
Big data y analítica de negocio

PID_00276617

Josep Curto

Tiempo mínimo de dedicación recomendado: 6 horas



**Josep Curto**

Licenciado en Matemáticas (UAB). Máster en Dirección de Empresas (International Executive MBA) en IE Business School. Máster en Business Intelligence y en Dirección y Gestión en Sistemas y Tecnologías de la Información por la UOC. Consultor, asesor y científico de datos. Profesor asociado en la Universitat Oberta de Catalunya (UOC). Profesor adjunto en IE Business School y ocasional en AGSM en University of New South Wales (UNSW). Miembro del consejo asesor de L'Escola L'Horitzó. Es autor de múltiples libros de divulgación, y ha vivido en cinco países y tres continentes.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por el profesor: Josep Curto

Primera edición: octubre 2020

© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)

Av. Tibidabo, 39-43, 08035 Barcelona

Autoría: Josep Curto

Producción: FUOC

Todos los derechos reservados

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita del titular de los derechos.

Índice

Introducción.....	5
Objetivos.....	8
1. Sistemas de información analíticos.....	9
1.1. Contexto actual	9
1.1.1. La evolución del dato	9
1.1.2. El reto de tomar decisiones en un entorno competitivo	12
1.1.3. Por qué algunas empresas compiten mejor que otras ...	13
1.2. Organizaciones orientadas al dato	14
1.2.1. Qué es una organización orientada al dato	15
1.2.2. Necesidades de una organización orientada al dato	15
1.2.3. Decálogo de una organización orientada al dato	19
1.3. Sistemas de información analíticos	20
1.3.1. <i>Business intelligence, business analytics y big data</i>	20
1.3.2. Diferencias entre <i>business intelligence, business analytics y big data</i>	25
1.3.3. La plataforma del dato	26
2. Beneficios, barreras y casos de uso.....	37
2.1. Beneficios organizacionales	37
2.1.1. Beneficios generales	37
2.1.2. Beneficios tangibles	38
2.1.3. Beneficios intangibles	38
2.2. Barreras	39
2.2.1. Barreras	39
2.2.2. ¿Cómo detectar que no existe una estrategia?	41
2.2.3. La necesidad de los centros de competencia	42
2.2.4. Buenas prácticas	43
2.3. Casos de uso	43
2.3.1. Toma de decisiones	44
2.3.2. Operaciones e inteligencia operacional	44
2.3.3. Validación de hipótesis y resolución de problemas	45
2.3.4. Productos y servicios de datos	46
2.3.5. Comercio de datos	47
2.3.6. Cómo identificar y discutir oportunidades analíticas ...	48
3. Fases, madurez y aspectos éticos.....	50
3.1. Fases	51
3.1.1. Diferentes caminos para llegar al mismo lugar	51

3.1.2.	Cómo empezar	53
3.1.3.	Cómo medir el éxito	53
3.2.	Madurez	54
3.2.1.	Modelos de madurez	54
3.2.2.	Competencias analíticas	57
3.2.3.	Pensamiento analítico	61
3.2.4.	Roles y perfiles	63
3.3.	Aspectos éticos	66
3.3.1.	Derivas, sesgos y riesgos	67
3.3.2.	Explicabilidad, interpretabilidad y equidad	69
3.3.3.	<i>Data ethics canvas</i>	72
Bibliografía		75

Introducción

Tal y como apunta Manuel Castells, vivimos en la era de la información: un periodo histórico caracterizado por una revolución centrada en las tecnologías digitales de la información y la comunicación. En este contexto, las organizaciones están viviendo una profunda transformación no solo desde la perspectiva de los sistemas de información (el conjunto de componentes que interaccionan entre sí con el propósito de alcanzar un objetivo determinado) o de la necesidad de lidiar con las nuevas estructuras en red, sino también en la adopción de una idea que está cambiando profundamente la gestión de las organizaciones: **el dato es el activo más relevante para competir en esta era.**

En esta dirección destaca que esta nueva era no solo afecta a las organizaciones, sino que es un cambio global que afecta profundamente a la sociedad. Varios aspectos se han combinado en este periodo histórico y destacamos algunos de ellos.

Durante las últimas décadas, la reducción en costes de hardware, software y servicios de las tecnologías de información y comunicaciones (TI¹) ha permitido a organizaciones e instituciones definir, desarrollar e implementar sistemas de información para el soporte de sus procesos de negocio.

La facilidad de acceso a sistemas de información de toda índole (desde sistemas operacionales hasta sistemas CRM²) ha permitido digitalizar paulatinamente los procesos de negocio y obtener datos estructurados sobre el rendimiento de la organización, al tiempo que ha abierto la puerta a analizar dichos datos.

Las organizaciones han usado diferentes estrategias para tomar mejores decisiones a partir de los datos. A medida que estas prácticas de análisis se han extendido, más y más empresas se han interesado por explotar de manera eficiente cualquiera de los ámbitos de la empresa (procesos, empleados, clientes y productos).

Con la democratización de internet, el advenimiento de las redes sociales o el despliegue de la internet de las cosas, las fuentes de datos vinculadas con el negocio han superado el perímetro corporativo y actualmente, para comprender lo que sucede, ya no es suficiente con analizar las fuentes de datos internas. Este fenómeno también ha producido, por un lado, una mayor riqueza y variedad de información que se debe considerar y, por otro, un incremento de la complejidad en la captura, el tratamiento, el análisis y la visualización de la información.

Lectura recomendada

Manuel Castells es el autor de la trilogía *La era de la información* (2006), donde profundiza sobre la transformación de la sociedad basada en las tecnologías digitales.

(1) **TI**: es el acrónimo de tecnologías de la información, que comprende todo lo que está vinculado con el almacenamiento, la protección, el procesamiento y la transmisión de la información. Cuando se incluye comunicaciones entonces se hace uso del acrónimo **TIC**.

(2) **CRM**: acrónimo de *customer relationship management*, que hace referencia a la gestión de la relación con clientes.

Internet de las cosas

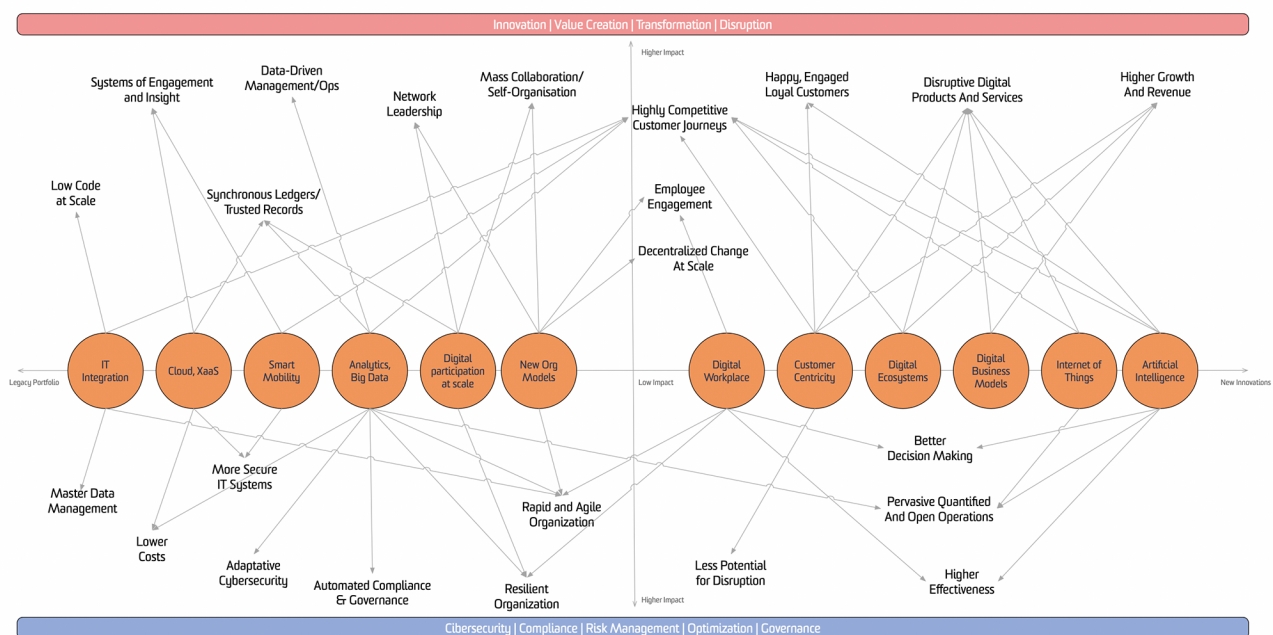
La internet de las cosas, también conocida como *Internet of Things*, hace referencia a la interconexión digital de objetos cotidianos mediante internet.

Esto se combina con el fenómeno de la transformación digital, que se caracteriza por la incorporación de las tecnologías de la información a los procesos de negocio con un foco **dual** (optimización e innovación). Cuando hablamos de **optimización** hacemos referencia a la creación y el despliegue de procesos inteligentes y automatizados con foco en la eficiencia y la optimización, mientras que cuando hablamos de **innovación**, hablamos de la generación y productivización de la innovación a través de la experimentación sistemática.

El gran reto de la empresa dual consiste en encontrar el balance adecuado entre las iniciativas que apuesta por el día a día y las que apuestan por horizontes futuros. Esta dualidad de objetivos se apoya en las múltiples tecnologías que conforman la transformación (desde *cloud computing* hasta las tecnologías de *big data* y analítica de negocio).

Las tecnologías pueden combinarse tanto para introducir casos de uso de optimización como de innovación, tal y como ilustra la figura 1.

Figura 1. Tecnologías en la transformación digital



Fuente: Josep Curto.

De forma natural, con independencia del nivel de madurez de la organización, los datos y la capacidad de explotarlos adecuadamente forman parte de las competencias críticas actuales y futuras.

Las organizaciones necesitan tomar decisiones cada vez más eficientes, rápidas y correctas para poder formar parte de un entorno altamente competitivo y en constante cambio. Tomar decisiones basándose en la experiencia o la intuición no es suficiente. Hay que fundamentarse en los datos y las evidencias. En definitiva, es necesario aprender una nueva manera de competir basada en la analítica de negocio y el *big data*.

A medida que estas estrategias se han hecho conocidas, las organizaciones las han ido implementado con menos fortuna de la esperada, tal y como apuntan los estudios de Aberdeen Group (Lock, 2012), Dresner Advisory Services (Dresner, 2012) o *Harvard Business Review* (Marchand y Peppard, 2013). Si las herramientas han ido madurando a lo largo de los últimos años, ¿cómo es que las organizaciones siguen teniendo tantos problemas en la implantación de este tipo de proyectos?

Como ya es conocido, este tipo de proyectos son complejos y las organizaciones se enfrentan a diversos retos. ¿Qué son los sistemas de información analíticos? ¿Qué tecnologías los conforman? ¿Cómo desplegar con éxito las iniciativas de explotación de datos? ¿Qué barreras presentan este tipo de proyectos? ¿Cómo gestionar adecuadamente este tipo de proyectos? ¿En qué situación se encuentra nuestra organización?

Respondiendo a las anteriores preguntas, el presente material busca capacitar a profesionales en los sistemas de información analíticos (centrándonos en *big data* y analítica de negocio) con el objetivo de desarrollar estrategias de transformación y orientación hacia el dato en el seno de su propia organización. Y, en consecuencia, dar pistas de qué elementos hay que tener en cuenta en la gestión de proyectos de analítica para maximizar sus posibilidades de éxito.

Si bien la obra es autocontenida en la medida de lo posible, a lo largo de este material se introducirán los conceptos necesarios para su seguimiento, como, por ejemplo, conocimientos sobre estrategia, gestión de las tecnologías de la información y comportamiento organizacional.

Referencias bibliográficas

Donald A. MARCHAND y Joe PEPPARD. Why IT Fumbles Analytics. En: *Harvard Business Review*. 2013.

Howard DRESNER. Wisdom of Crowds. En: *Business Intelligence Market Study*. Dresner Advisory Services, 2012.

Michael LOCK. *Managing the TCO of BI: The Path to ROI is Paved with Adoption*. Aberdeen Group, 2012.

Objetivos

Este material está dirigido a:

- Profesionales que buscan conocer las características que definen una organización orientada al dato.
- Consultores que quieren ayudar al desarrollo de una cultura analítica en una organización.
- Gestores que están interesados en la transformación de la toma de decisiones de su organización desde la intuición a la fundamentación basada en hechos y datos.

Y con el estudio y el trabajo de este módulo se deberán alcanzar los objetivos siguientes:

1. Conocer los diferentes sistemas de información analíticos (inteligencia de negocio, analítica de negocio, *big data*).
2. Entender los componentes que conforman los sistemas de información analíticos.
3. Conocer las barreras y los beneficios fundamentales de los sistemas de información analíticos.
4. Identificar casos de uso para los sistemas de información analíticos.
5. Conocer la estructura de los proyectos analíticos.
6. Conocer y aplicar modelos de madurez analíticos para determinar los siguientes pasos de una organización.
7. Aprender a identificar los aspectos éticos de transformarse en una organización orientada a los datos.

1. Sistemas de información analíticos

1.1. Contexto actual

Tal y como apunta Thomas H. Davenport (2005), en los últimos años se está observando la emergencia de una nueva forma de competir en el mercado basada en el uso extensivo de los datos, la analítica de negocio y la toma de decisiones basadas en evidencias.

Aunque el uso del dato para mejorar la toma de decisiones no es nuevo y el despliegue de estrategias como la inteligencia de negocio es cada vez más común en las organizaciones, en los últimos años está apareciendo una nueva hornada de empresas que buscan generar ventajas competitivas de una manera diferente: **mediante el uso del dato**. Ya no solo se busca entender mejor al cliente o los procesos de negocio, sino también poder posicionarse en el mercado de una forma más eficiente, clara y diferente.

A continuación, revisaremos ciertos factores que están empujando a las organizaciones a competir apoyándose en el dato como factor más relevante y diferencial en esta era.

1.1.1. La evolución del dato

Hablar de la evolución del dato significa hablar de un crecimiento en sus magnitudes físicas: volumen, variedad y velocidad. Lo que se conoce como **big data**. Actualmente este término se ha identificado como una de las principales tendencias de transformación del mercado. Tal y como apunta Brynjolfsson (2014), estamos en la era de la digitalización, en la que todo es susceptible de ser transformado en un dato digital.

Esta situación no es nueva. Mucho antes de la aparición de compañías como Yahoo (creada en 1994), Google (creada en 1998), Facebook (creada en 2004) o Twitter (creada en 2006), muchas organizaciones e instituciones ya trataban de gestionar sus datos corporativos mediante informes y bases de datos especializadas y de apoyar sus iniciativas de negocios en información relevante.

Por ejemplo, en 1992 Walmart tenía ya un *data warehouse* cuyo tamaño era de un terabyte³. En 2006 el tamaño de este almacén de datos corporativo alcanzó los 0,5 petabytes⁴, lo que permitía gestionar las actividades de compra de sus 100 millones de clientes en aquel momento. Dos años más tarde el volumen de

Referencia bibliográfica

T. DAVENPORT, D. COHEN y A. JACOBSON. *Competing on Analytics, Working Knowledge Research Report*. Babson Executive Education, 2005.

Referencia bibliográfica

Andrew MCAFEE y Erik BRYNJOLFSSON. *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*. Norton & Company, 2014.

⁽³⁾Un terabyte equivale a 10^{12} bytes.

⁽⁴⁾Un petabyte equivale a 10^{15} bytes.

datos llegó a los 2,5 petabytes. Está claro que, para Walmart, el *data warehouse* fue la pieza angular sobre la que la compañía definió su estrategia de *customer analytics*.

Parecida es la situación de eBay. Esta compañía tiene tres sistemas (<http://gigaom.com/2013/03/27/why-apple-ebay-and-walmart-have-some-of-the-biggest-data-warehouses-youve-ever-seen/>):

1) Un *data warehouse* principal cuyo tamaño es de 9,2 petabytes.

2) Un segundo sistema, denominado *singularity system* (<http://www.odbm.com/blog/2011/10/analytics-at-ebay-an-interview-with-tom-fastner/>), destinado a la información de los clics que se hacen en la web y cuyo tamaño es de 40 petabytes.

3) Un sistema basado en Hadoop cuyo tamaño es de 40 petabytes y que persigue analizar el comportamiento de clientes mediante técnicas de *machine learning*.

Además, cada vez más, las empresas buscan no solo analizar los datos generados durante sus actividades, sino también nuevos datos que anteriormente ni se planteaban. Esto es lo que está haciendo, por ejemplo, la compañía Disney (<http://www.rfid-blog.com/?p=1372>), que usa brazaletes equipados con sistemas de localización (GPS⁵ y NFC⁶) para monitorizar todo lo que hacen sus visitantes durante su estancia en los parques de entretenimiento, como Disneyland Paris.

En definitiva, muchas empresas ya están generando más datos mediante la integración de sensores y de aplicaciones móviles en sus productos y procesos de negocio y además están guardando más información o ampliando la cantidad de datos en el proceso de análisis, con el objetivo de conocer de manera pormenorizada el negocio y a los clientes.

A esto se le añade el hecho de que el consumidor se ha convertido en un generador de gran cantidad de datos mediante sus dispositivos móviles en forma de texto, audio, fotografías o datos de comportamiento y de geolocalización.

Las organizaciones persiguen combinar los datos internos (en sistemas corporativos, ficheros, documentos, conversaciones, transacciones, etc.) con datos externos (datos sociales, proveedores, internet de las cosas, etc.). De forma

Data warehouse

Data warehouse, o almacén de datos corporativo, es una base de datos especializada para la consulta de información.

Customer analytics

Customer analytics (o analítica de clientes) hace referencia a la captura, gestión, análisis y generación de valor estratégico de los datos de cliente de una organización.

Hadoop

Hadoop es un ecosistema de componentes que permite el procesamiento distribuido de grandes conjuntos de datos en grupos de computadoras utilizando modelos de programación simples. Fuente: <http://hadoop.apache.org/>.

Machine learning

Machine learning, o aprendizaje automático o de máquinas, es una rama de la inteligencia artificial cuyo objetivo es crear técnicas capaces de aprender de manera automática.

⁽⁵⁾ **GPS**: acrónimo de *global positioning system*, o sistema de posicionamiento global.

⁽⁶⁾ **NFC**: acrónimo de *near field communication*, o comunicación de campo cercano.

⁽⁷⁾ Un zettabyte equivale a 10^{21} bytes.

natural, el deseo de comprender más y mejor, combinado con una situación de creación exponencial de datos y la alta disponibilidad de estos datos, está generando la tormenta perfecta. Según IDC (2018), la cantidad de datos digitales llegará a los 90 zettabytes⁷ en el año 2025, superando ampliamente la capacidad para almacenar, capturar, analizar y procesar la información de las organizaciones si estas no toman las medidas adecuadas.

Referencia bibliográfica

IDC (John GANTZ, David REINSEL y John RYDNING). *The Digitization of the World: From Edge to Core*, 2018. Disponible en: <https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataa-ge-whitepaper.pdf>

Tal y como apuntan Parmar, Mackenzie, Cohn y Gann (2014), la importancia de la era en la que vivimos son las nuevas oportunidades que existen para expandir el negocio o crear nuevas fuentes de ingresos fundamentadas en tecnologías como *big data*, analítica o *cloud computing*. En esta dirección, comentan que las empresas están siguiendo diferentes patrones:

Referencia bibliográfica

Rashik PARMAR, Ian MAC-KENZIE, David COHN y David GANN. The new patterns of innovation. En: *Harvard Business Review*, 2014.

- **Aumentar los productos para generar datos:** las empresas añaden sensores a sus productos, como ha hecho Rolls Royce con sus motores, y utilizan su información para ofrecer nuevos servicios a los clientes, como por ejemplo el servicio de mantenimiento de Rolls Royce: Engine Health Management (EHM⁸), que traslada la idea de la telemetría de la Fórmula 1 al consumidor.
- **Digitalizar sus activos:** se transforman los procesos de creación o monitorización de productos para mejorar su rendimiento y comportamiento. Por ejemplo, en medicina, los cirujanos usan modelos tridimensionales antes de una intervención para reducir el impacto en el paciente e incrementar la precisión del proceso.
- **Combinar datos de múltiples industrias:** este es el caso de las iniciativas de tipo *smart cities* que se están observando en todo el mundo. Por ejemplo, en el caso de Santander (PDI, 2011-2013), destaca el proyecto Burba, que combina sensores en contenedores y en productos de supermercado para identificar el uso correcto de los contenedores de residuos.
- **Compartir y/o vender datos:** por ejemplo, Vodafone ayuda a TomTom a detectar problemas o congestiones de tráfico mediante los datos de los clientes que están conduciendo (<https://www.engadget.com/2006-10-27-vodafone-tomtom-partner-to-create-real-time-traffic-data-networ.html>).
- **Codificar un servicio:** digitalizar un proceso y transformarlo en una línea de negocio. Por ejemplo, Amazon Web Services surge de las iniciativas de gestión TI internas de la compañía.

Cloud computing

Cloud computing, o computación en la nube, hace referencia a recursos computacionales distribuidos y escalables disponibles bajo demanda.

⁽⁸⁾ **EHM:** sistemas de gestión del estado de un motor en tiempo real.

Referencia bibliográfica

Plan Director de Innovación. Ayuntamiento de Santander 2011-2013.

La clave de generar valor sobre el dato frecuentemente es el resultado de la combinación de algunos (o todos) de estos anteriores patrones.

1.1.2. El reto de tomar decisiones en un entorno competitivo

En la actualidad, se considera que muchos mercados se han transformado en mercados altamente competitivos como resultado de los avances en procesadores, bases de datos, sistemas de información, dispositivos, tecnologías de comunicación y redes. Este concepto, introducido por Richard A. D'Aveni, se define como:

«Un mercado competitivo es aquel en el que los competidores actúan rápidamente frente a los gustos de los consumidores, frecuentemente apoyándose en las tecnologías de la información».

