



Big Data: La disrupción del dato.

Editorial

Rafael Conde del Pozo
*(Digital & Innovation
Director, Vector ITC).*

La llegada de tecnologías como la computación en la nube, la mejora de la velocidad de la red y diferentes técnicas para analizar la información, ha generado la capacidad suficiente para convertir cantidades enormes de datos complejos en insights de gran valor añadido.

El objetivo es obtener conocimiento de negocio detallado, fiable y profundo; para poder tomar decisiones más inteligentes y con menor riesgo: el valor real del Big Data no está en acumular un gran volumen de datos, sino en extraer el máximo valor de ellos.

En este sentido, la IA es la gran tendencia dentro de todo el ecosistema Big Data; dado que una vez cubierta la parte de captura e infraestructura, el paso lógico es buscar soporte a la hora de analizar conjuntos de datos masivos y ser capaz de predecir escenarios mediante su análisis.

La evolución del Big Data pasa por aplicar algoritmos de Machine y Deep Learning a grandes volúmenes de información. El aprendizaje es posible gracias a la detección de patrones dentro de los conjuntos de datos, de forma que es viable obtener predicciones sobre situaciones futuras, extraer recomendaciones automáticas sobre que modelos de datos son más adecuados, y, finalmente, a través del aprendizaje continuo, generar insights que se conviertan en decisiones y resultados exitosos y fiables.

Rafael Conde del Pozo
(Digital & Innovation Director, Vector ITC).





Índice.

01 **6 - 27**

Big Data: Introducción

- Consideraciones Generales
- Estado del Arte: tendencias
- Big Data & Analytics aplicado al negocio
 - _ *Data Driven companies*
 - _ *Aplicaciones sectoriales*
 - _ *Business Performance Analytics*
 - _ *Customer Analytics: SMART KPI's*
 - _ *Big Data Technologies & Architecture*

03 **40 - 49**

Data Science

- Data Mining
- Analítica Avanzada
- Deep Learning: Modelos Predictivos

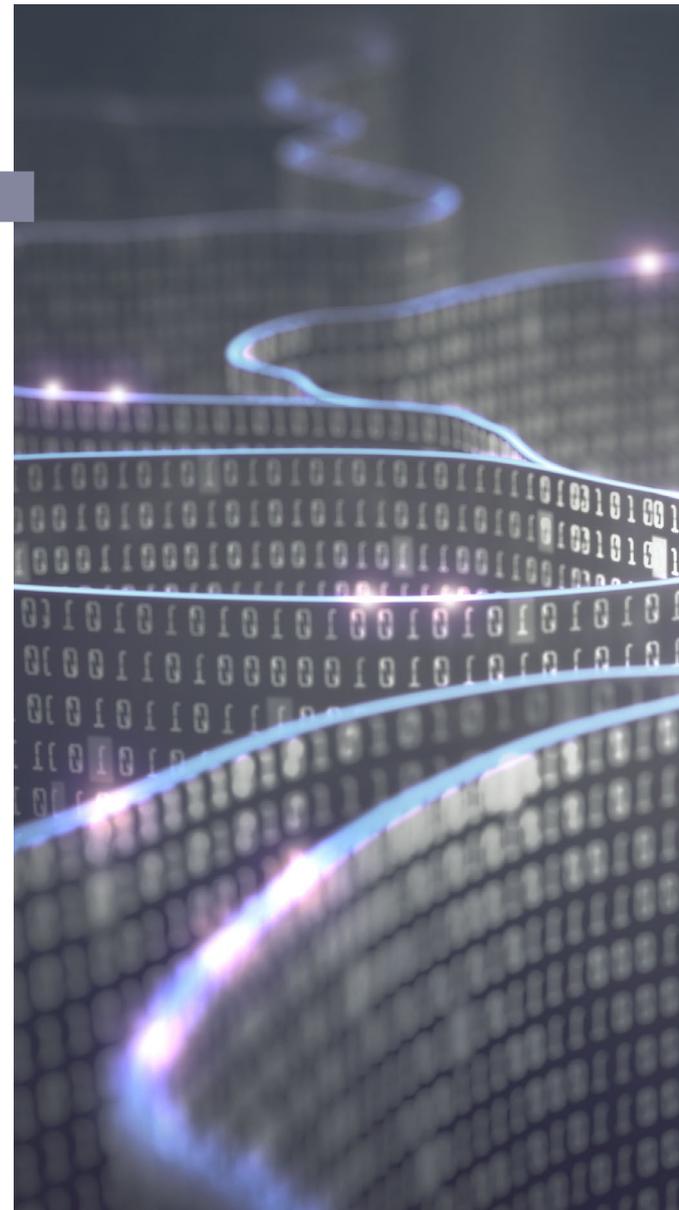
04 **50 - 53**

Conclusiones

02 **28 - 39**

Data Analytics & Visualization

- Data Management
- Data Governance
- Visualización de datos





Big Data: Introducción

01

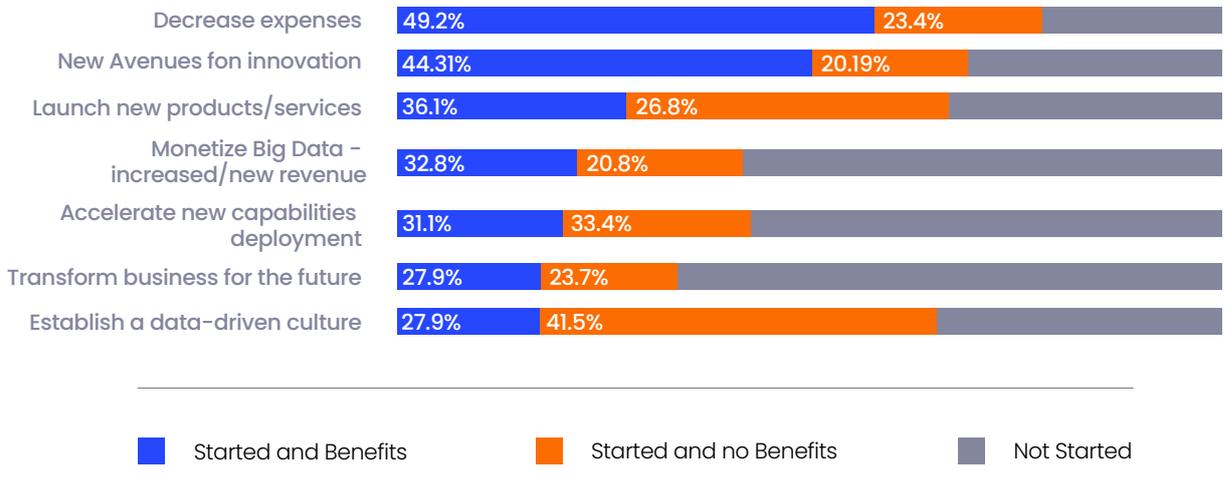
CONSIDERACIONES GENERALES

Durante los últimos años, el término *Big Data* está en boca de todos los profesionales y expertos de todos los sectores económicos y empresariales. De hecho, casi la mitad de las empresas españolas aseguran utilizarlo en su día a día. Sin embargo, la realidad es que el 65% de los proyectos basados en esta tecnología que se lanzan en España acaban fracasando. Bien por la escasa preparación tecnológica y formativa de las organizaciones, o bien por no contar con los expertos externos que ayuden a diseñar y monitorizar este tipo de innovaciones. A pesar de ello, la inversión en Big Data no para de crecer año tras año.

El Big Data es una de las tendencias con mayor impacto de las dos últimas décadas, y su rango de aplicaciones es altamente variado: desde la **toma de decisiones empresariales, pasando por la segmentación de clientes, hasta el aumento drástico de la productividad**. Por ejemplo, un 53% de las empresas a nivel mundial han mejorado sustancialmente sus servicios de Atención al Cliente gracias a esta tecnología. Otras muchas, gracias a herramientas como Hadoop y Spark han logrado analizar un volumen de datos mucho más alto y más rápidamente, lo que se traduce en un aumento destacado de la productividad de la fuerza laboral. También en la detección del fraude el Big Data ha demostrado su enorme valía: gracias al Machine Learning es posible detectar patrones y anomalías, especialmente útiles para bancos y aseguradoras, permitiéndoles detectar robos de tarjetas de crédito o compras fraudulentas antes de que el titular sea consciente de ello.

TODO ESTO SE TRADUCE, EN MUCHOS CASOS, EN UNA SIGNIFICATIVA REDUCCIÓN DE COSTES:

BIG DATA INITIATIVES AND SUCCESS RATE

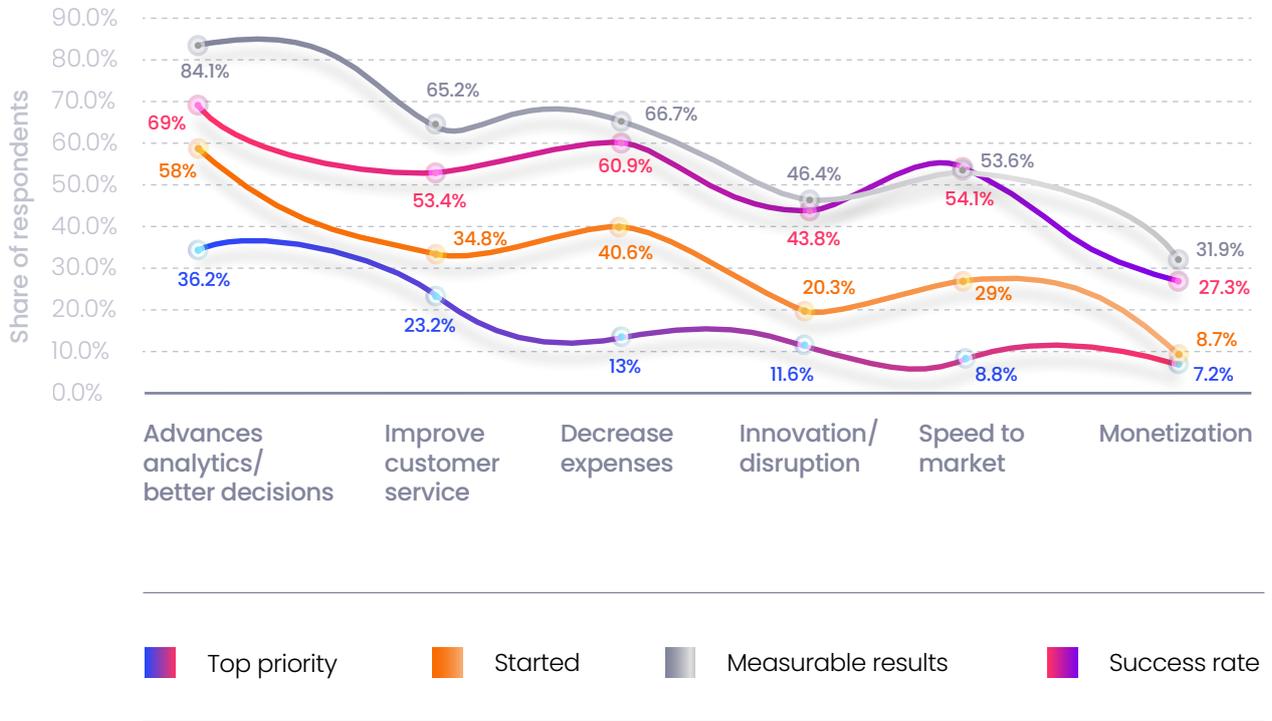


Big Data business initiatives underway; with successful results	Started	Success
Decrease expenses through operational cost efficiencies	72.6%	49.2%
Establish a data-driven culture	69.4%	27.9%
Create new avenues for innovation and disruption	64.5%	44.3%
Accelerate the speed with which new capabilities and services are deployed	64.5%	31.1%
Launch new product and service offerings	62.9%	36.1%
Monetize Big data through increased revenues and new revenue sources	54.8%	32.8%
Transform and reposition your business for the future	51.6%	27.9%

Fuente: NewVantage Venture

Las estadísticas muestran que el 84% de las empresas en Estados Unidos (el mercado con mayor implantación de Big Data) han lanzado proyectos basados en esta tecnología para lograr una mayor precisión y acelerar la toma de decisiones. Y la tasa de éxito es de las más elevadas, un 69%. Las siguientes aplicaciones más populares, con prácticamente un tercio de las empresas adoptándolas, son la mejora de la atención al cliente y la reducción de costes. En cambio, solo un tercio de ellas se han volcado en el uso de esta tecnología para la monetización, un campo menos maduro en el que todavía queda mucho por explorar, tecnológicamente y comercialmente.

