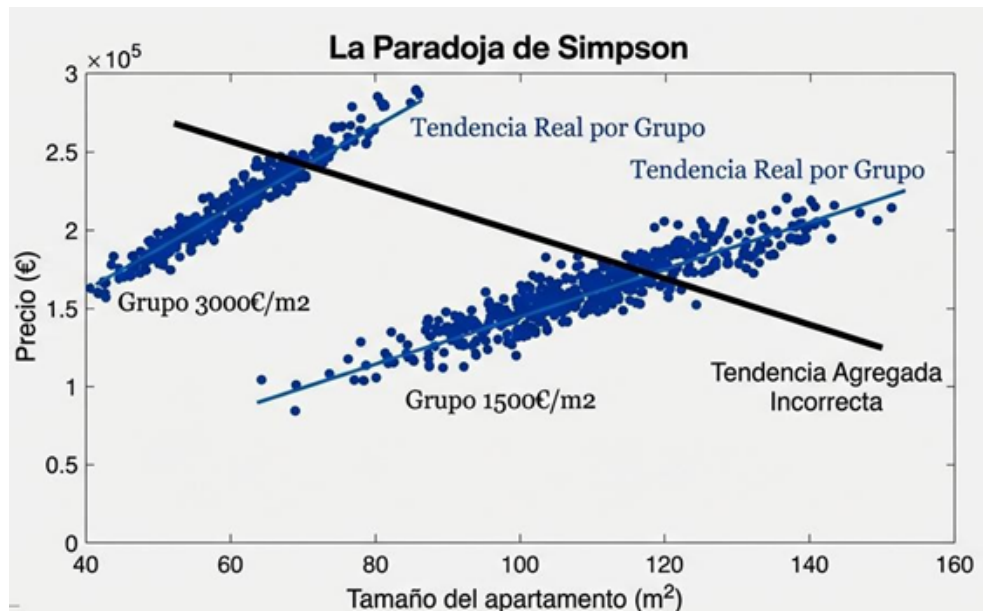
Encontrar si existen grupos es muy importante porque, si los ignoramos, los resultados del análisis pueden ser erróneos, como veremos en la sección siguiente.

### 6.2. LA IMPORTANCIA DE SEGMENTAR LOS DATOS

Consideremos la [figura 6.1](#), que corresponde a una muestra de datos de viviendas de una zona de Madrid hace algunos años. En el eje vertical se representa el precio de las viviendas y en el horizontal su tamaño en metros cuadrados. La figura muestra que existen claramente dos grupos de viviendas: unas más pequeñas en promedio y más

FIGURA 6.1

NECESIDAD DE SEGMENTAR LOS DATOS CUANDO HAY GRUPOS PARA EVITAR LA PARADOJA DE SIMPSON



Fuente: Elaboración propia.

caras, que corresponden a viviendas de construcción reciente y mayor calidad, y otras de mayor tamaño pero más antiguas y con necesidad de rehabilitación y, en consecuencia, más baratas.

Si se desconoce esta estructura y se utiliza la muestra completa para analizar la relación entre precio y tamaño mediante una regresión lineal, llegaremos a la conclusión de que, como indica la recta de la [figura 6.1](#), en promedio, cuanto mayor sea la vivienda, menor es su coste. Este resultado, que es obviamente falso al ser contradictorio con el comportamiento esperado del mercado, se produce al ignorar la existencia de dos grupos muy heterogéneos en los datos, con características muy distintas.

Este problema ha sido muy estudiado en estadística y se conoce como la paradoja de Simpson. Ocurre cuando una relación que se observa en varios grupos por separado se invierte o desaparece al combinar todos los datos en un solo conjunto. La paradoja aparece porque una variable adicional (variable de confusión, la edad de la vivienda en la [figura 6.1](#)) influye tanto en la variable explicativa como en la variable de interés, y al mezclar los datos oculta o distorsiona la relación real. En resumen, trabajar con datos agregados cuando existen grupos puede llevar a conclusiones erróneas.

Como segundo ejemplo, consideremos las notas de estudiantes en dos asignaturas: Historia y Lengua, y comparemos los datos de grupos de chicos y chicas. En Historia las chicas obtienen una nota media de 7,3, más alta que la de los chicos, que es 7,0, y lo mismo ocurre en Lengua, donde la nota media de las chicas es 8,0, más alta que la de los chicos, que es 7,8. Sin embargo, los chicos cursan más Lengua en una proporción del 70 % (donde las notas son más altas), mientras que las chicas cursan más Historia en la misma proporción del 70 % (donde las notas son más bajas). La nota promedio de los chicos será 7,56, más alta que la de las chicas, que es 7,5, y podemos erróneamente obtener conclusiones equivocadas respecto al rendimiento de ambos grupos.

Este ejemplo ilustra cómo trabajar con datos agregados sin identificar previamente los grupos puede llevar a conclusiones engañosas. Para evitar este problema, es fundamental detectar su existencia, para poder realizar comparaciones dentro de conjuntos homogéneos.

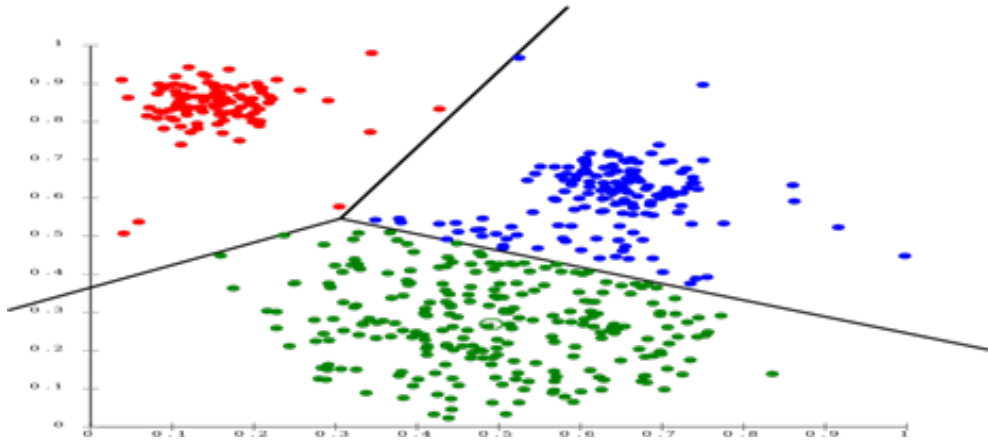
### 6.3. MÉTODOS DE PARTICIÓN

El primer enfoque para construir grupos es el denominado método de partición, donde comenzamos con un único conjunto con todos los datos que vamos dividiendo sucesivamente en un número prefijado de grupos. Suponemos que cada elemento a clasificar está definido por un conjunto de características o variables. Por ejemplo, queremos clasificar activos en función de variables que los definen, como rentabilidad, volatilidad, etc. Comenzamos estableciendo una medida de distancia entre activos, en función de las variables utilizadas; por ejemplo, hacemos una media ponderada de las variables de cada activo y calculamos las diferencias y, segundo, fijamos el número de grupos a construir. A continuación, se utiliza un algoritmo para iterativamente dividir los activos en ese número de grupos de manera que sean tan homogéneos como sea posible. Utilizamos la palabra algoritmo como el conjunto de pasos a realizar para obtener un objetivo. Por ejemplo, una receta de cocina muy detallada que produce siempre el mismo resultado es un algoritmo. La [figura 6.2](#) ilustra la aplicación de un algoritmo de partición en unos datos donde se concluye que hay tres grupos.

El algoritmo de partición más utilizado es el llamado de *k* medias, donde *k* indica el número de grupos. Se inicia seleccionando al azar *k* elementos como centros de los grupos y se asigna cada uno de los restantes elementos al grupo de cuyo centro está más próximo. Cada vez que se introduce un elemento en un grupo, se recalcula el nuevo centro del grupo y así se continúa viendo si es posible mejorar la clasificación de algún elemento, moviéndole a otro grupo de cuyo centro esté más próximo, hasta que esto no sea posible y tengamos todos los elementos clasificados en *k* grupos de manera que cada elemento esté en el grupo de cuyo centro esté más próximo. El proceso se repite desde distintos valores iniciales hasta que los grupos obtenidos finalmente sean siempre los mismos.

FIGURA 6.2

IDENTIFICACIÓN DE TRES GRUPOS CON UN ALGORITMO DE PARTICIÓN



Fuente: Elaboración propia.

La selección del número de grupos se efectúa comenzando con dos y aumentando su número hasta que no se obtenga una mejora importante en la medida de homogeneidad de la clasificación que se ha utilizado.

## 6.4. MÉTODOS JERÁRQUICOS

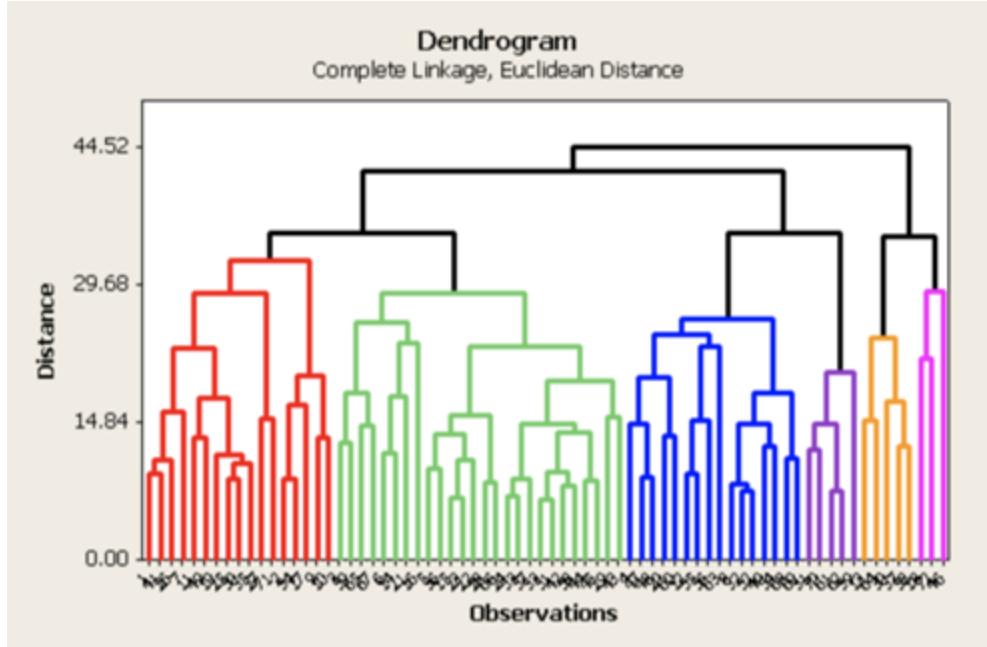
El segundo tipo de métodos para encontrar grupos es llamado de aglomeración, porque se comienza con los elementos individuales que se van uniendo a sus elementos más próximos hasta formar un grupo que engloba a todos. Esto se hace con un *algoritmo jerárquico*.

Supongamos de nuevo los activos financieros. Para ello, calcularíamos las distancias entre ellos utilizando alguna medida de distancias entre las variables y construiríamos una matriz de distancias entre todos los puntos. El algoritmo se inicia uniendo los dos elementos más cercanos y recalculando las distancias de los elementos restantes al nuevo creado. El proceso se repite uniendo de nuevo los más próximos y recalculando distancias, hasta tener un único grupo que englobe a todos los elementos.

El resultado de este proceso puede darse en forma gráfica en un *dendograma* como el que se presenta en la figura. En el eje horizontal, en la parte inferior de la [figura 6.3](#), se representan los elementos a clasificar y en el eje vertical las distancias a las que se unen por el algoritmo. Se observa cómo se van uniendo los elementos más próximos y formando nuevos elementos que se unen a su vez en etapas posteriores. El número

FIGURA 6.3

## IDENTIFICACIÓN DE SEIS GRUPOS CON UN ALGORITMO JERÁRQUICO



Fuente: Elaboración propia.

de grupos se selecciona estableciendo una distancia máxima de unión entre grupos. En este caso se han seleccionado seis grupos de elementos.

Los métodos jerárquicos tienen la ventaja de no requerir fijar *a priori* el número de grupos y de ofrecer una visión completa de la estructura de los datos, aunque pueden resultar computacionalmente costosos para grandes volúmenes de información.

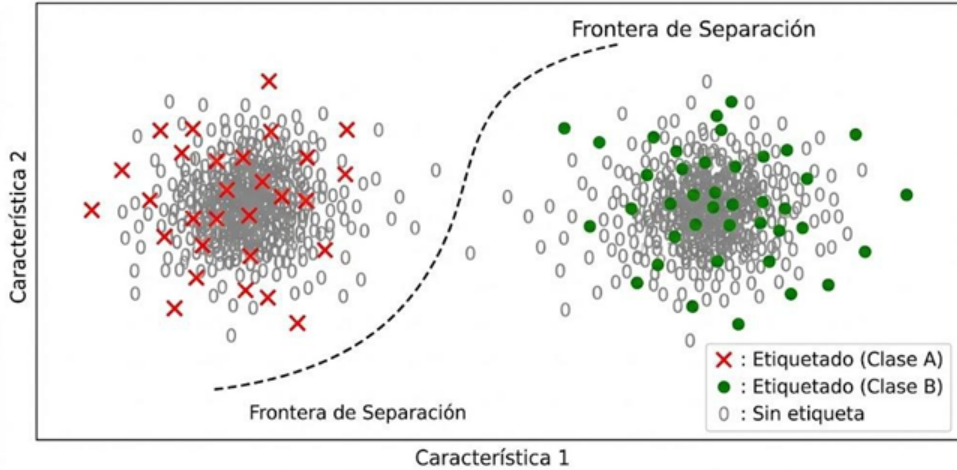
## 6.5. MÉTODOS MIXTOS

Las técnicas de clasificación supervisada y no supervisada están estrechamente relacionadas y, en muchos problemas reales, se combinan para mejorar los resultados. Este enfoque se conoce como aprendizaje semisupervisado o híbrido.

Consideremos la situación ilustrada en la [figura 6.4](#), donde se dispone de algunos datos etiquetados en dos grupos (puntos X en rojo y círculos en verde) y un conjunto mayor de datos sin etiqueta de grupo que están representados por ceros. Aunque los datos sin etiqueta no proporcionan información directa sobre la clase, su presencia ayuda a definir mejor la estructura de los grupos y la frontera de separación entre ellos.

FIGURA 6.4

CLASIFICACIÓN DE DATOS ETIQUETADOS Y SIN ETIQUETA



Fuente: Elaboración propia.

En muchos problemas de clasificación con muchos datos, asignar etiquetas es un trabajo difícil y costoso, y a veces imposible, si no se dispone de los datos. Sin embargo, mezclando unos pocos datos con etiquetas con otros muchos sin ellas, los grupos pueden aparecer más nítidos consiguiendo una clasificación más rápida y efectiva.

## 6.6. LECTURAS ADICIONALES SOBRE CLASIFICACIÓN

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.  
Presenta una visión completa de los métodos de clustering, incluyendo medidas de distancia, algoritmos de partición y enfoques jerárquicos, con numerosos ejemplos aplicados.
- Alpaydin, E. (2020). *Introduction to Machine Learning* (4ª ed.). MIT Press.  
Introducción muy didáctica a los principales algoritmos de Machine Learning. Explica clasificación supervisada y no supervisada de forma clara, con ejemplos y pseudocódigo de algoritmos clásicos.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.  
Aborda los métodos de *clustering* desde una perspectiva probabilística y conecta

la clasificación no supervisada con modelos estadísticos y de aprendizaje automático más generales.

- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley.  
Texto clásico dedicado exclusivamente al análisis de conglomerados, que aborda los fundamentos teóricos, los distintos tipos de algoritmos y los criterios para la selección del número de grupos.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.  
Referencia de aprendizaje profundo (*deep learning*). Aunque es más amplio, dedica capítulos a redes neuronales para clasificación, incluyendo softmax, aprendizaje supervisado y representaciones distribuidas.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.  
Obra de referencia fundamental que desarrolla en profundidad los métodos de clasificación no supervisada, los algoritmos de partición y los métodos jerárquicos, así como su relación con técnicas supervisadas.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.  
Introducción clara y aplicada al análisis de grupos, con especial atención a la robustez de los métodos y a la interpretación de los resultados.

## APÉNDICE 6

### Redes neuronales para clasificación no supervisada

Los datos disponibles para una red neuronal que construya *clústeres* son las variables de entrada, pero no tenemos las variables de salida que serían las clases en las que clasificamos los datos. Un método que se ha utilizado con redes profundas es aprender un espacio latente o factorial, donde los datos están organizados de forma más estructurada. Esto se conoce como *representation learning*. El espacio latente o factorial suele tener menor dimensión que el espacio original y captura factores relevantes de variación.

Para ello se utilizan redes neuronales llamadas *autoencoders*, donde la entrada y la salida tienen la misma dimensión y que tienen dos partes. En la primera, que se llama *Encoder*, entran los datos y produce una salida que es un vector de variables de dimensión menor que la entrada. La segunda parte es el *Decoder*, que toma como entrada la salida del *Encoder* y produce un vector de la misma dimensión que las variables originales. La red se entrena para que el error entre los datos iniciales y los reconstruidos sea mínimo.

