



Novedades en la ciencia de los datos

In this e-guide

- Seis razones por las que puede necesitar ciencia de datos como servicio
- Las habilidades de ciencia de datos con más demanda que necesitará
- Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos
- Obtener más contenido exclusivo de PRO+

La ciencia de datos es el estudio de dónde proviene la información, qué representa y cómo se puede convertir en un recurso valioso para la creación de estrategias empresariales y de TI. La extracción de grandes cantidades de datos estructurados y no estructurados para identificar patrones puede ayudar a una organización a controlar los costos, aumentar la eficiencia, reconocer nuevas oportunidades de mercado y aumentar la ventaja competitiva de la organización.

El campo de la ciencia de datos emplea matemáticas, estadística y disciplinas informáticas, e incorpora técnicas como el aprendizaje automático, el análisis de conglomerados, la extracción de datos y la visualización.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

Seis razones por las que puede necesitar ciencia de datos como servicio

Maria Korolov, Periodista de Tecnología

La [ciencia de datos](#) como servicio, o DSaaS, ha sido noticia durante un tiempo, pero se ha convertido en un tema candente una vez más. En marzo, AWS anunció una asociación con Change Healthcare para ofrecer DSaaS para el análisis de la salud.

Ese mismo mes, el proveedor de soluciones de movilidad Comviva anunció su propia oferta de DSaaS para ayudar a los proveedores de telecomunicaciones con el marketing. Y en febrero, el proveedor de servicios de datos Calligo adquirió el proveedor de análisis de datos Decisive Data para ampliar su oferta de DSaaS.

Ahora que la ciencia de datos como servicio está recibiendo más atención, es importante que las empresas sepan si deben invertir en este servicio y cuáles son las razones para considerarlo.

¿Qué es la ciencia de datos como servicio?

Según Anand Rao, socio y líder global de IA en PwC, la ciencia de datos como servicio es la tercerización de las actividades de ciencia de datos a un proveedor externo.

«El cliente proporciona los datos y el DSaaS proporciona la información de los datos al cliente», dijo.

Resulta especialmente útil para trabajos temporales, para cargas de trabajo repentinas o picos, [o para tareas estandarizadas como la realización de análisis en informes](#) mensuales o trimestrales, explicó.

In this e-guide

- Seis razones por las que puede necesitar ciencia de datos como servicio
- Las habilidades de ciencia de datos con más demanda que necesitará
- Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos
- Obtener más contenido exclusivo de PRO+

Una de las razones por las que no se oye hablar mucho de la ciencia de los datos como servicio es que el término puede significar muchas cosas diferentes. Abarca desde herramientas de análisis integradas en plataformas SaaS populares, como Salesforce, hasta proveedores especializados que ofrecen modelos preconstruidos para aplicaciones empresariales específicas que pueden personalizar y gestionar para los clientes, pasando por implantaciones de consultoría estándar.

Pero sea cual sea su forma, la DSaaS tiene mucho valor que ofrecer a las empresas, tengan o no un equipo interno de ciencia de datos.

DSaaS como IA integrada

Cada vez son más las aplicaciones empresariales [que incorporan funciones de análisis e inteligencia artificial](#), según Kjell Carlsson, analista principal de Forrester Research.

El propio equipo de ciencia de datos de una empresa puede ser capaz de construir algo más personalizado que estas aplicaciones, dijo.

«Pero es muy difícil poner esos modelos en manos del usuario final», dijo. «El usuario final ya está utilizando esas aplicaciones empresariales SaaS, y el modelo ya está ahí, integrado en su flujo de trabajo».

En muchos casos, las ventajas de contar con un modelo integrado que ya está ahí y con el que los usuarios pueden empezar a trabajar superan los posibles beneficios de crear un modelo a medida, según Carlsson.

Sin embargo, si los datos residen en diferentes sistemas, la escala del reto de integración podría requerir un esfuerzo interno.

«Cuantos más sistemas diferentes haya que vincular, menos probable es que se consuma ciencia de datos como servicio», afirma Carlsson.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

DSaaS como plataformas de inteligencia empresarial

Las [suites de inteligencia empresarial de uso general](#) que se ofrecen como servicio cuentan cada vez más con BI y análisis aumentados. Tableau Online, por ejemplo, ofrece análisis de autoservicio en la nube. Microsoft ofrece Power BI como servicio e IBM también ofrece sus herramientas de análisis como servicio.

