

ESTADISTICA

CIENCIA DE DATOS

CIENCIA DE DATOS

- **Definición de ciencia de datos**
- La ciencia de datos combina múltiples campos, como las estadísticas, los métodos científicos, la inteligencia artificial (IA) y el análisis de datos para extraer el valor de los datos. Los practicantes de la ciencia de datos se llaman científicos de datos y combinan una variedad de conocimientos para analizar los datos recopilados de la web, teléfonos inteligentes, clientes, sensores y otras fuentes para obtener información útil.
- La ciencia de datos abarca la preparación de los datos para el análisis, incluida la limpieza, la agregación y la manipulación de los datos para realizar análisis avanzados. Las aplicaciones analíticas y los científicos de datos pueden revisar los resultados para descubrir patrones y permitir que los líderes empresariales obtengan información fundamentada

PORQUE ? ES IMPORTANTE LA DS

- La ciencia de datos es uno de los campos más emocionantes que existen en la actualidad. Pero, ¿por qué es tan importante?
- Porque las empresas disponen un tesoro de datos sin aprovechar. Ahora que la tecnología moderna ha permitido la creación y el almacenamiento de cantidades cada vez mayores de información, el volumen de datos explotó. Se estima que el 90% de los datos en el mundo se crearon en los últimos dos años. Por ejemplo, los usuarios de Facebook suben 10 millones de fotos por hora.
- Pero estos datos frecuentemente solo están inmóviles en las bases de datos y los lagos de datos, básicamente sin tocar.
- La gran cantidad de datos recopilados y almacenados por estas tecnologías puede generar beneficios transformadores para las organizaciones y sociedades de todo el mundo, pero solo si sabemos interpretarlos. Ahí es donde entra en acción la ciencia de datos.
- La ciencia de datos revela tendencias y genera información que las empresas pueden utilizar para tomar mejores decisiones y crear productos y servicios más innovadores. Quizás lo más importante es que permite que los modelos de aprendizaje autónomo (ML) aprendan de las grandes cantidades de datos que se les suministran en vez de depender principalmente de los analistas de negocios para ver qué pueden descubrir a partir de los datos.
- Los datos son la base de la innovación, pero su valor proviene de la información que los científicos pueden extraer y luego utilizar a partir de estos.

DIFERENCIAS

- **¿Cuál es la diferencia entre la ciencia de datos, la inteligencia artificial y el aprendizaje autónomo?**
- Para comprender mejor la ciencia de datos (y cómo puede aprovecharla) es igual de importante conocer otros términos relacionados con el campo, como inteligencia artificial (IA) y aprendizaje autónomo. Frecuentemente, encontrará que estos términos se usan indistintamente, pero hay matices.
- Este es un breve resumen:
- **IA** significa hacer que una computadora imite de alguna manera el comportamiento humano.
- La **ciencia de datos** es un [subconjunto de la IA](#) que se refiere más a las áreas superpuestas de las estadísticas, los métodos científicos y el análisis de datos, que se utilizan todas para extraer significado y conocimientos de los datos.
- El [aprendizaje autónomo](#) es otro subconjunto de la IA y consiste en las técnicas que permiten que las computadoras descubran cosas a partir de los datos y realicen aplicaciones de IA. Y, por si acaso, incluimos otra definición.
- [Aprendizaje profundo](#), que es un subconjunto del aprendizaje autónomo que permite que las computadoras resuelvan problemas más complejos.

Cómo la ciencia de datos está transformando los negocios

- Las organizaciones están utilizando la ciencia de datos para convertir los datos en una ventaja competitiva al perfeccionar los productos y servicios. Algunos casos de uso de la ciencia de datos y el aprendizaje autónomo incluyen:
- Determinar la fuga de clientes analizando los datos que se recopilan de los centros de llamadas, para que el departamento de Marketing pueda tomar medidas a fin de retenerlos.
- Mejorar la eficiencia al analizar los patrones de tráfico, las condiciones climáticas y otros factores para que las empresas de logística puedan mejorar los tiempos de entrega y reducir los costos.
- Mejorar los diagnósticos de los pacientes mediante el análisis de los exámenes médicos y los síntomas informados para que los médicos puedan diagnosticar antes las enfermedades y tratarlas de manera más eficaz.
- Optimizar la cadena de suministro al predecir cuándo se producirán fallos en los equipos.
- Detectar los fraudes en los servicios financieros mediante el reconocimiento de los comportamientos sospechosos y las acciones anómalas.
- Mejorar las ventas al crear recomendaciones para los clientes basadas en las compras anteriores.

