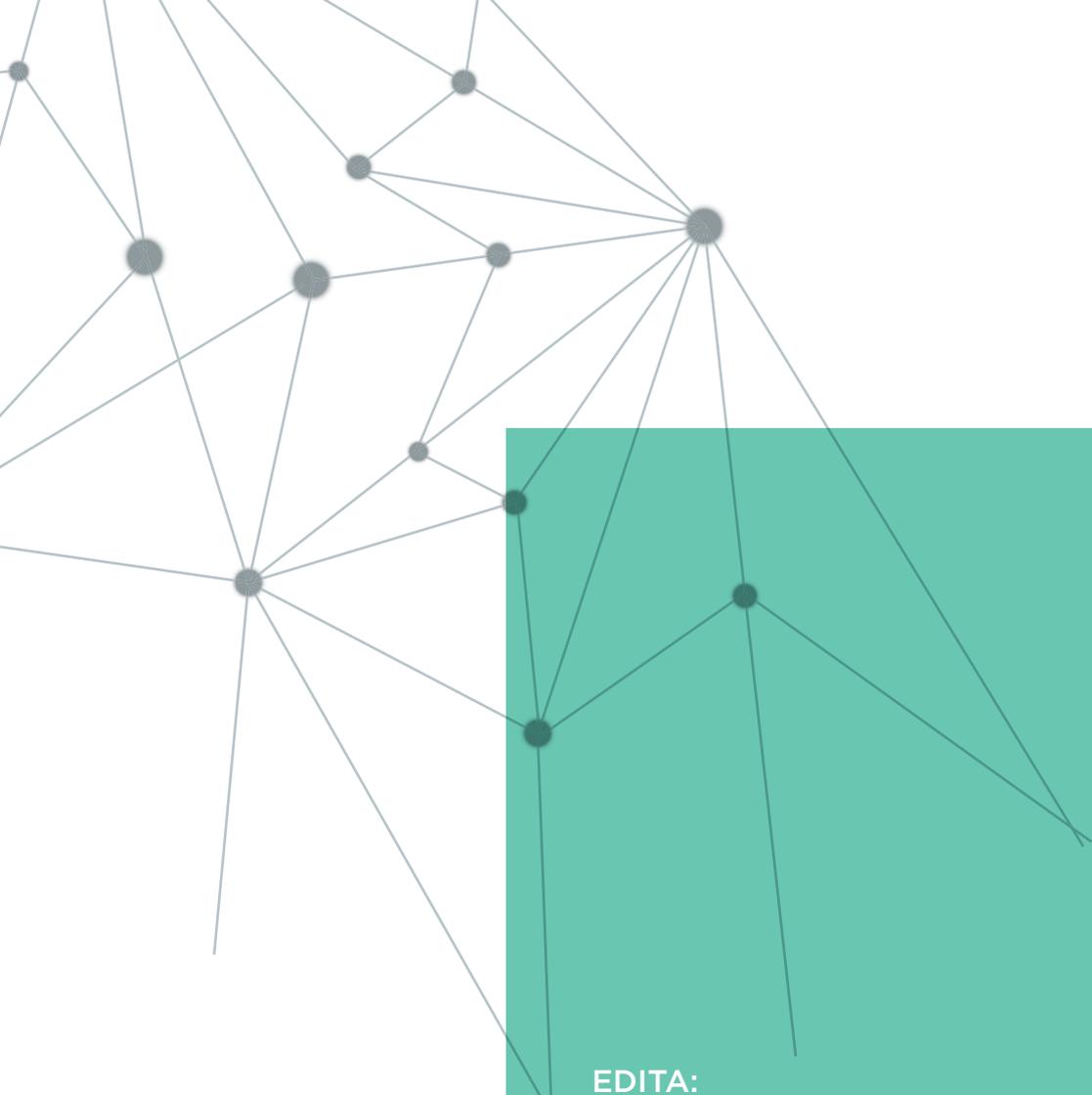




PARC CIENTÍFIC  
UNIVERSITAT DE VALÈNCIA

# Business Artificial Intelligence





**EDITA:**

Fundació Parc Científic Universitat de València  
c/ Catedrático Agustín Escardino, 9  
46980 Paterna (España)  
Telf: +34 963 543 841  
Kristin Suleng Furió  
Correo electrónico: comunicacion.pcu@uv.es

**COORDINADORES:**

Emilio Soria Olivas  
Mariano Serra Bondía

**AÑO:** 2021

**ISBN:**

978-84-09-36337-7

**DOI:**

10.7203/PCUV-9



UNIÓN EUROPEA

Fondo Europeo de  
Desarrollo Regional

Una manera de hacer Europa

*Actuación cofinanciada por la Unión Europea  
a través del Programa Operativo del  
Fondo Europeo de Desarrollo Regional (FEDER)  
de la Comunitat Valenciana 2014 - 2020*



**GENERALITAT  
VALENCIANA**



**AVI** AGÈNCIA VALENCIANA  
DE LA INNOVACIÓ

# PREÁMBULO

## BAI



Datos, datos y más datos... eso es lo que tenemos actualmente. Nuestros medios de entretenimiento generan datos (toda nuestra actividad queda grabada en las plataformas de entretenimiento como HBO o Netflix, por poner un par de ejemplos); nuestros medios de compra generan datos desde el momento de la compra (Amazon tiene su posición privilegiada gracias a esos datos) hasta el momento del pago (los bancos disponen de mucha información de nuestra actividad). A lo anterior hay que sumar los datos generados por nuestra actividad diaria (redes sociales, dispositivos vestibles como los sensores que marcan nuestro entrenamiento), los datos generados por los smartphones (y aquí abrimos el melón de los datos no estructurados como son las imágenes, vídeos, audios y texto), los datos generados por los dispositivos (el famoso Internet de las Cosas), entre otros. Si nos referimos a cantidades, y por dar solo unos números, existe una gran cantidad de infografías que muestran lo que ocurre en un minuto en Internet: en el año 2021, por minuto se tenían 200000 tweets, dos millones de visualizaciones en Twitch, se envían 197 millones de correos electrónicos y 69 millones de mensajes de texto, mientras que 1.6 millones de dólares se gastan en compras online.... Las cifras abruma. Con esta situación, se ha acuñado un término que ha ganado mucha publicidad en los últimos años, el Big Data.

Este término se usa para aquellos datos que tienen las siguientes 3 V:

- 1. VOLUMEN:** Se refiere al tamaño de los datos considerados de manera que la cantidad de datos supone un desafío importante para los sistemas de almacenamiento y procesamiento. Aunque no hay una distinción específica sobre el volumen de datos, normalmente puede variar desde los Terabytes (10<sup>12</sup> bytes) hasta los exabytes (10<sup>18</sup> bytes).
- 2. VELOCIDAD:** Los datos se generan a un ritmo muy rápido. La alta velocidad de los datos puede evaluarse por el hecho de que una gran proporción de los datos que se utilizan actualmente pertenecen al pasado reciente.
- 3. VARIEDAD:** Los datos considerados podrían obtenerse de numerosas fuentes como registros web, dispositivos del Internet de las Cosas (IoT), URLs, tweets de usuarios y patrones de búsqueda, etc. Asimismo, los datos pueden tener diferentes formatos, como valores separados por comas (CSV), tablas, documentos de texto y gráficos. Además, podrían ser estructurados o no estructurados.

Los humanos no podemos pensar que podemos estar en todos los procesos de generación y procesamiento de los datos. Según el famoso economista y premio Nobel Herbert Simon, existen dos tendencias: la riqueza de información que actualmente tenemos y que conlleva una pobreza de atención por nuestra parte. Nuestra capacidad para concentrarse en información de diferentes fuentes al mismo tiempo está destinada a fallar. Ahí están todos los estudios de la Economía del Comportamiento que demuestran nuestros sesgos e incapacidades para manejar grandes cantidades de información. Por todo ello, se hace necesario plantear sistemas automáticos para procesar la información que actualmente tenemos disponible.

Hasta ahora, para hacer este tipo de análisis se tenían las técnicas conocidas como *Business Intelligence (BI)*, que ayudaban a la toma de decisiones basándose en los datos analizados por los especialistas de las empresas. Estas técnicas presentan en la actualidad una serie de problemas derivados de los comentados anteriormente:

**ANÁLISIS DATOS ESTRUCTURADOS Y NO ESTRUCTURADOS.** Las plataformas tradicionales de BI se han centrado en gran medida en los datos estructurados (datos numéricos), pero hoy los usuarios requieren la capacidad de ver y analizar todo tipo de datos y en tiempo real. Ahora, los usuarios avanzados de herramientas de

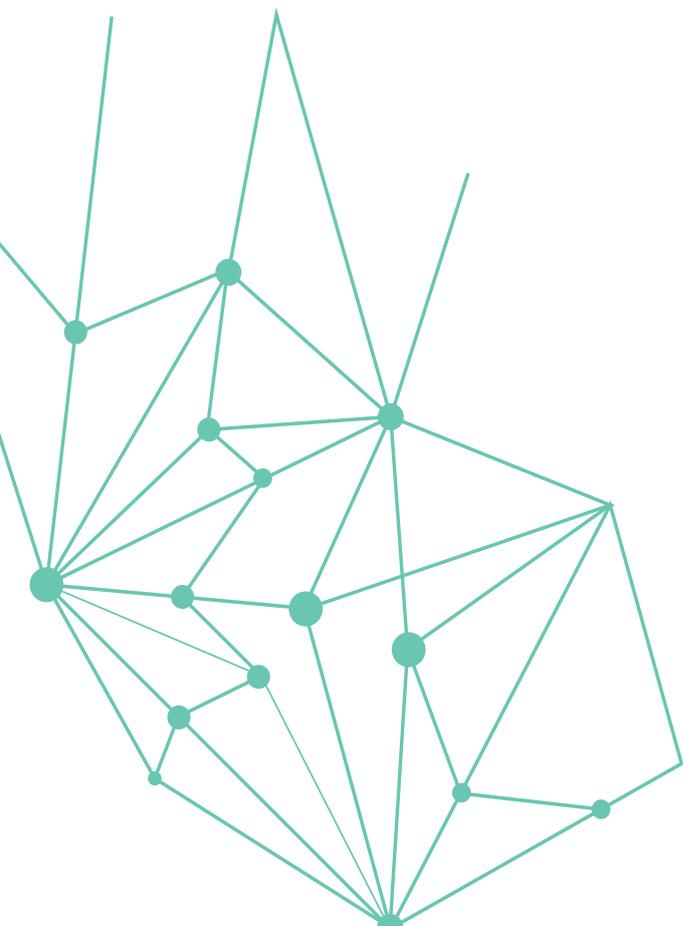


negocio de tipo BI necesitan acceder a todo tipo de información para tomar mejores decisiones. Según informes de empresas especializadas, en la actualidad las empresas obtienen un beneficio del análisis de estos datos no estructurados de hasta el 80%.

**NECESIDAD DE ANÁLISIS PREDICTIVOS.** Las herramientas tradicionales de BI suelen realizar análisis descriptivos de la información explicando el pasado, pero no estableciendo predicciones que pueden ayudar a gestionar de forma óptima las organizaciones y los negocios. Con estos modelos predictivos, las empresas pueden utilizar patrones y previsiones para obtener los siguientes pasos accionables utilizando sus datos.

**VISUALIZACIONES AVANZADAS.** La investigación en minería de datos visual ha crecido de forma exponencial en los últimos años. Nuevas formas de representación de todo tipo de datos han aparecido en los últimos años lo que ha impulsado la creación de nuevos tipos de cuadros de mando que son claves en el mundo del BI.

**FALTA DE PERSONALIZACIÓN.** Actualmente, los sistemas de BI funcionan con datos agregados unificando comportamientos y estableciendo conclusiones sobre valores promedios, pero se pierde el valor de los detalles que puede conducir al éxito a un determinado producto.



Las técnicas de Inteligencia Artificial (IA) actuales solucionan todos estos problemas. La IA ya es mejor que los humanos a la hora de identificar tendencias y extraer información de datos complejos. Estas capacidades no harán más que mejorar con el tiempo, por lo que se prevé una automatización de las tareas en un futuro próximo que permitan a los humanos centrarse en decisiones más complejas. Hay que destacar que estas tecnologías se han democratizado de tal manera que, donde antes existía una fuerte división entre dos grupos de usuarios (los que están lo suficientemente familiarizados con la ciencia de los datos como para obtener valor de los mismos, y los que no lo están) las técnicas actuales de IA salvan esa brecha ofreciendo conocimientos en un formato digerible para todos los usuarios, sean expertos o no. Entre los diferentes usos que puede tener la IA dentro de la inteligencia de negocio, cabe enumerar los siguientes:

- Predecir operaciones comerciales o previsiones de ventas basadas en los datos de los clientes y en los patrones de compra que puedan estimar con precisión los volúmenes de ventas, la gestión de inventarios y la demanda de productos, optimizando todo el conocimiento de negocio.
- Proporcionar a los compradores-online una experiencia personalizada basada en los conocimientos de datos impulsados por la IA derivados del comportamiento de los clientes anteriores; estos sistemas se conocen como recomendadores (no sólo de productos, sino que pueden gestionar la experiencia de usuario en la web).
- Optimizar la comunicación con los clientes usando *chatbots*, marketing automatizado por correo electrónico y mensajes en redes sociales que antes necesitaban agentes humanos y que ahora se pueden obtener por modelos generativos de lenguaje.
- Proporcionar ayuda en tiempo real a los clientes, donde las empresas pueden utilizar aplicaciones móviles movidas por la IA para comunicarse con un gran número de clientes en tiempo real o para proporcionar servicios personalizados a clientes individuales.
- Gestionar los crecientes volúmenes de datos no estructurados procedentes de diversas fuentes (incluso se pueden captar datos en Internet mediante el uso de técnicas de *web scraping*). Extracción de conocimiento de dichos datos aplicando técnicas que no se han usado previamente en BI como, por ejemplo, técnicas avanzadas de procesamiento de lenguaje natural.
- Obtener información, en tiempo real, de consumidores y proveedores, a partir de los datos de mercado que evolucionan rápidamente y que pueden ayudar a los directivos de las empresas a tomar decisiones clave en el momento adecuado.

El impacto económico que puede suponer el uso de las técnicas de IA en el campo de la BI es enorme además de la optimización continua de recursos que supone.

Esta monografía es un intento de avanzar por esa unión de los dos campos. En el primer capítulo, haremos un repaso a la unión de las dos técnicas y lo que supone este hecho; en el segundo, se realizará una revisión a los cuadros de mando, herramienta clásica en BI y una aplicación realizada para la administración pública. El texto se cerrará con un concepto que se está generalizando en BI, el de plataformas. Se explicará el desarrollo, uso y las ventajas de este tipo de herramientas, en este caso en un problema clínico. Esperamos que este sea otro elemento para la unión de estos dos campos de conocimiento.

Dr. Emilio Soria Olivas  
IDAL, Intelligent Data Analysis Laboratory  
<https://idal.uv.es/>  
Escuela Técnica Superior de Ingeniería  
Universitat de València

# ÍNDICE

CAP 1

**BI conoce a la IA**

8

CAP 2

**Cuadros  
de mando  
avanzados para  
la administración  
pública**

31

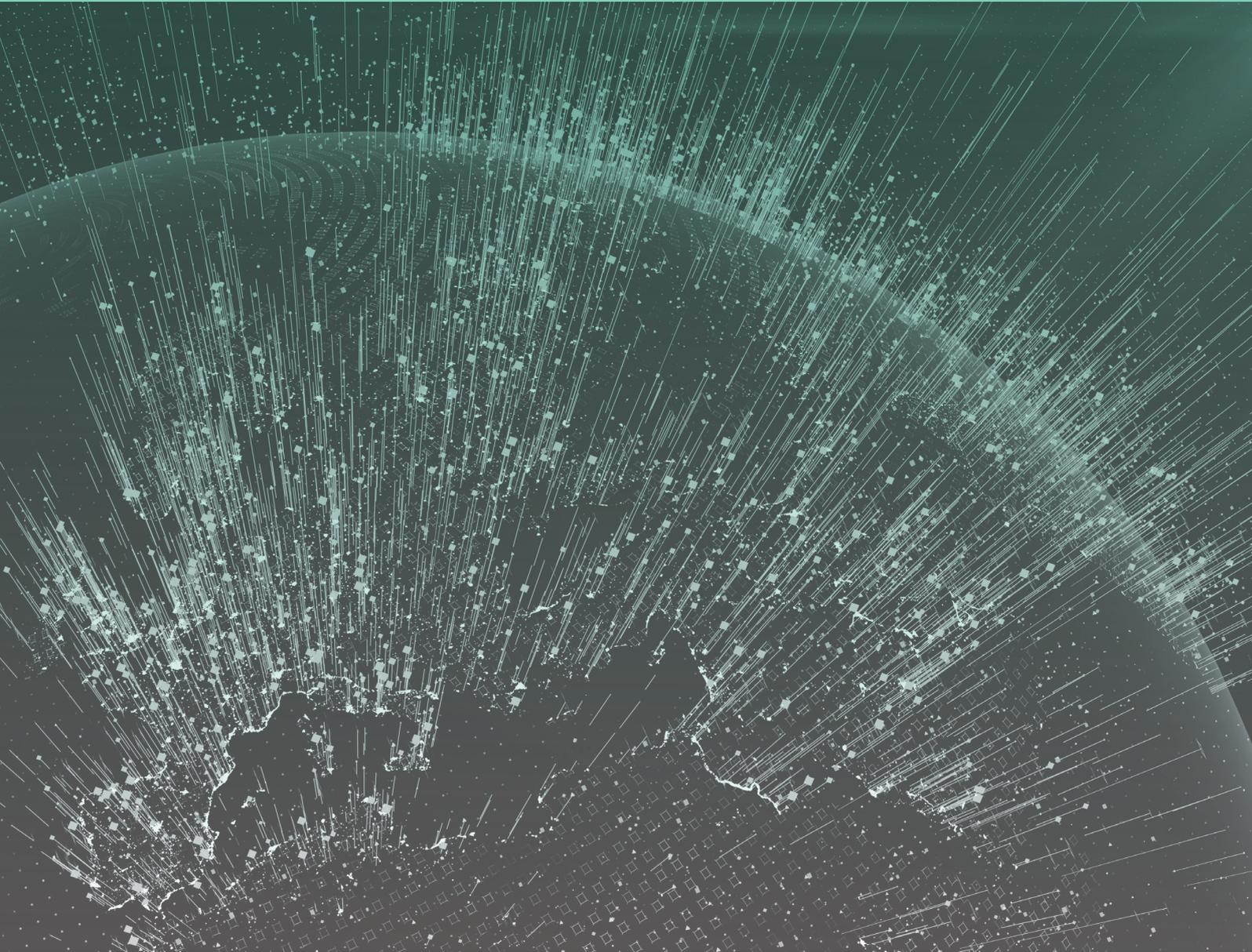
CAP 3

**Plataforma  
basada en  
IA para el  
diagnóstico  
avanzado en  
imagen médica**

65

# 1

## BI CONOCE A LA IA



Antonio José Serrano López  
Emilio Soria Olivas  
Fernando Mateo Jiménez

# 1

## BUSINESS INTELLIGENCE



En la actualidad, las empresas, tanto del sector público como del privado, se encuentran bajo una presión extraordinaria para ofrecer un rendimiento óptimo en todo momento, y por ello deben modernizarse y adaptarse a la situación del mercado. Para hacerlo, se requiere que una empresa sea capaz de tomar decisiones estratégicas, operativas y tácticas.

Estas decisiones, no obstante, implican analizar gran cantidad de datos, son complejas y se deben tomar rápidamente para que la empresa no pierda oportunidades de negocio. La calidad de los datos de partida para tomar estas decisiones es fundamental para tomar las decisiones correctas, como también es crucial la extracción adecuada de conocimiento a partir de los datos y su posterior análisis. Desde la década de 1950 existen métodos de ayuda a la decisión, que han ido cristalizando en lo que actualmente conocemos como inteligencia empresarial o *Business Intelligence* (BI).

En las últimas décadas, el contexto socioeconómico en el que se llevan a cabo las actividades económicas se ha conocido generalmente como la sociedad de la información y del conocimiento. Los profundos cambios que se han producido en los métodos de producción y en las relaciones económicas han llevado a un crecimiento en la importancia del intercambio de bienes intangibles, constituidos en su mayor parte por transferencias de información. La aceleración en el ritmo de los procesos de transformación actuales se debe a dos factores. El primero es la globalización, entendida como la interdependencia cada vez mayor entre las economías de los distintos países, que ha

Llevado al crecimiento de una única economía mundial caracterizada por un alto nivel de integración. El segundo son las nuevas tecnologías de la información, marcadas por la generalización del uso de Internet, de los smartphones, que han permitido transferencias de alta velocidad de grandes cantidades de datos y el uso generalizado de medios sofisticados de comunicación como las redes sociales.

En este escenario de rápida evolución, la gran cantidad de oportunidades de desarrollo empresarial no tiene precedentes. El fácil acceso a la información y al conocimiento ofrece varias ventajas para varios actores en el entorno socioeconómico: los individuos, que pueden obtener información más rápidamente, acceder a los servicios más fácilmente y llevar a cabo transacciones comerciales y bancarias en línea; las empresas, que pueden desarrollar productos y servicios innovadores que puedan satisfacer mejor las necesidades de los usuarios, y así lograr ventajas competitivas a partir de un uso más eficaz del conocimiento ganado; y, finalmente, la Administración pública, que puede mejorar los servicios proporcionados a los ciudadanos mediante el uso de aplicaciones de gobierno electrónico, como pagos en línea de contribuciones fiscales y herramientas de ciber salud, teniendo en cuenta el historial médico de cada paciente, mejorando así la calidad de los servicios sanitarios.

En cualquier organización, independientemente de su tamaño, las actividades comerciales incluyen la administración de grandes cantidades de datos de las condiciones comerciales internas y externas; todos estos datos relacionados con operaciones internas, publicidad, clientes, proveedores, activos económicos, etc., se acumulan históricamente y pueden ayudar a formar un conocimiento basado en la experiencia, para dar solución a problemas económicos complejos de vital importancia en la administración de la organización, mediante una toma de decisiones.



La definición precisa de BI, como ocurre con muchos otros conceptos relacionados con las nuevas tecnologías, es cambiante con el tiempo. En la literatura se pueden encontrar diferentes definiciones e interpretaciones del mismo concepto. Según Singh y Samalia, la BI constituye un marco que transforma la información en datos y después en aprendizaje, en consecuencia, mejorando el proceso básico de toma de decisiones de la empresa. Por otra parte, Nofal indica que la BI se caracteriza por ser un marco que recopila, cambia y muestra información organizada de diversas fuentes. La BI es un sistema y una respuesta que ayuda a los tomadores de decisiones a comprender las circunstancias económicas de la empresa.

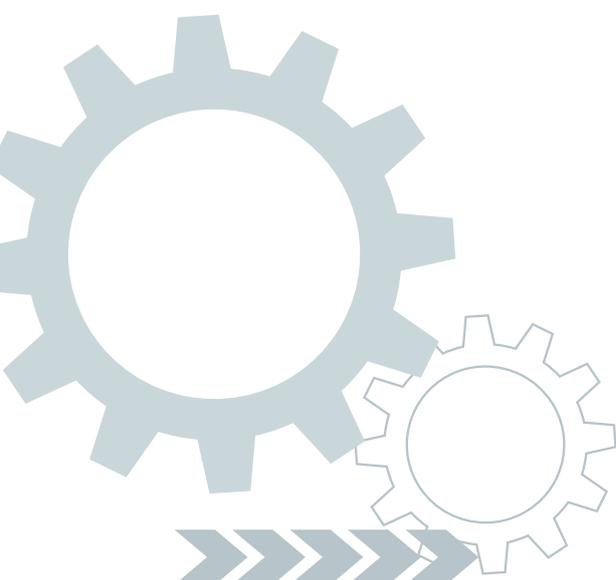
La BI también se describe como un conjunto de modelos numéricos y metodológicos de análisis, utilizados para extraer datos e información valiosa a partir de información en bruto. De manera similar, Wixom y Watson mencionan que la BI es una categoría amplia de tecnologías, aplicaciones y procesos para recopilar datos, almacenarlos, analizarlos y acceder a ellos para ayudar a sus usuarios a tomar mejores decisiones. Es posible actualizar los conocimientos proporcionados por las aplicaciones de BI, en particular mediante el uso de procedimientos de minería de información, para mejorar las predicciones y contribuir a una comprensión superior de las circunstancias económicas de cualquier organización.

La BI es un sistema y una respuesta que ayuda a los tomadores de decisiones a comprender las circunstancias económicas de la empresa.

---

En la literatura también se determina que la BI se asienta sobre tres pilares fundamentales: datos, información y conocimiento. Los datos, a su vez, se clasifican en tres categorías: estructurados, semiestructurados y no estructurados. La información se refiere al resultado de las actividades de extracción y procesamiento realizadas sobre los datos y, finalmente, el conocimiento está formado a partir de información que se utiliza para tomar decisiones y desarrollar las acciones correspondientes. Por tanto, podríamos decir que el conocimiento consiste en información de la que se hace uso en un dominio específico, y se ve reforzado por la experiencia y la competencia en la toma de decisiones.

- <sup>1</sup> Singh, H. y Samalia, H. V. (2014). "A business intelligence perspective for churn management". *Procedia-Social and Behavioral Sciences*, 109, 51-56.
- <sup>2</sup> Nofal, M. y Yusof, Z. (2013). "Integration of Business Intelligence and Enterprise Resource Planning within Organizations". *Technology*, Vol. 11, pp. 658-665.
- <sup>3</sup> Wixom, B. y Watson, H. 2010. *The BI-Based Organization*. *International Journal of Business Intelligence Research*, Vol. 1(1), pp. 12-24.
- <sup>4</sup> Vercellis, Carlo. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Politecnico di Milano, Italia; John Wiley & Sons Ltd.



Se puede representar la arquitectura de la BI mediante un diagrama genérico como el que se observa en la figura 1. En esta arquitectura se pueden distinguir varios componentes:

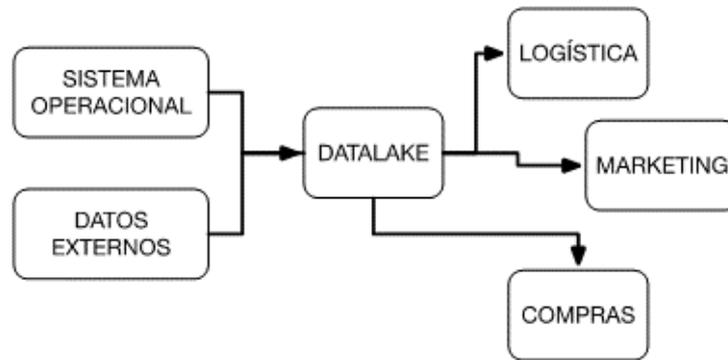


Figura 1. Arquitectura genérica de un sistema BI

**FUENTES DE DATOS.** En una primera etapa, es necesario recopilar e integrar los datos almacenados en las diversas fuentes primarias y secundarias, que son heterogéneas en origen y tipo. Las fuentes consisten en su mayor parte en datos pertenecientes a sistemas informáticos de la operativa habitual de la organización, pero también pueden incluir documentos no estructurados, como correos electrónicos y datos recibidos de proveedores externos. En términos generales, se requiere un gran esfuerzo para unificar e integrar las diferentes fuentes de datos.

### **ALMACENES DE DATOS (DATA-WAREHOUSES / DATA LAKES, ESTOS MÁS MODERNOS) Y MERCADOS DE DATOS (DATA-MARTS).**

Utilizando herramientas de extracción y transformación conocidas como Extract, Transform, Load (ETL), los datos que se originan en las diferentes fuentes se almacenan en bases de datos destinadas a apoyar los análisis de BI.

**METODOLOGÍAS DE BI.** Los datos se extraen finalmente y se utilizan para alimentar modelos matemáticos y métodos de análisis destinados a ayudar a la toma de decisiones en diferentes ámbitos. En un sistema de BI, se pueden implementar varias aplicaciones de apoyo a la toma de decisiones. Entre las posibles metodologías podríamos destacar: análisis de cubos multidimensionales, análisis exploratorio de datos, análisis de series temporales, modelos inductivos para minería de datos y modelos de optimización.

