



ESADE

RAMON LLULL UNIVERSITY

INSTITUTE FOR
DATA-DRIVEN
DECISIONS

THE
ESADE
MBA

MAYO
2018

Adopción e impacto del Big Data y Advanced Analytics en España

PRESENTACIÓN

Hace más de cuatro siglos, **la revolución científica cambió el paradigma de la concepción europea del conocimiento**. Como resume el historiador Yuval Noah Harari, el conocimiento pasó de entenderse como la combinación de las sagradas escrituras y el uso de la lógica para extraer su sentido exacto, a **buscar la respuesta en los datos empíricos y modelos matemáticos**.

Este cambio llevó a grandes avances y descubrimientos en todas las disciplinas científicas, pero no consiguió resolver las preguntas que, seguramente, sean las más relevantes que se ha planteado la humanidad: la distinción entre el bien y el mal, y el sentido de la vida. El conocimiento ético siguió bebiendo de las escrituras y fue evolucionando a terrenos más humanistas, dando lugar a referentes como, por ejemplo, la Declaración de Derechos Humanos de 1948.

En los últimos años, el pacto de conocimiento creado ha empezado a dar un salto muy relevante. **Los datos han evolucionado a Big Data**, es decir, conjuntos de datos que por su **gran tamaño, complejidad o variabilidad** requieren una capacidad de procesamiento muy rápida y la utilización de nuevas estructuras de bases de datos y nuevas técnicas analíticas para su explotación.

Los modelos matemáticos se han sofisticado enormemente, dando lugar a lo que llamamos **Advanced Analytics**. Se trata de nuevas técnicas de tratamiento, visualización y modelado de datos, más automatizadas, con mayor capacidad de predicción, generación de recomendaciones y descubrimiento de *insights*, y que además permiten trabajar con mayor volumen y diversidad de datos.

Empresas e instituciones de todos los sectores están siendo partícipes de esta transformación, que ha llegado para quedarse. Se trata de un cambio profundo que está **alterando de forma muy significativa la concepción del mundo y la realidad**, y que continuará haciéndolo de forma cada vez más rápida con **la irrupción de la inteligencia artificial y la robotización**.

El presente estudio nace con el objetivo **de conocer el estado de adopción e impacto que está teniendo el Big Data** en las empresas de nuestro país. Mediante un cuestionario dirigido a los responsables ejecutivos y técnicos de las mismas, se ha analizado a las compañías respecto a **cinco dimensiones**: 1) **visión estratégica y modelo organizativo**; 2) **inversión y creación de valor**; 3) **cultura y talento**; 4) **infraestructura de datos**; y 5) **prioridades a futuro**.

La publicación se estructura en **cuatro secciones** principales. La primera corresponde a un **resumen ejecutivo**, donde se exponen las conclusiones más relevantes del análisis del mercado español a nivel de Big Data y analítica de datos.

La **segunda** corresponde a los **resultados de cada una de las dimensiones analizadas**, tanto a nivel global como por sector o tamaño de compañía.

En la tercera sección, se exponen **ejemplos de interés y casos de negocio relacionados con la transformación analítica** de algunas compañías participantes del estudio.

Finalmente, en la última sección, titulada **“Impacto del Big Data en la sociedad”**, se presentan opiniones sobre **qué desafíos puede plantear** el uso masivo de datos y los avances en inteligencia artificial **a nivel social y humano**, y cómo pueden llegar a cambiar la definición misma de qué somos como sociedad.

Como pasó con la revolución científica, el nuevo conocimiento -y la gestión empresarial- basado en **Big Data y Advanced Analytics carece**, por sí mismo, **de un tamiz ético** que nos permita juzgar los avances que pueda conllevar, o los medios necesarios para conseguirlos. Por este motivo, hemos querido incorporar esta sección final de reflexiones, que tienen por objetivo plantear un debate público sobre el impacto y la trascendencia social de este cambio de paradigma.

ESADE Institute for Data-Driven Decisions
Senior Partners



Estudio realizado con el apoyo de ARCVI.

ÍNDICE

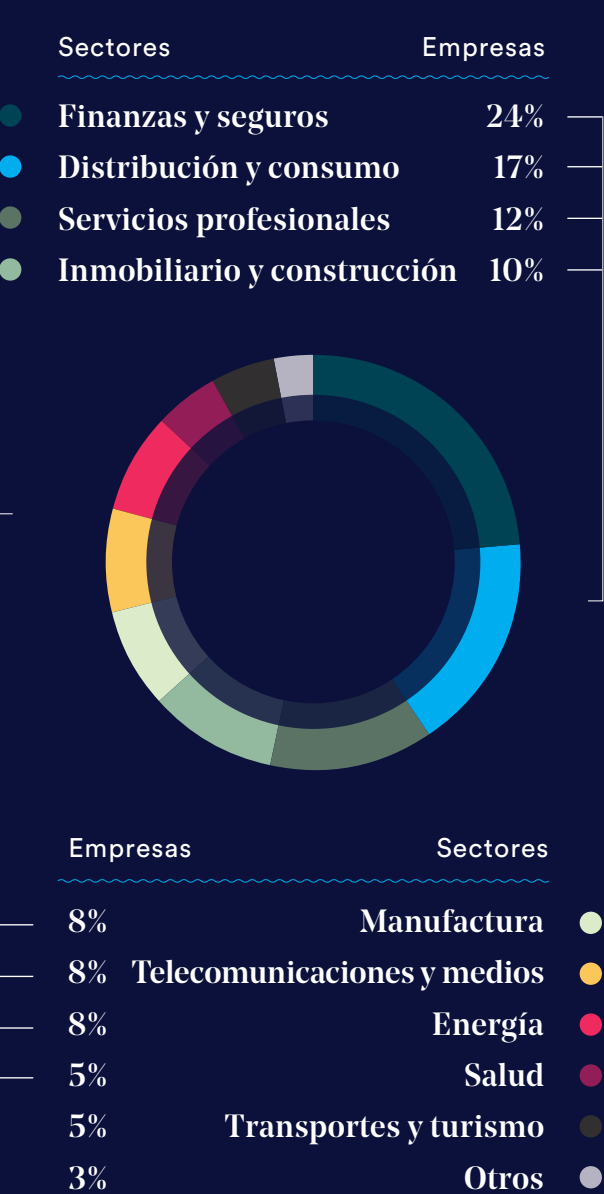
03	Presentación
06	Metodología
09	Resumen ejecutivo
10	Adopción e impacto del Big Data en las empresas
	Visión estratégica y modelo organizativo
	Nivel de inversión y creación de valor
	Cultura y talento
	Infraestructura de datos
	Estado de madurez y prioridades a futuro
42	Casos de uso
50	Impacto del Big Data en la sociedad

Metodología

La metodología utilizada para la obtención de la información es un cuestionario online, así como entrevistas personales realizadas a responsables de Big Data de las empresas participantes en el estudio.

La muestra es de 106 compañías que operan principalmente en España. Para aquellas de ámbito internacional, se han considerado los datos relativos al estado español.

Por sectores y volumen de facturación, las compañías se distribuyen según:



Empresas participantes

- ABERTIS INFRAESTRUCTURAS

ACCIONA, S.A.

ACM-CREDIT MUTUEL

AENA, S.A.

AIR EUROPA LINEAS AEREAS

APPLUS

AQUALOGY

ARAG

ASEPEYO

AXA

AZZURRO INNOVATION

BANCO SABADELL

BANCO SANTANDER

BANKIA

BASF ESPAÑOLA

BBVA

BLABLACAR

CAIXABANK

CARREFOUR

CASA VIVA

CBRE

CELLNEX TELECOM

CINESA

COCA COLA EUROPEAN PARTNERS

DAMM

DESIGUAL

DIGITAL ORIGIN

DKV SEGUROS

EBURY

EL CORTE INGLÉS

EVO BANCO

FAGOR EDERLAN GROUP

FERROVIAL

FINTONIC

FORCADELL CONSULTORES INMOBILIARIOS

FUJITSU

GAES

GB FOODS

GAS NATURAL FENOSA

GRUPO ANTOLIN

GRUPO EUROFRED

GRUPO GODÓ

GRUPO MÉMORA

GRUPO VIPS

HOLALUZ

IBERCAJA

IDEALISTA

IDILIA FOODS
- INDRA SISTEMAS

ING DIRECT

INSTITUTO DE INGENIERÍA DEL CONOCIMIENTO

IPSOS

LIBERBANK

LUCTA

MABRIAN TECHNOLOGIES

MANGO

MERCADONA

MULTIASISTENCIA

MUTUA MADRILEÑA

NETQUEST

OHL

ONEY

PEPSICO

PLAYGROUND

PRIMAGAS

RACC

RAKUTEN

RECYMET SYSTEMS

RED ELÉCTRICA DE ESPAÑA

REPSOL

RETAIL & TRADE MARKETING

RIBÉ SALAT

RICOH ESPAÑA

ROCHE DIAGNOSTICS

ROUSSELOT

SAMSUNG

SCHNEIDER ELECTRIC

SEAT

SEGURCAIXA ADESLAS

SEIDOR

SELLBYTEL GROUP

SIEMENS

SIMON

TELECABLE

TELEFÓNICA

THE COCKTAIL

TIENDA ANIMAL

TÜV RHEINLAND SERVICE GMBH

ULABOX

UNIBAIL-RODAMCO

UNIDAD EDITORIAL

VALTECNIC

VIDACAIXA

VIESGO ENERGÍA

VODAFONE

VOLSKWAGEN GROUP ESPAÑA DISTRIBUCIÓN

VUELING

WALLAPOP

WAYFAIR

WERFEN GROUP

YAENCONTRÉ

YAMAHA MOTOR EUROPE

ZANK

ZURICH ESPAÑA

Nota: No aparecen todas las participantes dado que algunas empresas han preferido mantener el anonimato.

Resumen Ejecutivo

El estado de la transformación analítica en las empresas

40%

No dispone de un rol específico de liderazgo en Big Data y Advanced Analytics (BDAA) o tiene menos de un año, aunque prevén crearlo y consolidarlo a corto plazo.



Los dos motivos principales para iniciar la transformación BDAA en las empresas:

68% 50%

Relación con el cliente



Automatización de procesos



85%

Han iniciado pilotos o proyectos con un éxito por encima de la media de otros procesos de transformación.



1,8M€

De presupuesto anual en las empresas de más de 200M de facturación.



55%

Consideran que no tienen una cultura de toma de decisiones basadas en datos.



7,7



De dificultad media en la obtención de perfiles analíticos. (siendo 0 muy fácil y 10 muy difícil)

50%

Utilizan *cloud* de forma completa o parcial para el almacenamiento de datos.



>250%

De incremento previsto en inversión y personal dedicado a Big Data en los próximos

3 años



Adopción e impacto del Big Data en las empresas

Visión estratégica y modelo organizativo

Los conceptos Big Data y Advanced Analytics (BDAA) irrumpieron en el mundo empresarial y en la sociedad hace pocos años, especialmente tras la publicación en 2011 del estudio *Big data: The next frontier for innovation, competition, and productivity*, del McKinsey Global Institute.

Desde ese momento, ambos conceptos empezaron a ser un tema en la agenda de los directivos de las compañías que fue ganando importancia hasta llegar a ser una prioridad. No en vano, ocupó varias portadas de las principales publicaciones económicas, como *Harvard Business Review* o *The Economist*.

Sin embargo, más allá de las definiciones técnicas, estos conceptos han ido adquiriendo diferentes connotaciones para los diferentes sectores económicos y compañías.

Foco en cliente, optimización de procesos e innovación

Más allá de las primeras publicaciones, las impresiones y estimaciones iniciales y las definiciones técnicas, las compañías han ido estableciendo sus propias visiones estratégicas sobre qué es y qué pueden representar los conceptos Big Data y Advanced Analytics para su negocio.

Preguntadas sobre la visión y las aportaciones esperadas de BDAA, las compañías en España vinculan Big Data y Advanced Analytics principalmente con la **mejora de la relación con el cliente** (~70% de las empresas). Se trata de una visión transversal en todos los sectores, que mayoritariamente coinciden en señalar este motivo entre los tres más relevantes para su incursión en BDAA (ver gráficos 1 y 2).

Las empresas encuestadas también coinciden, y de forma bastante transversal, en que otros factores relevantes para incorporar BDAA son la **reducción de costes y optimización de procesos**, así como la **mejora e innovación en productos y servicios**.

Casi el 70% de las empresas vinculan Big Data y Advanced Analytics con la mejora de la relación con el cliente

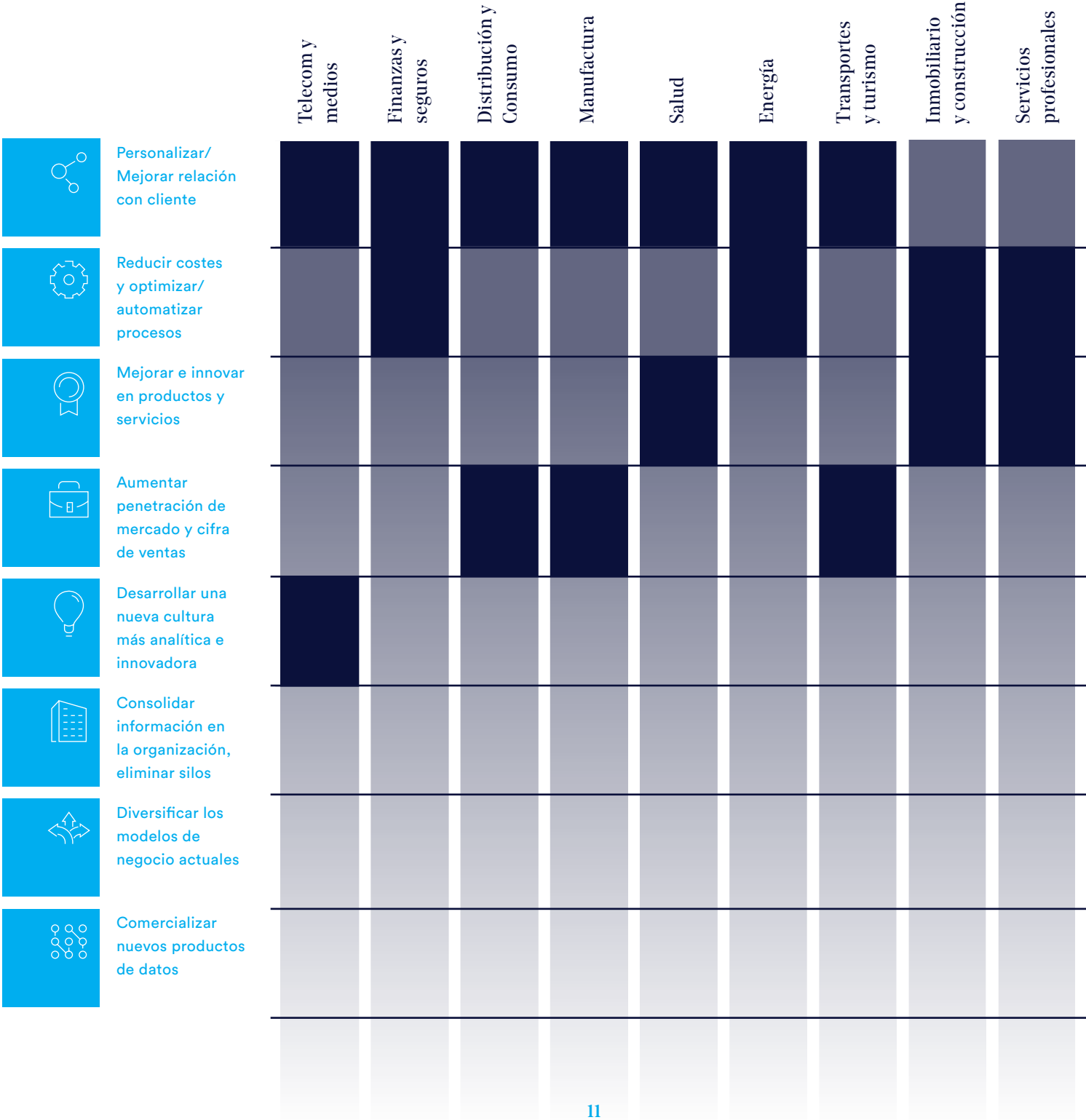
Gráfico 1: Principales factores que motivan la incursión en Big Data y Advanced Analytics a nivel global (% de empresas)



*Preguntas multirespuesta en base a los 3 factores más relevantes seleccionados por cada empresa.

La plasmación más evidente y gráfica de la importancia que está adquiriendo el Big Data en los factores *cliente* y *optimización* de negocio en las empresas son las palabras que utilizan para definir sus implicaciones estratégicas. En todos los sectores, *cliente* es una de las palabras más repetidas por las empresas encuestadas, junto con conceptos como *optimización*, *personalización*, *análisis*, *datos*, *valor* o *costes* (ver gráfico 5).

Gráfico 2: Principales factores que motivan la incursión en Big Data y Advanced Analytics por sector



Vinculación a la tecnología y la digitalización

En las empresas se aprecia también una fuerte vinculación de Big Data y Advanced Analytics a la digitalización, tanto de productos y servicios como de procesos.

Marketing Digital y Sistemas y Tecnologías de la Información son las funciones de negocio que concentran más usuarios de BDAA, en un ~60% y 44% de las empresas respectivamente (ver gráfico 3). Nuevamente, es una visión bastante transversal en todos los sectores, si bien algunos tienen otras funciones que concentran también de forma relevante a los usuarios de BDAA como, por ejemplo, la gestión del riesgo y fraude en Finanzas, o pricing y promociones en Distribución y Consumo (ver gráfico 4).

Marketing Digital y Sistemas y Tecnologías de la Información son las funciones de negocio que concentran más usuarios de Big Data y Advanced Analytics

A modo de ejemplo, es interesante señalar qué tipo de aportaciones destacan algunos de los responsables de las compañías participantes:

“Trabajamos en el desarrollo de soluciones BDAA (...) con especial dedicación al entorno digital”.