BIG DATA INITIATIVES AND SUCCESS RATES AMONG CORPORATIONS IN THE UNITED STATES AND WORLDWIDES, AS OF 2018



Measurable Results from Big Data/AI

	2017	2018	2019
Yes	48.4%	73.2%	62.2%
No Too Early To Tell	51.6%	26.8%	37.8%

Fuente: NewVantage Venture



No obstante, pese a las obvias ventajas que aporta el Big Data en cualquier organización, **todavía son muchos los desafíos** que obstaculizan el éxito de muchos proyectos en este campo. Entre ellos, destacan los siguientes:

NECESIDAD DE TALENTO

Los Científicos de Datos se encuentran entre los trabajadores más codiciados y mejor pagados en el sector IT. La falta de este tipo de profesionales ha sido el desafío número uno en el desarrollo del Big Data en los últimos tres años.

CALIDAD DE DATOS

Uno de los principales desafíos al trabajar con Big Data es que, antes de poder analizar los datos, los científicos y analistas necesitan asegurarse de que la información que están usando sea precisa, relevante y tenga el formato adecuado para el análisis. Aunque esta casuística ralentiza considerablemente el proceso de presentación de informes, es crítico abordar los problemas de calidad de los datos, para asegurar que las ideas generadas tengan valor y validez.

REGLAMENTACIÓN

Gran parte de la información relativa a las empresas es confidencial o personal, lo que significa que necesita cumplir ciertos estándares normativos propios de la industria o requisitos gubernamentales a la hora de tratar y almacenar los datos.

CIBERSEGURIDAD

El almacenamiento de datos, especialmente aquellos catalogados como confidenciales, puede hacer que las empresas sean un objetivo más atractivo para los ciberataques.

CAMBIO RÁPIDO

La tecnología cambia constante y rápidamente, por lo que las organizaciones se enfrentan a la posibilidad real de invertir en una herramienta concreta, y que se quede obsoleta a corto plazo.



ESTADO DEL ARTE: TENDENCIAS

El Big Data, con su potencial para determinar información valiosa para mejorar el proceso de toma de decisiones, ha despertado un gran interés entre académicos y profesionales del sector en los últimos años, convirtiéndose en una de las tendencias tecnológicas dominantes de esta década y uno de los segmentos más dinámicos e innovadores en el mercado de la tecnología de la información.

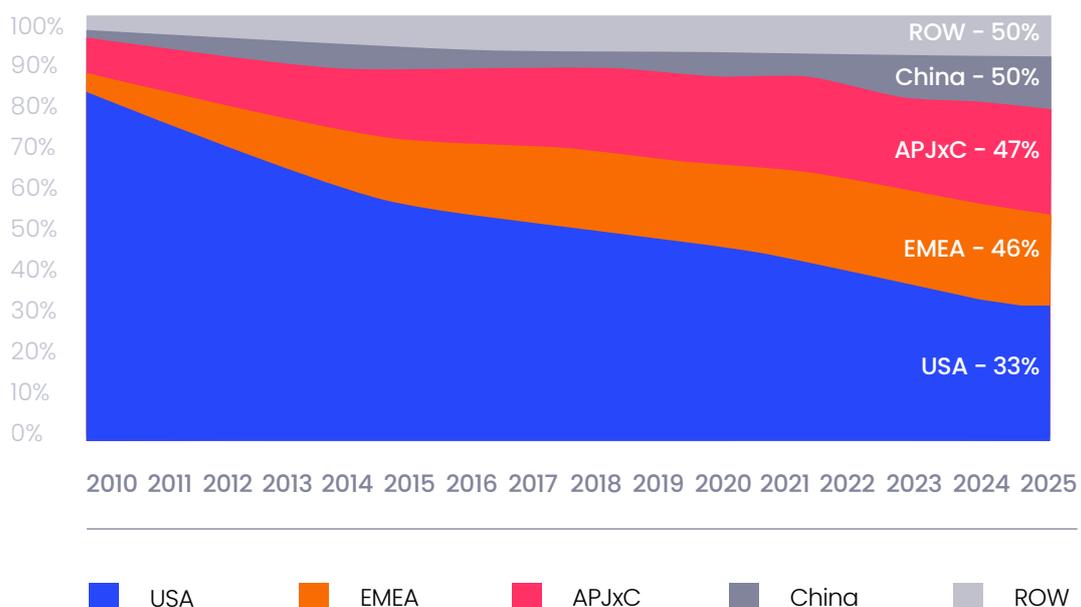
La magnitud de los datos generados y compartidos por empresas, Administraciones Públicas, numerosos sectores industriales y entidades

científicas ha aumentado en los últimos años llegando a los 33 Zettabytes en 2018, cifra que se espera que aumente a un total de 175 ZB en 2025. Es una cifra gigantesca que representa una oportunidad sin igual para la esfera económica.

Ante este nuevo panorama, los líderes empresariales se encontrarán con nuevas posibilidades para acceder a oportunidades de negocio únicas, potenciadas por el abrumador volumen de datos. Sin embargo, también se enfrentan a la **toma de decisiones estratégicas sobre la recopilación, utilización y ubicación** de los mismos.

Casi el 20% de la información en la esfera de datos global tendrá gran importancia para la vida cotidiana, y cerca de la mitad de estos pasarán a ser críticos para la subsistencia. De hecho, se prevé que para el año 2025, una persona promedio conectada a la red en cualquier parte del mundo interactuará con dispositivos conectados casi 4.800 veces al día, es decir, aproximadamente una interacción cada 18 segundos.

CLOUD STORAGE GROWTH AND SHARE BY REGION 2015-2025 CAGR



De todos los datos que se generen, el 49% serán almacenados en entornos Cloud públicos, lo que deja entrever la importancia que irá adquiriendo este entorno de almacenamiento.

Además, el carácter de estos datos será verdaderamente dinámico, pues el 30% se generarán en tiempo real.

En la actualidad, Estados Unidos lidera el ranking de los países que más datos generan.

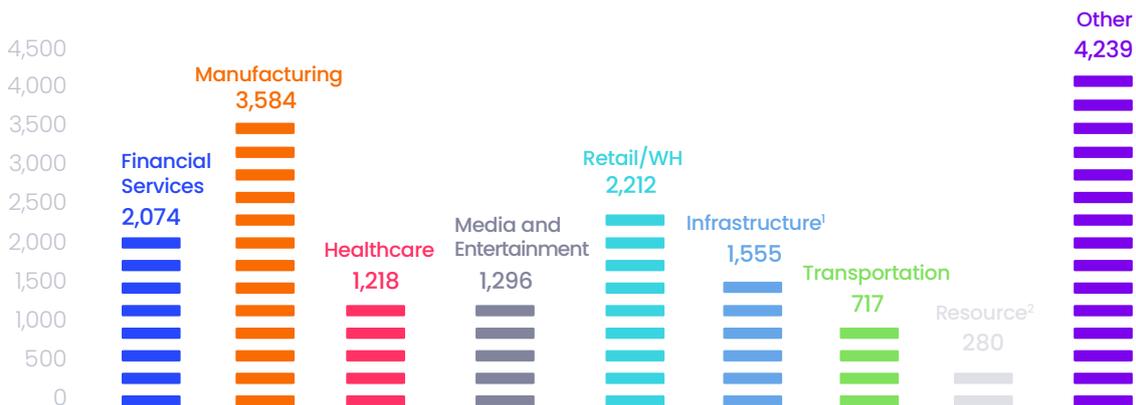
Según las estimaciones de la industria, se proyecta que la cantidad de datos almacenados en los EE. UU. crecerá de 0.9ZB en 2018 a 3.4ZB en 2025.

Junto con este aumento, el almacenamiento en Clouds públicas aumentará, ya que el 55% de los datos almacenados por las organizaciones y los consumidores se alojarán en ellas en 2025.

La evolución del tamaño, naturaleza, origen y almacenamiento de los datos determinan de forma decisiva el futuro del Big Data. Plantea excelentes oportunidades, dado que los **clientes aumentarán exponencialmente sus interacciones con dispositivos generadores de datos**, permitiendo a las empresas evolucionar sus productos y servicios prácticamente en tiempo

real; y también retos vitales, especialmente de índole tecnológica y formativa, como el desarrollo de herramientas que permitan procesar información a un ritmo todavía inimaginable, la adopción de los sistemas de Cloud pública por parte de empresas y administraciones, el desarrollo de marcos regulatorios adecuados que conjuguen la explotación del potencial de los datos con la protección de la privacidad de los usuarios, y un largo etcétera.

2018 ENTERPRISE DATASPHERE BY INDUSTRY (EB)



¹ Infrastructure includes Utilities, Telecommunications

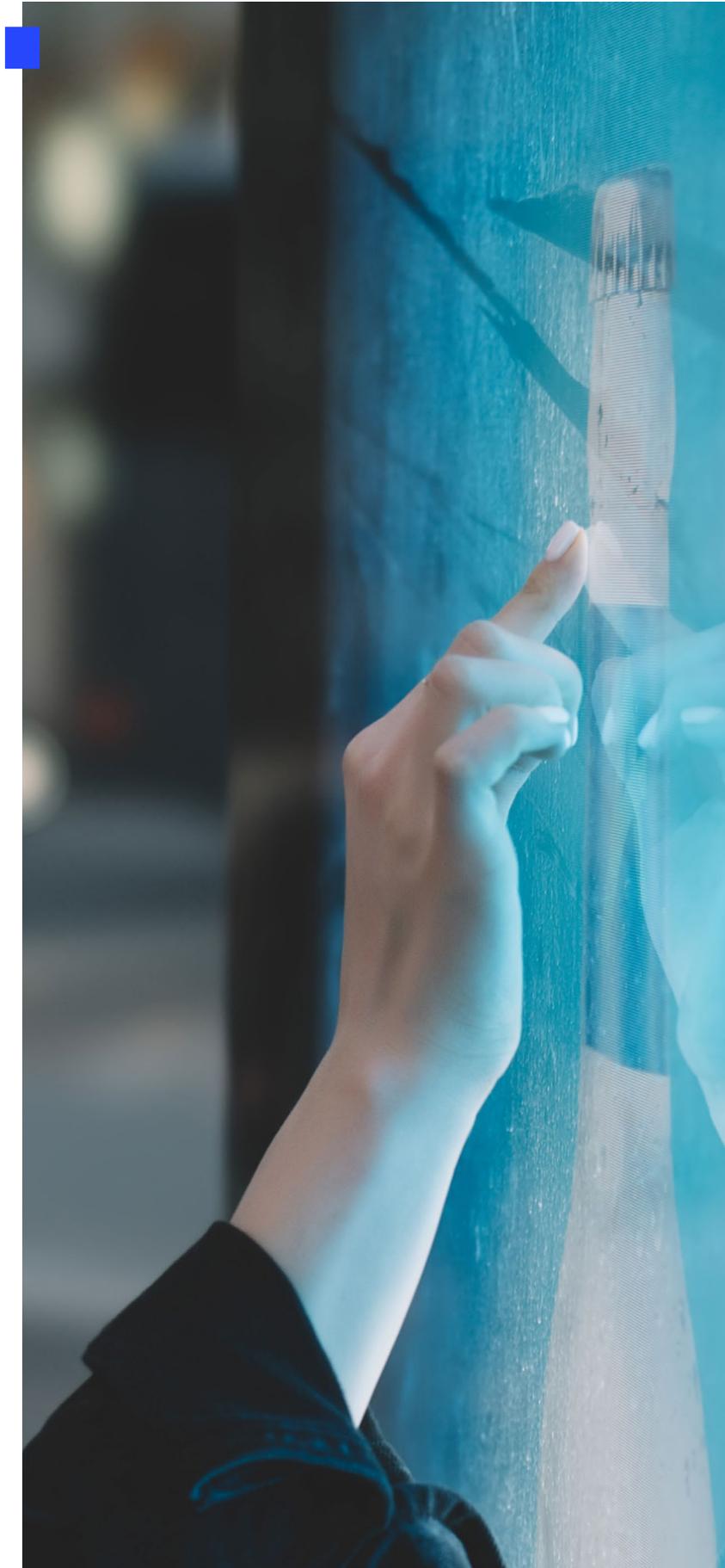
² Resource includes Oil and Gas (Mining), Transportation of oil & gas through pipelines or shipping, Resource industries, Petroleum and coal manufacturing/refining

Por otro lado, la adopción no se producirá de forma homogénea en las distintas industrias, e incluso en los distintos puntos de la cadena de valor dentro de cada industria. Por ejemplo, Ventas y Marketing, Investigación y Desarrollo (R&D) y Supply Chain Management (SCM) son las áreas que están haciendo las mayores contribuciones al crecimiento de la inversión en Big Data actualmente.