La implementación en software de los modelos de BI permite obtener información, analizarla, procesarla y reportarla a partir de bases de datos. El objetivo es utilizar el conocimiento extraído de esos datos extraídos para mejorar la eficiencia de una empresa, detectar posibles carencias y diferenciarse frente a la competencia.

Existen múltiples aplicaciones comerciales de diferentes desarrolladores para este propósito o englobadas en esta categoría y los principales tipos existentes son:

**GESTIÓN DE DATOS:** se centran en la obtención de la información, su procesamiento y exportación a otro sistema.

**DESCUBRIMIENTO:** su principal objetivo es la obtención, el análisis de nueva información y, mediante la aplicación de algoritmos, realizar predicciones o proyecciones.

**INFORMES:** permiten la creación de cuadros de mando para la representación de la información obtenida y procesada de una manera visual, organizada, clara y ajustada a las necesidades de la audiencia a la que va dirigida.

Las aplicaciones más utilizadas en la actualidad como herramientas BI son:

*Tableau (<https://www.tableau.com/>):* es una herramienta de visualización que permite acceder a los datos y analizarlos de manera simple y efectiva, crear cuadros de mando e informes, así como compartirlos de un modo sencillo.

*SAP BusinessObjects BI (<https://www.sap.com/products/bi-platform.html>):* plataforma integral de BI que posibilita obtener los datos, analizarlos y desarrollar modelos de aprendizaje automático. También es posible aplicar estos modelos para hacer predicciones y proyecciones de futuro, así como realizar cuadros de mando e informes.

*SAS BI (<https://www.stitchdata.com/analysis-tools/sas-bi/>):* herramienta especializada en la utilización de datos y métricas para, haciendo uso de modelos predictivos, asistir a la toma de decisiones de negocio.

MicroStrategy (<https://www.microstrategy.com/>): aplicación de BI cuyos puntos fuertes son el análisis de datos (por ejemplo, tendencias) y la definición y el desarrollo de cuadros de mando para la visualización de la información.

*QlikView (<https://www.qlik.com/us/products/qlikview>):* su punto fuerte es la búsqueda y el descubrimiento de datos de distintas fuentes, ofreciendo integración con multitud de aplicaciones como, por ejemplo, Salesforce, Teradata... Se trata de una plataforma personalizable, ya que cada usuario puede crear su propio espacio y procesar los datos de la manera que necesita.



*Microsoft Power BI (<https://powerbi.microsoft.com/>):* es una suite de herramientas de análisis y procesamiento de datos que provienen de diversos orígenes. Se trata de una plataforma potente para la visualización de la información y la creación de informes.

*Oracle BI (<https://www.oracle.com/business-analytics/business-intelligence/>):* software de BI cuyo punto fuerte es el descubrimiento de patrones y proporcionar una interfaz de visualización de datos sencilla e intuitiva. Ofrece las principales funcionalidades de las herramientas de BI, tales como cuadros de mando, inteligencia artificial, alertas...

# 2

## INTELIGENCIA ARTIFICIAL

*Artificial intelligence would be the ultimate version of Google. The ultimate search engine that would understand everything on the web. It would understand exactly what you wanted, and it would give you the right thing. We're nowhere near doing that now. However, we can get incrementally closer to that, and that is basically what we work on.*

El párrafo anterior es de Larry Page, cofundador de Google, lo enuncia en 2008 y responde muy bien a la situación actual en todos los campos del conocimiento. Si hay una palabra de moda actualmente es inteligencia artificial (IA), cualquier actividad humana (productiva, económica, social...) queda impactada por esta tecnología. Para entenderla vamos a hacer uso del siguiente gráfico, donde aparecen, en forma de subconjuntos, diferentes términos que es necesario conocer.



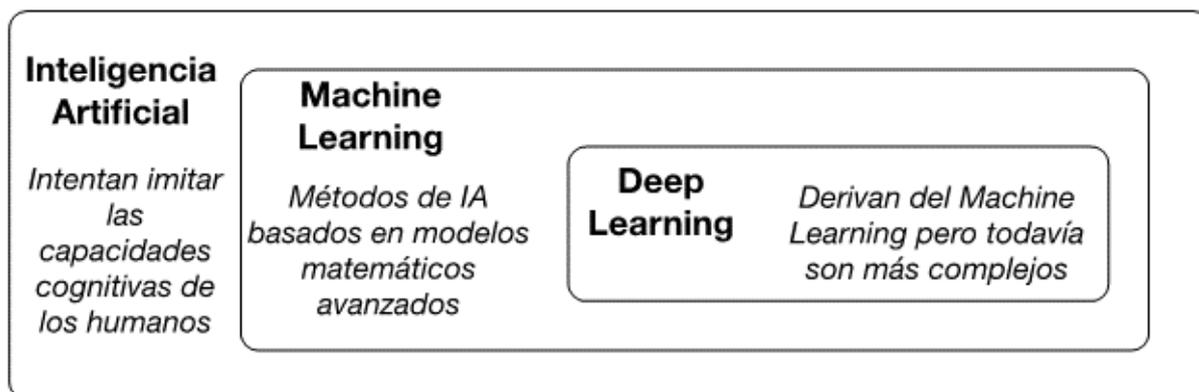


Figura 2. Esquema de la relación entre inteligencia artificial, Machine Learning (aprendizaje máquina) y Deep Learning (aprendizaje profundo).

El término inteligencia artificial nace de la mente de John McCarthy; en el verano de 1956 organiza una conferencia sobre este tema incipiente. Se trataba de una conferencia de dos meses de duración a la que se invitó a las principales figuras del nuevo campo emergente. Los objetivos eran ambiciosos y optimistas a la vez; la propuesta de la conferencia declaraba que *“se intentará encontrar la forma de hacer que las máquinas utilicen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados a los humanos y se mejoren a sí mismas”* y prometía que *“se podía lograr un avance significativo en uno o varios de estos problemas si un grupo de científicos cuidadosamente seleccionados trabajan juntos en él durante un verano”* (esto se conoce como exceso de optimismo). Desde sus comienzos hasta principios de los años 80, los enfoques seguidos en este campo son variados (reglas, agentes, lógica, etc.) y en esta década se empieza a plantear otra estrategia. Esta consiste en considerar todo problema de IA como un ajuste de un modelo matemático avanzado entre dos conjuntos de datos. Aquí aparece lo que se conoce como aprendizaje máquina (*Machine Learning*), un subcampo de la IA dedicado a los algoritmos que aprenden de los datos. Este aprendizaje puede dividirse en varias categorías. Por ejemplo, en el aprendizaje supervisado, se presenta a un algoritmo un conjunto de ejemplos previos etiquetados y tiene la ventaja de identificar asociaciones entre los datos y el resultado etiquetado, o la clasificación.

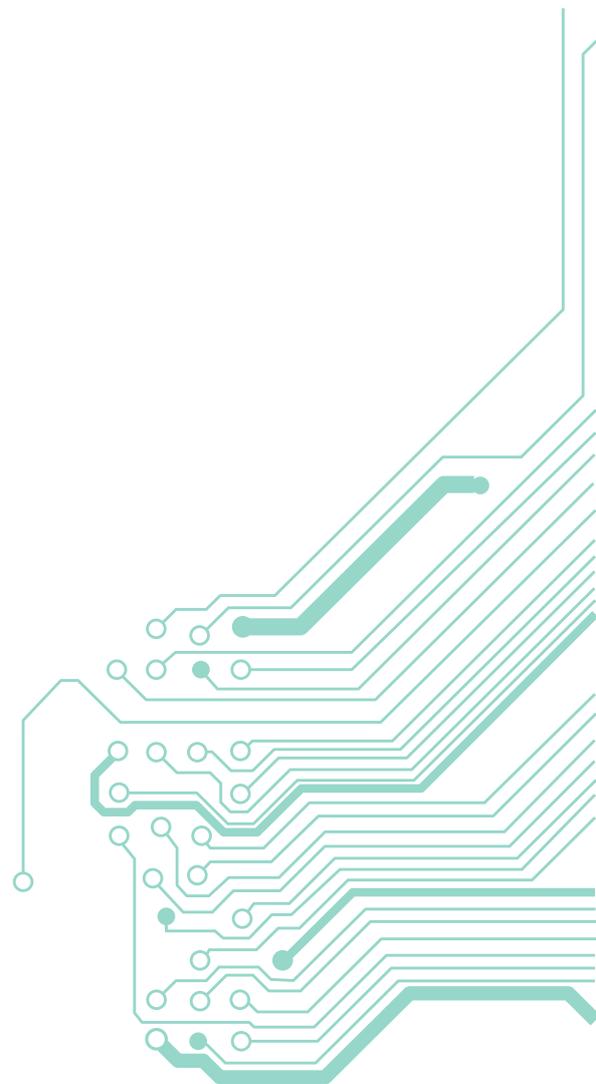
En el aprendizaje “no supervisado” no se dispone de conjuntos previos de ejemplos etiquetados, sino de datos sin etiquetar o sin clasificar. A modo de ejemplo el aprendizaje supervisado tiene como objetivo diferenciar las futuras imágenes de gatos y perros después de recibir muchas imágenes de gatos y perros, cada una etiquetada como “gato” o “perro”. El aprendizaje no supervisado, por el contrario, tiene como objetivo destacar las diferencias entre una gran colección de fotos de gatos y perros, sin esas etiquetas. Así pues, las técnicas de aprendizaje máquina tratan de utilizar las características o atributos de cada ejemplo para llegar a la etiqueta o clasificación correcta; por ejemplo, ¿qué características clave pueden utilizarse para distinguir una foto de un gato de la de un perro? A medida que se proporcionan más ejemplos de gatos y perros, los algoritmos de aprendizaje máquina ajustan los modelos que se alimentan de las características definidas por los expertos humanos.

El tercer subconjunto, dentro de los sistemas de aprendizaje máquina, se conoce como aprendizaje profundo (Deep Learning) y se sigue el mismo esquema que en el anterior caso; hay que ajustar unos modelos matemáticos extremadamente complejos a los conjuntos de datos que se tienen. Las diferencias que se tienen con el caso del aprendizaje máquina son la complejidad de los modelos (el número de parámetros crece en varios órdenes de magnitud) y que no necesita que ningún experto seleccione las características de entrada a los modelos. Esto conlleva la necesidad de usar un gran número de datos para ajustar estos modelos. Actualmente todo lo que se conoce como inteligencia artificial son sistemas de aprendizaje profundo.

Esta inteligencia artificial está cambiando nuestra forma de relacionarnos, comprar, viajar... Todo, absolutamente todo, está siendo modificado por estas tecnologías. No existe ningún campo del conocimiento donde esta tecnología no se aplique actualmente.

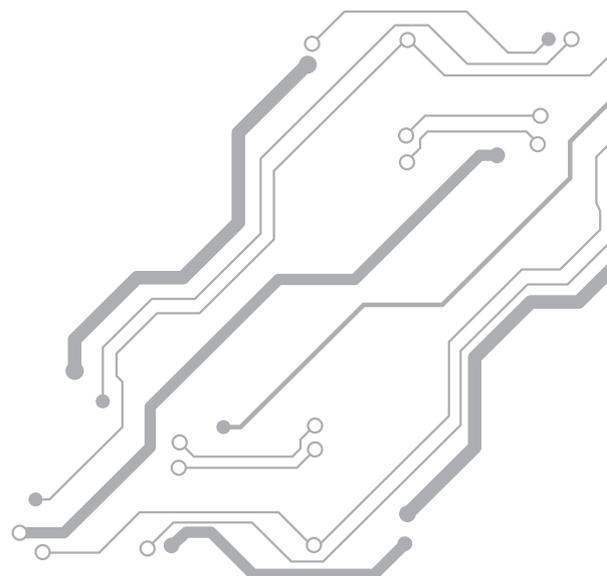
A esta situación se ha llegado por cuatro factores:

- 1 Tenemos un aumento sin precedentes en la cantidad de datos que manejamos, y la necesidad de conservarlos cuidadosamente y basar las decisiones en ellos. Es difícil decir exactamente cuántos datos almacenamos a diario. He aquí algunas cifras presentadas por IBM en 2018: a) 80.000.000 de resonancias magnéticas realizadas cada año; b) 600.000.000 de formas de malware



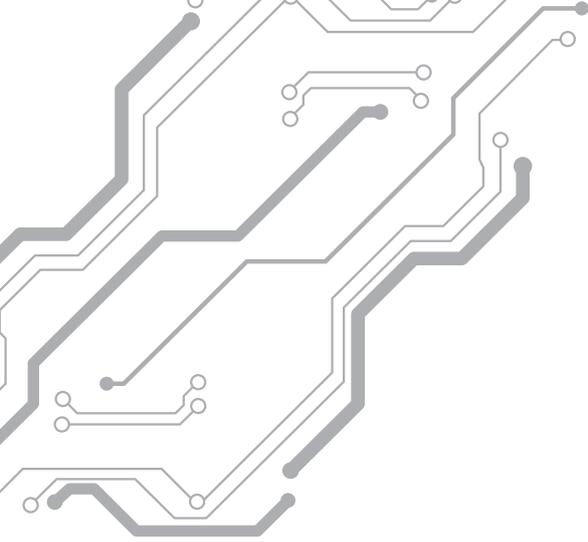
introducidas diariamente; c) 2.200.000.000 de lugares que generan previsiones meteorológicas cada 15 minutos. Otras obtenidas en 2019 (todas ellas durante un día) son: a) más de 115.000.000.000 de correos electrónicos enviados; b) más de 35.000.000 de fotos subidas a Instagram y c) más de 3.000.000.000 de búsquedas en Google. Existe una gran cantidad de infografías que muestran lo que ocurre en un minuto en Internet; en el año 2021 (por minuto) se tenían 200.000 tweets, 2 millones de visualizaciones en Twitch, se envían 197 millones de correos electrónicos, 69 millones de mensajes de texto, 1,6 millones de dólares que se gastan en compras online..., se tienen cifras que abruman. Además, hay que tener en cuenta que el concepto de dato ha cambiado. El dato concebido como número con un tamaño fijo ha cambiado; ese tipo de dato conocido como estructurado ha dado paso al no estructurado; aquí se engloban los datos tipo texto (correos electrónicos, comentarios en Tripadvisor, etc.); imágenes (que actualmente son el dato predominante en determinadas redes, por ejemplo, Instagram) o vídeos (cabe destacar el actual crecimiento de Twitch). Según una investigación realizada por Raconteur, es probable que alcancemos los 463 exabytes de datos producidos diariamente para el año 2025. Para que el lector se sitúe, un exabyte equivale a 20 veces todos los libros escritos hasta el año 2013. Este aumento de datos generados va a ser impulsado en gran medida por el aumento de los dispositivos que se conectan a la Internet de las cosas (IoT) en dispositivos vestibles (wearables), dispositivos inteligentes en el hogar y en la industria. Otra fuente clave de generación/almacenamiento de datos han sido los teléfonos inteligentes. Por ejemplo, las fotos tomadas por estos dispositivos, en tiempo real, pueden descargarse o subirse a Google Drive o a cualquier sitio web de redes sociales. Las aplicaciones inteligentes como el GPS y la navegación y el almacenamiento en la nube han hecho la vida de los seres humanos más fácil y cómoda. Servicios como iCloud y Google Drive

**Existe una gran cantidad de infografías que muestran lo que ocurre en un minuto en Internet; en el año 2021 (por minuto) se tenían 200.000 tweets, 2 millones de visualizaciones en Twitch, se envían 197 millones de correos electrónicos, 69 millones de mensajes de texto, 1,6 millones de dólares que se gastan en compras online..., se tienen cifras que abruman.**



utilizan la tecnología IoT para el almacenamiento, por lo que no es un problema guardar estos datos. La oportunidad de la IA consiste, simplemente, en que los humanos no pueden esperar, de ninguna manera, analizar ni siquiera una pequeña fracción de estos datos sin automatización. Aplicar el análisis de datos avanzado a gran escala es nuestra única esperanza. Yendo un paso más allá, también podemos esperar que, a menos que incorporemos la toma de decisiones automatizada en los sistemas, no podremos hacer frente al crecimiento de la demanda de toma de decisiones y de acción que el mundo actual demanda. De hecho, si pensamos en la forma en que procesamos los datos en nuestra mente (que es analógica) hay trabajos que demuestran que el consumo de información por parte de los seres humanos va por detrás de la velocidad a la que se crea la información mediante procesos digitales. Se ha descubierto que el consumo de información de la mente en bytes solo aumentó un 5,4% al año del periodo 1980 a 2008. Por todo esto, lo que está cobrando especial importancia son los datos que fluyen de máquina a máquina: datos creados para ser consumidos por otros dispositivos o algoritmos, no para ser consumidos por los humanos. Todos los dispositivos se interconectarán a través de los datos. Según el famoso economista y premio Nobel Herbert Simon, hay dos tendencias; por una parte, la riqueza de información que actualmente tenemos y que conlleva una pobreza de atención por nuestra parte. Nuestra atención y capacidad para concentrarse en información de diferentes fuentes al mismo tiempo está destinada a fallar, por lo que se hace necesario plantear sistemas automáticos para procesar la información que, actualmente, tenemos disponible.





**Evidencia económica de lo comentado es el incremento de valor de la empresa NVIDIA (uno de los principales fabricantes de GPU); entre enero de 2012 y enero de 2020 las acciones de NVIDIA se dispararon más de un 1.500%.**

- 2** El segundo factor es la potencia de procesamiento, aquí se ha tenido la famosa ley de Moore, que ha regido la evolución de los sistemas electrónicos durante los últimos 50 años. Esta ley, definida por Gordon Moore, actual presidente emérito de Intel, refleja que la capacidad de procesamiento de los ordenadores (el número de transistores en un circuito integrado precisamente) se duplica aproximadamente cada dos años. Esto corresponde a una tasa de crecimiento anual de aproximadamente el 30% en la capacidad de procesamiento, lo que conlleva que la capacidad de procesamiento de datos aumenta exponencialmente en el tiempo. Esta ley se ha mantenido durante unos 40 años y sirve como indicador del ritmo de avance y progreso en la industria informática. En este incremento de capacidades de procesamiento hay que destacar la aparición de las unidades de procesamiento gráfico, o GPU, diseñadas originalmente para acelerar los cálculos necesarios para representar gráficos de alta resolución. A partir de la década de 1990, estas unidades fueron especialmente importantes en las consolas de videojuegos de gama alta. Las GPU están optimizadas para realizar rápidamente un gran número de cálculos en paralelo. Una vez que los investigadores descubrieron que los cálculos requeridos por las aplicaciones de IA eran, en general, similares a los necesarios para optimizar gráficos, empezaron a recurrir en masa a las GPU, que rápidamente se convirtieron en la principal plataforma de hardware para la inteligencia artificial. Dado que es una cuestión clave la de la computación, la investigación en nuevos dispositivos no ha parado. Han aparecido un gran número de empresas que están planteando el diseño de los chips de una forma radicalmente nueva. Un nuevo tipo de chip se acerca mucho más a la imitación del cerebro, prescindiendo en gran medida de la capa de software que consume recursos e implementando los sistemas neuronales en el hardware. Estos diseños emergentes de chips “neuromórficos” implementan versiones de hardware de las neuronas directamente en el silicio. Evidencia económica de lo comentado es el incremento de valor de la empresa NVIDIA (uno de los principales fabricantes de GPU); entre enero de 2012 y enero de 2020 las acciones de NVIDIA se dispararon más de un 1.500%.

**3** El tercer factor es la computación en la nube (o cloud computing). La actual industria de la computación en nube comenzó en 2006 con el lanzamiento de Amazon Web Services, o AWS. La estrategia de Amazon consistía en aprovechar su experiencia en la construcción y la gestión de los enormes centros de datos que alimentaban su servicio de compras en línea, vendiendo a una amplia gama de clientes un acceso flexible a recursos informáticos alojados en instalaciones similares. Si nos fijamos en Amazon, su crecimiento ha sido espectacular; en 2016, AWS crecía tan rápido que los nuevos recursos informáticos que Amazon tenía que añadir a su sistema cada día equivalían aproximadamente a todo lo que la empresa tenía a finales de 2005. En 2018, operaba más de cien centros de datos ubicados en nueve países diferentes en todo el mundo. En 2019, los ingresos de AWS crecieron un treinta y siete por ciento, hasta los 8.200 millones de dólares, y el servicio en la nube representó alrededor del trece por ciento de los ingresos totales de la empresa. Si analizamos a todos los proveedores, según un estudio reciente, un noventa y cuatro por ciento de las organizaciones, desde corporaciones multinacionales hasta pequeñas y medianas empresas, utilizan ahora la computación en la nube. La computación en la nube está haciendo que tanto el almacenamiento como la capacidad de procesamiento estén ampliamente disponibles para todo el mundo, de modo que cualquiera puede acceder y hacer uso de complejas herramientas de IA a un precio que podríamos calificar de irrisorio. Hay que tener en cuenta que antes de la llegada de los proveedores de la nube, las empresas y las organizaciones necesitaban comprar y mantener sus propios servidores informáticos y software y emplear un equipo de especialistas para mantener y actualizar continuamente los sistemas. A partir de 2019 se tiene un nivel de acceso sin precedentes a servicios de aprendizaje máquina sofisticados a través de interfaces de programación de aplicaciones simplificadas en la nube. Amazon, Microsoft, Google e IBM ofrecen acceso a sus herramientas, y permiten a las organizaciones cargar datos, entrenar modelos y desplegar soluciones dentro de sus aplicaciones. La amplia disponibilidad de estas tecnologías como soluciones en la nube no solo reduce el nivel de conocimientos necesarios para implementar una aplicación impulsada por la IA, sino que además reduce el tiempo que se tarda en pasar de la idea al prototipo y al despliegue en lo que a producción se refiere. Además, se ofrecen modelos de IA ya desarrollados; se tienen, por ejemplo, paquetes para el reconocimiento del habla y el procesamiento del lenguaje natural, y motores de recomendación que pueden hacer sugerencias a los compradores de sitios online.

**Amazon, Microsoft, Google e IBM ofrecen acceso a sus herramientas, y permiten a las organizaciones cargar datos, entrenar modelos y desplegar soluciones dentro de sus aplicaciones.**

**4** El cuarto factor son las herramientas de código abierto, que están haciendo lo mismo en cuanto a las capacidades de nivel inferior. La IA se está democratizando de una forma que nunca antes había sido posible. En noviembre de 2015, TensorFlow (el marco de trabajo de Google para crear aplicaciones de aprendizaje automático) se publicó como una herramienta de código abierto. Desde entonces, el código abierto sobre herramientas de IA se ha disparado. Además, abundan los cursos en línea, tanto de pago como gratuitos, y la principal limitación para aprender sobre las técnicas de IA es la disponibilidad de tiempo de cada persona.

**Con estos cuatro factores, el crecimiento de estas técnicas ha sido espectacular.**

A día de hoy los sistemas de IA han superado a los humanos en un gran número de tareas; reconocer objetos en imágenes; reconocer sonidos, interpretar fotografías, etc. Cuando se han tenido datos de calidad y cantidad para entrenar un sistema de IA en una determinada tarea, siempre el sistema experto ha superado a los humanos. Eso es un hecho objetivo y no hay que entrar en controversias (¿es de verdad una inteligencia como la humana?, ¿nos llegarán a dominar?) que, actualmente, no conducen a nada y que limitan la potencia de estos sistemas. Tenemos que ser pragmáticos; un análisis de la consultora PwC comenta que la IA añadirá unos 15,7 billones de dólares a la economía mundial para el año 2030, y esto cobra más importancia cuando esperamos recuperarnos de la enorme crisis económica desatada por la pandemia del coronavirus. Quizás lo más importante es que la inteligencia artificial evolucionará hasta convertirse en una herramienta indispensable que será crucial para abordar los mayores retos a los que nos enfrentamos, como son el cambio climático y la degradación del medio ambiente, la (¿inevitable?) próxima pandemia, la escasez de bienes esenciales como son la energía y el agua dulce, la pobreza y la falta de acceso a la educación. Eso sí, esta tecnología conlleva sus peligros, que es necesario combatir y solucionar, pero no podemos permitirnos parar la investigación en este campo.

# 3

## BI2A. ¿UNA NUEVA DISCIPLINA?

Según el “cuadrante mágico”, de 2021, de la consultora Gartner, correspondiente a herramientas de *Business Intelligence* (BI), el mercado está gobernado por las propuestas Microsoft (con su PowerBI), Tableau y Qlik.

Su popularidad se basa en que permite crear visualizaciones de datos, también denominadas dashboard, para reflejar la situación de la empresa con gráficos interactivos y coloridos sin necesidad de disponer de grandes conocimientos de programación o de bases de datos. El mayor o menor éxito del producto final depende en gran medida de lo preparados que estén los datos de la empresa y de la habilidad en disponer y seleccionar unos gráficos frente a otros.

Esta realidad hace que, actualmente, las herramientas de BI se utilicen para lo que se conoce en el campo de la ciencia de datos como análisis descriptivo de los datos, y es la primera de las etapas de generación de conocimiento. Su objetivo consiste básicamente en estudiar el pasado con la información de la que se dispone y, en muchos casos, ni siquiera se describe la situación actual porque estas herramientas se nutren con información de hojas de cálculo o bases de datos que se actualizan diaria o semanalmente.

A pesar de que el conocimiento para la explotación de los datos es muy superior a lo que realmente se ha estado aplicando hasta el momento, la popularidad de términos como *big data* o inteligencia artificial ha propiciado que prácticamente todas las empresas busquen más valor para sus negocios a través de estas líneas de acción y están forzando que las herramientas de BI faciliten la aplicación de estas metodologías.

Hasta 2021, Microsoft es la empresa que mejor ha aprovechado su posición previa y la coyuntura actual para mantenerse como líder. A su popular herramienta PowerBI, le ha añadido componentes

de analítica aumentada, es decir, técnicas bien documentadas de preparación y exploración de datos, así como técnicas de analítica predictiva (*Machine Learning*) en forma de software como servicio (SaaS) dentro del sistema de computación en la nube de Microsoft denominado Azure.

La evolución de Microsoft ejemplifica muy bien los 4 factores que han permitido la revolución de la inteligencia artificial. Con su plataforma en la nube Azure, ofrece la posibilidad de trabajar con enormes cantidades de datos de manera distribuida y gran capacidad de cálculo, como otras grandes empresas tecnológicas como Amazon o Google, pero sorprendentemente la empresa que anteriormente había sido la defensora por excelencia del *software* privativo también permite complementar la programación con *software* libre basado en R o Python para aprovechar los grandes avances que se están consiguiendo con las herramientas de código abierto desarrolladas por la comunidad.

A medida que los científicos de datos comiencen a ocupar puestos dentro del ámbito de la *Business Intelligence* veremos la aplicación real de las técnicas de inteligencia

artificial. Aunque hay muchos profesionales que se están adaptando a la nueva realidad, los nuevos científicos de datos son idóneos para sacar el máximo rendimiento de las nuevas plataformas de BI porque están formados especialmente para extraer conocimiento y conocen en profundidad las etapas de la ciencia de datos, desde la gestión de todo tipo de datos hasta la aplicación de técnicas de IA y *machine learning* para el soporte a la toma de decisiones. Esta preparación les permite aprovechar las posibilidades más avanzadas de las actuales herramientas de BI, y de las que todavía no están disponibles, y ello da lugar a la nueva disciplina BI2A o el avance de *Business Intelligence* (BI) con el uso aplicado de la analítica de datos (A) y las técnicas de inteligencia artificial.

Y lo mejor está por llegar. Las técnicas más consolidadas de IA, que entran dentro de la denominación de análisis predictivo, se basan en modelos matemáticos para estimar (o predecir) información de la que todavía no disponemos. Por ejemplo, estimar el precio del oro la próxima semana, anticiparse a saber en dónde se estropeará una máquina o determinar el perfil de personalidad de un usuario de móvil.