0

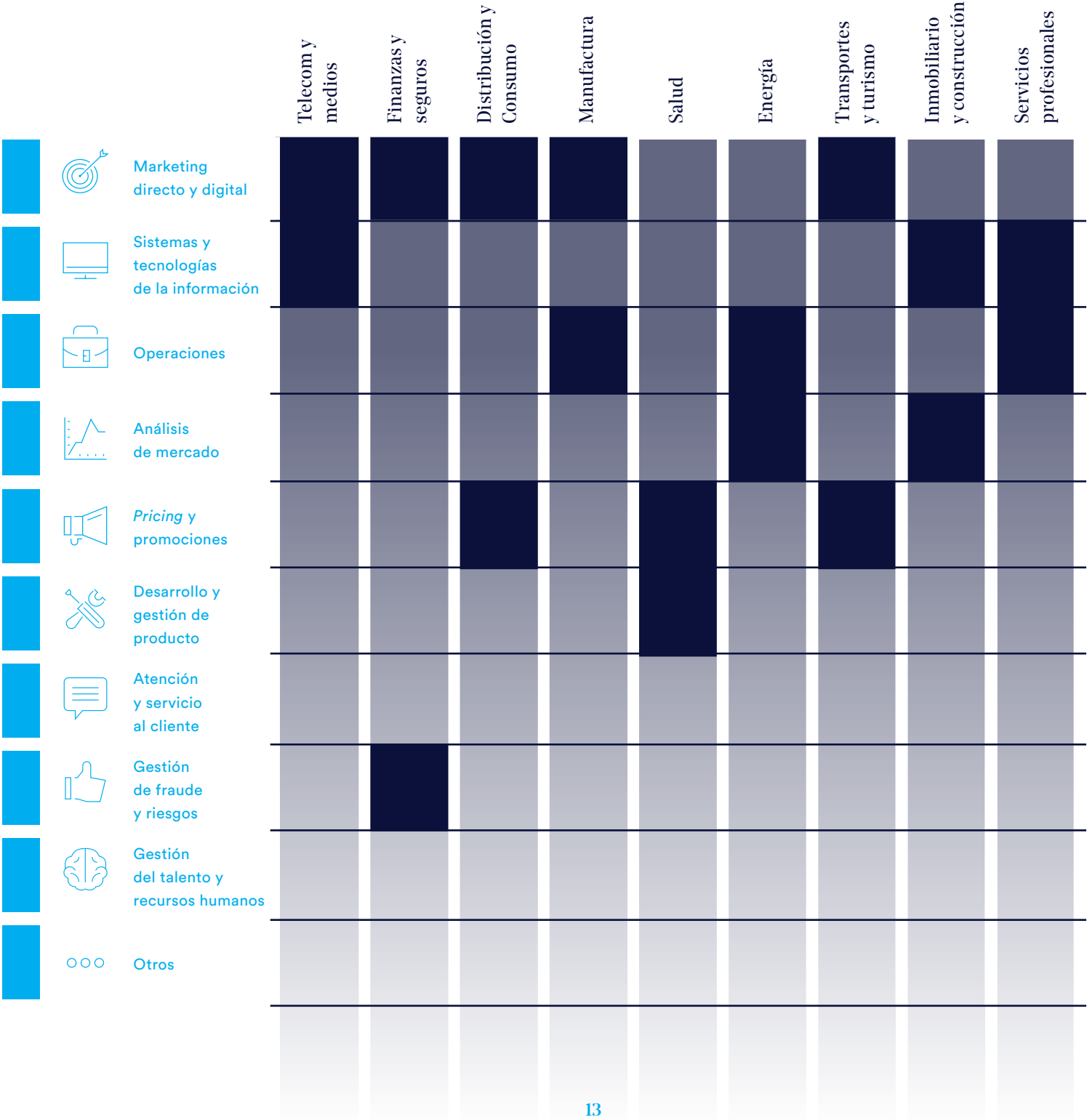
“Existen varias iniciativas BDAA relacionadas con la creación de software médico que permita asesorar en la toma de decisiones y monitorizar la evolución de pacientes”.



Gráfico 3: Funciones de negocio que concentran los principales usuarios de Big Data y Advanced Analytics a nivel global (% de empresas)

Nota: Pregunta multirespuesta de selección de 3 factores principales

Gráfico 4: Funciones de negocio que concentran los principales usuarios de Big Data y Advanced Analytics por sectores



Entre lo algorítmico y lo emocional

En la visión estratégica de Big Data Advanced Analytics en España conviven dos tendencias complementarias. Al ser preguntadas por qué representa BDAA, en las empresas se observa que la ejemplifican atributos como *modelos*, *algoritmos*, *eficiencia*, *análisis*, *técnicas*, *herramientas*, y tiene que ver con la parte más matemática, vinculada con la **optimización de procesos**, la **automatización** y la **eficiencia**.

Se trata de una visión presente en todos los sectores, dado que en todos existe terreno para este tipo de aportaciones de BDAA. No en vano, como señalábamos antes, un ~50% de las empresas cree que la **optimización y automatización** es una de las tres aportaciones más relevantes que puede tener BDAA, y en el 44% Sistemas y Tecnología se encuentran entre las funciones con más usuarios.

Por otro lado, existe una visión más **cálida**, **emocional**, **humana**. La ejemplifican atributos como *personalización*, *cliente*, *fidelización*, *conocimiento*, *comprensión*, y tiene que ver con la parte más cualitativa, vinculada con la **mejora en la relación y el conocimiento del cliente**, la **innovación y la creatividad**.

La combinación de las dos será una pieza clave a nivel estratégico para el éxito de las iniciativas BDAA y de los procesos que acompañan esta transformación, como la digitalización, la automatización y la robotización.

Gráfico 5: Wordcloud de
¿Qué es Big Data? por sectores



Telecomunicaciones y medios

cliente / **estrategia** /
personalización /
monetización / **adquisición** /
arpu / **digital** / **fidelización** /
audiencia / **análisis** / **datos** /
herramientas / **segmentación** /
desarrollo / **información** / **mejora** / **negocio** /
técnicas / **predicción** / **recomendación** /
toma de decisiones / **click** / **acciones**

Finanzas y seguros

cliente / **datos** /
modelos / **riesgo** /
comercial / **negocio** /
análisis / **servicios** / **gestión** /
mejora / **crédito** / **digital** /
propensión / **algoritmos** / **áreas** /
información / **aportaciones** / **banco** /
fraude / **mayor** / **predictivos** / **productos** /
relación / **capacidad** / **empleados** / **iniciativa**

Salud

médico /
análisis / **valor** /
asesorar / **prevención** /
absentismo / **beneficios** /
causas / **mutualistas** /
diagnóstico / **cliente** /
comercial / **datos** / **eficiencia** /
hospitales / **monitorizar** /
patrones / **servicio** /
toma de decisiones /
analytics / **clientes** / **relación** / **transformación**

Energía

clientes / **análisis** /
consumo / **datos** /
información / **procesos** /
energía / **ahorrar** / **demanda** /
mantenimiento / **mejora** / **mejorar** /
negocio / **rentabilidad** / **churn** /
predecir / **pricing** / **rendimiento** /
activos / **cliente** / **comportamiento** /
experiencia / **mejor** / **necesidades** /
optimización / **predicción** / **red**

Distribución y consumo

consumidor /
análisis / **negocio** /
marketing / **pricing** /
datos / **producto** /
optimización / **información** /
comportamiento / **herramientas** /
modelos / **procesos** / **comercial** /
eficiencia / **ingresos** / **objetivo** / **personalización** /
ventas / **toma de decisiones** / **canales** / **coste**

Servicios profesionales

datos / **clientes** /
negocio / **servicios** /
calidad / **desarrollo** /
marketing / **algoritmos** /
automatizar / **soluciones** /
áreas / **decisiones** / **diferentes** /
seguridad / **ventas** / **captación** /
crecimiento / **machine** /
learning / **análisis** / **comportamiento** /
conocimiento

Inmobiliario y construcción

cliente / **valor** /
analizar / **constructivos** /
negocio / **productos** /
conocimiento /
inteligencia / **servicio** /
mejora / **estrategia** / **gestión** /
potenciar / **comprensión** / **disponer** /
dispositivos / **eficiencia** / **experiencia** /
forma / **mayor** / **predecir** / **recursos** /
competitiva / **excelencia**

Manufactura

clientes / **valor** /
transversal / **calidad** /
digital / **servicios** /
producción / **asistencia** /
costes / **estratégica** / **motor** /
rendimiento / **compañía** / **datos** /
mercado / **mejorar** / **distribución** /
automatización / **capacidades** / **conocimiento** /
desarrollo / **fidelización**

Transportes y turismo

datos / **costes** /
optimización / **analítica** /
estacionalidad / **estrategia** /
marketing / **operaciones** /
rentabilidad /
negocio / **desarrollo** /
pasajeros / **viajes** /
decisiones / **empresa** / **procesos** /
ayuda / **conexiones** / **integrar** /
planificación / **red**

En un 40% de las compañías participantes en el estudio todavía no existe un rol de liderazgo específico en Big Data y Advanced Analytics o es inferior a un año

MODELO ORGANIZATIVO

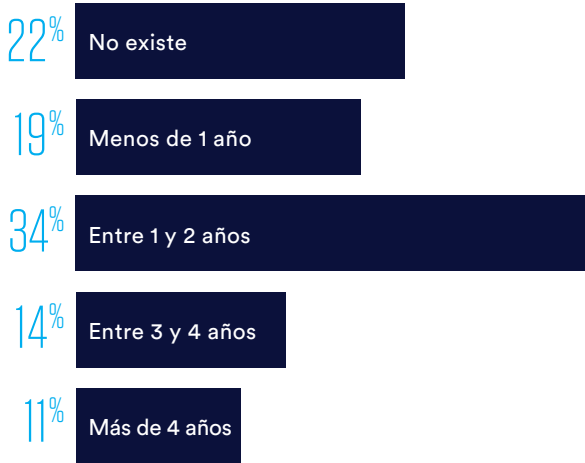
Trasladar la visión estratégica a un modelo organizativo que funcione es, sin duda, uno de los grandes desafíos que afrontan las compañías en su adopción de BDAA.

En primer lugar, porque, como cualquier cambio, conlleva **resistencias culturales y alteraciones en los equilibrios de poder** de diferentes áreas y funciones. En segundo lugar, porque todavía no existen suficientes referentes consolidados como para establecer mejores prácticas claras.

Lejos de la plena adopción, pero a buen ritmo

Transcurridos unos 7 años desde la irrupción de BDAA, en un **40% de las compañías** participantes en el estudio todavía **no existe un rol de liderazgo específico en BDAA, o tiene menos de un año**. Sólo una minoría, alrededor de un 25%, tiene liderazgos consolidados con 3 o más años de antigüedad (ver gráfico 6).

Gráfico 6: ¿Cuánto hace que fue creado el rol específico de liderazgo en Big Data y Advanced Analytics en su empresa? (% de empresas)



Este nivel de adopción no es especialmente lento. Los grandes cambios tecnológicos de la segunda mitad del siglo XX y de principios del XXI han necesitado unos 14 años para llegar a adopciones superiores al 80%. Por lo tanto, situarnos en una adopción del 60% en 7 años entra dentro de la normalidad.

Gráfico 7: Tiempo de adopción de innovaciones tecnológicas

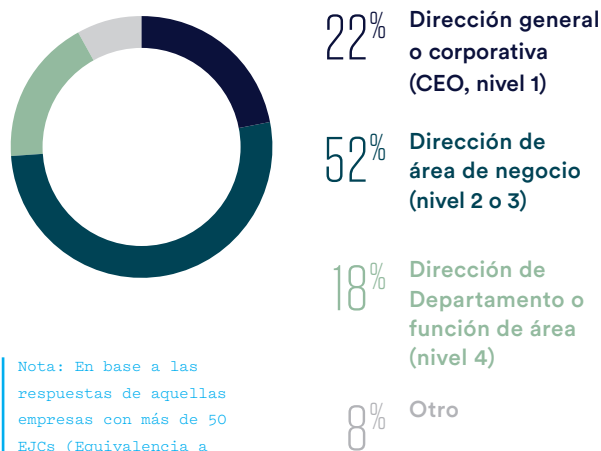
Año aparición	Años hasta 80% adopción	Tecnología
1950	10	TV
1960	15	Color TV
1972	22	Instrument Landing Systems
1984	20	PC
1989	15	Móviles
1990	18	Internet
2000	16	CRM-Cloud
2007	8	Smartphone

Fuente: Nicholas Felton, New York Times

Rango de dirección de área de negocio

Para la mayoría de las compañías, estos liderazgos en BDAA residen en general a nivel de área de negocio (nivel 2 o nivel 3), aunque es muy destacable que en un 22% de los casos ostenten el nivel de dirección general o corporativa (ver gráfico 8).

Gráfico 8: ¿En qué nivel reside el “liderazgo ejecutivo” en Big Data y Advanced Analytics en su organización? (% de empresas)



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC (Equivalencia a Jornada Completa)

En cuanto a funciones, las direcciones de **Tecnología y Digital lideran en un 35% de las compañías**, siendo relevante que **una de cada cinco compañías ha creado una dirección específica**, es decir, una figura exclusiva centrada en el liderazgo BDAA (ver gráfico 9). Cabe señalar que en las empresas con liderazgos más consolidados esta última opción es claramente mayoritaria (ver gráfico 9B).

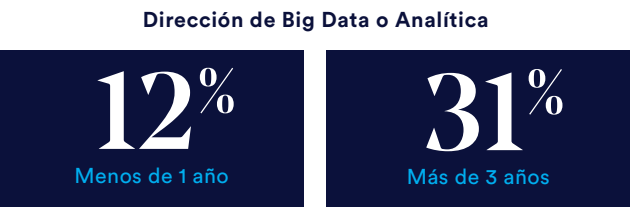
Gráfico 9: Función directiva que ostenta o impulsa el liderazgo ejecutivo en Big Data y Advanced Analytics (% de empresas)



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC (Equivalencia a Jornada Completa). 1. Ejemplos en esta categoría incluyen un responsable de estrategia o un responsable de gestión de clientes.

Una de cada 5 compañías ha creado una figura exclusiva centrada en el liderazgo en Big Data y Advanced Analytics

Gráfico 9B: Empresas con dirección propia y exclusiva en Big Data y Advanced Analytics (% de empresas)



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC

Mejorar el desempeño de otras áreas e influenciar en la visión estratégica a futuro

Muy en consonancia con la visión de qué representa y qué se espera de BDAA, las compañías **asignan mayoritariamente al rol de su liderazgo** las funciones de promover el análisis y el modelado para mejorar el desempeño de otras áreas, y participar en la definición de la nueva estrategia de la compañía a futuro, así como liderar su transformación (ver gráfico 10).

Gráfico 10: ¿Cuáles cree que deberían ser las funciones principales del Responsable de BDAA en su compañía durante los próximos 2 años? (% de empresas)



Nota: En base a las 3 funciones más relevantes seleccionadas por cada empresa.

Por el momento, diversidad de modelos organizativos

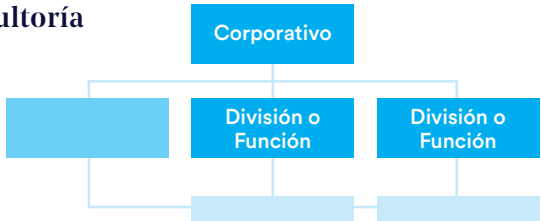
Llegados a este punto, una vez establecida una figura de liderazgo en BDAA y sus funciones, la pregunta clave para las compañías es cómo organizar a los perfiles dedicados.

Se han identificado **4 esquemas de organización principales** (ver figura Modelos organizativos de Big Data y Advanced Analytics), en línea con los establecidos por el experto internacional Thomas Davenport, que podríamos definir como:

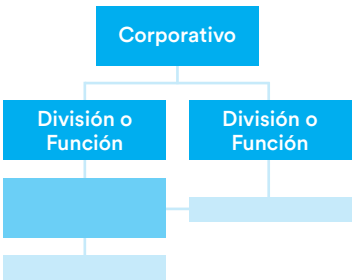
- 1. **División centralizada o consultoría:** una unidad propia e independiente de otras funciones y divisiones concentra a los perfiles analíticos de la compañía y su liderazgo, y ofrece servicio a otras áreas.
- 2. **Unidad funcional especialista:** un área perteneciente a la unidad que requiere de forma principal los servicios BDAA concentra a los perfiles analíticos de la compañía y su liderazgo, y ofrece servicio a otras áreas.
- 3. **COE:** diferentes unidades pueden contar con perfiles propios y proyectos en curso, pero todos ellos reportan también a una unidad corporativa global dónde reside el liderazgo BDAA.
- 4. **Descentralizado:** diferentes áreas disponen de perfiles y liderazgos propios, que ofrecen servicios a las propias áreas.

Modelos organizativos de Big Data y Advanced Analytics

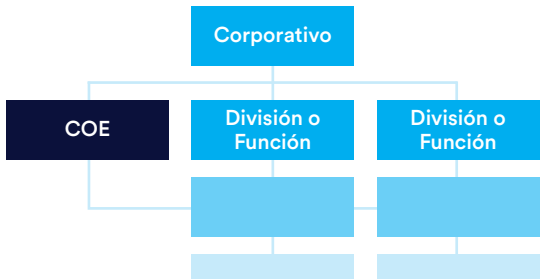
División centralizada o consultoría



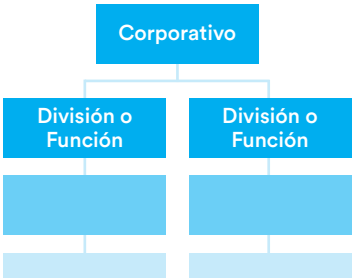
Unidad funcional especialista



Centro de Excelencia (COE)



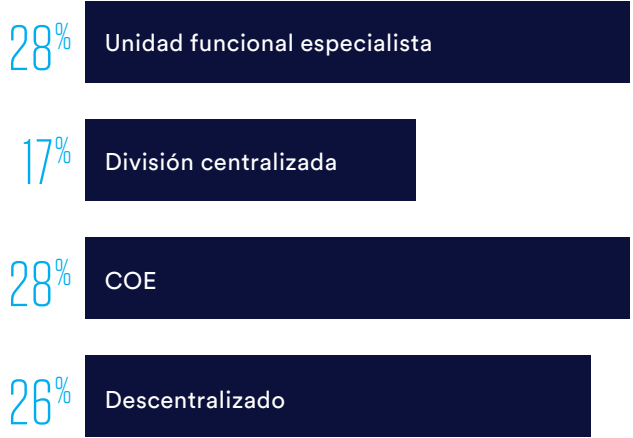
Descentralizado



Grupos analíticos Proyectos analíticos

La distribución de compañías según modelo organizativo es, a día de hoy, **muy equitativa**, con la salvedad de la división centralizada o consultoría, que tiene una presencia ligeramente inferior en las compañías.

Gráfico 11: Modelos organizativos de Big Data y Advanced Analytics (% de empresas)



Nota: Únicamente se incluyen aquellas empresas con más de 50 EJs, con capacidades y estructura definida

Nivel de inversión y creación de valor

La aparición de los conceptos Big Data y Advanced Analytics vino acompañada de estimaciones de impactos y potenciales de transformación muy significativos. Desde un incremento del 60% en el margen operativo en distribución y consumo, hasta 250.000M€ de ahorros y mejoras anuales en el sector público europeo para todos los sectores se presentaron grandes oportunidades [Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011].

Transcurridos los primeros años de adopción, es un buen momento para analizar en qué estado se encuentran las inversiones e iniciativas en BDAA, y qué impactos están teniendo.