Las representaciones latentes aprendidas por este procedimiento, que son las salidas del *Encoder*, pueden utilizarse para *clustering*. Un procedimiento típico es entrenar un *autoencoder*, extraer las representaciones latentes y aplicar un algoritmo como k-means en el espacio de las representaciones latentes o factores.

## CAPÍTULO VII

### El reconocimiento de voz

#### 7.1. INTRODUCCIÓN

Cuando los primeros ordenadores comenzaron a desarrollarse a mediados del siglo XX, la idea de que una máquina pudiera comprender el habla parecía ciencia ficción. Sin embargo, en este siglo, los sistemas de reconocimiento de voz han dejado de ser experimentos de laboratorio para convertirse en una tecnología cotidiana, integrada en teléfonos móviles, sistemas bancarios, automóviles y servicios de atención al cliente.

Los sistemas de reconocimiento de voz han evolucionado de forma especialmente rápida en los últimos quince años, aunque sus raíces se remontan mucho más atrás. Los primeros modelos estaban diseñados para reconocer el habla de una persona concreta, normalmente el propio usuario del sistema. Estos sistemas requerían entrenamiento previo: el usuario debía repetir palabras o frases para que el ordenador aprendiera cómo sonaba su voz.

El entrenamiento consistía en comparar el habla del usuario con grabaciones previamente almacenadas en bases de datos sonoras, producidas por hablantes nativos del idioma. El proceso era lento y frágil: pequeños cambios en la pronunciación, el tono o la velocidad al hablar podían provocar errores. Aun así, estos sistemas sentaron las bases conceptuales de los sistemas actuales, donde podemos interactuar de viva voz con muchos de los sistemas electrónicos que nos rodean, haciendo nuestra vida más cómoda y productiva.

#### 7.2. DIGITALIZAR EL SONIDO

El primer paso para que un ordenador entienda nuestra voz es convertir el sonido en datos digitales. Cuando hablamos, el micrófono capta las vibraciones sonoras y las transforma en señales eléctricas. Estas señales llegan al ordenador, que las convierte

en secuencias de números que el ordenador puede entender: la voz humana, continua y analógica, se transforma en un lenguaje matemático que el ordenador puede procesar. Cada fragmento de audio queda representado por vectores numéricos, que describen sus características acústicas esenciales. Este proceso se representa en la figura 7.1.

**FIGURA 7.1**

**REPRESENTACIÓN DIGITAL DEL SONIDO**



Fuente: Elaboración propia.

Una vez digitalizado el sonido, el sistema tiene que identificar las palabras. Los primeros métodos se basaban en el reconocimiento de fonemas, que son los sonidos básicos del idioma. La mayoría de los idiomas tienen entre 20 y 35 fonemas. Por ejemplo, el español tiene 24 fonemas (5 vocales y 19 consonantes) y es uno de los idiomas latinos más simples: el italiano tiene 34 fonemas y el francés 37. Los idiomas del norte de Europa son más complejos que los derivados del latín: el inglés tiene 44 fonemas y el danés llega hasta 50. Además, existen idiomas en África con 85 fonemas.

Un problema adicional es que en el idioma es frecuente que existan distintos acentos o formas de pronunciación, por lo que el sonido de un fonema varía y distintas personas lo emiten de forma diferente. Por tanto, hay que definir todos los sonidos que corresponden a cada fonema, que puede ser a veces un conjunto muy amplio, y que llamaremos la clase de sonidos del fonema. El objetivo es, por tanto, clasificar las señales digitales en estas clases de fonemas.

Este problema de clasificación se resolvió inicialmente con métodos locales, basados en una amplia base de datos de pronunciaciones del idioma. Estos diccionarios de palabras (*speech database*) se construyen grabando previamente a muchos locutores de un idioma. El sistema clasifica los conjuntos de fonemas dentro de ese diccionario para identificar, o escuchar, las palabras pronunciadas. Estos métodos mejoraban, además, con la experiencia de su uso, porque se puede ir ampliando la base de datos de sonidos al aplicar el clasificador cuando es posible comprobar si la clasificación ha sido correcta. Por esta razón, era frecuente que la máquina pidiese la confirmación al cliente sobre su comprensión para etiquetar los sonidos escuchados en la clase correcta e incorporarlos a la base de datos, mejorando su funcionamiento con la experiencia.

Los sonidos se agrupan con las técnicas de clasificación supervisada o discriminación que hemos visto en el capítulo V. Estas técnicas permitieron construir los primeros reconocedores de voz funcionales, aunque con limitaciones importantes: los fonemas cambian según el contexto, y los errores se acumulaban fácilmente.

La clasificación de fonemas, cuando la base de datos disponible no era muy grande, podía mejorarse añadiendo un sistema de video que analizase los labios del emisor para identificar los fonemas que utiliza. La evidencia demostró que al mezclar ambos métodos, el visual y el auditivo, se obtenía un sistema más potente. Véase Gong (1995), Padmanabhan and Premkumar (2015) y Nassif *et al.* (2019) para una descripción de los procedimientos utilizados antes de los grandes modelos de lenguaje para el reconocimiento de voz.

### 7.3. LOS MODELOS DE LENGUAJE

A partir de 2010 se produjo un cambio de paradigma con la utilización de las redes neuronales profundas (DNN) como sistema de aprendizaje del idioma. Los métodos de clasificación y predicción mediante estas redes comenzaron a sustituir a los modelos acústicos clásicos basados en mezclas estadísticas de distribuciones normales. Las redes neuronales profundas eran capaces de modelar relaciones no lineales complejas en los datos acústicos y ofrecían mejoras significativas en precisión.

Con el aumento de la capacidad de cálculo y la disponibilidad de grandes corpus de audio, el foco pasó de los fonemas a las palabras completas y, posteriormente, a las secuencias de palabras. Los sistemas modernos de reconocimiento de lenguaje se basan no solo en cómo suena una palabra, sino también con qué frecuencia aparece y en qué contextos. Esto se ha conseguido con la arquitectura de redes tipo *Transformer* que estudiaremos en el capítulo siguiente dedicado a los modelos de lenguaje, cuyo desarrollo ha marcado un punto de inflexión también en el reconocimiento de voz. Este sistema puede capturar dependencias a largo plazo de forma eficiente y paralelizable, facilitando su rápido cálculo. En reconocimiento de voz, han dado lugar a arquitecturas híbridas que combinan estructuras para capturar patrones locales del audio con mecanismos de atención, para tener en cuenta el contexto global donde se produce el sonido.

Los modelos de lenguaje permiten estimar la probabilidad de que una palabra aparezca después de otra. Por ejemplo, tras reconocer la palabra “casa”, el sistema asigna mayor probabilidad a palabras como “grande”, “nueva” o “antigua” que a combinaciones poco frecuentes como “lápiz” o “remo”. Este principio es similar al de los sistemas de texto predictivo utilizados en los teléfonos móviles.

Así, cuando el sistema reconoce una palabra, automáticamente espera las más probables a continuación de la escuchada y solo tiene que clasificarla entre las opciones ya

determinadas, en lugar de tener que buscar entre todas las posibles. De este modo, el sistema establece una estructura en forma de árbol donde, después de cada palabra, se busca entre las más probables, lo que mejora los resultados de los métodos de clasificación.

## 7.4. APLICACIONES

Los sistemas de reconocimiento de voz han dado lugar a los asistentes digitales que, además de contestar preguntas de los usuarios, como Alexa de Amazon o Siri de Apple, pueden efectuar acciones de asistencia como poner música, encender las luces o buscar el significado de una palabra en un diccionario. También comienzan a incorporarse en los sistemas de recomendación que describimos en el capítulo X.

Otras aplicaciones en nuestra vida diaria son, por ejemplo, los sistemas de atención automática al cliente, que son capaces de mantener interlocución con personas y facilitar la gestión de trámites. También, cada vez más, en la seguridad bancaria, para la detección del fraude, de comportamientos sospechosos o para el análisis de riesgos. De hecho, la voz se está convirtiendo en una forma segura de identificación personal, tan única como la huella dactilar o el reconocimiento facial.

El reconocimiento de voz continúa evolucionando. Cada interacción aporta nuevos datos, cada conversación mejora los modelos y cada avance en capacidad de cálculo abre nuevas posibilidades. La voz, que durante siglos fue el medio más natural de comunicación humana, se ha convertido también en uno de los puentes más importantes entre las personas y las máquinas.

## 7.5. LECTURAS ADICIONALES

A continuación, indicamos algunas referencias en orden cronológico para el lector interesado en conocer la evolución de los sistemas de reconocimiento de voz.

- Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357–366.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Prentice-Hall, Inc.

- Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. *Proceedings of the 23<sup>rd</sup> International Conference on Machine Learning* (369–376).
- Hinton, G. *et al.* (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29(6), 82–97.
- Hannun, A. *et al.* (2014). Deep Speech: Scaling up end-to-end speech recognition. *arXiv:1412.5567*.

Los sistemas actuales pueden consultarse en:

- Vaswani, A. *et al.* (2017). *Attention is All You Need*. *NeurIPS 2017*, 5998–6008.
- Gulati, A. *et al.* (2020). *Conformer: Convolution-augmented Transformer for speech recognition*. *Interspeech 2020*, 5036–5040.
- Baevski, A. *et al.* (2020). *wav2vec 2.0*. *NeurIPS 2020*, 12449–12460.
- Radford, A. *et al.* (2023). *Robust Speech Recognition via Large-Scale Weak Supervision*. *ICML 2023*, PMLR 202, 28492–28518.
- Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3<sup>rd</sup> ed. draft). Stanford University



# CAPÍTULO VIII

## Generación de lenguaje natural

### 8.1. INTRODUCCIÓN

Una de las aplicaciones de la IA que han despertado más interés en los últimos años son los sistemas de generación de lenguaje natural. Estos sistemas responden a nuestras preguntas de forma articulada, o redactan textos de acuerdo con nuestras indicaciones. Podemos plantearles peticiones por escrito o, también, oralmente, ya que los sistemas de generación van incorporando los métodos de análisis de voz comentados en el video anterior.

En este capítulo vamos a centrarnos en explicar los métodos que utilizan los grandes sistemas de procesamiento de lenguaje natural para la creación de textos. En el apéndice de este capítulo se presenta una autocomparación de las fortalezas y debilidades de tres de los sistemas más utilizados, ChatGPT, Gemini y DeepSeek, realizada por ellos mismos.

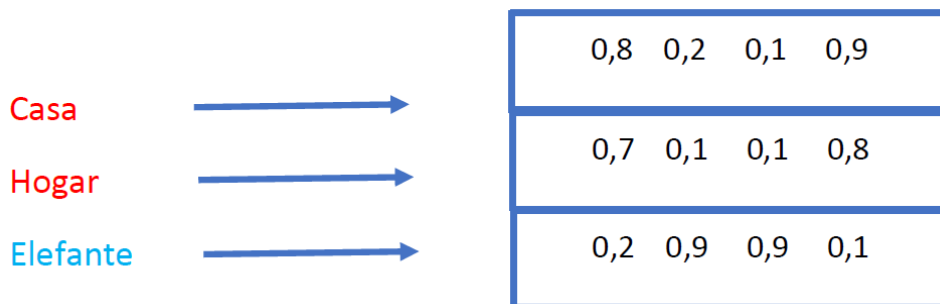
### 8.2. CONVERTIR PALABRAS EN VECTORES DE NÚMEROS

El primer paso para procesar lenguaje es convertir las palabras en secuencias de variables numéricas, que llamaremos vectores, que un ordenador pueda manejar. Este proceso se denomina vectorización de palabras, o *Word embedding* en inglés.

Estos vectores representan la situación de cada palabra con respecto a ciertos factores del lenguaje que se determinan empíricamente analizando mucho textos. Los valores de las coordenadas del vector en una palabra se asocian a factores semánticos del lenguaje. La asignación de estas coordenadas se realiza de manera que palabras que suelen aparecer juntas, porque se utilizan en el mismo contexto, tengan vectores próximos, o secuencias de los números que forman el vector, similares.

Diremos que una secuencia es similar a otra si tiene números similares en las mismas posiciones. Por ejemplo, la [figura 8.1](#) presenta tres secuencias de cuatro números

FIGURA 8.1  
VECTORIZACIÓN DE PALABRAS



Fuente: Elaboración propia.

correspondientes a tres palabras. Se observa que las dos primeras palabras están relacionadas, porque tienen secuencias de números similares, mientras que la tercera es muy distinta de las dos anteriores.

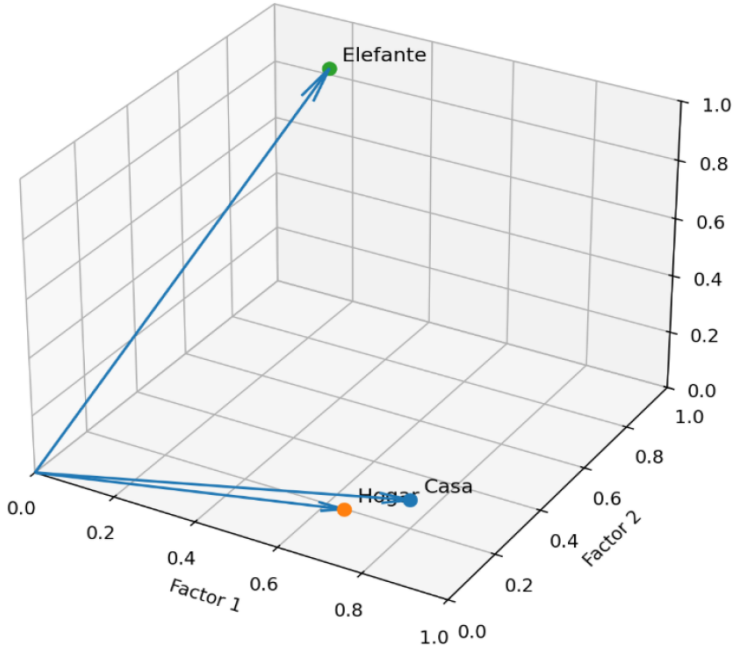
Los cuatro números asociados a estas palabras representan la valoración de la palabra en cuatro factores del lenguaje. Supongamos, por ejemplo, que el primer factor representa sensaciones positivas o neutras de la palabra, y casa y hogar se asocian a sensaciones positivas, con un alto valor para ese factor, y elefante a neutras, con bajo valor del factor. El segundo factor se relaciona con animales, y tendrá un el valor alto si la palabra se relaciona con ese contenido, como vemos en el valor para la palabra elefante. El tercero factor con movimiento, de nuevo alto para elefante y el cuarto con personas, con valores altos en casa y hogar. En la práctica se utilizan cientos o miles de factores distintos para codificar las palabras.

Las palabras asociadas con vectores próximos entre sí suelen aparecer cerca en un texto: se dice que tienen un contexto común. Por ejemplo, las palabras "casa" y "hogar" tienen un contexto común, porque aparecen frecuentemente escritas cerca de un conjunto de palabras similares. La [figura 8.2](#) presenta estos vectores en un espacio de dimensión tres para indicar su estructura de forma intuitiva.

El idioma español tiene unas 100.000 palabras, y necesitamos un gran conjunto de textos y documentos para poder calcular, como primer paso, la frecuencia relativa con la que aparece cada palabra del idioma y, además, la frecuencia relativa con la que cada palabra aparece cerca, es decir en la misma frase o separada por unas pocas palabras, de cada una de las otras palabras del idioma.

La forma inicial de realizar el *word embedding* fue asignar un vector fijo a cada palabra, con independencia del contexto, como los sistemas Word2Vec o GloVe. Para ello se consideraban todas las palabras del idioma, se definía un intervalo de proximidad, por

**FIGURA 8.2**  
**REPRESENTACIÓN VECTORIAL DE PALABRAS**



Fuente: Elaboración propia.

ejemplo, seis palabras, y se aplicaba en cada texto disponible para entrenamiento una ventana de longitud seis y se contaba para cada palabra las frecuencias relativas con las que el resto de las palabras del idioma aparecían próximas a ella. Con esta operación podemos asignar un *embedding* a cada palabra de dimensión igual al número de palabras del idioma y donde cada coordenada representa la proximidad o frecuencia de aparición de la palabra con respecto a todas las demás.

Este procedimiento simple no es eficiente, porque la matriz de relaciones entre las palabras tendrá muchos ceros y valores muy pequeños, ya que muchas palabras muy raramente aparecen con otras. Podemos resumir la información contenida en esta matriz de frecuencias utilizando factores asociados a las palabras, y representarlas por vectores de dimensión mucho menor cuyas coordenadas sean los valores en la palabra de estos factores semánticos o lingüísticos. Los factores se determinan viendo cuantas dimensiones o variables necesitamos para aproximar esta matriz por otra con el mismo número de filas pero con un conjunto de columnas mucho menor. Una manera de resolver este problema es utilizar los componentes principales, que son indicadores capaces de resumir la información contenida en una matriz de relaciones entre variables y considerar estos componentes como factores del lenguaje.