<sup>5</sup> Titulación de grado ofertada por varias universidades españolas, incluyendo la Universitat de València: <https://www.uv.es/grado-ciencia-datos/es/grado-ciencia-datos.html>.

Las nuevas metodologías de IA que se están desarrollando actualmente van un paso más allá. Nos ayudarán a determinar qué acciones son más interesantes para lograr nuestro objetivo. El análisis prescriptivo no solo analiza predictivamente, sino que aprovecha la estimación para aventurarse en el futuro y evaluar la relevancia de cada acción futura. El aprendizaje reforzado, que tan sorprendentes resultados está proporcionando, como la victoria sobre los humanos en el complejo juego del go, está esperando aplicarse en también en *Business Intelligence*.



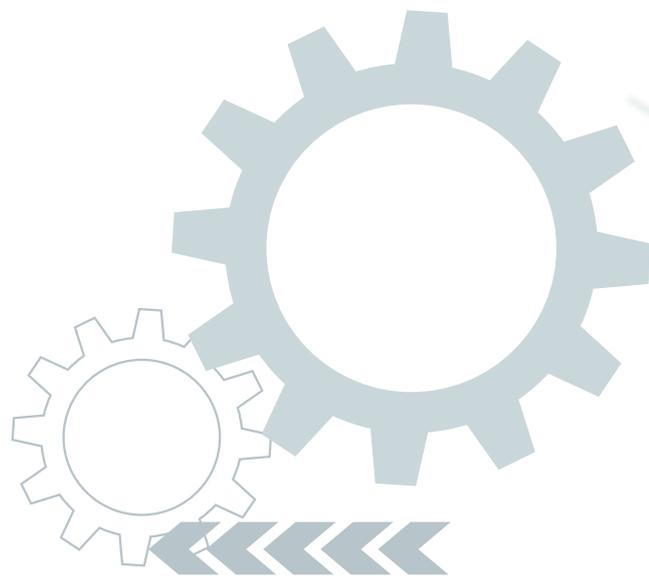
Figura 4. Relación entre la Business Intelligence clásica y la incorporación de analítica aumentada basada en las técnicas de inteligencia artificial entre las que se incluyen el análisis predictivo y prescriptivo.

La nueva disciplina de BI2A será la de la aplicación real de todas las etapas de ciencia de datos al campo de la *Business Intelligence* (figura 4). De manera simplificada, en las primeras etapas destacamos el análisis exploratorio, el procesamiento de los datos y el agrupamiento:

**ANÁLISIS EXPLORATORIO DE DATOS:** el propósito principal es resaltar las características relevantes de cada atributo contenido en un conjunto de datos, utilizando métodos gráficos y calculando estadísticas de resumen, e identificar la intensidad de las relaciones subyacentes entre los atributos. Ejemplos de análisis exploratorios incluyen: estadísticos univariados, gráficos de densidad/distribución estadística, medidas de correlación/covarianza, diagramas de dispersión, gráficos cuantil-cuantil (*q-q plots*), diagramas de caja (*box plots*), tablas de contingencia entre variables categóricas y proyecciones a espacios vectoriales de distinta dimensionalidad (*manifold learning*).

**PREPROCESAMIENTO DE LOS DATOS. INCLUYE VARIAS TÉCNICAS:** validación de datos, para identificar y eliminar anomalías e inconsistencias; integración y transformación de datos, para mejorar la precisión y la eficiencia de los algoritmos de aprendizaje; reducción y discretización del tamaño de los datos, para obtener un conjunto de datos con un número menor de atributos y registros, pero que sea tan informativo como el conjunto de datos original. Algunos ejemplos de técnicas de preprocesamiento incluyen: estandarización; filtrado; imputación de datos faltantes; remuestreo de los datos; técnicas de selección de características mediante métodos *filter*, que seleccionan los atributos en la fase previa a la creación de los modelos (por ejemplo, aquellos basados en la correlación con la variable

objetivo); métodos wrapper, que usan un modelo e iteran para llegar al conjunto de variables óptimo (*forward selection, backward elimination, forward/backward search*); y métodos *embedded*, que seleccionan los atributos en el proceso de creación del modelo de aprendizaje (por ejemplo, árboles de decisión). Otros métodos transforman las variables originales, por ejemplo, proyectándolas a las direcciones de máxima varianza (PCA).



**CLUSTERING O AGRUPAMIENTO:** se trata de una técnica no supervisada para detectar patrones comunes en los datos. Al definir métricas apropiadas y utilizando nociones de distancia y similitud entre pares de observaciones, el propósito de los métodos de *clustering* es la identificación de grupos homogéneos de observaciones, denominados *clusters*. Con respecto a la distancia específica seleccionada, las observaciones pertenecientes a cada grupo deben estar próximas entre sí y alejadas de las incluidas en otros grupos. Los algoritmos de clustering pertenecen a diferentes categorías atendiendo a la forma de realizar las agrupaciones: métodos de particionado basados en distancias (por ejemplo, *K-means*), basados en jerarquías (*clustering* jerárquico), basados en densidades (DBSCAN)...., aunque hay muchos más.

Típicamente, las siguientes etapas son las de modelado matemático del problema a resolver. Gran parte de estos modelos se pueden implementar en el aprendizaje máquina (Machine Learning, ML), ya sea mediante procesos de aprendizaje supervisado o no supervisado. Algunos ejemplos son:

**REGRESIÓN:** tratar con un conjunto de datos que consta de observaciones pasadas, para las cuales se conocen tanto el valor de los atributos explicativos como el valor de la variable objetivo de naturaleza numérica y continua. Ejemplos de métodos de regresión incluyen regresión lineal simple/múltiple, regresión Ridge, redes neuronales feed-forward de tipo perceptrón multicapa (MLP) y profundas (Deep Learning).

**MODELADO DE SERIES TEMPORALES:** el objetivo de los modelos para el análisis de series temporales es identificar cualquier patrón regular de observaciones relativas al pasado, con el propósito de hacer predicciones para períodos futuros. El análisis de series temporales tiene muchas aplicaciones en los dominios empresarial, financiero, socioeconómico, medioambiental e industrial. Algunos ejemplos de métodos de modelado son: modelo Naive, alisado exponencial, modelos autorregresivos y de media móvil (y combinaciones de ambos: ARMA, ARIMA...), además de redes neuronales y modelos recurrentes como *Long-Short Term Memory* (LSTM) y *Gate-Recurrent Units* (GRU).

**CLASIFICACIÓN:** los modelos de clasificación son métodos de aprendizaje supervisado para predecir el valor de un atributo objetivo categórico, a diferencia de los modelos de regresión que tratan con atributos numéricos. A partir de un conjunto de observaciones pasadas cuya clase objetivo es conocida, los modelos de clasificación se utilizan para generar un conjunto de reglas que permiten predecir la clase objetivo de ejemplos futuros. Algunos ejemplos son los árboles de decisión, los Random Forests, los modelos bayesianos, las redes neuronales y las *Support Vector Machines* (SVM).

**REGLAS DE ASOCIACIÓN:** son métodos que derivan reglas de asociación cuyo objetivo es identificar patrones regulares y recurrencias dentro de un gran conjunto de transacciones (listas de datos). Son bastante simples e intuitivos y se utilizan con frecuencia para investigar transacciones de ventas en el *Market Basket Analysis* para relacionar la compra de un producto con la de otro, investigar las rutas de navegación dentro de los sitios web (minería de web), compras con tarjeta de crédito y detección de fraudes.



# 4

## FUTURO DE LA BI2A

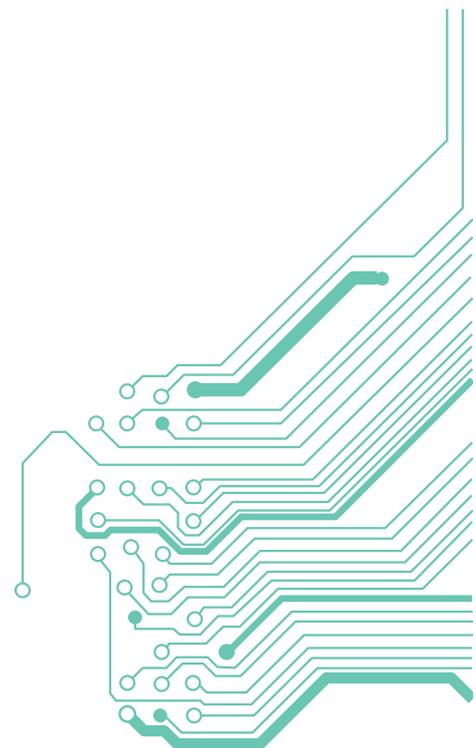
La formación de científicos de datos, así como la reconversión de otros titulados debido a la alta demanda de profesionales, supondrá el impulso definitivo a la incorporación de la analítica a los procesos del Business Intelligence y dará lugar a la BI2A.

Las plataformas de BI también evolucionan rápidamente para facilitar la aplicación las técnicas de IA con menos programación y, a pesar de todo, se estima que solo un 10% de los analistas en 2022 utilizarán el potencial de la tecnología disponible.

El futuro inmediato de la BI2A será sin duda de la explotación de los resultados basados en datos. Tanto es así que, en 2025, las historias de datos serán la forma más extendida de consumir analítica, y el 75% de las historias se generarán automáticamente. Hoy en día ya se pueden contratar las narrativas inteligentes (NLG) para transformar datos estructurados en narrativa escrita.

Otra de las grandes revoluciones será cuando la BI2A dé lugar al aprovechamiento masivo del lenguaje natural disponible en las empresas. Ya se pueden aplicar servicios ya desarrollados de detectores de idiomas, extractores de frases clave, moderación de contenido explícito y análisis de sentimientos, pero la verdadera revolución vendrá cuando se apliquen los avances que ya se están obteniendo con los modelos de lenguajes basados en aprendizaje profundo, como el famoso GPT-3, cuando se podrán resumir

**Tanto es así que, en 2025, las historias de datos serán la forma más extendida de consumir analítica, y el 75% de las historias se generarán automáticamente.**



Desde un punto de vista más general, el potencial de la IA permitirá ciertas mejoras:

**MEJORA DEL SERVICIO AL CLIENTE.** La capacidad de la IA para acelerar y personalizar el servicio al cliente se encuentra entre los principales beneficios que las empresas esperan obtener de la IA, clasificada como la número 2 entre las recompensas de la IA en un estudio de investigación de 2019 realizado por MIT Sloan Management y Boston Consulting Group.

**MEJORA DE LA SUPERVISIÓN.** La capacidad de la IA para procesar datos en tiempo real significa que las organizaciones pueden implementar un monitoreo casi instantáneo; por ejemplo, las plantas de producción están utilizando software de reconocimiento de imágenes y modelos de aprendizaje automático en los procesos de control de calidad para monitorear la producción y señalar problemas.

**DESARROLLO DE PRODUCTOS MÁS RÁPIDO.**

La IA permite acortar los ciclos de desarrollo y reducir el tiempo que transcurre entre el diseño y la comercialización para obtener un retorno de la inversión más rápido.

**MEJOR CALIDAD.** Las organizaciones esperan una reducción de los errores y una mayor adherencia a las normas de cumplimiento mediante el uso de la IA en tareas que antes se realizaban manualmente o con herramientas de automatización tradicionales, como la extracción, transformación y carga. La conciliación financiera es un área en la que el aprendizaje automático, por ejemplo, ha reducido sustancialmente los costes, el tiempo y los errores.

**INNOVACIÓN Y EXPANSIÓN DEL MODELO DE NEGOCIO.**

Los nativos digitales como Amazon, Airbnb, Uber y otros ciertamente han utilizado la IA para ayudar a implementar nuevos modelos de negocio. Según Andrew Ng, pionero en el desarrollo de la IA en Google y Baidu y actual director general y cofundador de Landing AI, las empresas tradicionales pueden encontrar difícil la transformación del modelo de negocio con IA. Ofreció un manual de seis pasos para poner en marcha la IA en las empresas tradicionales.



## BIBLIOGRAFÍA

- E. Alpaydin. Machine Learning. MIT Press, 2021.
- C. Howson. Successful Business Intelligence, Second Edition: Unlock the Value of BI & Big Data. McGraw-Hill, 2013.
- M. Iansiti, K. Lakhani. Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World. Ingram Publisher Services, 2020.
- B. Marr. La revolución inteligente. TEEL Editorial. 2020.
- N. Valigi, G. Mauro. Zero to AI. Manning Publications. 2020.

# 2

## Cuadros de mando avanzados para la administración pública



Borja Castillo Caballero  
Joan Vila Francés  
José Joaquín de Haro Navarro  
Pedro Cano Michelena

# 1

## INTRODUCCIÓN



Actualmente, la analítica de datos se ha establecido como un elemento clave en la toma de decisiones empresariales, ya que permite evaluar el funcionamiento de las compañías y la optimización de recursos. Sin embargo, su uso no está todavía muy asentado en las administraciones públicas. Las administraciones públicas funcionan en muchos aspectos como cualquier gran empresa, como por ejemplo en cuanto a número de empleados, y además trabajan con un elevado volumen de datos derivados de su propio funcionamiento y de los ciudadanos a los que dan servicio. Por tanto, se pueden beneficiar igualmente de la analítica de datos para mejorar su eficiencia y prestar un mejor servicio a la ciudadanía.

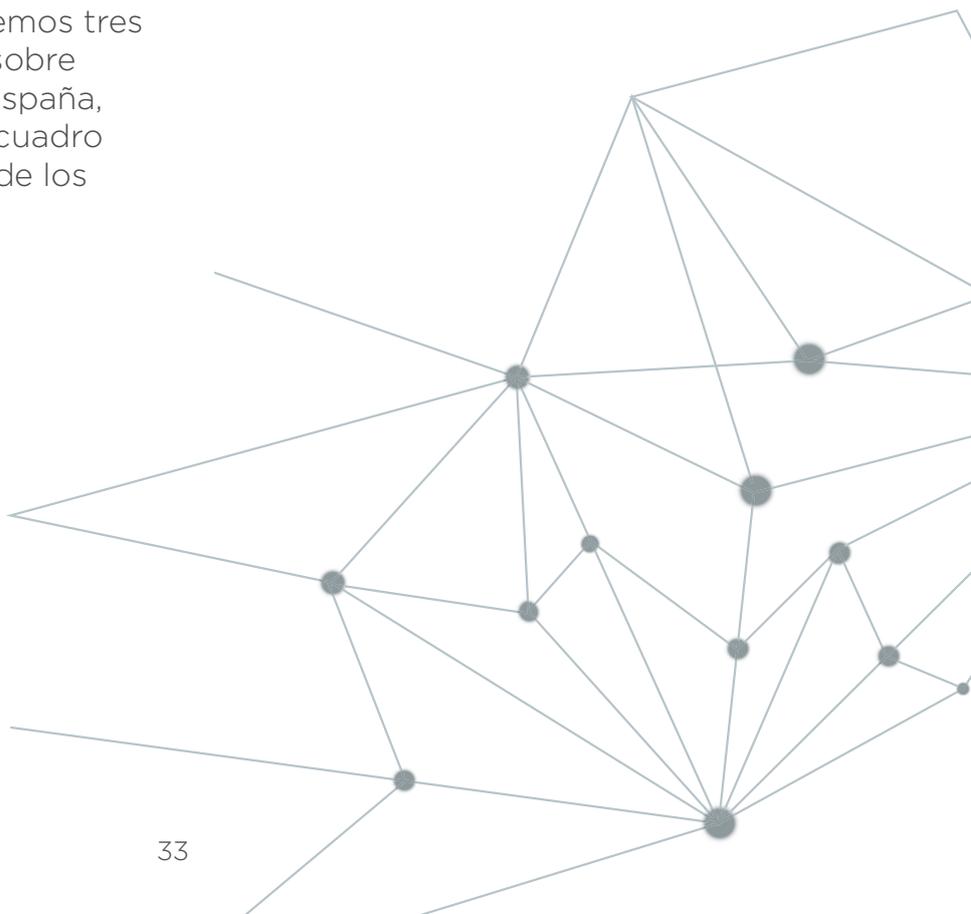
Dentro de la analítica de datos, los cuadros de mando constituyen una herramienta muy práctica para la interpretación visual de grandes volúmenes de datos. Un cuadro de mando se puede entender como una aplicación o un servicio web que muestra de manera dinámica y gráfica la información resumida de los datos de una organización o un proceso determinado. Los cuadros de mando aportan muchos beneficios sobre los tradicionales informes ejecutivos empresariales, pues en lugar de mostrar una foto fija en el tiempo y estática, permiten mostrar los datos en tiempo real y de una manera interactiva, en la que el usuario del cuadro de mando elige qué información quiere visualizar y en qué grado de detalle o profundidad.

Los usos de los cuadros de mando en la Administración Pública son muy diversos, aunque se pueden agrupar en dos categorías: aquellos destinados a mostrar información de gobernanza a la ciudadanía (p. ej., la rendición de cuentas mediante los portales de transparencia), y aquellos destinados a informar del



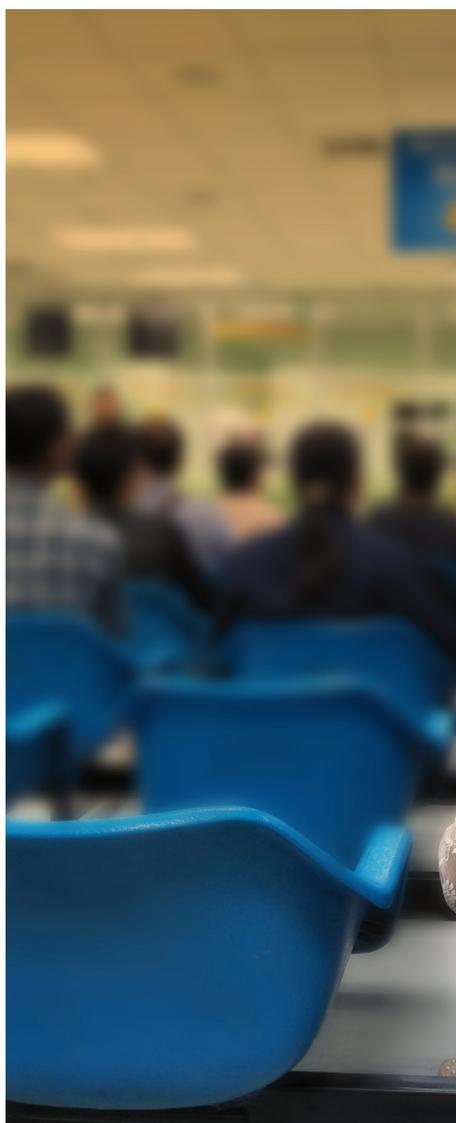
propio funcionamiento de la institución a la gerencia (p. ej., métricas de rendimiento de los procesos administrativos mediante KPI — Key Performance Index—).

En este capítulo vamos a repasar los conceptos más importantes de un cuadro de mando, y veremos su aplicación práctica a la Administración Pública. Como casos prácticos de uso, mostraremos tres ejemplos: un cuadro de mando sobre la evolución de la COVID-19 en España, un portal de transparencia y un cuadro de mando sobre el rendimiento de los procedimientos administrativos.



# 2

## LA ADMINISTRACIÓN PÚBLICA EN EL CONTEXTO ESPAÑOL



Para comprender la necesidad del uso de herramientas como los cuadros de mando para mejorar la gobernanza de las administraciones públicas, conviene hacer un repaso de la dimensión de la Administración Pública de España.

La Administración Pública de España es el aparato de gobierno y gestión de los intereses públicos españoles. Estructuralmente, la Administración Pública está formada por un conjunto de administraciones públicas divididas en administraciones territoriales (formadas por la Administración General del Estado, administraciones autonómicas y entidades locales), administraciones instrumentales o institucionales (organismos públicos) y las administraciones corporativas (colegios profesionales, cámaras de comercio, etc.).

El número de administraciones públicas en España es muy elevado. Según los datos del Ministerio de Hacienda y Función Pública, en el año 2021 había un total de 16.818 entes pertenecientes a la Administración Pública de España, como detalla la tabla 1.

**Tabla 1:** recuento de entes pertenecientes a la Administración Pública de España. **Fuente:** Inventario de entidades pertenecientes al sector de las Administraciones Públicas, Intervención General de la Administración del Estado (actualizado a 30 de julio de 2021).

SUBSECTOR	N.º DE ENTES
Administración Central	329
Administración Regional	1.331
Administración Local	15.129
Fondos de la Seguridad Social	29
<b>TOTAL</b>	<b>16.818</b>

El gasto presupuestario que manejan las administraciones públicas en España es, por tanto, también muy elevado. Según el INE, para el año 2019 (último año disponible) el gasto total de las administraciones públicas fue de 523.441 millones de euros. Como veremos más adelante, el análisis del gasto es una de las aplicaciones de los cuadros de mando que mayor beneficio pueden aportar a las administraciones públicas, tanto desde el punto de vista de la transparencia como del control interno.





# DATOS EN LAS ADMINISTRACIONES PÚBLICAS

Al igual que cualquier empresa del sector privado, las administraciones públicas manejan una gran cantidad de datos que son susceptibles de ser analizados mediante cuadros de mando. A diferencia de los datos de las empresas privadas, que se suelen limitar a sus clientes, las administraciones públicas recogen datos de toda la ciudadanía. Estos datos son de muy diversa índole, y pueden ser tanto estructurados como no estructurados. Entendemos por datos estructurados aquellos que se organizan con una forma tabular determinada y tienen un formato fijo (p. ej., una hoja de cálculo o una base de datos). Sin embargo, los datos no estructurados son aquellos que no siguen un formato determinado (p. ej., texto y contenidos multimedia). Los datos no estructurados son mucho más difíciles de analizar por sistemas automáticos, y es necesario aplicar previamente alguna técnica de extracción de conocimiento (minería de datos) para poder extraer características —en forma de información estructurada— de ellos.

A continuación, presentamos los tipos de datos más relevantes que manejan las administraciones públicas.

## DATOS DEMOGRÁFICOS

Cada administración pública recoge los datos de todos aquellos ciudadanos que están bajo su ámbito de actuación. Según la naturaleza de la administración pública en cuestión, estos datos pueden ser meramente demográficos (p. ej., el padrón) o incluir otro tipo de información (p. ej., datos de salud como en el ejemplo de la COVID-19 que veremos más adelante). Si bien el Instituto Nacional de Estadística (INE) ya recopila y agrega estos datos demográficos a nivel estatal, el uso de esta información por cada administración pública permite una mayor granularidad en la visualización de los datos.

Por tanto, el hecho de que cada administración pueda mostrar de forma agregada esta información puede ser útil a título informativo —permite la consulta ciudadana—, pero también como ayuda a la toma de decisiones por parte de los gestores de la entidad.

## ADMINISTRACIÓN ELECTRÓNICA

Actualmente, muchos trámites de los ciudadanos con la Administración se realizan de manera telemática, lo que permite mantener una información estructurada y siempre actualizada del volumen de tramitación. Por tanto, la generación de cuadros de mando que recopilen mediante gráficas y en tiempo real la información acerca de estos trámites es una aplicación directa y sencilla de implementar.

Adicionalmente, la mayoría de la contratación de las administraciones con sus proveedores se realiza de manera electrónica, y deja un registro digital de esta actividad. En este aspecto, los cuadros de mando son una herramienta idónea para la rendición de cuentas. En la actualidad, muchas administraciones han implantado los llamados portales de transparencia con esta finalidad.

En relación con la administración electrónica, los cuadros de mando no se limitan a mostrar la información de los trámites y los contratos realizados por cada administración, sino que además pueden mostrar información acerca de la eficiencia de estos procedimientos. Por ejemplo, en un cuadro de mando acerca de los trámites electrónicos se mostrarán no solo el volumen de tramitaciones solucionadas, de manera agregada y como serie temporal, sino también el tiempo de resolución por tipo de trámite, el desglose por unidades administrativas, etc. Esto permite a los gestores detectar los cuellos de botella y mejorar los procesos.

## RECURSOS HUMANOS

Las administraciones públicas emplean a una gran cantidad de personas. Según el último dato disponible (enero 2021), España cuenta con 2.710.405 empleados públicos, lo que convierte al Estado en el mayor empleador del país. La gestión de los recursos humanos implica llevar un control

sobre las tareas, las responsabilidades y la eficiencia de cada empleado. Tradicionalmente, las administraciones públicas no han sido muy resolutivas para evaluar a sus empleados, pero el uso de cuadros de mando puede ayudar a esta gestión, facilitando la información de manera gráfica de diversas métricas de rendimiento o KPI. A medida que se aumenta la digitalización del trabajo que desempeñan los empleados, estas métricas pueden incluir una información más detallada de su

rendimiento (p. ej., volumen de tramitación realizado, tiempo empleado por trámite, etc.).

Después, esa información se puede mostrar de manera individual o agregada con distintos grados de granularidad.

En el resto del capítulo se introduce de una manera más detallada el concepto de cuadro de mando, así como varios ejemplos de uso en el ámbito de las administraciones públicas.



# 3

## CUADROS DE MANDO



Pocos fenómenos caracterizan mejor nuestra era que el auge de las tecnologías de la información. Este auge ha generado un enorme torrente de datos que, sin las herramientas necesarias, serían virtualmente inaccesibles para su análisis y visualización. En este capítulo nos centraremos en analizar una de estas herramientas, los cuadros de mando.

### DEFINICIÓN

Un cuadro de mando es una interfaz gráfica de usuario que permite ver una serie de indicadores sobre los que tenemos interés de un solo vistazo. Normalmente, intentaremos que esta interfaz se pueda visualizar en una sola

pantalla y que, mediante las técnicas y las gráficas adecuadas, consigamos sintetizar grandes conjuntos de datos en un pequeño espacio de manera clara e inmediata. Para conseguir condensar tanta información en un lugar tan pequeño nos aprovecharemos de la interactividad que nos proporciona un ordenador más allá de, por ejemplo, un simple informe en papel. Aprovecharemos esta interactividad para, mediante las entradas del usuario, poder personalizar la información y las gráficas que se muestren por pantalla. Así pues, un buen cuadro de mando debe ser capaz tanto de mostrar un informe general de los datos como permitirnos realizar un análisis de aspectos concretos sobre en los que tengamos interés.

## TIPOS DE DATOS A VISUALIZAR

Los datos que podemos representar mediante cuadros de mando se pueden clasificar, siguiendo la estructura clásica de la estadística, en cualitativos o cuantitativos.

Los datos cualitativos describen características, identifican las observaciones dentro de una categoría, por lo que también se les llama datos *categoricos*. Si pueden ser ordenados (rango salarial, nivel educativo, curso escolar, **nivel de alerta** en una comunidad autónoma por el riesgo sanitario asociado a la COVID-19), son *ordinales*. Por el contrario, si no admiten orden alguno (sexo, etnicidad, lugar de nacimiento, nombre de una entidad pública), se llaman *nominales*.

Los datos cuantitativos expresan valores numéricos (peso, altura, número de pacientes en UCI, importe de un contrato público). Se dividen en discretos, si entre dos posibles valores distintos no puede existir otro, (cantidad de positivos en COVID-19 notificados los últimos 14 días), y continuos si entre dos valores distintos puede haber infinitos (peso, altura, tiempo de supervivencia).

La manera de proceder habitual en los cuadros de mando es filtrar y agrupar por variables cualitativas y representar gráficamente las variables cuantitativas asociadas a las primeras o el resultado de operar aritméticamente con ellas.

## TIPOS DE GRÁFICOS

Los tipos de gráficos que se pueden utilizar en un cuadro de mando pueden ser tan variados como nos podamos imaginar. Pero siempre han de comunicar la información de manera sencilla y eficaz. A continuación, describiremos algunos de los más usados y sus contextos más comunes.