Estado desigual de avance de las iniciativas Big Data, pero en general en primeros pilotos

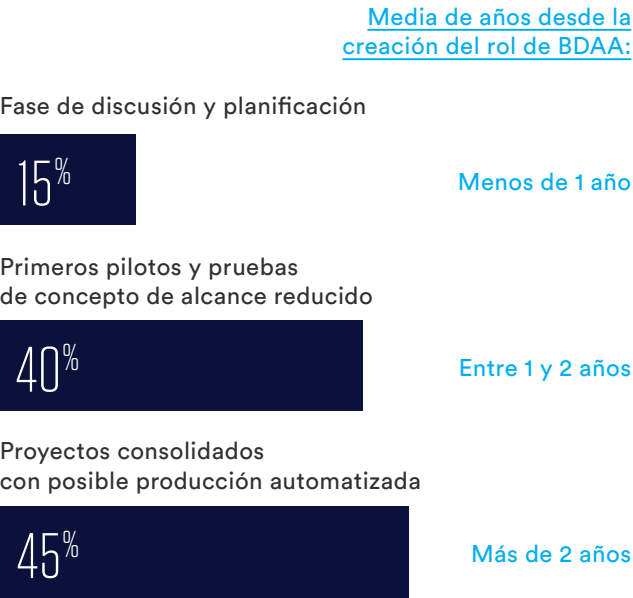
Existen diferencias evidentes entre el estado de avance de las iniciativas Big Data en las compañías españolas, muy relacionadas, como no podía ser de otra forma, con el tiempo que hace que crearon unidades y roles específicos.

Así, alrededor del 15% de las compañías se encuentran en fase de discusión y planificación, coincidiendo con la mayoría de empresas que todavía no tienen una figura de liderazgo en Big Data y Advanced Analytics.

El 45% de las empresas dispone de proyectos consolidados en curso e incluso en producción automatizada y en tiempo real

Hasta un 40% de las compañías están ya realizando primeros pilotos y pruebas de concepto de alcance reducido, mientras que el 45% restante dispone de proyectos consolidados en curso e incluso en producción automatizada y en tiempo real (ver gráfico 12). Este último grupo está integrado por compañías cuyas unidades de Big Data y Advanced Analytics fueron creadas hace unos 2 años, umbral que puede ser tomado como referencia para estimar el tiempo necesario para empezar a consolidar las iniciativas en las organizaciones.

Gráfico 12: Estado general de las iniciativas en Big Data y Advanced Analytics en las empresas (% de empresas)



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC's y más de 100M€ de facturación

Alta creación de valor, transversal a diferentes áreas

El impacto y beneficios que están teniendo y se espera que tengan las iniciativas en curso muestra dos conclusiones claras.

En primer lugar, las empresas están dejando atrás un cierto hype sobre Big Data. La mayoría de empresas sitúan el impacto obtenido en sus primeros proyectos de BDAA como alto o moderado, aunque sólo un 11% de ellas ha conseguido impactos muy altos disruptivos. Además, un 21% de las empresas encuestadas considera que todavía es demasiado pronto para poder evaluar los beneficios (ver gráfico 13).

En segundo lugar, tanto a nivel global como por sectores, las compañías apuntan a **beneficios transversales que involucran diferentes áreas** por igual (ver gráficos 14 y 15). Desde mejora en la personalización de productos y servicios, hasta desarrollo de nuevas líneas de negocio, pasando por pricing y ofertas comerciales, mejoras operacionales (previsión de consumos, gestión de stock, mantenimiento...), mejoras en la relación y comunicación, o identificación de oportunidades de venta cruzada y up-sell.

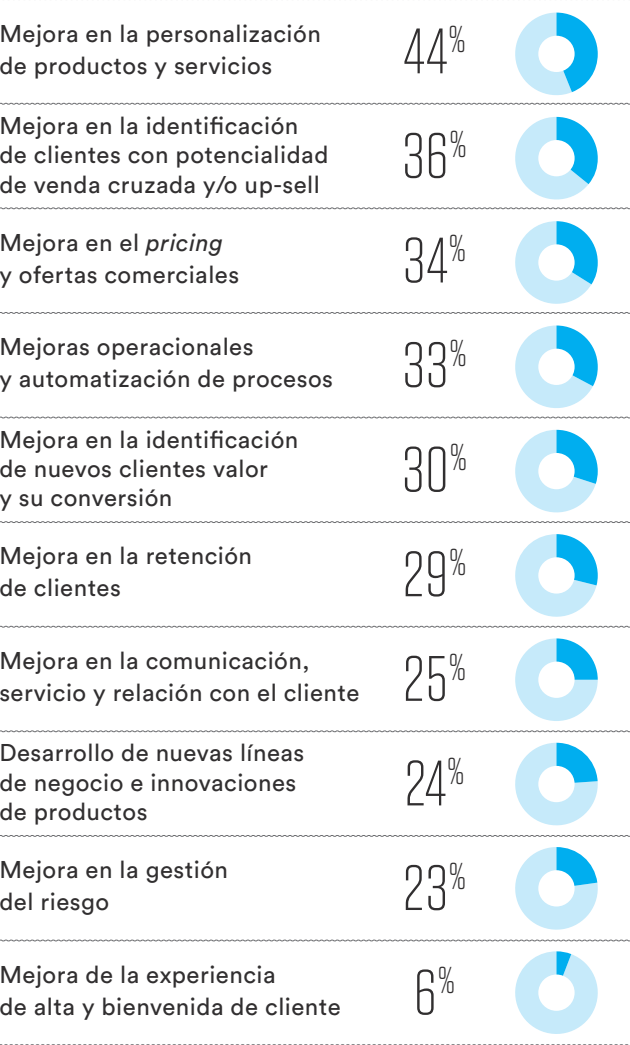
Gráfico 13: ¿Cómo valoraría el retorno de los proyectos lanzados durante los últimos 12 meses respecto a procesos típicos de transformación u optimización? (% de empresas)



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC's

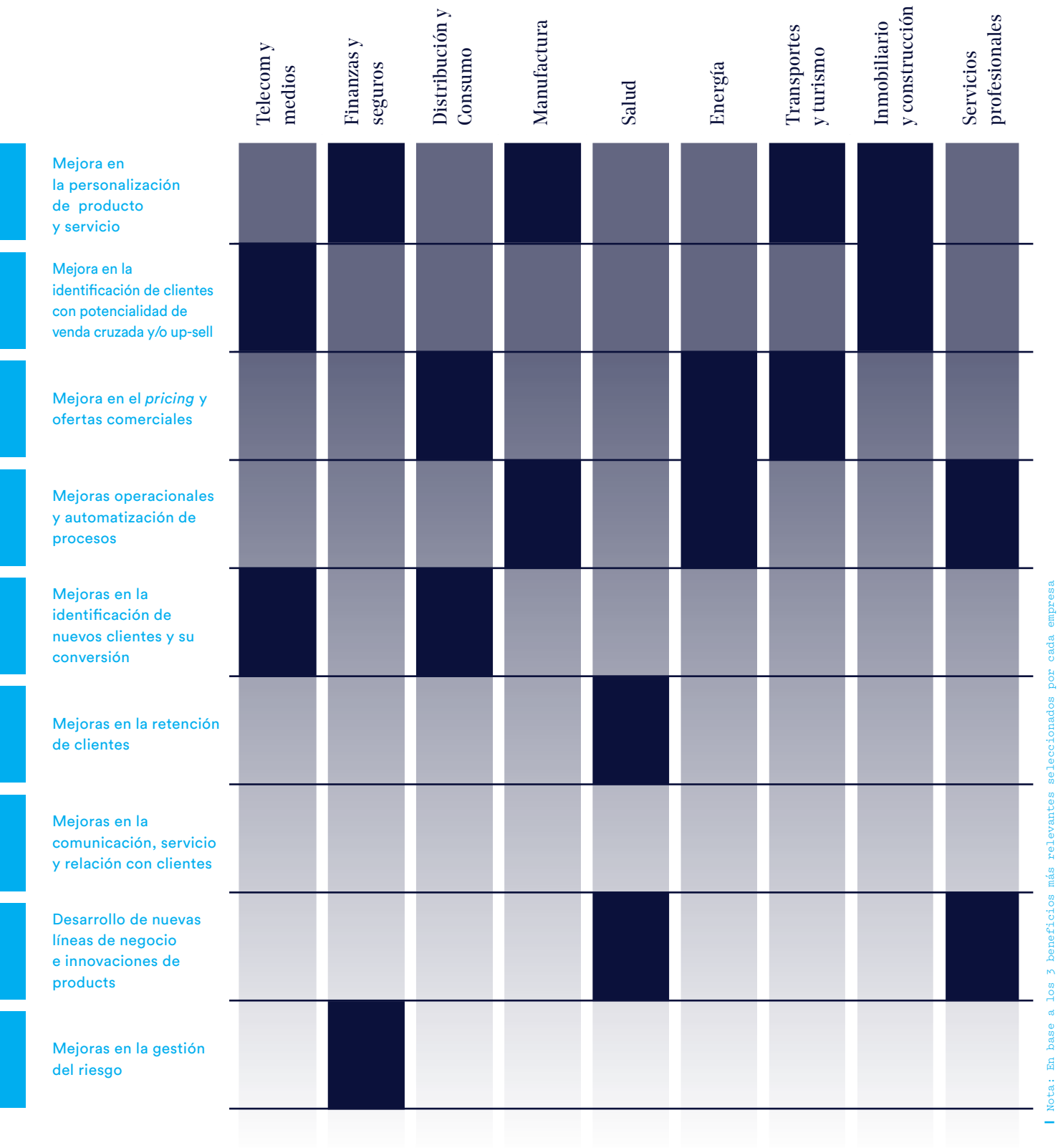
La mayoría de empresas sitúa el impacto obtenido en sus primeros proyectos de BDAA como alto o moderado

Gráfico 14: ¿Qué beneficios tangibles se esperan de las iniciativas BDAA? (% de empresas)



Preguntas multirespuesta en base a los 3 beneficios más relevantes seleccionados por cada empresa

Gráfico 15: Beneficios tangibles esperados por sector a partir de las iniciativas BDAA



Dificultad en la obtención de datos y falta de cultura analítica, principales barreras a la hora de capturar el valor de negocio

A pesar de este impacto moderado o alto y de la transversalidad de beneficios, las compañías manifiestan enfrentar algunos problemas a la hora de hacer tangibles los resultados de negocio de los proyectos.

Destacan especialmente dos factores, compartidos por el 60% de las compañías: la dificultad a la hora de obtener datos o garantizar la calidad de los mismos, y la falta de una cultura analítica arraigada en la compañía, que posibilite la comprensión de los modelos analíticos y enfoques planteados (ver gráfico 16).

Gráfico 16: ¿Cuáles considera que son las mayores barreras a la hora de tangibilizar proyectos BDAA a nivel de negocio? (% de empresas)



Preguntas multirespuesta en base a los 3 beneficios más relevantes seleccionados por cada empresa

El 60% de las compañías admiten dificultad a la hora de obtener datos o garantizar la calidad de los mismos

Además, la integración de los nuevos modelos, herramientas y procesos con los ya existentes, así como garantizar su integridad y minimizar los riesgos operacionales son otras de las barreras que más se destacan. No obstante, esta última barrera decae significativamente en importancia, hasta el 19%, en las empresas cuyas unidades BDAA fueron creadas hace al menos 3 años (ver gráfico 16B).

Gráfico 16B: Barreras a la hora de tangibilizar proyectos BDAA a nivel de negocio en las empresas con antigüedad mayor a 3 años (% de empresas)



Preguntas multirespuesta en base a los 3 beneficios más relevantes seleccionados por cada empresa

Presupuesto anual entre 300.000 y 1,8 millones de euros

La puesta en marcha de las unidades BDAA y el lanzamiento de primeras iniciativas ha requerido partidas presupuestarias de dos tipos fundamentales.

El primer grupo, **personal y procesos**, se refiere a los gastos asociados a capacitación y contratación de perfiles internos, y a la explotación de datos mediante análisis y modelos. El segundo, **infraestructuras e integración**, engloba todo aquello relacionado con el almacenamiento, adquisición e integración de datos, así como el cumplimiento de políticas de seguridad y privacidad.

Durante el último año, las empresas en España han dotado dichas unidades e iniciativas con un presupuesto que ha oscilado entre unos **300.000 euros para start-ups y empresas** de menos de 20M€ de facturación y **1,8M€ para grandes compañías** de más de 200M€ de ventas (ver gráfico 17).

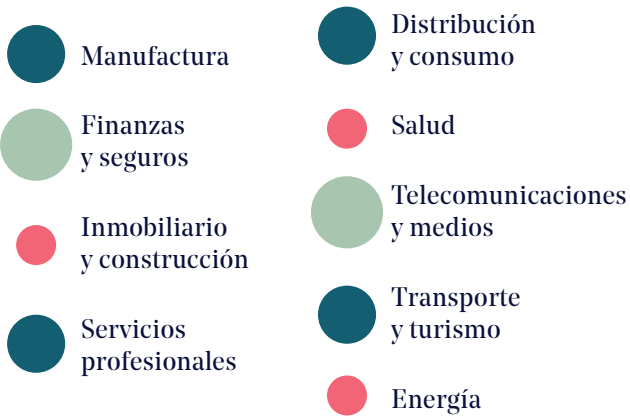
Gráfico 17: Presupuesto total destinado a iniciativas BDAA en el último año según tamaño de empresa (Millones de € invertidos por empresa)



Empresas pequeñas: menos de 20M facturación;
empresas medianas: 20M - 200M facturación;
empresas grandes: >200M facturación

Por sectores, **Telecomunicaciones y Finanzas son los que lideran la inversión** en España, con presupuestos que se sitúan más un 30% por encima de la media, en compañías de tamaño comparable. En el otro lado, **Energía y Salud son los que menos están invirtiendo** (ver gráfico 18).

Gráfico 18: Nivel de inversión por sector



Nota: Se han comparado presupuestos estandarizando por tamaño y considerando, para las multinacionales, el perímetro de inversión correspondiente a negocio generado en España.

Cultura y talento

La cultura es uno de los elementos más esenciales de una compañía. Impregna todos y cada uno de sus procesos y, sin embargo, o quizás por ser tan esencial, es invisible a los ojos.

Durante mucho tiempo lo analítico ha residido en un segundo plano de la cultura empresarial. La intuición, el olfato y la experiencia eran los elementos clave para la toma de decisiones. Algunos estudios relativamente recientes* destacaban que el 40% de las decisiones más relevantes no se basaban en datos sino en intuición.

Sin denostar el papel de la intuición, que es la suma de experiencia consolidada y talento, la investigación ha puesto de manifiesto los muchos sesgos y errores predecibles que puede conllevar. Investigadores como los Premio Nobel Daniel Kahneman y Richard Thaler, que se han centrado en construir marcos teóricos de los sesgos y errores de la intuición a partir de recoger evidencias empíricas, están teniendo gran influencia a nivel internacional en gobiernos y empresas.

Sólo el 45% de las empresas afirma tener una cultura analítica consolidada

Todavía lejos de tener culturas analíticas consolidadas

En España, las compañías están todavía en proceso de adopción de una cultura más analítica.

Sólo un 45% de las compañías está bastante o muy de acuerdo con definir su cultura como una cultura consolidada de toma de decisiones en base a datos y aprendizaje por prueba-error, frente el 26% que no está ni de acuerdo ni en desacuerdo, y el 29% que está nada o poco de acuerdo (ver gráfico 19).

Gráfico 19: Indique su grado de acuerdo: “En la compañía existe una cultura consolidada de toma de decisiones en base a datos y aprendizaje mediante prueba-error, valorando tanto los resultados negativos como los positivos (% de empresas)”



* [Accenture survey of 254 U.S. managers; ver “Most US Companies Say Business Analytics Still Future Goal, Not Present Reality”; Accenture press release, 2008]

La mayoría de empresas que defienden su cultura como muy analítica estiman que el impacto de sus proyectos BDAA es alto o muy alto

Culturas más analíticas tienden a obtener mejores resultados

Resulta relevante destacar que cuánto más analítica es la cultura de una compañía, mayor es el éxito y el retorno que obtiene en sus proyectos. Las **empresas que definen su cultura como muy analítica**, con decisiones basadas en datos a partir de la monitorización de resultados y análisis avanzados, estiman de forma mayoritaria (78% de las respuestas) el **impacto de sus proyectos BDAA como alto o muy alto**.

Por el contrario, en las empresas con **culturas poco analíticas**, con decisiones más basadas en intuición o experiencia, poco hábito de monitorización de resultados y realización de análisis, **sólo en un 36% de los casos** los proyectos BDAA han tenido un impacto alto o muy alto (ver gráfico 22).

Gráfico 22: Comparativa de retorno de proyecto según cultura empresarial (% de empresas con retornos altos o muy altos en proyectos BDAA)

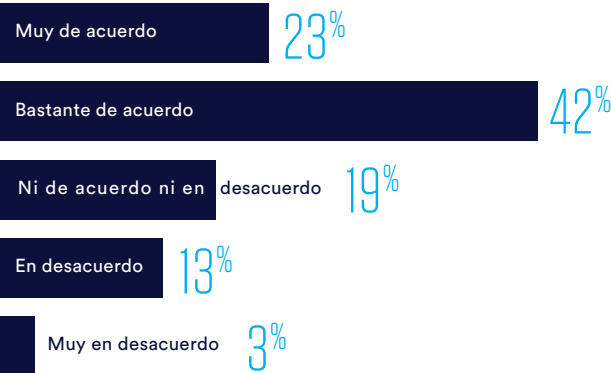
Éxito alto o transformador en proyectos Big Data



Nota: En base a las respuestas de aquellas empresas con más de 50 EJC's

Preguntadas por cómo intentan tomar decisiones y aprender a partir de los datos, el 65% de las compañías afirman que realizan una monitorización analítica de resultados, medición de indicadores de desempeño y creación de benchmarks - frente al 16% que está bastante o muy en desacuerdo (ver gráfico 20).

Gráfico 20: Indique su grado de acuerdo: “En la compañía existe una cultura consolidada de reporting de Key Performance Indicators (KPIs) y resultados, benchmarks internos y externos, con fuentes de datos y metodologías claras” (% empresas)



No obstante, sólo una de cada tres los realiza habitualmente (ver gráfico 21).

Gráfico 21: Indique su grado de acuerdo: “En la compañía existe una cultura consolidada de realizar análisis estadísticos multidimensionales y visualizaciones avanzadas de datos” (% de empresas)



Un talento escaso y difícil de obtener

Un elemento clave para la construcción y vertebración de esta cultura es disponer de personas con las capacidades necesarias: pensamiento analítico y cuantitativo, conocimiento técnico, creatividad, orientación a negocio y capacidad de trabajo en equipo.

Dentro de los perfiles de BDAA, destacan dos fundamentales: **data scientists**, encargados de la realización de análisis avanzados y modelados; y los **data managers**, responsables de aportar visión de negocio para orientar y liderar los análisis de los data scientists, y de extraer recomendaciones para la toma de decisiones.