Esta capacidad de rápida actuación provoca que estos mercados sean muy «agresivos», puesto que las reglas de la competencia están evolucionando todo el tiempo. Como apunta el propio D'Aveni:

«Las características de los negocios modernos no son tanto la estructura como el cambio dinámico. Veremos la interrupción de las ventajas competitivas con mucha más frecuencia en este entorno».

Lo que está detrás de esta afirmación es que las organizaciones que se enfrentan a un entorno extremadamente competitivo solo pueden crear ventajas competitivas temporales en lugar de crear ventajas competitivas sostenibles, como apuntaba Michael Porter (2008) y requieren una toma de decisiones más rápida. Esta nueva toma de decisiones debe ayudar a controlar y comprender la aceleración en procesos y operaciones.

Como resultado, las organizaciones viven en un entorno en el que los cambios suceden mucho más rápidamente y de forma continua.

Por ejemplo, si anteriormente el cliente debía desplazarse a una tienda a realizar una compra en unas determinadas horas del día, ahora es posible realizar dicha compra desde casa mediante un ordenador e incluso desde un teléfono inteligente (u otros dispositivos) fuera de los horarios de la tienda física.

Esto supone para la organización una monitorización continua de los procesos de negocio, una reducción del tiempo de respuesta para el cliente, la creación de propuestas de valor en tiempo real para capturar el interés del cliente y optimizar las operaciones para poder cumplir sus expectativas.

Como resultado, estos grandes cambios de base tecnológica provocan una gran turbulencia en aquellos sectores que se fundamentan o dependen profundamente de las tecnologías de la información, como apunta el estudio de McAfee y Brynjolfsson (2008). Los resultados de este estudio apuntan que el despliegue y uso eficiente de TI permite mejorar tanto los ingresos como los beneficios de las organizaciones, hecho que ha provocado, durante las últimas décadas, que:

Richard A. D'Aveni

Profesor de estrategia en Tuck School of Business, Dartmouth College y considerado como uno de los expertos mundiales en estrategia competitiva.

Referencia bibliográfica

Michael PORTER. The five forces that shape strategy. En: *Harvard Business Review*. 2008.

Referencia bibliográfica

Andrew MCAFEE y Erik BRYNJOLFSSON. Investing in the IT that makes a competitive difference. En: *Harvard Business Review*. 2008. Vol. 86, n.º 7/8.

- Las empresas que usan TI tengan mayores márgenes de beneficio que las que no los usan.
- Cada vez es más complicado mantenerse como líder de un mercado⁹ que depende de TI durante un largo periodo de tiempo, puesto que los competidores también despliegan las mismas tecnologías o equivalentes.

⁽⁹⁾En términos de cuota de mercado (en inglés, *market share*).

Estudios más recientes (Capgemini y MIT, 2012) incluso apuntan que el impacto de este tipo de entorno competitivo tiene mucho más calado. En particular:

- Las líneas entre la gestión estratégica y operacional se están diluyendo al tener que redefinir las reglas de competencia en tiempo casi real.
- Las tecnologías digitales están redefiniendo la competencia estratégica.
- La única forma de competir se basa en la experimentación y el prototipado rápido.
- Es necesario desarrollar una cultura que integre el proceso de innovación para poder transformar los experimentos y los prototipos en valor para el cliente.

Referencia bibliográfica

CAPGEMINI y MIT Center for Digital Business. *The Digital Advantage: How Digital Leaders Outperform their Peers in Every Industry*, 2012.

1.1.3. Por qué algunas empresas compiten mejor que otras

En el apartado anterior se ha introducido que la tecnología juega un papel fundamental en el nuevo entorno competitivo. Sin embargo, el fenómeno de la competencia extrema genera que las ventajas competitivas no sean perdurables. Si una empresa quiere destacar en este escenario, debe hacer una apuesta por la tecnología, y sin embargo teniendo en cuenta la miríada de nuevos avances tecnológicos y su potencial impacto, puede ser complicado dirimir cuál es la apuesta adecuada. Ha quedado patente que internet, *cloud computing* y los dispositivos móviles son parte de las tecnologías que deben ser consideradas y usadas dentro de una organización, pero vivimos en una carrera continua de disrupciones tecnológicas, tal y como apunta McKinsey (Manyika y otros, 2013).

Referencia bibliográfica

James MANYIKA, Michael CHUI, Jacques BUGHIN, Richard DOBBS, Peter BISSON y Alex MARRS. *Disruptive technologies: Advances that will transform life, business, and the global economy*. McKinsey Insight & Publications, 2013.

Según Brynjolfsson, Hitt y Kim (2011), de todas las tecnologías existentes, las empresas deben apostar preferentemente por aquellas vinculadas con el análisis de la información. Apuntan que aquellas compañías que usan análisis basado en datos en lugar de la intuición tienen una productividad y unos beneficios de un 5 a un 6 % más altos que sus competidores. Este tipo de estrategia ha ido calando entre las organizaciones y en 2013 observaron ya un aumento (Bloom y otros, 2013) significativo en el uso de esta estrategia.

La respuesta de por qué ciertas empresas compiten mejor que otras parece estar vinculada al hecho de que dichas compañías son empresas orientadas al dato.

Podemos matizar ahora el punto presentado en el apartado anterior:

La orientación al dato es la principal tecnología digital que está redefiniendo cómo compiten estratégicamente las empresas, pero debe combinarse con otras para alcanzar un grado de madurez tecnológico significativo.

La madurez tecnológica necesita tiempo para desarrollarse y lo que se está observando en estos momentos en muchas empresas es la culminación de años de inversión, trabajo y adquisición interna de competencias como en el caso de Walmart.

1.2. Organizaciones orientadas al dato

En los últimos años, estudios de IDC (2019) ponen en manifiesto que **big data** y **analítica de negocio** son actualmente dos de las principales iniciativas en las agendas corporativas y que están recibiendo una fuerte y continua inversión.

El uso combinado de estas estrategias de datos promete que la forma como las organizaciones operan va a transformarse profundamente para incrementar su productividad y beneficios. Sin embargo, es conocido que la gran mayoría de los proyectos de inteligencia de negocio, **big data** y analítica han fracasado (Dresner, 2012).

Esto pone en manifiesto que no es suficiente conocer la tecnología, las metodologías o la forma de desplegar proyectos de análisis de datos para tener éxito. Hay otro factor muy relevante: es necesario desarrollar e implantar una cultura diferente en la empresa.

A continuación, introduciremos el concepto de la organización orientada al dato, qué tipo de competencias requiere y qué supone generar una cultura de pensamiento crítico.

Referencia bibliográfica

Erik BRYNJOLFSSON, Lorin M. HITT y Heekyung Hellen KIM. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? 2011. Disponible vía SSRN: <https://ssrn.com/abstract=1819486> o <http://dx.doi.org/10.2139/ssrn.1819486>.

Nicholas BLOOM, Erik BRYNJOLFSSON, Lucia FOSTER, Ron S. JARMIN, Itay SAPORTA EKSTEN y John VAN REENEN. Management in America. En: *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-01*, 2013. Disponible vía SSRN: <http://ssrn.com/abstract=2200954>

Referencia bibliográfica

Worldwide Big Data and Analytics Software Forecast, 2019-2023. IDC, septiembre 2019.

Referencia bibliográfica

Howard DRESNER. *Wisdom of Crowds Business Intelligence Market Study*. Dresner Advisory Services, 2012.

1.2.1. Qué es una organización orientada al dato

Podemos definir este tipo de organizaciones como:

Una organización orientada al dato es aquella que se gestiona fundamentada en hechos y datos.

Extendiendo la definición anterior, según Patil y Manson (2015),

«Una **organización orientada al dato** adquiere, procesa y aprovecha los datos de manera oportuna para crear eficiencias, iterar y desarrollar nuevos productos y navegar por el panorama competitivo».

Referencia bibliográfica

D. J. PATIL y H. MASON. *Data Driven*. O'Reilly Media, Inc., 2015.

Esta descripción nos indica que este enfoque empresarial puede entenderse como una respuesta a un mercado altamente competitivo. La transformación hacia una organización orientada al dato requiere tres capacidades vinculadas entre ellas:

- En primer lugar, la organización debe ser capaz de **identificar, combinar y gestionar** múltiples fuentes de datos.
- En segundo lugar, la organización debe ser capaz de **construir modelos de análisis avanzado** para predecir y optimizar sus resultados.
- Tercero, y no por ello menos importante, debe poseer la **capacidad de transformar la organización** para que los datos y los modelos realmente deriven en mejores decisiones.

Estas capacidades deben estar sustentadas por una estrategia clara para el uso y el análisis de datos y el despliegue de arquitecturas y capacidades tecnológicas adecuadas, que conforman los sistemas de información analíticos en los que se fundamentan las organizaciones orientadas al dato.

1.2.2. Necesidades de una organización orientada al dato

Cuando hablamos de necesidades, estamos hablando en referencia a los datos corporativos. Es decir, estamos hablando de los datos mantenidos y relevantes para una organización, que incluyen aquellos relacionados con sus finanzas, empleados, clientes, proveedores y su actividad de negocio.

Ante el nuevo contexto, las organizaciones necesitan trabajar con el dato dejando atrás la noción de que la información de valor se encuentra tan solo en el seno de la organización. Este hecho obliga a pensar no solo en las magnitudes del dato, sino en el origen de partida de los datos. Por lo tanto, debemos hablar de datos internos y externos.

- **Datos internos:** hace referencia a datos que pertenecen a la organización. Dentro de los datos internos tenemos aquellos que ya existen o se crean en los propios sistemas de información de la organización (como pueden ser el ERP y/o el CRM), o bien que se están capturando y almacenando mediante mecanismos automáticos a través de diferentes estrategias, como el *crowdsourcing*, sensores y/o dispositivos de monitorización (como un podómetro con localización geográfica).
- **Datos externos:** hace referencia a datos de terceros y que deben ser conseguidos por la organización. Estas fuentes de datos pueden estar disponibles para su compra o ser de acceso libre. Los datos de libre acceso pueden ser a su vez de tres tipos: datos capturados mediante técnicas de *crowdsourcing*, datos de redes sociales (como pueden ser Facebook, Twitter o LinkedIn) y *open data*.

Existen muchas formas de clasificar los datos; según Reichert, Otto y Österle (2010), los datos de una organización pueden dividirse en datos maestros, datos transaccionales y datos de inventario.

- **Datos maestros:** hacen referencia a las características de los objetos básicos de negocio dentro de una organización. Este tipo incluye datos de materiales, productos, proveedores, clientes, empleados y activos.

Por ejemplo, respecto al cliente, su nombre y apellidos forman parte de los datos maestros.

- **Datos transaccionales:** describen las transacciones comerciales y representan la entrada y salida de los procesos empresariales. Se crea constantemente en el curso de las actividades empresariales, hace referencia a datos maestros e indica cambios en los datos de inventario.

Por ejemplo, imaginemos que tenemos una cafetería, la compra de un café y una galleta para desayunar y los datos vinculados a dicha compra (importe, hora, etc.) son datos transaccionales.

- **Datos de inventario:** hacen referencia a los niveles de existencias y cuentas.

Por ejemplo, si continuamos con el ejemplo de la cafetería, la cantidad en existencia de café, en qué fecha se compró o la fecha de caducidad son datos de inventario.

Cada uno de estos tipos de datos tiene características diferentes que podemos comparar y que nos permiten comprender mejor las diferencias entre ellos:

- **Referencia temporal:** se refiere a la dependencia del dato respecto a un determinado punto en el tiempo.
- **Frecuencia actualización:** se refiere a la frecuencia con la que un dato se actualiza o cambia durante un determinado periodo de tiempo.

Crowdsourcing

Cuando hablamos de *crowdsourcing* nos referimos al proceso de obtener servicios, ideas, contenido a través de la participación de una gran masa de personas.

Open data

Cuando hablamos de *open data* nos referimos a conjuntos de datos considerados que son un bien común, y que por ello son gratuitos, accesibles y bien estructurados para su descarga y análisis. Las tipologías son múltiples: de transporte, financieros, meteorológicos, estadísticos, científicos, culturales y geolocalizados.

Referencia bibliográfica

B. OTTO y A. REICHERT. Organizing master data management: findings from an expert survey. En: Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010. Págs. 106-110.

- **Volatilidad volumen:** se refiere al crecimiento del dato en volumen respecto al tiempo.
- **Independencia existencial:** se refiere al grado en el que el dato hace referencia a otros tipos de datos.

Y que se resumen en la tabla 1:

Tabla 1. Tipos de datos y características

Tipo	Referencia temporal	Frecuencia actualización	Volatilidad volumen	Independencia existencial
Datos maestros	Baja	Baja	Baja	Alta
Datos transaccionales	Alta	Baja	Alta	Baja
Datos de inventario	Alta	Alta	Baja	Baja

Fuente: Reichert, Otto y Österle.

Es necesario hablar de una última potencial fuente de datos, que, aunque podría incluirse dentro de la categoría de datos internos, no suele contemplarse dentro de las organizaciones. Estamos hablando del metadato y el valor asociado a este. Primero debemos definir qué es el metadato.

Se entiende por **metadatos** los datos estructurados y codificados que describen características de un objeto, dato o proceso de negocio (Conesa y Curto, 2016).

Referencia bibliográfica

J. CONESA y J. CURTO. *Introducción al Business Intelligence*. Barcelona: Editorial UOC, 2016.

Es decir, no es suficiente con generar valor a partir del dato, sino también de los metadatos vinculados a dicho dato. Podemos hablar de tres grandes categorías de metadatos:

- **Técnicos:** que describen los aspectos técnicos vinculados al dato, como pueden ser las magnitudes de este o, por ejemplo, los derechos de propiedad.
- **Operacionales:** que hacen referencia a los procesos de captura, transformación, almacenaje, análisis y visualización del dato, incluyendo las fórmulas de cálculo.
- **Atributos:** que hacen referencia a los atributos que enriquecen la información sobre el dato. Por ejemplo, en una fotografía encontramos aspectos como el dispositivo con el que se realizó.

El metadato abre la puerta a una nueva gama de análisis del valor y, sobre todo, a comprender de una manera mucho más profunda lo que sucede en una organización. Esta información no siempre es relevante para el usuario final, pero sí de suma importancia para el sistema que gestiona el dato.

Consideremos el correo electrónico, la llamada o el mensaje de texto que se usa para la comunicación tanto personal como profesional. En este caso, los metadatos son el horario, la fecha en la que se envió, la duración, los agentes que participan en la conversación y la localización desde que se conectó el usuario la última vez, entre otros. Esta información no revela el contenido de las comunicaciones, sino de las transacciones electrónicas mostrando sus patrones, relaciones y comportamientos.

Con independencia de la clasificación usada y de la tipología del dato, los usuarios de negocio y los distintos profesionales que trabajan con los datos tienen unas necesidades esenciales que deben ser cubiertas. Tal como ilustra la figura 2, los datos en una organización deben:

- **Estar disponibles:** los datos han sido capturados o generados y almacenados en un repositorio.
- **Ser accesibles:** los mecanismos adecuados de acceso a los datos almacenados en el repositorio están disponibles.
- **Tener la calidad adecuada:** los datos almacenados han sido validados y tienen la calidad adecuada para ser usados para los diferentes casos de uso.
- **Estar disponibles en el momento adecuado:** los datos han sido capturados en el momento correcto para el caso de uso.
- **Estar securizados:** los datos solo pueden ser accesible por las personas con derechos adecuados y existen los mecanismos pertinentes para su protección y securización.
- **Ser transformados en información o conocimiento:** los datos han sido transformados en información o conocimiento, o al menos han sido preparados para acelerar su uso.

Figura 2. Necesidades de los usuarios respecto a los datos



Fuente: Josep Curto.

1.2.3. Decálogo de una organización orientada al dato

Convertirse en una organización orientada al dato requiere transformar múltiples aspectos. Las necesidades descritas nos permiten intuir los aspectos primordiales que requieren este tipo de organizaciones.

Un mecanismo es considerar un decálogo similar al que incluimos en este material que cubre los principales aspectos:

- **Estrategia:** la organización usa los datos sistemáticamente para analizar problemas, tomar decisiones y crear mejores servicios para clientes, empleados y proveedores.
- **Responsabilidad:** todos los empleados son creadores y consumidores de datos, y, por lo tanto, responsables de su cuidado a lo largo de su ciclo de vida. De forma natural, cada dato debe tener un propietario.
- **Protección y ética:** la organización protege los datos de acuerdo con criterios legales y técnicos, así como considerando principios de sensibilidad y ética.
- **Relevancia:** los datos de la organización son clasificados y tratados según su propósito y el valor que aportan a los diferentes casos de uso.

- **Calidad:** la organización vela por la integridad, validez, completitud y precisión de los datos que se usan en relación con el caso de uso al que se vinculan.
- **Acceso:** la organización asegura la disponibilidad y facilita el acceso de los datos que necesita cada persona en el momento, lugar y formato adecuados.
- **Dato único:** cada dato de la organización tiene un significado único y compartido por todos, expresado y codificado formalmente, y no se duplica.
- **Políticas:** la organización establece y actualiza políticas y estándares organizativos y técnicos de gestión de los datos, basados en las buenas prácticas profesionales y adaptados al contexto de negocio y social.
- **Organización:** la dirección de la organización fomenta una cultura de datos a través de una función específica para diseñar y coordinar la estrategia y el gobierno de datos.
- **Innovación:** los servicios de datos son una palanca para la transformación de la organización para el desarrollo de nuevos productos y servicios.

1.3. Sistemas de información analíticos

La gestión de una organización se fundamenta en tomar decisiones adecuadas respecto a clientes, productos, empleados, proveedores y procesos de negocio. Por lo tanto, es necesario tener mecanismos que den soporte a una toma de decisiones eficiente. No es extraño que los CIO¹⁰ de las principales empresas del mundo destaquen que su principal prioridad tecnológica son este tipo de iniciativas.

⁽¹⁰⁾ **CIO:** siglas de *chief information officer*, que equivalen a hablar del director de sistemas de una organización.

Como ya sabemos, es necesario convertirse en una organización orientada a los datos, lo que supone dominar y hacer un uso extensivo de las tecnologías que permiten explotar adecuadamente el dato. En general, estamos hablando de la inteligencia de negocio, la analítica de negocio y el *big data*.

En esta sección definiremos estas tecnologías, sus diferencias y su encaje dentro de lo que se conoce como la plataforma del dato.

1.3.1. *Business intelligence, business analytics y big data*

El contexto de la sociedad de la información (Castells, 1999) ha propiciado la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor.

Referencia bibliográfica

M. CASTELLS. *La era de la información: economía, sociedad y cultura* (vol. 1). Chicago: Siglo XXI, 1999.

La **inteligencia de negocio** (o *business intelligence*) responde a dicha necesidad y podemos entender, en una primera aproximación, que es una evolución de los sistemas de soporte a las decisiones (*decision support systems*, DSS). Sin embargo, este concepto, que actualmente se considera crítico en la gran mayoría de las empresas, no es nuevo. En octubre de 1958 Hans Peter Luhn, investigador de IBM, acuñó el término en el artículo «A business intelligence system» (Luhn, 1958):

«La habilidad de aprehender las relaciones de hechos presentados de forma que guían las acciones hacia una meta deseada».

No es hasta 1989 cuando Howard Dresden, en dicho momento analista de Gartner, propone una definición formal de este concepto:

«Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos».

Desde entonces, el concepto del que estamos hablando ha evolucionado aunando diferentes tecnologías, metodologías y términos bajo su paraguas. En el contexto de este material usaremos la definición proporcionada por Davenport (2010):

«Se entiende por **inteligencia de negocio** una amplia categoría de habilidades, procesos, tecnologías, aplicaciones y prácticas para ayudar a los usuarios empresariales a tomar decisiones mejores y más eficientes».

En esencia, mediante la inteligencia de negocio, podemos romper con la siguiente máxima de William Thomson:

«Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar. Lo que no se mejora, se degrada siempre».

Para representar el componente analítico de la inteligencia de negocio, cada vez más prominente en la generación de valor, se introduce el concepto de **analítica de negocio** (o *business analytics*):

La cadena de valor empresarial

La cadena de valor empresarial, descrita y popularizada por Michael E. Porter en su obra *Competitive Advantage: Creating and sustaining superior performance*, es un modelo teórico que permite describir las actividades que generan valor en una organización.

Referencia bibliográfica

H. P. LUHN. A business intelligence system. En: *IBM Journal of research and development*. 1958. Vol. 2, n.º 4, págs. 314–319.

Gartner

Gartner es una de las principales empresas de prospección de mercado.