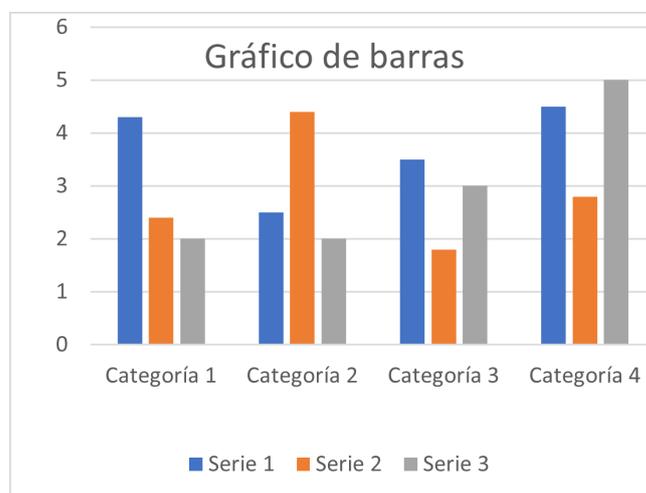
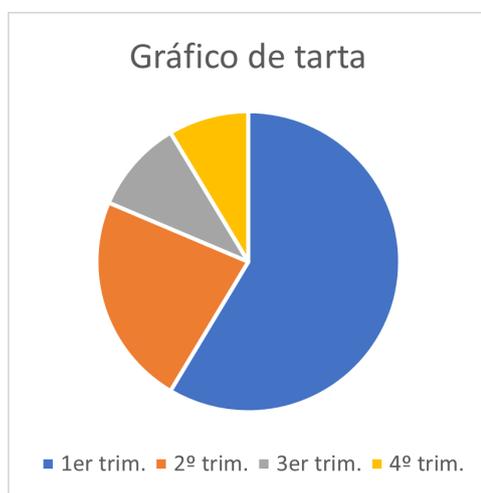
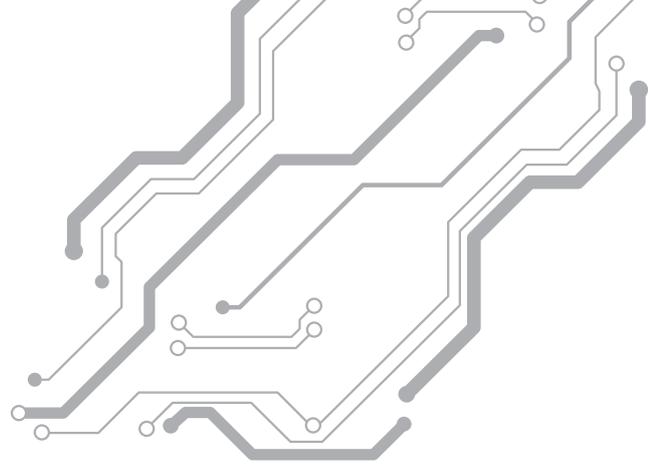


Ilustración 1



En la ilustración 1, podemos observar dos de los gráficos más comunes con los que nos encontraremos: el gráfico de tarta y el de barras, ambos utilizados para representar información categórica mediante áreas. Aunque cabe destacar que el gráfico de tarta no es muy recomendable, ya que presenta la información de manera menos accesible, es decir, es más fácil comparar el área de dos rectángulos contiguos que de dos segmentos de un círculo.

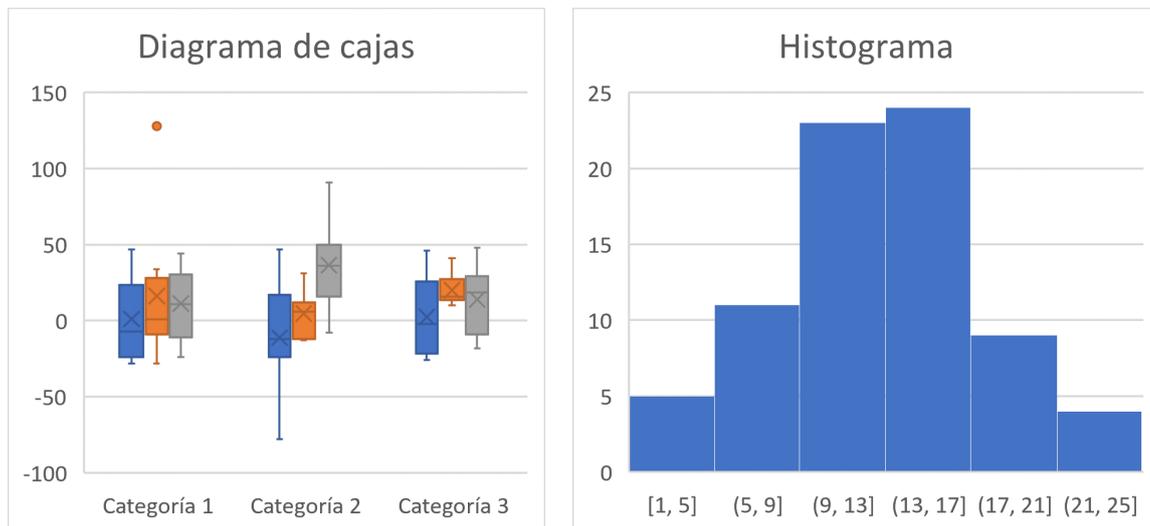


Ilustración 2

En cuanto a la ilustración 2, podemos encontrar dos formas de representar información continua: el diagrama de cajas y el histograma. El histograma nos permite ver la distribución de una variable continua mediante las alturas de los rectángulos. Estos rectángulos representan la cantidad de observaciones contenidas en los rangos indicados. Por otra parte, el diagrama de cajas representa la distribución de las variables continuas de una manera más compacta; mediante un rectángulo con bigotes. Cada uno de estos diagramas de cajas nos indica los cuantiles de la variable. Al tener un formato más compacto, podemos comparar distintas variables continuas a la vez, como podemos observar en la ilustración.



Ilustración 3

Otra de las gráficas más comunes con la que nos encontraremos son las series temporales. Estas se caracterizan por mostrar la evolución de una variable a lo largo del tiempo. En el ejemplo que mostramos, podemos ver cómo va cambiando el precio de un producto financiero.

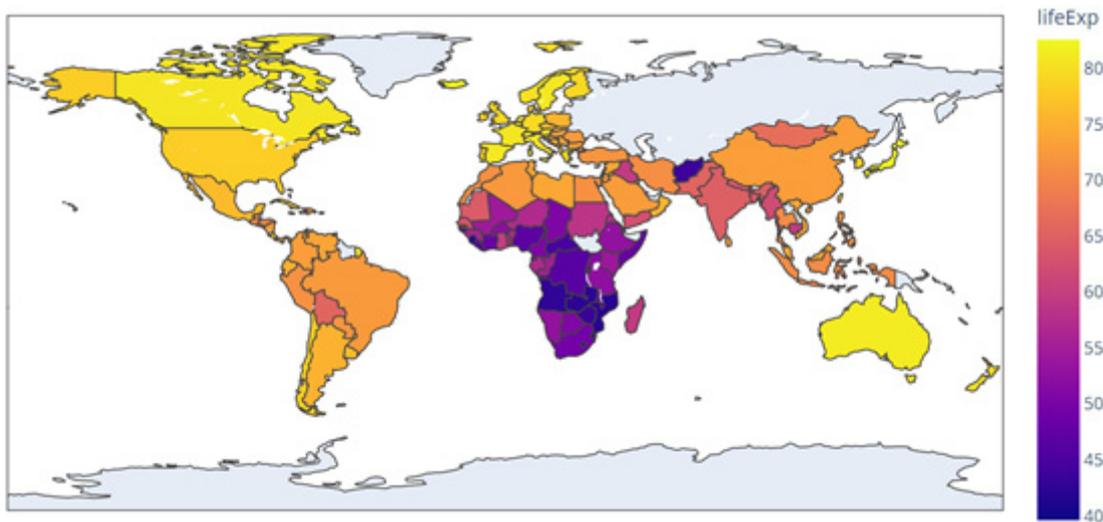


Ilustración 4

Uno de los mapas temáticos más utilizados es el mapa coroplético, caracterizado por colorear distintas regiones de distintos colores según algún índice de interés, como puede ser la población, la renta per cápita o, como en el caso representado, la esperanza de vida. La principal diferencia con el mapa de calor es que este sí que distingue áreas geográficas.

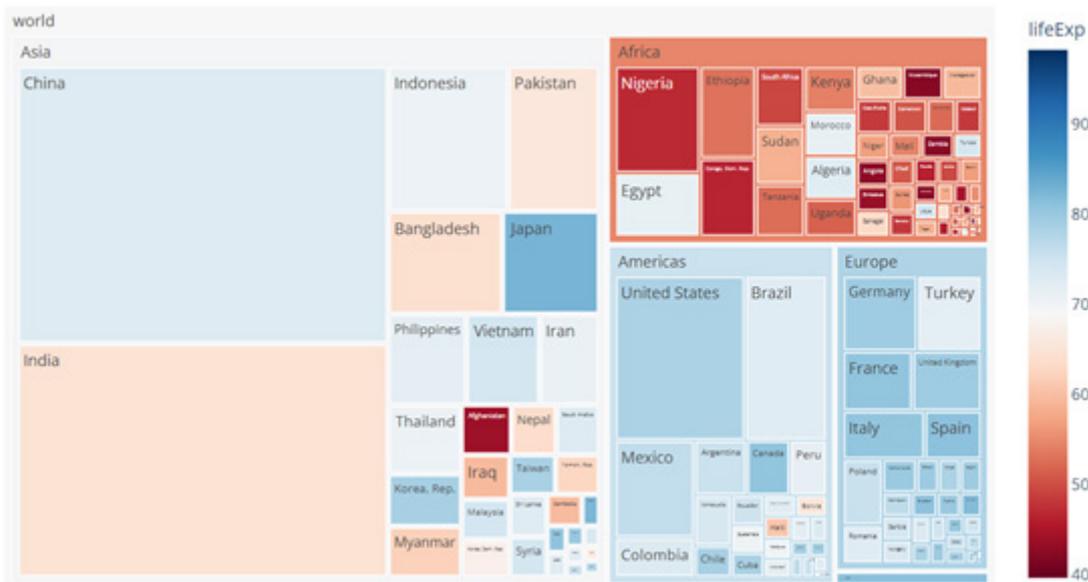


Ilustración 5

Por último, acabamos con una gráfica algo más compleja. El treemap es una forma de representar tres características simultáneamente. Por una parte, tenemos el tamaño del contenedor, por otra el color y por otra la jerarquía. En el caso de la ilustración 5, el tamaño representa la cantidad de población del país; el color, la esperanza de vida y la jerarquía viene dada por continentes y países. Los treemaps se ven muy beneficiados de la interactividad que les proporciona estar en un cuadro de mando, ya que esto permite a los usuarios centrarse solo en un nivel de la jerarquía, es decir, en nuestro ejemplo podríamos clicar en Asia para que esta se amplíe y se muestren solo los países que contiene.

## TECNOLOGÍAS

Las posibilidades a la hora de crear un cuadro de mando son amplias e incluyen alternativas de **software** libre y comercial. Empezaremos comentando algunas librerías, de distintos lenguajes de programación, especializadas en la producción de representaciones gráficas y páginas web interactivas.

• **Shiny**: paquete del lenguaje de programación R, desarrollado por **RStudio**, para la creación de páginas web interactivas. Nos permite combinar toda la potencia de R para el análisis de datos con la capacidad de generar cuadros de mando, sin la necesidad de tener grandes conocimientos de desarrollo web. Además, todo ello se puede hacer en un entorno de software libre y gratuito. Es muy versátil, pues existe un amplio catálogo de elementos de HTML (**shinythemes**, **shinyWidgets**) y de JavaScript (**shinyjs**) adaptados para su uso en Shiny, además de permitirnos modificar fácilmente el estilo CSS.

Otra ventaja es poder utilizar la gran variedad de librerías de representación gráfica de R (**ggplot2**, **plotly.R**, **highcharter**, **vcd**, **Leafleft**, etc.). El cuadro de mando sobre la COVID-19 realizado por el Centro Nacional de Epidemiología, que veremos más adelante, es un ejemplo de una herramienta desarrollada en Shiny.

·**Dash**: librería de Python más popular en la creación de cuadros de mandos. Desarrollada por **Plotly**, con una versión de *software* libre y gratuita, además de otra comercial llamada Dash Enterprise.

Como hacía Shiny, permite compaginar un lenguaje de programación muy popular en el análisis de datos con la capacidad de generar potentes herramientas de visualización interactiva. Quizá el entorno de librerías desarrolladas alrededor de Dash sea algo más limitado que el de Shiny, pero al estar basada en **React** (una importante biblioteca de JavaScript para la construcción de interfaces de usuarios, con una amplia comunidad y muchos ejemplos disponibles), es fácil adaptar elementos construidos en esta librería para utilizarlos en Python.

Otro de sus puntos fuertes es la buena integración que tiene con la librería de gráficas **plotly.py**, pues ambas son

producidas por Plotly. La herramienta de control interno de SEDIPUALB@, explicada al final del capítulo, ha sido producida en Dash.

Si se tiene experiencia en el desarrollo de páginas web al modo tradicional (utilizando HTML, CSS, JavaScript), las librerías de representación gráfica de JavaScript son también una opción interesante. Algunas son software libre y gratuito, como **D3.js**, **plotly.js**, **Leaflet**, **Highcharts** (que, además, se comercializa con licencias de pago). También es posible acceder a versiones gratuitas de librerías de pago, como es el caso de **amCharts**. En el Visor Presupuestario de la Generalitat Valenciana, que explicaremos más adelante, se muestran gráficas construidas con D3.js.

Además, existen algunas empresas que comercializan *software* para producir cuadros de mando utilizando una interfaz de usuario, sin la necesidad de programar. Ejemplos de ello son **Power BI**, de Microsoft, **Tableau y ArcGIS**, de Esri. El cuadro de mando sobre la evolución de la COVID de la Universidad Johns Hopkins y el de la Generalitat Valenciana, que se nombran al principio de la siguiente sección, son ejemplos generados con ArcGIS.



## CASOS DE USO EN LAS ADMINISTRACIONES PÚBLICAS

Presentamos a continuación tres ejemplos del uso de los cuadros de mando en la Administración Pública.

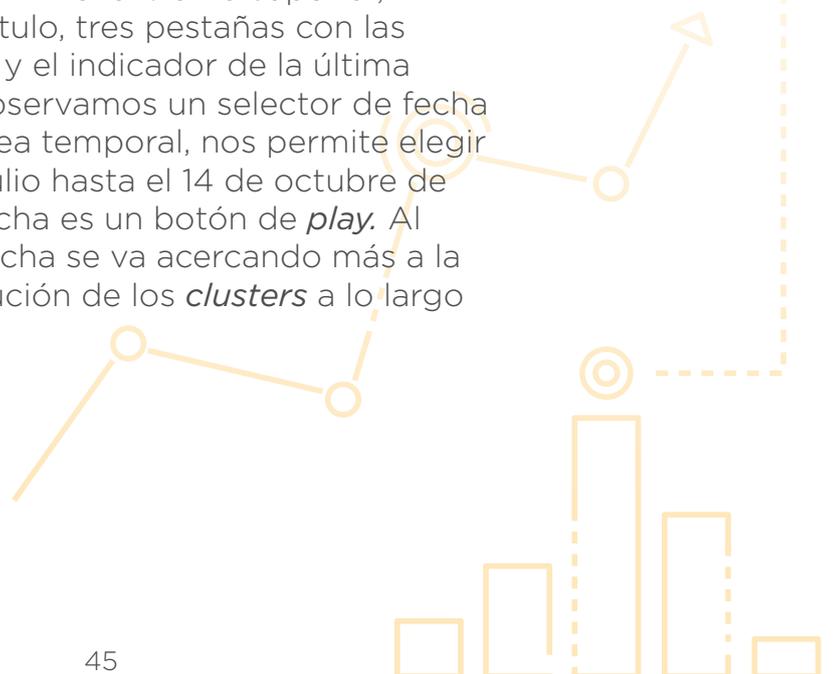
### COVID-19: VIGILANCIA Y EVOLUCIÓN EN ESPAÑA

La pandemia de COVID-19 generó una nueva necesidad de transmitir datos de manera rápida y sencilla al público general. Las gráficas que representaban la evolución de la transmisión, la ocupación de UCI o la mortalidad afloraron en todos los medios de comunicación. Consultar cuadros de mando como el de la **Universidad Johns Hopkins**, para conocer la evolución de la pandemia a nivel mundial, o como el de la **Generalitat Valenciana**, para saber en qué situación nos encontrábamos en nuestras localidades, se convirtió para muchos de nosotros en una actividad diaria.

Este último no es el único caso de cuadro de mando desarrollado por una entidad pública española, con el objetivo de dar información actualizada sobre la situación de la pandemia. Otro ejemplo es el desarrollado por el Centro Nacional de Epidemiología, centro propio del Instituto de Salud Carlos III, en colaboración con el Hospital 12 de Octubre de Madrid.

**La herramienta** permite detectar posibles brotes tanto a nivel nacional como autonómico, identificando aquellas zonas en las que, en un período de 7 días, se ha producido un número de casos estadísticamente superior al esperable en ese lugar y ese tiempo teniendo en cuenta las cifras nacionales. Además, es posible ver la evolución de estas zonas o **clusters** a lo largo del tiempo. Se ha construido en R, mediante las librerías Shiny y Leaflet.

En la ilustración 6, vemos el panel inicial del cuadro de mando, el dedicado a la información nacional. En el extremo superior, de izquierda a derecha, tenemos el título, tres pestañas con las que podemos navegar entre paneles y el indicador de la última actualización de los datos. Debajo observamos un selector de fecha que, deslizando el círculo sobre la línea temporal, nos permite elegir una semana concreta desde el 1 de julio hasta el 14 de octubre de 2021. El elemento junto a la última fecha es un botón de **play**. Al clicar en él, el valor del selector de fecha se va acercando más a la actualidad, y nos permite ver la evolución de los **clusters** a lo largo del tiempo.



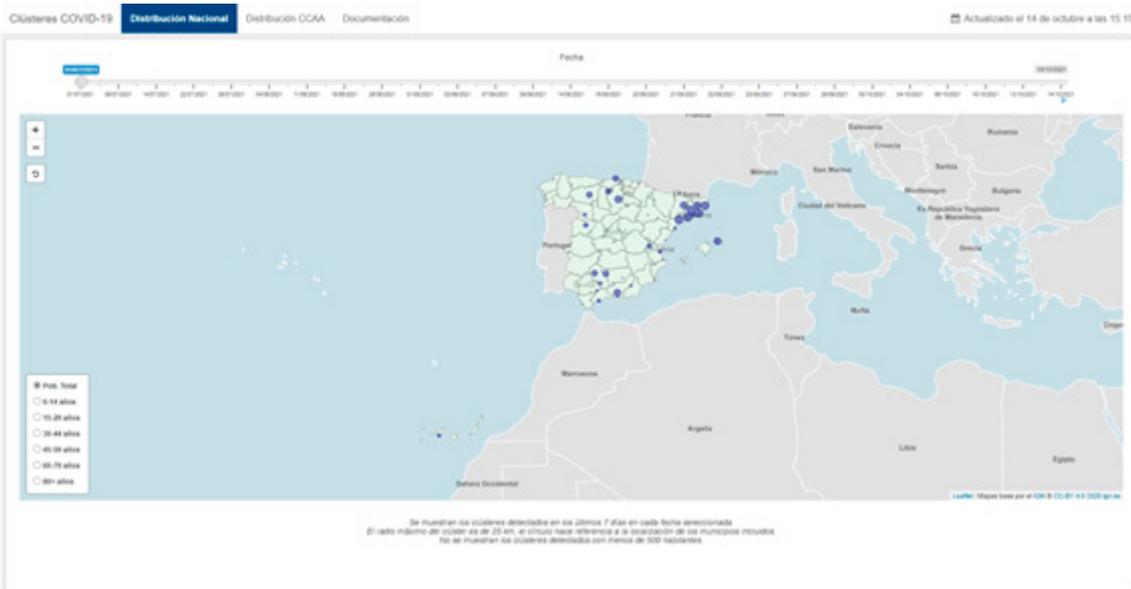


Ilustración 6

En la posición central de la pantalla vemos el mapa en el que se nos muestran los *clusters*. Dentro de él podemos hacer zoom, y mediante el selector de la esquina superior izquierda se nos permite observar los datos relativos a toda la población o a una franja de edad concreta. Los *clusters* son representados como círculos de color azul, ligeramente transparente, con radio máximo de 25 km y proporcional al número de municipios que contienen (para situar los municipios se considera el centro geográfico de los mismos).



Ilustración 7

En ilustración 7 tenemos una versión ampliada del mapa, en la que, además, vemos el efecto de situar el ratón sobre uno de los *clusters*. Nos aparece un resumen de la información asociada a él: el período al que se corresponden los datos (máximo una semana), la población en riesgo (el número total de habitantes residentes en el área coloreada), el número de municipios incluidos, la cantidad de casos observados, el número de casos que cabría esperar teniendo en cuenta los datos

nacionales y, por último, el RR: riesgo estimado dentro del *cluster* dividido por el riesgo estimado fuera del mismo. El riesgo estimado se define como el número de casos observados en una zona, entre el número de casos esperados en esa misma zona.

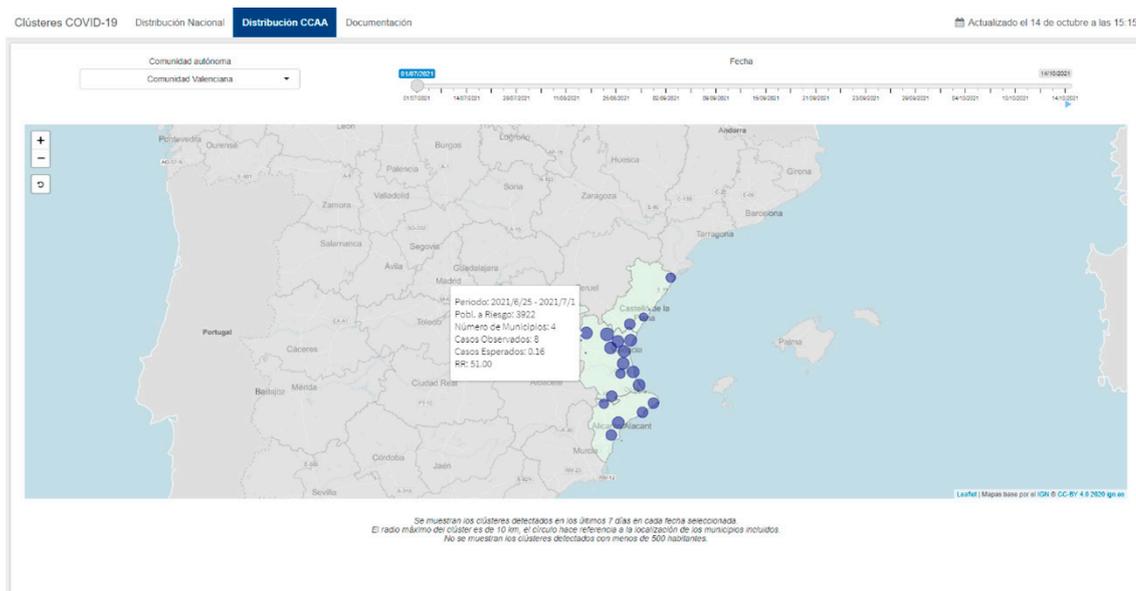


Ilustración 8

En el segundo panel (ilustración 8), en el que vamos a poder ver los clusters aplicados solo al territorio de cada comunidad autónoma, observamos como principales cambios la aparición de un selector desplegable que nos permite elegir la comunidad autónoma y la ausencia del selector de franjas de edad. Por lo demás, el funcionamiento es el mismo que en el panel anterior, salvo que el radio máximo de los clusters se ha reducido a 10 km.



**Panel Clústeres COVID-19**

Los resultados que se presentan en este Panel de Clústeres de COVID-19 en España se obtienen a partir de los casos individualizados declarados a la Red Nacional de Vigilancia Epidemiológica (RENAVE) a través de la plataforma SIVEIS (sistema de vigilancia de España) que gestiona el Centro Nacional de Epidemiología (CNE). Esta información procede de la encuesta epidemiológica para cada caso que las comunidades autónomas (CCAA) complementan ante la identificación de un caso de COVID-19.

En el panel se presentan los agrupamientos espacio-temporales detectados en los últimos 7 días, siguiendo la metodología propuesta por Kubonik, identificando los clústeres activos en España y en cada CCAA. Para ello, se utilizan los casos con fecha de inicio de síntomas o, en su defecto, la fecha de diagnóstico menos 3 días, para los casos asintomáticos se utiliza la fecha de diagnóstico. En aquellos casos en los que no se dispone de fecha de inicio de síntomas ni de diagnóstico se utiliza la fecha más próxima entre las fechas de consulta o de diagnóstico, aunque ocasionalmente se puede sustituir por la fecha de toma de muestras. Se consideran como casos aquellos que han sido confirmados por PCR o por pruebas de antigénicas. Con el fin de detectar clústeres a nivel municipal, los casos están agregados utilizando el municipio de residencia. Se utiliza la población del Padrón Municipal a 1 de Enero del 2020 del [Instituto Nacional de Estadística](#). Para los análisis nacionales se utilizan 5 grupos de edad y la población total, para los análisis por CCAA la población total de cada una de ellas.

Un clúster se identifica como un exceso de casos observados en el espacio y en el tiempo comparado con el resto del territorio. El tamaño máximo del clúster en el espacio se ha establecido de 25 km a nivel nacional y de 10 km por CCAA. El periodo temporal comprende los últimos 7 días desde la fecha seleccionada comparando con los 30 días previos. Se puede consultar toda la metodología empleada y su desarrollo en una [publicación](#) del CNE.

**Interpretación de los resultados**

Los resultados se muestran a nivel nacional y por CCAA quedando representados únicamente aquellos clústeres estadísticamente significativos con una probabilidad de ocurrencia debida al azar inferior al 5%. El tamaño del círculo hace referencia a la localización de los municipios incluidos en cada clúster, ordenando los clústeres que tengan menos de 500 habitantes. Como coordenadas del municipio se toma el centro geográfico de este.

En cada clúster detectado se presentan una serie de parámetros:

- **Periodo:** duración temporal del clúster (máximo 7 días anteriores al día consultado).
- **Presencia a riesgo:** en la suma de la población en el radio de análisis (radio 25 km a nivel nacional y 10 km por CCAA).
- **Número de municipios:** municipios que están incluidos dentro del clúster.
- **Casos observados:** número de casos observados en el espacio y en el tiempo dentro del clúster.
- **Casos esperados:** número de casos que se esperarían en ese espacio y en el tiempo teniendo en cuenta la tasa de incidencia general a nivel nacional o autonómico (población y casos) calculado para los 30 días previos.
- **RR:** es el riesgo estimado dentro del clúster dividido por el riesgo estimado fuera del mismo. Se calcula como los casos observados entre los casos esperados dentro del clúster dividido por los observados entre los esperados fuera del clúster.

Ilustración 9

Por último, en el tercer panel (ilustración 9) se nos ofrece información sobre la metodología utilizada y ayuda para interpretar los gráficos.

Esta herramienta resultó ser realmente útil para detectar la progresión de los distintos brotes que ocurrieron durante el verano de 2020, después del final del confinamiento total<sup>iii</sup>.

## RENDICIÓN DE CUENTAS: PORTALES DE TRANSPARENCIA

Las iniciativas de transparencia en la Administración surgen con la intención de evitar el ostracismo y las malas prácticas, con la puesta a disposición de información a la ciudadanía, de modo que puedan ejercer el necesario control de las instituciones. Estas medidas no pueden quedarse en el mero hecho de permitir que el ciudadano acceda a la información; además, se debe poder hacer fácilmente y presentando los datos de manera comprensible e, incluso, atractiva.

Con estos objetivos, la Generalitat Valenciana impulsó en 2015 su propio portal de transparencia, **GVA Oberta**, que desde 2018 incorpora el **Visor Presupuestario**, un cuadro de mando para visualizar quién y en qué se gasta el dinero de la Generalitat y de las entidades asociadas a ella.



Más concretamente, la herramienta permite realizar un estudio detallado de las previsiones de ingreso y gasto planificadas en el presupuesto anual aprobado por las Cortes Valencianas, sus posibles modificaciones y procesos de ejecución. Implementa una serie de representaciones gráficas, realizadas mediante la librería D3.js de JavaScript.