De este modo, se espera que los ingresos en el mercado mundial de Big Data para Software y Servicios aumentarán de 42.000 millones de dólares en 2018 a 103.000 millones en 2027, **alcanzando una tasa de crecimiento anual compuesta del 10,48%.**

Es relevante observar y destacar que la mayoría de las herramientas de Big Data son creaciones de las grandes compañías tecnológicas: los motores de búsqueda más conocidos (Google y Yahoo), y las Redes Sociales de mayores dimensiones (Facebook, Twitter y LinkedIn) donde los datos son masivos y en gran parte semiestructurados. A las que cabría añadir, por supuesto, Apple, y el gigante del e-Commerce, Amazon.

El claro liderazgo que ejercen en el ámbito del análisis de datos, facilitado sin duda por su condición de generadores masivos de datos de usuarios, prefigura la tendencia a la concentración de la propiedad en los dispositivos de hardware y software en los que se basa el Big Data.



BIG DATA & ANALYTICS APLICADO AL NEGOCIO

Las empresas saben que tienen que aumentar la inversión en tecnología para ser más ágiles y competitivas. En concreto, **el ritmo de inversión en Big Data está aumentando exponencialmente.**

El número de empresas que invirtieron más de 500 millones de dólares en Big Data ha aumentado del 12,7% en 2018 al 21,1% en 2019. Del mismo modo, el número de empresas que invierten más de 50 millones de dólares ha aumentado del 39,7% en 2018 al 55% en 2019.

Esta tendencia seguirá creciendo y el futuro se antoja prometedor para los actores del mercado tecnológico del Big Data, y para las compañías que apuesten primero por su aplicación sectorial. Por supuesto, esta afirmación incluye a España, donde el mercado local de Big Data continuará su crecimiento a un ritmo anual del 6% y superará los 540 millones de euros de facturación en 2021.

Sin embargo, la elevada inversión no siempre conlleva una correcta gestión de la misma. La falta de expertos en la materia y asesoramiento externo lleva a muchas empresas a la desorientación en torno a la aplicación óptima de las nuevas tecnologías en general, y del análisis de datos en particular. Existen, no obstante, plataformas que señalan el camino y otorgan una visión más clara.



No obstante, en la mayoría de los casos que se dan en España, la dificultad de adopción no se debe a la falta de plataformas tecnológicas que puedan fundamentar la transformación del segmento concreto de la cadena de valor de cada compañía, sino a la desorientación estratégica, a la falta de asesoramiento externo que aporte el talento, experiencia y conocimiento tecnológico y empresarial necesario para guiar la transformación.

DATA DRIVEN COMPANIES

Las compañías impulsadas por datos (Data Driven Companies) que utilizan la información en toda su organización para generar ventajas competitivas están creciendo a un ritmo de más del 30% anual, y están en camino de generar 1 800 millones de \$ para el 2021.

Convertirse en **compañías impulsadas por datos** ha sido un objetivo común para muchas empresas en la última década. Este tipo de empresas se basan en el análisis de datos para la toma de mejores decisiones, el desarrollo de nuevos productos y servicios, etc. No se trata de un apoyo accesorio que brindan herramientas concretas de Big Data, sino que la aplicación de dicha tecnología fundamenta todos los puntos de su cadena de valor. Y seguramente no es casualidad que gran parte de los productos, servicios, procesos y soluciones más innovadores que actualmente podemos encontrar en el mercado, hayan sido desarrollados por compañías de este tipo.

Los nombres de las empresas que han logrado una cuota de mercado indiscutible gracias a su apoyo en los datos son de sobra conocidos. **Amazon, Google y Netflix han utilizado con éxito los datos** para comprender a sus clientes, comercializar sus soluciones, mejorar los procesos operativos e incrementar los ingresos. El ejemplo más sencillo es Google (ahora Alphabet) que obtiene el 90% de sus 89 000 millones de

dólares de ingresos gracias a Google AdWords, el sistema de publicidad que segmenta clientes en base a datos más exitoso del mundo.

En cuanto a Netflix, es la primera compañía que ha utilizado el Big Data para analizar los hábitos de visualización de sus suscriptores y, en base a ellos, generar contenido audiovisual personalizado, convirtiéndose en la plataforma de contenido líder de Estados Unidos.

De este modo, vemos que gigantes tecnológicos como Amazon, Google o Netflix, no tienen en común su modelo de negocio, ni siquiera esencialmente su sector, sino el núcleo en torno al cual construyen su imperio empresarial, que es el análisis de datos de comportamiento del cliente, y la evolución de la cadena de valor en base a los resultados cuantitativos y cualitativos de dicho estudio.

Pero estos tres gigantes no son los únicos ejemplos de empresas que van más allá de saber quiénes son sus clientes y confían en el análisis de datos para obtener una imagen más precisa de lo que realmente quieren y necesitan.

De hecho, la mayoría de las empresas pertenecientes a mercados maduros han intentado tratar los datos como un activo importante, desarrollar su cultura corporativa hacia una dirección que esté orientada a los datos y ajustar sus estrategias para enfatizar los datos y el análisis, pero no todas consiguen su objetivo final, puesto que muchas de estas compañías están teniendo problemas para implementar este tipo de estrategias. El porcentaje de empresas que se identificaron como corporaciones impulsadas por los datos ha disminuido en los últimos 3 años, de 37,1% en 2017 a 32,4% en 2018 y a 31,0% este año.



APLICACIONES SECTORIALES

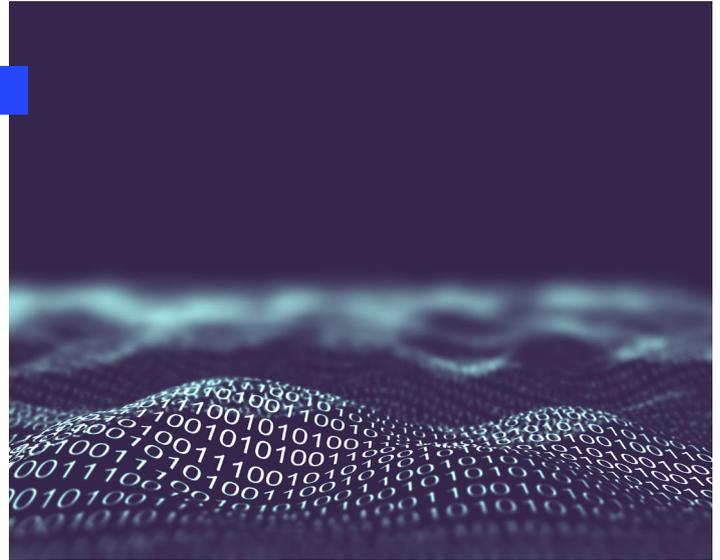
Existe una gran variedad de aplicaciones y soluciones Big Data para múltiples sectores, ya que esta tecnología tiene **suficiente potencial para irrumpir en casi cualquier industria**, pero algunos sectores tienen más probabilidades que otros de experimentar cambios significativos gracias al uso de esta tecnología.

INDUSTRIA AÉREA

Un avión crea 0.5 Terabytes de datos de media por cada vuelo que realiza. Toda esta información, combinada con datos sobre el clima, las interacciones del cliente con el servicio de atención al cliente, la plataforma de compra de billetes, etc., ofrece todo un ecosistema de datos que supone la mayor oportunidad de negocio para las aerolíneas. Especialmente en una industria altamente competitiva, donde el análisis de datos puede ser el instrumento fundamental para la fidelización del cliente. Además, el Big Data permitirá del mismo modo mejorar su eficiencia, reducir los retrasos y optimizar las técnicas de mantenimiento.

Un buen ejemplo es United Airlines, una compañía cuyos asistentes de vuelo cuentan con Wearables que permiten un acceso instantáneo al perfil digital del cliente que, en base a una serie de datos, como las especificaciones dietéticas, los asientos que acostumbra a seleccionar o sus interacciones pasadas con la aerolínea, generan una identidad personal, de forma que los empleados de la compañía pueden ofrecer una atención totalmente personalizada.

Otro ejemplo es Qantas Airways, una aerolínea internacional que utiliza el análisis de datos en una herramienta para pilotos. Esta nueva tecnología, llamada FlightPulse, proporciona recomendaciones en



tiempo real para optimizar la dirección de la aeronave, reduciendo el consumo de combustible y las emisiones. Los pilotos pueden revisar los datos de su propio procedimiento de despegue, crucero y aterrizaje, y la aplicación detecta los puntos susceptibles de optimización.

RECURSOS HUMANOS

El departamento de RR.HH. es un área fundamental de cualquier organización, y una de las más susceptibles de ser revolucionada por el Big Data. No solo en las funciones de localización y selección de personal, sino también para la optimización de la productividad y la prevención de la fuga de talento.



Fuente: AIHR



SALUD Y SANIDAD

Si bien es cierto que la industria de la salud adoptó el Big Data más tarde que otros sectores, los analistas señalan que 2019 será el año en el que las entidades de este sector adopten un enfoque más centrado en los datos. Su potencial es innegable: la predicción analítica permite mejorar la detección y el tratamiento de enfermedades, además de hacer más eficientes los procesos y reducir significativamente los costes.

El jefe de Servicio de Anatomía Patológica del Hospital Universitario Fundación Jiménez Díaz (HUFJD), el Dr. Federico Rojo, ha expresado que *“disponer de una amplia gama de información relacionada con imágenes, historia clínica y datos genómicos, permitirá contar con herramientas de Inteligencia Artificial para afrontar problemas de salud relevantes, mejorando la asistencia y el diagnóstico de los pacientes”*.

Actualmente, el campo de investigación más relevante es la aplicación de la **analítica de datos para detectar signos tempranos de la existencia de una enfermedad**, e incluso trabajar activamente en su prevención, para evitar que llegue a desencadenarse, especialmente entre poblaciones de riesgo. De este modo, se trabaja en estadios tempranos de la enfermedad, donde los tratamientos son mucho más eficaces y económicos.

Algunos ejemplos de esta incipiente revolución en el mundo de la salud son iniciativas como Horizonte 2020 a nivel europeo, o Savana en España, que se centran en diseñar algoritmos para mejorar el diagnóstico de enfermedades, y digitalizar los historiales clínicos de los pacientes para poder aplicar sobre ellos analítica de datos. No obstante, todos estos avances están planteando nuevos retos:

- **La privacidad de los datos:** al tratarse de datos clínicos, son datos especialmente sensibles, estrictamente regulados y con poca predisposición por parte de los usuarios a ser compartidos.

- **Formación necesaria:** los profesionales médicos no cuentan hoy en día con un conocimiento tecnológico suficiente para el manejo de este tipo de herramientas. La localización de personal adecuado para dirigir esta revolución tecnológica en los hospitales es ahora el mayor reto.

SECTOR ENERGÉTICO

El sector energético genera constantemente grandes cantidades de datos, gracias a los sensores, la transmisión inalámbrica y las tecnologías de computación Cloud en las que se basan muchos de los procesos de gestión. Hay más información que nunca sobre cómo se distribuye la energía, tanto en sentido cronológico como geográfico.

Todo ello contribuirá decisivamente a evitar accidentes en las plantas generadoras, y a optimizar



la producción energética. Los modelos predictivos y prescriptivos son especialmente útiles, ya que permiten conocer los resultados futuros y recomendar modificaciones para optimizarlos, **combinando factores ambientales, patrones climáticos o de consumo de energía.** Asimismo, con esta misma base de información, se están desarrollando ya herramientas que permiten prever los desastres medioambientales, pudiendo actuar antes de que sucedan.

La generación y planificación de la producción energética son también campos que se basarán al completo en el análisis de datos. Prever los picos y valles de consumo causados por acontecimientos climáticos o sociales con antelación es clave para optimizar la gestión de las plantas energéticas. Asimismo, también la energía renovable se verá especialmente beneficiada por el Big Data, pues se podrán elaborar previsiones de generación más precisas, teniendo en cuenta su dependencia de las condiciones climáticas, que determinan sus resultados. Todo ello gracias a datos meteorológicos, de consumo, datos GIS, etc.

SIEMENS Gamesa

RENEWABLE ENERGY

Por ejemplo, Siemens Gamesa, que aprovecha el Big Data para producir energía eólica de una forma más inteligente. Fue la primera empresa del mundo en instalar aerogeneradores inteligentes con sensores. Actualmente, cuenta con la mayor cantidad de datos históricos del sector, una base de datos que crece día a día con la recopilación de datos de más de 10 000 turbinas en todo el mundo. Pueden predecir y prevenir periodos de inactividad imprevistos, hacer que los parques eólicos se monitoricen en función de las condiciones y ampliar la vida útil de los aerogeneradores inteligentes.