Dave Costenaro, director de datos de Capacity, empresa de asistencia técnica impulsada por IA, utiliza opciones de DSaaS como Tableau para la analítica y AWS para el almacenamiento de datos.

«La infraestructura para la recopilación y el almacenamiento de datos puede subcontratarse fácilmente a varios proveedores de bases de datos en la nube», dijo.

Las empresas pueden subcontratar pasos individuales de la cadena de análisis a proveedores externos, dijo, incluso si tienen equipos internos para otras partes.

DSaaS como plataformas de desarrollo de IA

Proveedores como C3.ai ofrecen componentes y módulos de IA preconstruidos que las empresas pueden unir para construir sus propias aplicaciones predictivas, dijo Carlsson.

El uso de una plataforma externa de ciencia de datos puede tener sentido, incluso cuando una empresa tiene un equipo interno de ciencia de datos, porque permite la flexibilidad [para escalar los modelos hacia arriba y hacia abajo](#) según sea necesario y para poner en marcha los entornos de prueba rápidamente. También puede reducir los gastos de capital o los costos de licencia, y el proveedor es responsable de mantener la infraestructura y actualizarla.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

El uso de una plataforma DSaaS también puede proporcionar acceso a los algoritmos de ciencia de datos propiedad del proveedor, dijo Hugh Burgin, líder de datos e IA de Estados Unidos del Grupo de Servicios de Microsoft en EY.

Sin embargo, existen algunas desventajas.

«En un modelo alojado proporcionado por un proveedor, esto a menudo se puede sentir como una caja negra, donde la empresa tiene menos impacto y visibilidad sobre cómo funcionan las operaciones de datos y la ciencia de datos», dijo. «Para algunas empresas, la aceptación ejecutiva del uso de recursos externos podría ser un desafío también».

DSaaS como servicios de IA o de aprendizaje automático

Los principales proveedores de la nube [y las startups centradas en la IA](#) ofrecen modelos preentrenados para la visión, la búsqueda, las recomendaciones, el reconocimiento del habla, el procesamiento del lenguaje natural y otras tareas comunes de la ciencia de datos.

Puede ser mucho más rápido y fácil utilizar esas ofertas en lugar de construir un sistema en la empresa. Sin embargo, podría haber problemas de cumplimiento, privacidad o seguridad que pueden limitar el uso de proveedores externos, dijo Carlsson.

«El equipo de seguridad de la información puede ponerle obstáculos si los datos tienen que salir de sus muros», dijo.

DSaaS como aceleradores de IA

Algunas empresas de DSaaS tienen modelos preconstruidos que ajustan en función de los datos de una empresa concreta, y luego cobran mensualmente para mantener el modelo actualizado con los datos más recientes, dijo Carlsson.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«Ciertos casos de uso son tan comunes para muchas empresas diferentes que no va a ser un diferenciador para usted», dijo.

Según Doug Henschen, vicepresidente y analista principal de Constellation Research, hay una tendencia a que las empresas utilicen estas ofertas digitales y productivas para apoyar [el ciclo de vida de la ciencia de datos](#).

«En el lado del desarrollo, estamos viendo servicios de datos, por ejemplo, que apoyan el entrenamiento de modelos con datos orgánicos y sintéticos», dijo. «En el lado operativo, estamos viendo servicios de supervisión de modelos diseñados para optimizar la selección y el mantenimiento de modelos».

Los casos de uso más habituales son la detección de fraudes en la banca, la gestión de personal en la sanidad, la gestión de la curva en el sector de las telecomunicaciones y la optimización de precios [y la selección de clientes en el sector del comercio electrónico](#).

Henschen ha visto que las empresas utilizan equipos internos de ciencia de datos para experimentar e identificar oportunidades, y luego utilizan un proveedor de DSaaS para llevar esos proyectos a escala.

«He visto muchos acuerdos a largo plazo en los que las organizaciones siguen confiando en los proveedores de servicios, en lugar de intentar desarrollar la experiencia en ciencia de datos internamente», dijo. «Si vas a invertir en la construcción y el mantenimiento de la experiencia en ciencia de datos, debe ser algo que claramente ofrece un valor diferenciador para la empresa».