Ilustración 10

En la ilustración 10, observamos la pantalla que obtenemos al acceder por primera vez al visor. Recuadrado en gris, tenemos el inicio de la herramienta. A diferencia de otros muchos cuadros de mando, este ocupa más de una pantalla: consta de una serie de secciones o apartados dispuestos verticalmente, que varían según las elecciones que hagamos en el selector de clasificación que vemos en la parte central derecha.

Primero, nos encontramos con un panel (¿Cómo y en qué gastamos?) que nos introduce a la herramienta. En la esquina superior derecha, tanto del panel como de la página, tenemos un botón de información que abre un panel de ayuda (ilustración 11).

Esta misma estructura (panel, título, botón de ayuda) la encontraremos en la mayoría de las secciones siguientes. El sistema de ayuda es realmente útil para el usuario poco acostumbrado al vocabulario vinculado a la gestión presupuestaria.

**GVA Oberta**  
PORTAL DE TRANSPARENCIA DE LA GENERALITAT VALENCIANA

EL CONSELL ORGANIZACIÓN DE LA GVA SERVICIOS A LA CIUDADANÍA CIFRAS TRANSPARENCIA POR SECTORES TRANSPARENCIA EN LA GENERALITAT

**¿Cómo y en qué gastamos? Consulta y analiza los presupuestos de la Generalitat Valenciana**

Te mostramos de una manera clara como se distribuyen nuestros ingresos y gastos. Navega a través de las gráficas localizando cada partida del presupuesto. También puedes consultar esta información en forma de tabla, o descargar los datos abiertos.

FECHAS CENTROS CLASIFICACIÓN

Generalitat Valenciana  
Total Gastos - Gastos por clasificación funcional o programa

Presupuesto aprobado por las Cortes: 25.627,55 millones €

PRESUPUESTO	COMPROMETIDO	RECONOCIDO	PAGADO
25.627,55	14.888,60	10.018,87	10.769,89

## Presupuesto

El presupuesto es el plan económico y financiero anual de la Generalitat para conseguir el buen funcionamiento de todos los servicios públicos y los objetivos e inversiones del Equipo de Gobierno. El presupuesto tiene carácter anual y contiene las previsiones de ingresos y gastos del ejercicio. El presupuesto es la herramienta de planificación y programación, control, seguimiento y evaluación. Los presupuestos constituyen la expresión cifrada, conjunta y sistemática de las obligaciones y de los derechos económicos de la Generalitat durante el correspondiente ejercicio. Y por lo tanto contiene una información imprescindible para conocer, interpretar y auditar la gestión de la Generalitat.

### Ver también

- Cuentas Públicas

Ilustración 11

A continuación, tenemos el selector principal del panel de control, que consta de tres menús desplegable. **Fechas** permite seleccionar un mes desde el 2015 hasta la actualidad o el presupuesto anual de cada uno de los años dentro del mismo período. En **Centros**, podemos filtrar por organismos, y se puede elegir entre el total de entidades que conforman la Generalitat Valenciana, una sola de ellas o alguno de los organismos autónomos de la Generalitat. En **Clasificación** es posible elegir entre visualizar ingresos o gastos. En este último caso, también podemos decidir mediante qué clasificación se van a estructurar los datos que se representen después. Las tres clasificaciones posibles son: económica, funcional, y orgánica. Respectivamente, nos indica en qué se gasta, para qué y quién.

Después de haber seleccionado mayo de 2019, Generalitat Valenciana y gastos por clasificación funcional, observamos en la ilustración 12 lo que podría ser la sección 0 del cuadro de mando. Debajo del selector principal nos aparece un resumen de las elecciones que hemos hecho. Seguidamente, un panel en el que se nos indica el montante total del presupuesto aprobado por las cortes en 2019, y una serie de cifras destacadas. De izquierda a derecha representan los siguientes puntos: el total del presupuesto después de realizarse modificaciones a las partidas originales, es decir, la suma de los llamados **créditos definitivos**; la parte del crédito definitivo **ya comprometido** con terceros; la cantidad del crédito comprometido cuyo abono a los terceros es **reconocido** como debido por la Administración, una vez que los acreedores

cumplen sus obligaciones; el montante ya pagado por la Administración y, finalmente, el *gasto por habitante*. Las cifras representadas son las que se correspondían con estos conceptos hasta mayo de 2019.

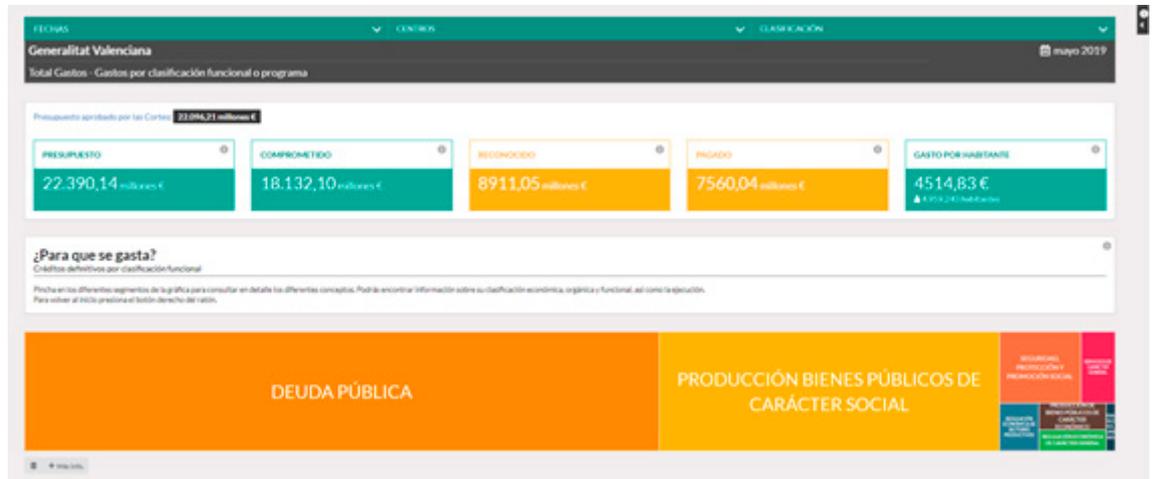


Ilustración 12

Debajo tenemos un *treemap*, en el que solo vamos a ver representadas dos dimensiones de la información: cada par recuadro-color representa una categoría dentro de la clasificación que hemos elegido, y al clicar en uno de ellos pasaremos a ver las subcategorías correspondientes, y así sucesivamente; el área de cada recuadro indica la porción del crédito definitivo asociado a la categoría, hasta mayo de 2015. Al pasar el ratón por encima de uno de los recuadros, se nos muestra el gasto relacionado con la categoría representada (ilustración 13). En la esquina inferior izquierda tenemos dos botones: el primero abre en una pestaña nueva los datos con los que se ha generado el *treemap* en la pantalla, y el siguiente abre el sistema de ayuda.



Ilustración 13

La característica más llamativa de este *treemap* es que también funciona como un selector. Como observamos en la ilustración 13, hemos escogido la categoría principal *Producción de Bienes Públicos de Carácter Social*, después la subcategoría *Educación*, y dentro de ella *Administración General de Educación*, y lo que se nos muestra son las subdivisiones de esta última selección. En la mayoría de las secciones siguientes, veremos que los datos representados son relativos a *Administración General de Educación* o a sus subdivisiones.

En la ilustración 14 observamos como, al bajar un poco más en la página, ha aparecido una barra de navegación acoplada al *treemap* que nos permite pasar de una sección a otra rápidamente. También vemos el apartado Presupuesto, que consiste en un listado de los elementos que aparecen en el *treemap* y los colores y los gastos asociados.



Ilustración 14

Al hacer clic en alguno de los elementos de la lista, se produce el mismo resultado que si clicáramos en el recuadro correspondiente del *treemap*. Esta funcionalidad permite seleccionar cualquier categoría fácilmente, aunque su área en el *treemap* sea muy pequeña.

En la sección *Gastos Cruzados* (ilustración 15), vamos a poder ver aquellas categorías de la clasificación económica (*¿En qué se gasta?*) y de la orgánica (*¿Quién gasta?*) en las que se incluyen los créditos que hemos elegido en el selector principal

y en el treemap. Observamos el mismo tipo de representación gráfica en ambos apartados: una barra horizontal dividida en áreas de colores distintos y de tamaños proporcionales al crédito definitivo en las categorías asociadas; debajo encontramos de nuevo una lista con las categorías, los colores y el montante de los créditos.

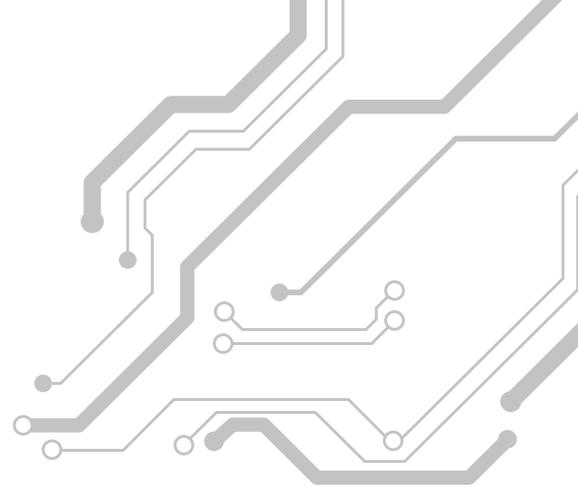


Ilustración 15

En el apartado *Ejecución por meses*, (ilustración 16) observamos cinco series temporales que representan la evolución mensual de los distintos tipos de crédito/gasto respectivos a la elección realizada en un principio. Pese a haber seleccionado mayo de 2019, aquí se nos muestran los datos de todo el 2019.



Ilustración 16



Complementando el gráfico hay una leyenda que nos indica que variable representa cada línea y, además, al pasar el ratón por encima del gráfico también nos muestra los valores en el mes en que nos situemos. En Ejecución por capítulos, (ilustración 17) vemos de nuevo un treemap con una lista acoplada. El funcionamiento, salvo la función de selector, es el mismo que en el anterior. La diferencia es que ahora las categorías son capítulos, el nivel principal en la Clasificación Económica. Nos enseña como el crédito comprometido, hasta mayo de 2019, en las partidas presupuestarias seleccionadas se distribuye en los distintos capítulos.



Ilustración 17

La penúltima sección nos muestra la evolución de los créditos definitivos, del mes elegido, asociados a las partidas presupuestarias seleccionadas, año a año desde 2016 a 2019. Es decir, en la gráfica de la ilustración 18, el tercer punto de las líneas representa el total del crédito definitivo en mayo de 2018, para la categoría correspondiente

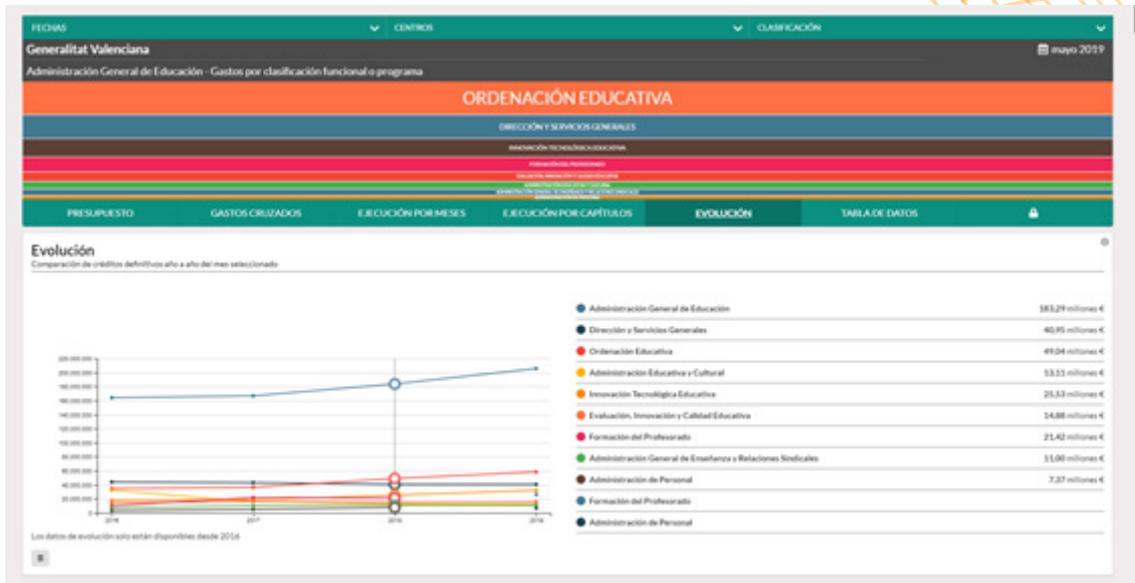


Ilustración 18

Por último, tenemos el único apartado que es totalmente independiente de las selecciones en el primer *treemap*. Como vemos en la ilustración 19, consiste en una tabla desplegable que nos indica el montante de los distintos tipos de crédito/gasto asociados a las categorías fijadas por la clasificación funcional. Es posible, por ejemplo, clicar en Servicios de *Carácter General* para ver los datos correspondientes a sus subdivisiones.

CLASIFICACIÓN	DESCRIPCIÓN	CRÉDITOS INICIALES	CRÉDITOS DEFINITIVOS	COMPROMETIDO	RECONOCIDO	PAGADO
0	Deuda Pública	5.762.676,2300 €	5.765.796,3900 €	5.697.327,3010 €	3.277.884,6791 €	3.277.882,9516 €
1	Servicios de Carácter General	722.775,0500 €	745.926,3042 €	512.277,8694 €	306.806,4835 €	129.505,5764 €
11	Alta Dirección de la Comunidad Autónoma y Gobierno	144.854,3900 €	148.578,8800 €	134.146,7200 €	48.452,5180 €	22.900,9064 €
12	Administración General	253.491,3000 €	260.265,6687 €	282.594,4292 €	45.522,0165 €	23.591,4212 €
13	Relaciones Externas	32.653,9500 €	32.745,6000 €	3.240,4123 €	457,0202 €	432,6267 €
14	Justicia	289.852,4100 €	304.336,1395 €	234.696,2963 €	92.374,9305 €	82.964,6218 €
141	Administración de Justicia	289.852,4100 €	304.336,1395 €	234.696,2963 €	92.374,9305 €	82.964,6218 €
2	Defensa, Protección Civil y Seguridad Ciudadana	119.940,8600 €	127.846,0820 €	127.306,2170 €	51.965,9820 €	1.566,5250 €
3	Seguridad, Protección y Promoción Social	1.707.154,6700 €	1.778.936,8081 €	1.245.083,2647 €	469.860,4470 €	365.234,3762 €
4	Producción Bienes Públicos de Carácter Social	524.188,0300 €	52.392,844370,62 €	9.549,008,9410 €	4.483,842,425,49 €	3.844,495,723,75 €
5	Producción de Bienes Públicos de Carácter Económico	512.905,7600 €	548.931,707,38 €	426.382,426,79 €	178.193,247,53 €	63.393,492,82 €
6	Regulación Económica de Carácter General	411.843,9800 €	430.118,990,31 €	529.595,451,58 €	53.000,051,29 €	33.906,197,17 €
7	Regulación Económica de Sectores Productivos	577.848,2600 €	578.705,369,48 €	405.207,784,75 €	181.496,007,21 €	49.122,080,74 €

Ilustración 19

Además, mediante el botón en la esquina inferior derecha podemos descargar un archivo CVS con todos los datos necesarios para generar la tabla.

## CONTROL INTERNO: MONITORIZACIÓN DE RECURSOS

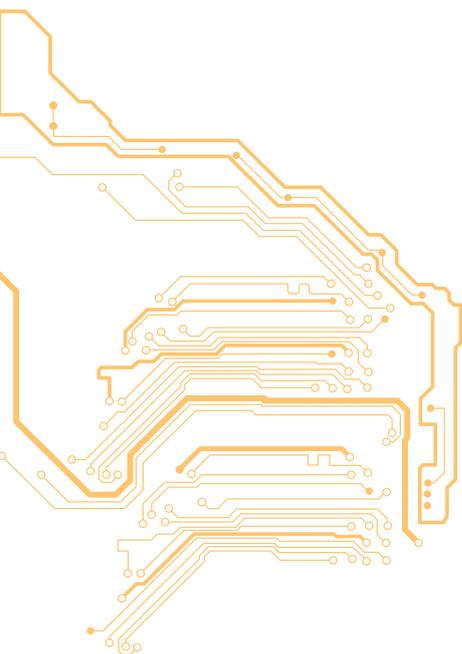
Gracias a la información que permiten extraer, podemos tomar mejores decisiones, controlar los cuellos de botella y medir el rendimiento de la entidad.

Otra de las facetas de las administraciones públicas en la que podemos sacar provecho de los cuadros de mando es la monitorización de recursos de ayuntamientos, diputaciones, parlamentos, etc. Gracias a la información que permiten extraer, podemos tomar mejores decisiones, controlar los cuellos de botella y medir el rendimiento de la entidad. Para ilustrar este ejemplo vamos a utilizar una herramienta, desarrollada por SEDIPUALB@ en colaboración con el grupo de investigación IDAL, encargada de ayudar a visualizar los recursos de las entidades con las que colabora. Pero antes de empezar con el cuadro de mando, aclaremos algunos conceptos.

**SEDIPUALB@** es una plataforma pública que permite realizar una gestión completamente digital de toda la actividad administrativa de las entidades de ámbito público. A continuación, presentamos un cuadro de mando que monitoriza los expedientes de un ayuntamiento que mantendremos como anónimo por protección de datos. Un expediente representa cualquier trámite administrativo gestionado a través de la entidad en cuestión, desde una solicitud de alta de padrón de habitante hasta una ampliación de crédito para un proyecto.

El cuadro de mando ha sido diseñado para ser lo más flexible posible. Por una parte, debe ser accesible para un usuario poco experimentado; sin embargo, si se desea, debe permitir hacer un análisis tan exhaustivo como se considere. Para cumplir estos dos puntos se ha optado por un diseño sencillo a simple vista, pero que encierra un gran potencial.

En la ilustración 20, tenemos la visión que obtenemos nada más entrar en la herramienta. Como se puede ver, la página está dividida en dos partes. La zona derecha, que contiene dos gráficas (la serie temporal de cantidad de expedientes y la del tiempo de gestión) y la zona izquierda, que contiene el título, un pequeño resumen de todos los expedientes que se han registrado en la plataforma, un rectángulo rojo (que de ahora en adelante llamaremos carta) y unos botones de control en la parte inferior.



### Expedientes

T. med. resolución: 112 días 8 horas  
 Total: 299206  
 Abiertos: 68202  
 Cerrados: 231004

**Seleccione los contenedores:**

Buscar...

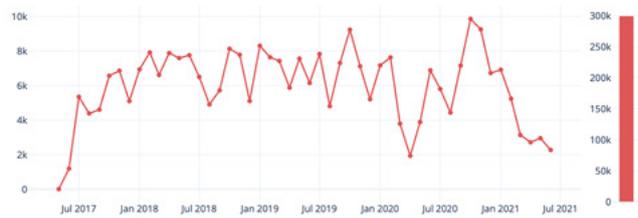
**Seleccione los procedimientos:**

Buscar...

**Seleccione el periodo:** 01/01/2018 - 31/12/2018

	T. med. resolución	Total	Abiertos	Cerrados
Se inician	233d 14h	299206	68202	231004
Se finalizan	172d 1h	21	0	21

### Cantidad de expedientes



### Tiempo de gestión



Inicio Finalizar +

Ilustración 20

Estas dos partes cumplen funciones bien distintas, pero estrechamente relacionadas. En la parte izquierda, seleccionamos el conjunto de datos sobre el que tenemos interés, y en la parte derecha visualizamos todas las características de este conjunto de datos. Así pues, empezamos entendiendo cómo se selecciona la información a visualizar. Para ello, centrémonos en la carta roja de la parte izquierda.

**Seleccione los contenedores:**

Buscar...

**Seleccione los procedimientos:**

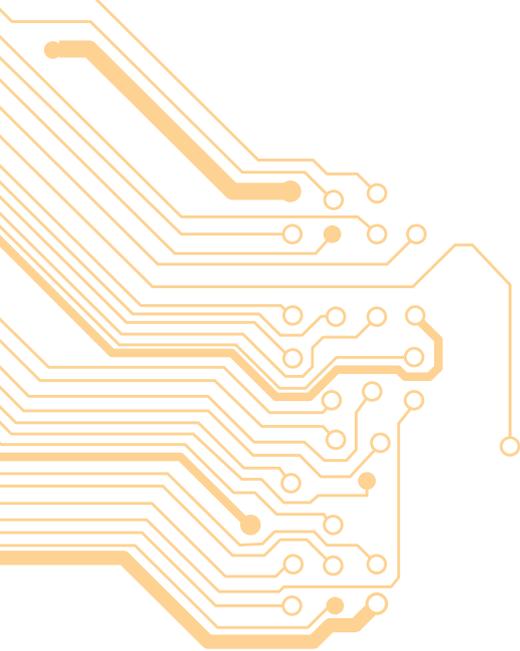
- × SALUD LABORAL - ACCIDENTES LABORALES
- × SALUD LABORAL - PREVENCION RIESGOS LABORALES

**Seleccione el periodo:** 01/01/2018 - 31/12/2018

	T. med. resolución	Total	Abiertos	Cerrados
Se inician	985d 3h	591	396	195
Se finalizan	172d 1h	21	0	21

Ilustración 21





Como podemos ver, esta contiene distintos selectores que nos permiten filtrar el conjunto total de expedientes. Por ejemplo, en la ilustración 21 estamos eligiendo los accidentes laborales y los expedientes relacionados con la prevención de riesgos laborales que se registraron en 2018. Cabe destacar que como el primer selector no tiene nada seleccionado, simplemente se ignora. Como podemos ver, en la parte inferior de la carta hay un resumen de los expedientes escogidos. Si nos fijamos, veremos que la fila correspondiente a los expedientes que finalizaron en 2018 está un poco traslúcida. Esto se debe a los botones de la parte inferior.



Ilustración 22

En la parte inferior izquierda del cuadro de mando hay 3 o 4 botones, dependiendo del escenario, ya que el botón de aplicar puede desaparecer como medida de seguridad. Empezando por la izquierda, tenemos un botón doble en el que seleccionamos si el rango de tiempo que hemos determinado en la carta filtra los expedientes que se inician o finalizan en el período seleccionado. Esta implementación sirve para que si, por ejemplo, estamos considerando un conjunto de contratos de un determinado período de tiempo, nos podamos centrar tanto en los contratos que se iniciaron en ese período como en los que finalizaron.

**Con lo que hemos descrito, en realidad, ya tendríamos un cuadro de mando suficiente. Podríamos seleccionar en la carta roja el conjunto de datos que queremos visualizar en el período que nosotros queramos y mediante los botones de la parte inferior podríamos decidir si centrarnos en los expedientes que empiezan o acaban en el período indicado. Después, pulsaríamos el botón de aplicar y en la parte derecha podríamos visualizar todas las características que quisiéramos del conjunto seleccionado. Pero a todo esto que hemos descrito le falta una parte importante, la capacidad de comparar.**

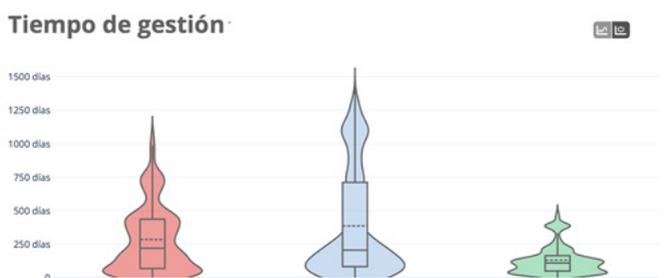
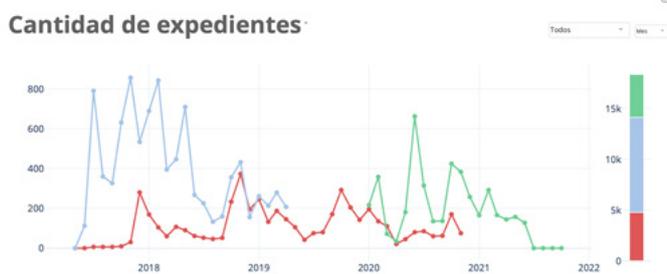
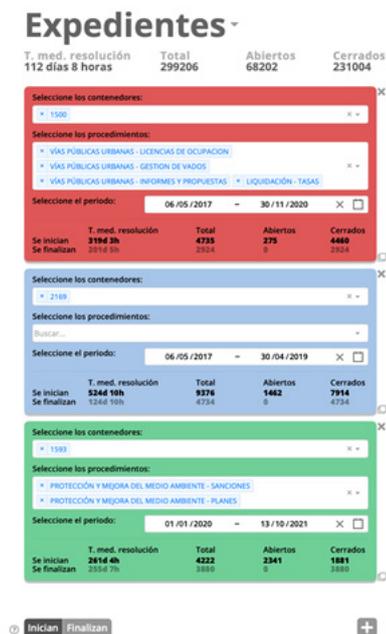


Ilustración 23

Esta capacidad de comparar viene dada por la posibilidad de crear múltiples cartas. Si pulsamos el botón en forma de cruz, en la parte derecha de la ilustración 22, añadimos una nueva carta que nos permite filtrar un nuevo conjunto de expedientes. Como se puede ver en la ilustración 23, esto permite que podamos visualizar simultáneamente distintos conjuntos de información para así poder comparar y hacer un análisis más profundo.

Si nos fijamos en la parte inferior izquierda de la ilustración 23, podemos ver que en la zona de los botones no está el botón de aplicar. Esto es debido a que el botón de aplicar solo aparece cuando la selección de la zona de las cartas no se corresponde con la información que muestran las gráficas de la derecha. Esto es una medida de seguridad para evitar posibles confusiones, ya que, como se puede ver en la ilustración 24, al eliminar el rango temporal de la carta azul la parte izquierda deja de corresponderse con la parte derecha hasta que volvamos a pulsar el botón de *aplicar*. Para hacer que esta situación sea evidente, la parte de las gráficas se difumina y aparece el botón de *aplicar*.

## Expedientes

T. med. resolución  
112 días 8 horas

Total  
299206

Abiertos  
68202

Cerrados  
231004

Selección de los contenedores: 1590

Selección de los procedimientos:  
- VÍAS PÚBLICAS URBANAS - LICENCIAS DE OCUPACIÓN  
- VÍAS PÚBLICAS URBANAS - GESTIÓN DE VADOS  
- VÍAS PÚBLICAS URBANAS - INFORMES Y PROPUESTAS - LIQUIDACIÓN - TASAS

Selección de el periodo: 06/05/2017 - 30/11/2020

T. med. resolución	Total	Abiertos	Cerrados
319d 3h	4735	275	4460
23h 6 3h	2324	0	2324

Se Inician Se Finalizan

Selección de los contenedores: 2189

Selección de los procedimientos: Buscar...