Referencia bibliográfica

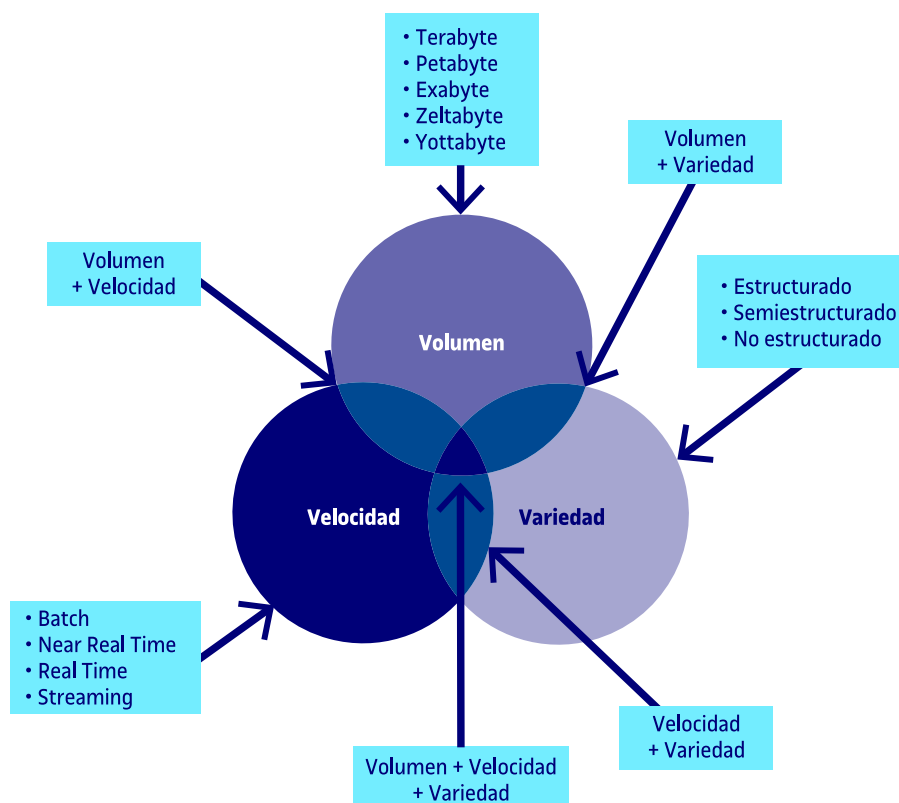
T. H. DAVENPORT, J. G. HARRIS y R. MORISON. *Analytics at work: Smarter decisions, better results*. Cambridge, MA: Harvard Business Press, 2010.

Analítica de negocio (BA) se define como el uso extensivo de datos, análisis estadísticos y cuantitativos, modelos explicativos y predictivos, y gestión basada en hechos para impulsar decisiones y acciones (Davenport y otros, 2010).

Es decir, a partir del rendimiento pasado de la organización, *business analytics* permite predecir comportamientos futuros, así como para detectar ocultos en la información.

La explosión en complejidad en los datos ha propiciado la emergencia de un nuevo concepto: **big data**. Cuando usamos este concepto (McAfee y Brynjolfsson, 2012) hacemos referencia a conjuntos de datos caracterizados por un gran volumen, alta velocidad y significativa variedad, tal y como ilustra la figura 3.

Figura 3. Magnitudes físicas del dato



Fuente: SmartDataCollective.

Por lo tanto, *big data* puede entenderse por sus características y los desafíos que los datos ponen en la capacidad informática (Janssen y otros, 2017).

Se entiende por **big data** el conjunto de estrategias, tecnologías y sistemas para el almacenamiento, el procesamiento, el análisis y la visualización de conjuntos de datos complejos.

Referencia bibliográfica

T. H. DAVENPORT, J. G. HARRIS y R. MORISON. *Analytics at work: Smarter decisions, better results*. Cambridge, MA: Harvard Business Press, 2010.

Referencia bibliográfica

Andrew MCAFEE y Erik BRYNJOLFSSON. Big data: The management revolution. En: *Harvard Business Review*. 2012. Vol. 90, n.º 10, págs. 60–68.

Referencia bibliográfica

M. JANSSEN, H. van der VOORT y A. WAHYUDI. Factors influencing big data decision-making quality. En: *Journal of Business Research*. 2017. Vol. 70, págs. 338–345.

Es necesario recordar que cuando hablamos de almacenamiento hablamos de los soportes físicos y de software que permiten guardar el dato en estructuras que representan su complejidad; que cuando hablamos de procesamiento nos estamos refiriendo a aquellas operaciones que permiten la ingestión, la transformación y la distribución del dato para adecuarlo para el consumo; que cuando hablamos de análisis hacemos referencia a las técnicas aplicadas para generar valor; y que cuando hablamos de visualización hacemos referencia a los mecanismos de consumo de información.

De acuerdo con el NIST⁽¹¹⁾, y en particular dentro de su grupo de trabajo de *big data* (más información en: <http://bigdatawg.nist.gov>), una forma de categorizar este concepto es mediante las necesidades de negocio:

⁽¹¹⁾**NIST:** acrónimo de *National Institute of Standards and Technology*, una institución americana que estudia, define y promueve estándares tecnológicos.

- El modelo de negocio **no se puede representar mediante una estructura de datos relacional** (es decir, mediante una base de datos relacional).

Estructura de datos relacional

Cuando hablamos de estructura de datos relacional hacemos referencia a un tipo de base de datos que permite establecer interconexiones o relaciones entre los datos guardados en tablas.

- El modelo de negocio necesita **ser escalable** por el crecimiento de datos respecto a su velocidad o volumen.

Escalabilidad

Cuando hablamos de *escalabilidad* nos referimos a la habilidad de un sistema, red o proceso para reaccionar y adaptarse sin perder calidad, o bien manejar el crecimiento continuo de trabajo de manera fluida, o bien para estar preparado para hacerse más grande sin perder calidad en los servicios ofrecidos. La escalabilidad horizontal está fundamentada en incrementos de nodos del sistema, proceso o red, mientras que la vertical consiste en añadir más recursos –memoria, disco duro y/o procesadores–.

La base de esta clasificación es poder identificar correctamente los diferentes escenarios que se derivan de estas condiciones ya sea mediante recursos internos o la ayuda de especialistas externos. La combinación de estos dos aspectos nos proporciona tres tipologías de *big data* y un escenario en el que no existe esta necesidad. Aunque es patente el valor que aporta el *big data*, no todos los problemas de una organización son necesariamente un problema de datos. Los tipos disponibles se resumen en la tabla 2.

Tabla 2. Tipos de *big data*

Tipo	Descripción
Tipo 1	Donde una estructura de datos no relacional es necesaria para el análisis de negocio.
Tipo 2	Donde es necesario aplicar estrategias de escalabilidad horizontal para procesar y analizar de manera eficiente el negocio.

Fuente: NIST.

Tipo	Descripción
Tipo 3	Donde es necesario procesar una estructura de datos no relacional mediante estrategias de escalabilidad horizontal para procesar y analizar de forma eficiente el negocio.

Fuente: NIST.

Por tanto, para una determinada necesidad de negocio, es posible identificar si estamos en un escenario de *big data* o no y si es necesario este tipo de tecnologías, hecho que cada vez más se erige como un punto relevante y de partida para la implementación de este tipo de proyectos. Esto se resume en la tabla 3.

Tabla 3. Autoevaluación de la existencia de *big data*

Volumen	Velocidad	Variedad	Escalabilidad horizontal	Estructura no relacional	Tipo de <i>big data</i>
No	No	No	No	No	No
No	No	Sí	No	Sí	Sí, Tipo 1
No	Sí	No	Sí	Quizá	Sí, Tipo 2
No	Sí	Sí	Sí	Quizá	Sí, Tipo 3
Sí	No	No	Sí	Quizá	Sí, Tipo 2
Sí	No	Sí	Sí	Sí	Sí, Tipo 3
Sí	Sí	No	Sí	Quizá	Sí, Tipo 2
Sí	Sí	Sí	Sí	Sí	Sí, Tipo 3

Fuente: NIST.

Esta tabla permite evaluar una necesidad de negocio. Una reflexión interesante es que dentro de una misma organización diferentes escenarios de *big data* pueden plantearse para resolver distintas necesidades de negocio, lo que en definitiva apunta que necesitaremos distintas tecnologías.

Caso: Infojobs

InfoJobs es uno de los principales portales de empleo en Europa a través de una plataforma en línea que conecta empresas (que publican ofertas) y personas (que buscan nuevas oportunidades para su carrera profesional).

La compañía estaba interesada en poder ofrecer recomendaciones analizando las trayectorias profesionales más habituales de sus usuarios (lo que también se conoce como crecimiento profesional). Esto se traduce en analizar una gran cantidad de información que debe estructurarse como una red que relaciona candidatos (más de 4 millones), experiencias profesionales (más de 12 millones) y capacidades (más de 18 millones).

Usando la tabla anterior, estamos en una situación en la que priman el volumen y la variedad, y, por lo tanto, es un escenario *big data* de tipo 3.

1.3.2. Diferencias entre *business intelligence*, *business analytics* y *big data*

Acabamos de definir lo que es la inteligencia de negocio, la analítica de negocio y el *big data*. Vamos a comparar estas estrategias respecto a diferentes factores: **herramientas**, **foco**, **uso**, **tipo del dato**, **complejidad del dato** y **alcance**. Además, indicamos su nivel de madurez en el mercado. La tabla 4 describe las diferencias entre estas estrategias:

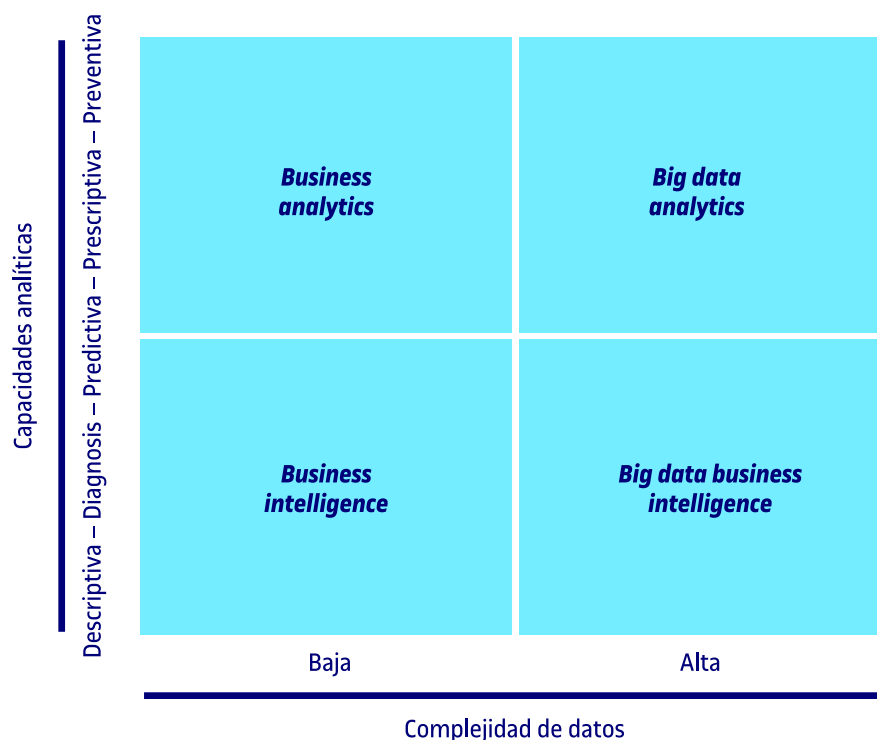
Tabla 4. Diferencias entre BI, BA y BD

	<i>Business intelligence (BI)</i>	<i>Business analytics (BA)</i>	<i>Big data (BD)</i>
Herramientas	Consultas, alertas <i>reporting</i> , OLAP, etc.	Clasificación <i>Clustering</i> Regresión	<i>Machine learning</i> , visualización, etc.
Foco	Qué y cómo pasó, cuántos, con qué frecuencia, cuál es el problema, qué es necesario hacer	Por qué está pasando, qué pasaría si todo continua igual, qué pasará a continuación, qué es lo mejor que puede pasar	Capturar, almacenar, procesar, analizar
Uso	Herramientas	• Proactivo • Predictivo • Prescriptivo	Todos
Tipo de dato	Estructurado	Estructurado Semiestructurado	Todo tipo
Complejidad del dato	Baja	Baja/media	Alta
Alcance	Dirección	Procesos	Vertical/procesos

Fuente: NIST.

Aunque queda claro que estas estrategias son diferentes, lo normal es que se combinen en los proyectos de explotación del dato. La figura 4 permite identificar casos de uso respecto a la complejidad del dato y las capacidades analíticas que se deben desarrollar en la organización.

Figura 4. Combinación de estrategias



Fuente: Josep Curto.

La lectura de esta gráfica se realiza a partir de sus ejes.

Por ejemplo, cuando necesitemos desarrollar capacidades analíticas descriptivas y la complejidad de dato sea alta, combinaremos *big data* con *business intelligence* formando un único sistema.

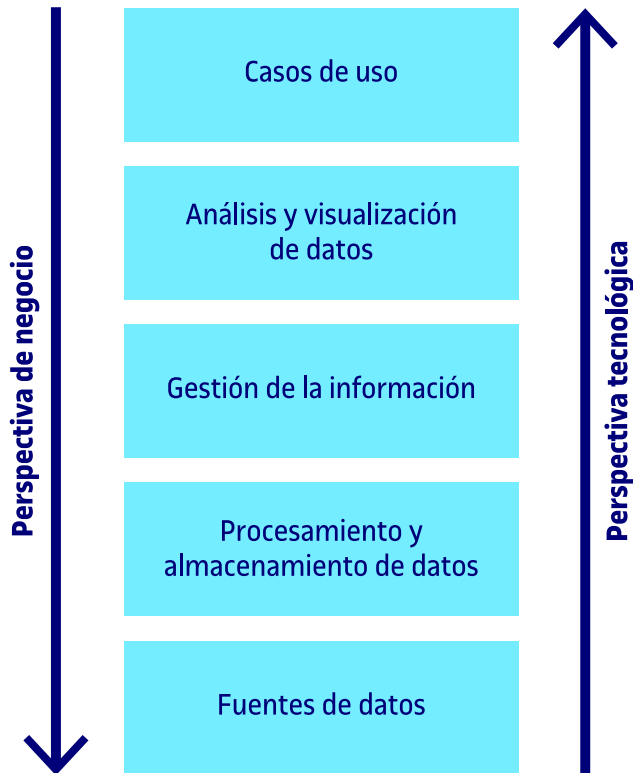
Una cadena de supermercados como, por ejemplo, Carrefour usa la inteligencia de negocio para comprender el rendimiento de cada uno de sus centros comerciales, usa la analítica de negocio para identificar los principales grupos de clientes (y las características que los definen) y usa *big data* (y técnicas de analítica) para implementar y desplegar un sistema de recomendación de productos para incentivar las compras.

Estas tres estrategias forman parte de un marco de referencia mucho más amplio con múltiples tecnologías. Es necesario, por lo tanto, introducir lo que se conoce como la plataforma del dato.

1.3.3. La plataforma del dato

La **plataforma del dato** es la plataforma tecnológica que soporta los casos de negocio que necesitan de las tecnologías que permiten procesar, almacenar, gestionar, analizar y visualizar datos. Esta plataforma debe aunar dos perspectivas: negocio (qué problema o necesidad se busca resolver en la organización) y tecnológica (qué tecnología permite dar una respuesta a la necesidad de negocio), tal y como ilustra la figura 5.

Figura 5. Perspectivas de negocio y tecnología



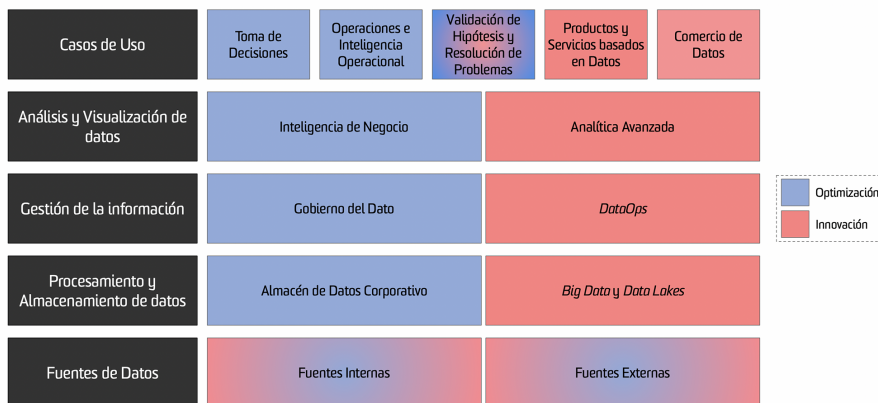
Fuente: Josep Curto.

La **plataforma del dato** soporta los casos de negocio de la organización alineada a las necesidades duales de optimización (azul) e innovación (rojo). Puede tener diferentes componentes distribuidas en distintas capas (tal y como ilustra la figura 6):

- **Casos de uso (*use cases*):** la organización tiene diferentes casos de uso donde puede aplicar *big data* y analítica de negocio. Estos son: toma de decisiones, inteligencia operacional y sistemas transaccionales, validación de hipótesis y resolución de problemas, creación de productos o servicios basados en datos, y comercio de datos.
- **Análisis y visualización de datos (*data analysis and visualization*):** en esta capa, la organización dispone de componentes tecnológicas para analizar el rendimiento pasado de la organización (inteligencia de negocio o *business intelligence*) y desplegar técnicas avanzadas de analítica (*advanced analytics*) que proporcionan capacidades predictivas y prescriptivas.
- **Gestión de la información (*information management*):** esta capa proporciona una gestión eficiente del dato como activo de valor (gobierno del dato), así como metodologías automatizadas orientadas a procesos, utilizadas por equipos analíticos y de datos, para mejorar la calidad y reducir el tiempo de ciclo de análisis de datos (DataOps).
- **Almacenamiento y procesamiento de datos (*data storing and processing*):** esta capa proporciona capacidades para el almacenamiento y procesamiento de fuentes estructuradas (almacén de datos corporativos o *enterprise data warehouse*) y *big data* / no estructuradas.

- **Fuentes de datos (*data sources*):** esta capa describe las fuentes de datos disponibles, distinguiendo en primera instancia internas frente a externas.

Figura 6. La plataforma del dato



Fuente: Josep Curto.

Cada una de las componentes tecnológicas dentro de las capas de análisis y visualización de datos, gestión de la información y procesamiento y almacenamiento de datos puede desplegarse en diferentes modalidades: **on-premise** (dentro del centro de datos de la organización), en la **nube** (*cloud computing*) o de forma **híbrida** combinando ambas. El departamento de analítica, a partir de las necesidades concretas del caso de uso, debe seleccionar las componentes tecnológicas adecuadas en cada capa (frecuentemente no son necesarias todas) y el mecanismo de despliegue.

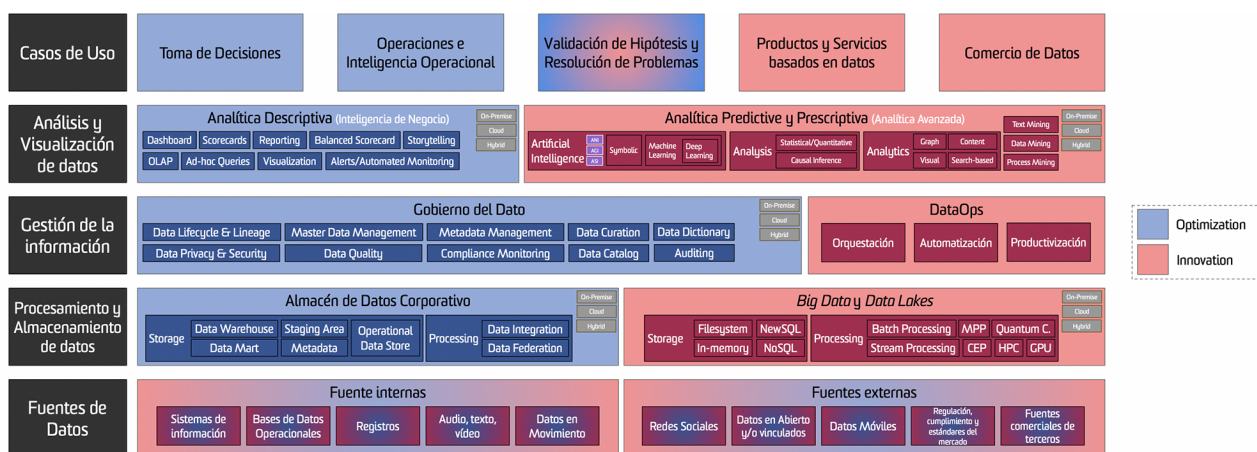
Es posible entrar en un mayor detalle en las tecnologías que forman cada una de las componentes (tal y como ilustra la figura 7):

- **Inteligencia de negocio:** incluye las tecnologías que permiten entender el rendimiento pasado de la organización, como, por ejemplo, cuadro de mando, listados de métricas, informes, cuadro de mando integral, visualización de datos, historias de datos, OLAP, consultas *ad hoc* y monitorización automática.
- **Analítica avanzada:** incluye las tecnologías que proporcionan capacidades predictivas y prescriptivas. Entre ellas destacan la inteligencia artificial y el aprendizaje automático, que introduciremos en el siguiente apartado.
- **Gobierno del dato:** incluye las tecnologías que posibilitan el ejercicio de autoridad, control y toma de decisiones compartida (planificación, vigilancia y aplicación) sobre la gestión de los activos de datos. Entre ellas destacan las componentes para calidad del dato o la gestión de datos maestros.
- **DataOps:** incluye las tecnologías que permiten la orquestación, la automatización y la productivización de proyectos analíticos. En esencia, se busca asegurar la escalabilidad y puesta en producción de las técnicas de análisis de datos.
- **Almacén de datos corporativo:** incluye las tecnologías para el almacenamiento (*storage*) y las de procesamiento (*processing*) para datos relacionales. Entre ellas destacan las técnicas de integración de datos, que permi-

ten preparar el dato para su consumo; o el almacén de datos, que permite guardar las magnitudes más relevantes de negocio para comprender su rendimiento.

- **Big data y data lakes:** incluye las tecnologías para almacenamiento (*storage*) y las de procesamiento (*processing*) para datos no relacionales. Entre ellas destacan NoSQL, que hace referencia a bases de datos que soportan datos dinámicos, por ejemplo, estructurados como un grafo.
- **Fuentes de datos:** incluye las fuentes de datos internas (como datos de sistemas de información, de bases de datos operacionales, de registros, audio, texto, vídeo y datos en movimiento –de sensores–) y externas (como datos de redes sociales, de móviles, en abierto y/o vinculados, relacionados con la regulación, el cumplimiento y estándares de mercado, y de fuentes comerciales de terceros).

Figura 7. La plataforma del dato (detalle completo)



Fuente: Josep Curto.