Desgraciadamente, en España el talento analítico es todavía muy escaso y difícil de obtener, tanto a nivel de data scientist como de data manager. Las compañías cuentan con **equipos medios de 5 a 20 personas** (gráfico 23), según si son empresas pequeñas y medianas (de hasta 200M€ de facturación) o grandes, y **valoran en general como bastante o muy difícil encontrar estos perfiles**. En media, puntúan con 7,7 la dificultad de encontrar data scientists y 7,6 la de encontrar data managers, siendo 1 muy fácil y 10 muy difícil (ver gráfico 24).

Gráfico 23: Número de data scientists y data managers FTEs según el tamaño de la empresa

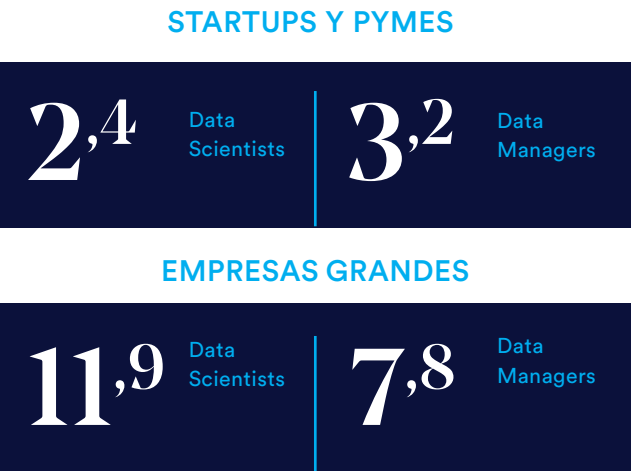


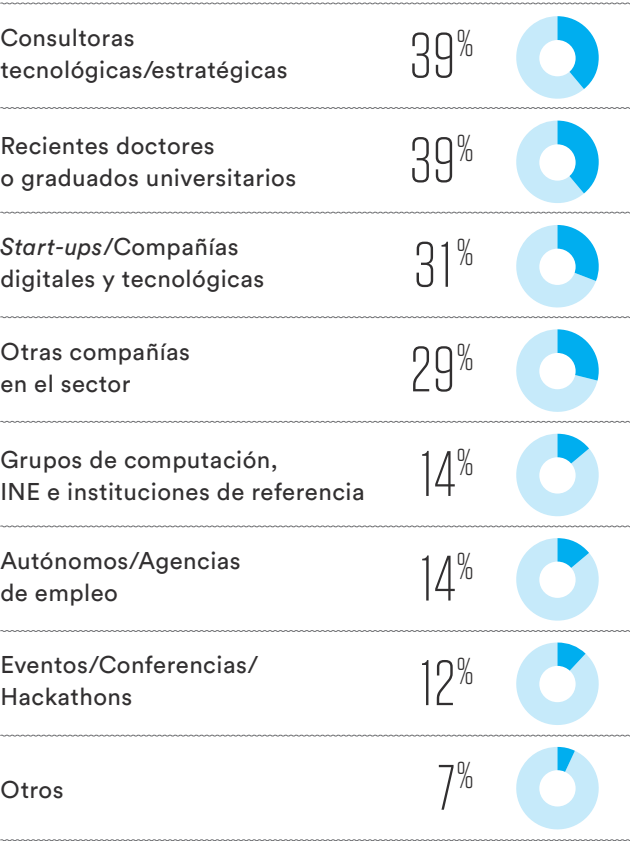
Gráfico 24: Puntúa de 0 (muy fácil) a 10 (muy difícil) de la dificultad de obtención de Data Scientists y Data Managers



Consultoras, start-ups y empresas digitales, principales fuentes de captación de talento

A la hora de adquirir estos perfiles, las compañías muestran una clara inclinación hacia la contratación de **profesionales con experiencia en consultoría estratégica y tecnológica o en start-ups y empresas digitales**. En el caso de los data scientists (ver gráfico 25), las universidades son también un foco de captación muy relevante, mientras que para los data managers (ver gráfico 26), dónde el factor experiencia y conocimiento de negocio es más relevante, gana importancia la capacitación de personal interno o la búsqueda en otras empresas del sector.

Gráfico 25: ¿Dónde está explorando y encontrando su compañía perfiles Data Scientist para su contratación? (% votos)



Nota: En base a las 3 respuestas más relevantes seleccionadas por cada empresa

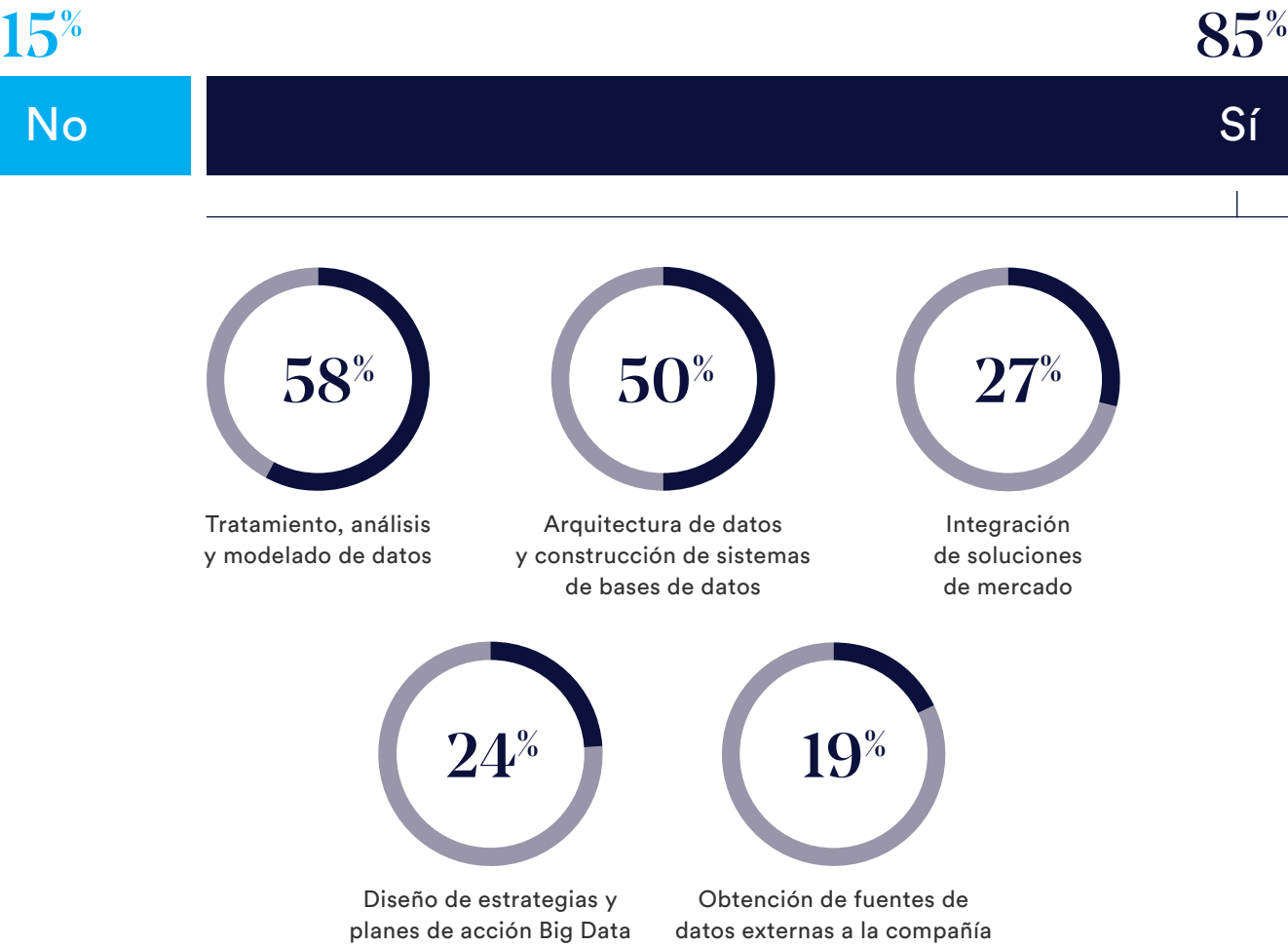
Gráfico 26: ¿Dónde está explorando y encontrando su compañía perfiles Data Manager para su contratación? (% votos)



Nota: En base a las 3 respuestas más relevantes seleccionadas por cada empresa

Finalmente, cabe señalar que la **gran mayoría de empresas cuenta también con el apoyo de colaboradores externos**. Más del 80% afirma haber desarrollado proyectos con proveedores o partners externos, en su gran mayoría expertos en análisis y modelado de datos, o en desarrollo de arquitectura y bases de datos (ver gráfico 27).

Gráfico 27: Colaboración con partners y proveedores externos BDAA durante el último año y área de colaboración (% de empresas)



¿Qué perfil tienen los data scientists?

Formación

- + Grado y/o Máster en Matemáticas, Estadística, Ingeniería, Física, Economía.
- + Formación complementaria en temas como Machine Learning, Data Science, Artificial Intelligence, Big Data o Data Mining, adquirida de forma reglada o en otros formatos como competiciones (por ejemplo en la plataforma Kaggle o la KDD Cup), hackathons, MOOCS...

Salario

- + Entre 25.000 y 40.000 euros en función del grado de experiencia.

Conocimientos específicos

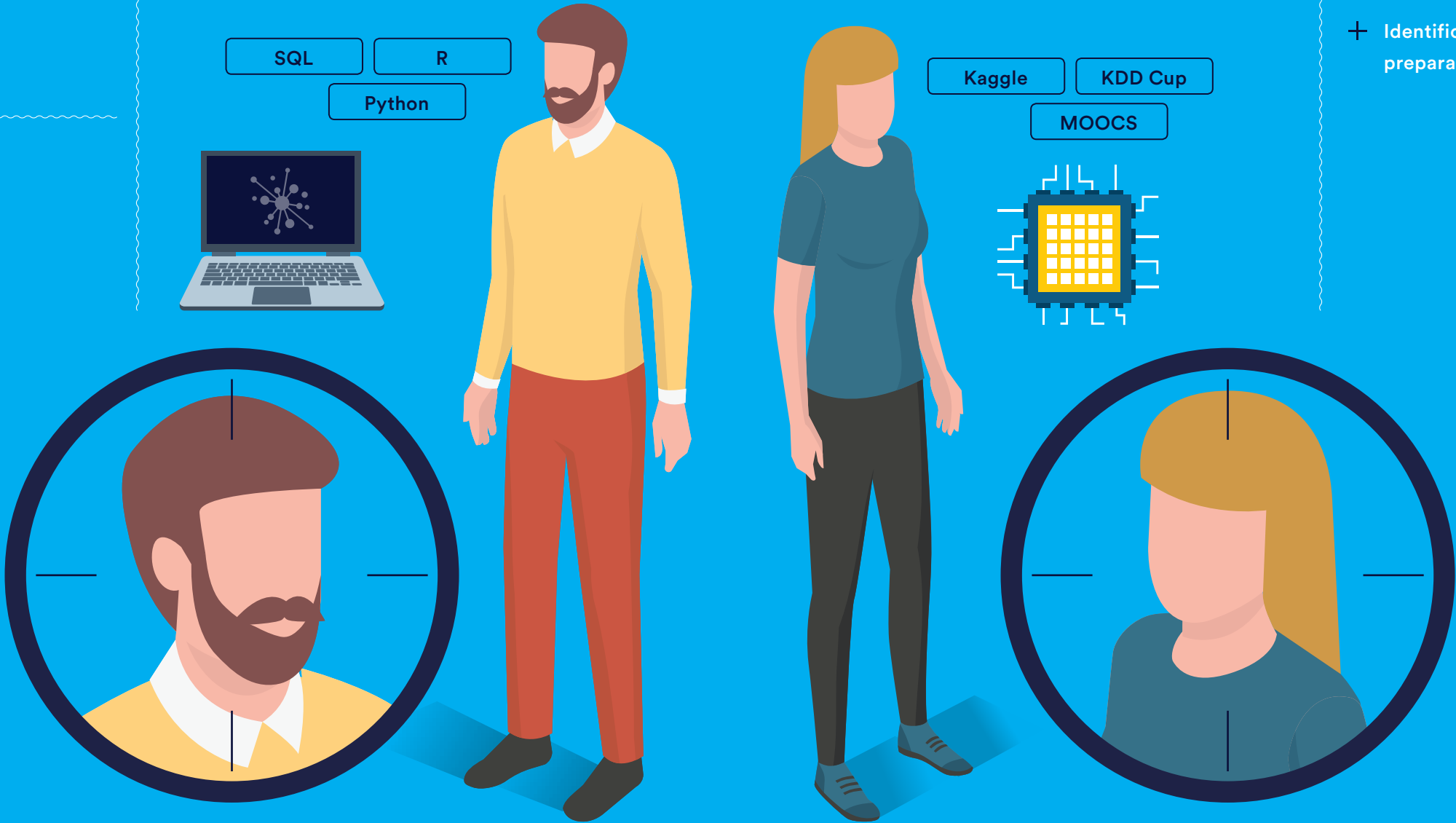
- + Diseño e implementación de modelos de regresión, de propensión, segmentación y simulación (p.e. regresiones logísticas, GBM, K-means...).
- + Conocimiento de lenguajes de programación y tratamiento de bases de datos como SQL, R o Python.

Habilidades

- + Pensamiento analítico y estructurado.
- + Creatividad para plantear y resolver problemas.
- + Capacidad de trabajo en equipo.
- + Motivación y voluntad de desarrollar visiones de negocio.
- + Capacidad de comunicar de forma clara y simple análisis complejos.

Funciones

- + Consolidación y tratamiento de diferentes fuentes de datos.
- + Realización, seguimiento y reporting de KPIs de negocio.
- + Análisis estadísticos multivariantes de las causas raíz de problemas de áreas de negocio.
- + Modelos de predicción (ventas, impago, baja de clientes...), segmentaciones y simulación de escenarios de negocio a futuro.
- + Identificación de oportunidades de mejora, preparación de tests y análisis de resultados.



Infraestructura de datos

Ninguna empresa puede ser realmente analítica sin disponer de datos de calidad ni de una buena infraestructura de almacenamiento y procesado. Este es, seguramente, el motivo principal que ha llevado a muchas empresas a asociar Big Data y Advanced Analytics a Sistemas, y por el que tantas veces se liga a conceptos como *data lake* o *cloud*.

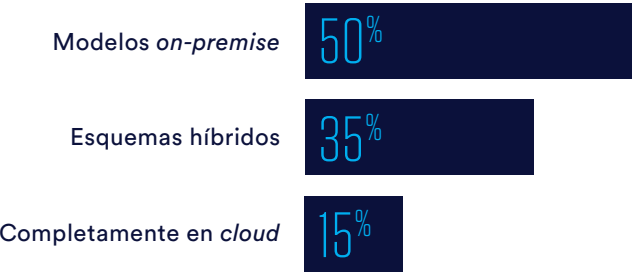
Predominio de arquitecturas *on-premise*, pero con fuerte presencia de *cloud*

En general, los esquemas de arquitectura se pueden clasificar en 3 grandes grupos: *on-premise*, que corresponden a aquellos donde las bases de datos y las herramientas para su explotación residen en servidores internos de la compañía; *cloud*, donde por el contrario, residen en servidores de terceros a los que se accede online y cuya capacidad puede adaptarse a petición; e híbridos, donde conviven los dos tipos de estructuras, en función de la sensibilidad, origen o uso de los datos.

Las compañías españolas optan todavía, a día de hoy, por mantener **modelos *on-premise*, en un 50% de los casos**. El 50% restante se distribuye entre *cloud* y esquemas híbridos, siendo que en estos últimos predominan los que son principalmente *on-premise*.

Cabe destacar que dentro de los modelos *on-premise*, la inmensa mayoría, tres de cada cuatro emplean modelos tradicionales de Business Intelligence, mientras que el 25% restante opta por arquitecturas propiamente Big Data como Spark o Hadoop.

Gráfico 28: Arquitecturas más usadas por las compañías para el almacenamiento de datos (% de empresas)



Nota: Los modelos *on-premise* se distribuyen entre un 69% de modelos tradicionales Business Intelligence y 31% arquitecturas Big Data. Los modelos híbridos se dividen en un 59% principalmente *on-premise* y un 41% principalmente en *cloud*

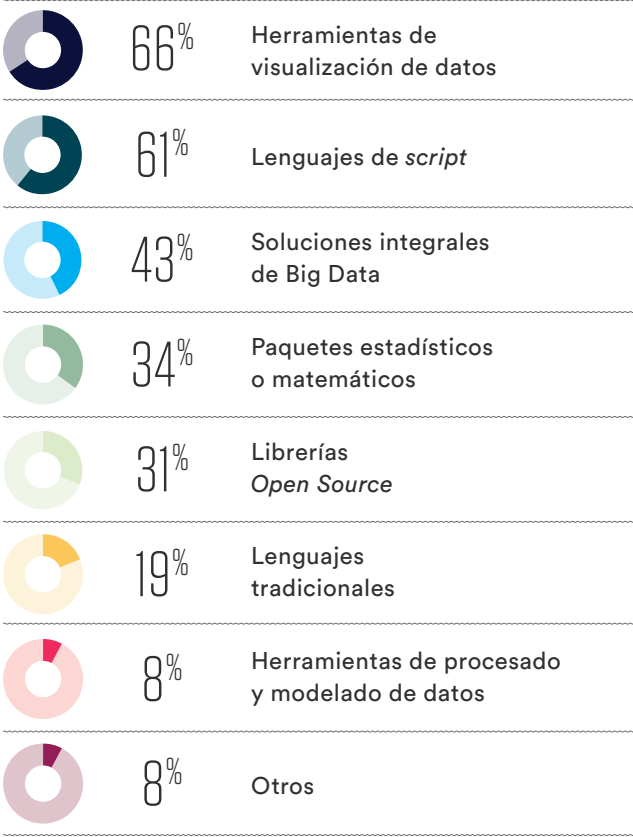
Las compañías españolas optan todavía por mantener modelos *on-premise*, en un 50% de los casos

Programación para el análisis, herramientas para la visualización

Para el tratamiento, **análisis y modelado**, el 60% de las compañías se decantan como opción preferida por **lenguajes de *script***, como R o Python. También son de uso extendido las soluciones integrales, como Azure o Big Data Oracle, y los paquetes estadísticos, como SAS o SPSS.

Respecto a la **exploración y visualización**, las **herramientas de fácil uso** e integración, como por ejemplo QlickView y Tableau, son utilizadas por **dos de cada tres compañías**.

Gráfico 29: Herramientas y soluciones utilizadas para la implantación de iniciativas Big Data en las empresas (% de empresas)



Los datos externos, todavía un recurso sin explotar para la mayoría de empresas

Una vez situados en cómo almacenan y explotan los datos las compañías, queda preguntarse lo más fundamental: cuáles usan y de dónde los obtienen.

Dentro de la gran variedad de fuentes disponibles, la **mayoría de empresas basa sus análisis y modelos Big Data en datos internos**. Así, casi el 90% declara usar datos transaccionales y de uso de productos y servicios, el 70% datos de clientes o prospectos y alrededor del 40% datos geo-localizados y de navegación en web y app.