SECTOR AUTOMOTRIZ

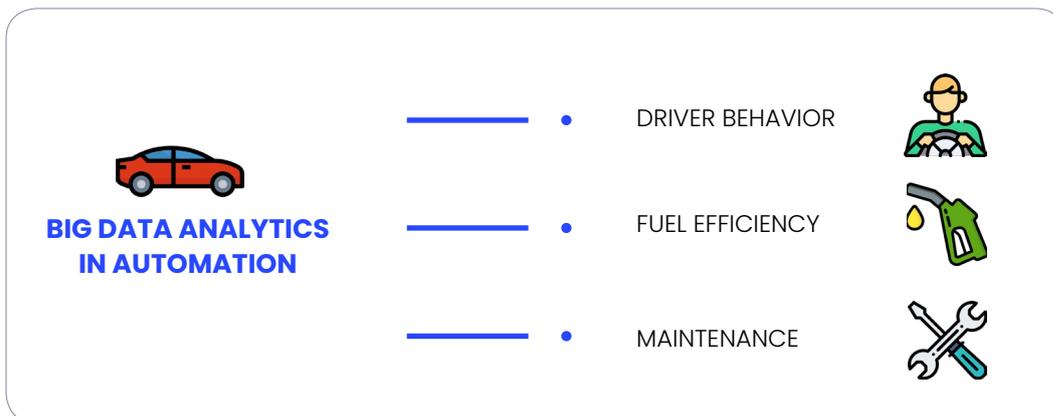
En la industria automotriz, el Big Data ha encontrado una enorme cantidad de aplicaciones, que van desde el diseño y la fabricación de vehículos, hasta el mantenimiento predictivo y la conducción autónoma. Las inversiones en Big Data aplicado al sector automotriz alcanzaron la cifra de 3 300 millones de dólares en 2018, y se espera que aumenten a un ritmo anual del 16% durante los próximos tres años.

La integración de dispositivos WiFi y conexión directa con la red en los automóviles ha abierto la puerta a una verdadera revolución en el sector, pues permite **instalar sensores en las distintas partes del vehículo y recabar millones de datos de forma constante**. Datos que, correctamente analizados, permiten al fabricante realizar diagnósticos remotos de averías, notificaciones automáticas de colisiones, optimizar los sistemas de seguridad, integrar seguros basados en el uso, o incluso fabricar modelos de coches autónomos.



Coche conectado. La instalación de sensores y la conexión a Internet permiten medir el rendimiento del motor y avisar preventivamente al conductor, por ejemplo, ante la necesidad de un cambio de aceite o líquido del radiador. Además, el coche conectado permite extraer patrones sobre el comportamiento del consumidor, especialmente útiles para perfiles publicitarios. Por ejemplo, puede existir un patrón entre el tipo de música que escucha y el tipo de restaurantes que visita, de modo que la publicidad que reciba sea especialmente personalizada.

Por otra parte, la recopilación de grandes cantidades de datos sobre la conducción permite optimizar la gestión de flotas, y la medición del desempeño de cada usuario según su forma de conducción, generando un perfil sobre el cual puede ofrecerse un seguro personalizado.



Conducción autónoma. Gracias a los sensores equipados en los vehículos, es posible medir todo tipo de variables: desde la posición, la velocidad, la dirección y el frenado; a los semáforos, proximidad peatonal y peligros en la carretera. Al hacer uso de estos datos, el vehículo puede tomar sus propias decisiones de conducción de forma más eficiente y segura que cualquier ser humano.

La primera ola de tecnología que apunta en esta dirección ya ha llegado al mercado generalista: sistemas de aparcamiento automático, asistencia en frenado y control de la velocidad, o incluso adelantamientos automáticos en algunos modelos de alta gama. Próximamente, la segunda ola tecnológica revolucionará de nuevo el sector automotriz, con los primeros vehículos totalmente autónomos en el mercado.



Mantenimiento

predictivo. El uso de analítica predictiva para prever problemas en las piezas de los vehículos es uno de los factores que propiciarán en mayor medida el ahorro de costes en el sector automotriz. El hecho de poder detectar posibles averías antes de que estas lleguen a producirse, notificando de forma remota al conductor, constituye la mayor revolución en el área del mantenimiento de automóviles.



Financiamiento de

automóviles. Uso del historial financiero del cliente y de sus preferencias para generar perfiles personalizados que, al combinarse con datos demográficos y geográficos, permiten obtener esquemas de financiación más adecuados, lo que se traducirá en un aumento de la cifra de negocio y, sobre todo, reducirá significativamente los impagos.



Mejoras en la cadena de

suministro. Los fabricantes de automóviles están utilizando el Big Data para comparar los costes, la fiabilidad y la calidad de los distintos componentes dentro de su cadena de suministro y determinar la mejor opción para sus modelos. Además, también utilizan esta tecnología para predecir la demanda, lo que ayuda a racionalizar los procesos de producción y adquisición, haciéndolos más eficientes y rentables.

Estas leyes se resumen en una denominada 0: **un robot no hará daño a la humanidad ni permitirá que, por inacción, esta sufra daño.** EE UU, Japón, China y Corea del Sur ya han comenzado a legislar. En Europa, de momento, se aplica la legislación sobre bienes defectuosos, pero sus normas pueden estar obsoletas para los productos automatizados o autónomos en sus concretas tareas. El propio Parlamento de Estrasburgo lo reconoce al reclamar un marco "adecuado, eficiente, transparente y coherente".

La Comisión Europea evalúa la directiva 85/374/EEC sobre productos defectuosos, la que actualmente se aplica para determinar su validez ante la programación, la computación en la nube, el Internet de las Cosas y los sistemas autónomos, así como sobre quién debe recaer la responsabilidad en relación de los daños causados por el comportamiento de un robot.

En la misma línea se sitúan las recomendaciones europeas, que advierten que, en cualquier caso, la regulación no debe impedir **el desarrollo de la robótica siempre que esta "esté al servicio de la humanidad"**. Es aquí donde el Parlamento Europeo ve en las leyes de Asimov la base del futuro de la robótica, en especial, en la norma 0, que engloba las restantes.



BUSINESS PERFORMANCE ANALYTICS

El Business Performance Analytics (BPA) implica el **uso sistemático de datos y métodos analíticos para la medición y la gestión del rendimiento**. Para toda corporación es una necesidad vital documentar los procesos operativos y, especialmente, hacerlo de extremo a extremo y documentar la cadena de valor en toda la compañía.

Precisamente este es el propósito del BPA, que supera los problemas tradicionales relacionados con los Sistemas de Gestión del Desempeño (PMS), como la sobrecarga de información, la dificultad para establecer relaciones causales entre las variables, y la ausencia de una visión holística de la organización. En este sentido, si bien en un primer momento fue utilizado especialmente por los desarrolladores, hoy en día está orientándose a comprender cómo funciona realmente el negocio desde la perspectiva de la dirección.

Los dos aspectos que están integrando más rápidamente el uso de estos métodos analíticos son la **definición del proceso comercial directamente conectado con las necesidades del usuario** según los perfiles de segmentación, y la toma de decisiones estratégicas. El objetivo actual es ampliar el rango de utilidad, y el 84% de las empresas del Fortune 1000 ya afirman haber lanzado proyectos de Analítica Avanzada destinados a mejorar la precisión y eficiencia de la toma de decisiones en base al BPA.

Además, se espera que la inversión en este tipo de herramientas aumente de los 5 300 millones de dólares en 2018 a los 19 400 en 2019, no solo en Estados Unidos sino a nivel global. El resultado parece estar perfilándose ya en la actualidad: las empresas están logrando reducir los gastos (en un 49,2% de las experiencias con BPA) y crear nuevas vías para la innovación (en un 44,3% de los casos). Todo ello redundará en una optimización de los procesos operacionales y la consiguiente reducción de costes.

Como se ha comentado, la continua mejora de los procesos operacionales es un desafío que requiere sistemas cada vez más complejos, y más los basados en tecnologías como el Big Data. Actualmente, la monitorización de la actividad empresarial, el análisis predictivo y el reconocimiento de patrones de comportamiento son la base de cualquier estrategia de gestión del rendimiento realmente eficiente.

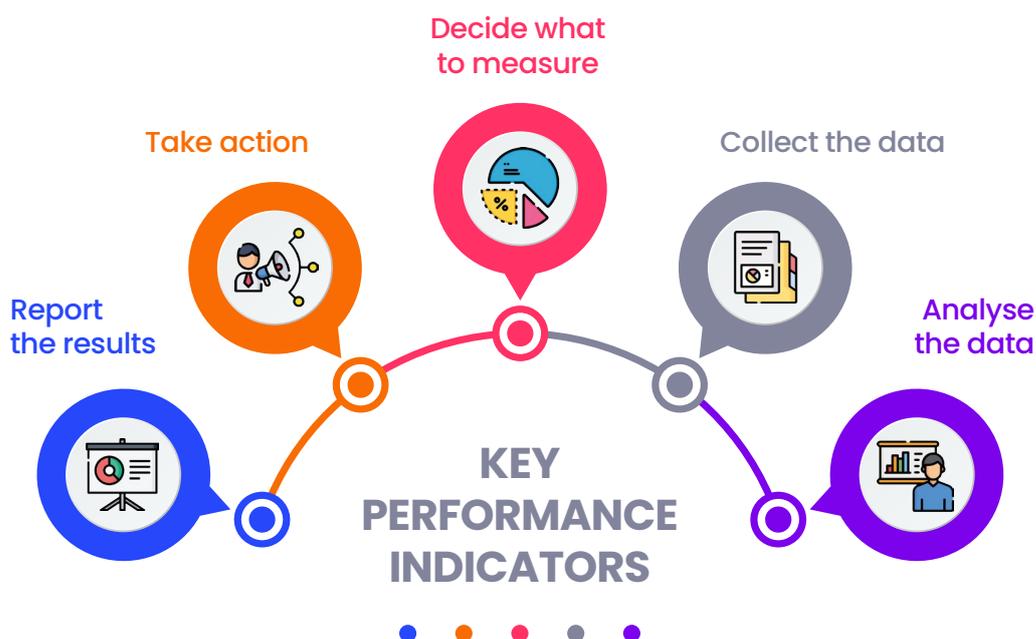
CUSTOMER ANALYTICS: SMART KPI'S

Los indicadores clave de rendimiento (KPI) son indicadores que miden objetivos específicos en empresas de cualquier sector, y se consideran una de las herramientas más importantes que tiene una compañía a su disposición a la hora de monitorizar su propia actividad.

Los KPI's difieren ampliamente en la forma en la que están diseñados y en lo que miden. Esto es crucial para comprender su efectividad y cómo deben implementarse y leerse. Por ejemplo, los indicadores altos pueden medir el desempeño general de un negocio de comercio electrónico, mientras que los indicadores bajos pueden medir las ventas de un producto específico o la producción de un departamento en particular.

Los puntos más susceptibles de monitorización en base a KPI's son los ingresos (beneficios medios, ingresos totales), empleo (rotación de empleados, rendimiento laboral), atención al cliente (tiempo promedio de las llamadas, índice de satisfacción), y Marketing (leads, ventas). Así, parece evidente que el equipo financiero realizará un seguimiento de los flujos de efectivo, del mismo modo que el equipo de ventas medirá la cifra de negocio. Y si bien los distintos departamentos tienen, por lo general, cierto grado de complementariedad, la realidad es que los KPI's que se establecen para cada uno son ciertamente distintos.

El primer problema del sistema clásico de KPI's es que, si bien se muestran útiles para la medición del rendimiento de una determinada área, resulta más complejo adquirir una visión amplia sobre el éxito de una iniciativa o de una empresa. A esta problemática se añade otra: la multiplicación del número de datos que genera cada actividad desde la irrupción de las nuevas tecnologías complica enormemente la labor de definir unos KPI's realmente claros y útiles para conocer el rendimiento empresarial.



En este escenario, surgen los **SMART KPI's, que son indicadores clave de rendimiento generados en base a la analítica de datos**, y que son específicos, medibles, realistas, relevantes y acotados en el tiempo. Estos indicadores, a diferencia de los KPI's tradicionales, permiten generar una serie de indicadores transversales a las distintas unidades profesionales, relacionándose con múltiples fuentes de datos, para obtener unos índices claros que fundamenten una visión holística del rendimiento empresarial.

Por ejemplo, el tráfico de cualquier página Web genera una enorme cantidad de indicadores (sesiones, tasa de rebote, tasa de conversión, páginas vistas, usuarios, etc.). Esta mezcla de información no aporta por sí misma valor a una compañía si su objetivo es monitorizar el rendimiento empresarial. Es necesario construir, sobre el análisis de esos datos, unos indicadores realmente inteligentes.

ACQUISITION - BEHAVIOR-CONVERSION (ABC)

A	B	C
Acquisition KPIs	Behavior KPIs	Conversions KPIs
Visitors	Bounce Rate	Deal Conversion Rate
Percentage of New Visitors	Pages per Visit	Deal Completions
Number of New Visitors	Average Visit Duration	Goal Value

Fuente: Intechnic



Por ejemplo, Google Analytics permite convertir todos esos datos en indicadores clave de rendimiento.