Selección de el periodo: dd/mm/aaaa - dd/mm/aaaa

T. med. resolución	Total	Abiertos	Cerrados
485d 7h	11587	2150	9437
55d 6 3h	9437	0	9437

Se Inician Se Finalizan

Selección de los contenedores: 1583

Selección de los procedimientos:  
- PROTECCIÓN Y MEJORA DEL MEDIO AMBIENTE - SANCIONES  
- PROTECCIÓN Y MEJORA DEL MEDIO AMBIENTE - PLANES

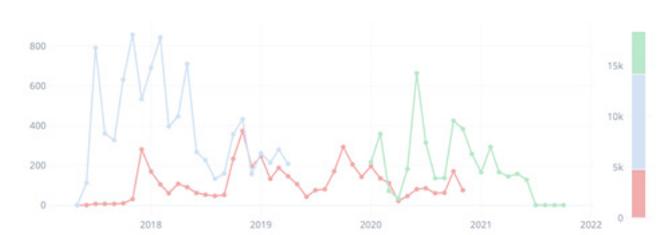
Selección de el periodo: 01/01/2020 - 13/10/2021

T. med. resolución	Total	Abiertos	Cerrados
281d 4h	4222	2341	1881
255d 7h	3300	0	3300

Se Inician Se Finalizan

Inician Finalizan Aplicar +

## Cantidad de expedientes



## Tiempo de gestión

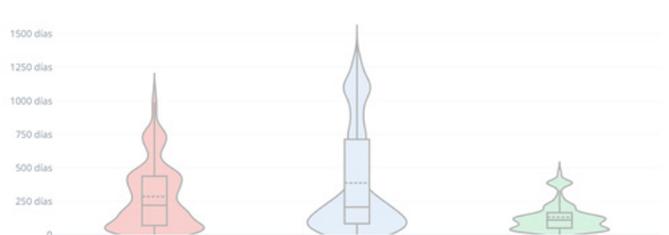


Ilustración 24

Por último, volvamos a la ilustración 21 para entender qué son los botones de la parte derecha de la carta. Si nos fijamos, se pueden apreciar dos iconos: el de arriba, en forma de equis, y el de abajo, en forma de dos cuadrados superpuestos. Como podemos suponer, el de arriba sirve para eliminar la carta en cuestión. Pero es el de abajo, el botón de duplicar la carta, el que merece nuestro interés. Este botón es otra forma de añadir una nueva carta con la diferencia de que, si ya hemos hecho una selección compleja de expedientes en una carta y queremos comparar ese conjunto de información con un conjunto de expedientes similar, pero que cambia algunos matices, podemos evitarnos empezar una carta de cero simplemente duplicando la carta original y modificando los matices necesarios.

Ahora que ya entendemos las distintas herramientas de las que disponemos para seleccionar y comparar conjuntos de expedientes, centrémonos en la parte derecha, es decir, en las visualizaciones a las que tenemos acceso.

Si pensamos en el tipo de dato que es un expediente, nos damos cuenta de que este puede ser tan variado que hay muy pocos aspectos comunes a todos ellos. Por ejemplo, puede que algunos expedientes tengan asociado un presupuesto, pero habrá un alto porcentaje que no lo tenga. En las representaciones visuales queremos ver las características que comparten todos los expedientes. Por esto, como vamos a ver, las gráficas solo se centran en los dos aspectos comunes a todos los expedientes: el tiempo de gestión y la cantidad.

## Cantidad de expedientes

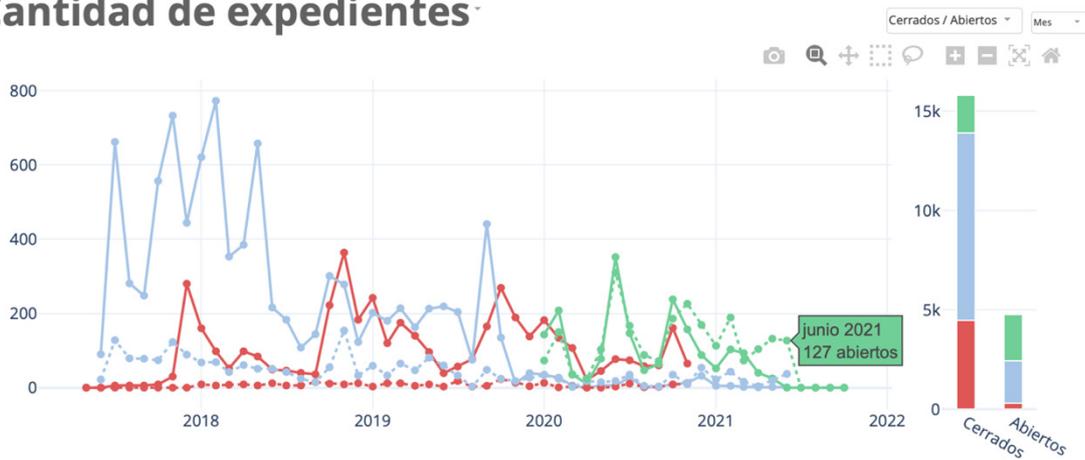


Ilustración 25

En la ilustración 25, podemos ver la primera de las representaciones visuales formada por dos gráficas: una serie temporal y un gráfico de barras apiladas. En esta representación visualizamos la cantidad de expedientes que hay en cada una de las cartas y su distribución a lo largo del tiempo. Como podemos observar, se utilizan distintos colores, pero no hay una leyenda. Esto es porque las cartas de la parte izquierda del cuadro de mando, con sus colores, cumplen esta función.

Si nos fijamos en la esquina superior derecha, podemos ver un par de selectores. En el primero de ellos, podemos seleccionar qué expedientes queremos visualizar: cerrados, abiertos, todos o cerrados y abiertos a la vez. En el segundo selector, decidimos cómo queremos agrupar los puntos de la serie temporal, es decir, podemos elegir que cada punto represente la suma de los expedientes de un mes, una semana o un día. Además, todas las gráficas que vamos a ver en este cuadro de mando son interactivas. Esta interactividad vendrá dada normalmente por cuadros emergentes que nos permiten extraer información por donde pasamos el ratón. Por ejemplo, en la ilustración 25 podemos ver un cuadro verde que nos indica que en junio de 2021 se iniciaron 127 expedientes de la carta verde que siguen abiertos.

## Tiempo de gestión

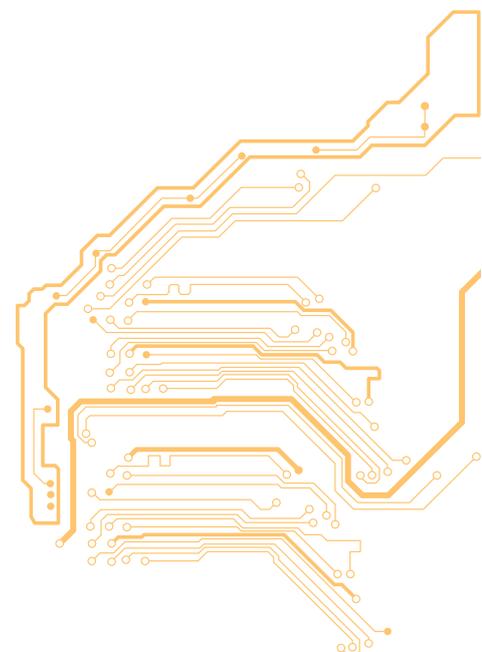


Ilustración 26

La siguiente gráfica que vamos a analizar, que podemos ver en la ilustración 26, es la encargada de representar los tiempos de gestión de los expedientes. Como se puede apreciar, aquí también tenemos una serie temporal como en la ilustración 25, solo que en este caso los puntos representan el tiempo medio de gestión de los expedientes y únicamente se consideran los expedientes cerrados.

Otro de los cambios que hay implementados en esta gráfica es la presencia del área gris de la esquina superior derecha. Esta área es la encargada de mostrarnos la obligada tendencia decreciente de la gráfica; es decir, si calculamos el tiempo de gestión de un expediente iniciado hace una semana, sabemos que, como máximo, podrá tener un tiempo de gestión de una semana. Sin embargo, si cogemos un expediente iniciado hace un año, sabemos que su tiempo de gestión podría ser de hasta un año. Así pues, el área gris representa el conjunto donde, por definición, no pueden estar los puntos de la gráfica.

En la esquina superior derecha de la ilustración 26 podemos observar un par de botones. Estos dos botones sirven para alternar entre las dos gráficas referentes al tiempo de gestión de los expedientes. El primer botón es para la serie temporal y el segundo botón es para el diagrama de violín. El diagrama de violín, que podemos ver en la ilustración 23, nos permite ver, a la vez, las distribuciones de los tiempos de gestión de los expedientes de las cartas.



## Procedimientos

Cerrados



Ilustración 27

También cabe destacar los *treemaps* del cuadro de mando. Recordemos que en cada carta hay dos selectores; uno para filtrar el contenedor y otro para el procedimiento al que pertenecen los expedientes. En la ilustración 27 se puede observar el treemap que nos permite ver la distribución de los procedimientos de cada carta. Si nos fijamos, se pueden ver tres marcos con los colores de las cartas: azul, rojo y verde. Y dentro de cada uno de estos marcos, podemos ver por tamaños y escala de colores los procedimientos. El tamaño viene dado por la cantidad de expedientes que contiene ese procedimiento, y el color, por el tiempo medio de gestión. Para los contenedores se sigue la misma estrategia.

Ahora bien, si volvemos a la ilustración 20 y nos fijamos en la parte derecha del cuadro de mando, podríamos pensar que solo caben dos representaciones gráficas; una arriba y otra abajo. Pero esto no es así, ya que cada título es, además, un selector que nos permite escoger la representación gráfica que nosotros queramos. Así pues, podemos elegir las visualizaciones de la parte derecha a nuestro gusto.

Por último, cabe destacar la modularidad de la página. La estructura del cuadro de mando que hemos descrito es fácilmente adaptable a un tipo de dato similar a los expedientes. Por ejemplo, si nos quisiéramos centrar en los contratos de una entidad, podríamos simplemente añadir algunas opciones a las cartas, como los adjudicatarios o el tipo de contratación, y añadir algunas gráficas útiles como las relacionadas con el importe de los contratos. En general, la página que hemos descrito, caracterizada por el sistema de cartas para seleccionar la información y la parte derecha para seleccionar las visualizaciones convenientes, es fácilmente exportable.



## REFERENCIAS

- Instituto Nacional de Estadística, Clasificación del gasto de las Administraciones Públicas por funciones COFOG (publicado el 23 de diciembre de 2020), [https://www.ine.es/daco/daco42/cne15/aappgastcofog95\\_19.xlsx](https://www.ine.es/daco/daco42/cne15/aappgastcofog95_19.xlsx)
- <https://www.mptfp.gob.es/portal/prensa/actualidad/noticias/2021/06/20210630.html>
- “Desarrollan una nueva herramienta para facilitar la vigilancia de la COVID-19 a tiempo real”, <https://www.ciberesp.es/noticias/desarrollan-una-nueva-herramienta-para-facilitar-la-vigilancia-de-la-covid-19-a-tiempo-real>

# 3

## Plataforma basada en inteligencia artificial para el análisis de imágenes médicas



Ana Jiménez-Pastor  
Angel Alberich-Bayarri  
Eduardo Camacho-Ramos  
Eduardo Ibor Crespo  
Fabio García Castro  
José Sánchez-García

# 1

## INTRODUCCIÓN



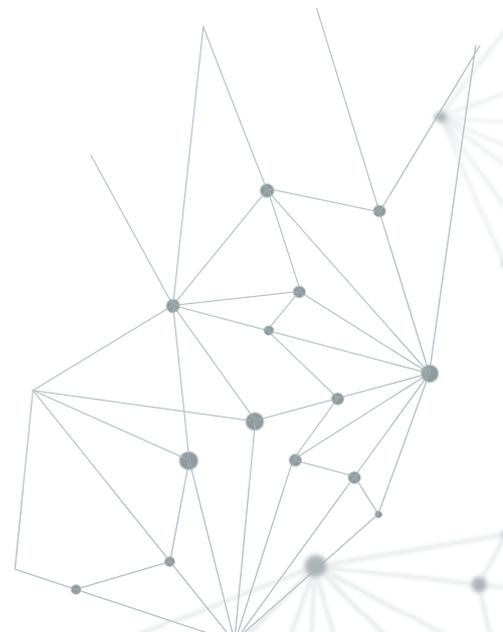
Los biomarcadores de imagen son características objetivamente medibles a partir del procesamiento de imágenes médicas. Para ser considerados biomarcadores y no simples características o parámetros, deben comportarse como indicadores de procesos fisiológicos y patológicos que suceden en el cuerpo humano, y son también de gran ayuda en el estudio de la respuesta terapéutica a un fármaco. Se resuelven en el espacio (podemos evaluar su distribución vóxel a vóxel en mapas paramétricos) y en el tiempo (podemos analizar sus cambios longitudinales). En los últimos años, se ha demostrado que los biomarcadores de imagen ofrecen información complementaria muy útil al diagnóstico radiológico tradicional, que hoy en día se basa principalmente en la observación visual y no en la cuantificación (Mazón et al.,2018).

El desarrollo de un biomarcador de imagen implica definir su relación con la realidad objetiva (estructural, fisiológica, biológica o molecular) y controlar su validez técnica con los criterios de evaluación finales. Los principales resultados que deben predecirse están relacionados con el fenotipado, la elección del tratamiento, la respuesta al mismo y la supervivencia. El aprendizaje automático y el procesamiento de imagen permiten extraer distintos tipos de información, desde segmentaciones automáticas de órganos o detección de lesiones, como por ejemplo lesiones tumorales (Salad et al., 2021), hasta diferentes medidas, como puede ser la cuantificación de grasa y hierro hepáticos a partir de resonancia magnética (RM). La combinación de estas técnicas para la obtención de biomarcadores de imagen, junto con algoritmos de inteligencia artificial (IA), ayudan a correlacionar los resultados cuantitativos con el posible diagnóstico y pronóstico del paciente. Estas técnicas necesitan un vehículo para poder utilizarse en la práctica



clínica y ensayos en que la idea principal es mantener el flujo de trabajo, automatizar los diferentes procesos y proporcionar valor.

Hoy en día existen múltiples soluciones para el análisis de la imagen médica radiológica; sin embargo, muchas de ellas se encuentran centradas en una única patología y hacen que los servicios de radiología deban integrar y familiarizarse con múltiples interfaces gráficas, resultando ineficiente para el flujo radiológico. Por lo tanto, se propone una plataforma que integre métodos de análisis de múltiples áreas anatómicas para que todos los miembros de un servicio de radiología puedan hacer uso de la misma solución.



# 2

## NECESIDADES



Las enfermedades y los tratamientos producen cambios sutiles en órganos y tejidos que, en ocasiones, permanecen ocultos a simple vista en la imagen radiológica, especialmente en los estados más tempranos de la enfermedad. Todo esto puede retrasar el diagnóstico de la enfermedad y hacer que se detecte en estadios más avanzados. Existen numerosos ejemplos de este retraso diagnóstico en diferentes escenarios clínicos: neurología, reumatología, oncología, entre otros. Además, en el sector de la industria farmacéutica, determinados ensayos clínicos se alargan en el tiempo por la falta de detalle de los objetivos primarios y secundarios seleccionados en la actualidad para evidenciar la respuesta al tratamiento. Por otro lado, el informe radiológico sigue realizándose de manera cualitativa, lo que dificulta la gradación de la enfermedad de manera objetiva. Todo ello, unido a las desviaciones de los protocolos de imagen y los retrasos en la transferencia de datos, podría dar lugar a fallos en el diagnóstico y ensayos clínicos costosos e ineficientes. Por tanto, existe la necesidad de cuantificar los hallazgos de forma objetiva, automática y reproducible. Las técnicas avanzadas de visión por computador unidas a modelos de IA pueden aportar grandes beneficios, automatizando tareas laboriosas o proporcionando información que no puede obtener el ojo humano, como la segmentación de áreas anatómicas o hallazgos radiológicos, el cribado de estudios o la cuantificación. Estos algoritmos pueden ser encapsulados en diferentes módulos de análisis orientados a patologías o áreas anatómicas específicas, y, a su vez, estos módulos pueden ser ofrecidos tanto a la comunidad científica como a la clínica por medio de una plataforma basada en IA para el análisis de imágenes médicas.



Esta plataforma, pese a contar con la opción de ser instalada en un centro de forma local, no debe estar pensada para ser instalada en un único dispositivo, sino que debe ser accesible desde cualquier lugar por medio de un ordenador, aprovechando la escalabilidad que ofrecen las infraestructuras en la nube. La tecnología en la nube para la centralización de la información ha tenido un gran crecimiento en los últimos años, que se ha acrecentado por la pandemia y ha generado nuevas necesidades en el contexto actual para la búsqueda de alternativas que permitan responder a la demanda tecnológica. La gran capacidad de almacenamiento, la rapidez y la inmediatez son algunas de las ventajas que han contribuido a mejorar la atención médica, y brindan la posibilidad de realizar proyectos multicéntricos, involucrando hospitales, clínicas, investigadores y personal médico de todo el mundo, y conectándolos entre sí. Por otro lado, también se orientaría a la teleradiología, y ello permitiría visualizar, consultar y realizar un diagnóstico a distancia. Esto favorece la productividad, en circunstancias de aumento del número de imágenes para diagnóstico, complementar la escasez de médicos radiólogos para interpretarlos o la inmediatez necesaria para el diagnóstico. Por tanto, una plataforma de análisis plantea nuevos escenarios de trabajo, de colaboración y de cooperación entre profesionales.

# 3

## OBJETIVO



Una plataforma de análisis es el nexo que permite llegar a pacientes de todo el mundo reduciendo los costes económicos, humanos y de tiempo, y pretende aumentar la objetividad y la importancia de los datos resultantes del procesamiento de imágenes para proporcionar diagnósticos mucho más precisos y tempranos. El objetivo principal es predecir y anticipar los diagnósticos para mitigar los efectos de determinadas enfermedades y evaluar precozmente la respuesta a los tratamientos.

Con ella se pretende cubrir la necesidad del cálculo de biomarcadores de imagen médica mediante el procesamiento avanzado de imágenes radiológicas dentro de dos áreas principales: en los ensayos clínicos de nuevos medicamentos, que implican estudios con imagen médica; y en los servicios de radiología de hospitales y centros de diagnóstico, con el objetivo de mejorar la salud de los seres humanos detectando los cambios producidos por las enfermedades y los medicamentos en el organismo.

Además, esta plataforma no debe permitir únicamente el análisis de la propia imagen, sino ofrecer un espacio para la centralización de la información de los pacientes, desde la información clínica hasta la genómica, pasando por analíticas y resultados de anatomía patológica, y esta funcionalidad resulta de especial relevancia en la ejecución de proyectos multicéntricos y multiómicos en los que hospitales e investigadores de diferentes instituciones deben trabajar sobre la misma información.

# 4

## DESCRIPCIÓN TECNOLÓGICA

### ESTRUCTURA DE LA PLATAFORMA

Los servicios necesarios que debe ofrecer una plataforma de análisis de imagen pueden ser cubiertos por tres elementos principales. En primer lugar, la interfaz gráfica, donde el usuario puede interactuar con la plataforma mediante la subida de estudios de imagen, visualizarlos, analizarlos o revisar los resultados obtenidos. En segundo lugar, el almacenamiento, donde encontramos todos los archivos con los que la plataforma interactúa, como las imágenes, los archivos de configuración y los resultados de los análisis. Y, en tercer lugar, la base de datos, que representa el lugar donde la aplicación almacena la información estructurada, como la información del paciente, las cabeceras DICOM, las rutas de los archivos o los formularios con datos adicionales de los pacientes. Para la integración de la plataforma sin interferir en el flujo de trabajo clínico, los módulos basados en IA deben actuar de manera automática, asegurando que esta integración sea lo más fluida posible (figura 1).

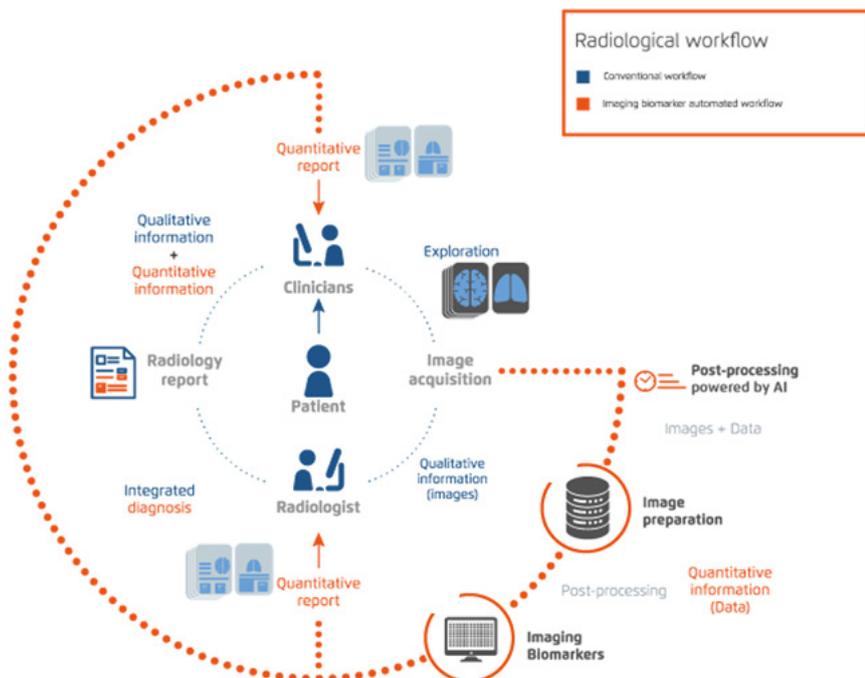
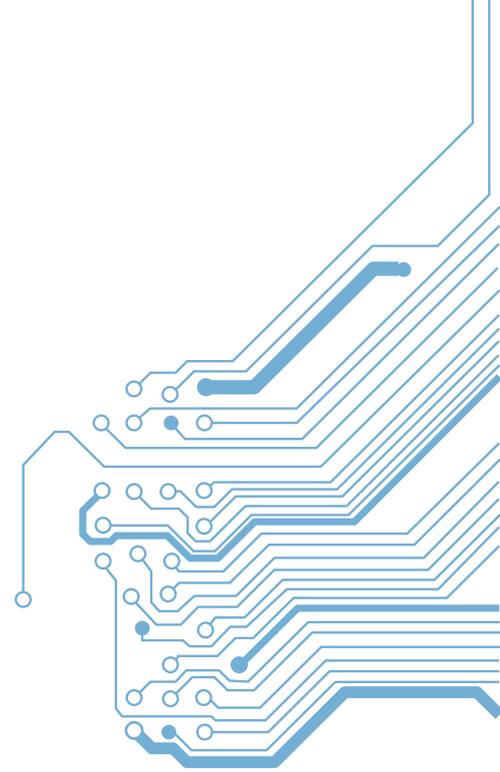


Figura 1: flujo del sistema



Las bases de datos tienen como objetivo gestionar todos los datos almacenados, facilitando su manejo, mejorando su organización y evitando la redundancia. Por tanto, su importancia reside en la gestión adecuada de los mismos para simplificar y agilizar procesos, aumentando la productividad. Estas bases pueden ser tanto relacionales como no relacionales, ambas con sus ventajas e inconvenientes. Para la descripción de esta plataforma, la base de datos expuesta como caso práctico es una base de datos no relacional llamada MongoDB.

MongoDB utiliza una organización orientada a objetos en lugar de un modelo relacional y emplea una colección de documentos como unidad básica de datos, que se almacena en un formato JSON (*JavaScript Object Notation*). MongoDB es conocida por su alta escalabilidad tanto en almacenamiento como en rendimiento, y es capaz de gestionar grandes cantidades de datos, sin comprometer el acceso a los mismos, gracias a la distribución de los conjuntos de datos en múltiples máquinas virtuales. Además, no requiere esquemas predefinidos, por lo que los registros de datos pueden ser creados, recuperados y editados por el usuario con la máxima flexibilidad.

La plataforma permite al usuario encontrar y descargar estudios, buscando directamente en el PACS mediante operaciones DICOM de *Query/Retrieve* en las instalaciones locales. DICOM (*Digital Imaging and Communications in Medicine*) es un protocolo estándar para la gestión y la transmisión de imágenes médicas y datos relacionados, desarrollado

originalmente por la National Electrical Manufacturers Association (NEMA) y el American College of Radiology (ACR) (Bidgood W. D., 1997). Además, es el estándar internacional para comunicar y gestionar imágenes y datos médicos con la misión de garantizar la interoperabilidad de los sistemas utilizados para producir, almacenar, compartir, mostrar, enviar, consultar, procesar, recuperar e imprimir imágenes médicas, así como de gestionar los flujos de trabajo relacionados. La operación *Query/Retrieve* permite buscar estudios desde el PACS mediante los diferentes filtros como son nombre del paciente, ID del paciente, ID del estudio de imagen, descripción del estudio, modalidad del estudio o fecha de adquisición, y muestra una lista de estudios que cumplen los criterios establecidos. Una vez detectado el estudio de interés, este debe descargarse del PACS y almacenarse en la plataforma de análisis de imagen.

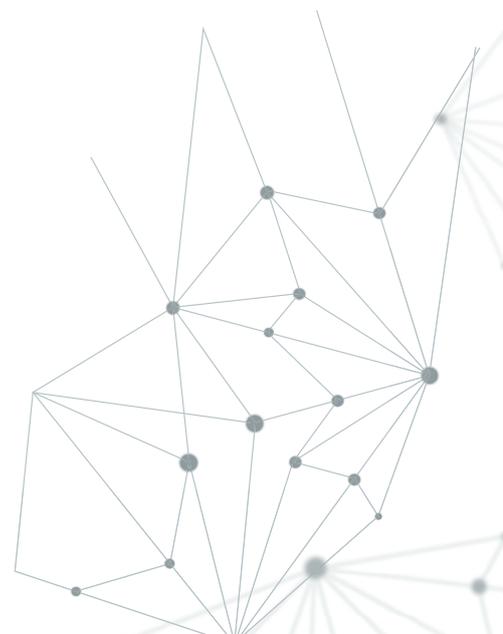
La plataforma también requiere un *software* que actúe como conector DICOM y que permita la comunicación bidireccional entre esta y la estación

de trabajo, y así no solo ser capaz de obtener los estudios, sino también de integrar cualquier módulo de análisis dentro del flujo clínico de un centro. La recuperación automática de estudios específicos del PACS se realiza por medio del motor de reglas siempre que se cumplan una serie de criterios predefinidos. Estas reglas se basan en establecer criterios de filtrado de cabeceras DICOM específicas que deben contener o deben ser iguales a unos valores concretos. Además de los filtros, cada regla define una frecuencia de consulta y acciones a realizar cuando un estudio coincide con los criterios de filtrado. El motor de reglas supervisa continuamente el PACS según la frecuencia de consulta definida y, cuando se cumplen los criterios de filtrado, activa la recuperación del estudio y las acciones adicionales definidas para la regla. Todo ello es configurado por el personal técnico en la fase de instalación y puede ser modificado posteriormente.

**Para proceder al análisis de imagen, la plataforma debe ofrecer la posibilidad de realizar la subida de estudios de imagen. En el proceso de carga, el usuario puede cargar las imágenes directamente desde un directorio local, sólo con una acción de arrastrar y soltar. Las imágenes DICOM, además de contener la propia imagen, incluyen una serie de metadatos, entre los que encontramos información relacionada con el paciente, el centro donde se ha realizado el estudio, los parámetros empleados durante la adquisición de la imagen, entre otros. Por lo tanto, durante el proceso de subida de imágenes, especialmente cuando se almacenan en la nube, estas imágenes deben anonimizarse mediante la eliminación de cualquier metadato de carácter personal y que permita la reidentificación del paciente. Así, el usuario debe introducir un identificador que será empleado para la reidentificación del paciente dentro de la plataforma.**

Adicionalmente, puede que otro tipo de información acerca del estudio o del paciente sea relevante y, por tanto, deba almacenarse en la base de datos. De este modo, se ofrece la posibilidad de rellenar un formulario de datos (DTF - *Data Transmittal Form*), adaptable a cada proyecto según las necesidades y los intereses específicos.

La plataforma cuenta también con una herramienta de visualización de imágenes en formato DICOM. Este visor, además de la propia visualización de las imágenes, permite la anotación de las mismas mediante la delineación de regiones de interés o segmentaciones. Estas regiones de interés serán empleadas tanto para el entrenamiento de los modelos de IA como para delimitar las regiones a analizar



de cara a la ejecución del algoritmo en entornos de producción. Además, permite interactuar con los resultados generados por los módulos de análisis. Esta interacción permite al usuario realizar correcciones sobre los resultados que serán empleadas para el reentrenamiento de los modelos.

En una misma plataforma se incluyen algoritmos para el análisis de cerebro, tórax, hueso y cartílago, abdomen y procesos oncológicos, entre otros. Estos algoritmos se presentan en forma de distintos módulos de análisis que funcionan bajo unas condiciones de entrada específicas definidas y pueden ser lanzados tanto de forma manual desde la plataforma como de forma automática, configurando el motor de reglas para recibir los casos en las condiciones definidas, lanzar el análisis y devolver el resultado para que el radiólogo disponga de dicha información a la hora de realizar el informe.