Gráfico 30: ¿Qué fuentes de datos se consideran principalmente en las iniciativas BDAA?



La mayoría de empresas basa sus análisis y modelos Big Data en datos internos

Por el contrario, **menos de la mitad utiliza datos externos**, como *open-data*, información de administraciones y organismos públicos (ej: BOE o catastro) o datos detallados de competidores (ej.: *scrapping* web de precios). Fuentes más novedosas y desestructuradas, como redes sociales, imágenes y vídeos, son claramente la asignatura pendiente: menos del 25% de empresas las incorpora a día de hoy en sus proyectos.



Estado de madurez y prioridades a futuro

Las diferentes dimensiones analizadas ofrecen una foto actual del estado de madurez en analítica y Big Data de las compañías españolas.

Se han identificado cinco grandes niveles de madurez, que se resumen en la siguiente tabla.

Madurez Big Data

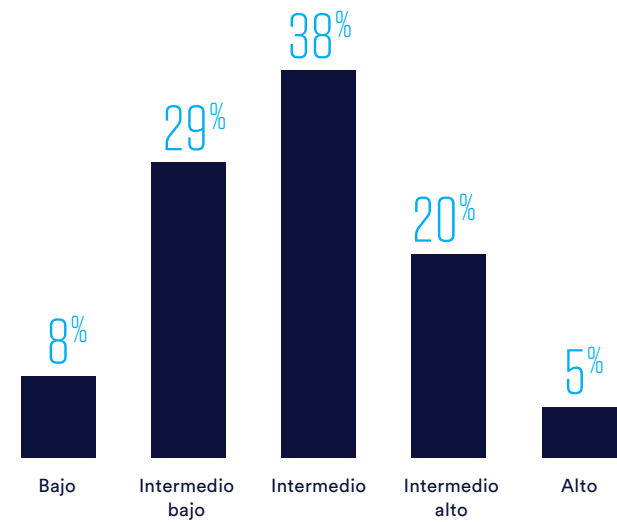
	BAJA	INTERMEDIA BAJA	INTERMEDIA	INTERMEDIA ALTA	ALTA
 Visión y modelo organizativo	<ul style="list-style-type: none"> - Sin liderazgo específico - Sin modelo organizativo definido 	<ul style="list-style-type: none"> - Liderazgo ejecutivo poco consolidado (<1 año, nivel n-3, n-4) centrado en consolidación de datos y reporting - Modelo organizativo no consolidado, en general, descentralizado o unidad especialista 	<ul style="list-style-type: none"> - Liderazgo ejecutivo en consolidación, con 1 o 2 años de antigüedad, en rango n-2 a n-3, centrado en modelado y apoyo a algunas áreas - Modelo organizativo poco consolidado, de cualquier tipo 	<ul style="list-style-type: none"> - Liderazgo ejecutivo consolidado con 3 o más años de antigüedad, y en rango n-1 o n-2, que promueve iniciativas de mejora, proyectos de innovación y generación de nuevas líneas de negocio - Modelo organizativo bastante consolidado 	<ul style="list-style-type: none"> - Liderazgo ejecutivo consolidado, en general, con 3 o más años de antigüedad, y en rango n-1 o n-2, capaz de influir decisivamente en la estrategia y transformación de la compañía - Modelo organizativo consolidado
 Inversión y creación de valor	<ul style="list-style-type: none"> - Proyectos e iniciativas todavía en discusión 	<ul style="list-style-type: none"> - Primeros pilotos en curso, con alcance reducido centrado en alguna unidad o área de negocio específica 	<ul style="list-style-type: none"> - Proyectos en curso, transversales a unas pocas áreas de negocio, con primeros resultados preliminares 	<ul style="list-style-type: none"> - Proyectos ya consolidados, con impactos tangibles 	<ul style="list-style-type: none"> - Múltiples proyectos consolidados y sistemas en producción automática y en tiempo-real, integrados en los procesos de la compañía
 Cultura y talento	<ul style="list-style-type: none"> - Sin perfiles analíticos en la compañía - Cultura basada en experiencia e intuición, y poco en medición de resultados 	<ul style="list-style-type: none"> - Con algún data scientist en la compañía - Análisis y visualizaciones básicas 	<ul style="list-style-type: none"> - Pequeño equipo de BDAA con data managers - Análisis básicos, reporting y <i>benchmarks</i>, y algún análisis o modelo multivariante 	<ul style="list-style-type: none"> - Equipo o equipos de analítica con presencia o dando soporte a varias líneas de negocio y participando en la toma regular y activa de decisiones - Análisis multivariantes avanzados y algunos modelos predictivos 	<ul style="list-style-type: none"> - Equipo/s de analítica totalmente consolidado/s y extendido/s en la compañía - Realización de análisis multivariantes, modelos predictivos, simulaciones estadísticas y visualizaciones avanzadas para la toma de decisiones
 Infraestructura	<ul style="list-style-type: none"> - Modelos <i>on-premise</i> - Uso de pocas fuentes de datos, esencialmente internas (transaccional y de cliente) 	<ul style="list-style-type: none"> - Arquitectura principalmente <i>on-premise</i> - Uso de diferentes fuentes de datos, esencialmente internos - Uso exclusivo de paquetes estadísticos y herramientas de procesado para la exploración de datos 	<ul style="list-style-type: none"> - Arquitecturas híbridas, mayormente <i>on-premise</i> - Uso de datos internos y alguna fuente de datos externa - Explotación con diferentes paquetes estadísticos y soluciones integrales Big Data, y con lenguajes de programación 	<ul style="list-style-type: none"> - Arquitecturas híbridas, con mayor presencia cloud - Uso habitual de algunas fuentes de datos externas - Explotación con lenguajes tradicionales y/o de <i>script</i> como apoyo de las herramientas BD 	<ul style="list-style-type: none"> - Arquitecturas híbridas con mayor peso en cloud, o completamente cloud - Explotación de muchas fuentes de datos tanto internas como externas - Uso completo de lenguajes tradicionales y/o de <i>script</i> y otras herramientas según tipo de análisis y modelo

Nivel de madurez entre intermedio-bajo e intermedio

Como se ha ido observando a lo largo del informe, la mayoría de empresas han iniciado ya procesos de transformación analítica y Big Data. Por este motivo, sólo un 8% se encuentra todavía en un nivel de adopción bajo, es decir, sin figura de liderazgo, sin prácticamente ningún perfil dedicado, con una cultura poco basada en toma de decisiones mediante datos, y con sistemas tecnológicos clásicos y pocas fuentes de datos explotadas.

No obstante, aunque sean pocas las que no han empezado, son todavía menos las que están ya muy consolidadas (~5% tienen una madurez alta), y de forma general las empresas en España se encuentran en un estadio intermedio-bajo o intermedio.

Gráfico 31: Empresas según score (% de empresas)



Prioridades de futuro

Conocido el estado actual, el último punto por tratar es cómo prevén evolucionar las compañías en el futuro más inmediato, en los próximos 3 años.

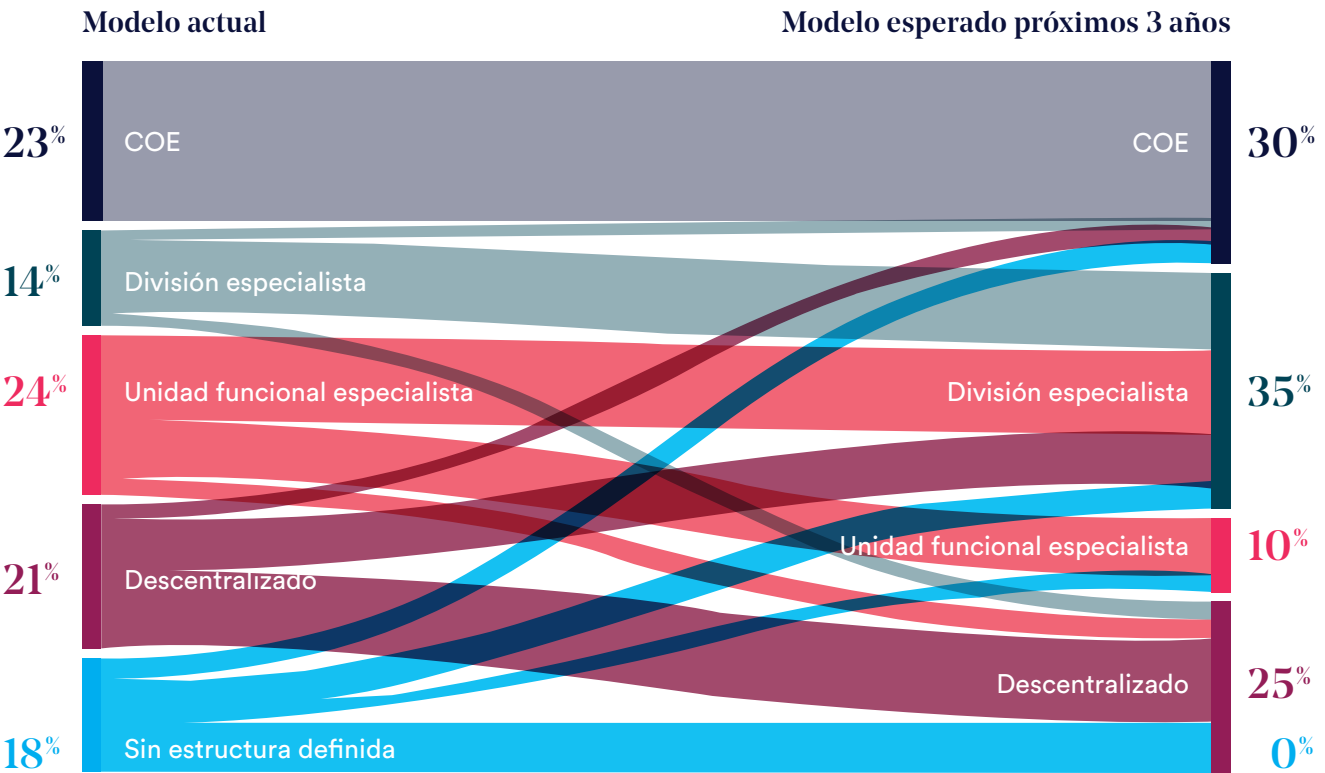
Las empresas sitúan tres elementos como prioridades clave: la consolidación (o creación) de liderazgos y modelos organizativos, la consecución de proyectos y el desarrollo de sistemas automatizados, y el refuerzo de equipos e implantación de culturas fuertemente analíticas.

Hacia la plena adopción y consolidación de modelos organizativos

En primer lugar, a corto plazo las empresas prevén consolidar los liderazgos y modelos organizativos de sus unidades BDAA.

Por un lado, aquellas que no disponen en la actualidad de capacidades o estructura específica, prevén desarrollarlas, en general y de forma inicial en un modelo descentralizado. Por otro lado, entre las organizaciones que ya disponen de dichas unidades, prevén evolucionarlas principalmente hacia Centros de Excelencia y Divisiones Especialistas.

Gráfico 32: Evolución del modelo organizativo

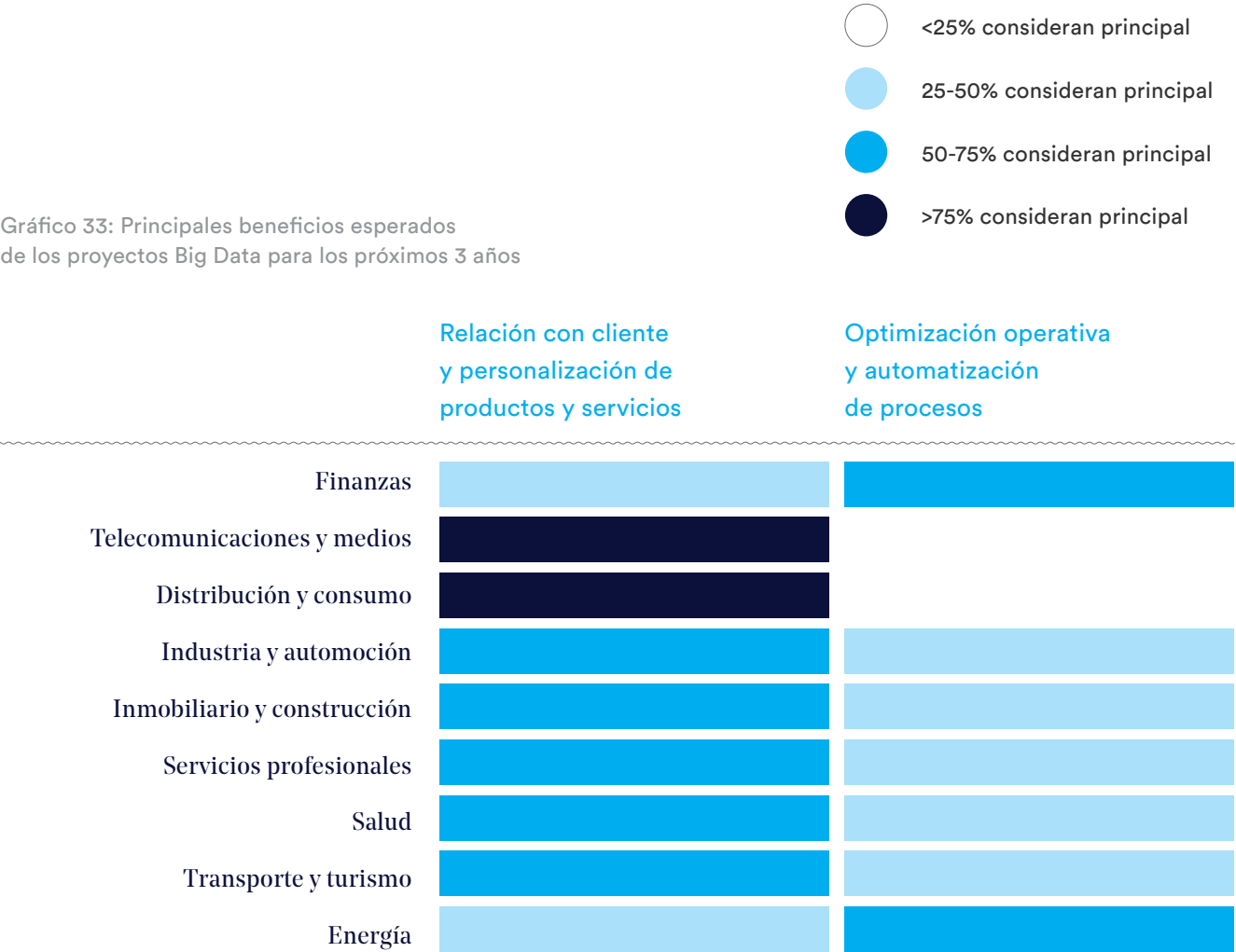


Sistemas automatizados con foco en el cliente

En segundo lugar, y en consonancia con la evolución organizativa, todas las empresas prevén estar lanzando proyectos Big Data, y una mayoría de ellas, el 55%, tener **sistemas en producción automática integrados** en su operativa real.

Estos proyectos y sistemas seguirán capturando beneficios en el terreno de la optimización de procesos y las mejoras operacionales, pero se centrarán principalmente en la **mejora de experiencia, comunicación y relación con cliente, y la personalización de productos y servicios.**

Gráfico 33: Principales beneficios esperados de los proyectos Big Data para los próximos 3 años

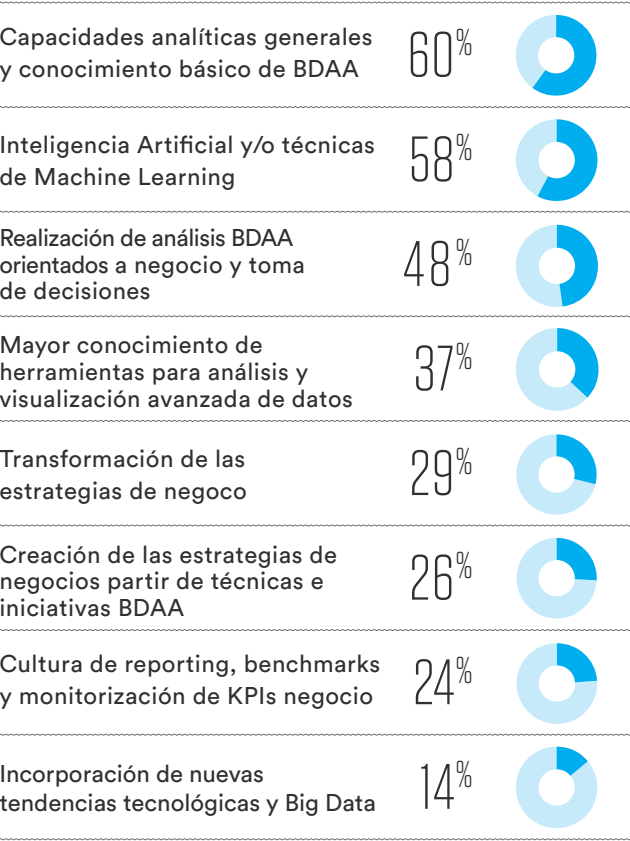


Avance hacia culturas más analíticas y fuerte crecimiento de los perfiles analíticos

Finalmente, las empresas en España esperan **avanzar rápidamente hacia culturas mucho más analíticas**, extendiendo las capacidades analíticas y el conocimiento de BDAA en la organización, incorporando técnicas más avanzadas (*Machine Learning*, inteligencia artificial) y enfocando sus análisis a la mejora de toma de decisiones de negocio. (ver gráfico 34).

Además, prevén aumentar de forma muy significativa el número de perfiles analíticos, **multiplicando por 3 el número de data scientists y por 2,5 el de data managers.**

Gráfico 34: Elementos culturales y capacidades analíticas a consolidar en las organizaciones durante los próximos 3 años (% de empresas)

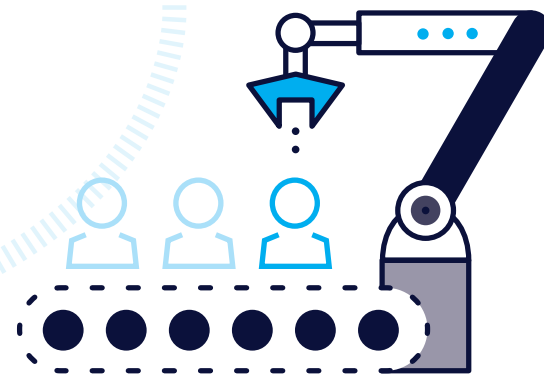


Nota: En base a las 3 respuestas más relevantes seleccionadas por cada empresa

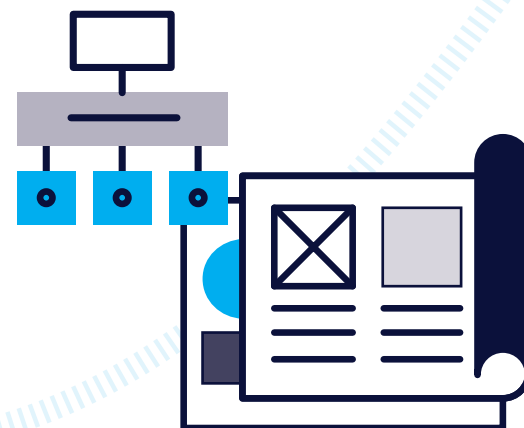


¿Qué pasos siguieron las compañías que valoran más positivamente sus iniciativas BDAA?