DSaaS como consultoría

Las empresas de servicios profesionales también están encantadas de construir modelos de IA desde cero. Incluso las empresas que tienen equipos internos de ciencia de datos pueden querer traer a una empresa de consultoría para hacer proyectos especializados.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«Incluso si tienes a un científico de datos en la plantilla, un científico de datos que sepa cómo construir modelos de visión es raro», dijo Carlsson. «Alguien que sepa cómo hacer un modelo de reconocimiento de voz es extremadamente raro».

Además, la construcción del modelo también requiere datos de entrenamiento etiquetados, que una empresa podría no tener o solo podría obtener con un gran esfuerzo.

Una situación habitual es la de una empresa que necesita una extracción de texto inteligente para hacer frente a flujos de trabajo que implican una gran cantidad de documentos escaneados. La ingestión de esos documentos, la extracción de los datos en un formato que la empresa pueda utilizar y el análisis de textos no es un conjunto de capacidades que la mayoría de las empresas puedan tener internamente.

También es probable que las empresas recurran a consultores para proyectos puntuales.

«La ciencia de datos como servicio puede proporcionar un golpe rápido para resolver un problema empresarial rápido», dijo Chandana Gopal, director de investigación del futuro de la inteligencia en IDC. «Es capaz de ofrecer resultados de negocio de la manera más rápida [cuando no se tienen las habilidades en la empresa](#)».

La subcontratación también puede tener sentido si una empresa no sabe si el proyecto será un éxito o no. No tiene sentido contratar a un equipo de científicos de datos solo para descubrir que los datos no dan el resultado que se busca.

«Entonces se podría contratar a un equipo interno una vez que se sepa que hay un beneficio probado en el uso de científicos de datos», dijo Gopal.

Pero, incluso entonces, una empresa de consultoría puede ayudar a cubrir el vacío hasta que una empresa contrate y forme a sus propios empleados.

«[Y la escasez de competencias es tan profunda](#) y grande que las organizaciones pueden acabar recurriendo a proveedores de servicios a largo plazo porque no pueden contratar gente», añadió Gopal.

In this e-guide

-
- Seis razones por las que puede necesitar ciencia de datos como servicio
-
- Las habilidades de ciencia de datos con más demanda que necesitará
-
- Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos
-
- Obtener más contenido exclusivo de PRO+

El tamaño de la empresa también supone una gran diferencia, dijo Dan Simion, vicepresidente de IA y análisis de Capgemini.

«Si se trata de una empresa pequeña, la ciencia de datos como servicio puede encajar mejor que para una empresa a gran escala que ya tiene su equipo interno de ciencia de datos en funcionamiento», dijo.

Una de estas empresas es DayaMed, una startup con sede en Nevada que está creando una aplicación de salud para móviles. En marzo, la empresa anunció un proyecto piloto con el Departamento de Asuntos de los Veteranos de EE.UU. centrado en la gestión y el cumplimiento de la medicación, dijo el director general de la empresa, Justin Daya.

Para conseguir esa ciencia de datos y la IA –así como la integración en una aplicación móvil escalable– DayaMed recurrió a SenecaGlobal por ciencia de datos como servicio.

«Ahora tenemos planes para expandirnos a nivel nacional a múltiples sistemas de salud, farmacias, [organizaciones de atención responsable], planes de salud y otros clientes», dijo Daya.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

Las habilidades de ciencia de datos con más demanda que necesitará

Maria Korolov

Según el «Informe de empleos emergentes» de LinkedIn en Estados Unidos de este mes, los científicos de datos tuvieron el tercer mayor crecimiento anual, con un 37 %, y han estado en la cima del informe durante los últimos tres años. Eso pone a los científicos de datos, y las habilidades de ciencia de datos, en gran demanda.

Las habilidades requeridas para trabajos de científicos de datos incluyen aprendizaje automático, Python, R y Apache Spark, según el [informe de LinkedIn](#). Pero una vez que se tiene esos fundamentos, ¿qué hace un científico de datos a continuación para prepararse para el futuro de su carrera? Los expertos recomiendan estas habilidades primarias para los científicos de datos.