Los métodos actuales calculan los factores orientados a la predicción de nuevas palabras utilizando una gran masa de datos como entrenamiento. Se basan en la idea de los *Transformers*, que es un tipo de red neuronal capaz de encontrar la relación entre palabras y asignarlas un *embedding* variable que depende del contexto donde aparecen, haciendo su representación mucho más eficaz. Además, el *embedding* de las palabras se adapta al objetivo del análisis. Por ejemplo, es distinto si queremos generar lenguaje prediciendo la siguiente palabra de una secuencia dadas las anteriores, que si queremos traducir un texto.

Los modelos generativos actuales van generando las palabras en función de un contexto definido por la pregunta que les hacemos. El *embedding* se hace teniendo en cuenta las probabilidades de cada palabra condicionada a las que forman la pregunta y su longitud, o número de coordenadas del vector, depende del sistema. Por ejemplo, los modelos actuales tienen una dimensión de *embedding* de entre 4.000 y 12.000 dimensiones aproximadamente.

### 8.3. LAS REDES TRANSFORMERS

Para generar texto como respuesta a una pregunta, o para generar un documento sobre una petición, el sistema toma las palabras de entrada para definir el contexto, que consiste en el amplio conjunto de palabras que suelen aparecer cerca de las usadas para la pregunta. Después toma una ventana de recuerdo de cinco o seis palabras y va eligiendo las siguientes palabras, sorteando entre las posibles palabras que existen en la base de datos que han aparecido a continuación de las ya seleccionadas.

Se comienza con una de las palabras utilizadas en la petición. Dada la primera palabra, la segunda se selecciona al azar entre las palabras de contexto que pueden seguirla, sorteando aleatoriamente entre ellas utilizando las probabilidades o frecuencias relativas de la base de datos del idioma para las palabras que constituyen el contexto.

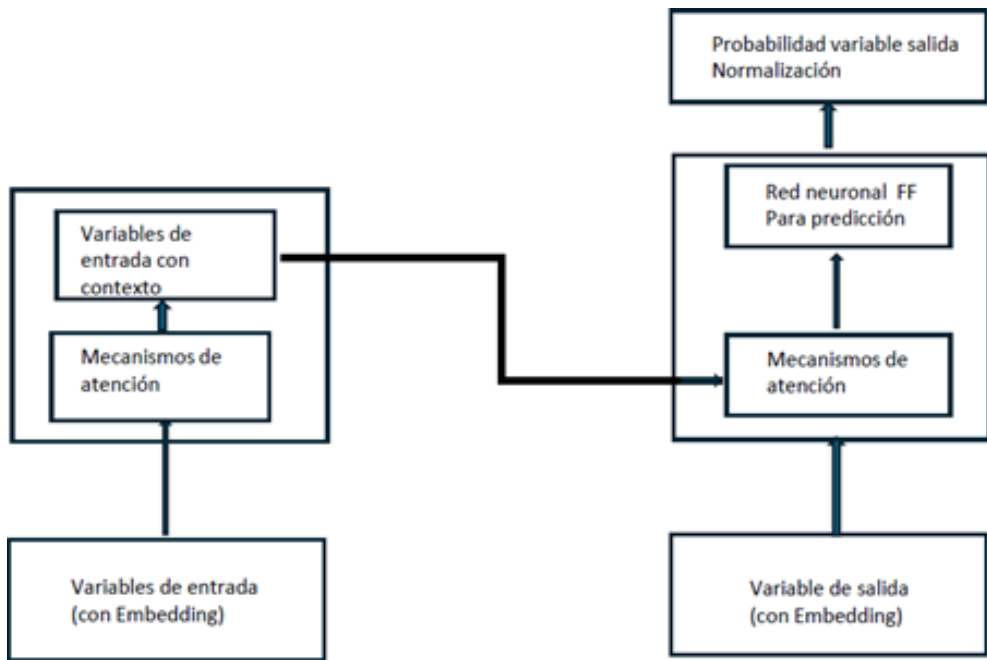
Por ejemplo, si tenemos la palabra "casa" y en la base de textos disponible esta palabra aparece 10 veces y de estas en 3 va seguida de la palabra "pequeña", en 5 de la palabra "grande" y en 2 de la palabra "confortable", el generador simula sacar una bola de una urna con el 30 % de las bolas con la palabra "pequeña", el 50 % con la palabra "grande" y el 20 % con la palabra "confortable". Supongamos que el sorteo produce la palabra "grande".

Entonces tenemos "casa grande" y la tercera palabra se genera utilizando otro sorteo con las frecuencias relativas de las palabras que van después de "casa grande", y así sucesivamente.

Naturalmente, para llevar esto a la práctica con todas las secuencias posibles, necesitamos una gran potencia de cálculo, tiempo de computación y métodos matemáticos para acelerar el proceso, como ocurre con ChatGPT.

Las **figuras 8.1** y **8.2** resumen el funcionamiento de los modelos generativos de lenguaje. Estos gráficos tratan de explicar la lógica general, y no los detalles de cada uno de los sistemas. Las variables de entrada son las palabras de la frase que definen el contexto de generación de las palabras siguientes y a cada una se le asigna un *embedding* inicial general que va a modificarse para adaptarla al contexto. Estas palabras se interpretan con el mecanismo de atención, o red *Transformer*, por su nombre inicial debido a Vaswani *et al.* (2017). Su esquema de funcionamiento se presenta en la **figura 8.2**.

**FIGURA 8.1**  
**ESQUEMA DE FUNCIONAMIENTO SIMPLIFICADO DE UN LLM**

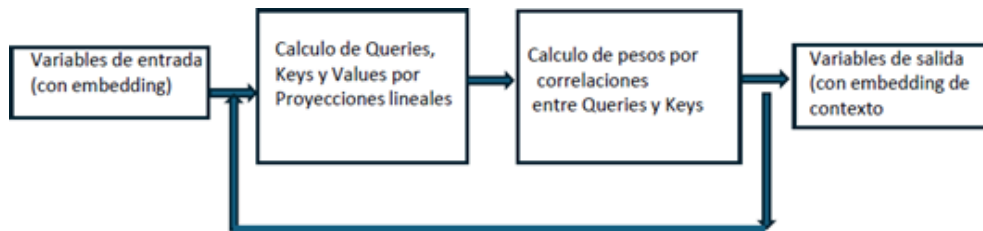


Fuente: Elaboración propia.

El contexto modifica la representación de las palabras en función de las que la acompañan. Para cada palabra se definen tres vectores de variables con dimensión menor que el *embedding* inicial, y que se refieren a distintas partes de este, es decir, a distintos factores semánticos. Estos vectores de variables son el *Value*, o el contenido de la palabra, que va a ser la modificación de la parte del *embedding* de la palabra correspondiente

FIGURA 8.2

EL MECANISMO DE ATENCIÓN DEL *TRANSFORMER*



Fuente: Elaboración propia.

a los factores semánticos seleccionados. El vector de *Queries*, o preguntas, que define la información sobre cada palabra que queremos buscar en las otras palabras del contexto, y el vector de *Keys*, o de respuestas, que indica la información que estas otras palabras pueden aportar a la palabra que estamos considerando. Utilizando las correlaciones entre las *Keys* y las *Queries*, entre las preguntas y las respuestas, se modifican las partes del *embedding* definidas en el *Value*, adaptando su significado al contexto. Esto se repite muchas veces para distintos tipos de *Keys*, *Queries* y *Values*, que afectan a distintas partes del *embedding* inicial, es decir, a distintos factores semánticos, y con ello se obtiene la representación vectorial final de cada palabra adaptada al contexto.

Una vez adaptadas las palabras de entrada, se hace lo mismo con las posibles palabras de salida o de continuación de la entrada, para adaptar su *embedding* al contexto de entrada mediante otro mecanismo de atención que tiene en cuenta la relación entre las palabras de entrada y salida. Después, toda esta información pasa a una red neuronal del tipo ANN *Feed-Forward*, que se utiliza para prever las probabilidades de las siguientes posibles palabras condicionadas a las anteriores de la frase.

Los modelos de lenguaje, en lugar de las palabras, utilizan *tokens* que pueden ser palabras, pero también divisiones o partes de las palabras y símbolos. Por ejemplo, los verbos se convierten generalmente en dos *token*: Comiendo se convierte en los *token* com y iendo. En promedio, las palabras producen 1,3 *token*. De esta manera, el manejo del lenguaje es más eficiente.

## 8.4. EJEMPLO DE UN SISTEMA GENERATIVO

Es interesante ver las respuestas de los sistemas generativos cuando les preguntamos por su funcionamiento. El resumen realizado por todos ellos es similar y, por brevedad, presentamos a continuación la respuesta de uno de los sistemas más populares, que describe su funcionamiento como sigue:

1. Entrenamiento masivo: Fui entrenado con una gran cantidad de textos públicos para aprender el uso del lenguaje, gramática, estilo y el contexto de las palabras. Este entrenamiento me permite generar respuestas coherentes y relevantes a lo que dices.
2. Predicción de palabras: Cuando escribes algo, analizo la secuencia de palabras que me proporcionas y trato de predecir qué respuesta sería adecuada. Lo hago palabra por palabra o bloque de palabras, generando la próxima palabra en función de lo que es más probable que siga a la anterior según patrones aprendidos.
3. Contexto: También considero el contexto de nuestra conversación. Si me haces una pregunta o mencionas algo que requiere un análisis más profundo, utilizo la información anterior para construir una respuesta que tenga sentido y que continúe de manera lógica la conversación.
4. Corrección automática: Mi respuesta está estructurada para ser gramaticalmente correcta y con sentido lógico, ajustando el estilo o el tono de la conversación para que se ajuste a lo que parezca más apropiado en cada momento.

## 8.5. CONCLUSIONES

Los sistemas generativos son sistemas estadísticos que generan las palabras más probables que pueden aparecer a continuación de las anteriores. Por lo tanto, son muy útiles para hacer resúmenes, obtener información estructurada sobre un tema y realizar tareas repetitivas. Sin embargo, no son buscadores de internet, sino predictores de palabras. Aunque buscan y utilizan información en internet, no nos informan de lo encontrado, como los buscadores, sino que elaboran la información para prever la respuesta y, en ese proceso, pueden fácilmente distorsionar el contenido o inventar datos importantes, como nombres y fechas, proporcionando una respuesta que contiene errores y datos equivocados.

Tampoco son programas de cálculo realizados de forma segura y fiable, y si les pedimos operaciones complejas, pueden darnos fácilmente resultados falsos. Por ejemplo, si pedimos multiplicaciones con números muy altos, con frecuencia cometen errores. Es, por tanto, fundamental que el usuario compruebe siempre por sí mismo los resultados obtenidos, sin confiar ciegamente en la respuesta proporcionada por el sistema generativo.

## APÉNDICE 8

En este apéndice se presentan las comparaciones realizadas por tres de los sistemas generativos más utilizados, ChatGPT , Gemini y DeepSeek, de las fortalezas y debilidades de estos y otros programas generativos actualmente disponibles para los usuarios de internet.

TABLA 8.1

PROPIEDADES DE LOS MODELOS GENERATIVOS DE LENGUAJE MÁS UTILIZADOS. REALIZADOS CON GPT-4.

Modelo	Razonamiento	Código	Conversación	Apertura
GPT-4	★★★★★★	★★★★★★	★★★★★★	✗
Claude	★★★★	★★★	★★★★★★	✗
Gemini	★★★	★★★	★★★★	✗
LLaMA	★★★★	★★★★	★★★	✓
Mistral	★★★	★★★	★★★	✓
DeepSeek	★★★★★★	★★★★★★	★★★	✓

TABLA 8.2

COMPARATIVA DE TRES MODELOS REALIZADA POR GEMINI

Característica	ChatGPT (OpenAI)	DeepSeek (R1/V3)	Gemini (Google)
Modelos principales	GPT-4o, o1 (preview), o1-mini.	DeepSeek-V3, DeepSeek-R1.	Gemini 1.5 Pro, 1.5 Flash.
Punto fuerte clave	Razonamiento y coherencia. Es el modelo más "redondo" y pulido para uso general y escritura creativa.	Eficiencia y código. Rendimiento nivel GPT-4 pero gratuito/barato y de código abierto (Open Weights).	Contexto y ecosistema. Ventana de contexto de hasta 2 Millones de tokens y comprensión de video nativa.
Ventajas	La serie o1 es excelente en lógica compleja.  Modo de voz avanzado.	Open Source: Puedes descargarlo y usarlo en privado.  Coste de API ridículamente bajo.	Ventana de Contexto: Puedes subir libros enteros o bases de código gigantes.  Integración total con Google (Docs, Drive, Gmail).
	Generación de imágenes (DALL-E 3).  Gran tienda de GPTs personalizados.	Excelente en matemáticas y programación.  Transparencia técnica (Papers).	Multimodalidad real (ver video/escuchar audio).  —

**TABLA 8.2 (continuación)**

**COMPARATIVA DE TRES MODELOS REALIZADA POR GEMINI**

Característica	ChatGPT (OpenAI)	DeepSeek (R1 / V3)	Gemini (Google)
Inconvenientes	Precio: La API es cara.	Privacidad: Origen chino (preocupación para algunas empresas occidentales).	A veces peca de exceso de cautela (se niega a responder por seguridad).
	"Caja Negra": No sabemos cómo se entrena.	Censura específica en temas políticos sensibles en China.	La generación de imágenes de personas ha tenido polémicas.
	Restricciones fuertes (Guardrails) a veces molestas.	Menor capacidad creativa/literaria que GPT/Gemini.	Menor disponibilidad de "razonamiento puro" (tipo o1) públicamente.
	Memoria limitada comparada con Gemini.	—	—
Mejor para...	Uso diario general, escritura creativa, brainstorming y tareas que requieran la mejor experiencia de usuario.	Programadores, investigadores, matemáticas, y empresas que quieren montar su propia IA local sin pagar API.	Analizar documentos largos (PDFs de 500 pág), analizar videos, y trabajar conectado a tu Google Workspace.

**TABLA 8.3**

**COMPARACIÓN HECHA POR DEEPSEEK DE CHATGPT, GEMINI Y DEEPSEEK**

Característica	ChatGPT (OpenAI)	Gemini (Google)	Deepseek (DeepSeek-AI)
Desarrollador	OpenAI	Google (DeepMind)	DeepSeek-AI
Modelo más avanzado	GPT-4 Turbo	Gemini Ultra	Deepseek V3
Modalidades admitidas	Texto, imagen (entrada), voz (solo ChatGPT-4o)	Texto, imagen, audio, video	Principalmente texto (con capacidad de cargar imágenes/archivos)
Costo básico	Gratuito con limitaciones, Plus: \$20/mes	Gratuito con limitaciones, Advanced: \$19.99/mes	<b>Completamente gratuito</b> (sin límites conocidos)
Contexto ventana	128K tokens (GPT-4 Turbo)	1M tokens (Gemini 1.5)	128K-1M tokens según modelo

**TABLA 8.3 (continuación)**

**COMPARACIÓN HECHA POR DEEPSEEK DE CHATGPT, GEMINI Y DEEPSEEK**

Característica	ChatGPT (OpenAI)	Gemini (Google)	Deepseek (DeepSeek-AI)
Ventajas principales	Calidad de lenguaje natural excepcional	Integración nativa con Google Workspace	<b>Completamente gratuita sin restricciones</b>
	Ecosistema amplio de herramientas y API	Multimodalidad avanzada	Buen rendimiento en tareas de razonamiento
Desventajas principales	Integración con Microsoft y otros servicios	Rendimiento destacado en benchmarks	Opción de búsqueda web integrada
	Versión gratuita limitada	Menos opciones de personalización	Menor reconocimiento de marca
Disponibilidad idiomas	Sin búsqueda web en versión gratuita	A veces demasiado cauteloso en respuestas	Interfaz menos pulida
	Historial de sesiones corto en versión gratuita	Integración con Google puede ser limitante	Capacidades multimodales más limitadas
Capacidades de programación	Multilingüe excelente (100+ idiomas)	Multilingüe muy bueno	Buena comprensión multilingüe (especialmente inglés y chino)
Precisión factual	Excelente, con soporte para múltiples lenguajes	Muy bueno, integración con Colab	Muy competente, especialmente en Python y desarrollo web
Personalización	Buena, pero puede alucinar información	Excelente, integra datos de Google Search	Buena, con opción de búsqueda web para verificación
Velocidad de respuesta	Guarda preferencias, memoria limitada	Guarda algunas preferencias	Menos opciones de personalización
	Rápida en versión gratuita, muy rápida en Plus	Variable, a veces más lenta en tareas complejas	Generalmente rápida, incluso en tareas complejas

Como conclusión, ChatGPT parece ser la opción más equilibrada y madura en su versión de pago para usuarios que buscan calidad de lenguaje natural y estabilidad. Gemini destaca por su integración con servicios de Google y capacidades multimodales avanzadas. Deepseek ofrece capacidades avanzadas completamente gratuitas, siendo la mejor opción para usuarios con presupuesto limitado o que requieren uso intensivo sin costos.

# CAPÍTULO IX

## Reconocimiento de imágenes

### 9.1. INTRODUCCIÓN

Se denomina visión por ordenador al reconocimiento de una imagen digital. Casi todas las cámaras actuales nos proporcionan imágenes digitales, pero también podemos escanear una imagen antigua y convertirla a digital. Enseñar a un ordenador a “ver” ha sido uno de los retos más difíciles de la informática. Mientras que procesar texto o números encajaba bien con la lógica digital, las imágenes planteaban un problema distinto: una imagen digital, como veremos, no es más que un conjunto muy grande de números con relaciones muy complejas entre sí. El reto consistía en transformar esa enorme cantidad de datos en conceptos visuales: bordes, formas, objetos, escenas.