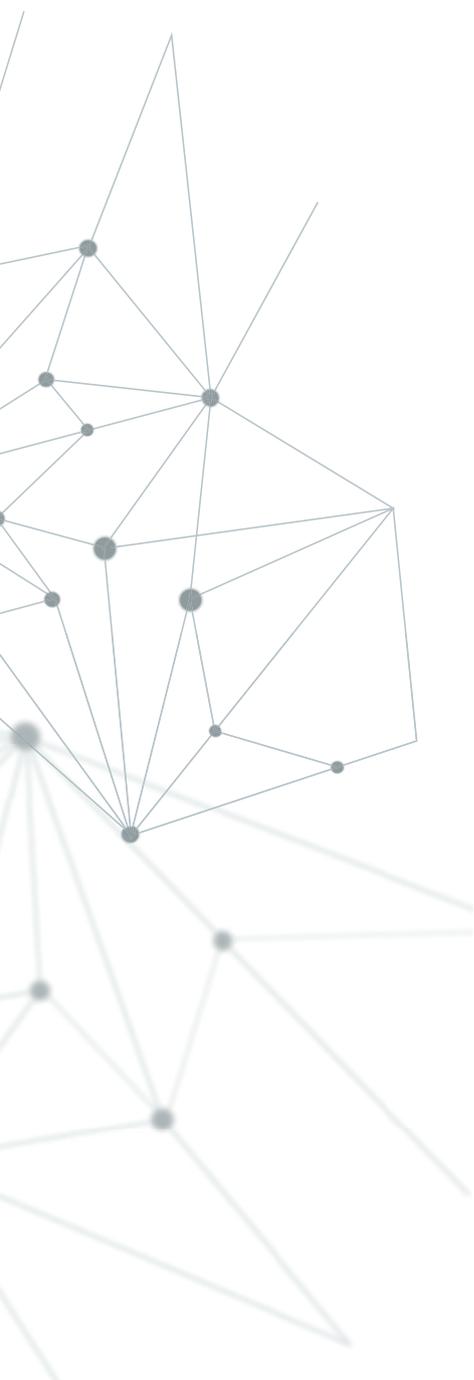
## CONEXIÓN DICOM Y GESTIÓN DE LA INFORMACIÓN

La plataforma cumple con el estándar DICOM tanto en cuanto a comunicación como en cuanto a datos, y recibe los estudios de imagen de los hospitales, o centros radiológicos. A continuación, el sistema analiza las imágenes y obtiene medidas cuantitativas, que se almacenan en una base de datos y se muestran al usuario desde la interfaz. Mediante el uso de un *software* de conexión, la plataforma de imágenes se integra con el PACS del hospital, y se pueden consultar y recuperar estudios médicos, procesarlos, lanzar análisis sobre los mismos y almacenar los informes resultantes como objetos DICOM en el PACS.

Este *software* de conexión permite cerrar la brecha entre los sistemas informáticos de los hospitales y la plataforma, situándose dentro de los hospitales y las clínicas y estableciendo todas las comunicaciones necesarias para la gestión de objetos DICOM. El conector se integra dentro del flujo de trabajo radiológico gracias a sus capacidades para consultar el PACS, recuperar estudios y para configurar los procesos de análisis en un enfoque basado en un motor de reglas. Además, convierte los informes generados en objetos DICOM y los almacena en el PACS como una nueva serie dentro del estudio original utilizando la operación C-STORE DICOM.

Esta plataforma está pensada para ser utilizada por profesionales capacitados, incluyendo radiólogos y oncólogos. En el escenario del entorno clínico, los usuarios no tienen que preocuparse por dónde está el estudio o el informe. En su lugar, estas cuestiones son transparentes, y no es necesario realizar ninguna acción para iniciar





un análisis, dado que el motor de reglas hace precisamente ese trabajo. Una vez terminado el análisis, los usuarios pueden acceder a los resultados tanto desde la interfaz web como desde el PACS. El sistema brinda la posibilidad de lanzar nuevos análisis desde la plataforma o utilizar la información cuantitativa proporcionada por los análisis ya finalizados para rellenar el informe radiológico. Los resultados obtenidos pueden exportarse en diferentes formatos para mayor comodidad del usuario, como PDF para el informe, o Excel y CSV para los biomarcadores extraídos.

### **SEGURIDAD: ACCESO A LOS DATOS Y PROTECCIÓN DE LA INFORMACIÓN**

En relación con los datos, es crítico poder garantizar la privacidad de los usuarios y la protección de sus datos personales siguiendo las normativas existentes, como el RGPD. La anonimización de datos define la metodología y el conjunto de buenas prácticas y técnicas que reducen el riesgo de identificación de personas, la irreversibilidad del proceso de anonimización y la auditoría de la explotación de los datos anonimizados. Para ello, en el paso previo a la subida de imágenes a la plataforma, se realiza el reemplazo de ciertos atributos de las cabeceras DICOM para las imágenes subidas por versiones cifradas, lo que impide la identificación directa del individuo, y se protege así su privacidad.

A la hora de proteger el acceso a los datos confidenciales, es necesario tener en cuenta los tres posibles estados que se pueden presentar: información en reposo, en tránsito y en uso.

La información en reposo se refiere a aquellos datos a los que no se está accediendo, que no se están usando ni procesando y que se encuentran alojados en el sistema de persistencia (en nuestro caso: base de datos y almacenamiento de ficheros). La información en tránsito es aquellos datos que están siendo transferidos de un sistema a otro (p. ej., la transmisión de datos entre la aplicación y la base de datos). Por último, la información en uso representa aquellos datos a los que se está accediendo, que están siendo usados o procesados por la aplicación (p. ej., las imágenes visualizadas a través del visor web).



Además de proporcionar seguridad física y control de acceso a los datos, la securización de los datos en reposo se realiza mediante el cifrado de los mismos, de manera que resulte inviable (o altamente costoso) acceder a ellos mediante ataques de fuerza bruta. Actualmente, el cifrado en reposo también puede ser requerido por la necesidad de cumplir con determinadas normativas y estándares industriales y gubernamentales, tales como HIPAA, PCI y FedRAMP. En este aspecto, existen diferentes niveles de cifrado para este tipo de datos, y los más habituales son en cuanto a disco duro, fichero y base de datos.

La protección de datos en tránsito debe ser obligatoria para cualquier tráfico de red que requiera autenticación o incluya datos que no sean de acceso público. Esto se consigue mediante el cifrado de los datos de extremo a extremo, asegurando que los datos en movimiento se encuentren protegidos durante su transmisión mediante una conexión HTTPS. Los protocolos más habituales para el cifrado de estos datos son SSL (Secure Sockets Layer) o TLS (Transport Layer Security), que permiten la transferencia segura de datos mediante certificado entre servidores o sistemas.

Por último, para la protección de datos en uso es necesario implementar medidas previas al acceso del usuario al contenido. En este estado la información se encuentra más vulnerable, ya que debe encontrarse descifrada para poder acceder a ella, de manera que la seguridad recaerá en mayor medida en el control de la autenticación y autorización de los usuarios. Es necesario aplicar fuertes medidas para la gestión de identidad de los usuarios (incluyendo elementos como el doble factor de autenticación) y herramientas de control de acceso condicional o basadas en el rol (Rol-Based Access Control, RBAC).

# 5

## INTERFAZ DE USUARIO Y HERRAMIENTAS

### ACCESO Y GESTIÓN DE CASOS

Todos los usuarios registrados en la plataforma son gestionados a través de una vista dedicada dentro del panel de administración. Esta vista muestra una tabla con todos los ya registrados y permite ver sus datos, sus casos y gestionar su información. Los usuarios en la plataforma se definen por su nombre, apellidos, nombre de usuario, dirección de correo electrónico y centro, y para cada uno el personal de administración selecciona su rol y su idioma por defecto. Toda la información es editable excepto el nombre de usuario, que permanece constante para cada uno como identificador único.

Todos los casos subidos son importados a la plataforma; sin embargo, cada usuario solo puede acceder a los suyos. Aquel con permisos de administrador puede acceder a todos ellos desde el apartado “administración”, donde puede revisar y evaluar todos los casos. Desde esta vista se pueden comprobar las características de cada uno, como el nombre del estudio, su modalidad, la fecha de adquisición, el número de series e imágenes, la edad y el sexo del sujeto, entre otras. También es posible realizar acciones sobre estos estudios como acceder al entorno de análisis o descargar sus archivos DICOM. Además, también aparece una columna que define el estado actual de los análisis ejecutados. Este estado de ejecución es editable solo por los usuarios administradores para corregir manualmente posibles errores ocurridos durante la ejecución y que no se hayan actualizado correctamente en la vista del caso. Los casos que aparecen en esta tabla pueden agruparse según diferentes niveles de jerarquía para mayor comodidad del usuario a la hora de revisar los estudios.

### VISOR DICOM

Este visor permite navegar por las diferentes imágenes pertenecientes a un paciente, proporciona una visualización de los datos cargados permite, por ejemplo, modificar su contraste o elegir un mapa de colores específico. Dado que los datos pueden ser de 3 dimensiones, se puede



navegar a través de ellos, mostrando sus diferentes cortes y permitiendo reconstruir el resto de vistas. Además, se pueden realizar diversas acciones como acercar o alejar la imagen, ya que algunas estructuras pueden ser demasiado pequeñas para ser vistas correctamente, cambiar la ventana o visualizar los metadatos.

La cuantificación de los datos será posible a través de formas específicas como rectángulo, elipse, o incluso a mano alzada, que permiten guardar y cargar de nuevo regiones de interés (ROI) a través de la API web. Además, ofrece la posibilidad de seleccionar la ventana en la imagen, es decir, la modificación de la anchura de la ventana (el rango de valores de los píxeles a mostrar) y el centro de la ventana (el centro del rango de valores de los píxeles), lo que permite adaptar la vista para mejorar la visualización de zonas anatómicas específicas. Los metadatos se presentan como una superposición de la imagen y proporcionan información de contexto que es útil para entender la imagen y que debe ser rápidamente accesible. Adicionalmente se proporciona una lista de todas las ROI dibujadas, en el caso de que existan o se hayan creado, y permite cambiar la visibilidad de cada una y acceder al corte donde se dibuja mediante una lista de etiquetas predefinidas y con posibilidad de crear otras nuevas. Por un lado, este visor proporciona una herramienta de anotación de casos, y se emplea para crear conjuntos de datos con los que entrenar y desarrollar modelos de IA para segmentación automática, detección o clasificación. Por otro lado, permite la corrección de segmentaciones generadas como salida por los modelos integrados en los módulos de análisis, favoreciendo la implementación de métodos basados en aprendizaje continuo, y la monitorización, impidiendo la degradación de los modelos y adaptándose a las necesidades de los usuarios (figura 2).

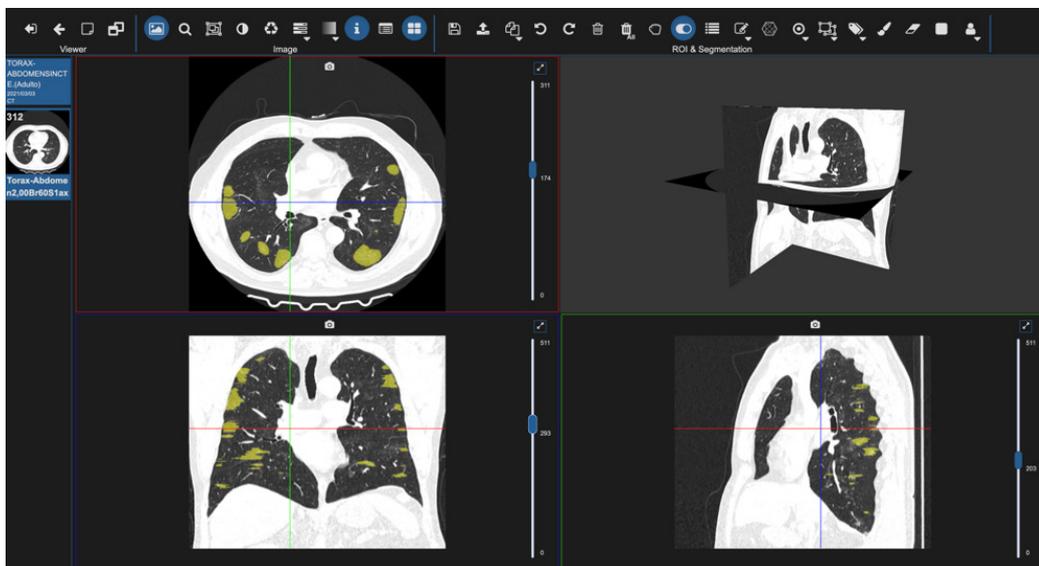


Figura 1: ejemplo de opciones de visualización y edición del visor DICOM. En amarillo, máscaras utilizadas para el etiquetado de lesiones en el pulmón en pacientes de COVID-19.

El sistema permite el etiquetado de los casos mediante plantillas radiológicas de informe estructurado, es decir, un documento que contiene un resumen de la información del paciente, los resultados de los algoritmos lanzados y una conclusión.

## ANÁLISIS

Una vez un estudio está cargado en la plataforma, cualquiera de los módulos integrados en ella puede ser analizado siempre y cuando cumpla las condiciones de entrada necesarias para el algoritmo. Para ello se selecciona el caso, el módulo de análisis que se desea ejecutar y la serie estándar, que será la serie del estudio objetivo para el mismo.

En resumen, una vez seleccionado el módulo, emparejadas las series y definidos los parámetros de entrada, se ejecuta el análisis. Esta acción, por un lado, almacena la información del análisis ejecutado en la base de datos y, por otro, llama al programador de trabajos para ejecutarlo. La barra de estado del caso indica el progreso del análisis, haciendo una estimación del porcentaje completado y, una vez finalizado, los resultados pueden ser revisados desde la propia plataforma (figura 3).

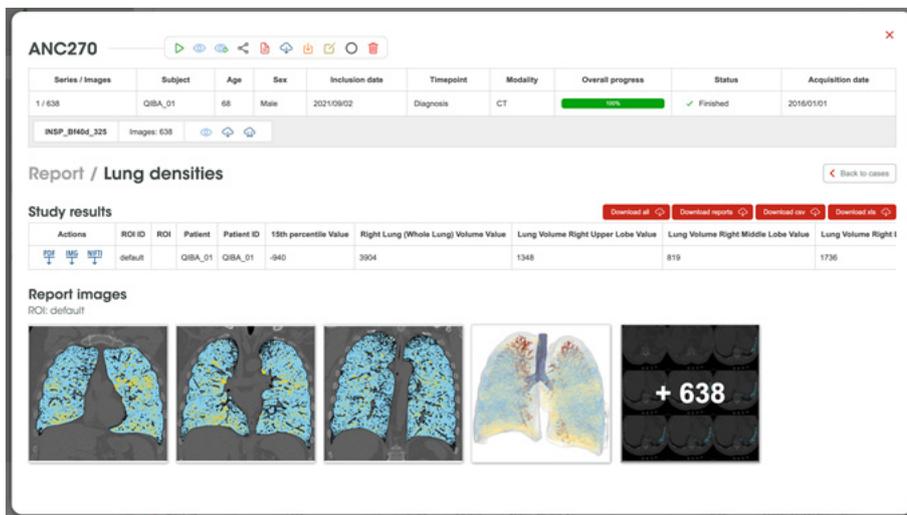


Figura 3: resultados del análisis lanzado sobre el estudio



## INFORME ESTRUCTURADO

La metodología para informar los estudios de imagen radiológica se ha basado tradicionalmente en el uso de texto libre. El texto libre, por definición, impide una estructura regular y concisa en los informes redactados por el radiólogo, lo que aumenta la variabilidad a la hora de informar los hallazgos de un mismo estudio. Sin embargo, si los hallazgos radiológicos y el diagnóstico están dotados de una estructura, conocida y pactada de antemano, permitirá disponer de informes más concisos y completos, a la vez que se mejorará la eficacia en el tiempo. Por lo tanto, para su utilidad en la práctica clínica, los informes radiológicos deben estar estructurados, utilizar una terminología estandarizada y transmitir información procesable. Para generar estos informes, los módulos de análisis anteriormente descritos permiten obtener información cuantitativa que puede ser incluida en el informe radiológico para aportar valor. En definitiva, los informes estructurados pueden conducir a un diagnóstico más rápido al mejorar la comunicación entre radiólogos y clínicos, aumentar la eficacia de los informes, reducir los costes y aumentar los niveles de satisfacción de los clínicos.

Los informes estructurados pueden clasificarse, según su contenido, formato y tipo de datos, en un modelo en texto libre dividido en secciones; un modelo con plantillas con un formato destacado; y un informe estructurado que incluye menús desplegables con preguntas, varias respuestas posibles y léxico estandarizado. La información clínica, técnica y del paciente puede extraerse directamente de las etiquetas de cabecera DICOM del estudio. Además, el informe puede diseñarse para añadir información obtenida de la aplicación de técnicas computacionales a las imágenes, es decir, los biomarcadores de imagen extraídos gracias a los módulos de análisis disponibles. Añadir este tipo de información al informe aporta valor al radiólogo para el diagnóstico y al clínico para decidir el mejor tratamiento o estrategia de seguimiento (figura 4).

The screenshot displays a software interface for a prostate MRI report. At the top, it is titled 'Prostate structured report'. The main section is 'PI-RADS section', which includes several dropdown menus: 'Appearance' set to 'Heterogeneous', 'Lesions and PI-RADS scoring' set to '1', 'Calculation method' set to 'With all sequences', and 'Dominant lesion' set to 'DL 1'. Below these are three anatomical diagrams of the prostate labeled 'Base', 'Middle', and 'Apex', with a red highlight on the 'Base' diagram. A 'Restart' button is located below the diagrams. At the bottom, there is a table with the following data:

Dominant lesion	T2	DWI	DCE	Diameter (mm)	Lesion center	PI-RADS
DL1	2	4	+	21	Peripheral	4

Below the table is the 'TNM classification' section, with '(T) Primary tumor' set to 'T4'.

Figura 4: ejemplo de informe estructurado para un caso analizado por el módulo de análisis de próstata

## MINERÍA DE DATOS

La herramienta de minería de datos (Dataminer) está diseñada para realizar análisis estadísticos sobre datos extraídos del análisis de imágenes, lo que permite a los médicos e investigadores obtener información sobre la enfermedad y detectar problemas clínicos y grupos de interés para la investigación y la explotación científica. Esta herramienta acelera en gran medida el análisis exploratorio de datos al permitir a los usuarios filtrar los datos rápidamente y visualizar las tendencias o los grupos. Además, puede utilizarse con toda la gama de módulos de análisis de imágenes disponibles en la plataforma, y permite visualizar los resultados en diferentes formas, tales como nube de puntos, gráfico de barras (figura 5), histograma, mapa de calor, matriz de correlación, etc. Esta herramienta muestra no solo las variables del informe, sino todos los biomarcadores de imagen que han sido extraídos de los análisis, así como también las variables clínicas que han sido añadidas al estudio, y ello permite una amplia oportunidad de explotación de datos.

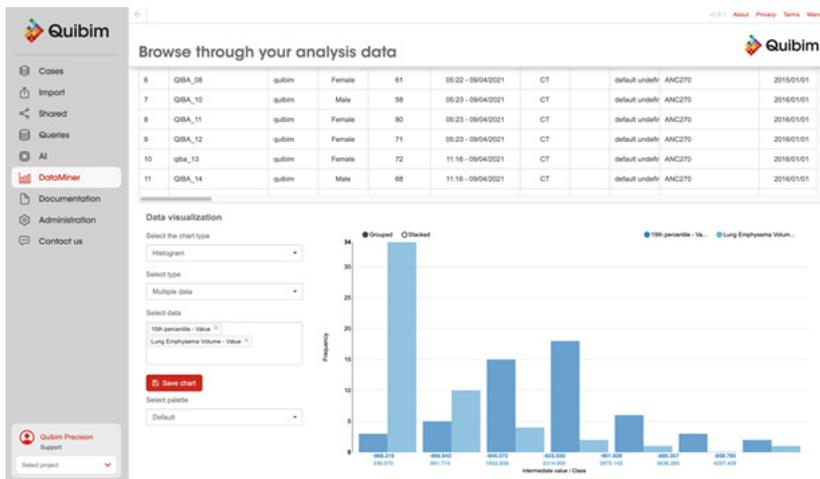


Figura 5: minería de datos para análisis de bajas densidades pulmonares (n = 50)

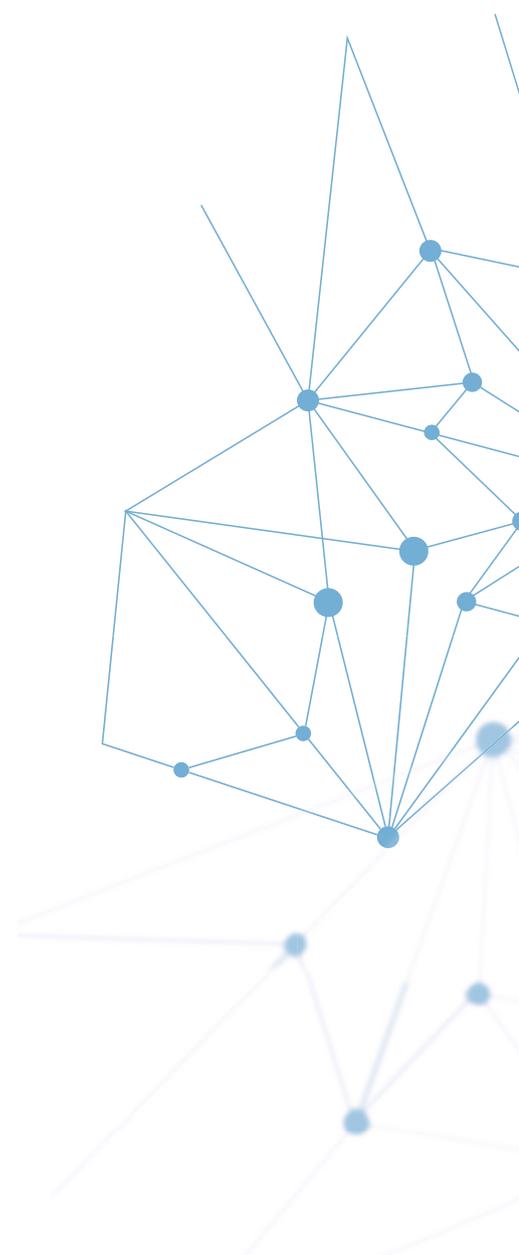
Aunque esta herramienta está disponible para su uso independientemente del número de casos cargados y analizados, cuyo principal objetivo estudiar el comportamiento de los datos o descubrir tendencias en las patologías objeto de estudio, el Dataminer adquiere valor cuando se ha analizado un gran número de casos y se han generado suficientes datos para observar tendencias.



El conjunto de prácticas de DevOps para incorporar un alto nivel de automatización permanente y supervisión constante al desarrollo de las aplicaciones comprende los métodos de implementación, integración y distribución continua (CI/CD). Incorpora la automatización continua y el control permanente en todo el ciclo de vida de las aplicaciones, desde las etapas de integración y prueba hasta las de distribución e implementación (Azeri, 2019). La integración continua (CI) permite que varios desarrolladores trabajen de forma simultánea en diferentes funciones de la misma aplicación; por otro lado, la implementación y la distribución continua (CD) automatiza las pruebas de validación y liberación del código, así como el lanzamiento de la aplicación a producción con los cambios recientes que acaban de pasar esas pruebas.

Por otro lado, es importante resaltar el uso de la tecnología Docker (Boettiger, 2014), que permite crear contenedores Linux usándolos como máquinas virtuales extremadamente livianas y modulares. Su aplicación en sistemas basados en la nube se optimiza gracias a su gran flexibilidad a la hora de implementarlos y otorga la capacidad de ejecutar varios procesos y aplicaciones por separado de forma simultánea, haciendo un mejor uso de toda la infraestructura creada. Docker ofrece un modelo de implementación basado en imágenes, lo que permite compartir una aplicación, en este caso referida a un módulo de análisis, con todas sus dependencias en varios entornos, automatizando la implementación de los conjuntos combinados de procesos que la constituyen en un entorno de contenedores.

Las principales ventajas de esta tecnología serían: en primer lugar, la modularidad, que permite reparar o actualizar partes específicas de una aplicación. En segundo lugar, la facilidad de implementación, que prioriza la eficiencia y reduce el tiempo de desarrollo. En tercer lugar, el control de versiones y su distribución, ya que una imagen docker está formada por varias capas y cada vez que se hace una modificación de la misma, se crea una nueva y se añade al registro de cambios. Y, por último, la restauración, que permite volver a una versión anterior en caso de que la actual presente comportamientos no deseados.





**La rama del MLOps requiere aprender cómo construir un ciclo de vida de aprendizaje automático (Machine Learning) donde, tanto el código como la nueva información, se integran constantemente a una aplicación. Todo ello sin perder de vista que en el centro de dicha aplicación existe un modelo no determinístico y que requiere de conocimiento especializado para ser mantenido.**

Para comenzar a hacer MLOps hay que aceptar la premisa de que una aplicación exitosa de aprendizaje automático involucra muchos más componentes que el modelo estadístico creado. El trabajo como MLOps es el dar soporte para que este modelo sea fácil de reentrenar, validar, desplegar y monitorizar, todo esto con la finalidad de garantizar que su desempeño a través del tiempo se mantenga y siga sirviendo al propósito para el cual fue creado en primera instancia. Las diferentes partes de un ciclo de vida de aprendizaje automático pueden ser resumidas en: gestión de los datos, desarrollo del modelo, puesta en producción y monitorización.

## **GESTIÓN DE DATOS**

Comprende el manejo de bases de datos, la extracción, el almacenamiento de características y la actualización continua de los mismos conforme a lo que sucede en producción. De esta forma, la plataforma recibe nuevos datos, permite etiquetarlos y además realiza análisis que extraen una serie de características (biomarcadores) que describen estos datos, almacenándolos y permitiendo así actualizar los conjuntos de datos continuamente.

## **DESARROLLO DEL MODELO**

Consiste en el entrenamiento y la validación de modelos de IA usando los conjuntos de datos generados a partir de las actualizaciones en la base de datos del sistema. Una parte crucial de esta fase es el control de versiones, es decir, almacenar y versionar los modelos generados, lo que facilita organizar y mantener un seguimiento de los modelos entrenados. Para ello, en la plataforma, estos modelos son incluidos en un contenedor docker permitiendo su portabilidad a cualquier plataforma y definiendo el tag (etiqueta) de la imagen generada con la versión del módulo para mantener la trazabilidad. Estas imágenes son generadas a partir de los repositorios de cada uno de los módulos de análisis basados en Git (A Successful Git Branching Model, 2010) para organización del código, el cual permite la paralelización del desarrollo mediante ramas independientes para la preparación, el mantenimiento y la publicación de versiones del algoritmo, así como soporte para reparación de errores en cualquier momento.

## PUESTA EN PRODUCCIÓN

Esta fase comprende el proceso en el que el modelo de IA desarrollado se integra en una cadena de procesos de análisis junto con otras piezas de *software* para generar un resultado que aporte valor y proporcione la salida en la forma deseada. Además, deben realizarse procesos de testeo a distintos niveles para asegurar que todo funciona correctamente. La plataforma incorpora estos modelos que extraen características específicas de las imágenes junto con otros componentes, como pueden ser modelos que permiten obtener segmentaciones de zonas anatómicas concretas, funciones que obtienen parámetros con métodos de visión por computador tradicionales o funciones que permiten la generación de resultados como reconstrucciones 3D o formateo de variables para ser visualizadas en la plataforma.

## MONITORIZACIÓN

Un modelo de aprendizaje automático captura y simula una versión del mundo y el mundo está en constante cambio, por lo que, con el tiempo, si no se actualizan estos modelos, van perdiendo precisión. Además, no solo hay que prestar atención al modelo de IA, sino también al sistema completo, adaptándose a las necesidades de los usuarios. Para ello, la plataforma, gracias al almacenamiento continuo de nuevos datos y del *feedback* continuo recibido de los usuarios, permite realizar reentrenamientos y modificaciones del sistema periódicos que impiden la degradación de los módulos de análisis.