Inteligencia artificial y aprendizaje automático

Dentro de la plataforma del dato, y en particular, dentro de la componente de analítica avanzada, nos encontramos con la inteligencia artificial y el aprendizaje automático. Estos conceptos son actualmente el principal foco de las principales organizaciones orientadas al dato, como Amazon, Apple, Google o Facebook. Primero debemos definir este concepto:

Se entiende por **inteligencia artificial** la teoría y el desarrollo de sistemas informáticos capaces de realizar tareas que normalmente requieren inteligencia humana.

Cuando hablamos de tareas que requieren inteligencia humana, podemos pensar en la percepción visual, el reconocimiento de voz, la toma de decisiones y la traducción entre idiomas. Dentro de la inteligencia artificial tenemos diferentes versiones:

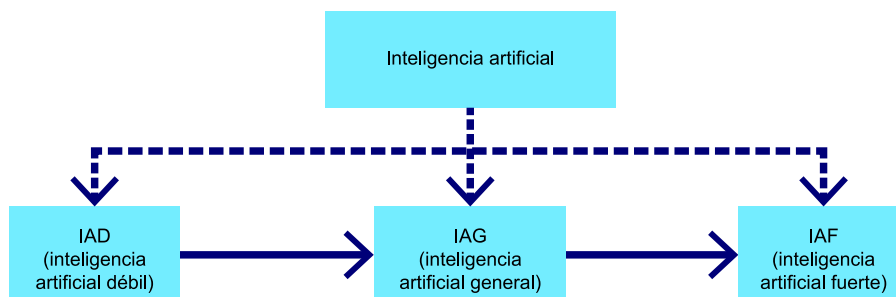
- **Inteligencia artificial débil (IAD):** se especializa en una tarea. Como, por ejemplo, los asistentes virtuales (como Siri o Alexa). En la actualidad, todos los casos de uso desplegados son de este tipo.
- **Inteligencia artificial general (IAG):** es cuando se ha llegado a un nivel cognitivo equivalente a un humano simultáneamente en múltiples dominios, como el procesamiento del lenguaje, el procesamiento de imágenes, el funcionamiento computacional y el razonamiento, etc.
- **Inteligencia artificial fuerte (IAF):** es cuando se superan las capacidades humanas en todos los aspectos, como, por ejemplo, toma de decisiones racionales o construir relaciones emocionales, entre otras. En este nivel, las capacidades cognitivas evolucionarán a una gran velocidad incrementando el *gap* entre seres humanos y entidades con inteligencia artificial.

Como se ilustra en la figura 8, hablamos de un proceso secuencial (IAD → IAG → IAF).

Observación

Cabe comentar que IAG e IAF solo son conceptos teóricos.

Figura 8. Amplitud de inteligencia

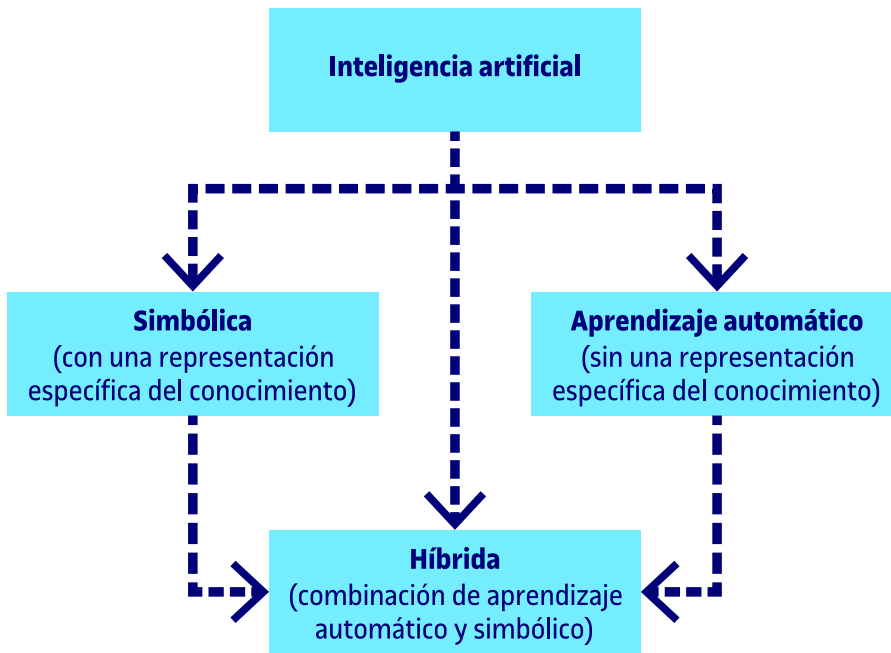


Fuente: Josep Curto.

Existe también otra perspectiva para comprender la inteligencia artificial, que es la capacidad de aprendizaje. Podemos considerar principalmente dos:

- **Simbólica:** requiere una representación específica y completa del conocimiento. En este enfoque necesitamos símbolos que describen la realidad y/o los problemas a los que nos enfrentamos, y consiste en encontrar mecanismos para la manipulación de dichos símbolos en la resolución de problemas (como podrían ser reglas de actuación).
- **Aprendizaje automático:** no requiere una representación específica y completa del conocimiento. En este enfoque tenemos observaciones de los eventos de la realidad y, a partir de dichas observaciones, se busca el modelo que responde a un problema específico.
- **Híbrido:** en la actualidad están emergiendo enfoques que combinan las dos capacidades de aprendizaje. El motivo principal para combinar los enfoques es que el aprendizaje automático, aunque ha logrado grandes éxitos, está orientado principalmente a la inteligencia artificial débil.

Figura 9. Capacidad de aprendizaje



Fuente: Josep Curto.

Lo que estamos comentando es que la gran mayoría de los casos de uso de analítica avanzada que están recibiendo mayor atención actualmente se fundamentan en aprendizaje automático. Según Arthur Samuel, podemos definir este concepto como:

«Se entiende por **aprendizaje automático** el campo de estudio que brinda a las computadoras la capacidad de aprender sin ser programado explícitamente».

Es decir, estamos hablando del descubrimiento y aprendizaje automático de patrones por una computadora:

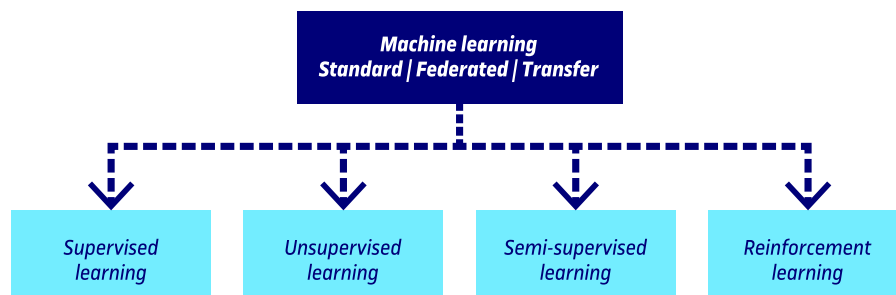
- **Aprendizaje:** a partir de datos y/o modelo inicial, la computadora aprende cómo mejorar el modelo.
- **Automático:** sin la necesidad de intervención de personas, más allá del modelo o los datos iniciales.

Actualmente, el aprendizaje automático incluye diferentes procesos de aprendizaje, como ilustra la figura 10:

- **Aprendizaje supervisado (*supervised learning*):** donde existen datos etiquetados de partida y los algoritmos aprenden a crear predicciones a partir de estos.
- **Aprendizaje no supervisado (*unsupervised learning*):** donde no existen datos etiquetados de partida y los algoritmos aprenden la estructura inherente que contienen dichos datos de partida.

- **Aprendizaje semisupervisado (*semi-supervised learning*):** donde existen algunos datos etiquetados de partida, aunque en su mayoría no lo están y se combinan elementos de aprendizaje supervisado y no supervisado.
- **Aprendizaje por refuerzos (*reinforcement learning*):** donde el algoritmo aprende qué hacer, cómo asignar situaciones a acciones, para maximizar una señal de recompensa numérica.

Figura 10. Tipos de aprendizaje automático



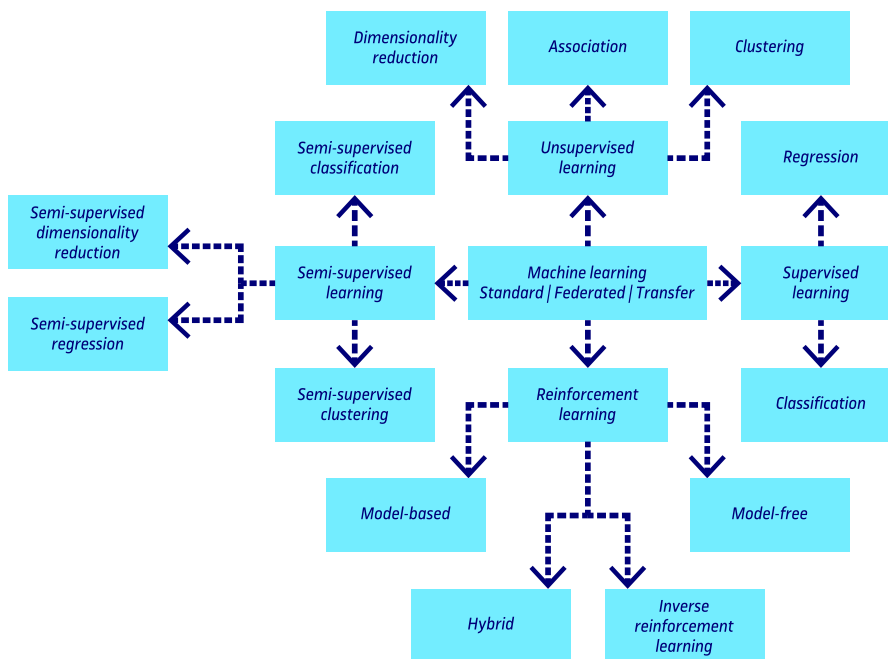
Fuente: Josep Curto.

Cada uno de estos procesos de aprendizaje permite responder a diferentes preguntas de negocio, como ilustra la figura 11:

- **Aprendizaje supervisado**
 - **Clasificación:** permite predecir resultados categóricos.
 - **Regresión:** permite predecir valores numéricos.
- **Aprendizaje no supervisado**
 - **Agrupación:** permite encontrar grupos de objetos similares.
 - **Reducción de dimensionalidad:** permite reducir la cantidad de variables usadas.
 - **Asociación:** permite encontrar reglas o, en general, recomendaciones.
- **Aprendizaje semisupervisado**
 - **Clasificación semisupervisada:** permite mejorar la clasificación supervisada minimizando los errores en los ejemplos etiquetados, pero también debe ser compatible con la distribución de entrada de instancias no etiquetadas.
 - **Agrupación semisupervisada:** a partir de los mínimos datos etiquetados, permite obtener agrupaciones mejor definidas que las obtenidas de datos no etiquetados.
 - **Reducción de dimensionalidad semisupervisada:** además de abundantes ejemplos sin etiquetar, el conocimiento del dominio en forma de restricciones por pares está disponible.
 - **Regresión semisupervisada:** se refiere al caso en el que la variable de salida tiene un valor real y hay una mezcla de datos etiquetados y no etiquetados.
- **Aprendizaje por refuerzos**

- **Basado en modelo:** el agente debe aprender un modelo de cómo funciona el entorno a partir de sus observaciones y luego planificar acciones utilizando ese modelo.
- **Sin modelo:** el agente no aprende un modelo del entorno a partir de observaciones, es un enfoque de prueba y error.
- **Híbrido:** combina elementos basados en modelo y sin modelo.
- **Aprendizaje de refuerzo inverso:** consiste en aprender los objetivos, valores o recompensas de un agente observando su comportamiento.

Figura 11. Tipos de aprendizaje automático



Fuente: Josep Curto.

Ciencia de datos

Uno de los motivos para implementar la plataforma de datos es introducir la disciplina de la ciencia de datos dentro de la organización.

Se entiende por **ciencia de datos** la disciplina aplicada que busca resolver problemas de negocio mediante soluciones que combinan matemáticas, estadística e ingeniería de software de forma que son reproducibles, replicables, interpretables y se ajustan al propósito de negocio.

En esencia, en esta disciplina, el equipo experto debe combinar adecuadamente las componentes de la plataforma del dato para resolver una necesidad de negocio. Es necesario destacar que, en la ciencia de datos, es predominante el uso de la analítica avanzada y *dataops*.

Opciones para la plataforma del dato

En el momento de considerar una plataforma de datos, las organizaciones tienen diferentes opciones estratégicas:

- **Construir (*build*):** la organización se responsabiliza de diseñar, integrar, mantener y securizar la plataforma tecnológica.
- **Comprar (*buy*):** la organización adquiere una solución integrada de terceros como plataforma tecnológica.
- **Alquilar (*lease*):** la organización alquila (en el formato SaaS, *software as a service*) la plataforma del dato.
- **Socio (*partner*):** la organización se fundamenta en la plataforma del socio tecnológico de referencia.

Tal y como apuntan Gandhi, Thota, Kuchembuck y Swartz (2018), cada opción tiene sus ventajas e inconvenientes, que se resumen en la tabla 5:

Tabla 5. Fortalezas y limitaciones

Estrategia	Fortalezas	Limitaciones
Construir	• Las soluciones internas siempre serán las más adaptadas a las necesidades de vuestra empresa. • Las soluciones que requieren altos niveles de seguridad y control invitan a asociaciones externas limitadas (aunque los expertos en seguridad pueden mitigar el riesgo).	• Construir supone una inversión significativa, por lo que es necesaria una evaluación de la infraestructura tecnológica, las capacidades de talento y los recursos para el desarrollo y mantenimiento continuos.
Comprar	• Inversión reducida en comparación con la construcción interna. • Las soluciones estándar pueden satisfacer necesidades estratégicas mediante desarrollo interno.	• Puede ser un compromiso tan significativo como la estrategia de construcción. • Requiere una comprensión profunda de las funciones, flexibilidades y limitaciones de las soluciones estándar.
Alquilar	• La mercantilización y la sofisticación de las soluciones tecnológicas significa que hay muchas opciones para respaldar la estrategia. • Compromiso reducido a largo plazo. • Se incentiva a los proveedores a proporcionar servicios de consulta para garantizar que las soluciones satisfagan las necesidades.	• Las opciones de alquiler disponibles podrían no cumplir con los requisitos organizacionales. • La proliferación de proveedores puede complicar la operacionalización.
Socio	• Evita el fuerte compromiso requerido para construir o comprar una plataforma. • Permite la personalización que un modelo de alquiler no siempre proporciona.	• Requiere capacidades de gestión de proveedores y de inteligencia del mercado.

Fuente: Gandhi, Thota, Kuchembuck y Swartz.

Referencia bibliográfica

S. GANDHI, B. THOTA, R. KUCHEMBUCK y J. SWARTZ. Demystifying data monetization. En: *MIT Sloan Management Review*, 2018. Disponible en: <https://sloanreview.mit.edu/article/demystifying-data-monetization/>

Opciones para la estrategia tecnológica

Aquellas organizaciones que buscan convertirse en una organización orientada al dato deben decidir qué tipo de estrategia seguir. Tradicionalmente, las organizaciones han tenido diferentes enfoques respecto a la tecnología:

- **TI como fuente de ventajas competitivas:** la organización se apalanca en la tecnología para generar nuevos modelos de negocio, para mejorar procesos y liderar el mercado. Este tipo de organizaciones busca un control de TI, por lo que tiene un equipo de expertos capaz de gestionar y desarrollar tanto proyectos externos como trabajar con terceros siempre y cuando no haya posibilidades de pérdida de valor.
- **TI es una comodidad (también conocido como *commodity* en inglés):** la organización considera que la tecnología es un mero soporte a negocio y frecuentemente cualquier desarrollo o implementación tecnológica se sustenta en proveedores.

Al mismo tiempo, la **complejidad organizacional** es un factor relevante en la adquisición tecnológica. Aquellas empresas de baja complejidad, bien porque son pequeñas o porque tienen poca presencia global, pueden permitirse tener una estrategia diferente que una empresa con una alta complejidad.

Por otro lado, en el momento de seleccionar una plataforma de explotación de datos (o una colección de soluciones), las organizaciones deben tener en cuenta que existen diversos aspectos que hay que valorar, como son el tipo de despliegue de la solución o el tipo de solución.

- **Despliegue/implantación:** las soluciones TI pueden desplegarse internamente (máximo control), en la nube (control limitado, mayor flexibilidad) o como servicio (menor control, máxima flexibilidad).
- **Tipo de solución:** una plataforma de explotación de datos puede ser genérica o de *best-of-breed*. Frecuentemente se usa el concepto *best-of-breed* para denominar una solución que está formada por la composición de distintas herramientas, siendo cada una de ellas la mejor de su clase para las soluciones de nicho. Una solución *best-of-breed* ofrece la mejor solución disponible, mientras que una genérica cubrirá muchos de los casos de uso para la explotación de datos. Cabe comentar que dentro de las plataformas genéricas pueden estar enfocadas a la gran empresa (denominadas *tier I*) o a empresas medianas o pequeñas (denominadas *tier II* o *III*). La siguiente tabla compara ambos tipos de solución.

Tabla 6. *Best-of-breed* frente a genérica

	<i>Bestof-breed</i>	Genérica
Compra	Mayor precisión	Más barata, pero falta de transparencia de cada componente.

Fuente: Josep Curto.

	<i>Bestof-breed</i>	Genérica
Formación	Tiende a ser más cara y soportada por una comunidad más pequeña.	El conocimiento tiende a ser más generalizado, y se puede reutilizar en otros proyectos.
Administración y soporte	Frecuentemente más simple y accesible al usuario final. Menor integración con ecosistemas de terceros.	Orientada a aplicaciones que afectan a toda la organización. Mayor integración en el ecosistema propio.
Funcionalidad	La mejor disponible, pero no cubre todas las necesidades generales.	Alto rango; sin embargo, no presenta funciones específicas en algunas áreas.

Fuente: Josep Curto.

2. Beneficios, barreras y casos de uso

2.1. Beneficios organizacionales

La implementación de *big data* y analítica proporciona una serie de beneficios que es necesario conocer. Vamos a diferenciar los beneficios generales, tangibles e intangibles.

2.1.1. Beneficios generales

La implantación de los sistemas de información analíticos (y en particular, la plataforma del dato) y transformarse en una organización orientada al dato proporciona diversos beneficios generales, entre los que podemos destacar:

- **Crear un círculo virtuoso de la información:** los datos se transforman en información que genera conocimiento, este conocimiento permite tomar mejores decisiones y cada decisión se traduce en resultados que generan nuevos datos.
- **Visión única:** los datos se guardan de forma consistente, regular y representando los eventos relevantes para la organización. Esto permite crear una visión única, conformada, histórica, persistente y de calidad de toda la información relevante para la organización.
- **Métricas relevantes:** la implementación de este tipo de proyectos permite identificar, crear, manejar y mantener métricas, indicadores clave de rendimiento (*key performance indicator*, KPI) e indicadores clave de metas (*key goal indicator*, KGI), fundamentales para la empresa.
- **Acceso:** habilitar el acceso a información actualizada tanto a nivel agregado como en detalle.
- **Alinear la organización:** la implementación de este tipo de proyectos permite reducir el diferencial de orientación de negocio entre el departamento TI y la organización.
- **Incrementar el conocimiento organizacional:** la implementación de este tipo de iniciativas permite desarrollar una mejor comprensión y documentación de los sistemas de información en el contexto de una organización.
- **Incremento de la competitividad:** transformarse en una organización orientada al dato proporciona una mejora de la competitividad de la organización y de los resultados de negocio al ser capaces de :
 - Diferenciar lo relevante sobre lo superfluo.
 - Acceder más rápido a la información.
 - Tener mayor agilidad en la toma de las decisiones.

2.1.2. Beneficios tangibles

Entre los beneficios tangibles que se pueden conseguir podemos destacar:

- Incrementar el valor para la organización desplegando estrategias orientadas al cliente fundamentadas en evidencias.
- Crear nuevos productos basados en las preferencias de clientes, empleados y proveedores.
- Identificar las propiedades de productos que realmente satisfacen las necesidades de clientes, empleados y proveedores.
- Incrementar la efectividad de las acciones de marketing generando más ingresos por acción.
- Automatizar estrategias de precio basadas en evidencias.
- Reducir el tiempo y los costes vinculados a la toma de decisiones al reducir la latencia en la captura, el procesamiento y el análisis del dato.
- Mejorar los costes de retención y adquisición de clientes y empleados al tener estrategias que identifican sus características y su previsión al abandono.
- Reducir el riesgo asociado al cliente al identificar perfiles por impago o fraude.
- Mejorar la productividad de la fuerza de venta a partir de la optimización de los programas de incentivos.
- Analizar y poder optimizar cadena de suministro.

2.1.3. Beneficios intangibles

Entre los beneficios intangibles podemos destacar:

- Aportar conocimiento y herramientas a la dirección para la toma de decisiones sobre clientes, mercado y competencia.
- Detectar cambios en el mercado y adquirir habilidad para reaccionar a estos con rapidez, gracias a su conocimiento.
- Capacitar a la organización para la creación de experiencias de cliente y empleados basándose en un conocimiento profundo de las interacciones internas y externas.
- Posibilitar la visión 360° del cliente al poder analizar los datos generados en todos los canales (en línea y fuera de línea).

Cabe destacar que, a mayor cantidad de datos integrados en la plataforma del dato disponibles para su análisis y reutilización, mayor riqueza de conclusiones se obtendrá y, por tanto, mayor valor generado a partir de los datos.

2.2. Barreras

Transformarse en una organización orientada al dato no es un proceso sencillo o corto, tal y como apunta Thomas H. Davenport (2005). En dicho proceso de cambio, las empresas deben superar múltiples retos y crear una estructura adecuada que sustenta y apoya la creación de una cultura analítica.

Conocer las barreras existentes, así como una falta de estrategia, es sumamente importante.