En los años setenta y ochenta, los primeros enfoques de visión artificial se basaban en reglas diseñadas a mano. Los investigadores programaban detectores de bordes, esquinas o regiones uniformes, esperando que la combinación de estas piezas permitiera reconocer objetos. Sin embargo, estos sistemas eran frágiles: funcionaban bien en condiciones muy controladas, pero fallaban ante cambios de iluminación, rotación, ruido o variaciones naturales de los objetos.

A finales del siglo pasado surgió una idea clave como consecuencia de disponer de grandes conjuntos de imágenes digitales. En lugar de decirle explícitamente al ordenador qué debe buscar, entrenarle para aprender estadísticamente a partir de ejemplos. En ese contexto aparecen las redes neuronales convolucionales, inspiradas en cómo el sistema visual biológico procesa la información. Estas redes marcaron un punto de inflexión en la clasificación de imágenes y, con el tiempo, se convertirían en la tecnología dominante en visión por computador.

Estas redes nos las encontramos diariamente cuando nuestro teléfono móvil organiza las fotos por temas: viajes, comida, amigos... Es un ejemplo de cómo la inteligencia artificial se está utilizando para reconocer y clasificar imágenes. En este capítulo explicaremos cómo funciona esta capacidad.

## 9.2. DIGITALIZACIÓN DE IMÁGENES

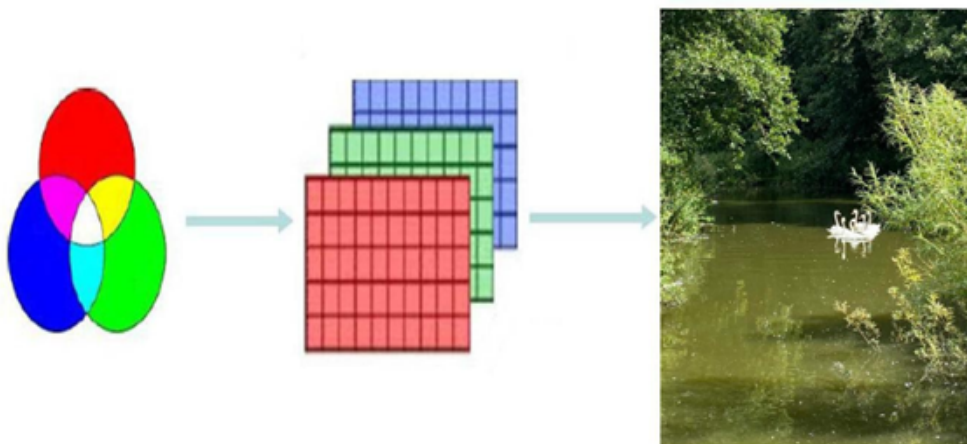
Una imagen digital es un conjunto de muchos números dispuestos en filas y columnas en una o varias estructuras rectangulares que llamamos matrices. Cada uno de estos números indica el color de una pequeña superficie cuadrada de la imagen, que llamamos píxel. Al observar el conjunto de tonalidades de estos píxeles que forman la imagen, nuestra retina integra el color de todos ellos para verla.

Estas imágenes digitales se denominan de mapa de bits, o bitmap, y su resolución se define por el número de píxeles que contiene, bien en densidad, píxeles por pulgada (ppi) o por centímetro (pcm), o bien en número de píxeles totales. Por ejemplo, una resolución de un megapíxel equivale a un millón de píxeles, o a un cuadrado con 1.000 píxeles en cada lado.

Una imagen de un tamaño dado será tanto más clara cuantos más píxeles contenga, y si aumentamos su tamaño, manteniendo el mismo número de píxeles, comenzará a verse borrosa. El ojo humano, como el de otros seres vivos, tiene un límite en la cantidad de píxeles que puede apreciar, aunque este límite es muy alto con relación a las cámaras actuales. El ojo no toma imágenes fijas, como una cámara, sino que está en continuo movimiento. Se estima que el ojo humano promedio tiene alrededor de 576 megapíxeles de resolución.

FIGURA 9.1

CONSTRUCCIÓN DE UNA IMAGEN DIGITAL CON EL SISTEMA RGB



Fuente: Elaboración propia.

Cada píxel tiene un color uniforme y la sensación de color en la imagen se forma como resultado de integrar visualmente, en la retina, las variaciones de color y luminosidad entre píxeles vecinos.

El color de cada píxel puede definirse de dos formas principales. En las figuras en blanco y negro, o en tonos de grises, el color se define por una única matriz que contiene un número asociado para cada píxel, que es la intensidad de gris. En las figuras en color, el método más utilizado, llamado RGB, es utilizar la intensidad de tres colores básicos, rojo (Red), verde (Green) y azul (Blue), y construir tres matrices del mismo tamaño. El número que contienen en cada casilla representa la intensidad de uno de los tres colores R, G y B, y la combinación de estos tres números define el color asociado a cada píxel de la imagen.

Para clasificar imágenes se utilizan sus píxeles, que pueden ser muchos, y normalmente los archivos *bitmap* se comprimen para procesarlos. Entre los sistemas de compresión más utilizados están el formato JPG, que desecha primero información no visible, y el PNG, que crea archivos algo mayores que el JPG.

### 9.3. CLASIFICACIÓN DE IMÁGENES CON REDES DE CONVOLUCIÓN

La intuición básica detrás de una red convolucional es sencilla, pero poderosa: las imágenes tienen estructura espacial local. En otras palabras, los píxeles cercanos entre sí suelen estar relacionados. Un borde, por ejemplo, se manifiesta como un cambio de intensidad entre píxeles vecinos. Una textura se caracteriza por patrones repetidos en pequeñas regiones. Un ojo ocupa una zona concreta del rostro y tiene una estructura reconocible, independientemente de dónde aparezca en la imagen. Las redes convolucionales aprovechan esta idea procesando la imagen por partes, en lugar de tratarla como un todo.

Para trabajar con imágenes, el primer paso es identificar los diferentes elementos que contienen: personas, montañas, animales, etc. Cada objeto queda definido por su contorno, que contiene píxeles con estructura similar. Para encontrar los objetos se utilizan un tipo de redes neuronales llamadas de convolución, que dividen la imagen en partes o bloques y hacen un suavizado o promedio ponderado de los píxeles de las partes. Luego utilizan la información en estos bloques para encontrar líneas y contornos de la imagen que luego se comparan con los que existen en una base de imágenes de objetos para identificarlos y clasificar la imagen dentro de una clase de objetos definidos.

El corazón de una red convolucional son las capas de convolución. Estas capas utilizan pequeños filtros (también llamados núcleos) que recorren la imagen original o las representaciones intermedias. Cada filtro es sensible a un patrón concreto. Por ejemplo, en el apéndice se estudia con detalle cómo funciona un filtro concreto para identificar líneas.

Al principio del entrenamiento, estos filtros no “saben” nada: se inicializan de forma casi aleatoria. A medida que la red ve ejemplos, los filtros se ajustan automáticamente para responder a estructuras útiles.

Inicialmente, los filtros aprenden a detectar patrones muy simples, como bordes horizontales, verticales o diagonales. En capas intermedias, combinan estos bordes para formar esquinas, curvas o texturas. Más adelante, los patrones se vuelven más abstractos: partes de objetos, configuraciones complejas o incluso objetos completos.

El resultado de aplicar un conjunto de filtros a una imagen es un conjunto de mapas de características. Cada mapa resalta las regiones donde aparece un determinado patrón.

Es importante notar que la red no trabaja con una sola representación de la imagen, sino con muchas en paralelo. Cada mapa de características ofrece una “interpretación parcial” del contenido visual. Juntas, estas interpretaciones permiten a la red construir una descripción rica y multidimensional de la imagen.

Entre capas convolucionales suelen aparecer capas de agrupamiento o *pooling*. Su función es reducir la resolución espacial de los mapas de características, conservando la información más relevante. La idea es simple: en lugar de recordar cada detalle exacto, la red se queda con una versión más compacta que indica si un patrón está o no presente, no exactamente dónde. Esto introduce una cierta invariancia a pequeños desplazamientos o deformaciones del objeto.

Tras varias capas convolucionales y de agrupamiento, la red ha transformado la imagen original en una representación de alto nivel. En este punto, suele añadirse una o varias capas (similares a las redes neuronales FF clásicas), cuya función es combinar toda la información aprendida y tomar una decisión.

Estas capas finales ya no trabajan directamente con píxeles, sino con conceptos visuales implícitos: formas, partes, relaciones espaciales. El resultado es una probabilidad para cada clase posible, y la red selecciona la más probable como salida.

Por ejemplo, en imágenes de personas, comparamos los bloques de píxeles suavizados con una base de datos con imágenes de caras, narices o cabellos. Una vez identificadas las características de la imagen, podemos clasificar la imagen completa, asignándola a grupos de categorías, como coches, edificios o retratos.

## 9.4. EJEMPLO DE GENERACIÓN DE IMÁGENES

De la misma forma que hemos visto cómo generar texto con IA, también podemos generar imágenes. Hemos pedido a varios programas de IA generar una imagen de un matemático trabajando, y el resultado combinará las imágenes de entrenamiento que existen en la base de datos utilizada por el sistema, pero mezclando características de muchas imágenes, de manera que la imagen proporcionada sea siempre algo distinta de las ya existentes.

El proceso es similar al de generar texto: se resume cada imagen en un conjunto de variables, y cuando se pide una nueva, se mezclan elementos de las imágenes existentes de acuerdo con sus probabilidades. Así se crea una imagen nueva, pero basada en las que ya teníamos.

Como ejemplo presentamos tres imágenes. Las dos primeras, creadas con los programas Gemini y ChatGPT. A ambos les he hecho la misma petición: "Genera una imagen de un investigador en inteligencia artificial". El primer programa ha sido mucho más rápido y las respuestas de ambos están en las [figuras 9.2](#) y [9.3](#). Como referencia, hemos hecho la misma solicitud a Leonardo AI, que es otro programa gratuito de generación de imágenes, y el resultado se presenta en la [figura 9.4](#).

**FIGURA 9.2**

**IMAGEN CREADA POR GEMINI DE UN INVESTIGADOR DE IA**



FIGURA 9.3

IMAGEN CREADA POR CHATGPT DE UN INVESTIGADOR DE IA



FIGURA 9.4

IMAGEN CREADA POR LEONARDO AI DE UN INVESTIGADOR DE IA



## 9.5. CONCLUSIONES

Las redes neuronales convolucionales transformaron la forma en que los ordenadores interpretan imágenes. Al abandonar las reglas programadas a mano y adoptar un aprendizaje jerárquico basado en datos, estas redes lograron cerrar una brecha histórica entre la percepción humana y la artificial.

Hoy en día, cuando un sistema identifica un objeto en una imagen con precisión sorprendente, lo hace gracias a una cascada de filtros aprendidos que, capa a capa, convierten píxeles en significado. Comprender esta idea —más que los detalles matemáticos— es clave para entender por qué las redes convolucionales se han convertido en uno de los pilares fundamentales de la inteligencia artificial moderna.

Aunque este capítulo se ha centrado en la clasificación de imágenes, las redes convolucionales se utilizan hoy en muchas otras tareas: detección de objetos, análisis médico, conducción autónoma o reconocimiento facial. En todos los casos, la idea central es la misma: aprender automáticamente jerarquías de representación visual a partir de datos. La clasificación de imágenes ha sido el primer gran éxito visible de este enfoque, pero no será el último.

## 9.6. LECTURAS ADICIONALES

- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193–202.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). *Backpropagation applied to handwritten zip code recognition*. *Neural Computation*, 1(4), 541–551.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *CVPR 2009*, 248–255.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *NeurIPS 2012*, 1097–1105.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *ICLR 2015*.

- Szegedy, C. et al. (2015). Going deeper with convolutions. CVPR 2015, 1–9.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. CVPR 2016, 770–778.

## APÉNDICE 9

Supongamos una imagen en blanco y negro. Para simplificar, la imagen es de escala de grises (blanco y negro). Cada píxel tiene un valor de intensidad: 0 si negro y 255, el valor máximo, si blanco. Los valores intermedios son tonalidades grises. Un filtro clásico para detectar líneas horizontales es este (3×3):

$$\begin{array}{ccc} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{array}$$

Este pequeño bloque de números es el filtro convolucional (o kernel). La fila central tiene valores positivos grandes y las filas superior e inferior tienen valores negativos. El filtro “se activa” cuando la fila central es clara (valores altos) y arriba y abajo son más oscuras. Esto es lo que ocurre en una línea horizontal clara sobre fondo oscuro (o al revés, según el signo). El filtro se coloca sobre un pequeño fragmento de la imagen, por ejemplo sobre estos 9 píxeles:

$$\begin{array}{ccc} 10 & 12 & 11 \\ 200 & 210 & 205 \\ 15 & 14 & 13 \end{array}$$

Aquí vemos claramente una franja clara en el centro. El filtro multiplica el valor del píxel de la imagen por el número del filtro que tiene encima y suma todos los resultados, produciendo un solo número de salida. Por ejemplo:

$$\begin{aligned} & (10 \times -1) + (12 \times -1) + (11 \times -1) + (200 \times 2) + (210 \times 2) + (205 \times 2) + (15 \\ & \quad \times -1) + (14 \times -1) \\ & (13 \times -1) = 1155 \end{aligned}$$

La fila central aporta un valor muy grande positivo y las filas superior e inferior restan poco. El resultado final es alto, indicando la detección de una línea horizontal en esa parte de la imagen. Ese número se guarda en la imagen de salida, justo en la posición central del filtro. Luego el filtro se mueve un píxel a la derecha y repite el proceso, recorriendo toda la imagen como una ventana deslizante.

Observemos que si no hay línea y el filtro cae sobre una zona uniforme, por ejemplo:

```
120 122 121
119 121 120
118 119 117
```

Aquí no hay una franja clara central y los valores son parecidos en todas las filas. El filtro da un resultado cercano a cero y guarda "Aquí no hay línea horizontal". Después de recorrer toda la imagen, se obtiene una nueva de menor tamaño donde los valores altos indican: "aquí hay una línea horizontal" y los bajos "aquí no hay línea horizontal". Esta nueva imagen es un mapa de características.

Si la imagen es RGB en color, el filtro se aplica a cada matriz de píxeles por separado y se obtienen tres resultados (uno por color). Estos valores se suman y se proporciona un único valor de salida.

En la práctica, los filtros no se programan y la red los aprende automáticamente durante el entrenamiento; algunos filtros se especializan en bordes, otros en colores, otros en texturas, etc.



# CAPÍTULO X

## Sistemas de recomendación

### 10.1. INTRODUCCIÓN

La selección de productos por parte de un cliente dentro de una oferta cada vez más amplia se ha convertido en un problema central del comercio digital. A medida que crecen los catálogos de productos —libros, películas, música, cursos, hoteles o inversiones— resulta cada vez más difícil para el usuario encontrar aquello que realmente se ajusta a sus intereses. En este contexto, surgen los sistemas de recomendación, cuyo objetivo principal es reducir la sobrecarga de información y facilitar la toma de decisiones a los clientes.

Un sistema de recomendación analiza el comportamiento, las preferencias y, en algunos casos, el contexto del usuario para sugerirle productos que probablemente le resulten atractivos. El rápido crecimiento de Internet, del comercio electrónico y de las plataformas de contenidos ha propiciado la proliferación de estos sistemas, que hoy son una pieza clave del *marketing* digital y de las experiencias de los usuarios.

Desde un punto de vista histórico, los primeros sistemas de recomendación aparecen a comienzos de los años noventa, coincidiendo con la disponibilidad de grandes bases de datos de clientes y transacciones. Con el tiempo, estos sistemas han evolucionado desde métodos simples basados en reglas hasta modelos complejos que integran aprendizaje automático, técnicas estadísticas y, más recientemente, aprendizaje profundo.

### 10.2. CLASIFICACIÓN DE LOS SISTEMAS DE RECOMENDACIÓN

Los sistemas de recomendación pueden clasificarse en tres categorías, dependiendo del procedimiento utilizado para construirlos. Los primeros y más simples son los sistemas de reconocimiento por contenido, que se basan en la historia pasada de compras similares del cliente. Este es el tipo de sistema que recomienda, por ejemplo, un fondo de inversión de bajo riesgo a clientes que en el pasado han elegido este tipo de producto.

Es el procedimiento más simple y el que utilizaría un vendedor familiarizado con las compras previas de un cliente.

El segundo tipo de sistemas de recomendación requiere más información porque recomienda, además de productos parecidos a los ya adquiridos por el cliente, otros productos que han comprado otros clientes que tienen características similares al cliente objetivo. Pueden, por tanto, proponer artículos en los que el cliente nunca ha pensado porque desconoce su existencia. Estos métodos, que se llaman de filtros colaborativos, intentan anticiparse a nuestros deseos y sugerir productos cuya existencia ignoramos.

Finalmente, los sistemas híbridos tratan de combinar toda la información disponible sobre el cliente: su historial de compras y también los datos de productos comprados por otros clientes del mismo grupo que el cliente objetivo, para el que se generan las recomendaciones.