Conocimiento del negocio

Los científicos de datos más buscados no solo saben cómo usar la tecnología, sino cuándo usarla y por qué, dicen los expertos. Eso requiere una comprensión de los beneficios comerciales de los datos, así como la capacidad de comunicar estos beneficios a ejecutivos y colegas en otras áreas de la empresa.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«He visto que muchos científicos de datos tienen una buena comprensión de los modelos y no tienen ningún impacto comercial», dijo Andrew Fast, científico jefe de datos de CounterFlow Inc., una compañía de seguridad cibernética con sede en Charlottesville, Virginia.

Para abordar el problema, algunos profesionales se están concentrando en casos de uso específicos.

«[Vemos que el mercado se está volviendo más especializado](#)», dijo Fast. «En lugar de que alguien sea bueno en ciencia de datos, se podría decir que alguien está interesado en ciencia de datos para marketing o ciencia de datos para IoT. Desarrollar una especialidad o enfocarse en un área en particular lo ayudará a obtener tracción a medida que la ciencia de datos y el análisis crecen en popularidad. Tener un área que pueda llamar suya es muy valiosa para el éxito».

También puede ser útil comprender las finanzas y cómo los proyectos de [ciencia de datos](#) pueden ayudar a los resultados de una empresa.

«Muchas personas técnicas y científicas no sienten nada por el dinero ni la economía», dijo Michael Feindt, asesor estratégico y fundador de Blue Yonder, parte de JDA Software Group, una compañía alemana de software de inteligencia artificial.

En Europa en particular, dijo, parece que los científicos de datos piensan que el dinero es algo malo. Un MBA u otro programa de educación formal puede ayudar, dijo Feindt, pero el mejor lugar para aprender sobre negocios es en su trabajo actual.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«Para las personas que son inteligentes y analíticas, aprender sobre negocios es bastante fácil», dijo. «Hay que querer aprenderlo, pero no ser demasiado arrogante y respetar a las personas que no son demasiado científicas».

Las habilidades blandas marcan el camino

[Fuera del conocimiento técnico y la familiaridad comercial, la buena comunicación y otras habilidades blandas](#) están en la parte superior de la lista para las habilidades necesarias de ciencia de datos en el mercado laboral.

En True Fit Corp., una compañía de recomendaciones de moda impulsada por IA, son las habilidades no técnicas las que marcan la diferencia.

«Se ha hecho mucho hincapié en las habilidades técnicas, en el aprendizaje de un algoritmo específico o una clase de algoritmos», dijo Rhonda Textor, directora de ciencia de datos de True Fit. «Lo complicado es que si no entiendes el problema que estás resolviendo y no entiendes cómo funciona el negocio del que eres parte, cómo gana dinero, cómo se alimenta tu algoritmo en el panorama general para hacer eso, entonces no importa si se comprende todo sobre las redes neuronales porque no podrá aplicarlo a los problemas correctos».

La compañía busca buenas habilidades de comunicación, dijo. «Y también buscamos personas que no solo entiendan la ciencia de los datos, sino que puedan asignar eso al valor comercial y al impacto comercial».

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

Incluso las empresas más orientadas técnicamente, como el proveedor de ciberseguridad de San Francisco, Kenna Security, están comenzando a preferir a los científicos de datos que pueden ver el panorama general en lugar de perseguir el último algoritmo.

«Ya sea [aprendizaje profundo](#) o bosques aleatorios o regresiones, esas conversaciones se están volviendo menos interesantes», dijo Michael Roytman, científico jefe de datos de Kenna Security. «Lo que más me entusiasma es descubrir la forma correcta de estructurar la investigación, la ciencia de datos y el desarrollo juntos, de modo que desde el inicio hasta la producción, un nuevo producto pueda ver la luz del día más rápido. Para mí, todo se trata de las habilidades blandas: comunicación, personas y equipos».

Se trata de la ejecución

Los científicos de datos que saben cómo trabajar con ingenieros de datos, o ellos mismos tienen fuertes habilidades de ingeniería, pueden marcar la diferencia entre simplemente tener un proyecto piloto de exhibición y entregar un ROI sólido con una implementación exitosa.

[Esto no significa que los científicos de datos necesariamente tengan que volver a capacitarse como ingenieros](#) de software, dijo Fast de CounterFlow.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«Es difícil esperar que un científico de datos sea bueno en ambas cosas», dijo. «Pero necesitan entender el lenguaje del desarrollo de software».