2.2.1. Barreras

Algunos de los retos y barreras existentes son comunes a cualquier tipo de proyecto. Por ejemplo,

- No se ha definido claramente el alcance, los requisitos de negocio y los beneficios.
- Falta de gestión en el proyecto.
- Falta de comunicación entre TI y negocio.
- Falta de involucración por parte del usuario.
- Falta de propiedad del proyecto por parte de negocio.
- Inversión/presupuesto inadecuado.
- Planificación inadecuada o inexistente.
- Alcance del cambio no considerado.
- Problemas de relación entre departamentos involucrados en el proyecto.
- Falta de objetivos organizacionales.
- Falta de cambio cultural.
- Complejidad tecnológica.
- Foco en la tecnología, en lugar de en el impacto de negocio.
- Los datos necesarios no son capturados, fáciles de acceder o procesar.
- Calidad de datos baja o inexistente.
- Tecnología inadecuada para el proyecto.
- Capacidades TI limitadas.
- Fabricante/integrador erróneo o no alineado a las necesidades de la organización.

En definitiva, todo el mundo es responsable de que este tipo de proyectos fracase. Cabe destacar que pueden existir motivos personales, como, por ejemplo:

- Preferencia de la intuición sobre el dato.
- Interés por amañar el sistema por interés propio.
- Continuar con la toma de decisiones basada en las opiniones de terceros.

Dado que existen tantas barreras, tiene sentido clasificarlas en grandes bloques. En particular, las podemos clasificar en:

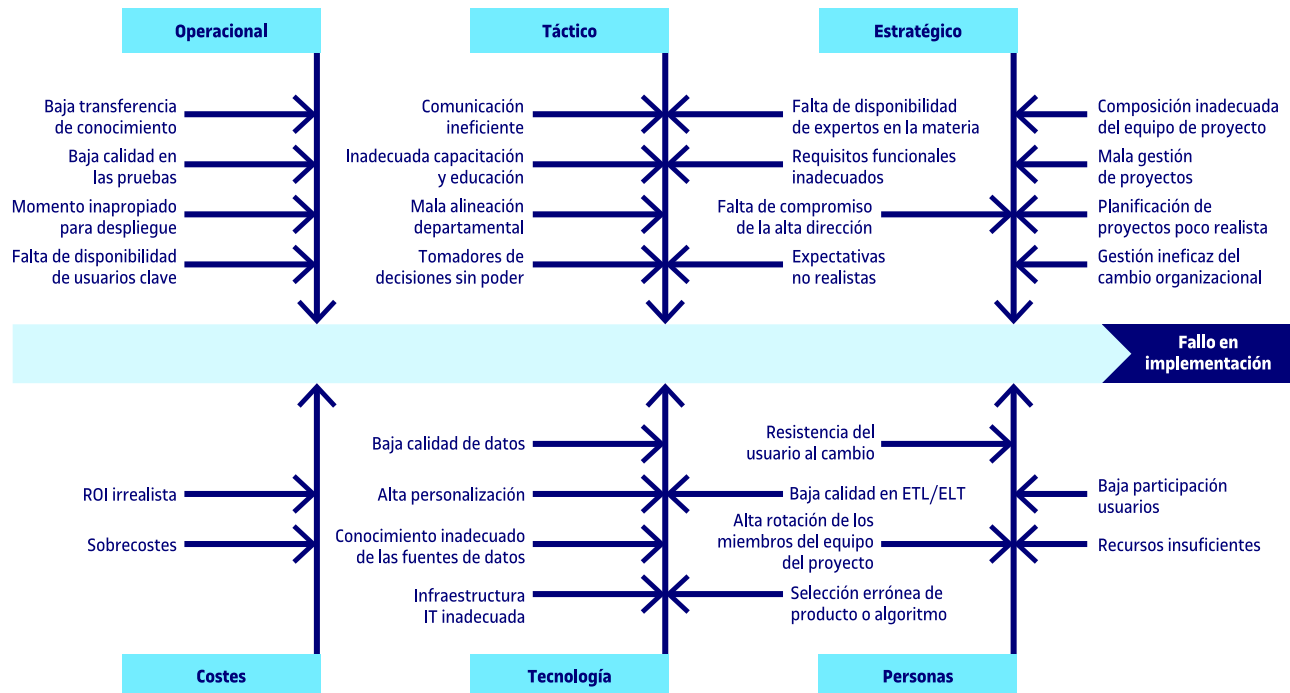
Referencia bibliográfica

T. DAVENPORT, D. COHEN y A. JACOBSON. *Competing on Analytics, Working Knowledge Research Report*. Babson Executive Education, 2005.

- **Operacionales:** vinculadas con aspectos operacionales, como baja transferencia de conocimiento (de negocio y/o de la solución), baja calidad en las pruebas (de rendimiento, calidad), momento no adecuado de despliegue o falta de disponibilidad de usuarios clave.
- **Tácticas:** vinculadas con aspectos tácticos, como comunicación ineficiente, inadecuada capacitación o entrenamiento, mala alineación departamental, personas clave sin poder de decisión real, falta de disponibilidad de los expertos en la materia, requisitos funcionales inadecuados o expectativas irrealistas, entre otros.
- **Estratégicas:** vinculadas con aspectos estratégicos, como la composición inadecuada del equipo de proyecto, mala gestión de proyectos, planificación de proyectos poco realista, falta de compromiso en la alta dirección o gestión poco eficaz del cambio organizacional, entre otros.
- **Costes:** vinculadas con aspectos financieros del proyecto, como, por ejemplo, el retorno de la inversión (ROI, *return on investment*) irreal o presupuesto incorrecto, entre otros.
- **Tecnología:** vinculadas con aspectos de tecnología, como, por ejemplo, baja calidad en los procesos de extracción, transformación y carga de los datos (ETL, *extract, transform and load*), baja calidad de datos, exceso en personalización de la solución, conocimiento insuficiente de las fuentes de datos o dimensionamiento no adecuado de la infraestructura tecnológica, entre otros.
- **Personas:** vinculadas con aspectos como, por ejemplo, la baja participación o recursos insuficientes dedicados al proyecto, entre otros.

La figura 12 ilustra las múltiples barreras que pueden generar fallos en la implementación.

Figura 12. Barreras en la implementación



Fuente: Josep Curto.

2.2.2. ¿Cómo detectar que no existe una estrategia?

Desplegar un proyecto de *big data* y analítica de negocio en el seno de una organización no es un proceso sencillo. Las buenas prácticas indican que, para llegar a buen puerto, es necesario tener una estrategia analítica suficiente que coordine las tecnologías, el uso, los procesos de madurez y la metodología que usar.

Es posible detectar que no existe una estrategia definida a través de los siguientes puntos y percepciones en el seno de una organización:

- Los usuarios identifican el departamento de informática (*information technology*, IT) o departamento IT como el origen de sus problemas en los sistemas de información analíticos.
- La dirección considera que la inteligencia de negocio es otro centro de coste.
- El departamento de IT continúa preguntando a los usuarios finales sobre las necesidades de los informes.
- Los sistemas de información analíticos están soportados por *help desk*.
- No es posible medir el uso de los sistemas de información analíticos.
- No es posible medir el retorno de la inversión (*return on invest*, ROI) del proyecto de los sistemas de información analíticos.
- No hay un plan para desarrollar, contratar, retener y hacer crecer el equipo de *big data* y analítica de negocio.
- No se conoce si la empresa tiene una estrategia para el *big data* y analítica de negocio.
- No existe un responsable funcional (o bien el asignado no es el adecuado).

- No existe un centro de competencia.
- Existen múltiples soluciones en la organización distribuidas en diferentes departamentos que repiten funcionalidad.
- No hay un plan de formación real y consistente de uso de las herramientas.
- Alguien cree que es un éxito que la información consolidada esté a disposición de los usuarios finales al cabo de dos semanas.
- Los usuarios creen que los datos y/o la información de sistemas de información analíticos no es correcta.
- No existe una cultura analítica en la que el dato y los hechos sean relevantes para tomar decisiones sea cual sea el nivel de la organización.

2.2.3. La necesidad de los centros de competencia

Teniendo en cuenta las potenciales barreras para el despliegue de este tipo de proyectos, los diferentes roles que participan en ellos, así como las distintas opciones para desarrollar una organización orientada al dato, queda patente que es necesario definir una estrategia precisa y adecuada para evitar fallos y una mayor complejidad en la organización.

Algunas organizaciones están empezando a considerar la creación de un grupo organizativo que se encargue de diseñar y ejecutar una estrategia tecnológica de explotación de datos a largo plazo. Es decir, la creación de un **centro de competencias**.

Un centro de competencias es una estructura organizacional usada para coordinar competencias TI dentro de una organización.

El centro de competencias tiene diversas responsabilidades:

- Proporcionar conocimiento experto en proyectos.
- Ser fuente de conocimiento y recursos.
- Ayudar a identificar claros objetivos de negocio.
- Reconocer que los proyectos de analítica requieren gestión del cambio y participar para que dicho cambio se haga realidad.
- Involucrar a las personas necesarias para el cambio.
- Habilitar transparencia en tiempo real.
- Tomar acciones con respecto a proyectos y definición de la cultura orientada al dato y evidencias.
- Evitar la pérdida de control en proyectos.
- Asegurar el apoyo continuo por parte de la dirección.
- Asegurar que se despliega la tecnología adecuada para el proyecto.
- Desplegar una metodología de trabajo práctica, pragmática y simple.
- Buscar el camino de menor resistencia para el proyecto.

El centro de competencias tiene diferentes componentes:

- **Tecnología:** se encarga de probar nuevas soluciones, de validar la tecnología, de identificar nuevas aplicaciones para la organización y de determinar su impacto en la organización, así como del soporte y del mantenimiento.
- **Consultoría:** se encarga del desarrollo o la consultoría de los proyectos o de la gestión de terceros, así como de la gestión del dato.
- **Educación:** se encarga de formar y educar a la organización en la cultura analítica y en las soluciones desplegadas con el objetivo de favorecer el cambio.
- **Administración:** se encarga de definir la gobernanza de la explotación de datos, lo que supone definir políticas, estándares y guías que seguir. Además, se ocupa de la vigilancia continua del rendimiento de las iniciativas y los comentarios de mejora por parte de los usuarios.

Estas funciones requieren que el centro de competencia tenga poder de decisión para gestionar proveedores tecnológicos e integradores, y que se puedan definir acuerdos marco de colaboración a largo plazo.

2.2.4. Buenas prácticas

Teniendo en cuenta las posibles barreras que pueden tener este tipo de proyectos, es necesario observar ciertas buenas prácticas que pueden ayudar al despliegue de estas iniciativas:

- Asegurar el apoyo y el compromiso de la dirección en el proyecto.
- Crear y establecer un contexto de colaboración y comunicación entre el equipo que desarrolla el proyecto y los beneficiarios.
- Foco en el liderazgo, la planificación y la gestión del proyecto.
- Desarrollo coordinado de este tipo de proyectos dentro de una iniciativa mayor de desarrollo de una cultura de toma de decisiones orientada a las evidencias y fundamentada en el dato.
- Identificar iniciativas realistas con objetivos medibles.
- Establecer y mantener un enfoque pragmático del uso de *big data* y analítica de negocio para evitar el exceso de análisis.
- Poner especial atención en la formación de los beneficiarios.
- Seguir una metodología de despliegue de proyectos de analítica.

2.3. Casos de uso

Siguiendo la estructura de la plataforma del dato, vamos a introducir diferentes casos de uso siguiendo los movimientos organizacionales:

- toma de decisiones
- operaciones e inteligencia operacional
- validación de hipótesis y resolución de problemas
- productos y servicios basados en datos

- comercio de datos

Vamos a explicar en detalle cada uno de estos movimientos organizacionales.

2.3.1. Toma de decisiones

El primer caso de uso es la toma de decisiones.

Esta aproximación consiste en la ampliación de las capacidades tradicionales de toma de decisiones mediante las tecnologías de la plataforma del dato.

Caso: NH

NH es una cadena hotelera con más cuatrocientos hoteles en veinticinco países. Dentro de la estrategia de mejorar el servicio para sus clientes, la compañía cada año selecciona diversos hoteles sobre los que hará mejoras. Las mejoras que se realizan van desde la ampliación del personal hasta la creación de nuevas instalaciones. Para tomar la decisión de «dónde es necesario invertir en este periodo», NH se ha fundamentado tradicionalmente en dos fuentes de datos:

- Los datos financieros consolidados en el *data warehouse* de la compañía.
- Una serie de encuestas realizadas a los clientes para conocer su satisfacción de los servicios e instalaciones del hotel. Estas encuestas no son exhaustivas y no cubren todos los hoteles ni todos los clientes por los costes asociados con su realización.

En los últimos años, en NH se han dado cuenta de que, para conocer la satisfacción del cliente, así como las áreas de mejora, la información relevante se encuentra más allá del perímetro de la organización. Los clientes de NH comparten sus impresiones a través de diferentes canales, como pueden ser TripAdvisor, Yelp o Expedia. Es decir, estamos hablando de fuentes de datos externas a la organización y además no estructuradas o con diferentes formatos.

El enfoque de la organización ha sido complementar la información financiera en el *data warehouse* con información externa que se almacena y se procesa con tecnologías de *big data* y minería de texto (para extraer los comentarios relevantes para la mejora de los hoteles), y que permite mejorar y complementar la toma de decisiones en un proceso ya existente.

2.3.2. Operaciones e inteligencia operacional

El segundo caso de uso son las operaciones y la inteligencia operacional, que suceden en tiempo real.

Esta aproximación consiste en la aplicación de estas tecnologías en el ámbito de operaciones tanto para el control y el análisis de proceso de negocio, como para el diseño y la implementación de sistemas transaccionales.

Este segundo escenario trasciende de la toma de decisiones y permite entender por qué las tecnologías analíticas están llamadas a ser muy relevantes dentro de las tecnologías de información. Es previsible que se integren de modo natural en múltiples aplicaciones.

Por lo tanto, estamos hablando, por un lado, de sistemas de inteligencia y detección de patrones en tiempo real y, por el otro, de sistemas operacionales que o bien por sus necesidades en escalabilidad o por su complejidad en el esquema de los datos ya no se fundamentan en tecnologías relacionales.

Caso: Santander/CaixaBank

Uno de los puntos más relevantes para muchas organizaciones es cuando interactúan con sus clientes. Es lo que llamamos momentos de la verdad. Entre estos momentos destaca cuando el cliente se pone en contacto con una organización para la resolución de una incidencia. En épocas anteriores se ha automatizado el proceso (mediante sistemas de respuesta predefinida) o se ha dividido el servicio en diversas capas dentro y fuera de la organización para tener diferentes niveles de servicio.

En el contexto financiero, son varias las entidades financieras españolas que ya usan *speech analytics* para entender las emociones de sus clientes durante sus interacciones en una llamada. Es posible, por lo tanto, detectar cuándo va a disminuir la satisfacción del cliente y actuar consecuentemente.

2.3.3. Validación de hipótesis y resolución de problemas

Uno de los escenarios más importantes es la validación de hipótesis y resolución de problemas.

Este escenario consiste en encontrar soluciones para problemas de negocio que no han sido anteriormente abordados en una organización y para los cuales no hay preguntas predefinidas.

Es decir, se busca conocer qué ha sucedido, qué factores son los más relevantes y el porqué. Es necesario crear hipótesis y validarlas a través de la técnica más adecuada y eficiente. Este tipo de aplicación es el equivalente a tener a Sherlock Holmes en casa. Es, en definitiva, un entorno que debe ser lo suficientemente flexible para funcionar en diferentes escenarios de necesidades.

El resultado puede ser una solución puntual o una propuesta que pase a convertirse en uno de los otros escenarios.

Caso: Sky

Sky es una empresa que produce y distribuye contenidos de vídeo tanto en directo como bajo demanda con presencia en diversos países europeos. La distribución de contenidos de vídeo se realiza a través de redes informáticas conocidas como *content delivery networks*. Estas redes están formadas por múltiples elementos tanto pertenecientes a la propia compañía como a terceros, que deben asegurar que la distribución se realiza manteniendo el nivel de calidad contratado por el cliente. Frecuentemente, el proceso de transmisión de vídeo a través de la red se controla y se mide para asegurar su correcto funcionamiento. Este es el caso de Sky. Sin embargo, meses atrás tuvo un gran fallo en su *content delivery network* durante la transmisión de la jornada futbolística en el fin de semana que producía errores de acceso al sistema, congelación de la imagen o transmisión de imágenes en baja calidad. A pesar de tener un sistema de monitorización, Sky no conocía los motivos por los que había sucedido esta caída de calidad en el servicio.

El enfoque de la organización ha sido contratar a un experto para investigar lo sucedido. Lo que se traduce en este caso en trabajar con millones de registros en formato log e investigar las causas del error. Tras aplicar lo que se conoce como *root cause analysis*, una técnica para encontrar de forma sistemática los factores que provocan un fallo, a un conjunto de datos almacenados en tecnologías de *big data* por su tamaño, se encontraron las razones del error y se propusieron una serie de mejoras para la red.

2.3.4. Productos y servicios de datos

El cuarto escenario de uso es la creación de productos y servicios basados en datos.

El dato se transforma en la pieza angular para mejorar la experiencia de uso del producto y servicio o para el diseño y despliegue de este.

Estamos hablando de modelos de negocio en los que el dato y los algoritmos analíticos generan valor tanto para el cliente como para la organización, y por ello modifican todos los aspectos primordiales del modelo de negocio.

Caso: Nest

Nest es una empresa que produce dispositivos inteligentes, en particular termostatos, detectores de humo y cámaras de vigilancia, entre otros. Fue adquirida por Google en 2014.

Los productos creados por Nest son un ejemplo de producto basado en datos y algoritmos. Por ejemplo, el termostato contiene diferentes tipos de sensores para detectar a las personas y los animales presentes en el hogar, así como la temperatura de este. Además, va registrando las preferencias de las personas que viven en casa. A saber, a qué hora están en casa, cuántas personas, en qué habitaciones se encuentran y cuál es la temperatura a la que prefieren su casa. Y lo combina con información contextual como dónde se encuentra la casa y en qué época del año. Con todos estos datos, se crea un perfil de preferencias y a partir de un cierto momento, el dispositivo empieza a trabajar de forma automática. Este proceso automático permite reducir el coste energético.

Existe también otro potencial beneficiario de estos datos, aunque en formato agregado: las empresas productoras y distribuidoras de energía. A partir de los datos de todos los usuarios de Nest en aquellas zonas geográficas en las que ofrecen servicio, pueden conocer las necesidades energéticas y, por lo tanto, ajustar la demanda.

2.3.5. Comercio de datos

El último escenario de uso es el comercio de dato, también referenciado como **monetización de datos**.

El dato se prepara para su venta a terceros.

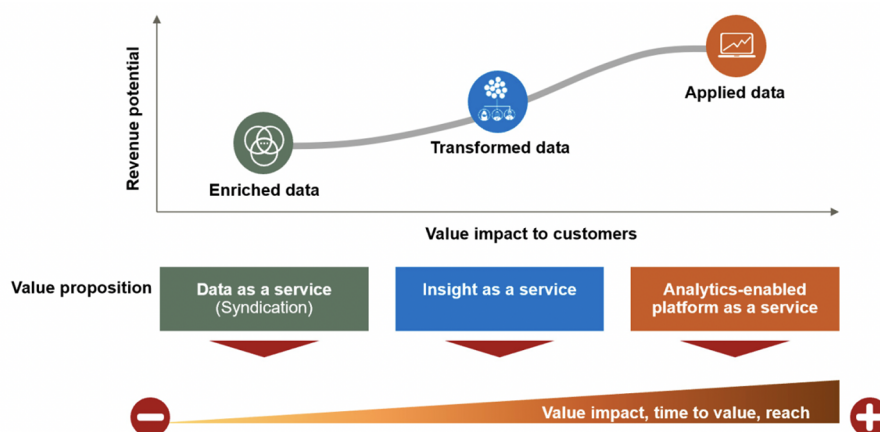
Esto puede incluir diversos procesos, como agregación, transformación y distribución del dato o, en el caso de contener información sensible, enmascarar dichos datos para que el conjunto final contenga datos anónimos. Este tipo de uso también puede derivar en diseñar una plataforma *ad hoc*. El dato puede comercializarse en bruto o en forma de conocimiento.

En el caso de que la organización apueste por monetizar estos activos puede considerar:

- **Datos como servicio:** consiste en ofrecer datos propios enriquecidos para el consumo de los clientes de la organización.
- **Conocimiento como servicio:** consiste en ofrecer análisis (fundamentados en datos enriquecidos) para el consumo de los clientes de la organización.
- **Plataforma como servicio:** consiste en ofrecer los datos enriquecidos y los mecanismos de análisis para el consumo de los clientes de la organización.

Según Gandhi, Thota, Kuchembuck y Swartz (2018), estamos hablando de un espectro de opciones que incrementan la propuesta de valor, pero que requieren un mayor esfuerzo organizativo, como ilustra la figura 13:

Figura 13. Monetización del dato



Fuente: Thota, Kuchembuck y Swartz.

Referencia bibliográfica

S. GANDHI, B. THOTA, R. KUCHEMBUCK y J. SWARTZ. Demystifying data monetization. En: *MIT Sloan Management Review*, 2018. Págs. 1–9.

Caso: Vodafone/TomTom

Vodafone es una conocida compañía que proporciona servicios de telecomunicaciones a nivel mundial. TomTom es una compañía que ofrece productos y servicios de GPS.

Los servicios de TomTom permiten conocer la ruta óptima a un conductor. La calidad de este servicio depende de trabajar con datos actualizados y contextualizados, que incluyan accidentes o atascos de tráfico. Por ello, TomTom compra datos de terceros como Vodafone.