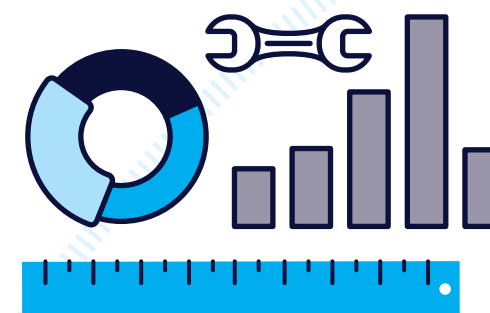
1. Identificar a una persona de **mando intermedio** en la compañía con **mentalidad analítica y capacidad de entender e interpretar modelos**, que pueda liderar a nivel operativo primeros pilotos y educar a otras personas de la compañía.



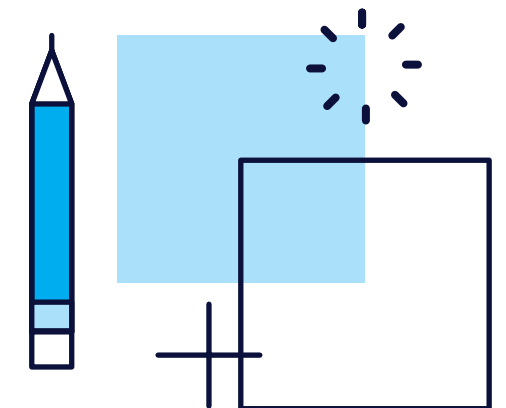
2. Identificar **prioridades de negocio** en la compañía donde se cumpla alguna de las siguientes casuísticas: (i) **no se hayan utilizado técnicas de modelado** avanzadas; (ii) se disponga de **datos diferenciales respecto a competidores**, sin explotar; (iii) existan **nuevas fuentes de datos internas o externas** que la compañía pueda incorporar a corto plazo.



3. Definir un **primer proyecto de alcance reducido**, de 8 a 12 semanas, sobre alguna de las prioridades identificadas y que implique de forma directa a una sola área (o a unas pocas).



4. Centrar los **objetivos del proyecto** en la **generación de primeros insights y quick-wins**, es decir, pequeñas transformaciones y optimizaciones **que puedan testearse a corto plazo y generar resultados** para convencer al resto de la organización.



5. Empezar con **muestras parciales de los datos** cuando no sea posible obtener de forma rápida y fácil todos los que se deberían tener para poder lanzar la iniciativa de forma completa (ej.: empezar con una muestra de clientes de una zona, considerar sólo unos determinados productos, un segmento concreto, etc.)



6. Focalizar los **esfuerzos iniciales** - tanto de inversión como de recursos - de la transformación Big Data en realizar **primeros pilotos más** que en construir grandes arquitecturas de almacenamiento y explotación de datos.

Casos de uso

_Asepeyo

_Cinesa

_Fintonic

_Fujitsu

_Mutua Madrileña

_Netquest

_Zurich Seguros

Asepeyo



Big Data, una herramienta clave para mejorar la toma de decisiones

Asepeyo es una Mutua de Accidentes de Trabajo colaboradora con la Seguridad Social con más de 280.000 empresas asociadas y 2,3 millones de trabajadores protegidos. El reto de la mutua es realizar una óptima gestión de los recursos públicos administrados a la vez que prestar un servicio de calidad a sus pacientes y un valor añadido diferencial, respetando el marco regulatorio, a sus empresas asociadas.

A finales de 2015, Asepeyo lanzó un **proyecto de Big Data** que tuvo como objetivo inicial la creación de una sola base de datos corporativa con más de 1.000 millones de registros que se ha ido enriqueciendo progresivamente con nuevas variables y nuevos diccionarios. Esta fuente de información permite diseñar cuadros de mando transversales y tener un mejor conocimiento de nuestros clientes.

En esta línea, durante 2017 se lanzó el **servicio ADD** (“Análisis Dinámico de Datos”) que consiste en una plataforma web de explotación de datos a disposición de nuestras empresas mutualistas. Uno de los mayores costes de las empresas es el relacionado con el absentismo de su plantilla. El ADD ofrece a la empresa un **instrumento de análisis personalizado** de todos sus indicadores relacionados con el absentismo de modo que les permite conocer sus puntos débiles y diseñar planes de acción para reducirlo. Es decir, les capacita para la **toma de decisiones** en base a al análisis detallado de la información. A mes de febrero 2018 esta aplicación tiene un total de 1.258 usuarios autorizados para más de 900 grandes empresas mutualistas de Asepeyo.

Constantino Perea, director de relaciones externas de la mutua, explica que “a diferencia de los informes estáticos cargados de números, el ADD es una **herramienta interactiva y visual** que faculta a la dirección de RRHH de la empresa para profundizar en el origen de su absentismo. Ello les permite establecer políticas de empresa saludable que incidan directamente en la salud

de sus trabajadores”. El comportamiento de los índices de las empresas que implantan estos modelos mejora un 5% los índices de sus respectivos sectores de actividad. El total de sesiones de clientes de la herramienta en 2017 fue de 2.226 y la duración media de cada sesión se situó en 42 minutos.

“Otro ámbito relacionado con el Big Data es el análisis de la experiencia de navegación de los clientes en la web, las Apps y la Oficina Virtual. Recogemos múltiple información por perfil y funcionalidad de modo que evolucionamos estas herramientas en un modelo de mejora continua. La potenciación de aquellas áreas más demandadas ha mejorado el uso de los canales digitales con cliente en más del 19%”, afirma Perea.

En el futuro más inmediato, Asepeyo está trabajando en aprovechar el Big Data para desarrollar modelos predictivos de duración de la baja. Ello permitirá identificar desde el mismo momento en que se produce la baja aquellos casos que requieren un mejor estudio para establecer los medios que permitan acortar el periodo de absentismo. “El conocer *a priori* los casos que se desviarán de la media nos permite adelantar la implantación de técnicas y soluciones que tendrán sus efectos optimizando los recursos dedicados a su seguimiento”.

Cinesa



Generando propuestas de valor mediante el análisis de datos

Cinesa, exhibidor de cine líder en España y perteneciente a Odeon Cinemas Group y a AMC (líderes también a nivel europeo y mundial), proyecta más de 500 películas cada año y recibe más de 21 millones de visitas.

Su **programa de fidelización**, CinesaCard, es uno de los mayores de España con más de 1.5 millones de miembros y, como todos los planes de fidelización, se enfrenta al desafío constante de generar un **valor diferencial para el cliente** cada año. Según Belén Peñalver, responsable del programa, “las exhibidoras de cine, en el contexto de mercado actual, nos diferenciamos por ofrecer al cliente una experiencia de cine diferente a la que pueda tener en casa, tanto en tipo de contenidos, como en tecnología de pantalla y sonido y, como no, la oferta de productos de bar para acompañar una experiencia de cine completa”.

Desde 2016, Cinesa empezó un proceso de transformación en la relación con los miembros de CinesaCard. La clave de la transformación: conocer mejor a los clientes mediante el **análisis de datos** para construir **propuestas personalizadas** y ofrecer a cada cliente los beneficios que mejor encajen con sus preferencias. Andrés Ripoll, responsable de Customer Intelligence de la compañía, explica que “la aplicación de técnicas Big Data nos ha permitido dar un salto muy relevante en nuestro conocimiento y trato con el cliente. Podemos mejorar la selección de nuestras carteleras, y recomendar a cada cliente los estrenos que más gustarán, mejorando la eficacia de nuestras comunicaciones”.

Además de la selección y recomendación de contenidos, Cinesa ha puesto las técnicas Big Data a trabajar para mejorar los beneficios de CinesaCard y ofrecer ofertas y **propuestas de valor segmentadas**. Peñalver señala que “gracias a los datos aportados por la Cinesacard, perfilamos a los clientes según lo que consumen, lo que les motiva y lo que están dispuestos a pagar.”

Peñalver destaca dos ejemplos muy concretos de adaptación de la propuesta de valor a perfiles identificados mediante análisis de datos. Por un lado, para “las familias que tienen un presupuesto más apretado, nos interesa que se beneficien de buenos precios y promociones y que puedan visitarnos con mayor frecuencia”. Por otro lado, “sabemos que los *seniors* tienen más tiempo, pero no pueden consumir productos de bar. Por ello, ofrecemos formatos de ópera y otro contenido a un precio de entrada superior, pero de valor muy diferencial para este segmento”.

En el futuro más inmediato, Cinesa continuará centrando sus esfuerzos Big Data en desarrollar propuestas de contenido y bar, y ofertas más personalizadas para sus clientes. Como apunta Ripoll: “a medio y largo plazo, estamos seguros de que los datos serán completamente disruptivos. No sólo en la relación con cliente, sino también en lo que entendemos por cine. ¿Podremos adaptar al gusto de cada persona toda la experiencia desde que piensa en ir al cine hasta que abandona las instalaciones? ¿Empezarán a pesar significativamente otro tipo de contenidos y actividades? ¿Qué papel jugará la realidad virtual en las películas del futuro?”.

Fintonic



Modelos predictivos al servicio de los usuarios

Fintonic nació en 2011 para **ayudar a la gente a mejorar sus finanzas**, facilitando el control de las cuentas y el acceso a productos financieros. Actualmente cuenta con más de **400.000 usuarios activos**, que usan sus servicios gratuitos disponibles en versiones app y web. Además de España, está también disponible en **Chile y próximamente en México**.

Fintonic permite a sus usuarios agregar de forma anónima sus posiciones financieras en más de 80 entidades, categorizando sus movimientos diariamente y **generando alertas e informes personalizados** (por ejemplo, “tu cuenta está cerca de llegar a descubierto”, “has gastado 300 € en restaurantes este mes”). Actualmente, 750 millones de transacciones están almacenadas y disponibles para su consulta por los usuarios. Esta ingente cantidad de información implica un **uso intensivo de soluciones de procesamiento** (Big Data) y **análisis de datos** (Advanced Analytics).

El uso de estas tecnologías tiene como objetivo principal **ofrecer mejores funcionalidades a los usuarios**, facilitando su gestión diaria de las finanzas de forma más sencilla y transparente. Específicamente, Fintonic ha desarrollado tres **modelos predictivos** cuya complejidad y desempeño se sitúan entre lo más avanzado del sector: la categorización de transacciones, el perfilado de riesgo de crédito y la recomendación de productos.

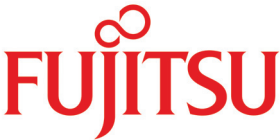
El **modelo de categorización** permite identificar de forma automática la categoría de gasto a la que pertenece cada movimiento en cuenta o tarjeta (ej. Restaurante, Farmacia, Supermercado, etc.). Se trata de un modelo altamente complejo dada la naturaleza del problema: un 60% de operaciones son relativamente fáciles de identificar con un análisis de palabras clave (ej. Nómina) mientras que el 40% restante es extremadamente difícil debido a un efecto de “*long tail*” (hay +700.000 comercios identificados). El uso de

algoritmos específicos de identificación de patrones junto a un proceso de entrenamiento constante de los modelos ha permitido optimizar el proceso, pasando de una precisión del 72% a un 92% en 6 meses.

El **perfilado de riesgo de crédito**, FinScore, es un producto único en España que ofrece a los usuarios una valoración transparente de su perfil crediticio, es decir, conocer “cómo le ven los bancos”. El modelo estadístico que soporta el FinScore emplea más de 130 variables sintéticas y algoritmos de Machine Learning combinados (por ejemplo: regresiones logísticas y modelos de *boosting*) que permiten inferir el riesgo potencial que tiene un usuario de no poder hacer frente a deudas financieras en el futuro. El modelo ha sido auditado por varias entidades financieras como base de su proceso de concesión de préstamos.

Los **modelos de propensión al uso de productos** son clave para ofrecer una experiencia atractiva para el usuario, facilitándole el acceso a productos que son realmente de su interés. No sólo permiten identificar a aquellos usuarios que están potencialmente interesando en un producto (por ejemplo: contratación de seguros) e incluso en la gama dentro del producto, sino que además permiten dirigir las comunicaciones, ofertas comerciales y descuentos únicamente hacia aquellos usuarios con mayor interés por los mismos.

En definitiva, Fintonic, como start-up financiera de nueva generación, pone las **metodologías analíticas más avanzadas disponibles en la actualidad al servicio de los usuarios**, desarrollando modelos que mejoran tanto la experiencia de uso (categorización, propensión) como la transparencia (FinScore).



Sistema de inteligencia para la gestión de pliegos y solicitud de propuestas (RPF)

Fujitsu Global es uno de los proveedores IT más importantes del mundo, con presencia en más de 70 países. El mercado principal de la empresa fuera de Japón es Europa, Oriente Medio, India y África (EMEIA), por lo cual Fujitsu Global dedica mucho tiempo y recursos a la valoración y redacción de respuestas a pliegos y RFP de sus clientes.

Dentro la estrategia global de la empresa para mejorar su eficiencia operativa y ganar en competitividad en el mercado, la inteligencia artificial y el *big data* juegan un papel esencial como palanca de transformación gracias a la automatización de los procesos y a la inteligencia aportada por el uso de técnicas de analítica avanzada.

Fujitsu Global ha construido una solución a partir de las capacidades que le proporciona su *framework Sholark* (ver imagen), que ya está siendo utilizando para la industrialización de varios supuestos de uso de inteligencia semántica.

Sholark es una solución modular, basada en tecnologías del *big data* de nueva generación y en un motor semántico, en el cual se utilizan técnicas de Procesamiento de Lenguaje Natural (NLP) y aprendizaje automático (*machine learning*), junto con una base de datos de grafos, lo que posibilita el tratamiento y el análisis avanzado de información no estructurada, especialmente texto. Además, la plataforma dispone de una interfaz web que abstraer a los usuarios de la necesidad de conocer este tipo de técnicas y tecnologías.

En este sentido, el sistema avanzado para la gestión de pliegos y RFP dispone de las siguientes funcionalidades principales:

- **Módulo de reconocimiento de entidades.** Sirve para identificar la información básica de los documentos, de modo que se puedan relacionar los requerimientos de los diferentes pliegos con las capacidades y los recursos de la organización.

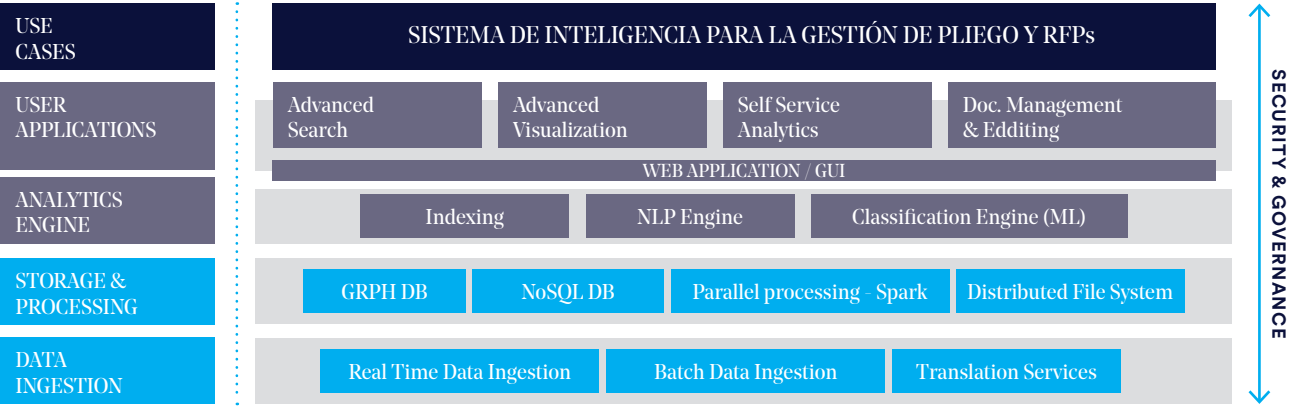
- **Módulo de resúmenes automáticos.** Ayuda a comprender la documentación y a realizar una primera valoración de la misma.
- **Módulo de clasificación.** Clasifica de forma automática, con la posibilidad de corrección, los pliegos en distintas categorías (tipología, ámbito, presupuesto, etc.) y les asigna un umbral de adecuación y un riesgo o un potencial (relacionado con el *portfolio* y las capacidades de la empresa, el historial de éxitos, etc.).
- **Módulo de auditoría y análisis** (visualización avanzada) para el control del ciclo de preventa.
- **Módulo de búsqueda avanzada.** Permite buscar documentos, conceptos, etc., relacionados en función de los filtros que decide el usuario.

Todo ello permite gestionar el ciclo de preventa de una forma mucho más eficaz, automatizando parte de las tareas que se realizan a día de hoy de forma manual. Además, gracias al uso de técnicas de aprendizaje automático, al usuario final puede corregir y mejorar partes del ciclo.

Gracias a la implementación de esta solución, Fujitsu Global está consiguiendo importantes beneficios. En primer lugar, reducir el tiempo de respuesta y los equipos necesarios para la gestión del ciclo de preventa. En segundo lugar, mejorar el control del ciclo *end-to-end* al basarse el proceso en datos y aumentar la ratio de éxito (más pliegos y RFP ganados). Disponer de una solución escalable y modular aplicable a otras áreas de la empresa para mejorar la eficiencia interna. Y, por último, digitalizar parte de los procesos de la empresa y dotarla de una herramienta sencilla, que permite a los usuarios actuales utilizarla con una mínima formación.

Durante el año 2017, solo en España se gestionaron más de 2.000 oportunidades de negocio, de las cuales cerca del 75 % tenían un importe superior a los 10.000 €. Para gestionarlas, se necesitaron más de 25 empleados a jornada completa, y la empresa dedicó unas 75.000 horas a darles respuesta o a descartarlas. En total, la ratio de éxito de estas oportunidades fue del 80 %.

Arquitectura funcional de Sholark



El Big Data, pieza clave en la digitalización

Mutua Madrileña es el 2º grupo asegurador de no vida en España, con más de 10 millones de asegurados y unos ingresos de 5.000 millones en primas.

Hace tres años el grupo situó la transformación digital y Big Data en el **centro de su plan estratégico**, creando para ello una dirección de Transformación Digital dentro del Comité Ejecutivo.

Dicha transformación está conllevando cambios muy relevantes en el sector asegurador, especialmente en el proceso de contratación. Por un lado, el canal on-line desplaza cada vez con más fuerza a los canales directos y a las oficinas en la búsqueda de información y contratación de pólizas. En 2017, las contrataciones iniciadas en la web representaron el 50% de las nuevas primas de Mutua Madrileña, mientras que en 2012 sólo eran del 30-40%.