Para ello se basa en lo que denomina ABC: Acquisition (cantidad de tráfico al sitio web), Behavior (nivel de compromiso con el sitio web), y Conversion (conversión de visitantes en clientes). De este modo, lo que antes era una multitud de valores complejos, ahora son tres SMART KPI's que permiten monitorizar el rendimiento.

De forma paralela a estos indicadores inteligentes, han surgido otros medidores, los OKRs (Objetivos y Resultados Clave), que durante los últimos años han alcanzado una enorme popularidad, en parte por su uso por parte de Google. La principal diferencia entre estos y los SMART KPI's, es que mientras los primeros miden los objetivos y el grado de cumplimiento de estos, los segundos monitorizan el rendimiento para conocer el estado real de la compañía, sin vinculación inmediata con los objetivos.

Estos indicadores **permiten mantener a toda la compañía comprometida en una misma dirección, y son claros y específicos**. Sirven como base para una planificación empresarial inteligente, y permiten la adaptación en tiempo real según el grado de cumplimiento o según las variaciones en los KPI's de partida.

BIG DATA TECHNOLOGIES & ARCHITECTURE

Actualmente, las plataformas que han dominado el sector de las nuevas tecnologías durante los últimos 30 años no pueden soportar la carga de trabajo y datos necesarios para impulsar el crecimiento empresarial. Por ello, la reestructuración de la infraestructura TI es una misión realmente compleja, pero que inevitablemente debe afrontarse.

Hasta hoy, las organizaciones han tenido serias dificultades para explotar el potencial de sus datos, debido a la multitud de formatos, las limitaciones de sus BB.DD., la falta de sistemas que combinen insights de múltiples fuentes, etc. Todos estos problemas tienen solución en el marco de las nuevas arquitecturas y tecnologías de la información.

Ignorar el enorme potencial del Big Data para mejorar la toma de decisiones, la captación y fidelización de clientes, y la optimización de los procesos operativos, es una decisión errónea y con importantes consecuencias negativas.

Pero para poder utilizar el análisis de datos con todo su potencial, las empresas necesitan **herramientas adecuadas para almacenar, procesar y analizar la información que producen** diariamente, con el fin de obtener resultados.

Los cuatro segmentos de la infraestructura tecnológica que deben transformarse son la extracción de datos, el almacenamiento, el análisis y la visualización. Cada uno de estos elementos cuenta con herramientas específicas, y una arquitectura de datos concreta.



NoSQL DATABASES

En el caso de las bases de datos, frente al modelo tradicional, basado en sistemas estructurados y relacionales, emerge NoSQL, un modelo simple desde la perspectiva técnica que incrementa enormemente la velocidad de procesamiento y permite el almacenamiento de datos cuantitativos y cualitativos de muy diversa naturaleza, además de establecer entre ellos relaciones multidimensionales. Se calcula que más de un tercio de las compañías que han impulsado un proyecto de Big Data utilizan ya sistemas NoSQL.

ALL IN THE NoSQL FAMILY

Las bases de datos NoSQL están orientadas a gestionar grandes conjuntos de datos variados y actualizados con frecuencia, a menudo en sistemas distribuidos o en la nube. Evitan los esquemas rígidos asociados a las bases de datos relacionales. Pero las arquitecturas en sí varían y se dividen en cuatro clasificaciones principales.



Documentar las bases de datos:

Almacenaje de elementos de datos en estructuras similares a documentos que codifican información en formatos como JSON. Los usos comunes incluyen administración de contenido y monitoreo de aplicaciones web y móviles.

EJEMPLOS: Couchbase Server, CouchDB, MarkLogic, MongoDB



Bases de datos de gráficos:

Enfatizar las conexiones entre los elementos de datos, almacenando "nodos" relacionados en gráficos para acelerar la consulta. Los usos comunes incluyen motores de recomendación y aplicaciones geoespaciales.

EJEMPLOS: Couchbase Server, CouchDB, MarkLogic, MongoDB



Bases de datos de valor clave:

Uso de un modelo de datos simple que combine una clave única y su valor asociado para almacenar elementos de datos. Los usos comunes incluyen el almacenamiento de datos de flujo de clics y registros de aplicaciones.

EJEMPLOS: Aerospike, DynamoDB, Redis, Riak



Wide column stores:

También llamadas bases de datos de estilo de tabla, almacenan datos en tablas que pueden tener un gran número de columnas. Los usos comunes incluyen búsquedas en internet y otras aplicaciones web a gran escala.

EJEMPLOS: Accumulo, Cassandra, HBase, Hypertable, SimpleDB

Fuente: TechTarget

PLATAFORMAS DE TIEMPO REAL

Responder a las necesidades de los clientes en tiempo real es una tarea crítica que repercute cada vez más decisivamente sobre la fidelización. Pero para lograr un objetivo de tal envergadura, son necesarias herramientas de Software capaces de procesar los datos en tiempo real y formular una prescripción basada en el perfil personalizado.

Algunas herramientas populares en este campo son:



Servicio administrado, escalable y basado en Cloud que permite el procesamiento en tiempo real de grandes flujos de datos.



Sistema de mensajería distribuida que integra aplicaciones y flujos de datos de diversas fuentes.



Apache Flink

Motor que proporciona facilidades para el cálculo distribuido sobre grandes flujos de datos.



Sistema para recuperar streams de datos en tiempo real desde múltiples fuentes de manera distribuida, tolerante a fallos y en alta disponibilidad.

DOCKER & CONTAINERS

Los containers generan beneficios significativos para desarrolladores y operadores. El enfoque tradicional para el aislamiento de infraestructuras era la partición estática, la asignación de un segmento de recursos fijo y separado (ya sea un servidor físico o una máquina virtual) para cada carga de trabajo.

En cambio, el gran beneficio de la “tecnología de contenedores” es su capacidad para crear un nuevo tipo de aislamiento. El mismo container puede ejecutarse en hardware básico en un centro de datos local o en una máquina virtual en una Cloud pública, sin necesidad de cambio alguno. Esta es la verdadera movilidad de la carga de trabajo.





Data Analytics & Visualization

02

Campo emergente relacionado con el análisis, el modelado y la visualización de datos complejos. Ante el creciente número de datos y la mayor complejidad de los patrones que de ellos se extraen, surge la necesidad de **clarificar y procesar visualmente esos datos, facilitando la comprensión y la transmisión.**

DATA MANAGEMENT

Debido a la necesidad por parte de las compañías de obtener información en tiempo real, se espera que el mercado global del *Data Management* se expanda en un 12% CAGR llegando a los 146 mil millones de dólares para 2023.

A nivel geográfico el mercado de gestión de datos empresariales está segmentado en cuatro regiones diferentes: Norteamérica, Asia Pacífico, Europa y el resto del mundo. Actualmente se anticipa que

Asia Pacífico será la región de mayor crecimiento en el mercado de gestión de datos empresariales durante los próximos cinco años. Especialmente India, Japón, Corea del Sur y China, que lideran la inversión en este campo.

Por otro lado, las principales tendencias en la gestión de datos implican cambios sustanciales que no obstante todavía pueden tardar años en desarrollarse. La infraestructura actual para almacenar datos es, en la

mayoría de las empresas, extraordinariamente ineficiente.

Por ello es fundamental la **búsqueda de métodos alternativos de almacenamiento**, ya que las corporaciones están en posesión de más datos que nunca (se producen diariamente 2,5 quintillones de bytes). Esos datos no pueden ser aprovechados para obtener información que revierta en ingresos o ahorro de costes sin un adecuado almacenamiento.

Al margen de las limitaciones técnicas, también existen riesgos de privacidad cuando se confía en un servicio con información (especialmente la que es confidencial). Algunas de las modificaciones que están realizándose en las infraestructuras de datos de las empresas son:



MODERNIZACIÓN DE HADOOP

Para obtener más valor de Hadoop, muchos usuarios están modernizando ciertos aspectos de sus implementaciones.

Por ejemplo, para reducir el coste y simplificar el mantenimiento de los clústeres en las instalaciones, algunas organizaciones están migrando a Hadoop a Cloud.

De manera similar, se seguirán utilizando las herramientas de Hadoop del ecosistema de Apache, pero cambiarán el sistema de archivos distribuidos de Hadoop por algún tipo de almacenamiento de objetos basado en Cloud.

Una modernización popular entre todos los usuarios de Hadoop (ya sea para casos de uso analíticos o no) es migrar de *MapReduce* a *Spark* para una mejor gestión de recursos, procesamiento de microparches y rendimiento de baja latencia.

EL REMIX DEL DATA LAKE

Para alojar los datos de la mejor forma posible, la tendencia con los Data Lakes (al igual que con los almacenes de datos) es **avanzar hacia una arquitectura multiplataforma**.

Ahí es donde el "Lake" se distribuye físicamente entre

los clústeres de Hadoop y las bases de datos relacionales, que pueden implementarse en las instalaciones, en las plataformas Cloud o en una combinación híbrida de ambas.

En ese sentido, los propietarios de Data Lakes están revisando su elección de plataformas de datos para profundizar en el

llamado *Data Lake híbrido*.

Para hacer que los datos dispares y diversos parezcan más simples de lo que son, las técnicas de virtualización son útiles (incluso requeridas), por lo que un *Data Lake híbrido* es también un *Data Lake virtual*, similar a un almacén de datos virtual.

EL CATÁLOGO DE DATOS

En general, el término “metadato” hace referencia a los metadatos técnicos que las herramientas utilizan para interactuar con gran variedad de aplicaciones y plataformas de datos. Como complemento, la catalogación de datos generalmente describe las características de los mismos, como el dominio, la calidad, el linaje y las estadísticas de perfil de un conjunto específico.

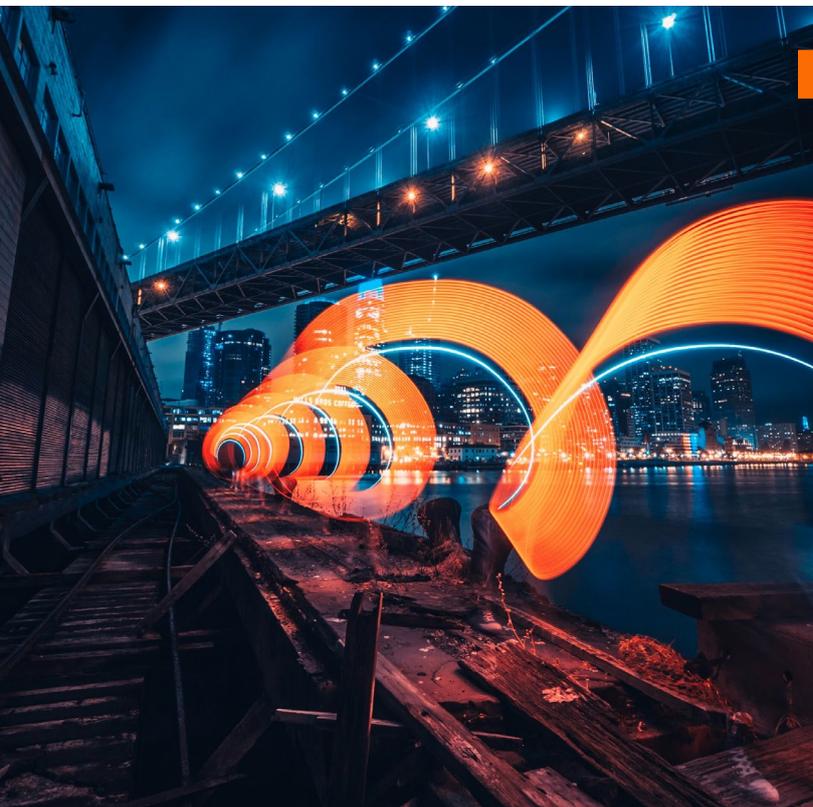
Los rasgos pueden ser subjetivos, como cuando los usuarios etiquetan datos para calificar su nivel de confianza, sensibilidad de cumplimiento o facilidad de uso. De esta manera, **los analistas y usuarios similares pueden enfocar sus búsquedas y consultas** por dominio (por ejemplo, solo mirar los datos de los clientes), el riesgo de cumplimiento y por la confiabilidad o utilidad (según la calificación de otros usuarios).



IOT COMO LA PRÓXIMA OLA DE BIG DATA

El auge del paradigma del IoT permite a una organización monitorizar, medir y ajustar una larga lista de procesos, clientes, instalaciones, etc., con niveles de detalle, precisión y velocidad sin precedentes. Sin embargo, este planteamiento requiere de una inversión considerable en tecnologías de la información, con un claro foco en una moderna infraestructura de administración de datos.

El desafío es crear soluciones que se puedan **escalar mientras se captura, se procesa y se responde** a los datos en tiempo real.