En este caso particular, Vodafone comercializa los datos agregados, anónimos y geo-localizados de los usuarios de su red. En el caso de tener una gran acumulación de usuarios en un mismo lugar (y estar dicho lugar en una carretera), esto se traduce en una situación de atasco o accidente, por lo que TomTom puede usar esta información para proponer una ruta alternativa y ofrecer una mejor experiencia de cliente.

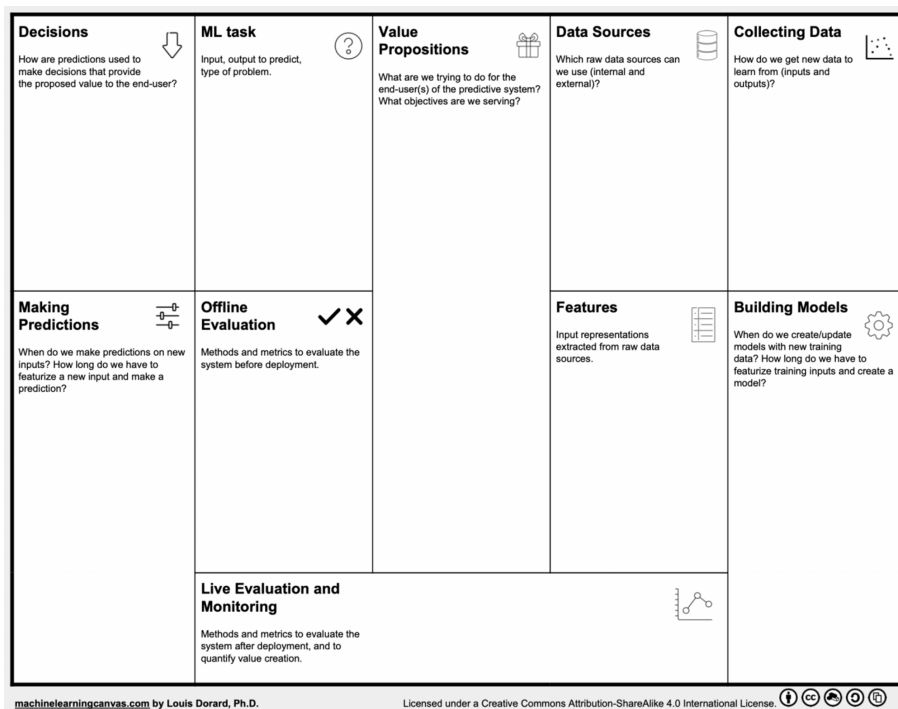
2.3.6. Cómo identificar y discutir oportunidades analíticas

Aquellas organizaciones interesadas en identificar y discutir oportunidades analíticas tienen a su disposición el *machine learning canvas* (que sigue una estructura similar al *business model canvas*) creado por Louis Dorard. Este modelo permite discutir una oportunidad en función de diferentes dimensiones, como ilustra la figura 14:

Enlace recomendado

Para consultar el documento original de Louis Dorard, podéis acceder a: <https://www.louisdorard.com/machine-learning-canvas>

Figura 14. *Machine learning canvas*



Fuente: Louis Dorard.

- decisiones
- tareas analíticas (o de forma particular tareas de *machine learning*)
- propuesta de valor
- fuentes de datos
- proceso de recopilación de datos
- predicciones que realizar

- variables relevantes para el modelo
- construcción del modelo de negocio
- proceso de evaluación fuera de línea (en desarrollo)
- proceso de evaluación en vivo y monitorización (en producción)

Existen otros enfoques basados en el *business model canvas*, como el de Ivan Marin (2019).

Referencia bibliográfica

I. MARIN Data Science and Development Team Remote Communication: the use of the Machine Learning Canvas. En: 2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE), Montreal, QC, Canadá, 2019, págs. 18–21, doi: 10.1109/ICGSE.2019.00018.

3. Fases, madurez y aspectos éticos

Mejorar el posicionamiento de una organización ha sido objeto de estudio durante las últimas décadas tanto dentro como fuera de la academia. Cuando una organización cuenta con una mejor posición que sus rivales, decimos que tiene una **ventaja competitiva** que permite asegurar a los clientes y defenderse contra las fuerzas competitivas (Porter, 1998).

- Por ejemplo, Delta Airlines, aerolínea de Estados Unidos, ha sido considerada durante muchos años un ejemplo en el uso de la eficiencia operacional para destacar por encima de otras compañías en su mismo mercado.
- Otro ejemplo es Apple, que ha destacado durante varios años por ofrecer productos altamente integrados a nivel de software y hardware (como iPod, iPhone o AirPods) con diseños orientados al cliente como mecanismo de diferenciación.

Referencia bibliográfica

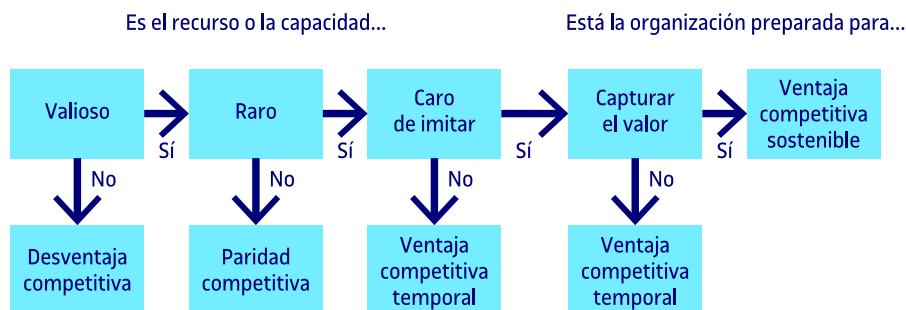
Michael PORTER. *Competitive Advantage Creating and Sustaining Superior Performance*. Free Press, 1998.

Tradicionalmente, las ventajas competitivas se han fundamentado en:

- **Liderazgo en coste:** la organización es capaz de producir un producto o proporcionar unos servicios a menor coste que sus competidores.
- **Diferenciación:** la organización es capaz de ofrecer un servicio o producto superior a los clientes. La diferenciación frecuentemente es el resultado de optimizar la cadena de valor del mercado en el que se opera o de realizar cambios en el alcance competitivo.

A mediados de los años ochenta emerge un modelo que considera que los recursos y las capacidades que tiene la organización son la clave para generar un rendimiento superior al de sus competidores (Wernerfelt, 1984). En este enfoque estratégico, es necesario poder determinar si los recursos y las capacidades son relevantes, lo que proporciona el marco de referencia que se ilustra en la figura 15.

Figura 15. Perspectiva organizacional basada en recursos



Fuente: B. Wernerfelt.

Referencia bibliográfica

B. WERNERFELT. A Resource-Based View of the Firm. En: *Strategic Management Journal*. 1984. Vol. 5, n.º 2, págs. 171-180.

El marco anterior busca identificar y clasificar los activos de una organización (financieros, físicos, humanos, tecnológicos, de reputación u organizacionales) en función de su valor.

El objetivo es identificar las fortalezas y debilidades en relación con sus competidores y, a través de ello, encontrar nuevas oportunidades o un uso más eficiente de estas.

De hecho, este mismo marco sirve para analizar las capacidades de una organización puesto que una capacidad puede entenderse como un recurso intangible.

Está claro que el principal objetivo de esta estrategia es generar ventajas competitivas sostenibles fundamentadas en:

- **Duración**, es decir, la velocidad a la que un recurso se deprecia o se convierte en obsoleto.
- **Trasparencia**, es decir, a menor transparencia en el recurso, menor capacidad de que sea imitado por un competidor.
- **Transferencia**, es decir, a menor capacidad de transferencia de recursos – por ejemplo, no se pueden adquirir–, menor capacidad de que sea imitado por un competidor.
- **Replicación**, es decir, a mayor complejidad de procesos y capacidades organizacionales, mayor será la dificultad de replicación.

Que una organización tenga recursos y capacidades que sean susceptibles de generar ventajas competitivas no significa que realmente las genere. En definitiva, es necesario diseñar una estrategia que identifique los recursos y las capacidades de la organización, que mida su valor, que maximice su explotación y que, además, se realice de forma continua para aumentar este tipo de recursos dentro de la organización.

Por lo tanto, una organización orientada al dato persigue convertir el dato en un recurso **valioso, raro, caro de imitar y que sea capaz de capturar el valor de forma recurrente** para crear ventajas competitivas sostenibles. Y necesita definir una estrategia clara de progresión.

3.1. Fases

Llegar a ser una organización orientada al dato no es un proceso sencillo, pues requiere una transformación a múltiples niveles: dato, organización, liderazgo, objetivos, tecnología y personas.

3.1.1. Diferentes caminos para llegar al mismo lugar

Cuando nos hacemos la pregunta de cómo transformarnos en este tipo de organización, la respuesta inicial es que existen diferentes caminos para llegar al mismo lugar. Aunque pueda parecer una respuesta clásica de un consultor de negocio, para cada organización el proceso de transformarse en una organiza-

ción orientada al dato será un proceso único. Esto es así por diversos motivos, pero el principal de ellos es que el punto de partida de la organización para cada una será diferente. Este punto de partida debe considerarse en torno a tres áreas:

- **El nivel de las competencias analíticas, soluciones y herramientas:** es decir, el conocimiento que tiene la organización respecto a técnicas, modelos y algoritmos de analítica y análisis de datos, así como el nivel de maestría en soluciones y herramientas disponibles en el mercado para desarrollar este tipo de proyectos.
- **El dominio en la gestión de la información:** el nivel de conocimiento y políticas establecidas de gestión respecto a la calidad de datos, la creación de datos maestros, la gestión de la accesibilidad, disponibilidad, seguridad, etc.
- **El alcance de la cultura de orientación al dato:** hace referencia al impacto y el nivel de coordinación de estas iniciativas (desde el nivel de las líneas de negocio hasta toda la organización).

Las organizaciones con aspiraciones tienden a invertir en competencias en analítica y herramientas. A medida que se implementan proyectos se encuentran con limitaciones que se resuelven mediante competencias en la gestión de la información. Por ello, es necesario incluir en cada proyecto aspectos del gobierno del dato (como calidad o datos maestros para progresar en esta competencia).

También es importante remarcar que muchas organizaciones inician su camino analítico mediante inversiones en departamentos o líneas de negocio específicas, frecuentemente no coordinadas. Progresivamente deben coordinarse dichas iniciativas, por ejemplo, a través del centro de competencias. La figura 16 ilustra estos diferentes caminos.

Figura 16. Caminos de evolución



Fuente: Thomas Davenport.

Otros aspectos que van a influir en este proceso son la evolución histórica de la organización, el sector en el que opera o la estrategia general que tiene la organización respecto a las TI, que influye en considerar la tecnología como una palanca para el negocio o tan solo como un departamento de soporte.

3.1.2. Cómo empezar

Aunque no existe una estrategia 100 % infalible para convertirse en una organización orientada al dato, existen varios pasos relevantes:

- **Análisis de la situación actual:** consiste en evaluar la situación actual de la organización en términos de la madurez y las capacidades analíticas.
- **Definición de la situación deseada:** definir la situación deseada (considerando los casos de uso que implementar).
- **Definición de la hoja de ruta:** identificar los pasos que desde la situación actual permiten llegar a la situación deseada.

Con independencia de cómo empezar, se recomiendan las tres siguientes iniciativas (parte de las mejores prácticas):

- Crear un **programa de alfabetización de datos** y vincularlo a un plan de **gestión del cambio** (para desarrollar una cultura de datos: conocimiento, actitud y hábitos).
- Iniciar el **programa de gobierno de datos** (para aumentar el dominio de la gestión de la información).
- Crear y evolucionar el **equipo experto en analítica de negocio y big data** (para aumentar el dominio de análisis y herramientas).

3.1.3. Cómo medir el éxito

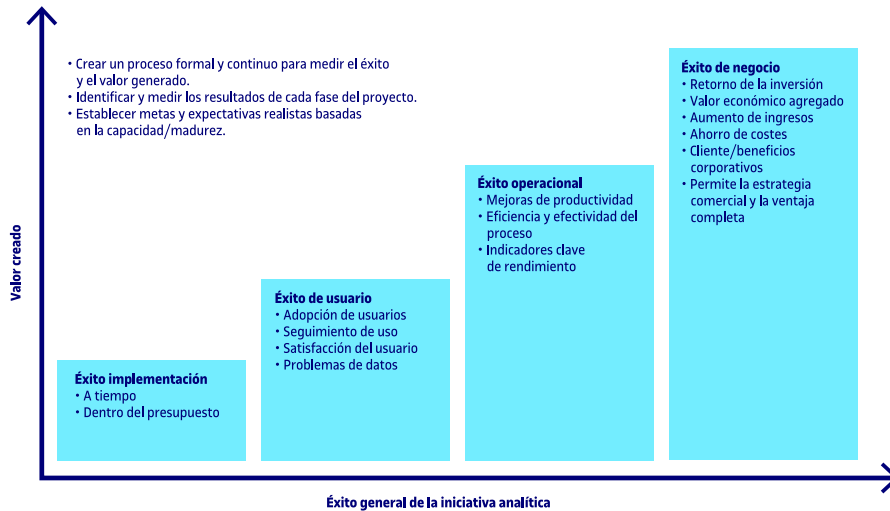
Desde el principio es importante saber el impacto y el éxito de nuestras iniciativas analíticas. Para entender qué significa tener éxito, es necesario poder medirlo. Existen cuatro dimensiones que es posible usar para medir el éxito de nuestras iniciativas:

- **Éxito en la implementación:** consiste en que la implementación del proyecto de *big data* y analítica de negocio ha sido implementada sin desviaciones graves de tiempo o en presupuesto.
- **Éxito respecto al usuario:** consiste en que los usuarios/beneficiarios del proyecto de *big data* y analítica de negocio han adoptado la solución, hacen un uso recurrente de los servicios y están satisfechos.
- **Éxito operacional en la perspectiva operacional:** consiste en que se logran cambios operacionales significativos en la organización, como mejoras en la productividad, en la eficiencia y efectividad de los procesos de negocio, así como en el uso de indicadores clave de rendimiento.
- **Éxito respecto al negocio:** consiste en que se logran cambios de profundo calado en la organización, como tener un retorno de la inversión, generar

valor económico agregado, lograr aumento de ingresos, conseguir ahorro de costes, producir beneficios corporativos o captar clientes, entre otros.

La figura 17 ilustra el valor creado respecto al éxito general de la iniciativa analítica. Al ser necesario tener en cuenta todas estas dimensiones de éxito, es normal que la percepción de fracaso en estos proyectos sea bastante alta.

Figura 17. Marco de referencia para el éxito



Fuente: Josep Curto.

3.2. Madurez

Empresas que actualmente se apoyan en los datos y los hechos en múltiples procesos de negocio, como Netflix, han seguido un largo proceso de muchos años para llegar al grado de madurez que presentan actualmente. Es, por lo tanto, necesario introducir qué es un modelo de madurez en el contexto de analítica.

Por ejemplo, Netflix usa los ránquines de películas y series más descargadas mediante P2P para la compra de derechos y ampliar el catálogo disponible. También usa las preferencias para su sistema de recomendación o para la creación de nuevos productos.

3.2.1. Modelos de madurez

Como hemos comentado, uno de los puntos de partida es entender la situación actual de la organización. Una de las herramientas de análisis más útiles en este sentido son los **modelos de madurez**, que permiten identificar áreas de mejora y proporcionan una perspectiva temporal de estas.

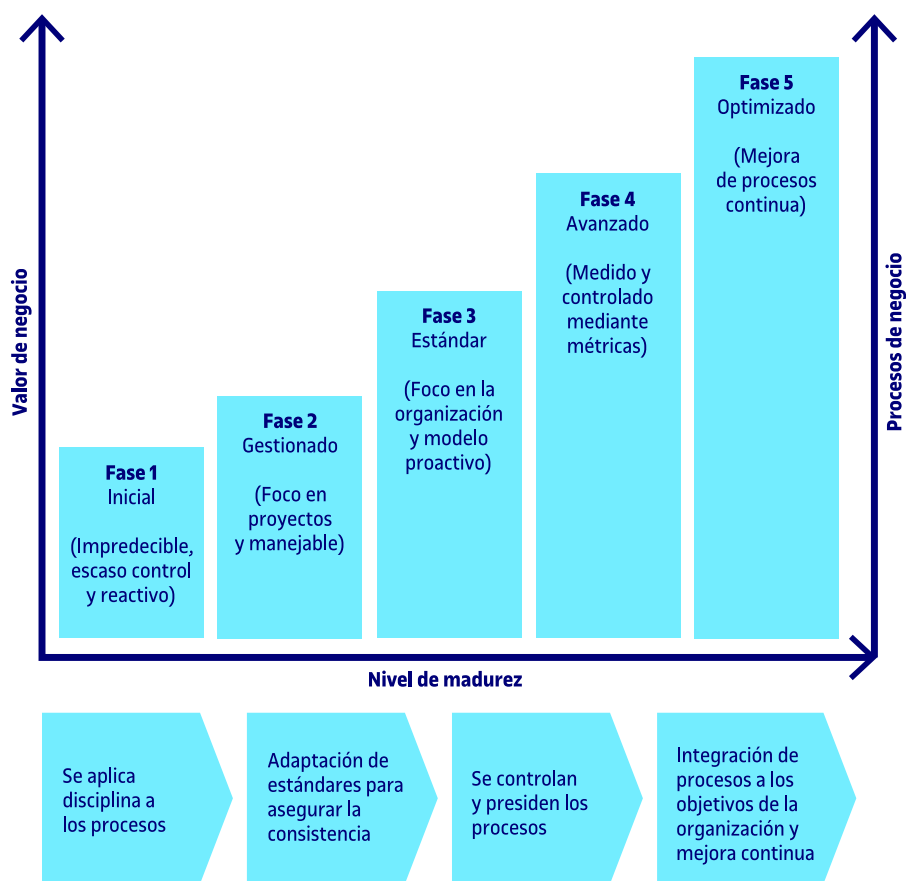
En general, los modelos de madurez consisten en definir una escala de estados posibles de un proceso y un conjunto de métricas que permiten establecer en qué estado se encuentra un proceso en un momento determinado.

Los cinco estados de la escala de valoración son:

- 1) **Inicial:** el proceso no se controla y sus resultados son impredecibles.
- 2) **Repetitivo:** hay experiencias disponibles y se usan para aprender e imitar.
- 3) **Definido:** el proceso está completamente definido y es comprendido por quienes deben ejecutarlo.
- 4) **Gestionado:** el proceso está bajo control y se toman medidas de sus características más significativas.
- 5) **Óptimo:** se trabaja en la mejora continua del proceso.

La figura 18 representa el impacto de las diferentes fases del modelo de madurez respecto al valor de negocio.

Figura 18. Impacto del modelo de madurez



Fuente: Josep Curto.

Esta estructura ha sido adaptada por Davenport, Harris y Morison (2010), considerando una matriz en la que se evalúan seis factores de éxito respecto a cinco fases de madurez, que son:

- **Fase 1: Incapacidad analítica.** La organización no tiene ningún tipo de interés en el uso de analítica ni dispone de capacidades analíticas.
- **Fase 2: Aplicación de analítica de forma localizada.** La organización ha empezado a desplegar proyectos de manera independiente y descoordinada que responden a necesidades muy específicas.
- **Fase 3: Aspiración por conversión en una compañía analítica.** La organización empieza a dar los primeros pasos para coordinar tareas de forma conjunta.
- **Fase 4: Compañía analítica.** La organización coordina las iniciativas analíticas en la organización.
- **Fase 5: Competición basada en analítica.** La organización ha embebido la cultura analítica en la organización y tiene articulada una estrategia completamente coordinada.

Como es posible imaginar, una misma organización puede encontrarse en diferentes niveles de madurez respecto a cada uno de los factores de éxito. Esto permite identificar áreas de mejora en la organización y diseñar una hoja de ruta (o *roadmap*) específica para cada organización.

Para cada fase y cada factor de éxito debemos describir los aspectos más relevantes.

Tabla 7. Modelo de madurez analítica

Factor de éxito	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Datos	Inconsistentes, baja calidad, organización de datos pobre	Datos usables en silos funcionales o por procesos	La organización empieza a crear un repositorio central de datos	Datos integrados y precisos, que son comunes en un repositorio central	Incesante búsqueda de nuevos datos y métricas
Organización	No existe / No aplica	Silos de datos, tecnología y conocimiento	Fases iniciales de un enfoque para toda la empresa	Datos, tecnología y análisis clave están centralizados o en red	Todos los recursos son gestionados de forma central o en red
Liderazgo	No hay conciencia, no interesa	Solo a nivel de función o proceso	Los líderes empiezan a reconocer la importancia de los análisis	Apoyo por parte de la dirección a la competencia analítica	Fuerte liderazgo para la competencia analítica
Objetivos	No existe / No aplica	Múltiples objetivos inconexos que pueden no ser de importancia estratégica	Esfuerzos analíticos coalescentes detrás de un pequeño conjunto de objetivos	Actividad analítica centrada en torno a ciertas áreas clave	La estrategia y las ventajas competitivas se fundamentan en la analítica

Fuente: Davenport, Harris y Morison.

Referencia bibliográfica

Thomas H. DAVENPORT, Jaenne G. HARRIS y Robert MORISON. *Analytics at Work: Smarter Decisions, Better Results*. En: *Harvard Business Review Press*, 2010.

DELTA

Este modelo se denomina DELTTA (*data, enterprise, leadership, targets, technology y analysts*), si bien ha sido adaptado al castellano para este material y, por lo tanto, su nombre no encaja con el del modelo original.

Factor de éxito	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Tecnología	Uso mayoritario de Excel como herramienta de análisis	Despliegue de soluciones independientes a nivel de proceso o función	Inicio del despliegue de iniciativas a nivel corporativo	Uso coordinado de tecnologías comunes en ciertas áreas clave	Búsqueda continua de nuevas tecnologías que desplegar a nivel de organización
Personas	Capacidades analíticas limitadas y ligadas a una función específica	Grupos aislados de analistas sin comunicación	Afluencia de analistas en las áreas clave	Analistas altamente capacitados organizados centralmente o en red	Equipo de profesionales de alto calibre y en ampliación continua

Fuente: Davenport, Harris y Morison.