Habilidades de plataforma a prueba de futuro

El lenguaje de programación Python continuará siendo una habilidad fundamental en el futuro previsible, y los científicos de datos pueden aumentar eso con nuevas técnicas como redes neuronales, aprendizaje profundo, redes adversas y aprendizaje de transferencia.

También hay un creciente énfasis [para los científicos de datos en las implementaciones basadas en la nube](#), que permiten una fácil transferencia de las pruebas a los entornos de producción, escalado automático y rentable, y bibliotecas integradas y fáciles de usar de algoritmos y datos de capacitación. Además, las plataformas se integran bien con las fuentes de datos más populares, como los conjuntos de datos disponibles comercialmente, así como los datos obtenidos de sistemas internos o socios comerciales.

Las tres pilas principales pertenecen a Google, Microsoft y Amazon. También hay plataformas dedicadas a la ciencia de datos, como Databricks, DataRobot y Domino Data Lab, que permiten a los científicos de datos colaborar en la construcción e implementación de modelos. Por lo tanto, puede beneficiar a los científicos de datos familiarizarse con estas plataformas.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

«Estas herramientas están tratando de convertirse en una ventanilla única para todo lo que se quiere hacer, y tienen las integraciones correctas, la funcionalidad correcta y son muy flexibles», dijo Ivaylo Bahtchevanov, jefe de ciencia de datos en ForgeRock, una empresa de ciberseguridad con sede en San Francisco.

Lo que es más importante, las plataformas en la nube hacen que sea fácil para las empresas comenzar de a poco y luego escalar fácilmente.

«Puede expandir fácilmente la cantidad de servidores o recursos que necesita sin tener que ir a comprar hardware real», dijo Charles Ng, vicepresidente de IA empresarial en Appier, una compañía de tecnología de marketing basada en IA con sede en Taiwán.

«Las plataformas de inteligencia artificial y aprendizaje automático están cambiando drásticamente», dijo Myke Miller, decana del Deloitte Cloud Institute, parte de Deloitte Consulting LLP.

[Las nuevas plataformas simplifican el trabajo del científico de datos](#), dijo, y agregan una enorme cantidad de automatización a las secuencias de ciencia de datos. Los científicos de datos necesitan aprender a usar las nuevas plataformas, pero después de eso, las plataformas realmente reducen la necesidad de ciertas habilidades específicas de ciencia de datos.

«La necesidad de Python o Firehose o cualquiera de los productos componentes está disminuyendo», dijo Miller. «Y la necesidad de colaboración y comprensión del problema comercial continúa aumentando».

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos

Bridget Botelho, Senior News Director

Los proyectos de ciencias de datos pueden proporcionar un valor comercial inmenso, guiando a las empresas hacia mayores ingresos u operaciones mejoradas, pero también pueden ser perjudiciales si se hacen mal.

Digamos que alguien en un equipo de marketing que se imagina un científico de datos ciudadanos utiliza Domo, Google Analytics o una herramienta de inteligencia de negocios fácil de usar para obtener información sobre las ventas. Los datos muestran que el equipo alcanzará sus números, y presenta esa información a los altos ejecutivos.

Sin embargo, esos datos pueden no ser exactos, ya que no se puede ver la metodología detrás de las herramientas de caja negra, dijo Ian Swanson, fundador y CEO de DataScience, un proveedor de servicios y plataforma de ciencia de datos en California.

La construcción de modelos predictivos exactos requiere personalización y conocimiento de diversas metodologías y enfoques, que se aplican en función del escenario y del conjunto de datos. En otras palabras, si conecta datos en una herramienta de análisis predictivo de caja negra, y confía en que ha aplicado la metodología y el enfoque adecuados para analizar sus datos, está tomando una apuesta. Los mejores equipos de ciencias de datos nunca harían eso, y tampoco debería hacerlo usted.

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

"Podría ser genial sacar algunas ideas de [herramientas de caja negra], pero puede ser realmente peligroso", dijo Swanson. "¿Debería un ciudadano científico de datos tomar decisiones que impactan a la empresa? Ninguna gran empresa quiere [eso]".

Swanson discutió tres peligros comunes a tener en cuenta durante los proyectos de ciencia de datos, incluyendo problemas de modelos predictivos y calidad de datos, la importancia del linaje de datos, y cómo asegurarse de que tiene el flujo de trabajo de análisis adecuado.