### 10.3. RECOMENDACIÓN POR CONTENIDO

Los sistemas basados en contenido son los más simples conceptualmente. Funcionan de manera similar a un vendedor que conoce bien a un cliente y le recomienda productos parecidos a los que o bien ha comprado o bien se ha interesado o valorado positivamente en el pasado.

Estos sistemas construyen un perfil del usuario a partir de su historial: compras anteriores, productos consultados, valoraciones, tiempo de permanencia, etc. Paralelamente, cada producto se describe mediante un conjunto de características: tipo, categoría, precio, riesgo, palabras clave o atributos técnicos.

La recomendación se genera comparando el perfil del usuario con las características de los productos disponibles. Si un cliente ha comprado repetidamente artículos para deporte, el sistema tenderá a recomendarle productos similares.

Sus ventajas son que no requieren información de otros usuarios, funcionan bien incluso con pocos clientes, son interpretables, siendo fácil de explicar por qué se recomienda un producto y respetan la privacidad, al no cruzar datos entre usuarios.

Sus inconvenientes son que tienden a recomendar siempre “más de lo mismo”, dependen mucho de cómo se describan los productos y tienen poca capacidad de descubrimiento o de sorpresa. Estos sistemas son especialmente adecuados cuando los productos tienen descripciones ricas y bien estructuradas, como noticias, artículos técnicos o productos financieros.

## 10.4. RECOMENDACIÓN POR FILTRO COLABORATIVO

Los sistemas de filtrado colaborativo suponen un salto conceptual importante. En lugar de centrarse únicamente en el historial del usuario, utilizan la información agregada de muchos usuarios para generar recomendaciones. Su idea central es sencilla: usuarios con comportamientos similares suelen tener gustos similares.

Estos métodos comenzaron a utilizarse de forma sistemática en los años noventa. Uno de los hitos más influyentes fue el sistema desarrollado por Amazon, que introdujo la recomendación basada no en la similitud entre clientes, sino en la similitud entre productos.

En el filtrado colaborativo basado en usuarios, a cada usuario se le asigna un conjunto de “vecinos”: otros usuarios con patrones de compra o valoración similares. Se identifican usuarios similares al cliente objetivo; se analiza qué productos han comprado o valorado positivamente esos usuarios y se recomiendan aquellos productos que el cliente aún no ha adquirido.

Las ventajas de este sistema son que introduce descubrimiento de productos nuevos y, al no requerir descripciones detalladas de los productos, puede encontrar preferencias implícitas difíciles de modelar. Sus inconvenientes son que requiere muchos usuarios y es muy sensible a datos escasos y presenta el problema de “arranque en frío” para nuevos usuarios.

En el filtrado colaborativo basado en productos, se parte de los datos totales de compras disponibles y se construye una matriz donde aparecen todos los productos y clientes, y también sus compras. Un ejemplo simple de esta matriz se presenta en la [tabla 10.1](#). Con estos datos podemos calcular probabilidades de compra de productos y de pares de productos. Dos productos X e Y están relacionados cuando su probabilidad de compra conjunta es mayor que el producto de las probabilidades de sus compras individuales.

Partiendo de esta idea básica, se establece una asociación entre productos que, naturalmente, requiere ajustes para tener en cuenta los tiempos de compra, la disponibilidad, etc. De esta manera, para cada cliente se crea una base de productos “vecinos” a los ya comprados, que es la base de las recomendaciones.

**TABLA 10.1**

**CÁLCULO DE LAS PROBABILIDADES DE COMPRA**

	X=Planta	Y=Fertilizante	Z=Martillo	E=Escoba
Ciente 1	Sí	Sí	no	No
Ciente 2	No	No	Sí	No
...				...
...				...
...				...
Ciente 100	No	Sí	No	No
	40	36	50	60

$$\text{Prob}(X) = 40/100 = 0,4$$

$$\text{Prob}(X,Y) = 30/100 = 0,3$$

$$\text{Prob}(Y) = 36/100 = 0,36$$

$$\text{Prob}(X,Z) = 15/100 = 0,15$$

$\text{Prob}(X,Y) = 0,3$  como:  $0,3 > 0,4 \times 0,36 = 0,14$ , *dependencia* entre X e Y, productos vecinos  
 $\text{Prob}(X,Z) = 0,15$  como  $0,15 < 0,4 \times 0,5 = 0,2$  no *dependencia* entre X y Z, no vecinos

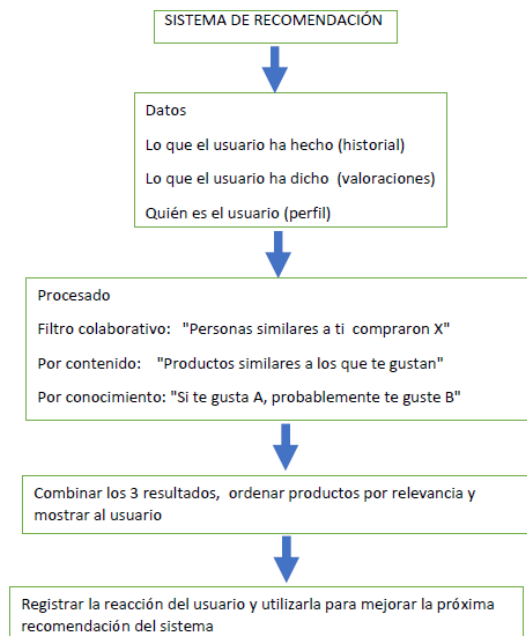
Una ventaja añadida es que los cálculos intensivos, las probabilidades de compra y las medidas de asociación se realizan *offline*, con lo que la recomendación puede generarse de forma rápida en el momento de hacer la sugerencia durante la conexión.

Los sistemas de recomendación han sido criticados por el uso de información confidencial, especialmente en el campo de la salud, donde este tema es muy sensible. A pesar de las críticas, el éxito de estos sistemas ha contribuido a su amplia difusión. Amazon estima que un 30 % de sus ventas se generan por los sistemas de recomendación y Netflix asegura que el 80 % de las películas que ven sus usuarios son las recomendadas.

La [figura 10.2](#) resume en un diagrama el proceso de recomendación por filtro colaborativo, y ha sido creada manualmente por el autor. La [figura 10.3](#) resume este mismo filtro representado con un programa generativo de IA.

**FIGURA 10.2**

**EL PROCESO DE RECOMENDACIÓN**



Fuente: Elaboración propia.

FIGURA 10.3

EL PROCESO DE RECOMENDACIÓN CREADO POR GEMINI



## 10.5. OTROS SISTEMAS DE RECOMENDACIÓN

Los sistemas híbridos son muy utilizados en la actualidad, combinando métodos de predicción, basados en modelos, con técnicas de *cluster* y asociación, para clientes y productos. Los sistemas de recomendación se han demostrado, además, útiles para desarrollar la lealtad del cliente.

Una evolución natural del filtrado colaborativo consiste en utilizar modelos de factores latentes. La idea es que tanto usuarios como productos pueden representarse en un espacio abstracto de dimensiones reducidas que capturan gustos, estilos o características implícitas. El sistema aprende estas representaciones automáticamente a partir de los datos de interacción.

## 10.6. CONCLUSIONES

Los sistemas de recomendación se han convertido en una infraestructura invisible pero esencial del mundo digital. Al guiar nuestras elecciones, influyen en lo que compramos, vemos y escuchamos. Comprender cómo funcionan, así como sus ventajas e inconvenientes, es clave para diseñarlos y utilizarlos de forma responsable

Estos sistemas están incorporando el uso de voz y combinándose con los asistentes o agentes digitales para proporcionar recomendaciones de forma verbal a solicitudes diversas del cliente, como encontrar un restaurante vegetariano, reservar una habitación en un hotel o comprar las entradas de un espectáculo.

## 10.7. LECTURAS ADICIONALES

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370.
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30–37.
- Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). *Mining of Massive Datasets*. Cambridge University Press.
- Resnick, P., & Varian, H. R. (1997). *Recommender systems*. *Communications of the ACM*, 40(3), 56–58.
- Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook* (2nd ed.). Springer.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International World Wide Web Conference (WWW 2001)*, 285–295.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. *The Adaptive Web, Lecture Notes in Computer Science*, vol. 4321, Springer, 291–324.

# CAPÍTULO XI

## La inteligencia artificial en la educación

### 11.1. INTRODUCCIÓN

En este capítulo analizamos tres herramientas fundamentales mediante las cuales la inteligencia artificial (IA) puede contribuir de forma destacada a la mejora de los procesos de enseñanza y aprendizaje:

- los tutores virtuales inteligentes,
- los sistemas de asistencia al profesor basados en el análisis del aprendizaje,
- los nuevos sistemas de evaluación, continuos, personalizados y formativos.

Estas tres herramientas no deben entenderse como elementos aislados, sino como partes de un ecosistema educativo que transforma progresivamente el papel del estudiante, del docente y de la propia institución educativa.

### 11.2. TUTORES VIRTUALES

Desde hace décadas sabemos que una de las formas más eficaces de aprender es hacerlo con un profesor particular que sea experto en la materia y capaz de adaptar el ritmo, el nivel de dificultad y los ejemplos a los conocimientos previos del estudiante. Este tipo de enseñanza personalizada permite detectar errores conceptuales, reforzar los puntos débiles y mantener la motivación.

Sin embargo, se trata de un modelo con alto coste y difícil de aplicar en general a toda la población. La mayoría de los sistemas educativos no pueden ofrecer un tutor personal a cada estudiante. Aquí es donde los tutores virtuales basados en IA aparecen como una alternativa prometedora.

Los primeros sistemas de tutoría asistida por ordenador surgieron en la segunda mitad del siglo XX, bajo el nombre de *Intelligent Tutoring Systems*. Aquellos sistemas eran rígidos, con reglas preprogramadas y escasa capacidad de adaptación real. La situación cambia radicalmente con la aparición de grandes plataformas educativas digitales y, más recientemente, con los modelos generativos de lenguaje.

Salman Khan, un ingeniero de origen indio que estudió en EE. UU., comenzó hace más de veinte años a grabar videos para explicar matemáticas a su sobrina. Sus videos estaban planteados como una clase particular interactiva, donde después de explicar cada concepto se comprobaba su asimilación por el estudiante antes de pasar al siguiente.

Estos videos circularon por la web con tanto éxito que le llevaron a crear la Khan Academy, una plataforma de aprendizaje gratuita con millones de usuarios de todo el mundo que le valió la concesión del premio Princesa de Asturias de Cooperación 2019. Su siguiente desarrollo, la plataforma gratuita Khanmigo (el nombre viene de “conmigo”, en español) integra el uso de la inteligencia artificial utilizando ChatGPT en el aprendizaje con tutores virtuales para los estudiantes y herramientas de seguimiento para los profesores.

Los tutores virtuales modernos no se limitan a mostrar contenidos. Su funcionamiento se basa en varios principios clave:

- Interacción constante: el tutor explica un concepto y plantea preguntas antes de avanzar.
- Seguimiento continuo: registra cada respuesta, cada error y el tiempo dedicado a cada actividad.
- Análisis del error: no solo detecta si la respuesta es incorrecta, sino qué tipo de error se ha cometido.
- Adaptación dinámica: ajusta ejemplos, ejercicios y explicaciones al perfil del estudiante y a su conocimiento a medida que avanza en la materia.
- Personalización: tiene en cuenta intereses, ritmo de aprendizaje y estilos cognitivos.

Con el tiempo, el tutor construye un modelo cada vez más preciso del estudiante, lo que le permite anticipar dificultades y proponer refuerzos específicos.

Los tutores virtuales no sustituyen a las clases presenciales, sino que transforman su contenido. El aula deja de ser el lugar donde los alumnos reciben clases magistrales y

evoluciona hacia un escenario en el que discutir las dudas sobre la materia estudiada e interactuar con compañeros y profesores, resolviendo problemas en grupo. Este enfoque es la filosofía del concepto de *clase invertida*, donde el aprendizaje individual se realiza fuera del aula y el tiempo presencial se reserva para la interacción con los otros, la discusión y otras actividades de valor pedagógico.

### 11.3. ASISTENCIA AL PROFESOR

La inteligencia artificial puede ayudar también al docente, suministrándole información sobre el proceso seguido por cada estudiante para adaptar el ritmo y los ejercicios que propone, tanto para casa como en el aula. Tradicionalmente, el docente ha tenido una información limitada sobre el proceso de aprendizaje individual de cada alumno, especialmente en grupos numerosos. Los tutores virtuales y plataformas educativas basadas en IA generan una enorme cantidad de datos sobre la actividad del estudiante, que es proporcionada al profesor como información pedagógicamente relevante. De esta manera, el profesor tiene información individual sobre el progreso de cada estudiante y puede adaptar sus enseñanzas al ritmo medio de la clase de forma más precisa.

Como un ejemplo muy simple de la información que puede proporcionar un tutor virtual al profesor, la **figura 11.1** representa el proceso seguido por un grupo de estudiantes online a lo largo del tiempo utilizado para realizar un conjunto de ejercicios.

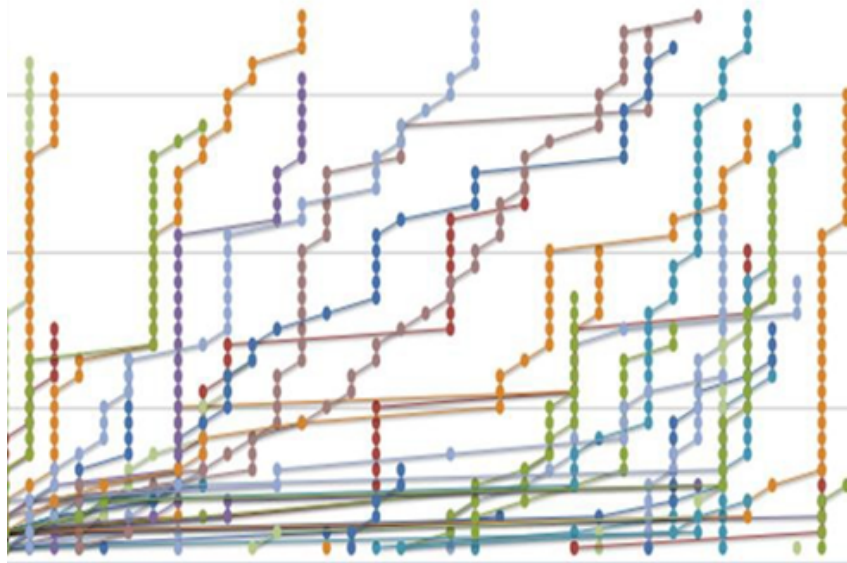
El eje vertical de la figura señala las actividades que los estudiantes han realizado y el horizontal, el tiempo invertido en completarlas. Las líneas, en distintos colores, representan las trayectorias de los estudiantes. Hay estudiantes que progresan muy rápido, representados por las líneas muy verticales, y otros más lentos con tramos horizontales, que son los que marcan actividades donde los estudiantes han tenido que invertir más tiempo repasando material anterior para completarlas.

El análisis detallado de estos gráficos de actividad, que incluyen otra mucha información, indica al profesor los distintos patrones de aprendizaje en su clase, así como los estudiantes en riesgo de fracaso. Con su ayuda, el docente puede cambiar su papel de transmisor de información a facilitador del aprendizaje de cada estudiante, mediante:

- Identificación de patrones comunes de error.
- Detección de estudiantes en riesgo de abandono o fracaso.
- Ajuste del ritmo y velocidad del curso.

FIGURA 11.1

PROCESO DE APRENDIZAJE *ONLINE* DE UN GRUPO DE ESTUDIANTES



Fuente: Elaboración propia.

- Diseños de actividades diferenciadas para los estudiantes con problemas o aquellos que avanzan más rápido.

Gracias a estos sistemas, el docente puede pasar de una enseñanza uniforme a una enseñanza más personalizada basada en los datos y adaptada a los estudiantes, sin perder su papel central y utilizando las herramientas proporcionadas por la IA para tomar mejores decisiones pedagógicas.

#### 11.4. EVALUACIÓN DEL APRENDIZAJE

La evaluación ha sido tradicionalmente uno de los aspectos más problemáticos del sistema educativo. Los exámenes puntuales ofrecen una visión parcial del aprendizaje y están influidos por factores externos como el estrés o las circunstancias personales.

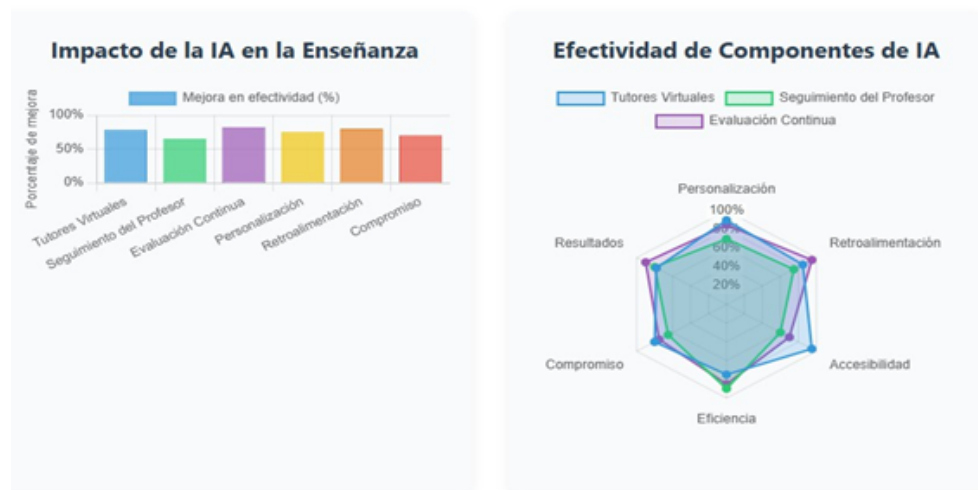
Los tutores virtuales permiten avanzar hacia una evaluación continua, basada en el seguimiento detallado del trabajo del estudiante a lo largo del tiempo, de manera que sea posible evaluar todo el proceso seguido por el estudiante, y no solo su resultado.