Por otro lado, al ofrecer una mayor facilidad para realizar cotizaciones y compararlas, especialmente con los agregadores, el canal digital ha obligado a las compañías de seguros a **establecer precios con la mayor precisión posible** para poder seguir siendo competitivas.

Es por este motivo que, dentro de la dirección de Transformación Digital, Mutua Madrileña situó a un equipo de Big Data formado por 5 *data scientists*. En coordinación con los equipos de sistemas y actuariales – requisito indispensable para satisfacer los requerimientos regulatorios del sector-, el equipo de Big Data ha estado trabajando durante los últimos tres años en la obtención de nuevas fuentes de información y en su explotación para la transformación de la relación con los mutualistas, personalizando ofertas y productos, y la mejora de las campañas de captación on-line.

Una de las últimas iniciativas lideradas por el equipo de Big Data ha sido la **creación de un scoring digital**. Construido con más de 20 variables de comportamiento digital y socio-demográficas, este score permite perfilar a las visitas web en diferentes grupos para identificar los nichos de clientes más interesantes para la compañía y ofrecer tarifas personalizadas en tiempo real, durante la cotización.

La personalización y optimización de ofertas y tarifas a partir de los modelos Big Data ha mejorado significativamente la eficacia y eficiencia del canal digital. En el caso de pólizas de vida riesgo, por ejemplo, el crecimiento experimentado el primer año superó el 100%, con casi un 20% de mejora en la prima media. Además, en los diferentes ramos, los scorings digitales han permitido reducir en más de un 25% los costes de captación de nuevas pólizas

Netquest



Netquest, datos de comportamiento digital

Netquest es la empresa líder en Iberoamérica en recolección de datos para fines de investigación de mercados a través de personas. Para ello cuenta con paneles online, comunidades de personas que aceptan compartir su opinión a través de encuestas a cambio de incentivos. Netquest cuenta con 2,2 millones de participantes en 21 países, que completan cerca de 24 millones de encuestas anuales. Más de 10.000 empresas en todo el mundo toman decisiones basadas en datos Netquest.

En 2015, Netquest decidió ofrecer un nuevo tipo de dato a los investigadores: información acerca del **comportamiento digital del consumidor**. Para ello, una parte de los participantes en encuestas aceptó instalar en sus dispositivos de navegación (PCs, *smartphones* y *tablets*) una aplicación que rastrea su actividad digital: páginas webs visitadas, términos de búsqueda, Apps usadas, etc. Actualmente, 50.000 personas de países como España, México o Estados Unidos confían a Netquest información acerca de qué hacen en Internet.

Los **datos de comportamiento digital** son de una naturaleza muy diferente a los datos de encuesta. Un usuario típico interacciona diariamente cerca de 250 páginas webs diarias a través de diferentes dispositivos. Una muestra de tan solo 1.000 personas genera más de 7,5 millones de registros mensuales de comportamiento. La recolección y tratamiento diario de toda esta información es extremadamente exigente, requiriendo la adopción de tecnologías y servicios aptos para manejar Big Data como *Amazon Web Services*, *Hadoop* o *Python*.

Toda esta información es tratada por investigadores de mercado que combinan lo que dicen los consumidores (encuestas) con lo que hacen online (comportamiento digital), logrando una **mirada integral del consumidor** y proporcionando valiosos *insights*, como por ejemplo, que desde finales de 2017 más del 20% de la población internauta visita regularmente supermercados online, o que usuarios con perfiles en LinkedIn realizan entre 3 y 4 veces más visitas y búsquedas relacionadas con productos financieros y tienen entorno a un 35% más renta que la media nacional.

En 2016 Netquest fue un paso más allá incorporando la medición de datos de comercio electrónico con el lanzamiento de Netrica. Para los fabricantes, Netrica permite, entre otros, identificar dónde se venden sus productos en internet, a qué precio, cuál es su market share o hacer *benchmarks* contra competidores. Para sitios de e-commerce, da respuesta cuestiones como ¿cuál es mi *market share* dentro de mi categoría?, ¿qué cantidad y tipo de productos venden mis competidores?, ¿qué conversión de ventas logran? o ¿dónde están captando a sus clientes?

Actualmente, los datos de comportamiento digital ya representan un 16% del negocio de Netquest, siendo la principal apuesta estratégica de la compañía para los próximos años.

Zurich Seguros



Detección de fraude con “ZFINDER SUITE”

Zurich se fundó en Suiza en 1872 como entidad reaseguradora del campo marino bajo el nombre «*Versicherungs-Verein*». Doce años más tarde, en 1884, abrió su primera oficina en Barcelona. Desde sus inicios, la vinculación con el mercado asegurador español se basó en la aportación por parte de la compañía de productos y técnicas de gestión novedosas. Actualmente Zurich España tiene cerca de 2.000 empleados que damos servicio a 2,4 millones de clientes, consolidados como líderes en solidez y solvencia.

A continuación, presentamos el proyecto que nos ha permitido desarrollar nuestro nuevo producto para la detección de fraude. Este proyecto es el resultado de la colaboración de tres equipos: IT, SIU (empresa) y *Advanced Analytics*.

ZFinder Suite consiste en dos herramientas complementarias: ZFinder y ZFinderVision. ZFinder es una herramienta de detección temprana de fraude en el ciclo de vida de la tramitación de siniestros que, en una primera fase, se ha implementado para el ramo de hogar. El mismo entra en funcionamiento cuando el siniestro se apertura, enviándose a la Unidad de Investigación Especial (UIE) en el momento en que se superan ciertos umbrales de probabilidad. ZfinderVisual es una herramienta de visualización de los datos que ayuda a la UIE durante todo el proceso de investigación de fraude. ZFinder usa la arquitectura Big Data desarrollada por el equipo local de IT en Zurich, mientras que ZfinderVisual utiliza un *plugin* de fuente abierta para la visualización de datos.

Anteriormente, la UIE operaba con *check lists* manuales y reglas basadas en la experiencia y la intuición. Sin embargo, de los casos sospechosos detectados de esta forma, solo el 48% de los mismos resultaban ser fraudulentos. En dicho momento, no existía ningún tipo de herramienta automática. Si bien actualmente en el mercado existen una variedad de opciones, el desempeño de estas herramientas no suele resultar del todo satisfactorio o demostrado y los costos tienden a ser muy elevados.

El objetivo es pues desarrollar un modelo predictivo que permita:

1. Incrementar el número hallado de casos fraudulentos.
2. Reducir la proporción de casos investigados que no son fraudulentos.

Adicionalmente, se ha considerado deseable una solución visual complementaria que permitiera apoyar a la UIE durante el proceso de investigación.

Enfoque: Esencialmente, la detección de fraude se ha modelizado como una predicción de patrones anómalos. En general, los problemas basados en anomalías se caracterizan por la presencia de un pequeño número de datos etiquetados y una gran masa de datos no revelados. Un ejemplo clásico suele ser la detección de fraude: hay casos de fraude identificados, pero los demás casos no investigados no se sabe si son fraudulentos o no. En nuestro caso, sobre una base de siniestros cerrados en los últimos dos años, solo un 0.8% de los datos pudo ser identificado como fraude efectivo. El restante del 99.2% casos pueden interpretarse como casos no revelados, dado que son casos que no fueron investigados en su momento. Para el análisis se han utilizado unas 430 millones de observaciones y más de 1,200 variables. En esta primera fase solo se han utilizado datos estructurados. Respecto a la predicción, se han aplicado técnicas avanzadas de aprendizaje automático. El modelo predictivo se ha desarrollado utilizando el lenguaje de fuente abierto Python.

Resultados: El modelo obtuvo inicialmente un *recall* del 96% y una precisión del 83%. Adicionalmente, se ha realizado un test *out-of-sample*, que ha arrojado un *recall* del 100%. Sin embargo, los resultados más interesantes han surgido de aquellos siniestros que en el test no se han identificado como fraudulentos, pero que el modelo ha predicho que sí lo eran. Para estos casos, el modelo ha identificado, además de los casos de fraude revelado, un doble de casos potenciales. Estos nuevos casos se han analizado posteriormente por parte del equipo de UIE y ha resultado que aproximadamente 90% de los mismos poseían cierta evidencia de fraude.

Impacto Esperado en los KPIs: Indicadores financieros - Aumentado el ahorro al haber aumentado el número de casos fraudulentos detectados; Eficiencia en los procesos - Incrementado la productividad debido a la disminución de la proporción de casos investigados que no son fraudulentos, la simplificación interna con la automatización y el uso de herramientas visuales; Satisfacción del cliente - Mejora en los tiempos de resolución de siniestros no fraudulentos.

Impacto del Big Data en la sociedad

El Big Data como fenómeno disruptivo para la sociedad del siglo XXI: singularidades y riesgos desde una perspectiva social.

Por David Murillo

¿Cómo podemos evaluar el conocimiento generado a partir de Big Data?

Por Carlos Teixidó Román & Ferran A. Mazaira Font

La intencionalidad como factor clave en la toma de decisiones basadas en Big Data

Por Marc Le Menestrel

Retos y oportunidades en la toma de decisiones algorítmicas basadas en datos: puntos claves para avanzar hacia el bien social

Por Nuria Oliver



El Big Data como fenómeno disruptivo para la sociedad del siglo XXI: singularidades y riesgos desde una perspectiva social

Por David Murillo,

Profesor Asociado, Departamento de Ciencias Sociales de ESADE

El auge del Big Data, definido como la obtención de grandes cantidades de datos y la extensión de las técnicas avanzadas que permiten su uso, conlleva un **fuerte impacto a nivel social** del que no siempre somos conscientes. Como cualquier fenómeno disruptivo podemos entenderlo como aquel cambio que no tiene marcha a atrás y del que podemos esperar con certeza un fuerte impacto transformador. Desde la perspectiva social, ¿tenemos motivos para estar preocupados?

Unos de los grandes cambios introducidos por la extensión de los algoritmos inteligentes en la llamada ola digital es su gran **capacidad de ser replicables y escalables**. Al igual que sucede con todas las innovaciones debemos estar atentos a los dilemas éticos que su uso generalizado va a conllevar. Un ejemplo conocido es la programación del algoritmo que se requiere para programar el coche inteligente. Con independencia del deseo de sus programadores, de una manera activa o implícita, este algoritmo deberá tomar decisiones según criterios morales, pero ¿vamos a seguir un patrón de conducta que, en caso de accidente, minimice el daño causado, el número de víctimas, la vida del conductor o la reputación de la marca? ¿Cómo vamos a medir este daño?

Durante los últimos años ya hemos recogido mucha información sobre el impacto social del Big Data para casos menos extremos que el anterior. Sobresale en primer lugar, la **desigualdad en el acceso a los datos**. Si analizamos las dinámicas de mercado generadas por las grandes plataformas digitales vemos que estamos entrando en un nuevo modelo de competencia donde no todos los agentes (públicos, privados, tercer sector) tendrán acceso a los datos de igual forma ni en similar magnitud. ¿Qué tipo de mercados estamos creando en un mundo donde, si seguimos la estela de lo que sucede en Estados Unidos, Facebook y Google concentran dos terceras partes de la publicidad digital y capturan el 99% del crecimiento de este mercado?

Estamos entrando en un nuevo modelo de competencia donde no todos los agentes (públicos, privados, tercer sector) tendrán acceso a los datos de igual forma ni en similar magnitud

¿Quién puede acceder a la información que generamos en el espacio digital y quién termina beneficiándose de ella? ¿En qué lugar deja a los usuarios y qué marco normativo se establece para proteger nuestra privacidad, la propiedad o la portabilidad de nuestros datos? La Ley Europea que entra en vigor en mayo de 2018 (EU GDPR) aborda algunos de estos elementos, aunque sabemos que la tecnología siempre va a desarrollarse a una velocidad superior a la capacidad de la ley de dar respuesta a los nuevos retos planteados. ¿Cómo debemos incorporar las expectativas sociales y los dilemas éticos en el análisis y manipulación de grandes cantidades de datos?

Hoy en día, con el código IP de un usuario de internet o con rastreadores de nuestra huella digital, una empresa de venta online como Amazon puede hacer una estimación de la capacidad de compra de sus clientes y definir estrategias de *pricing* personalizado que desbordan el marco legal existente además de redefinir en su aplicación práctica los marcos de competencia de una forma radical. ¿Sigue siendo el cliente el rey y el propósito último de una actividad empresarial cuando una empresa puede beneficiarse de los datos de sus clientes para personalizar precios en su contra? ¿Cómo este fenómeno va a verse aumentado y multiplicado con la extensión del Big Data?

Otro patrón de cambio, imperceptible para muchos pero importante de destacar, sería el **grado de confianza en los datos y la capacidad de revertir procesos de toma de decisiones automatizadas y generalizadas**. El Flash Crash de mayo 2010 implicó un desplome de los índices bursátiles de hasta 8 puntos por sistemas de *trading* automatizados. Fenómenos como el sucedido en los mercados de capitales van a ser cada vez más frecuentes. Cómo decía hace poco un directivo en el aula, realmente ¿la solución a problemas como este pasa por desconectar el ordenador? El famoso artículo de Chris Anderson (2008) en *Wired*, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, nos sitúa ante un panorama desolador desde el punto de vista humanístico. ¿Realmente no nos hace falta conocer el funcionamiento del mundo porque los datos, por ellos mismos, van a permitirnos llegar al conocimiento profundo de los fenómenos económicos y sociales? ¿Qué tipo de directivos y de ciudadanos estamos creando con la generalización de una fe absoluta en la capacidad de los datos de dar respuesta a los grandes desafíos del mundo?

Los sistemas de *predictive policing*, que se empiezan a extender, permiten anticipar las posibilidades de que surja el crimen en una zona determinada

Como las **premisas básicas** de los programadores influyen en el mundo es otro elemento de análisis ante el auge del uso generalizado del Big Data. Los sistemas de *predictive policing*, que se empiezan a extender, permiten anticipar las posibilidades de que surja el crimen en una zona determinada tal y como anticipaba la

película *Minority Report* (2002). El algoritmo por sí mismo permite mejorar la eficiencia de los recursos policiales y reducir el crimen en determinadas zonas urbanas. Menos conocido es la extensión de la estigmatización y el racismo por prácticas automatizadas que discriminan de facto a personas de color. **¿Somos conscientes de las dinámicas sociales que desencadenamos con la automatización de decisiones de este tipo?** Una solución para resolver este tipo de problema es la transparencia en los algoritmos, pero ¿cómo vamos a llegar a ella si el grueso del valor de una aplicación o una plataforma digital proviene precisamente de la capacidad predictiva de sus algoritmos que es su propiedad intelectual?

Por último, remarquemos la clara tendencia a la **hiperespecialización**, que nos lleva la revolución digital por ella misma. Una hiperespecialización que, en contraste, permite anticipar un incremento de la demanda de profesionales pluridisciplinarios, capaces de utilizar su experiencia como valor añadido para ir más allá de los silos, de los múltiples subsegmentos del conocimiento científico, y resituar las organizaciones en el centro del debate social y abiertos al diálogo con el mundo. Citemos para terminar (¿cómo no?) una frase de Steve Jobs: “*El futuro son la tecnología y las humanidades. La tecnología nos dará las pautas de cómo hacer las cosas, pero sin las humanidades no conoceremos las trampas que nos ponemos a nosotros mismos.*”

Con independencia de que sea cierta o no, es la combinación de las dos aproximaciones, científica y humanística, la única que nos dará un mundo mejor y organizaciones más sostenibles en el tiempo y responsables.

¿Cómo podemos evaluar el conocimiento generado a partir del Big Data?

Por Carlos Teixidó Román, MSc Filosofía de la Ciencia, Aix-Marseille Université
& Ferran A. Mazaira Font, colaborador académico en ESADE

Para conocer el mundo natural, **la ciencia moderna se sirve de la recolección y el análisis de datos**, así como de la elaboración de teorías y modelos matemáticos que expliquen los fenómenos observados y permitan predecir el curso de futuros experimentos. De esta forma, convertimos nuestras observaciones e intuiciones sobre la naturaleza en *conocimiento objetivo*.

Dado el **parecido que guardan los métodos científicos con las técnicas de análisis y creación de modelos Big Data** que utilizan las empresas, es razonable pensar que estas últimas generan a su vez conocimiento objetivo. Sin embargo, la evaluación de algunos ejemplos nos permitirá ver que esta asociación de ideas no es tan simple como parece y que más bien es necesario plantearse **¿qué tipo de conocimiento generan las prácticas actuales de Big Data en las empresas?**

Una de las premisas más extendidas en las empresas es que **a mayor cantidad de datos y modelos más complejos, más y mejor conocimiento del cliente y del mercado**. Esto hace que no sólo se valore positivamente la abundancia de datos para realizar análisis y producir modelos predictivos, sino que también se **desdén el papel nocivo que puede tener la acumulación de variables, los sesgos que pueden tener los datos o las dificultades que puede entrañar interpretar correctamente modelos muy complejos**.

Veamos tres casos ilustrativos de este fenómeno. Un primer ejemplo que podemos utilizar es el de Netflix. En 2006, la compañía organizó una competición para diseñar un sistema de recomendación más eficiente para sus usuarios. Los ganadores combinaron quinientos modelos predictivos distintos. Esta práctica, conocida como ensamblaje de modelos, no es rara en las empresas y se basa en agregar distintos modelos preexistentes que dan lugar a uno mayor y más complejo. El problema es que los resultados de esta técnica son a menudo tan

complejos que no se pueden interpretar por falta de parámetros, y además su incremento de complejidad no se traduce en una mejora notable de calidad predictiva. **En el caso de Netflix, la diferencia de calidad entre un modelo y la combinación de quinientos era menor del dos por ciento**.

Otro ejemplo relevante, por la notoriedad de la compañía, es el de **Wikipedia**. A pesar de disponer de más de 46 millones de entradas, la generación de contenido está completamente sesgada. **Sólo uno de cada diez artículos está escrito por mujeres**, y hay una sobre-ponderación hacia temas técnicos, occidentales y dominados por hombres. Por lo tanto, cualquier análisis o estudio que pretendiera entender sobre cómo construimos los ciudadanos de todo el mundo artículos enciclopédicos y qué temas nos interesan, basado en de entradas en Wikipedia, sería muy poco representativo.

Un último ejemplo interesante es el del mercado asegurador. Los modelos predictivos de baja de clientes han ganado mucha importancia en el sector, especialmente a raíz del incremento de competencia producido por la digitalización y el auge de la actividad comercializadora de la banca, en un contexto de tipos de interés negativos y de reducción de activos tóxicos. En la construcción de **modelos Big Data para la baja, es habitual encontrarse con que una de las variables que parece tener más potencial predictivo es la prima pagada del asegurado**. Cuanto mayor es la prima, más fidelizado está un cliente. Siendo así, podría parecer que subiendo las primas a los clientes que pagan menos, las compañías lograrían más fidelidad. Sin embargo, **no se trata de una relación causal**. No es que cuanto mayor es la prima menos probable es que un cliente abandone la aseguradora, sino clientes que dan más importancia al seguro y tienen menos probabilidad de cancelar sus pólizas, tienden a contratar mayores coberturas y pagar primas más altas.