DISTRIBUYENDO DATOS CON BLOCKCHAIN

Para almacenar datos en este contexto de rápido desarrollo, es imperativo que se creen bases de datos más resistentes y eficientes. Deben ser entornos altamente seguros y expertos en atender las necesidades de las aplicaciones en los campos del IoT y de la IA. Para este propósito, la tecnología Blockchain es la solución ideal.

En lugar de ser ejecutado por una entidad como Amazon o Google, la integridad de Blockchain está asegurada por nodos en la red que sincronizan copias de la base de datos. Desde un punto de vista de seguridad, este modelo es increíblemente difícil de comprometer, ya que requiere que una parte obtenga el control de la mayoría de los nodos para poder alterar las entradas en el libro de registros mayor.

Dado que los nodos están distribuidos y funcionan de igual a igual, la posibilidad de que se formen cuellos de botella es inexistente. Sin embargo, una de las características más importantes de los sistemas Blockchain es la inmutabilidad: una vez que se agrega una entrada a la base de datos, no se puede eliminar.

*How
Blockchain
Can Solve
The World 's
Coming Data
Management
Problem.*



Fuente: Youtube

En definitiva, existe un tema común a todas las tendencias mencionadas: la necesidad de una infraestructura que contenga mucha más información de la que estamos acostumbrados a almacenar. Las bases de datos arcaicas que se han utilizado hasta ahora resultan claramente insuficientes. Es hora de utilizar nuevas tecnologías para crear nuevos métodos, **bases de datos más fuertes, más seguras, más rápidas y escalables** para impulsar verdaderamente la innovación. Y el éxito de dicha transformación pasa necesariamente por la inversión en tecnología y la captación de expertos, así como por contar con empresas externas que puedan aportar la experiencia necesaria.



DATA GOVERNANCE

La capacidad de la empresa para sacar provecho de los procesos impulsados por los datos, para cosechar el rendimiento de los datos como un activo organizativo, depende del Data Governance, que sustenta la base de la gestión de datos *en sí*. Sin embargo, el Data Governance 1.0 ha dejado de ser efectivo, por ello, los profesionales se preguntan cuáles son sus límites y en qué consiste el Data Governance 2.0.

Los componentes del Data Governance (Stewardship, Lineage, Data Quality, etc.) son necesarios para poder ver más allá de la reducida visión de una única unidad de negocio, así como para **desbloquear el valor estratégico de la toma de decisiones basadas en datos** como un activo organizativo. Las organizaciones deben adoptar el Data Governance 2.0, centrado en la operacionalización y la solución de problemas de negocio.

Además, el concepto clásico de Data Governance debe cambiar para garantizar que la visión de los profesionales se refleje en la sostenibilidad. Sus componentes deben integrar como socios del ecosistema tanto a los productores de datos como a los consumidores de los mismos. En resumen, el Data Governance permite a una organización una mejor gestión de varios aspectos de los datos: la disponibilidad, la facilidad de uso, la integridad y la seguridad.



Data Governance 1.0

- _ DM Policy & Standards
- _ Driving Controls
- _ Enterprise Glossary Metadata
- _ Enterprise-Driven



Data Governance 2.0

- _ Operationalization
- _ Collaborative
- _ Driving Innovation
- _ Business Solutions-Driven Metadata

Data Governance 1.0 > 2.0

Evolving Data Governance from a perception of a "Policy & Compliance" function to delivering tangible value and innovation to the business

En el caso más importante de una tecnología basada en la analítica de datos, la Inteligencia Artificial, los proyectos no deberían iniciarse hasta contar con datos precisos y adecuados.

Al utilizar las tecnologías adecuadas, el Data Governance puede **impulsar los valores comerciales y respaldar la transformación de la IA**, incluyendo una mayor seguridad, la prevención de pérdida, la integridad y el linaje de los datos.

Para que la calidad de los datos sea efectiva, las políticas y la metodología deben estar definidas previamente, y todos los servicios de datos o mecanismos de recopilación deben hacer cumplir estos principios. Los perfiles deben mantenerse actualizados, y deben abordarse las incoherencias entre los sistemas y los protocolos para comunicar los cambios entre los propietarios de los datos.



VISUALIZACIÓN DE DATOS

La visualización de datos es la presentación de datos en un formato pictórico o gráfico, para que los profesionales puedan evaluar los análisis presentados de forma visual, y comprender conceptos innovadores o identificar nuevos patrones. Con la visualización interactiva, el concepto va un paso más allá, utilizando la tecnología para profundizar en tablas y gráficos con más detalles, cambiando interactivamente los datos según se ven y se procesan.

Independientemente de la industria o el tamaño, todos los tipos de empresas pueden utilizar la visualización para:



Comprender información rápidamente

Al utilizar representaciones gráficas de información comercial, pueden verse grandes cantidades de datos de manera clara y cohesiva, y sacar conclusiones de esa información. Dado que es significativamente más rápido analizar la información en formato gráfico (en lugar de analizarla en hojas de cálculo), las empresas pueden abordar los problemas o responder preguntas de manera más oportuna.



Identificar relaciones y patrones

Es posible reconocer parámetros altamente correlacionados a partir de grandes cantidades de datos complejos si se presentan gráficamente. La identificación de esas relaciones ayuda a las organizaciones a centrarse en las áreas con mayor probabilidad de influencia en sus objetivos más importantes.



Identificar tendencias emergentes

El uso de la visualización de datos para descubrir tendencias, tanto en el negocio como en el mercado, puede aportar una ventaja sobre la competencia. Es fácil detectar los valores atípicos que afectan la calidad del producto o la rotación de clientes, y abordar los problemas antes de que adquieran gravedad.



Comunicarse con el entorno

Una vez descubiertas nuevas perspectivas gracias al análisis visual, el siguiente paso es comunicar esas percepciones a los demás. Los cuadros, gráficos u otras representaciones de datos visualmente impactantes transmiten el mensaje rápidamente y de forma atractiva.

GUIDED VISUALIZATIONS FOR CHARTS AND GRAPHICS

WHAT WOULD YOU LIKE TO SHOW?

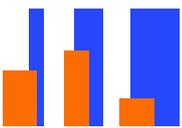
COMPARISON

Among Items

Over Items

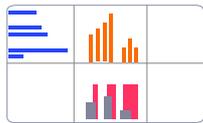
Two Variables per Item

Variable Width Column Chart



One Variable per Item

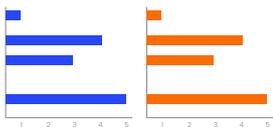
Many categories
Table or Table with Embedded Charts



Few Categories

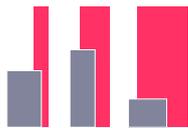
Many Items

Bar Chart



Few Items

Column Chart



Many Periods

Cyclical Data

Circular Area Chart



Non - Cyclical Data

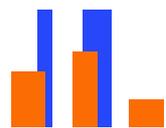
Line Chart



Few Periods

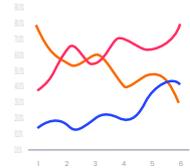
Single or Few Categories

Column Chart



Many Categories

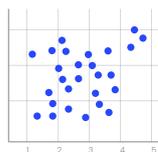
Line Chart



RELATIONSHIP

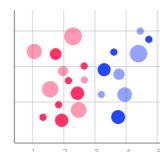
Two Variables

Scatter Chart



Three Variables

Bubble Chart



GUIDED VISUALIZATIONS FOR CHARTS AND GRAPHICS

WHAT WOULD YOU LIKE TO SHOW?

COMPOSITION

Changing Over Time

Static

Many Periods

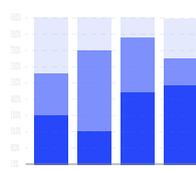
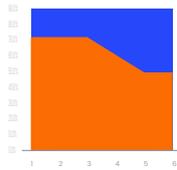
Few Periods

Only Relative Differences Matter

Only Relative Differences Matter

Stacked 100% Area Chart

Stacked 100% Column Chart

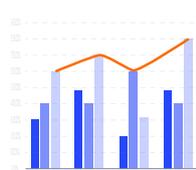
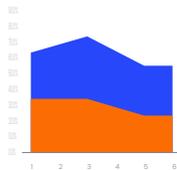


Relative and Absolute Differences Matter

Relative and Absolute Differences Matter

Stacked Area Chart

Stacked Column Chart



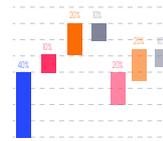
Simple Share of Total

Pie Chart



Accumulation or Subtract to Total

Waterfall Chart



Components of Components

Stacked 100% Column Chart with Subcomponents



DISTRIBUTION

Single Variables

Two Variables

Three Variables

Many Data Points

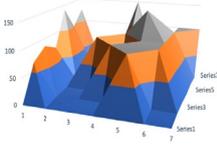
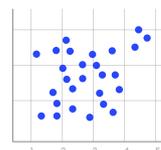
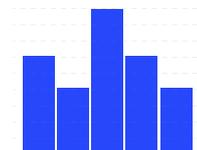
Few Data Points

Scatter Chart

3D Area Chart

Line Histogram

Column Histogram





Hasta este momento, las empresas y organizaciones se han centrado cada vez más en los datos que pueden obtener de sus clientes y servicios, realizando un seguimiento de casi todas las áreas empresariales.

Esto ha generado una oportunidad de negocio para los proveedores de herramientas de visualización de datos, que han proliferado ampliamente, como por ejemplo Tableau, Alteryx, Infogram, ChartBlocks, Datawrapper, Plotly, Visual.ly, etc.

Entre estas herramientas destacan Tableau y Alteryx, pues son consideradas un estándar de la industria y han dominado el mercado de análisis de datos desde hace algunos años.

Tableau ayuda a los analistas a llevar a cabo actividades de inteligencia de negocio y visualización de datos.

Los usuarios pueden generar paneles de control y visualizaciones de datos extremadamente visuales. Su interfaz interactiva ayuda a generar informes rápidamente donde es posible descender hasta un nivel muy profundo de detalle.



What is Tableau?

Fuente: Youtube





Alteryx, por su parte, es una herramienta utilizada en el análisis de datos que también brinda información significativa al personal de nivel ejecutivo.

Con una interfaz sencilla, el usuario puede extraer los datos, transformarlos y cargarlos en la herramienta.

La combinación de varias de estas herramientas puede brindar un servicio de alta calidad a las empresas, y facilita que los altos ejecutivos puedan tomar decisiones basadas en el conocimiento que proporcionan los datos.

En cualquier caso, es evidente que la visualización de los datos tiene el poder de cambiar el futuro y e influir decisivamente en la forma en la que se actúa a nivel empresarial.



The Alteryx Platform

Fuente: Youtube

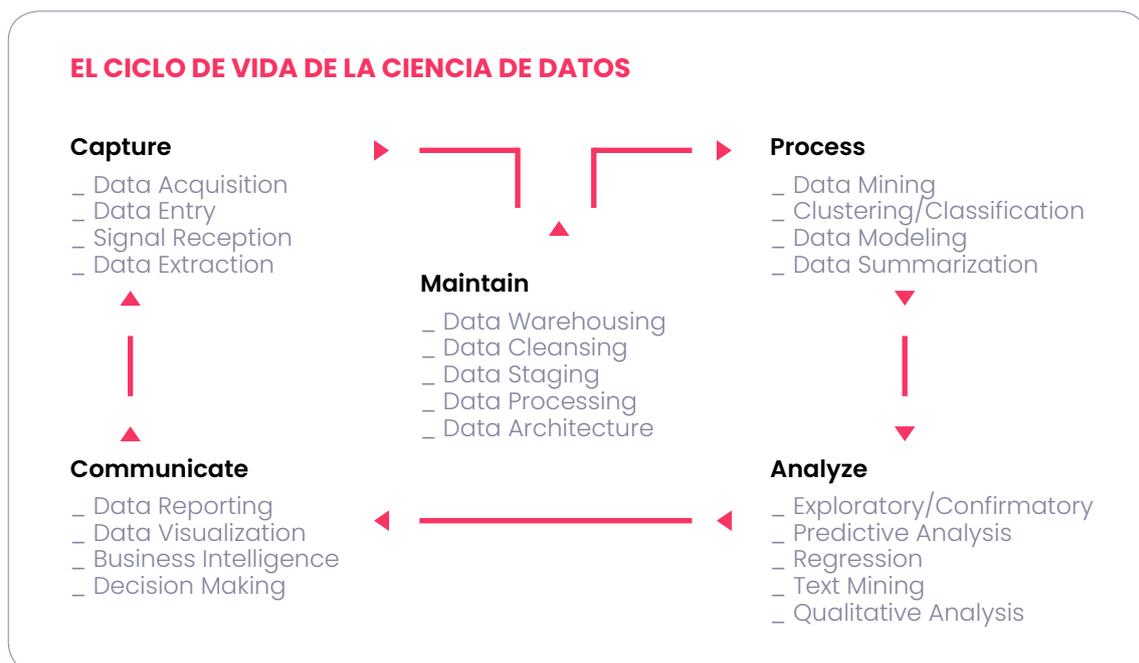


Data Science

03

Para impulsar la inteligencia empresarial, los Data Scientists deben dominar todo el espectro del ciclo de vida de los datos, así como poseer un alto nivel de flexibilidad y comprensión para maximizar el rendimiento en cada fase del proceso. Deben identificar preguntas relevantes, recopilar datos de multitud de fuentes diferentes, organizar la información, traducir los resultados en soluciones y comunicar sus hallazgos de una manera que afecte positivamente a las decisiones comerciales. Estas habilidades son necesarias en casi todas las industrias, lo que hace que los científicos expertos en datos sean cada vez más valiosos para las empresas.