Analizando el proceso de cómo se resuelven los ejercicios, los sistemas de IA pueden evaluar, entre otros aspectos, la coherencia de las respuestas proporcionadas por el

estudiante, la evolución de su razonamiento y la consistencia en la forma y estilo de resolución de los problemas. Con estos datos podemos detectar, por ejemplo, si una respuesta ha sido elaborada por el propio alumno o si ha sido copiada directamente de un sistema generativo. Más allá del control, esta información es valiosa para comprender el nivel real de comprensión de los estudiantes.

FIGURA 11.2

IMPACTO Y EFECTIVIDAD DE IA EN LA ENSEÑANZA



Fuente: La figura 11.2 realizada por Gemini resume los datos de varios estudios sobre el impacto de la IA en la educación y sus efectos beneficiosos en varios aspectos considerados.

Al final del curso, el sistema puede generar un informe global del desempeño de cada estudiante, comparándolo con el resto del grupo y proponiendo una calificación orientativa. En caso de duda, puede incluso sugerir preguntas específicas para que el profesor pueda comprobar el dominio real de los contenidos de cada estudiante.

Es fundamental subrayar que la decisión final debe recaer siempre en el profesor, manteniendo la evaluación como un acto pedagógico y no puramente automático.

## 11.5. CONCLUSIÓN

Los sistemas generativos de IA fueron inicialmente percibidos por muchos centros educativos como una amenaza, lo que llevó a restricciones y prohibiciones. Sin embargo, la evolución de los tutores virtuales muestra que la IA puede integrarse de forma constructiva en la educación.

Lejos de sustituir a docentes y estudiantes, estas herramientas ofrecen la oportunidad de avanzar hacia una educación más personalizada, flexible y creativa, donde el profesor recupera su papel como guía y facilitador del aprendizaje profundo. La **figura 11.3** resume el análisis realizado en este capítulo.

**FIGURA 11.3**

**LOS BENEFICIOS DE LA IA EN LA ENSEÑANZA**



Fuente: Elaboración propia.

## 11.6. LECTURAS ADICIONALES

El reciente libro de Khan (2024) es muy recomendable.

- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Khan, S. (2012). *The One World Schoolhouse: Education Reimagined*. Twelve/Hachette.
- Khan, S. (2024). *Brave New Words. How AI will Revolutionize Education*. Penguin Random House.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.

- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599.
- UNESCO. (2021). AI and Education: Guidance for Policy-makers. UNESCO Publishing. Woolf, B. P. (2010). *Building Intelligent Interactive Tutors*. Morgan Kaufmann.



## CAPÍTULO XII

### Inteligencia artificial en finanzas

#### 12.1. INTRODUCCIÓN

Durante gran parte del siglo XX, las decisiones financieras se basaron en la experiencia, la intuición y el conocimiento experto de analistas, gestores y directivos. Aunque los datos siempre han sido importantes, su volumen y complejidad limitaban el grado en que podían ser utilizados de forma sistemática. En la última década, esta situación ha cambiado radicalmente.

Hoy, los datos, los algoritmos y el aprendizaje automático están transformando la forma en que las empresas financieras toman decisiones. Bancos, aseguradoras, gestoras de fondos y empresas *fintech* utilizan la inteligencia artificial (IA) para analizar millones de transacciones, anticipar riesgos, personalizar servicios y automatizar procesos que antes requerían grandes equipos

En este capítulo exploraremos cómo la inteligencia artificial está revolucionando el mundo financiero en dos grandes frentes: mejorando la eficiencia operativa y fortaleciendo la seguridad de sus operaciones. Al mismo tiempo, veremos que, junto a sus enormes ventajas, el uso de la IA también supone importantes riesgos y desafíos.

#### 12.2. LA EFICIENCIA EN LAS FINANZAS

La eficiencia ha sido siempre una ventaja competitiva clave en el sector financiero. Desde la automatización de la contabilidad hasta la introducción de sistemas electrónicos de pago, cada avance tecnológico ha permitido reducir costes, acelerar procesos y mejorar la fiabilidad de las operaciones.

Uno de los primeros ámbitos donde la IA ha demostrado su valor es la automatización de tareas repetitivas y de alto volumen. Procesos como la verificación de documentos,

la validación de identidades, la gestión de solicitudes de crédito o la conciliación de transacciones pueden realizarse hoy de forma automática mediante sistemas de visión artificial, procesamiento del lenguaje natural y modelos de clasificación. Esto reduce significativamente los tiempos de respuesta y los costes operativos, al tiempo que disminuye los errores humanos.

La inteligencia artificial representa un salto cualitativo porque no se limita a automatizar tareas mecánicas, sino que permite automatizar decisiones. Esto es especialmente relevante en un sector donde cada decisión —conceder un crédito, aprobar una transacción, invertir en un activo— tiene consecuencias económicas directas.

En las áreas de inversión y tesorería, los algoritmos para predecir y clasificar, que hemos estudiado en capítulos anteriores, son capaces, analizando grandes volúmenes de datos, de detectar patrones, anticipar tendencias del mercado y optimizar carteras en tiempo real.

Los asistentes virtuales y chatbots se han convertido en un elemento habitual de la atención al cliente en el sector financiero. Estos sistemas pueden atender miles de consultas diarias de forma simultánea, las 24 horas del día, respondiendo a preguntas sobre saldos, movimientos, productos financieros o incidencias técnicas.

A diferencia de los antiguos sistemas de respuesta automática, los chatbots actuales utilizan modelos de lenguaje capaces de mantener conversaciones naturales, comprender el contexto y adaptarse al perfil del cliente. Además, pueden dirigir automáticamente la consulta al personal humano cuando detectan situaciones complejas o sensibles.

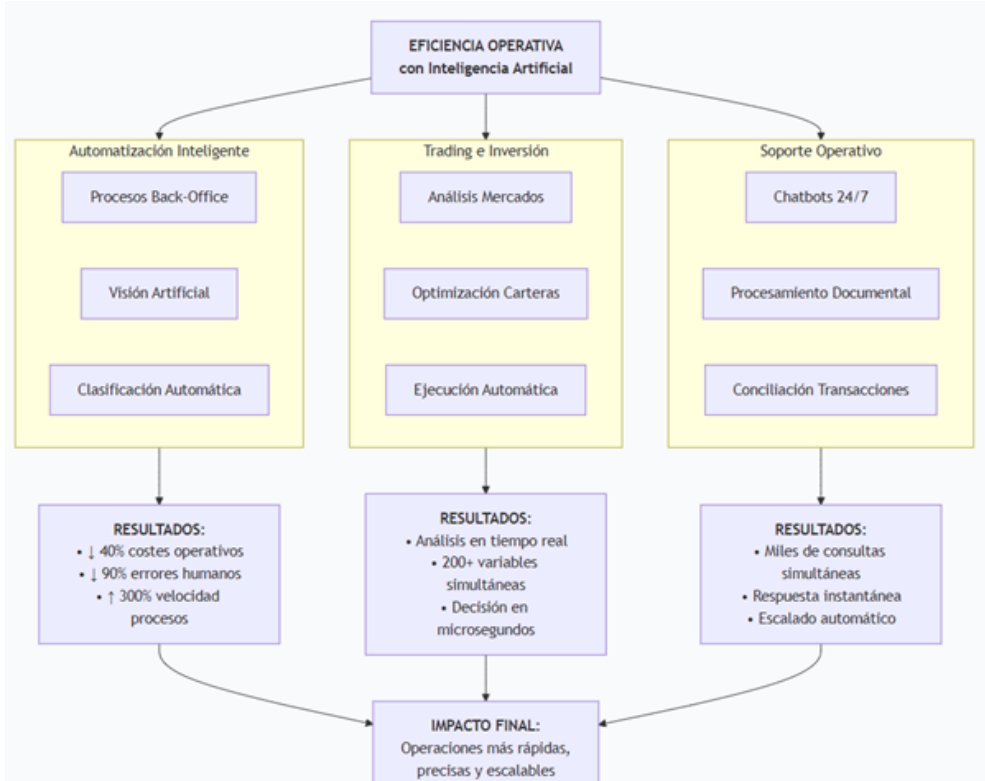
En las áreas de inversión, *trading* y tesorería, la IA se utiliza para analizar grandes volúmenes de datos financieros, económicos y de mercado en tiempo real. Los algoritmos de predicción y clasificación permiten detectar patrones ocultos en series temporales, anticipar cambios de tendencia, optimizar carteras según distintos criterios de riesgo y rentabilidad y reaccionar rápidamente ante eventos inesperados.

Aunque estos sistemas no eliminan la incertidumbre inherente a los mercados, sí permiten una gestión más informada y ágil, especialmente en entornos de alta volatilidad.

Así, las empresas financieras pueden predecir los cambios y prepararse para adaptarse a ellos antes de que ocurran. La [figura 12.1](#) resume las razones del aumento de la eficiencia con IA que hemos presentado en esta sección.

FIGURA 12.1

EL AUMENTO DE LA EFICIENCIA CON IA



Fuente: Elaboración propia con ayuda de Gemini, Chatgtp and DeepSeek.

### 12.3. GESTIÓN DEL RIESGO Y SEGURIDAD FINANCIERA

Además de mejorar la eficiencia, los métodos de inteligencia artificial están siendo muy útiles para aumentar la seguridad mediante la gestión del riesgo y la detección de fraudes. Los modelos predictivos permiten identificar señales tempranas de impago, movimientos inusuales en cuentas y operaciones potencialmente fraudulentas incluso antes de que ocurran.

La IA ha transformado este ámbito al permitir pasar de una gestión reactiva del riesgo a una gestión predictiva y preventiva.

Uno de sus usos más extendidos en finanzas es la detección del fraude. Cada cliente genera un patrón de comportamiento financiero: tipos de comercio, importes habitua-

les, localización geográfica, frecuencia de uso de tarjetas, horarios, etc. Cuando una transacción se desvía significativamente de este patrón, el sistema puede generar una alerta, solicitar verificaciones adicionales o bloquear temporalmente la operación.

Este enfoque permite detectar fraudes en tiempo real, incluso cuando los importes individuales son pequeños, si el patrón global resulta sospechoso.

Como ejemplo, un banco europeo redujo fraudes con tarjetas en un 65 % tras implementar un sistema que analiza no solo transacciones individuales, sino patrones relacionales entre 50.000 comercios afiliados.

La IA permite analizar las relaciones económicas entre clientes de una entidad financiera, como transferencias recurrentes entre ellos, por cuotas o pagos de alquiler, ingresos regulares o puntuales, etc. Esta información puede ser muy valiosa para identificar riesgos sistémicos, detectar comportamientos anómalos y mejorar la predicción de impagos. Como veremos en el capítulo siguiente, las características de los interlocutores financieros de un cliente son a menudo excelentes predictores de su comportamiento futuro.

Una de las ventajas de la IA es su capacidad de aprender continuamente. Cada nuevo dato mejora el modelo, ajustando sus parámetros y reduciendo progresivamente los errores. Esto convierte la gestión del riesgo en un proceso dinámico, donde los modelos evolucionan con el entorno económico y los hábitos de los clientes, reduciendo la dependencia exclusiva del juicio humano. La [figura 12.2](#) resume estas ventajas para mejorar la seguridad.

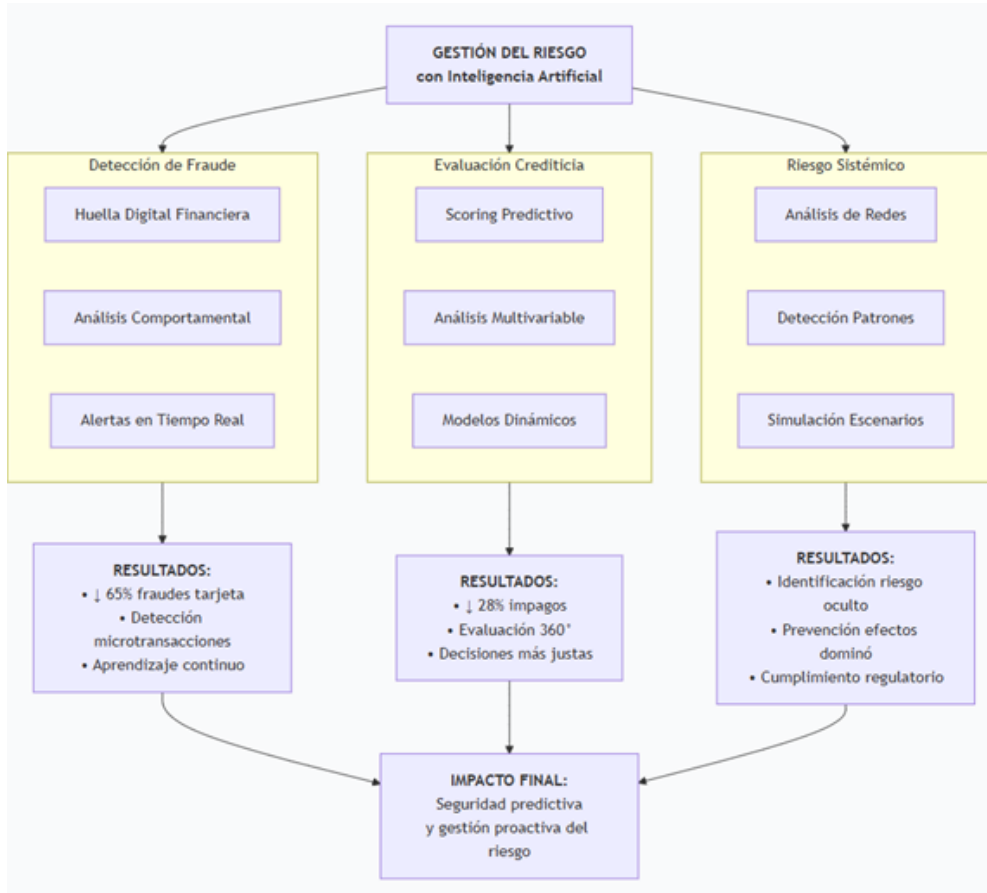
## 12.4. PERSONALIZACIÓN DE LOS PROCEDIMIENTOS

Paradójicamente, la digitalización masiva y la automatización bancaria debilitaron inicialmente la relación entre las entidades financieras y sus clientes, que había sido muy personalizada en el pasado y se convirtió en más impersonal. Para evitarlo, los bancos concentraron su atención en los mejores clientes, prestando un servicio más lejano a la mayoría de los usuarios. La IA permite ahora recuperar, a gran escala, un nivel de personalización similar al que existía cuando el gestor conocía personalmente a cada cliente y aplicárselo a todos ellos.

Esta personalización tiene dos componentes. El primero, conocer y tener en cuenta la situación financiera actual y prevista de cada cliente, que se enriquece comparándola con la experiencia de clientes similares. De esta forma, se pueden ofrecer créditos

FIGURA 12.2

LA GESTIÓN DEL RIESGO CON IA



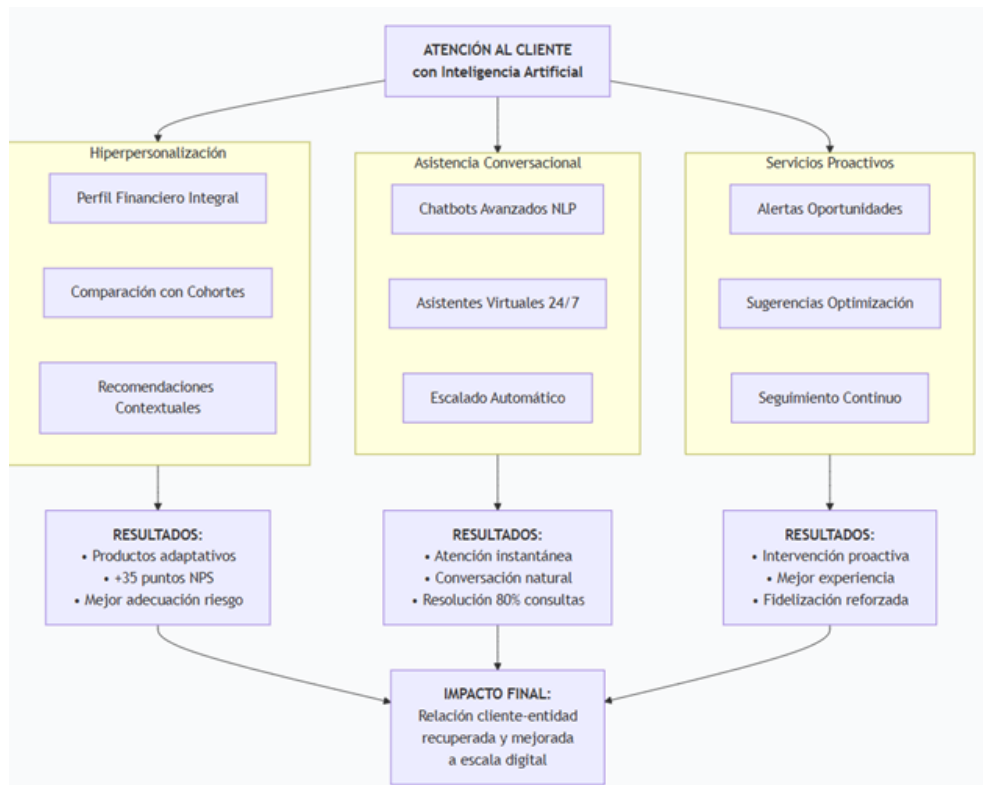
Fuente: Elaboración propia con ayuda de Gemini, Chatgpt and DeepSeek.

ajustados a la capacidad real de pago, seguros alineados con los riesgos efectivos y recomendaciones de inversión coherentes con el perfil financiero y vital del usuario. El segundo componente es estar disponible para atender al cliente cuando lo solicite.