Mala calidad de los datos

El éxito de un proyecto de ciencia de datos comienza con buenos datos. Si los datos que entran en un modelo predictivo son malos, las salidas predictivas no serán exactas.

Con eso, un primer paso crítico en el modelado predictivo es explorar y evaluar la calidad de los datos, determinar cuánta limpieza de datos se requiere y llevar los datos a un formato utilizable, dijo Swanson.

Los equipos de ciencias de los datos también deben comprobar que el tipo correcto de datos está incluso allí. Tomemos la elección presidencial de los Estados Unidos en 2016. Los modelos predictivos [señalaron a Hillary Clinton](#) como la ganadora, y claramente esas predicciones estaban equivocadas. Una razón, según Swanson y otros expertos en ciencias de datos, era que había voces críticas que faltaban en los datos. Además, los datos introducidos en los modelos predictivos no pudieron ser validados;

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

algunos votantes pudieron haber dicho que planeaban votar en una dirección, pero finalmente votaron de manera diferente.

Falta de linaje de datos

Un equipo de ciencia de datos debe ser capaz de seguir el ciclo de vida de los datos que utiliza, incluyendo el origen de los datos y cómo se recolectaron. El equipo también debe ser capaz de explicar lo que encontró durante la fase de exploración de datos, la metodología y el proceso de análisis, y cómo los equipos de negocios de la empresa podrán utilizar la información.

Sin un claro linaje de datos, los ejecutivos pueden no confiar en los datos, y en su lugar pueden optar por apoyarse más en su intuición que en el análisis de datos, dijo Swanson.

"Conectamos los puntos, así que cuando las partes interesadas ejecutivas ven los datos, les mostramos cómo los datos y los resultados fueron encontrados y cómo se pueden usar en los productos", dijo. "Todos los puntos necesitan ser puestos juntos".

Equipos analíticos de torre de marfil

A menudo, los equipos de ciencia de datos están centralizados, y no tienen flujos de trabajo integrados con los equipos de negocios. Swanson recomienda [que los](#)

In this e-guide

- [Seis razones por las que puede necesitar ciencia de datos como servicio](#)
- [Las habilidades de ciencia de datos con más demanda que necesitará](#)
- [Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos](#)
- [Obtener más contenido exclusivo de PRO+](#)

[científicos de datos se integren en los equipos de negocios](#) para asegurarse de que entienden los problemas que necesitan ser resueltos, y para trabajar juntos para averiguar cómo la producción analítica predictiva se puede hacer productiva (si esa es la meta). La integración de los equipos también ayuda a los equipos de ciencia de datos y de negocios a identificar oportunidades analíticas y a aprovechar el conocimiento institucional.

"Vemos problemas de integridad de datos y [problemas en] la elección del algoritmo correcto, pero lo más importante es el flujo de trabajo: si resolvemos este problema de esta manera, ¿se puede utilizar?", dijo Swanson. "No tener a los interesados de negocios en la mesa es una trampa común. Si estás tratando de resolver un problema para marketing, ¿están las personas de marketing en la mesa con los científicos de datos?".

De particular importancia es que los equipos de ciencias de datos trabajen con el equipo de ingeniería que armoniza los modelos predictivos y los pone en producción, añadió. Si esos dos equipos no usan el mismo lenguaje, los proyectos de ciencia de datos podrían estar muertos antes de llegar.

In this e-guide

- Seis razones por las que puede necesitar ciencia de datos como servicio
- Las habilidades de ciencia de datos con más demanda que necesitará
- Tres fallas de modelado predictivo que estropean los proyectos de ciencia de datos
- Obtener más contenido exclusivo de PRO+

Obtener más contenido exclusivo de PRO+

Como miembro de PRO+, tienes acceso a todo el portafolio de más de 140 sitios web de TechTarget. El acceso a PRO+ te dirige a “contenidos exclusivos para miembros platino” que están garantizados para ahorrarte tiempo y esfuerzo de tener que rastrear dicho contenido premium por tu cuenta, ayudándote en última instancia a resolver tus desafíos más difíciles de TI de manera más efectiva y rápida que nunca.

Aprovecha tu membresía al máximo visitando:
<https://www.computerweekly.com/es/eproducts>

Images; stock.adobe.com

© 2021 TechTarget. Ninguna parte de esta publicación puede ser transmitida o reproducida de ninguna forma o por ningún medio sin el permiso escrito del editor.