En cualquier caso, dentro de la cadena de tratamiento de datos, existen tres procesos a tener especialmente en cuenta: **Data Mining, Analítica Avanzada y Deep Learning**.



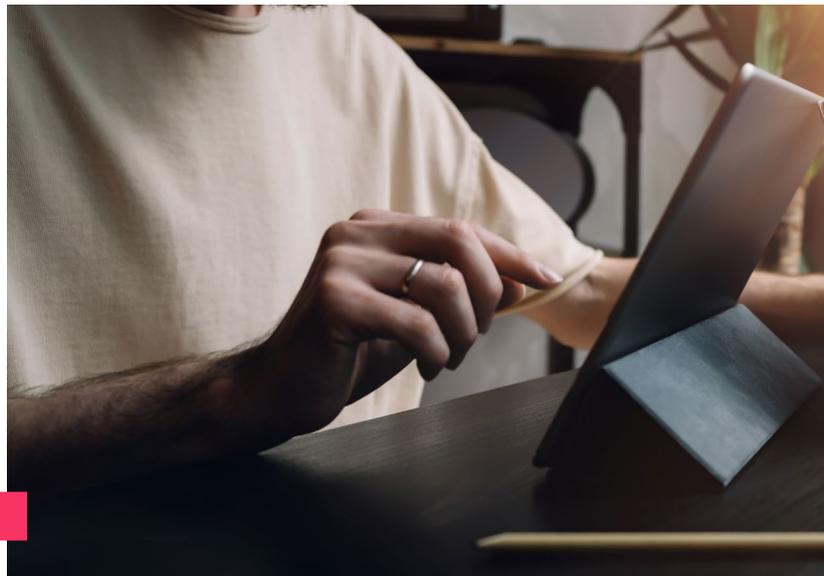
DATA MINING

El Data Mining tiene como finalidad revelar información que de otra forma no podría ser encontrada. Es una de las técnicas más empleadas hoy en día en numerosos ámbitos, sobre todo dentro del Marketing Digital.

Como en todo proceso, la minería de datos debe llevarse a cabo en diferentes fases: comprensión de lo buscado y del problema a resolver; determinación, captación y limpieza de los datos necesitados; creación de los modelos matemáticos; y validación y comunicación de los resultados.

Esta tecnología ha abierto todo un mundo de posibilidades para los negocios, ya que relaciona millones de datos aislados y sirve a las empresas, por ejemplo, para **detectar patrones de conducta en sus clientes y predecir su comportamiento**. Su objetivo principal es generar nuevas oportunidades de mercado.

La capacidad predictiva de la minería de datos ha cambiado el diseño de las estrategias empresariales, ya permite entender el presente para anticiparse al futuro. Algunos ejemplos de aplicación sectorial de esta técnica en la actualidad son los siguientes:



MARKETING

Analizando las relaciones entre parámetros como la edad de los clientes, el género, los gustos, etc., es posible adivinar su comportamiento para dirigir campañas personalizadas de fidelización o captación. El *Data Mining* también predice qué usuarios pueden darse de baja de un servicio, qué les interesa según sus búsquedas o qué debe incluir una lista de correo para lograr una tasa de respuesta mayor.

COMERCIO MINORISTA

Los supermercados, por ejemplo, emplean los patrones de compra conjunta para identificar asociaciones de productos y decidir cómo situarlos en los diferentes pasillos y estanterías de los lineales. El *Data Mining* detecta además qué ofertas son las más valoradas por los clientes e incrementa la venta en la cola de caja.



TELEVISIÓN Y RADIO

Multitud de medios de comunicación aplican la minería de datos a sus registros de audiencia en televisión online (IPTV) y radio. Estos sistemas recaban y analizan en tiempo real información anónima de las visualizaciones, las retransmisiones y la programación de los canales. Gracias al *Data Mining* se pueden emitir recomendaciones personalizadas a los radioyentes y telespectadores, conocer en directo sus intereses y su actividad, y entender mejor su conducta. Las cadenas obtienen además, conocimiento muy valioso para sus anunciantes, que aprovechan estos datos para llegar con más precisión a sus clientes potenciales.

BANCA

Los bancos recurren a la minería de datos para entender mejor los riesgos del mercado. Es habitual que se aplique a la calificación crediticia (rating) y a sistemas inteligentes antifraude para analizar transacciones, movimientos de tarjetas, patrones de compra y datos financieros de los clientes.

El *Data Mining* también **permite a la banca conocer más sobre las preferencias o hábitos** en Internet de los usuarios para optimizar el retorno de sus campañas de Marketing, estudiar el rendimiento de los canales de venta o gestionar las obligaciones de cumplimiento de las regulaciones.



MEDICINA

La minería de datos favorece diagnósticos más precisos. Al contar con toda la información del paciente —historial, examen físico y patrones de terapias anteriores— se pueden prescribir tratamientos más efectivos. También posibilita una gestión más eficaz, eficiente y económica de los recursos sanitarios al identificar riesgos, predecir enfermedades en ciertos segmentos de la población o pronosticar la duración del ingreso hospitalario.

Detectar fraudes e irregularidades y estrechar vínculos con los pacientes al ahondar en el conocimiento de sus necesidades son también ventajas de emplear el *Data Mining* en este sector.

Aunque esta tecnología ya se está utilizando, su potencial no ha sido explorado aun por completo. Esto se debe en gran medida, a la existencia de obstáculos que dificultan el desarrollo tecnológico y la adopción por parte de las empresas.

Entre esos retos, podemos destacar:

- **Mala calidad de la recopilación de datos:** datos ruidosos y sucios, valores de datos mal colocados, inexactos o incorrectos, volumen insuficiente y representación deficiente en el muestreo.
- **Integración redundante de datos de varias fuentes no marcadas:** pueden estar representados de varias formas: datos numéricos, archivos multimedia, datos de interacción social o datos de ubicación geográfica.
- Proliferación de problemas de seguridad y preocupación por la privacidad por parte de organizaciones privadas y gubernamentales y por parte de los usuarios.
- Datos que exceden los límites estáticos, son sensibles a los costes o simplemente no están equilibrados.
- Dificultad para acceder a diferentes tipos de datos y su falta de disponibilidad.
- En función de la cantidad de bases de datos existentes se puede tardar bastante tiempo en procesar toda la información. En estas situaciones surgen desafíos relativos a la mano de obra, al consumo del tiempo y a los resultados financieros.
- El análisis de grandes conjuntos de datos requiere en ciertos momentos de diferentes métodos de distribución.
- El alto coste del software de recolección de datos y del hardware utilizado para acumular y organizar grandes cantidades de datos de diferentes sectores.



A pesar de estos retos, algunas de las organizaciones más grandes del mundo ya están **utilizando la minería de datos para aumentar sus ingresos y disminuir sus costes**, y para relacionarse de un modo más eficaz con los clientes.

The Amazon logo, consisting of the word "amazon" in a lowercase, sans-serif font, with a curved orange arrow underneath it that points from the letter 'a' to the letter 'z'.

Como por ejemplo Amazon, con su sistema de recomendaciones personalizadas en base a las preferencias de compra de usuarios con perfiles o conductas similares.

ANALÍTICA AVANZADA

Las técnicas de Analítica Avanzada permiten mejorar y optimizar la toma de decisiones, minimizar los costes y maximizar la eficiencia de los procesos, dado que el **análisis de datos en tiempo real permite obtener prescripciones, frente a los modelos meramente descriptivos.**

Esta tendencia afectará profundamente a la mayor parte de los sectores empresariales, aunque todavía está en fase de adopción en función de la industria: por ejemplo, menos del 30% de empresas del sector financiero utiliza actualmente sistemas de este tipo, pero se erige como una prioridad de inversión para el 87% de las entidades aseguradoras.

El potencial es enorme: la mejora del margen bruto de las compañías industriales podría ser del 40% gracias a la integración de la Analítica Avanzada.



LA ANALÍTICA AVANZADA EN LOS DIFERENTES SECTORES DE ACTIVIDAD

Sector Público

Sólo el 10% de las instituciones públicas como educación o sanidad **utilizan técnicas analíticas** avanzadas para apoyar su toma de decisiones.

Turismo

Usar analítica avanzada en hostelería y restauración **reduce los costes entre un 5 y un 20%** de la facturación total.

Retail

Dentro de los próximos tres años, **más de 3/4 de las empresas** del sector retail **piensan invertir** en soluciones de analítica avanzada.

Industria

El uso de la analítica avanzada puede suponer **mejores de hasta un 40%** en el margen bruto total de la compañía.



Banca

Menos del **30% de las empresas** del sector financiero **utiliza actualmente** analítica avanzada de datos.

Seguros

La analítica avanzada ya es una **prioridad de inversión para el 87%** de los ejecutivos del sector seguros.

Energía

El uso de analítica avanzada podría **ahorrar hasta un 30%** en **emisiones y costes energéticos**.

Logística y transportes

El **43% de los expertos** del sector creen que el uso de analítica avanzada será la **principal vía de cambio y mejora** en los próximos años.

La Analítica Avanzada es la fusión de la Analítica Descriptiva, Predictiva y Prescriptiva.

La primera muestra las causas de la configuración actual de un determinado proceso, ofreciendo una imagen general de lo que ocurre en la compañía, y estableciendo la cadena de causalidad entre las distintas variables.

En segundo lugar, la Analítica Predictiva permite transformar esas métricas descriptivas en un conjunto de previsiones, pronósticos y tendencias muy precisas, que anticipan lo que sucederá en relación con un proyecto o empresa en el futuro. La mayor parte de empresas que exploran la analítica de datos, alcanzan este nivel y utilizan las previsiones para fundamentar su toma de decisiones.

Pero las técnicas de Analítica Avanzada permiten dar un paso más. Si a estos pronósticos realizados gracias a la analítica predictiva, se aplica analítica prescriptiva, se pueden identificar las decisiones óptimas a tomar teniendo en cuenta los grandes volúmenes de datos y las infinitas variables y restricciones existentes, automatizando así el proceso de toma de decisiones.

DEEP LEARNING: MODELOS PREDICTIVOS

El *Deep Learning* corresponde al tercer nivel (después de la Inteligencia Artificial y el Machine Learning) en la evolución de los sistemas de IA, los cuales se acercan cada vez más al reto de conseguir sistemas informáticos que de forma autónoma y sin apenas intervención humana sean capaces de imitar comportamientos y razonamientos propios del ser humano.



Se calcula que el mercado del Deep Learning superará los 18.000 millones de dólares para 2024, con un CAGR del 42%.

Sus algoritmos tienen un enorme potencial y recogen datos como vídeos, imágenes, grabaciones de audio y texto para hacer predicciones aplicables, entre otros, al ámbito empresarial.

Se trata de una realidad especialmente palpable en sistemas como el asistente virtual de Amazon, que utiliza el aprendizaje profundo para comprender el lenguaje humano, o el propio traductor de Google.

El Deep Learning, por el momento, es lo más parecido que existe al pensamiento humano.

De hecho, estas redes neuronales son capaces de generar **procesos que incluyen el pensamiento abstracto** (lenguaje propio y razonado) para el reconocimiento de patrones u objetos.

De hecho, en un futuro no muy lejano, será posible dotar a estos sistemas de capacidad creativa.

LAS APLICACIONES DEL *DEEP LEARNING* SON MUY DIVERSAS SEGÚN LOS DISTINTOS SECTORES:

- **Conducción Autónoma:**

detección automática de señales, peatones y semáforos, en sus múltiples formas y ubicaciones.

- **Sector Aeroespacial y de Defensa:**

identificación de objetos y zonas desde satélites, midiendo, por ejemplo, el nivel de seguridad de una determinada área.