Este enfoque mejora la personalización del cliente, permite una atención continua a sus dudas y necesidades y conduce a ofrecerle oportunidades que desconoce. Como resultado se obtienen clientes más satisfechos y leales a la entidad financiera. Además, este enfoque reduce el riesgo tanto para la entidad como para el usuario. En la [figura 12.3](#) se describe esta estrategia basada en IA.

FIGURA 12.3

LA ATENCIÓN AL CLIENTE CON IA



Fuente: Elaboración propia con ayuda de Gemini, Chatgtp and DeepSeek.

## 12.5. DESAFÍOS REGULATORIOS Y ÉTICOS

Los sistemas de IA se entrenan con datos históricos, que pueden reflejar sesgos sociales, económicos o culturales. Un ejemplo conocido es el de algunos sistemas de selección automática que discriminaban a mujeres debido a los patrones presentes en los datos de entrenamiento. En el ámbito financiero, estos sesgos pueden traducirse en discriminación en el acceso al crédito, a seguros o a productos de inversión.

Uno de los mayores desafíos es la explicabilidad de los modelos. Las decisiones financieras tienen consecuencias importantes, y es esencial poder explicar por qué se ha concedido o denegado un crédito, o por qué se ha activado una alerta de fraude.

Por ello, los reguladores exigen cada vez más la llamada trazabilidad de las decisiones, es decir, entender exactamente cómo el algoritmo aplicado ha llegado a la decisión final.

Para garantizar esta trazabilidad hay que someter a auditorías a los algoritmos e introducir controles humanos en los procesos críticos antes de aplicar la decisión obtenida.

Las instituciones nacionales e internacionales están desarrollando marcos regulatorios que promueven un uso responsable de la IA en finanzas, protegiendo la privacidad de los clientes y garantizando la equidad y la rendición de cuentas. Por ejemplo, el Parlamento Europeo aprobó, el 13-3-2024, un reglamento sobre el uso de sistemas de inteligencia artificial en la Unión Europea (UE) que sitúa al sector bancario como de alto riesgo.

## 12.6. CONCLUSIONES

La inteligencia artificial en las finanzas es ya una realidad consolidada. Su uso permite operaciones más eficientes, seguras y personalizadas, al tiempo que introduce nuevos desafíos éticos y regulatorios.

Aunque, como hemos comentado, el uso de la IA no está exento de riesgos, también representa una oportunidad para construir un sistema financiero más robusto, transparente y centrado en el cliente. En el próximo capítulo veremos con mayor detalle y mediante ejemplos cómo estas tecnologías se están integrando en los bancos para mejorar simultáneamente la eficiencia y la gestión del riesgo.

## 12.7. LECTURAS ADICIONALES

- Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*, 61, 101646.
- Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in Finance: a comprehensive review through bibliometric and content analysis. *SN Business & Economics*, 4(2), 23.
- Bouveret, A. (2018). Cyber risk for the financial sector: a framework for quantitative assessment. *IMF Working Papers*.
- European Central Bank. (2020). Artificial intelligence in financial services. *ECB Occasional Paper*.
- Gomber, P., Koch, J.-A., & Siering, M. (2017). Digital Finance and FinTech: current research and future research directions. *Journal of Business Economics*, 87, 537–580.

- Hilpisch, Y. (2020). *Artificial intelligence in finance*. O'Reilly Media.
- Mayer-Schönberger, V., & Ramge, T. (2018). *Reinventing Capitalism in the Age of Big Data*. Basic Books.
- OECD. (2021). *Artificial Intelligence, Big Data and Financial Stability*. OECD Publishing.
- Weber, P., Carl, K. V., & Hinz, O. (2023). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, 74(2), 867-907.

## CAPÍTULO XIII

### Inteligencia artificial en la banca

La banca ha sido históricamente uno de los sectores más intensivos en información. Cada pago, cada transferencia, cada crédito concedido o devuelto genera un registro que refleja la actividad económica de personas y empresas. Durante décadas, gran parte de esta información se utilizó de forma agregada y retrospectiva, principalmente para cumplir requisitos contables y regulatorios.

La digitalización completa de las actividades bancarias ha cambiado radicalmente este escenario. Hoy, los bancos disponen de grandes bases de datos detalladas y actualizadas en tiempo casi real, lo que abre la puerta al uso sistemático de técnicas de inteligencia artificial (IA) para mejorar tanto la eficiencia operativa como la seguridad financiera.

En este capítulo analizamos dos aplicaciones concretas de la IA en la banca. En primer lugar, veremos cómo el análisis de redes de clientes permite comprender mejor la estructura económica del banco y tomar decisiones más informadas. En segundo lugar, estudiaremos cómo los modelos predictivos basados en redes neuronales permiten estimar con mayor precisión la probabilidad de que un cliente entre en mora, incorporando no solo su historial individual, sino también su posición dentro de la red financiera.

#### 13.1. REDES DE CLIENTES

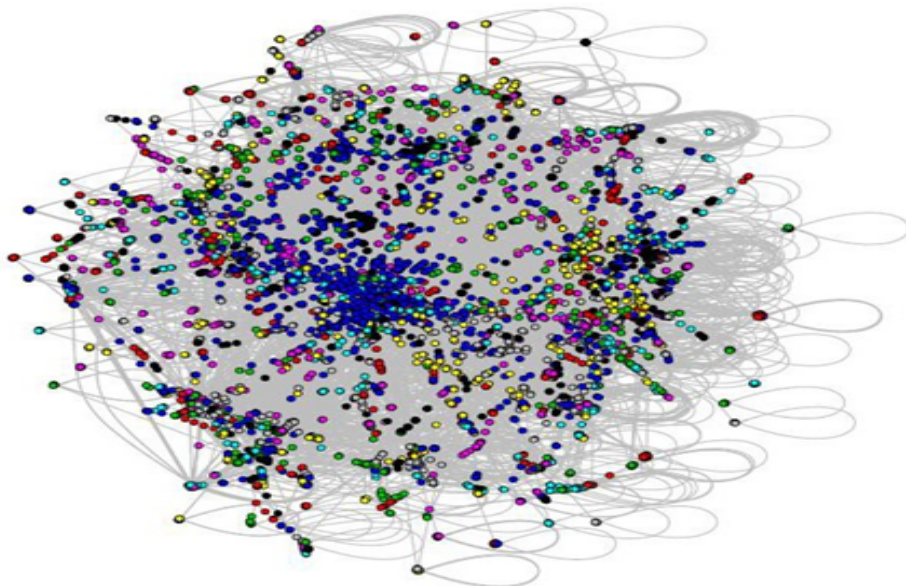
Tradicionalmente, el análisis bancario se ha centrado en el cliente individual: ingresos, gastos, activos, deudas y comportamiento pasado. Sin embargo, ningún cliente opera de forma aislada. Las personas y empresas están conectadas entre sí mediante pagos, transferencias, alquileres, nóminas, préstamos y relaciones comerciales.

Una red de clientes es una representación gráfica de estas relaciones económicas. En ella, cada persona física o jurídica se representa como un nodo; las relaciones económicas entre dos personas se representan como conexiones o arcos y a cada conexión se le asocia información sobre el volumen, frecuencia y tipo de transacciones. Este enfoque relacional permite pasar de una visión fragmentada a una visión sistémica de la actividad bancaria.

Por ejemplo, la **figura 13.1** muestra la red de clientes de un gran banco en una zona definida. Cada persona física o jurídica se representa como un nodo, que son los pequeños círculos de colores que se ven en el gráfico. Los colores en la red ayudan a distinguir distintos tipos de clientes o comunidades económicas.

**FIGURA 13.1**

**UNA RED DE CLIENTES DE UN BANCO**



*Fuente:* Elaboración propia.

Los nodos se unen por conexiones llamadas arcos, a los que se asocian el conjunto de transferencias o intercambios económicos entre los dos nodos en un periodo de tiempo determinado.

Este tipo de redes se utiliza en inteligencia artificial para descubrir grupos de clientes, identificar patrones en los tipos de relación y analizar la importancia de cada nodo dentro del sistema.

## **13.2. MEJORA DE LA EFICIENCIA MEDIANTE EL ANÁLISIS DE REDES**

El análisis de la red de clientes efectuado por la IA ofrece una importante herramienta para mejorar la eficiencia del banco, ya que le permite identificar comunidades o grupos de cliente y también clasificar a cada cliente por su actividad en la red.

Las comunidades de clientes agrupan nodos que mantienen relaciones económicas más intensas entre sí que con el resto de la red. Estas comunidades pueden corresponder a sectores económicos, zonas geográficas, cadenas de suministro o grupos familiares o

empresariales. La IA permite detectar estas estructuras automáticamente, sin necesidad de conocimiento previo, lo que facilita una segmentación más precisa de los clientes.

Tradicionalmente, los clientes se han clasificado en los bancos por sus activos, pero las redes permiten clasificar también a los clientes por su importancia estratégica. El valor de un cliente para un banco no depende únicamente de sus activos actuales, sino de la importancia de las consecuencias económicas de su pérdida. Un cliente puede ser estratégico por su posición dentro de la red, aunque su volumen de negocio individual sea moderado. Por ejemplo, un cliente que conecta dos grandes grupos económicos puede ser especialmente valioso, ya que su salida podría debilitar la relación del banco con ambos grupos.

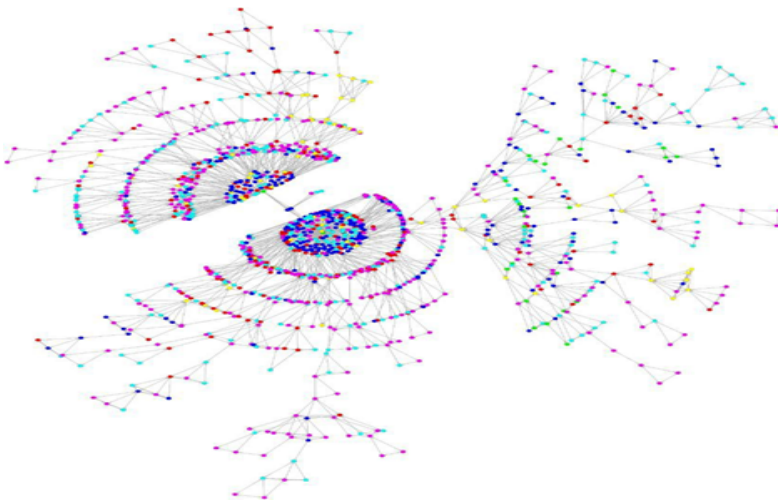
El análisis de redes permite identificar nodos muy conectados, detectar clientes que actúan como puentes entre comunidades y estimar el impacto que tendría la pérdida de un cliente sobre el conjunto del sistema, que indica su valor estratégico para el banco.

Además de analizar clientes actuales, la red incluye clientes potenciales, es decir, personas físicas y jurídicas con las que los clientes del banco mantienen relaciones económicas. Estos clientes potenciales pueden incorporarse a la red y la IA puede identificar aquellos nodos que, por su posición estratégica, representan una oportunidad especialmente atractiva para el banco. Este enfoque permite priorizar acciones comerciales y diseñar estrategias de captación basadas en evidencia, en lugar de campañas genéricas.

Por ejemplo, en la **figura 13.2** se representa una parte de una red de clientes donde vemos que aparece un nodo que conecta a dos comunidades amplias. Este cliente puede ser de gran valor estratégico para el banco.

**FIGURA 13.2**

**DOS GRUPOS DE CLIENTES CONECTADOS POR UN CLIENTE CLAVE EN EL RED**



*Fuente:* Elaboración propia.

El conocimiento detallado de la red facilita también recomendaciones personalizadas de productos financieros, así como anticipación de cambios en la estructura de la red.

### 13.3. CONTROL DEL RIESGO DE CRÉDITO

Uno de los mayores riesgos para una entidad bancaria es el riesgo de crédito, es decir, la posibilidad de que un cliente no pueda devolver un préstamo. La probabilidad de que esto ocurra se denomina la probabilidad de mora y su estimación es esencial para fijar las condiciones de un crédito, calcular provisiones por posibles impagos y estimar el riesgo esperado, es decir, la cantidad que el banco podría perder. También es necesario su cálculo para cumplir con los requisitos regulatorios establecidos para el riesgo de crédito.

Tradicionalmente, esta probabilidad se estimaba analizando datos históricos del cliente y del tipo de crédito, como, por ejemplo, los servicios contratados con el banco, los recursos del cliente y los cambios en su situación económica a lo largo del tiempo. Estos modelos, aunque útiles, tenían una capacidad limitada para capturar relaciones complejas y efectos indirectos.

Hoy, gracias a nuestra alta capacidad de cálculo, podemos incorporar mucha más información, muy detallada, sobre las actividades del cliente, ya que toda transacción o actividad deja huella en el sistema que puede incorporarse como variable. La digitalización de esta información permite tener en cuenta miles de variables relacionadas con la actividad financiera del cliente: frecuencia de transacciones, regularidad de ingresos, cambios en el comportamiento, patrones temporales y estacionales, etc.

A esta información se añade ahora un elemento clave: la posición del cliente dentro de la red financiera, que nos permite incluir variables sobre sus conexiones con clientes que han entrado, o tengan altas probabilidades de entrar, en mora.

Los estudios empíricos muestran que el riesgo financiero no es independiente entre clientes. Las dificultades económicas tienden a propagarse a través de la red de relaciones. Algunas variables especialmente relevantes son el número de conexiones directas con clientes que ya han entrado en mora, el número de clientes morosos conectados a dos pasos y la intensidad de las relaciones económicas con estos clientes.

Esto ocurre porque los comportamientos financieros suelen propagarse por la red: las dificultades económicas de algunos nodos afectan a otros con los que tienen relaciones. De hecho, el banco puede predecir dificultades de sus clientes antes de que ellos las perciban. Por ejemplo, si un cliente recibe una parte importante de su renta por un alquiler y su inquilino entra en mora, es probable que el problema se traslade al propietario.

Los modelos basados en redes neuronales son especialmente adecuados para este tipo de problemas, ya que pueden capturar relaciones no lineales, interacciones entre variables y efectos indirectos dentro de la red. Estos modelos permiten combinar información individual y relacional de forma no lineal para generar predicciones más precisas que los enfoques tradicionales. Sus predicciones proporcionan al banco la posibilidad de detectar riesgos antes de que se materialicen y, también, de ajustar las condiciones de crédito de forma dinámica ante cambios imprevistos.

### 13.4. LA NECESIDAD DE SUPERVISIÓN HUMANA

La importancia de las consecuencias y la potencia y rapidez de los modelos de decisión basados en IA requieren que su uso deba estar siempre acompañado de supervisión humana. Las decisiones automáticas en banca tienen consecuencias importantes para personas y empresas, por lo que los procedimientos automáticos basados en algoritmos tienen que ser auditables y las decisiones explicables. Esto requiere una combinación de IA y juicio humano para asegurar decisiones éticas y responsables.

La adopción de IA en una organización no es solo un cambio tecnológico, sino también organizativo, ya que hay que contratar o formar a personas con nuevos perfiles profesionales para manejar los sistemas de decisión. Para ello es fundamental la colaboración entre expertos en datos y financieros, de manera que se establezca una cultura orientada a la toma de decisiones basada en datos, que garantice los derechos de los clientes y la equidad y responsabilidad en las decisiones, proteja la privacidad de los clientes y esté libre de sesgos algorítmicos. Los bancos que logren integrar estos elementos estarán mejor posicionados en un entorno cada vez más competitivo.

### 13.5. CONCLUSIONES

La inteligencia artificial está transformando profundamente la banca. El análisis de redes de clientes y los modelos predictivos basados en redes neuronales permiten mejorar simultáneamente la eficiencia, la seguridad y la calidad del servicio.

La clave del éxito no reside únicamente en la tecnología, sino en su integración responsable dentro de los procesos bancarios, combinando el poder de la IA con la experiencia y el criterio humano.