Cabe comentar que existen otros modelos de madurez, como el propuesto por LaValle, Lesser, Shockley, Hopkins y Kruschwitz (2011). A diferencia del anterior, este modelo solo considera tres fases: aspiracional, experiencia y transformado respecto a cinco aspectos: conocimiento funcional, retos de negocio, barreras organizacionales, gestión del dato y despliegue de analítica. Por lo que el modelo de madurez debe considerarse como una referencia que puede ajustarse a la realidad de la propia organización.

Para completar el modelo de madurez y tener una visión completa de la situación actual, es necesario recopilar datos hasta conseguir información de calidad suficiente (en un proceso iterativo). Los métodos de obtención de información (Navarro, 2006) más frecuentes son:

- entrevistas,
- reuniones y grupos focales,
- cuestionarios,
- análisis de documentos,
- herramientas de generación de informes, y
- observación.

3.2.2. Competencias analíticas

Uno de los aspectos relevantes para convertirse en una organización orientada al dato son las competencias necesarias.

Inicialmente podemos considerar que las competencias relevantes solo están vinculadas al análisis:

- **Análisis descriptivo:** la organización es capaz de entender qué pasó. Por ejemplo, en las interacciones de cliente: qué han comprado o cuántos clientes han dejado de usar los servicios de la compañía.

Referencia bibliográfica

Steve LAVALLE, Eric LESSER, Rebecca SHOCKLEY, Michael S. HOPKINS y Nina KRUSCHWITZ. Big Data, Analytics and the Path From Insights to Value. En: *MIT SLOAN Management Review*, 2011.

Lectura recomendada

Para conocer más sobre estos métodos generales de consultoría, recomendamos la lectura de:

Xavier NAVARRO ARNAL. *Consultoría en los Sistemas de Información*. Universitat Oberta de Catalunya, 2006.

- **Análisis de diagnóstico:** la organización es capaz de entender las razones por las que sucede un evento. Por ejemplo, las interacciones con los clientes: por qué los clientes compran un determinado producto o han dejado de pagar por nuestros servicios.
- **Análisis predictivo:** la organización es capaz de predecir ciertas interacciones futuras. Por ejemplo, qué clientes tienen la intención de abandonar los servicios de la organización.
- **Análisis prescriptivo:** la organización es capaz de tomar decisiones vinculadas con las interacciones futuras basadas en escenarios. Por ejemplo, identificar a los clientes a los que aplicar estrategias de retención.
- **Análisis preventivo:** la organización es capaz de actuar con antelación a las necesidades futuras. Por ejemplo, enviando ofertas a los clientes antes de que se definan sus necesidades.

Cabe comentar que este enfoque no es suficiente, puesto que deja de lado múltiples aspectos relevantes. Es necesario considerar una encuesta de evaluación de capacidades analíticas mucho más amplia. De nuevo existen múltiples versiones. En estos materiales consideraremos la **encuesta creada por Hindle y Vidgen (2018)**, que tiene en consideración las siguientes dimensiones:

- datos: calidad, privacidad y seguridad, uso;
- organización: estrategia, cultura y estructura;
- proceso;
- personas: equipo experto (en *big data* y analítica), recursos humanos;
- tecnología; y
- valor.

Los elementos de la encuesta se evalúan mediante una escala del 1 (completamente en desacuerdo) al 7 (completamente de acuerdo) y, para cada una de las dimensiones, se calcula el valor promedio de los elementos. La siguiente tabla recoge los elementos analizados:

Tabla 8. Evaluación de competencias

Datos - calidad	
DQ1	Nuestros datos son accesibles.
DQ2	Nuestros datos están en el nivel correcto (de agregación).
DQ3	Nuestros datos son precisos.
DQ4	Nuestros datos son consistentes.
DQ5	Nuestros datos son oportunos.
DQ6	Nuestros datos son creíbles.
Datos - privacidad y seguridad	

Fuente: Hindle y Vidgen.

Referencia bibliográfica

Giles A. HINDLE y Richard VIDGEN. Developing a business analytics methodology: A case study in the foodbank sector. En: *European Journal of Operational Research*. 2018. Vol. 268.3, págs. 836–851.

DE1	Establecemos claramente qué datos se recopilan y almacenan en relación con los interesados (por ejemplo, clientes, empleados).
DE2	Establecemos claramente cómo se utilizan los datos recopilados y almacenados.
DE3	Establecemos claramente cómo los datos que se recopilan y almacenan podrían compartirse con terceros.
DE4	Solo usamos los datos para el propósito para el que se han recopilado (es decir, no reutilizamos los datos recopilados con un propósito para otro propósito).
DE5	Tenemos un conocimiento profundo del entorno legal y regulatorio de nuestro uso de datos.
DE6	Nuestros datos se almacenan de manera segura (por ejemplo, encriptados).
Datos - uso	
DU1	Podemos compartir datos de manera efectiva con socios.
DU2	Hacemos uso efectivo de datos externos (por ejemplo, calificaciones crediticias).
DU3	Hacemos un uso efectivo de los datos abiertos (p. ej., datos meteorológicos y censales).
Organización - estrategia	
OS1	Tenemos una estrategia analítica claramente articulada.
OS2	Nuestra estrategia de análisis deja en claro cómo y dónde se crea el valor.
OS3	Nuestra estrategia analítica se comunica eficazmente en toda la organización.
OS4	Nuestra estrategia de análisis está alineada con nuestra estrategia de negocios.
Organización - cultura	
OC1	Las personas de nuestra organización entienden el valor de la toma de decisiones basada en datos y evidencias.
OC2	Las decisiones clave se toman utilizando evidencia en lugar de «intuición» en nuestra organización.
OC3	Nuestra alta gerencia apoya totalmente el uso de análisis de negocios.
Organización - estructura	
OO1	Big data y analítica de negocio tienen un lugar claro y apropiado en la estructura de nuestra organización.
OO2	El equipo de análisis tiene un profundo conocimiento de nuestro negocio.
OO3	La función analítica tiene un fuerte liderazgo.
OO4	Estamos dispuestos a compartir el valor creado por el análisis con socios, clientes y otros.

OO5	Nos asociamos de manera efectiva con organismos externos (por ejemplo, universidades, asociaciones profesionales) para complementar nuestra experiencia en análisis.
Proceso	
PR1	Tenemos un proceso definido para garantizar que los algoritmos se utilicen éticamente.
PR2	Podemos identificar consecuencias no deseadas que pueden surgir del uso de algoritmos en nuestra toma de decisiones.
PR3	Tenemos un proceso definido para garantizar que nuestro uso de análisis no dañe nuestra reputación o marca.
PR4	Utilizamos métodos ágiles de manera efectiva para crear aplicaciones de análisis.
PR5	Podemos equilibrar el uso de análisis para la explotación (mejorando la forma en que se hacen las cosas) y la exploración (haciendo cosas nuevas).
PR6	Podemos medir el valor creado a partir de proyectos de análisis empresarial.
PR7	Gestionamos nuestros proyectos analíticos como una cartera de inversiones.
PR8	Contamos con un proceso efectivo para la gestión del riesgo del proyecto analítico.
Personas - equipo experto (en <i>big data</i> y analítica)	
HA1	Nuestro equipo experto es curioso.
HA2	Nuestro equipo experto está orientado a los problemas.
HA3	Nuestro equipo experto utiliza las herramientas apropiadas para resolver problemas comerciales.
HA4	Nuestro equipo experto a menudo propone soluciones a problemas comerciales existentes.
HA5	Nuestro equipo experto a menudo propone nuevos usos de análisis para abordar nuevas oportunidades de negocio.
HA6	Nuestro equipo experto tiene excelentes habilidades técnicas de TI (por ejemplo, R, SAS, Tableau).
HA7	Nuestro equipo experto tiene excelentes habilidades estadísticas y de modelado.
Personas - recursos humanos	
HM1	No tenemos dificultades para contratar personal cualificado de análisis de datos.
HM2	Nuestro personal de análisis tiene excelentes oportunidades de progreso profesional.
HM3	No tenemos dificultades para retener al personal de análisis de datos.
Tecnología	

TE1	Tenemos plataformas de datos empresariales adecuadas (por ejemplo, almacenamiento de datos, almacenamiento en la nube).
TE2	Tenemos un conjunto de herramientas adecuado para el personal de análisis de datos (por ejemplo, R, SAS).
TE3	Tenemos excelentes herramientas para la visualización de datos por parte del usuario final.
Valor	
VA1	Mediante nuestro uso de análisis podemos tomar mejores decisiones.
VA2	Como resultado del uso de análisis, hemos construido relaciones más sólidas con los clientes.
VA3	<i>Big data</i> y analítica de negocio nos han permitido mejorar la experiencia del cliente.
VA4	Nuestro uso de análisis nos ha permitido innovar nuestros productos y servicios.
VA5	Mediante nuestro uso de análisis, podemos obtener mejores resultados que nuestros competidores.
VA6	Utilizamos análisis para crear valor para el bien público y para fines comerciales.
VA7	Mediante nuestro uso de análisis podemos ahorrar costes.
VA8	Mediante el uso de análisis podemos mejorar nuestra rentabilidad.
VA9	Mediante nuestro uso de análisis podemos mejorar nuestra productividad.

Fuente: Hindle y Vidgen.

La combinación del modelo de madurez, la evaluación de competencias e información específica de nuestra organización (como, por ejemplo, casos de uso) permite construir una hoja de ruta que incluya acciones concretas, reales y medibles.

3.2.3. Pensamiento analítico

Como en el caso de la aplicación de políticas, recomendaciones y soluciones para la seguridad en el contexto de una organización, el eslabón más débil de la cadena del desarrollo de una cultura analítica son las personas. Las personas en el ámbito de la cultura analítica se pueden dividir en dos tipos:

- **Analistas (frecuentemente denominados *quants*):** expertos y conocedores de métodos, técnicas, herramientas y soluciones de análisis de datos. Esto incluye expertos en cada una de las componentes de la plataforma del dato.

- **Usuarios de negocio (frecuentemente llamados *non-quants*):** profesionales que hacen uso de las soluciones analíticas implementadas en la organización.

De manera natural, los primeros han desarrollado una sensibilidad natural al pensamiento analítico, pero no tiene por qué estar extendida en los segundos. Primero es necesario definir el pensamiento analítico:

Entendemos el **pensamiento analítico** como la capacidad de definir un problema y plantear las preguntas adecuadas para comprender y resolver dicho problema.

Si una empresa quiere transformar su cultura analítica, debe no solo ser capaz de tener *quants* en su organización, sino que además los *non-quants* deben ser capaces de desarrollar y usar pensamiento analítico. En ningún caso esto significa que todo el mundo deba convertirse en un analista ni en un experto, pero sí requiere un cambio en actitud, hábitos y conocimientos.

- **Actitud:** se debe desarrollar una actitud abierta a trabajar con datos, información y conocimiento, así como a insistir en la calidad de estándares en las evidencias. En definitiva, se persigue un cambio de mentalidad respecto al uso de evidencias y datos en la toma de decisiones, fundamentada en:
 - Perder el miedo al uso de datos y evidencias.
 - Tener una actitud proactiva respecto a lo que se desconoce para familiarizarse con métodos y técnicas.
 - Mantener una actitud de curiosidad respecto a métodos, técnicas y resultados.
- **Hábitos:** tener una actitud analítica es el punto de partida para crear el pensamiento crítico, pero no es suficiente. Es necesario crear hábitos de comportamiento para consolidar buenas prácticas y transformar el pensamiento crítico en una práctica habitual. Esto supone:
 - Pedir los datos que sustentan ideas, intuiciones, hechos, teorías y observaciones.
 - No creer los datos de forma unilateral. El dato debe ser relevante, preciso e interpretado correctamente para apoyar una teoría.
 - Discutir y argumentar las relaciones de causalidad. Aunque ciertos hechos estén correlacionados, eso no significa automáticamente que haya una relación de causalidad.
 - Hacer preguntas para entender el problema y el proceso de resolución.
 - Usar los resultados analíticos para la toma de decisiones.
- **Conocimientos:** aunque el usuario de negocio no debe conocer en profundidad los modelos, las técnicas y los algoritmos, debe tener claro para qué sirven, qué proporcionan y cuándo son aplicables. Este conocimien-

to, en definitiva, ayudará *a posteriori* a entender qué preguntas se pueden realizar y si estas pueden responderse o no. Para poder adquirir este conocimiento, el usuario de negocio debe leer libros vinculados con la práctica de la analítica. Esto significa leer libros o recibir formación genérica para usuarios de, al menos, estadística aplicada a la toma de decisiones, inteligencia de negocio, analítica de negocio, *big data* y ciencia de datos.

Como es posible imaginar, actitud, hábitos y conocimientos se retroalimentan. Una actitud analítica reforzada por hábitos permitirá su consolidación. Al mismo tiempo, los hábitos generarán de modo natural un conocimiento y pensamiento crítico sobre los procesos analíticos, lo que ampliará los conocimientos de la organización. A medida que se consoliden los conocimientos, la actitud analítica se consolidará al tener un mayor grado de confianza en el pensamiento analítico.

3.2.4. Roles y perfiles

Desde hace años, los perfiles que ayudan a la explotación del dato han ido adquiriendo relevancia e importancia en las organizaciones. En los últimos años, con la emergencia del *big data* y la analítica de negocio como dos de las principales tendencias que están transformando las organizaciones, nuevos perfiles están siendo demandados con las acreditaciones y los conocimientos adecuados a las necesidades de captura, procesamiento, almacenamiento, análisis y visualización de la información.

Múltiples estudios como los de McKinsey (2011) o e-skills UK (2013) apuntan que existe una gran demanda de perfiles analíticos de diferente calado en todo el mundo.

Por ejemplo, e-skills apuntaba que se necesitaban 69.000 perfiles en Reino Unido para 2017 y McKinsey que se necesitaban de 140.000 a 190.000 perfiles con un profundo conocimiento analítico y más de 1,5 millones de profesionales capaces de tomar decisiones fundamentadas en hechos y datos.

Aunque es evidente que hay una fuerte demanda de perfiles analíticos, no está claro qué tipo de capacidades son las que se requieren en la nueva era de análisis de datos y estamos observando una explosión de perfiles que se unen a los que ya existían tradicionalmente.

Podemos distinguir estos tipos de perfiles:

- **Perfiles de negocio**, que son los *stakeholders* del proyecto.
- **Perfiles tecnológicos**, que tienen los conocimientos adecuados para el despliegue de proyectos de explotación de negocio. Podemos distinguir dos tipos:
 - **Perfiles tradicionales**, que tienen conocimientos en inteligencia de negocio y analítica de negocio. Este tipo de perfiles suele ya existir en la organización.

Referencias bibliográficas

Big Data: the next frontier for innovation, competition and productivity. McKinsey, 2011. Disponible en: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Big Data Analytics: Adoption and Employment Trends, 2012-2017. e-skills UK. 2013. Esponsorizado por SAS. Disponible en: <http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf>

- **Nuevos perfiles**, cuya área de competencia está vinculada al resto de los componentes tecnológicos de la plataforma del dato. En general, este tipo de perfiles no existe aún en la organización.

Dentro de cada categoría tenemos distintos tipos de perfiles, tal y como presenta la tabla 9.

Tabla 9. Perfiles tradicionales frente a nuevos perfiles

Negocio	Perfiles tradicionales	Nuevos perfiles
Usuarios de negocio Patrocinador del proyecto Gerente de proyecto Analista de negocio	Director BI/DW Gerente de proyectos Analista de calidad de datos Arquitecto de datos / DBA Gerente de metadatos Arquitecto/desarrollador ETL Analista BI Arquitecto BI / desarrollador de aplicaciones o portales / arquitecto TI Gerente de seguridad Gerente del equipo de pruebas Especialista minería de datos / estadística Formador	Científico del dato Especialistas en visualización Ingenieros del dato Arquitectos <i>cloud computing</i> Chief data officer

Fuente: Josep Curto.

Vamos a entrar en detalle en cada uno de estos perfiles.

Perfiles de negocio

Como ya se ha comentado, los perfiles de negocio son los patrocinadores y clientes del proyecto. Podemos distinguir diversos perfiles:

- **Patrocinador del proyecto:** este perfil es el que apoya el desarrollo del proyecto. Debe pertenecer a la dirección para poder influenciar a todos los participantes con el fin de que el proyecto llegue a buen puerto. Además, debe tener capacidad de liberar recursos (en tiempo y dinero) para el proyecto.
- **Gerente del proyecto:** es el responsable de la ejecución del proyecto por parte de negocio. Entre sus tareas está la de monitorizar que la evolución del proyecto sigue los plazos establecidos y cumple los requisitos definidos.
- **Analista de negocio:** participa en el proyecto proporcionando los requisitos y las necesidades de negocio al ser experto en el negocio.
- **Usuarios de negocio:** son aquellos que van a hacer uso de las herramientas o soluciones de explotación de datos. Este tipo de usuarios deben ser formados en la solución desplegada. Además, frecuentemente algunos de ellos participa en el proyecto como analista de negocio.

En función de la magnitud y el alcance del proyecto, las responsabilidades de las figuras anteriores recaerán en una única persona o en varias.

Perfiles tradicionales

Desde el área de TI o, en el caso de existir, el área de inteligencia de negocios y/o analítica de negocio de la organización existen también diversos perfiles que pueden participar en este tipo de proyecto.

- **Director BI/DW:** es el responsable último dentro de la organización de todas las iniciativas de explotación de datos. Si no hay un área específica de BI o DW, frecuentemente coincide con el director de sistemas.
- **Gerente de proyectos:** es el responsable de un proyecto específico desde la perspectiva de implementación. Es el interlocutor con el gerente del proyecto por parte de negocio.
- **Analista de calidad de datos:** es el responsable de la calidad de datos del proyecto.
- **Arquitecto de datos / DBA:** es el responsable de la estructura de datos para el proyecto.
- **Gerente de metadatos:** es el responsable de los metadatos del proyecto. En el caso de BI, los metadatos de la factoría de información corporativa.
- **Arquitecto / desarrollador ETL:** es el responsable de crear los procesos ETL necesarios para el proyecto.
- **Analista BI:** es el responsable de plasmar las necesidades de negocio en una determinada solución de BI. Entre sus tareas está hablar con el analista de negocio y crear informes, vistas OLAP o cuadros de mando.
- **Arquitecto BI / desarrollador de aplicaciones o portales / arquitecto TI:** es el responsable de dimensionar la plataforma de BI.
- **Gerente de seguridad:** es el responsable de la seguridad de la plataforma y su integración dentro de la arquitectura de seguridad de la organización.
- **Gerente del equipo de pruebas:** es el responsable de las pruebas de rendimiento y usabilidad de la solución.
- **Especialista minería de datos / estadística:** es el responsable de crear los modelos necesarios para el análisis de datos en el caso de que el proyecto incluya fases que los requieran.
- **Formador:** es el responsable de la formación de los usuarios de negocio.

De nuevo, como en el caso anterior, las responsabilidades de estos perfiles pueden estar agregadas en una única persona.

Nuevos perfiles

Las nuevas tendencias con base tecnológica están haciendo emerger nuevos perfiles en el contexto de la explotación del dato. Aunque no está claro si van a aparecer más perfiles, entre ellos destacamos los siguientes:

- **Científico del dato:** este perfil, también conocido como *data scientist*, procesa datos en bruto e identifica qué tipo de análisis se debe aplicar para con el dato. Combina diferentes tipos de conocimientos, como las matemáticas, la estadística, la inteligencia artificial o el procesamiento de lenguaje

natural. De hecho, dentro de su abanico de conocimientos suele poseer también conocimiento de negocio y la habilidad de usar diversos lenguajes/herramientas para manipular datos.

- **Especialistas en visualización:** este perfil es especialista en la contextualización del dato y su representación a través de técnicas de visualización para que la comunicación a los usuarios de negocio sea óptima.
- **Ingenieros del dato:** este perfil, también conocido como el arquitecto del dato, tiene conocimiento en las tecnologías de *big data* y cuál de ellas es la que se ajusta a las necesidades del proyecto. Además, debe diseñar la evolución del sistema para que este pueda continuar dando el servicio para el que ha sido diseñado.
- **Arquitectos *cloud computing*:** es el que permite diseñar y desplegar arquitecturas de explotación a través de *cloud computing*. Entre sus responsabilidades está la de crear el acceso del dato como servicio.
- **Chief data officer:** es el responsable de todas las iniciativas de explotación de datos vinculadas con la ciencia de los datos y *big data*. Debe tener responsabilidad directiva y jugar el papel de agente del cambio dentro de la organización. Se trata de un perfil con conocimientos de gestión y técnicos.

Como en los casos anteriores, en función del proyecto y de la madurez de la organización estos perfiles pueden estar concentrados en una o varias personas.

Perfiles como el de científico del dato, que aúnan conocimientos de diferentes disciplinas, son complicados de conseguir en estos momentos. Las empresas se están decantando en constituir equipos multidisciplinares que de manera coordinada cubren todas las necesidades requeridas.

3.3. Aspectos éticos

La **ética de datos** (o *data ethics*) es una rama de la ética que evalúa las prácticas de datos con el potencial de impactar negativamente en las personas y la sociedad: en la recopilación, el intercambio y el uso de datos.

En el contexto de la ética de datos, es necesario tener en cuenta la regulación actual, los diferentes sesgos existentes, cómo mitigarlos y cómo identificarlos.

En este apartado hablaremos de estos problemas, cómo entenderlos e introduciremos un marco de referencia que podemos usar.

3.3.1. Derivas, sesgos y riesgos

¿Qué es la deriva conceptual?