Si bien los tres casos son distintos, todos muestran que **no siempre una mayor acumulación de datos y complejidad de modelos aporta un resultado positivo** a las empresas. En el primero, Netflix no obtiene una mejor comprensión de sus clientes ya que su modelo no es interpretable. En el segundo, por muy grande que sea el volumen de artículos de Wikipedia, no conseguimos superar el sesgo masculino y occidental. En el tercero, la falta de reflexión sobre el significado de los datos utilizados y la confusión entre correlación y causalidad podría conllevar a errores de ejecución con resultados contraproducentes para las compañías.

Así pues, si asimilar mayores datos y modelos más sofisticados a mayor conocimiento no es necesariamente cierto, **¿podemos encontrar criterios que ayuden a las empresas a desarrollar mejor conocimiento?**

En uno de los artículos más influyentes en epistemología del último siglo, «Objectivity, Value Judgment and Theory Choice», **Thomas Kuhn distingue cinco características** que deberían perseguir las teorías científicas: **exactitud** (encaje con los datos observados), **coherencia** (de la propia teoría, y con el resto de teorías consideradas como válidas), **alcance**, **simplicidad y utilidad** (capacidad de suscitar nuevas preguntas y relaciones desconocidas anteriormente). Todos ellos son criterios habituales para la validación y selección de modelos en ciencia, aunque el peso relativo de cada uno de ellos es una elección en el proceso científico.

Tanto en el desarrollo de una teoría científica como en el de cualquier evaluación racional de la realidad, pese a que unos criterios pesarán siempre por encima de otros, tenemos que ser conscientes de nuestras decisiones y de qué características privilegiamos por encima de otras. La exactitud o precisión predictiva puede ser el criterio más importante en la elaboración de nuestros

modelos. Pero casos como el de Netflix nos obligan a preguntarnos cómo debemos ponderar exactitud y simplicidad. Casos como el de Wikipedia nos recuerdan hasta qué punto deberíamos medir bien el alcance -si los datos son sesgados, el alcance es reducido- y casos como el de seguros, ponen de manifiesto la importancia de la coherencia.

Del mismo modo que la ética nos ayuda a definir los límites de lo que debe de ser el Big Data para la sociedad, la filosofía de la ciencia puede sernos de gran utilidad para reflexionar sobre nuestros métodos de selección de datos, creación e interpretación de modelos. Sólo así podremos **entender el tipo de conocimiento que estamos generando, ser conscientes de los valores que privilegiamos con nuestros métodos y de sus limitaciones**.

La intencionalidad como factor clave en la toma de decisiones basadas en Big Data*

Por Marc Le Menestrel, PhD

Associate Professor, Department of Economics and Business, Universitat Pompeu Fabra

Visiting Professor of Corporate Governance and Sustainability, INSEAD

El uso del *Big data* refuerza el conocimiento del que se dispone para tomar decisiones. Ello no significa que con *big data* se vayan a tomar mejores decisiones ya que, **no es siempre cierto que con un conocimiento más amplio de un tema se puede tomar la mejor decisión**. A veces, el hecho de disponer de demasiada información repercute en la forma en que tomamos decisiones, y nos impide realizar una valoración emocional y ética correcta ya que perdemos la orientación intuitiva. Por tanto, en este caso, pese a disponer de una información muy amplia, habría que retroceder, simplificar la cuestión y tomar una decisión teniendo en cuenta los factores más relevantes por considerar.

Lo más importante es la intención, qué información se va a utilizar para tomar decisiones y qué influencia queremos lograr. Por ejemplo, si nos fijamos en el ámbito de la política, los políticos pueden decidir a quién influenciar para aumentar sus probabilidades de ser elegidos. Sin embargo, ¿el elector realmente puede estar seguro de que está votando a quién realmente más le interesa? De hecho, esto es un riesgo por la democracia. Lo mismo ocurre cuando se quiere conseguir cualquier otro tipo de objetivo basándose en *big data* e inteligencia, como por ejemplo inteligencia social o militar. Disponer de información masiva siempre beneficia a determinadas instituciones y corporaciones, pero puede cuestionarse si ello mejora **los derechos o la seguridad de los ciudadanos**. Por otro lado, cabe destacar que el uso del *big data* puede mejorar nuestra comprensión sistemática de la sociedad, pero raramente esta comprensión es utilizada con un propósito de beneficio social. Por lo tanto, lo más importante es la intención que se tenga al utilizar *big data*.

En segundo lugar, la toma de decisiones incluye, por su naturaleza, valores, emociones e intenciones éticas, pero actualmente, como consecuencia de la utilización de sistemas automáticos, se está observando una carencia de todos estos elementos. En esta línea, la tecnología asociada al *big data* podría reflejar los múltiples valores que intervienen en la toma de decisiones de las empresas y así ayudar a entender el comportamiento de los consumidores. Pero ello no es automáticamente así: se está acentuando la falta de consideraciones éticas a la hora de tomar decisiones. Sin embargo, **el riesgo está aquí**. Se ha convertido en una forma de que disponen las empresas para manipular mejor, en beneficio propio, a los consumidores en vez de darles respuesta a sus verdaderas necesidades actuales. Con todo, cabe señalar que esta tecnología incluye técnicas realmente excelentes para hacer análisis cualitativos, y no solo cuantitativos, de los datos a partir de expresiones faciales y emociones, aunque dichas técnicas se utilizan con el fin de manipular aún más. El mundo necesita más utilizar el corazón y el alma para la toma de decisiones que basarse en datos.

Así pues, no está claro que la aplicación del *big data* en las empresas logre contribuir a mejorar su ética, puesto que se trata de un concepto que, a primera vista, parece más sencillo de lo que realmente es. **La cuestión depende de la responsabilidad y de la intencionalidad de cada empresa**, de si se desea utilizar o no el *big data* para ayudar a detectar los abusos éticos del sistema, basándose en el consumo de bienes. Una solución para que las empresas puedan mostrar si realmente tienen una verdadera preocupación social es la transparencia, la información abierta. Al respecto, existen sólidos argumentos que defienden que el *big data* ha de ser totalmente abierto. Así, siempre habrá la posibilidad de explorar cómo se está utilizando la información.

Las principales preocupaciones potenciales sobre el uso de esta tecnología revolucionaria son que se trata de un método que reforzará el poder de los negocios, las empresas y los políticos sobre la sociedad. La naturaleza de la tecnología puede hacer que la legislación política sea débil para regular la protección de los individuos frente a ella. Por este motivo, el *big data* como negocio no es un factor natural en la armonía de la gobernanza global. Claramente, **la tecnología big data todavía no se ha desarrollado para mejorar la seguridad del mundo**, aunque mucha gente así lo crea. Además, también puede ser un factor de influencia indebida sobre algunas empresas o de algunos países sobre otros. Aquí se deberían analizar las cuestiones relativas a la soberanía de los datos. Sería interesante que nos preguntáramos y exploráramos profundamente por qué la privacidad es un tema ético. Es necesario debatir más abiertamente sobre temas relacionados con el control social, la manipulación social o la ingeniería del consentimiento, entre otros. También sería interesante explorar las formas de alineación de los ciudadanos, porque abundan y tendrían que ser una cuestión de preocupación constante.

En conclusión, factores de riesgo como la manipulación de grupos de usuarios o la violación de la privacidad se extienden con la proliferación del *big data*. En general, se están utilizando cada vez más datos estadísticos para influenciar en las decisiones individuales a partir de lo que se hace a escala colectiva, lo cual dificulta al individuo que pueda tomar su propia decisión. A fin de cuentas, **no todo se tendría que basar en los estudios analíticos**, sino considerar también la intención, el propósito y los valores. Y siempre deberíamos preguntarnos a nosotros mismos: “¿A quién estoy sirviendo?”

*Nota: este texto se ha desarrollado a partir de una entrevista realizada al Prof. Le Menestrel

Retos y oportunidades en la toma de decisiones algorítmicas basadas en datos: puntos claves para avanzar hacia el bien social ¹

Por Nuria Oliver, PhD

Chief Data Scientist en DataPop Alliance

Director of Research in Data Science en Vodafone

La disponibilidad de cantidades ingentes de datos sobre el comportamiento humano está transformando el mundo en el que vivimos. Hoy contamos con cantidades masivas de datos que podemos utilizar para entrenar algoritmos que permiten a investigadores, empresas, gobiernos y otros actores del sector público abordar problemas complejos. Decisiones con impacto tanto individual como colectivo que previamente eran tomadas por humanos —con frecuencia expertos— son tomadas hoy en día por algoritmos, incluyendo decisiones relativas a la contratación de personas, la concesión de créditos y préstamos, las sentencias judiciales o la compra-venta de acciones en bolsa. Las decisiones algorítmicas basadas en datos tienen el potencial de mejorar la eficiencia del gobierno y de la provisión de servicios al optimizar procesos burocráticos, proporcionar *feedback* en tiempo real y predecir resultados. La historia ha demostrado que las decisiones humanas están sujetas a conflictos de interés, corrupción y sesgos, lo que ha resultado en procesos y resultados injustos y/o ineficientes. Por tanto, el interés hacia el uso de algoritmos basados en datos puede interpretarse como el resultado de una **demanda de mayor objetividad en la toma de decisiones** y de un conocimiento más profundo de nuestro comportamiento y necesidades tanto individuales como colectivas.

Sin embargo, la toma de decisiones basadas en algoritmos entrenados con datos **no está exenta de limitaciones** que es necesario abordar para conseguir que el impacto de esta nueva manera en la toma de decisiones sea positivo, incluyendo:

1. **Discriminación:** La discriminación algorítmica puede surgir de distintas fuentes. En primer lugar, los datos utilizados para entrenar los algoritmos pueden tener sesgos que den lugar a decisiones discriminatorias. En segundo lugar, la discriminación puede surgir como consecuencia del uso de un determinado algoritmo. La categorización, por ejemplo, puede ser considerada una forma de discriminación directa, ya que utiliza algoritmos para hacer un tratamiento dispar de las diferentes clases. En tercer lugar, los algoritmos pueden resultar en discriminación como resultado del mal uso de ciertos modelos en diferentes contextos. En cuarto lugar, datos sesgados pueden ser usados tanto como evidencia para el entrenamiento de algoritmos como prueba de su eficacia. Además, el uso de procesos de decisión algorítmicos basados en datos pueden implicar que se les denieguen oportunidades a personas no por sus propias acciones, sino por las acciones de otros con los que comparten ciertas características. Por ejemplo, algunas compañías de tarjetas de crédito han reducido el límite de crédito de sus clientes no por el historial de pagos del propio cliente, sino como resultado del análisis del comportamiento de otros clientes con un historial de pagos deficiente que ha-

bían comprado en los mismos establecimientos donde el cliente había comprado. En la literatura se han propuesto diferentes soluciones para afrontar la discriminación algorítmica y maximizar la justicia. Sin embargo, me gustaría subrayar la urgencia para que expertos y expertas de distintos campos (incluyendo el derecho, la ética, la informática, la filosofía y las ciencias políticas) **inventen, evalúen y validen en el mundo real diferentes métricas de justicia algorítmica** para diferentes tareas. Además de esta investigación empírica, es necesario proponer un marco de modelado teórico —avalado por la evidencia empírica— que ayude a los usuarios de dichos algoritmos a asegurarse de que las decisiones tomadas son lo más justas posible.

2. **Falta de transparencia/opacidad:** La transparencia hace referencia a la cualidad de poder entender un modelo computacional y por ello puede ser un mecanismo que contribuya a la atribución de responsabilidad de las consecuencias del uso de dicho modelo. Un modelo es transparente si una persona puede observar y entender con facilidad el modelo entero. Por ello, sería deseable que los modelos tuviesen una complejidad computacional

baja. Burrell propone tres tipos distintos de opacidad —i.e. falta de transparencia— en las decisiones algorítmicas: (1) opacidad intencional, cuyo objetivo es la protección de la propiedad intelectual de los inventores de los algoritmos. Este tipo de opacidad podría mitigarse con legislación que obligaría al uso de sistemas de software abierto. La nueva Regulación Europea de Protección de Datos (GDPR) con su *derecho a una explicación* es un ejemplo de este tipo de legislación. Sin embargo, intereses comerciales y gubernamentales poderosos pueden dificultar la eliminación de este tipo de opacidad; (2) opacidad de conocimiento, debido al hecho de que la gran mayoría de las personas carecen de las habilidades técnicas para entender cómo funcionan los algoritmos y modelos computacionales construidos a partir de los datos. Este tipo de opacidad se vería atenuada con programas educativos que incluyesen una asignatura de pensamiento computacional y permitiendo que expertos independientes pudiesen aconsejar a aquellos afectados por los procesos de decisión algorítmicos basados en datos; y (3) opacidad intrínseca, que surge por la naturaleza de ciertos métodos de aprendizaje por ordenador (e.g. modelos de aprendizaje profundo). Esta opacidad es bien conocida en la comunidad de investigación de aprendizaje computacional y también se conoce como el problema de la interpretabilidad.

3. **Violación de la privacidad:** Informes y estudios se han enfocado en el mal uso de datos personales proporcionados por los usuarios de servicios y en la agregación de datos de diferentes fuentes por parte de entidades como los *data brokers* con implicaciones directas en la privacidad de las personas. Un elemento que a menudo no se tiene en cuenta es que los avances en los algoritmos combinados con la disponibilidad de nuevas fuentes de datos sobre el comportamiento humano (e.g. datos de social media) permiten la **inferencia de información privada** (e.g. orientación sexual, incli-

naciones políticas, nivel de educación, estabilidad emocional) que nunca ha sido explícitamente revelada por las personas. Este elemento es esencial para entender las implicaciones del uso de algoritmos como ha quedado patente en el escándalo reciente de Facebook/Cambridge Analytica².

4. **Alfabetismo digital:** Es de gran importancia que dediquemos recursos a programas de **alfabetismo digital y computacional** para todos los ciudadanos, desde los niños hasta las personas mayores. Si no lo hacemos, será muy difícil, sino imposible que seamos capaces como sociedad de tomar decisiones sobre tecnologías que no entendemos. En ese sentido, el libro “Los nativos digitales no existen”³ enfatiza la necesidad de enseñar pensamiento computacional y a hacer un buen uso de la tecnología a niños y adolescentes.

5. **Responsabilidad difusa:** Conforme cada vez más decisiones que afectan a millones de personas son tomadas automáticamente por algoritmos, es necesario que haya claridad con relación a quien es responsable de las consecuencias de dichas decisiones. La transparencia suele considerarse un factor fundamental para contribuir a la **claridad de atribución de responsabilidad**. Sin embargo, la transparencia y las auditorías no son suficientes para garantizar una clara responsabilidad. De hecho, en un artículo reciente, Kroll et al. han propuesto el uso de métodos computacionales para proporcionar claridad con relación a la atribución de responsabilidad, incluso cuando parte de la información esta oculta.

1. Este artículo esta basado en los artículos [33] y [34] de la bibliografía

2. https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

3. <https://www.amazon.es/Los-nativos-digitales-existen-colecci%C3%B3n/dp/8423426599>



6. **Falta de marcos éticos:** Los procesos de toma de decisiones algorítmicas basadas en datos generan dilemas éticos importantes con relación a cuáles deberían ser las acciones adecuadas a tomar en base a las inferencias llevadas a cabo por los algoritmos. Por ello, es fundamental que las decisiones sean tomadas de acuerdo con un **marco ético** claramente definido y aceptado. Hay varios ejemplos de principios éticos propuestos en la literatura para este propósito ⁴⁵ y de institutos y centros de investigación creados con este fin, como el Digital Ethics Lab en la Universidad de Oxford o el AI Now Institute en NYU. Sin embargo, todavía es un área activa de investigación y no hay un método único para la incorporación de principios éticos en los procesos algorítmicos de decisión basados en datos. Es importante destacar la importancia de que todos los desarrolladores y profesionales que trabajen en el desarrollo y uso de algoritmos para la toma de decisiones se comporten de acuerdo a un claro Código de Conducta y de Ética definido por las organizaciones en las que trabajan.

7. **Falta de diversidad:** Dada la variedad de casos de uso en los que pueden aplicarse algoritmos para la toma de decisiones, es importante reflexionar sobre la frecuente falta de diversidad en los equipos que generan dichos algoritmos. Hasta ahora, el desarrollo de los algoritmos basados en datos y en técnicas de inteligencia artificial para la toma de decisiones han sido desarrollados por grupos homogéneos de profesionales de la informática. En el futuro, deberíamos **asegurarnos de que los equipos son diversos** tanto con relación a las áreas de conocimiento como con relación a factores demográficos –y en particular de género, donde el porcentaje de mujeres profesionales en la informática es menor de un 20% en muchas empresas tecnológicas.

Para concluir, me gustaría destacar tres requisitos centrados en las personas que considero de vital importancia para permitir la disrupción positiva de las decisiones algorítmicas basadas en datos, como ilustra la Figura 1: (1) propiedad y gestión de los datos centrados en las personas; (2) transparencia, responsabilidad, respeto a la privacidad, ética, justicia y diversidad en los algoritmos, como he explicado anteriormente; y (3) *living labs* para experimentar políticas centradas en los datos. Este tercer requerimiento consiste en desarrollar laboratorios para experimentar las decisiones tomadas a partir del manejo de información, es decir, para experimentar y co-crear políticas y soluciones basadas en datos, y a su vez consensuadas con los humanos.

En conclusión, nos encontramos en un momento sin precedentes en la historia de la humanidad, donde disponemos de una gran cantidad de datos sobre el comportamiento humano. La toma de decisiones basada en Big Data y la utilización de algoritmos inteligentes nos ofrece grandes oportunidades, en tanto en cuanto tengamos en cuenta los riesgos y limitaciones que ello comporta.

Por ello será solamente cuando respetemos los tres requisitos ilustrados en la Figura 1, seremos capaces de avanzar y pasar de una posible tiranía de los datos y de los algoritmos a un modelo de gobernanza democrática basado en los datos, por y para las personas. El potencial para tener impacto positivo es inmenso. No deberíamos desaprovechar esta oportunidad.

Figura 1: Resumen de los requisitos para conseguir que la toma de decisiones algorítmicas basadas en datos sea positiva



4. <https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/>

5. <https://futureoflife.org/ai-principles/>

Dirección ejecutiva

Núria Agell, Profesora, ESADE

Manu Carricano, Profesor Asociado, ESADE

Luis Vives, Profesor titular, ESADE

Análisis de resultados

Ferran Mazaira i Font

Oriol Cosp i Arqué

Juan José Garau Luis

Gonzalo de la Torre Carazo

Miguel Teixidó

Diseño

Vänster and Lei | www.vansterandlei.com

© 2018 ESADE Business & Law School

www.esade.edu/d3

www.esade.edu/ftmba



ESADE

RAMON LLULL UNIVERSITY

INSTITUTE FOR
DATA-DRIVEN
DECISIONS