### 13.6. LECTURAS ADICIONALES

- Alt, R., Fridgen, G., & Chang, Y. (2024). The future of fintech—Towards ubiquitous financial services. *Electronic Markets*, 34(1), 3.

- Aziz, L. A. R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. *Reviews of Contemporary Business Analytics*, 6(1), 110-132.
- Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
- Gyau, E. B., Appiah, M., Gyamfi, B. A., Achie, T., & Naeem, M. A. (2024). Transforming banking: Examining the role of AI technology innovation in boosting banks financial performance. *International Review of Financial Analysis*, 96, 103700.
- Paramesha, M., Rane, N., & Rane, J. (2024). Artificial intelligence, machine learning, deep learning, and blockchain in financial and banking services: A comprehensive review. *Machine Learning, Deep Learning, and Blockchain in Financial and Banking Services: A Comprehensive Review*.
- Rahman, M., Ming, T. H., Baigh, T. A., & Sarker, M. (2023). Adoption of artificial intelligence in banking services: an empirical analysis. *International Journal of Emerging Markets*, 18(10), 4270-4300.
- Sadok, H., Sakka, F., & El Maknoui, M. E. H. (2022). Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance*, 10(1), 2023262.

# CAPÍTULO XIV

## Conclusiones

### 14.1. EL FUTURO DE LA IA

A lo largo de este libro hemos visto que la inteligencia artificial no es una tecnología aislada ni un simple conjunto de algoritmos. Es una forma nueva de producir conocimiento, tomar decisiones y organizar la acción colectiva. Por ello, su impacto va mucho más allá de la eficiencia técnica o la automatización de tareas. Joseph Weizenbaum, uno de los pioneros de la informática que nos hemos encontrado en el capítulo I, dijo que “lo que hacemos con los ordenadores refleja quiénes somos y qué tipo de sociedad queremos ser” (Weizenbaum, 1976). Cada sistema de IA incorpora supuestos sobre el mundo: qué datos se consideran relevantes, qué variables se priorizan y qué resultados se optimizan. En este sentido, la IA no es neutral: refleja las prioridades, los valores y también las limitaciones de las empresas que han construido los algoritmos y de las sociedades que los aplican. Como advierte Cathy O’Neil, los algoritmos no son neutrales: son opiniones incrustadas en código (O’Neil, 2016). El futuro de la inteligencia artificial dependerá menos de la potencia de cálculo que de decisiones humanas conscientes sobre su uso y sus límites.

Comenzando con los cambios positivos que podría aportar la IA, uno de los principales podría ser acelerar la producción de conocimiento científico. Tradicionalmente, la ciencia ha avanzado formulando hipótesis que luego se contrastan experimentalmente. Hoy, en muchos campos, el volumen y la complejidad de los datos hacen que el descubrimiento comience incluso antes de que exista una hipótesis explícita. Como ha señalado uno de los pioneros de la IA, Yann LeCun, gran parte de la inteligencia no consiste en razonamiento simbólico, sino en la capacidad de aprender regularidades del mundo (LeCun, 2019). Los sistemas de IA destacan precisamente en esta tarea: detectar patrones que escapan a la exploración humana directa mediante métodos estadísticos automáticos de análisis de datos masivos.

Este aspecto es una de las grandes promesas de la IA. Sin embargo, como advirtió Norbert Wiener, una máquina puede ser peligrosa no por lo que hace mal, sino por

lo bien que hace exactamente lo que se le pide (Wiener, 1950). En disciplinas como la biología molecular, la física de materiales o la climatología, la IA identifica relaciones que luego los científicos deben interpretar. Esto introduce una tensión nueva en la ciencia: modelos que predicen con gran precisión sin ofrecer explicaciones comprensibles. ¿Podemos aceptar resultados científicos que funcionan aunque no sepamos exactamente por qué? El reto no es solo hacer ciencia más rápido, sino preservar la capacidad de cuestionar los resultados generados automáticamente y ser capaces de convertirlos en conocimiento científico entendible y contrastable.

Existen riesgos importantes cuando se producen avances en la predicción de un fenómeno que no van acompañados de una mejor comprensión de este. Existe amplia evidencia de que un sistema que optimiza patrones basado en las regularidades encontradas en los datos analizados puede reforzar sesgos existentes, dar un peso desmedido a fenómenos atípicos y orientar la investigación futura hacia lo que es fácilmente medible, y no hacia lo conceptualmente relevante.

Como ejemplos de estos riesgos, comentaremos el caso de las predicciones de gripe realizadas por Google Trends en 2012-2013, utilizando los datos de búsquedas en la red sobre la gripe, que sobreestimó enormemente los casos al utilizar las correlaciones encontradas sin un modelo epidemiológico capaz de explicarla y que puede consultarse en Lazer *et al.* (2014). Otro caso célebre es el algoritmo utilizado en hospitales de EE. UU. para asignar recursos médicos que subestimaba sistemáticamente la gravedad de los pacientes afroamericanos, descrito en Obermeyer *et al.* (2019), porque predecía los costes sanitarios que habían sido tradicionalmente más bajos para los pacientes negros que para los blancos. En tercer lugar, el trabajo de Bender *et al.* (2021) ilustra los riesgos, frecuentemente estudiados en la literatura estadística, de basar decisiones científicas en correlaciones y patrones aprendidos, pero no comprendidos científicamente.

Un problema adicional en el desarrollo de la IA son los enormes recursos energéticos y ambientales que requiere. Véase, por ejemplo, Bolón-Canedo *et al.* (2024). La creciente introducción de la IA en las empresas puede acarrear importantes problemas medioambientales y de consumo de recursos naturales, ya que el entrenamiento y funcionamiento de estos sistemas requiere enormes cantidades de energía, en muchos casos procedente de fuentes no renovables. Los grandes centros de datos que sostienen la IA demandan electricidad constante y sistemas de refrigeración intensivos, lo que incrementa las emisiones de gases de efecto invernadero y la huella de carbono global. Además, la fabricación del *hardware* necesario —como servidores, chips y dispositivos de almacenamiento— implica la extracción de minerales escasos, el uso intensivo de agua y la generación de residuos electrónicos difíciles de reciclar. Si este crecimiento no va acompañado de regulaciones, eficiencia energética y una transición real hacia energías limpias, la expansión de la IA podría agravar la presión sobre los ecosistemas y acelerar el agotamiento de recursos naturales clave.

Un debate importante es cómo se reparten socialmente los beneficios de la IA. Actualmente, las grandes empresas tecnológicas están poniendo recursos gigantescos en el avance de la IA, y acumulando el beneficio de la comercialización de sus aplicaciones. Al combinar la recogida masiva de datos con su análisis mediante técnicas de IA, consiguen una situación de dominio tecnológico como no se había producido antes en el mundo. Esta es una gran diferencia con el desarrollo del conocimiento en el pasado, realizado en gran medida por las universidades y centros públicos de investigación, donde los descubrimientos se hacen públicos y se ponen a disposición de toda la humanidad. La concentración de los avances en IA y su explotación comercial en un grupo reducido de empresas está contribuyendo a incrementar la desigualdad y reducir la competencia de forma muy preocupante, en este siglo. La amenaza es que continúe haciéndolo en el futuro, creando una sociedad más polarizada y con más tensiones internas. Por ejemplo, existe amplia evidencia de que la unión de la globalización y el avance de la IA en este siglo está llevando a una sociedad más desigual, donde la participación del trabajo en la renta está disminuyendo frente a la parte del capital. Las grandes inversiones empresariales en IA reducen la mano de obra y favorecen los beneficios empresariales.

La política puede también verse muy influida por el desarrollo de la IA si cada vez más las decisiones públicas se apoyan en sistemas automáticos: asignación de recursos, evaluación de riesgos, detección de fraude o diseño de políticas sociales. Herbert Simon ya advertía que quien controla la información controla la decisión (Simon, 1960). En el contexto actual, esta afirmación puede extenderse a los algoritmos que procesan dicha información y delimitan qué opciones aparecen como razonables o viables.

La promesa de decisiones “objetivas” basadas en datos convive con un riesgo profundo: trasladar decisiones políticas a sistemas técnicos opacos, difíciles de auditar y alejados del debate democrático. Cuando una decisión se justifica con un “lo dice el algoritmo”, la responsabilidad se diluye.

Uno de los grandes desafíos hacia el futuro es la relación entre IA y democracia. El control de los ciudadanos por su actividad en las redes sociales puede utilizarse para dirigir su voto y su participación política, como ha ilustrado Da Empoli (2020). Los experimentos secuenciales automáticos para medir nuestras reacciones a las informaciones recibidas y enviarnos supuestos hechos, con frecuencia inventados, para condicionar nuestras opiniones, influyen ya en lo que vemos, leemos y compartimos, configurando de forma indirecta la esfera pública. Al optimizar métricas como la atención o la interacción, estos sistemas pueden reforzar burbujas informativas, polarización y desinformación. La IA no crea estos problemas desde cero, pero puede amplificarlos a una escala inédita, obligando a replantear el papel de la tecnología en las democracias contemporáneas.

Uno de los peligros más sutiles del avance de la IA es la delegación progresiva de responsabilidades humanas. A medida que los sistemas parecen funcionar bien, aumenta

la tentación de aceptar sus resultados sin cuestionarlos. Stuart Russell lo expresa con claridad al señalar que el verdadero problema no es que las máquinas sean demasiado inteligentes, sino que persigan objetivos que no coinciden exactamente con los nuestros (Russell, 2019). Delegar decisiones implica aceptar objetivos implícitos que quizá nunca hemos discutido de forma explícita. La responsabilidad moral, el juicio y la deliberación colectiva no pueden automatizarse sin pérdida.

El futuro de la IA exige algo más que innovación técnica: exige gobernanza. Como subraya Timnit Gebru, no basta con preguntar si un sistema funciona, sino para quién funciona y a costa de quién (Gebru *et al.*, 2021). Regular la IA no significa frenar el progreso, sino orientar su desarrollo hacia la transparencia, la explicabilidad y la protección de la privacidad, asegurando un control sobre sus aplicaciones, encaminado a garantizar el bien común y el progreso de la sociedad.

## 14.2. LECTURAS RECOMENDADAS

A continuación, indicamos algunos libros básicos, la mayoría publicados en español, sobre los fundamentos y posibilidades de la IA que amplían y complementan la introducción incluida en este libro. Se añade un breve comentario sobre cada uno para guiar al lector.

- Boden, M. A. (2017). *Inteligencia artificial*. Turner, Madrid.  
Una buena introducción que ofrece el equilibrio necesario entre la historia de la IA y sus fundamentos filosóficos básicos.
- Crawford, K. (2022). *Atlas de inteligencia artificial: Poder, política y costos planetarios*. Fondo de Cultura Económica, Madrid.  
Un estudio crítico sobre el ejercicio del poder y los costes planetarios ocultos tras el desarrollo de la inteligencia artificial.
- Kai-Fu Lee. (2020). *Superpotencias de la inteligencia artificial: China, Silicon Valley y el nuevo orden mundial*. Deusto, Barcelona.  
Obra fundamental para entender la competencia geopolítica entre China y Estados Unidos y el nuevo orden mundial que esta tecnología impone.
- Kneusel, R. T. (2024). *Cómo funciona la inteligencia artificial: De la magia a la ciencia*. Marcombo.  
Explica bien la base científica y matemática que sustenta lo que a menudo percibimos como magia.

- Larson, E. J. (2022). *El mito de la inteligencia artificial: Por qué las máquinas no pueden pensar como nosotros*. Shackleton Books, Barcelona.  
Explica con gran claridad por qué la inteligencia de las máquinas sigue estando radicalmente alejada de la intuición y el pensamiento humano.
- López de Mántaras, R. y Meseguer, P. (2017). *Inteligencia Artificial*. Los Libros De La Catarata.  
Un libro muy recomendable. El primer autor es uno de los pioneros de la IA en España con reputación internacional y gran divulgador.
- Mitchell M. (2024). *Inteligencia Artificial. Guía para seres pensantes*. Titivillus.
- O'Neil, C. (2018). *Armas de destrucción matemática: Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán Swing.  
Obra imprescindible para entender cómo el uso de algoritmos y *big data* puede reforzar desigualdades sociales y poner en riesgo la democracia sin una supervisión ética adecuada.  
Libro muy competente y ameno por una gran investigadora de EE. UU. del Santa Fe Institute. Proporciona una versión muy completa y sugerente de la IA.
- Ortega Candel, J. M. (2025). *Inteligencia artificial: Investigaciones, aplicaciones y avances*. Anaya Multimedia.  
Ofrece una recopilación detallada y más técnica que la de este libro sobre las herramientas de IA y los avances prácticos en diversos sectores de aplicación.
- Pickover, C. A. (2025). *Inteligencia artificial: Una visión cronológica* (edición actualizada). Ediciones El Libro.
- Rodríguez Muiños, C. y Sánchez Lasheras, F. (2022). *Modelos de Inteligencia Artificial*. Marcombo.  
Detalla los modelos matemáticos adaptados a un módulo profesional de IA para la detección de patrones en grandes volúmenes de datos.

- Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta Editorial.  
Esta obra ofrece una introducción práctica y directa, ideal para entender el impacto inmediato de la tecnología en nuestra vida diaria y profesional.
- Salazar García, I., y Benjamins, V. R. (2020). *El mito del algoritmo: Cuentos y cuentas de la Inteligencia Artificial*. Anaya Multimedia.  
Un texto adecuado para desmitificar la idea de que las máquinas piensan, explicando con claridad qué pueden y qué no pueden hacer realmente los algoritmos.
- Sigman, M., y Bilinkis, S. (2023). *Artificial: La nueva inteligencia y el contorno de lo humano*. Debate/Penguin Random House.  
Explora la relación entre la nueva inteligencia y la esencia humana, analizando cómo redefine nuestros límites y el contorno de lo que nos hace únicos.
- Simon J.D. Prince, S. (2024). *Understanding Deep Learning*. The MIT Press.  
Una explicación técnica pero excelente de los fundamentos estadísticos y de machine learning para la IA.
- Suleyman, M. y Bhaskar, M. (2024). *La ola que viene: Tecnología, poder y el gran dilema del siglo XXI*. Debate, Barcelona.  
Analiza los riesgos y dilemas de poder y control que plantea la IA, presentándola como la tecnología más transformadora y disruptiva del siglo XXI.
- Tegmark, M. (2018). *Vida 3.0: Ser humano en la era de la inteligencia artificial*. Taurus, Barcelona.  
Una reflexión profunda sobre cómo la inteligencia artificial afectará la definición misma del ser humano y nuestro futuro colectivo.
- Torres, J. (2023). *La inteligencia artificial explicada a los humanos*. Plataforma Editorial.  
Una lectura diseñada para explicar conceptos complejos de forma sencilla pero rigurosa, adaptada específicamente para el público general y escrita por un reconocido investigador español sobre IA.

## Referencias del capítulo

BENDER, E. M., GEBRU, T., McMILLAN-MAJOR, A., and SHMITCHELL, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

BOLÓN-CANEDO, V., MORÁN-FERNÁNDEZ, L., CANCELA, B., and ALONSO-BETANZOS, A. (2024). A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, 599, 128096.

DA EMPOLI, G. (2020). *Los ingenieros del caos*. Anaya.

GEBRU, T., & OTHERS (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92.

LAZER, D., KENNEDY, R., KING, G., and VESPIGNANI, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, 343(6176), 1203-1205.

LECUN, Y. (2019). *On the measure of intelligence*. arXiv preprint arXiv:1911.01547.

O'NEIL, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.

OBERMEYER, Z., POWERS, B., VOGELI, C., and MULLAINATHAN, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.

NG, A. (2017). *AI is the new electricity*. Stanford University/Baidu Talk.

RUSSELL, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

SIMON, H. A. (1960). *The New Science of Management Decision*. Harper & Brothers.

WEIZENBAUM, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman and Company.

WIENER, N. (1950). *The Human Use of Human Beings: Cybernetics and Society*. Houghton Mifflin.



## Sobre el autor



### **Daniel Peña**

Daniel Peña es Profesor emérito de Estadística en la Universidad Carlos III de Madrid, Presidente de la Comisión de Inteligencia Artificial de la Real Academia de Ciencias Exactas, Físicas y Naturales de España y Coordinador del área de Inteligencia Artificial y Big Data en Funcas.

Ha sido catedrático de Estadística en las universidades Politécnica de Madrid, Wisconsin-Madison, Chicago y Carlos III de Madrid, donde fue miembro de la Comisión Gestora, Director del departamento de Estadística y Rector. Su investigación se ha centrado en la predicción con series temporales y el análisis de grandes masas de datos y es Fellow de la American Statistical Association y The Institute of Mathematical Statistics. Ha recibido varios premios por su labor de investigación, como el Primer Premio Nacional de Estadística, El Premio Jaime I en ciencias económicas, el Youden Prize al mejor artículo publicado en *Technometrics* y la Medalla de la Sociedad Española de Estadística e Investigación Operativa.



Funcas  
Caballero de Gracia, 28  
28013 Madrid  
Teléfono: 91 596 54 81

[publica@funcas.es](mailto:publica@funcas.es)  
[www.funcas.es](http://www.funcas.es)

ISBN 979-13-87770-21-1