La deriva conceptual en el aprendizaje automático y la minería de datos se refiere al cambio en las relaciones entre los datos de entrada y salida en el problema subyacente a lo largo del tiempo.

El primer desafío es detectar cuándo ocurre la deriva del concepto. El segundo desafío es mantener los patrones actualizados sin inducirlos desde cero.

Un problema difícil con el aprendizaje en muchos dominios del mundo real es que el concepto de interés puede depender de algún contexto oculto, no dado explícitamente en forma de características predictivas.

Un cambio en la distribución de datos se denomina deriva de concepto virtual. Si bien la deriva del concepto virtual puede ocurrir solo, la mayoría de las veces es un enlace a la deriva del concepto.

Existen dos tipos de desviaciones conceptuales:

- 1) Repentino (abrupto, instantáneo).
- 2) Gradual (moderado o lento).

En el contexto de una organización orientada al dato, podemos considerar el siguiente marco para pensar sobre la deriva conceptual:

- **Suposición futura:** el experto en analítica debe hacer una suposición sobre la fuente de datos futura.
- **Tipo de cambio:** el experto en analítica necesita identificar posibles patrones de cambio.
- **Adaptabilidad:** según el tipo de cambio y la suposición futura, un experto en analítica elige los mecanismos que hacen que el algoritmo sea adaptativo.
- **Selección de modelo:** el experto en analítica necesita un criterio para elegir una parametrización particular del algoritmo seleccionado en cada paso de tiempo.

Aparte de tener un marco de trabajo para la deriva, necesitamos también un plan de acción. A continuación, revisamos las opciones disponibles:

- **No hacer nada (modelo estático):** la forma más común es no manejarlo en absoluto y asumir que los datos no cambian.
- **Introducir capacidades de monitoreo:**
 - **Habilidad.** Se refiere a la distribución de una característica particular, covarianza entre variables, etc.

- **Concepto de detección de deriva.** Trazar la habilidad del modelo estático a lo largo del tiempo y si la habilidad cae o cambia, tal vez se esté produciendo una deriva del concepto y se requiera alguna intervención.
 - **Rendimiento de referencia.** Usar la habilidad del modelo estático como línea de base para comparar con cualquier intervención que realice.
-
- **Reinstalar periódicamente.** Una buena intervención de primer nivel es actualizar periódicamente su modelo estático con datos históricos más recientes.
 - **Actualización periódica.** Esta es una eficiencia sobre el enfoque anterior (ajuste periódico), en la que, en lugar de descartar completamente el modelo estático, el estado existente se utiliza como punto de partida para un proceso de ajuste que actualiza el ajuste del modelo utilizando una muestra de los datos históricos más recientes.
 - **Datos de peso.** En este caso, puede usar una ponderación que sea inversamente proporcional a la antigüedad de los datos, de modo que se preste más atención a los datos más recientes (mayor peso) y se preste menos atención a los menos recientes (menor peso).
 - **Aprender el cambio.** Se puede usar un enfoque de conjunto cuando el modelo estático se deja intacto, pero un nuevo modelo aprende a corregir las predicciones del modelo estático en función de las relaciones en los datos más recientes.
 - **Detectar y elegir modelo.** Para algunos dominios problemáticos, es posible diseñar sistemas para detectar cambios y elegir un modelo específico y diferente para hacer predicciones.
 - **Preparación de datos.** En series de tiempo (y problemas similares), es común preparar los datos de tal manera que se eliminen los cambios sistemáticos en los datos a lo largo del tiempo, como las tendencias y la estacionalidad por diferenciación.

Sesgos y riesgos

La deriva conceptual no es el único aspecto que se debe tener en cuenta en este tipo de proyectos. Según Ntousti, Fafalios y Gadiraju (2020), debemos desarrollar un marco de actuación que nos permita:

- **Entender el riesgo**, que puede tener causas sociotécnicas, aparecer en los datos o crearse durante la aplicación de la técnica de análisis.
- **Atenuar el riesgo**, que puede realizarse durante el preprocesamiento de datos, durante la aplicación de la técnica de análisis o *a posteriori*.
- **Comprender el sesgo** a través del análisis de los datos, describiendo los sesgos –mediante un marco de referencia–, y usando mecanismos de interpretabilidad y explicabilidad.

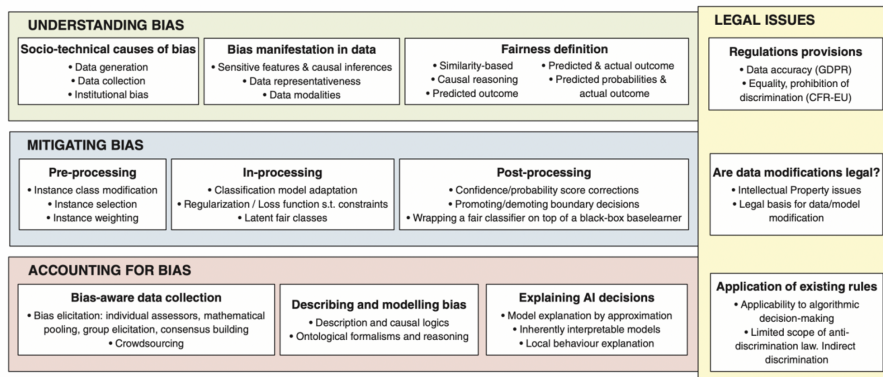
Referencia bibliográfica

E. NTOUTSI, P. FAFALIOS, U. GADIRAJU *et al.* Bias in data-driven artificial intelligence systems. An introductory survey. En: *WIREs Data Mining Knowl Discovery*, 2020. e1356. <https://doi.org/10.1002/widm.1356>

La aplicación de los algoritmos debe al mismo tiempo considerar el cumplimiento de las regulaciones vigentes y conocer sus límites.

La figura 19 recopila los puntos que se deben considerar para conocer los sesgos y sus riesgos:

Figura 19. Marco de referencia del sesgo



Fuente: Ntousti, Fafalios y Gadiraju.

3.3.2. Explicabilidad, interpretabilidad y equidad

Creemos que las máquinas son árbitros neutrales: entidades frías y calculadoras que siempre toman la decisión correcta, que pueden ver patrones que nuestras mentes humanas no pueden o no quieren ver. Como hemos discutido, esta afirmación no es cierta y, en ciertos escenarios, la toma de decisiones mediante algoritmos se transforma en una forma de amplificar, extender y hacer inescrutables los prejuicios y la discriminación que prevalecen en la sociedad.

Es necesario, por lo tanto, conocer no solo las respuestas que podemos responder a través de los algoritmos, sino también comprender por qué se toman ciertas decisiones.

Para ello, es importante considerar dentro de los proyectos analíticos la explicabilidad y la interpretabilidad (Molnar, 2019), como ya hemos introducido en el apartado anterior.

Explicabilidad

La **explicabilidad** es el grado en el que los valores característicos de una instancia están relacionados con la predicción de su modelo de tal manera que los humanos lo entiendan.

En términos básicos, es la comprensión de la pregunta «¿por qué está sucediendo esto?».

Lectura recomendada

Para profundizar más sobre el tema, recomendamos la lectura de: C. MOLNAR. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2019. Disponible en: <https://christophm.github.io/interpretable-ml-book>

Existen diferentes métodos de explicabilidad, que deben tener las siguientes propiedades:

- **Portabilidad:** define la gama de modelos de aprendizaje automático donde se puede utilizar el método de explicación.
- **Poder expresivo:** se define como la estructura de una explicación que un método puede generar.
- **Translucidez:** significa cuánto depende el método de explicación del modelo de aprendizaje automático. Los métodos de baja translucidez tienden a tener una mayor portabilidad.
- **Complejidad algorítmica:** define la complejidad computacional de un método donde se generan las explicaciones.
- **Fidelidad:** la alta fidelidad se considera una de las propiedades importantes de una explicación, ya que la baja fidelidad carece de la explicación del modelo de aprendizaje automático.

Interpretabilidad

La **interpretabilidad** se define como la cantidad de predicciones consistentes del resultado de un modelo sin tratar de conocer las razones detrás de la escena.

Es más fácil saber la razón detrás de ciertas decisiones o predicciones si la interpretabilidad de un modelo de aprendizaje automático es mayor.

La interpretabilidad puede evaluarse de diferentes maneras:

- **Evaluación del nivel de aplicación:** esta es básicamente la tarea real. Significa poner la explicación en el producto y el usuario final hará todas las pruebas.
- **Evaluación a nivel humano:** esta es una tarea simple que puede considerarse como una evaluación de nivel de aplicación simplificada. En este caso, los experimentos se llevan a cabo por laicos, haciendo que los experimentos sean más baratos y los evaluadores se puedan encontrar fácilmente.
- **Evaluación de nivel de función:** este es un enfoque en el que una persona anónima ya evalúa la clase de modelo. Este enfoque también se conoce como una tarea proxy.

¿Qué es la equidad algorítmica?

Escándalos como los de Cambridge Analytica (<https://www.theguardian.com/news/series/cambridge-analytica-files>) han mostrado que **se pueden usar los datos para afectar a la sociedad** y han abierto la puerta a hablar de justicia y equidad en el uso de *big data* y analítica de negocio. Como respuesta ha

surgido un creciente interés en la equidad algorítmica: cómo podemos razonar sistemáticamente sobre los sesgos en los algoritmos y cómo podemos hacer que nuestros algoritmos sean justos.

Según Lipton, Chouldechova y McAuley (2017), la **equidad algorítmica** presenta múltiples facetas:

- **Paridad de tratamiento:** un clasificador debe ser ciego a una característica protegida dada. Es decir, se trata de justicia a través de la falta de conciencia.
- **Paridad de impacto:** la fracción de personas que reciben una decisión positiva debe ser igual en los diferentes grupos. Esto también se conoce como paridad demográfica, paridad estadística o independencia de la clase protegida y el puntaje.
- **Paridad de clasificación:** alguna medida dada de error de clasificación es igual en todos los grupos definidos por los atributos protegidos. También denominada igualdad de oportunidades.
- **Calibración:** los resultados son independientes de los atributos protegidos condicionados a la puntuación de riesgo. Es decir, la realidad se ajusta al puntaje de riesgo.
- **Paridad estadística (o paridad demográfica):** un modelo generalmente se considera justo si las diferencias en la distribución de las predicciones del modelo condicionadas a la variable protegida no exceden un umbral predeterminado, medido por alguna noción apropiada de distancia entre las distribuciones de probabilidad.

Por medio de la equidad podemos responder a las siguientes preguntas:

- **¿Se alcanza la paridad demográfica?** La paridad demográfica se logra si el número absoluto de predicciones positivas en los subgrupos está cerca uno del otro.
- **¿Se alcanza la paridad proporcional?** La paridad proporcional se logra si la proporción de predicciones positivas en los subgrupos son cercanas entre sí.
- **¿Se logran las cuotas igualadas?** Las probabilidades igualadas se alcanzan si las sensibilidades (verdaderos positivos divididos por todos los positivos) en los subgrupos están cerca una de la otra.
- **¿Se logra la paridad de tasa predictiva?** La paridad de tasa predictiva se logra si las precisiones o valores predictivos positivos (verdaderos positivos divididos por todos los positivos predichos) en los subgrupos están cerca uno del otro.
- **¿Se alcanza la paridad de precisión?** La paridad de precisión se logra si las precisiones (todas las clasificadas con precisión divididas por todas las predicciones) en los subgrupos están cercanas entre sí.
- **¿Se alcanza la paridad de tasa de falsos negativos?** La paridad de tasa de falsos negativos se logra si las tasas de falsos negativos (división de falsos negativos con todos los positivos) en los subgrupos están cercanas entre sí.

Referencia bibliográfica

Zachary C. LIPTON, Alexandra CHOULDECHOVA y Julian MCAULEY. Does Mitigating ML's Impact Disparity Require Treatment Disparity? En: ArXiv:1711.07076 [Cs, Stat], November 19, 2017. Disponible en: <https://theodi.org/article/data-et-hics-canvas/>.

- **¿Se alcanza la paridad de tasa de falsos positivos?** La paridad de tasa de falsos positivos se consigue si las tasas de falsos positivos (división de falsos positivos con todos los negativos) en los subgrupos están cercanas entre sí.
- **¿Se alcanza la paridad de tasa positiva negativa?** La paridad del valor predictivo negativo se logra si los valores predictivos negativos (división de negativos verdaderos con todos los negativos predichos) en los subgrupos están cerca uno del otro.
- **¿Se alcanza la paridad de especificidad?** La paridad de especificidad se logra si las especificidades (división de negativos verdaderos con todos los negativos) en los subgrupos están cercanas entre sí.

3.3.3. *Data ethics canvas*

Teniendo en cuenta los sesgos y riesgos de los proyectos analíticos, resulta sumamente importante tener un marco de referencia para discutir los aspectos éticos.

Data ethics canvas (<https://theodi.org/article/data-ethics-canvas/>), creado por el Open Data Institute, es un marco de referencia que permite discutir diferentes aspectos relevantes para entender el impacto de las iniciativas analíticas. Consiste en una batería de preguntas que cubren múltiples aspectos:

- **Compartir estos datos:** ¿Con quién los compartimos? ¿Por qué? ¿Cómo?
- **Leyes, políticas y clasificación:** ¿Estamos en línea con el RGPD, con la Ley de Protección de Datos de 2018 y con la política de clasificación y manejo de la información cooperativa? ¿Es confidencial la información?
- **Marco ético existente:** ¿Se alinea con los valores éticos de la cooperativa?
- **Propósito para usar estos datos:** ¿Existe alguna razón buena y mutuamente beneficiosa para recopilar o usar los datos? ¿La recopilación de datos mejora las cosas para los miembros o podemos obtener información sobre los productos a partir de ellos?
- **Fuentes de datos:** ¿De dónde vienen los datos?
- **Limitaciones de las fuentes de datos:** ¿Los datos son de baja calidad o tienen un sesgo conocido?
- **Compartir datos con otros:** ¿Con quién los compartimos? ¿Por qué? ¿Cómo?
- **Contexto ético y legislativo:** ¿Se alinea con los valores éticos de organización? ¿Estamos en línea con el RGPD? ¿Es confidencial la información?
- **Derechos referentes a fuentes de datos:** ¿Tenemos permiso para hacer lo que planeamos hacer con ellos?
- **Comunicar su propósito:** Cuando solicitamos datos, ¿estamos explicando por qué de la manera más adecuada?
- **Efectos positivos en las personas:** ¿Cómo podemos aumentar el impacto positivo del proyecto y cómo podemos medirlo?
- **Efectos negativos en las personas:** ¿Hay algún punto donde existe la posibilidad de una violación de datos?

- **Minimizar el impacto negativo:** ¿Dónde podemos reducir el daño y cómo podemos medir el impacto?
- **Compromiso con las personas:** Describa cómo las personas pueden interactuar con usted y su proyecto, si las personas afectadas pueden apelar o solicitar cambios en el producto o servicio.
- **Franqueza y transparencia:** ¿Dónde están los riesgos financieros y de reputación?
- **Implementación continua:** ¿Cómo revisamos de forma continua los aspectos éticos?
- **Revisión e iteración:** ¿Cuándo deberíamos volver a usar este marco de referencia?
- **Tus acciones:** ¿Qué sucede después? ¿Quién está haciendo qué y dónde deben buscar ayuda si es necesario?

Responder a las preguntas que conforman el *data ethics canvas* nos permite:

- Identificar cómo se define e implementa la ética dentro de una organización.
- Revisar el concepto de ética de los datos de nuestra organización.
- Entender cómo se procesan los datos de manera ética.
- Crear una fotografía de la situación actual en cualquiera de nuestros proyectos.

En la figura 20 podemos encontrar las preguntas que pertenecen a cada bloque (se recomienda acceder directamente a la página del marco de referencia para descargar la última versión en formato Word):

Figura 20. Data ethics canvas



Fuente: ODI.

Bibliografía

Big Data Analytics: Adoption and Employment Trends, 2012-2017. e-skills UK. 2013. Esponsorizado por SAS. Disponible en: <http://www.sas.com/offices/europe/uk/downloads/bigdata/eskills/eskills.pdf>

Big Data: the next frontier for innovation, competition and productivity. McKinsey, 2011. Disponible en: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

BLOOM, Nicholas, BRYNJOLFSSON, Erik, FOSTER, Lucia, JARMIN, Ron S., SAPORTA EKSTEN, Itay y VAN REENEN, John. Management in America. En: *US Census Bureau Center for Economic Studies, Paper No. CES-WP-13-01*, 2013. Disponible vía SSRN: <http://ssrn.com/abstract=2200954>

BRYNJOLFSSON, Erik, HITT, Lorin M. y KIM, Heekyung Hellen. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? 2011. Disponible vía SSRN: <https://ssrn.com/abstract=1819486> o <http://dx.doi.org/10.2139/ssrn.1819486>

CAPGEMINI y MIT Center for Digital Business. *The Digital Advantage: How Digital Leaders Outperform their Peers in Every Industry*, 2012.

CASTELLS, M. *La era de la información: economía, sociedad y cultura* (vol. 1). Chicago: Siglo XXI, 1999.

CONESA, J. y CURTO, J. *Introducción al Business Intelligence*. Barcelona: Editorial UOC, 2016.

DAVENPORT, T. H., HARRIS, J. G. y MORISON, R. *Analytics at work: Smarter decisions, better results*. Cambridge, MA: Harvard Business Press, 2010.

DAVENPORT, T., COHEN, D. y JACOBSON, A. *Competing on Analytics, Working Knowledge Research Report*. Babson Executive Education, 2005.

DAVENPORT, Thomas H., HARRIS, Jaenne G., MORISON, Robert. *Analytics at Work: Smarter Decisions, Better Results*. Harvard Business Review Press, 2010.

DRESNER, Howard. *Wisdom of Crowds Business Intelligence Market Study*. Dresner Advisory Services, 2012.

GANDHI, S., THOTA, B., KUCHEMBUCK, R. y SWARTZ, J. Demystifying data monetization. En: *MIT Sloan Management Review*. 2018. Disponible en: <https://sloanreview.mit.edu/article/demystifying-data-monetization/>

HINDLE, Giles A. y VIDGEN, Richard. Developing a business analytics methodology: A case study in the foodbank sector. En: *European Journal of Operational Research*. 2018. Vol. 268.3, págs. 836–851.

IDC (John GANTZ, David REINSEL y John RYDNING). *The Digitization of the World: From Edge to Core*, 2018. Disponible en: <https://www.seagate.com/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

JANSSEN, M., van der VOORT, H. y WAHYUDI, A. Factors influencing big data decision-making quality. En: *Journal of Business Research*. 2017. Vol. 70, págs. 338–345.

LAVALLE, Steve, LESSER, Eric, SHOCKLEY, Rebecca, HOPKINS, Michael S. y KRUSCHWITZ, Nina. Big Data, Analytics and the Path From Insights to Value. En: *MIT SLOAN Management Review*. 2011.

LIPTON, Zachary C., CHOULDECHOVA, Alexandra y MCAULEY, Julian. Does Mitigating ML's Impact Disparity Require Treatment Disparity? En: ArXiv:1711.07076 [Cs, Stat]. November 19, 2017. Disponible en: <https://theodi.org/article/data-ethics-canvas/>

LOCK, Michael. *Managing the TCO of BI: The Path to ROI is Paved with Adoption*. Aberdeen Group, 2012.

LUHN, H. P. A business intelligence system. En: *IBM Journal of research and development*. 1958. Vol. 2, n.º 4, págs. 314–319.

MANYIKA, James, CHUI, Michael, BUGHIN, Jacques, DOBBS Richard, BISSON, Peter y MARRS, Alex. *Disruptive technologies: Advances that will transform life, business, and the global economy*. McKinsey Insight & Publications, 2013.

MARCHAND, Donald A. y PEPPARD, Joe. Why IT Fumbles Analytics. En: *Harvard Business Review*. 2013.

MARIN, I. Data Science and Development Team Remote Communication: the use of the Machine Learning Canvas. En: 2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE), Montreal, QC, Canada, 2019, págs. 18–21, doi: 10.1109/ICGSE.2019.00018.

MCAFEE, Andrew y BRYNJOLFSSON, Erik. *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*. Norton & Company, 2014.

MCAFEE, Andrew y BRYNJOLFSSON, Erik. Big data: The management revolution. En: *Harvard Business Review*. 2012. Vol. 90, n.º 10, págs. 60–68.

MCAFEE, Andrew y BRYNJOLFSSON, Erik. Investing in the IT that makes a competitive difference. En: *Harvard Business Review*. 2008. Vol. 86, n.º 7/8.

MOLNAR, C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2019. Disponible en: <https://christophm.github.io/interpretable-ml-book>

NAVARRO ARNAL, Xavier. *Consultoría en los Sistemas de Información*. Universitat Oberta de Catalunya, 2006.

NTOUTSI, E., FAFALIOS, P., GADIRAJU, U. Bias in data-driven artificial intelligence systems. An introductory survey. En: *et al. WIREs Data Mining Knowl Discovery*. 2020. e1356. <https://doi.org/10.1002/widm.1356>

OTTO, B. y REICHERT, A. Organizing master data management: findings from an expert survey. En: Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010. Págs. 106–110.

PARMAR, Rashik, MACKENZIE, Ian, COHN, David y GANN, David. The new patterns of innovation. En: *Harvard Business Review*, 2014.

PATIL, D. J. y MASON, H. *Data Driven*. O'Reilly Media, Inc., 2015.

ORTER, Michael. *Competitive Advantage Creating and Sustaining Superior Performance*. Free Press, 1998.

PORTER, Michael. The five forces that shape strategy. En: *Harvard Business Review*, 2008.

WERNERFELT, B. A Resource-Based View of the Firm. *Strategic Management Journal*. 1984. Vol. 5, n.º 2, págs. 171–180.

Worldwide Big Data and Analytics Software Forecast, 2019-2023. IDC, septiembre 2019.