- **Investigación Médica:**

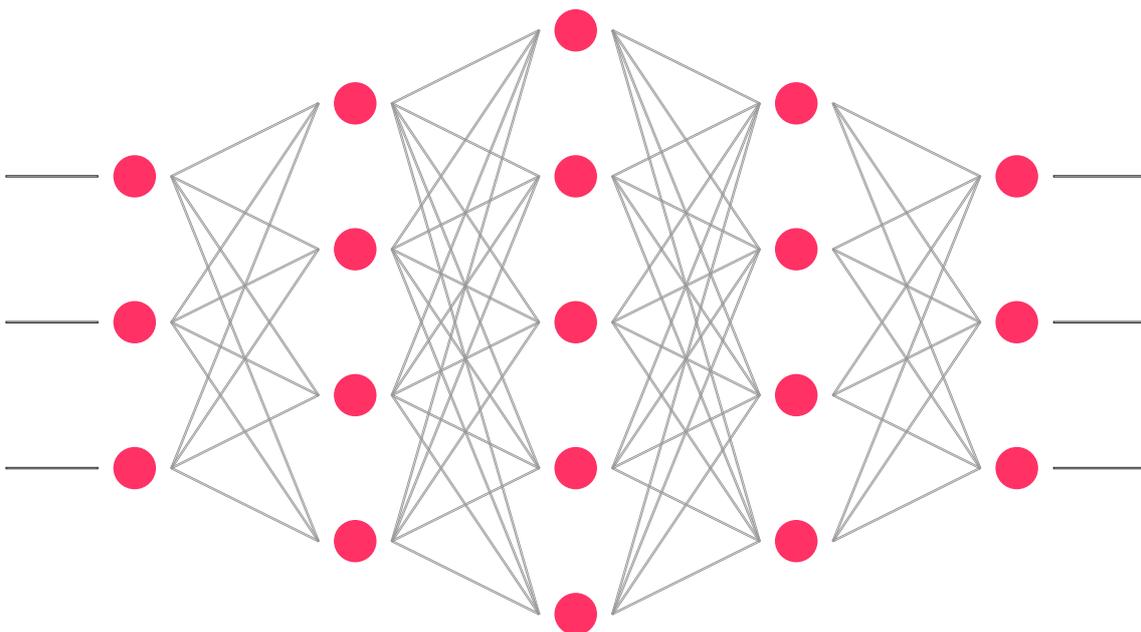
detección de células cancerígenas de forma inteligente. Por ejemplo, profesionales de la UCLA han construido un microscopio avanzado que produce un conjunto de datos multidimensional empleado para entrenar una aplicación a fin de identificar con precisión dichas células.

- **Automatización Industrial:**

el *Deep Learning* está ayudando a mejorar la seguridad de los trabajadores en entornos con maquinaria pesada, gracias a la detección automática de personas u objetos cuando se encuentran a una distancia no segura de las máquinas.

- **Electrónica (CES):**

audición automatizada y traducción del habla. Por ejemplo, los dispositivos de asistencia doméstica que responden a la voz y conocen las preferencias de los usuarios.



Asimismo, entre las principales tendencias que durante este 2019 marcarán la evolución del *Deep Learning*, y que tendrán un impacto significativo sobre el sector empresarial, podemos observar las siguientes:



Sesgo que influye en la IA

El sesgo es por ahora una característica inherente a prácticamente cualquier decisión. Por supuesto, se da en el caso de las decisiones humanas, pero también en las basadas en aprendizaje profundo, cuyo sesgo compromete en ocasiones la integridad de los datos. Acaba por repercutir negativamente en las aplicaciones reales, incluidos los algoritmos de reconocimiento facial o las campañas de Marketing basadas en perfiles especializados, por ejemplo. Por ello, será uno de los principales campos en los que se centrarán los expertos durante los próximos años.



Hacer de la IA una realidad

Los robots basados en el aprendizaje profundo son el futuro, pero todavía queda mucho para llegar al estadio de madurez. Los inversores lo saben y algunos de ellos prefieren dirigir sus inversiones a escenarios más plausibles a corto plazo. Para evitar que esto suceda, se debe comprender que la Inteligencia Artificial, el *Machine Learning* y el *Deep Learning* forman ya parte de nuestra vida cotidiana, y que ya son fuente de rentabilidad a día de hoy.



Retos legales

La adaptabilidad de la IA en las industrias estrictamente reguladas es un impedimento importante, y los expertos coinciden en señalar que en 2019 se abrirán vías de investigación para ofrecer marcos regulatorios más favorables para la adopción de este tipo de innovaciones tecnológicas.



Capacidades de adopción en Cloud

Este 2019 será el año en el que las empresas buscarán mejoras en sus infraestructuras tecnológicas, y migrarán a servicios de alojamiento en Cloud, mucho más adecuados para sostener la adopción del *Deep Learning*. A medida que las empresas comiencen a experimentar con estos sistemas, aumentará la demanda de herramientas y técnicas más personalizadas para el soporte Cloud, y para casos de aprendizaje profundo complejos y muy específicos, como la fusión de imágenes de satélites con datos financieros para mejorar las capacidades comerciales, por ejemplo.

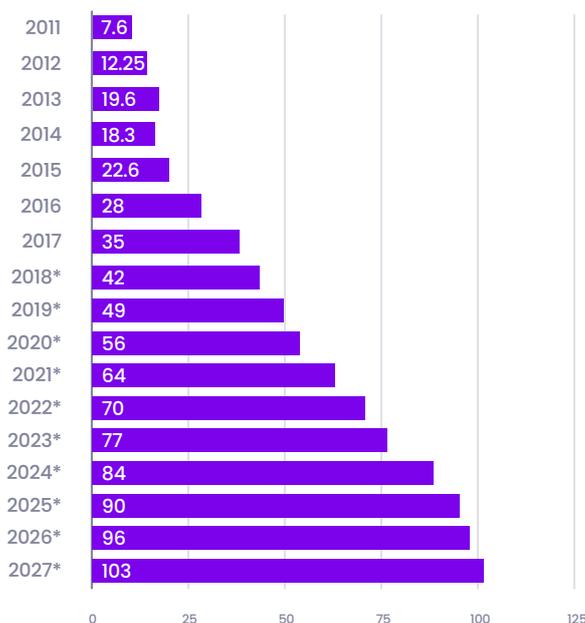


Conclusiones 04

EL ANÁLISIS DE DATOS SERÁ UNO DE LOS PRINCIPALES PILARES TANTO DE LAS TRANSFORMACIONES TECNOLÓGICAS COMO DE LA EVOLUCIÓN DE LOS MODELOS DE NEGOCIO TRADICIONALES DURANTE LAS PRÓXIMAS DÉCADAS. EL VALOR QUE LOS DATOS PUEDEN APORTAR, SI SON CORRECTAMENTE CAPTADOS, ALMACENADOS Y ANALIZADOS, ES COMPARABLE AL IMPACTO DE LA SEGUNDA O TERCERA REVOLUCIÓN INDUSTRIAL.

El tamaño del mercado del Big data ha crecido, desde los 7,6 mil millones de dólares en el año 2011, hasta los 42 mil millones en 2018. Estamos hablando de una cifra de negocio que se ha quintuplicado en apenas siete años, con un crecimiento sostenido y de enorme volumen con pocos precedentes en la historia económica reciente.

BIG DATA MARKET SIZE REVENUE FORECAST WORLDWIDE FROM 2011 TO 2027 (IN BILLION U.S. DOLLARS)



Additional Information

Worldwide; Wikibon; 2014 to 2018

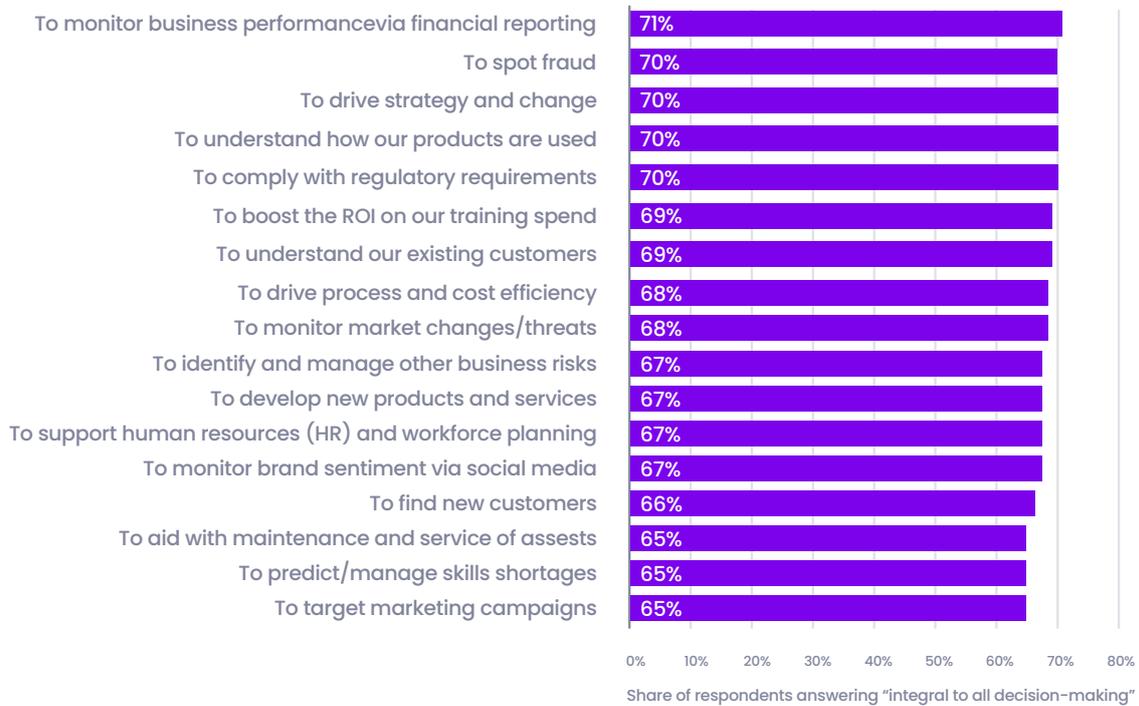
Source

Wikibon; SiliconANGLE
© Statista 2019

Pero las previsiones indican que en 2027, el tamaño de este mercado podría alcanzar los 103 mil millones de dólares, lo que supone más del doble del tamaño actual. Este dato permite poner en contexto el futuro potencial de esta tendencia tecnológica y de negocio, cuya implementación todavía está incompleta y fragmentada en la mayoría de los mercados, y en la que aun existen múltiples vías abiertas de investigación y desarrollo.

El análisis de datos, como cualquier tendencia tecnológica novedosa que tiene éxito, ha superado diversas fases de expansión: ha sido ampliamente conceptualizada; es **un pilar estratégico fundamental en la mayoría de grandes empresas, existen múltiples herramientas tecnológicas avanzadas que facilitan su adopción, y los casos de éxito se cuentan por centenares.**

INTEGRATION OF DATA AND ANALYTICS IN AREAS OF BUSINESS DECISION-MAKING WORLDWIDE, AS OF JULY 2016



Additional Information

Worldwide; Forrester Research; July 2016; 2,165 respondents;

Decision-makers responsible for setting strategy for, or management of, business intelligence, data analytics, data warehousing, data management/big data management initiatives

Source

Forrester Research © Statista 2018



En definitiva, el objetivo que se debe perseguir no es integrar la analítica de datos en un departamento particular o para optimizar una función concreta; sino implantar un modelo data-driven que abarque todas las áreas de la organización y que suponga un beneficio claro para la cadena de valor, la estrategia empresarial y que optimice los procesos operativos y los modelos de negocio.

Por ello, el verdadero reto de las empresas es transformarse en *data-driven companies*, aprovechando el potencial de la información y de los datos en toda su organización. Para lograr este cometido, es esencial diseñar planes de transformación integral que hayan sido analizados por expertos internos y externos, y dotarlos de los recursos económicos, tecnológicos y humanos necesarios.

El futuro inmediato pasa por comprender la profundidad de esta revolución tecnológica y transformar integralmente la compañía, maximizando su rendimiento, constituyéndose como *early adopter* y generando así una ventaja competitiva verdaderamente significativa; o bien adoptar una política reactiva, aceptando el cambio tan pronto como este se imponga como urgente e inevitable. Algo que, parece, no está muy lejos de suceder.



**Somos expertos en servicios,
productos y soluciones digitales,
en innovación tecnológica y en
procesos ágiles de transformación.**

Con un equipo humano de **más de 2.500 profesionales**, tenemos presencia a nivel nacional e internacional, con sedes en EE. UU., Perú, Brasil, Chile, México, Colombia, Paraguay, Reino Unido y Alemania.

¿Nuestra misión? Generar el mayor valor a nuestros clientes, a los sectores económicos y al conjunto de la sociedad, mediante el diseño y desarrollo de iniciativas **basadas en tecnología de vanguardia**, con el objetivo de accionar el cambio digital de una manera disruptiva.

www.vectoritcgroup.com