Cuando se habla de MLOps es necesario entenderlo no como un proceso con un principio y un fin, sino como un ciclo. El objetivo de MLOps es gestionar el despliegue de modelos de aprendizaje automático y aprendizaje profundo en entornos de producción a gran escala. Estos modelos tienen que actualizarse con el tiempo incluso si funcionan correctamente, lo que conlleva más iteraciones en el proceso. Por lo tanto, hay que supervisar los modelos en producción y refrescar el modelo mediante el reentrenamiento si el rendimiento del modelo cae por debajo de un determinado criterio. Por lo tanto, podemos pensar en MLOps como CI + CD (DevOps tradicional) + CT (entrenamiento continuo). De igual forma, también habrá veces en que del desarrollo del modelo haya de regresar a la parte de almacenamiento y procesamiento de datos, u otras en las que el monitoreo motive a revisar la etapa de desarrollo del modelo. En definitiva, se trata de recibir constante retroalimentación para mejorar el producto.



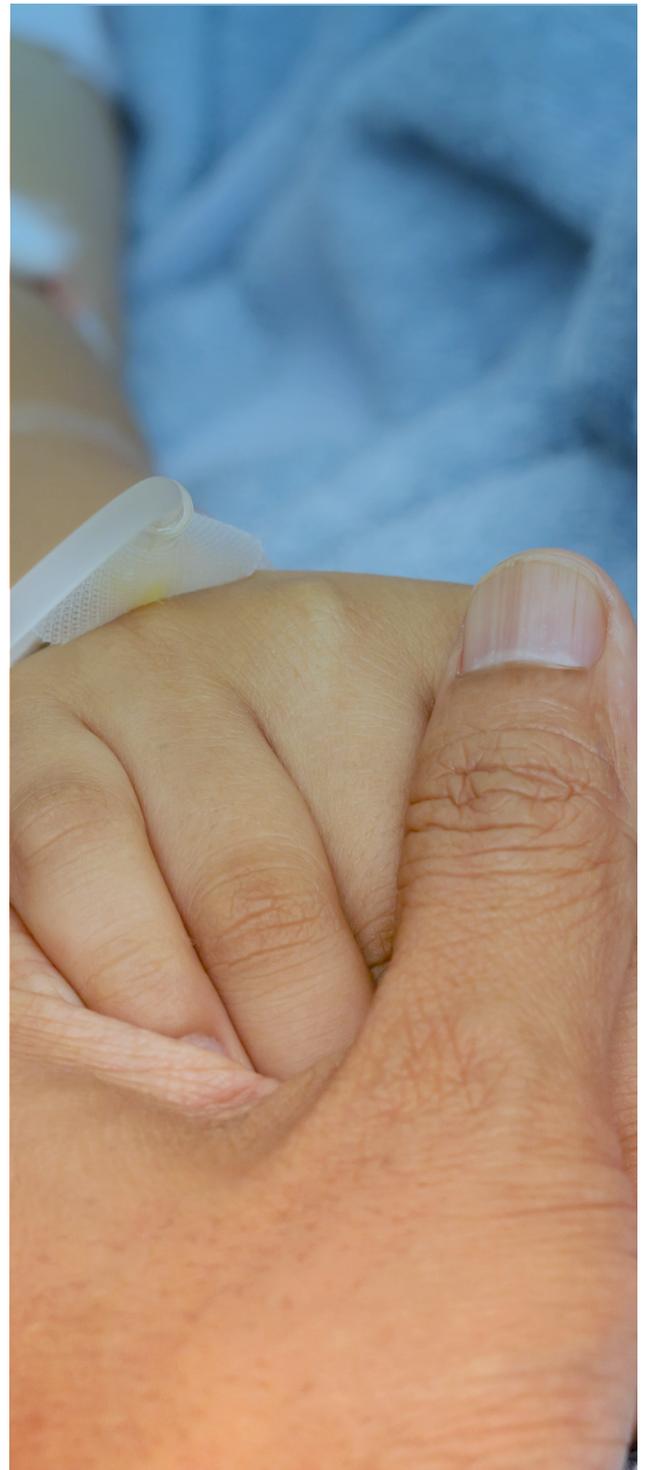
# 7

## CASOS PRÁCTICOS

### PRIMAGE

El proyecto PRIMAGE (Martí et al., 2020) consiste en el análisis predictivo in silico multiescala para apoyar el diagnóstico y el pronóstico personalizado del cáncer, con la utilización de biomarcadores de imagen. Para ello se ha implementado una plataforma en la nube para la centralización de la información asociada a cada paciente y el análisis de dicha información. Concretamente, se están estudiando dos cánceres pediátricos: neuroblastoma y glioma pontino intrínseco difuso (DIPG).

PRIMAGE es uno de los mayores y más ambiciosos proyectos de investigación sobre imágenes médicas, inteligencia artificial y tratamiento del cáncer infantil. Un proyecto de cuatro años financiado por la Comisión Europea y que cuenta con 16 entidades europeas en el consorcio, entre ellos la Sociedad Europea de Oncología Pediátrica. El proyecto se construye como un estudio observacional in silico que incluye conjuntos de datos anonimizados para el entrenamiento y la validación de algoritmos de aprendizaje automático. La plataforma ofrece asistencia clínica precisa para el diagnóstico, el pronóstico y el estudio de la eficacia terapéutica basada en el uso de biomarcadores de imagen y modelos multiescala para la simulación del crecimiento tumoral. Para sacar el mayor partido a toda la información disponible asociada a los pacientes se están desarrollando herramientas de visualización avanzada y



entrenando modelos basados en IA para la predicción de variables como el tiempo de supervivencia o si el paciente va a recaer o progresar en la enfermedad.

Los modelos in silico requieren de importantes recursos informáticos y de almacenamiento de datos para su procesamiento.

La plataforma ofrece una solución informática a medida, combinando HPC

(High Performance Computing) a gran escala y recursos informáticos versátiles de computación en la nube para lograr una eficiencia y fiabilidad óptimas.

La infraestructura se basa en última instancia en un sistema de ayuda al diagnóstico que se diseña para la gestión del cáncer con una funcionalidad y usabilidad avanzadas con un enfoque centrado en el usuario guiado por los clínicos.

Los datos clínicos almacenados son recogidos por los centros y las asociaciones de oncología pediátrica que participan en el proyecto donde, además, esto se realiza bajo procedimientos administrativos y contractuales muy estrictos para garantizar el cumplimiento legal y ético del Reglamento General de Protección de Datos en Europa. Todos los casos incluidos en el repositorio de datos, tanto para el desarrollo de modelos como para la validación de la plataforma, tienen que ser aprobados por los comités de ética de sus respectivos centros.

Al final del proyecto, la plataforma desarrollada (Quibim, 2018) apoyará el diagnóstico, la estratificación del tratamiento y el pronóstico de los pacientes. El prototipo de apoyo a la toma de decisiones se construye y se valida en los cánceres NB y DIPG y los resultados estarán disponibles para la comunidad científica y para hacer transferencia de aprendizaje en otros tumores sólidos malignos. Dada la rareza de estos tumores, la cooperación internacional en red es esencial para aglutinar los datos relevantes del mundo real, para lo cual el nexo de unión es esta plataforma de IA, lo que en última instancia facilita la identificación de herramientas clínicas eficaces.





## IMAGING COVID-19

La plataforma de Imaging COVID-19 fue una iniciativa para generar una gran base de datos multicentro lo suficientemente heterogénea como para representar la situación a nivel global. Esta recopilación de datos permitió desarrollar un modelo de aprendizaje profundo para la detección y la clasificación automatizada de COVID-19 en las exploraciones de tomografía computarizada (TC) y evaluar la gravedad de la enfermedad en los pacientes mediante la cuantificación de la afectación pulmonar. Esta plataforma se presentó como una solución de investigación, abierta y gratuita para la comunidad científica y médica; una plataforma web avanzada para el almacenamiento y la gestión centralizada de imágenes, y su análisis cuantitativo.

El brote de COVID-19 (síndrome respiratorio agudo severo coronavirus 2, SARS-CoV-2) ha pasado por varias fases en el primer año de pandemia, en lo que respecta a los conocimientos clínicos y las pruebas científicas disponibles para la evaluación de la enfermedad. Inicialmente, la enfermedad se identificó como un síndrome similar a la gripe debido a la coincidencia de algunos síntomas como la fiebre y la tos. Posteriormente, se observó un mayor número de reproducciones en los pacientes con COVID-19, así como un mayor número de ingresos en la unidad de cuidados intensivos (UCI), una mayor carga de daños en el parénquima pulmonar y una mayor tasa de mortalidad. Aunque en la actualidad la enfermedad se diagnostica bien en el ámbito clínico mediante la reacción en cadena de la polimerasa de transcripción inversa (RT-PCR), las primeras publicaciones evidenciaban la heterogeneidad y la limitada sensibilidad/especificidad de la técnica (Fang et al., 2020).

Debido a los síntomas respiratorios de la enfermedad, la necesidad de caracterizar la afectación de los pulmones y la accesibilidad mundial a modalidades radiológicas estándar como la radiografía convencional y la tomografía computarizada, la radiología desempeñó un papel fundamental para el diagnóstico rápido, la estadificación y el control de la gravedad del paciente (Piroth et al., 2021).

El principal problema al que se enfrentaba el diagnóstico, más allá de la escasez de recursos, era identificar a los pacientes positivos y diferenciarlos no solo de un paciente sin signos de infección, sino también de una neumonía convencional. En una fase temprana, la infección por COVID-19 presentaba unos hallazgos radiológicos específicos que se perdían a medida que la enfermedad avanzaba y daba lugar a una neumonía en el paciente (figura 6); una tarea muy complicada incluso para los más experimentados. La inteligencia artificial, concretamente las redes neuronales convolucionales, podían ayudar en esta tarea aprendiendo a identificar patrones muy específicos de estos hallazgos, algunos de los cuales podían escapar incluso al ojo humano, para clasificar al paciente. De ahí que las técnicas de imagen se consideraran una alternativa diagnóstica en el manejo de los pacientes de urgencias, especialmente en aquellos casos en los que el acceso a las pruebas de PCR era limitado o existía la sospecha clínica de que se obtuviera un falso negativo en la misma. Por otro lado, la TC es la técnica de imagen más sensible y específica; por tanto, para abordar esta necesidad, se propuso una plataforma basada en la nube para recopilar casos de todo el mundo y generar una base de datos grande y heterogénea para construir modelos predictivos basados en IA, y proporcionar así una herramienta de ayuda al diagnóstico.

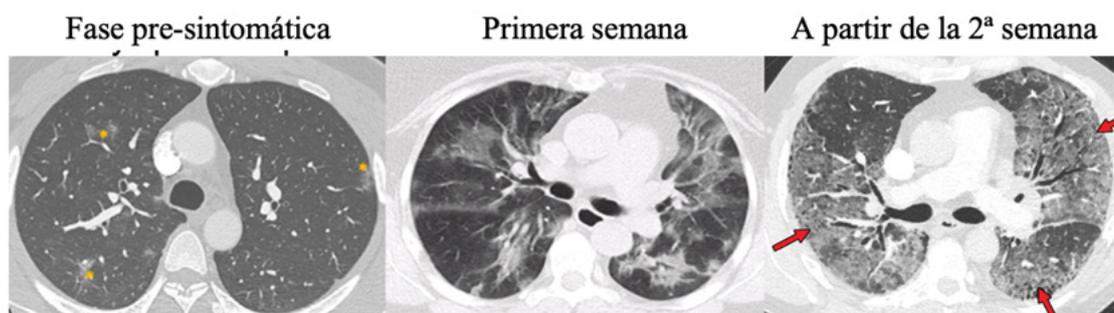
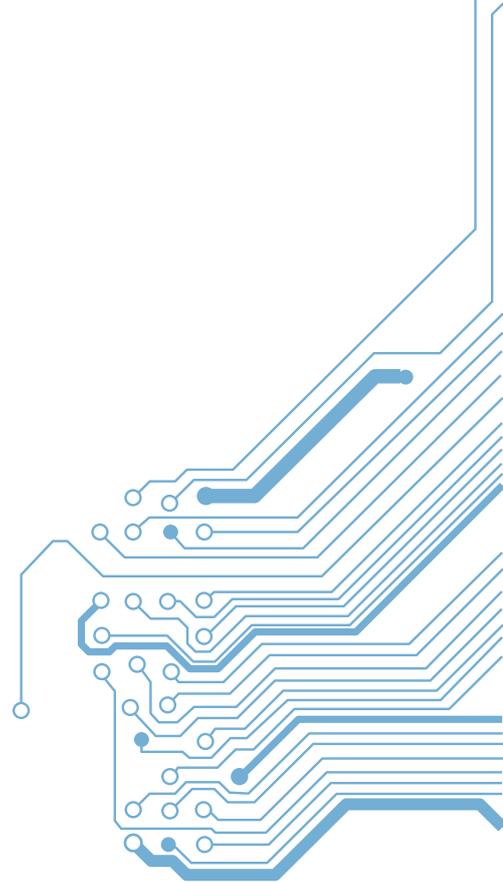


Figura 6: progresión de las lesiones relativas a COVID-19

El origen de la idea fue el hecho de que diferentes científicos y radiólogos compartieron en diferentes plataformas casos que llegaban a sus centros y hospitales y que mostraban estos hallazgos característicos que se observaban como patrones recurrentes en los pacientes que padecían la enfermedad (figura 7). De ahí surgió la primera hipótesis de poder detectar estos casos mediante la creación de un modelo basado en inteligencia artificial, concretamente en redes neuronales convolucionales que aprendieran de estos patrones y fueran capaces de diferenciar los casos normales o los pacientes con otras infecciones pulmonares de los casos positivos de COVID-19.

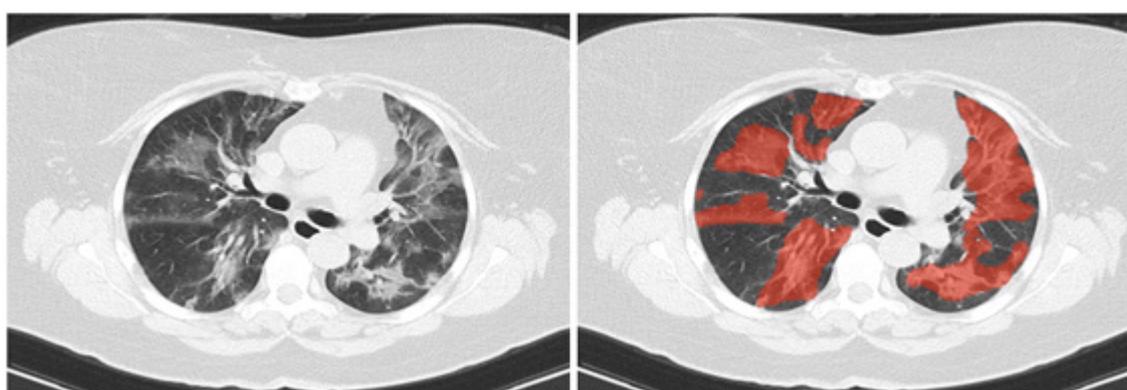


Figura 7: segmentación manual hallazgos relativos a COVID-19

Esta herramienta podía servir y ayudar en la detección de COVID-19, el diagnóstico, la elección de terapias y el seguimiento del tratamiento, basándose en técnicas de IA de última generación que son capaces de aprender de los patrones de la imagen. La plataforma se creó con el objetivo de generar una gran base de datos, con la que aportar conocimiento y respuestas a las numerosas incógnitas que plantea esta enfermedad.

La plataforma Imaging COVID-19 AI fue ofrecida a la comunidad científica y médica para mejorar la investigación y acelerar el descubrimiento de patrones de imagen de COVID-19 a través del aprendizaje profundo. Los algoritmos de IA robustos solo pueden crearse con un número representativo de muestras y los casos de COVID-19 estaban aislados en hospitales específicos, por lo que centralizar todos los estudios posibles para generar una gran base de datos fue el primer paso.

Una vez lanzada, la comunidad científica no tardó en comenzar a subir los primeros estudios. Estos estudios se subían una vez anonimizados y se solicitaba rellenar un formulario para completar la subida, en el que el usuario especificaba información como: síntomas, estado, resultado de la PCR y

diagnóstico final. Una vez completado este proceso, el estudio se almacenaba en la base de datos y se podía tanto acceder al visor para visualizar las imágenes como lanzar un análisis del estudio. Gracias a esta plataforma, se recopiló una gran base de datos (figura 8, figura 9) con la que generar modelos de IA que permitieran detectar la enfermedad y cuantificar la afectación pulmonar.

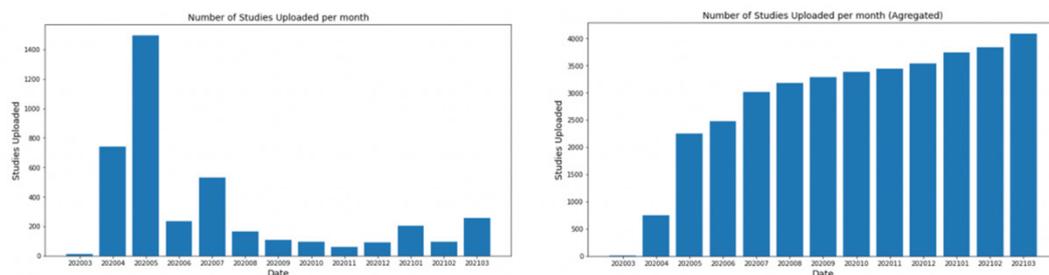


Figura 8: número de estudios subidos a la plataforma

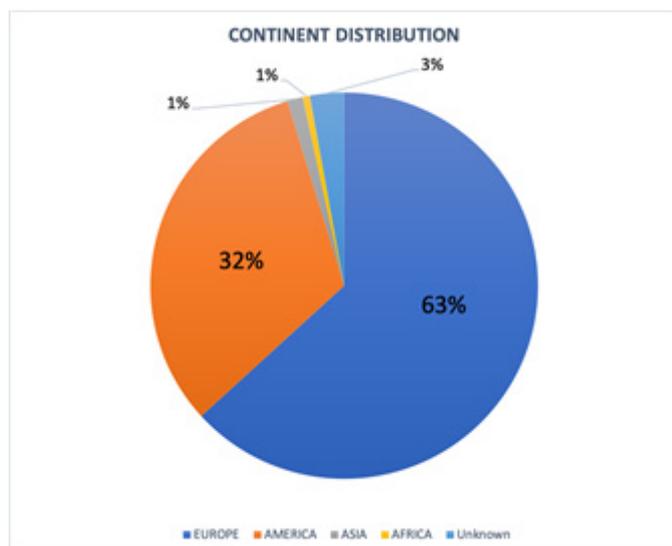


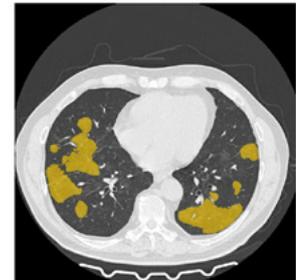
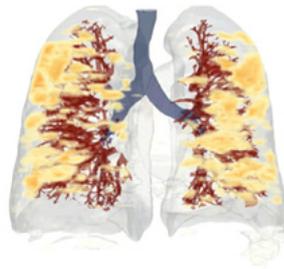
Figura 9: distribución de los estudios subidos por continente

En concreto, la plataforma hizo posible el desarrollo de dos modelos independientes que serían unidos en la herramienta final para proporcionar un resultado conjunto. Por un lado, un modelo de clasificación para la detección de pacientes positivos de COVID-19, diferenciándolos de pacientes con neumonías convencionales y de sujetos sanos. Por otro lado, un modelo de segmentación que permitiera cuantificar los hallazgos radiológicos de los pacientes positivos, lo cual ofrece la posibilidad de analizar la afectación pulmonar y la gravedad del paciente. Estos modelos, una vez

## Chest Imaging COVID-19 Analyzer

VERSION 2.0.1-research

Imaging Center	qubim_generic_01	Patient Name	olimpo_lab
Modality	CT	Patient ID	olimpo_lab
Study Description	TORAX-ABDOMEN SIN CTE.(Ad...	Patient Sex	M
Study Date	20210303	Birthdate	undefined



entrenados y validados, fueron incluidos en un módulo de análisis, un sistema que, dado una adquisición de TC a la entrada, clasificaba la muestra, cuantificaba la afectación en caso de que fuera positiva y presentaba los resultados en un informe estructurado (figura 10).

### Lung Disease Classification

<b>COVID-19 probability</b>	
<b>0.58</b>	
<b>Other Type of Pulmonary Infection</b>	
<b>0.27</b>	
<b>No Imaging Signs of Infection</b>	
<b>0.15</b>	

### Quantitative Analysis

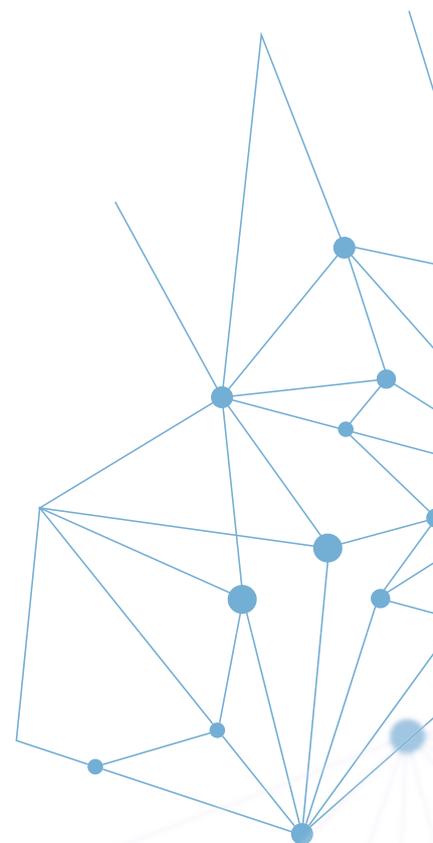
Lobes Quant.	Lung (mL)	Opac. (mL)	Opac. (%)	R - L Quant.	Lung (mL)	Opac. (mL)	Opac. (%)
Left Upper	1336	83	6	Left Lung	2945	252	8
Left Lower	1609	169	10				
Right Upper	989	101	10	Right Lung	3411	254	7
Right Middle	787	37	4				
Right Lower	1633	115	7				

Total Lung Volume (mL)	Affected Lung Volume (mL)	Affected Lung Percentage (%)	Severity Index (out of 25)
6356	506	7.97	9
<b>Infection:</b>		<b>Potential COVID-19</b>	
<b>Involvement:</b>		<b>Moderate</b>	

Figura 10: ejemplo de informe para el módulo Imaging COVID-19 Analyzer

Not cleared for clinical use  
Qubim S.L. - Quantitative Imaging Biomarkers in Medicine  
Edificio Europa - Avenida Aragon 30, 13th Floor, Office I - J. Valencia (Spain)

Finalmente, esta herramienta (Imaging COVID-19 Analyzer) se puso disponible en la plataforma de Imaging COVID-19 y actualmente está siendo usada por más de 60 instituciones alrededor del mundo, que comprenden centros de investigación, clínicas y hospitales.



# 8

## CASOS PRÁCTICOS

En esencia, una plataforma de análisis de imágenes médicas proporciona herramientas únicas para obtener información cuantitativa y ayudar a la mejora del diagnóstico en radiología para un gran abanico de áreas anatómicas. Esta proporciona herramientas de IA, conformando diferentes módulos de análisis desarrollados usando grandes volúmenes de datos. Permite además reunir datos heterogéneos, incluyendo imágenes procedentes de distintos hospitales, escáneres, protocolos de adquisición y variación en la calidad de la imagen, que se incluyen en cada conjunto de datos utilizado para entrenar cada modelo. Además, refleja la necesidad de sistemas que centralicen la imagen médica para su análisis y explotación de resultados, ya que gracias a esta recopilación de datos es posible desarrollar y reentrenar estos modelos basados en inteligencia artificial que aprenden desde los patrones más generales hasta los más específicos. Para ello, se resalta la importancia de la metodología y los procesos MLOps en torno al diseño, la construcción y el despliegue de modelos de ML en producción. Estos modelos tienen el potencial de ser utilizados para identificar, cuantificar o representar diferentes zonas anatómicas o hallazgos radiológicos, y aportan valor en la práctica clínica, tanto en términos de tiempo como de precisión diagnóstica. Esto genera un gran impacto no solo en el cribado de urgencias, donde la identificación de patologías debe realizarse de forma rápida y lo más precisa posible, sino también en la decisión sobre el tratamiento del paciente, hecho que mejora la evaluación del diagnóstico y el control de la enfermedad a lo largo de su evolución y también en los ensayos clínicos, haciéndolos menos costosos y eficientes.

## REFERENCIAS

- “A successful Git branching model”. (5 de enero de 2010). Nvie.Com. <https://nvie.com/posts/a-successful-git-branching-model/>
- Azeri, I. (15 de septiembre de 2019). “What is CI/CD?”. Mabl. <https://www.mabl.com/blog/what-is-cicd>
- Bidgood, W. D.; Jr, Horii; S. C., Prior; F. W., y Van Syckle, D. E. (1997). “Understanding and using DICOM, the data interchange standard for biomedical imaging”. *Journal of the American Medical Informatics Association: JAMIA*, 4(3), 199–212. <https://doi.org/10.1136/jamia.1997.0040199>
- Boettiger, Carl. (2014). “An introduction to Docker for reproducible research, with examples from the R environment”. *ACM SIGOPS Oper. Syst. Rev.* 49. 10.1145/2723872.2723882.
- DeepLearning.AI. (2 de julio de 2021). “Machine Learning Engineering for Production (MLOps) Specialization”. <https://www.deeplearning.ai/program/machine-learning-engineering-for-production-mlops/>
- Fang, Y.; Zhang, H. y Xie, J. (19 de febrero de 2020). “Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR”. *Radiology*, 296(2). <https://doi.org/10.1148/radiol.2020200432>
- Martí-Bonmatí, L.; Alberich-Bayarri, Á.; Ladenstein, R.; Blanquer, I.; Segrelles, J. D.; Cerdá-Alberich, L.; Gkontra, P.; Hero, B.; García-Aznar, J. M.; Keim, D.; Jentner, W.; Seymour, K.; Jiménez-Pastor, A.; González-Valverde, I.; Martínez de Las Heras, B.; Essiaf, S.; Walker, D.; Rochette, M.; Bubak, M.; Mestres, J.; ... Neri, E. (2020). “PRIMAGE project: predictive *in silico* multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers”. *European radiology experimental*, 4(1), 22. <https://doi.org/10.1186/s41747-020-00150-9>

- Mazón, M.; Vázquez Costa, J. F.; Ten-Esteve, A.; & Martí-Bonmatí, L. (2018). "Imaging Biomarkers for the Diagnosis and Prognosis of Neurodegenerative Diseases. The Example of Amyotrophic Lateral Sclerosis". *Frontiers in neuroscience*, 12, 784. <https://doi.org/10.3389/fnins.2018.00784>
- Piroth, L.; Cottenet, J. y Mariet, A.-S. (1 de marzo de 2021). "Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study". *The Lancet*, 9(3), 251-259. [https://doi.org/10.1016/S2213-2600\(20\)30527-0](https://doi.org/10.1016/S2213-2600(20)30527-0)
- Quibim. (2018). Quibim Precision | "Quantitative imaging biomarkers in medicine. Quibim Precision". <https://primage.quibim.com/login/#/signin>
- Sadad, T.; Rehman, A.; Munir, A.; Saba, T.; Tariq, U.; Ayesha, N. y Abbasi, R. (2021). "Brain tumor detection and multi-classification using advanced deep learning techniques". *Microscopy research and technique*, 84(6), 1296-1308. <https://doi.org/10.1002/jemt.23688>



PARC CIENTÍFIC  
UNIVERSITAT DE VALÈNCIA



Actuación cofinanciada por la Unión Europea a través del Programa Operativo del Fondo Europeo de Desarrollo Regional (FEDER) de la Comunidad Valenciana 2014-2020