Cómo se lleva a cabo la ciencia de dato

- El proceso de analizar y utilizar los datos es iterativo más que lineal, pero este es el flujo normal del ciclo de vida de la ciencia de datos para un proyecto de modelado de datos:
- **Planificación:** Definir un proyecto y sus posibles resultados.
- **Construir un modelo de datos:** Los científicos de datos frecuentemente usan una variedad de bibliotecas de código abierto o herramientas en la base de datos para construir modelos de aprendizaje autónomo. A menudo, los usuarios necesitan API para que los ayuden con la ingestión de datos, la visualización y creación de perfiles de datos o la ingeniería de funciones. Necesitan las herramientas adecuadas, así como acceso a los datos correctos y otros recursos como la capacidad de proceso.
- **Evaluar un modelo:** Los científicos de datos deben lograr un alto porcentaje de exactitud en sus modelos antes de poder implementarlos con confianza. La evaluación del modelo habitualmente genera un conjunto completo de métricas de evaluación y visualizaciones para medir el rendimiento del modelo frente a los datos nuevos y también para clasificarlos a lo largo del tiempo a fin de permitir un comportamiento óptimo en la producción. La evaluación del modelo va más allá del rendimiento en bruto para tener en cuenta el comportamiento de referencia esperado.
- **Explicar los modelos:** No siempre hemos sido capaces de explicar la mecánica interna de los resultados de los modelos de aprendizaje autónomo en términos humanos, pero esto es cada vez más importante. Los científicos de datos desean recibir explicaciones automatizadas de la ponderación relativa y la importancia de los factores que intervienen en la generación de una predicción, junto con detalles explicativos específicos del modelo sobre las predicciones del modelo.
- **Implementar un modelo:** Tomar un modelo de aprendizaje autónomo entrenado e implementarlo en los sistemas correctos es frecuentemente un proceso difícil y laborioso. Esto se puede simplificar operacionalizando los modelos como API escalables y seguras, o usando modelos de aprendizaje autónomo dentro de la base de datos.
- **Monitorear los modelos:** Desafortunadamente, la implementación del modelo no es el paso final. Los modelos siempre deben monitorearse después de la implementación para garantizar que funcionen correctamente. Con el paso del tiempo, los datos con los que se entrenó el modelo pueden quedar obsoletos para las predicciones futuras. En la detección de fraudes, por ejemplo, los delincuentes siempre encuentran nuevas formas de piratear las cuentas.

Herramientas para la ciencia de datos

- Crear, evaluar, implementar y monitorear los modelos de aprendizaje autónomo puede ser un proceso complejo. Es por eso que la cantidad de herramientas de ciencia de datos ha aumentado. Los científicos de datos utilizan muchos tipos de herramientas, pero una de las más comunes son los cuadernos de código abierto, que son aplicaciones web para escribir y ejecutar código, visualizar datos y ver resultados, todo dentro de un mismo entorno.
- Algunos de los cuadernos más populares son Jupyter, RStudio y Zeppelin. Los cuadernos son muy útiles para realizar análisis, pero presentan ciertas limitaciones cuando los científicos de datos tienen que trabajar en equipo. Para resolver este problema, se crearon las plataformas de ciencia de datos.
- Para determinar qué herramienta de ciencia de datos es adecuada para usted, es importante formular las siguientes preguntas: ¿Qué tipo de lenguajes utilizan sus científicos de datos? ¿Qué tipo de métodos de trabajo prefieren? ¿Qué tipo de fuentes de datos usan?
- Por ejemplo, algunos usuarios prefieren tener un servicio independiente de la fuente de datos que utilice bibliotecas de código abierto. Otros prefieren la velocidad de los algoritmos de aprendizaje autónomo en la base de datos.

¿Quién supervisa el proceso de ciencia de datos?

- En la mayoría de las organizaciones, los proyectos de ciencia de datos suelen estar supervisados por tres tipos de administradores:
- **Directores comerciales:** Estos directores trabajan con el equipo de ciencia de datos para definir el problema y desarrollar una estrategia para el análisis. Pueden ser los jefes de una línea de negocios como Marketing, Finanzas o Ventas y contar con un equipo de ciencia de datos directamente subordinado. Trabajan codo a codo con los directores de Ciencia de Datos y Tecnología Informática para garantizar que se concreten los proyectos.
- **Directores de Tecnología Informática:** Los directores sénior de Tecnología Informática son responsables de la infraestructura y de la arquitectura que asistirá las operaciones de ciencia de datos. Supervisan continuamente las operaciones y el uso de los recursos para garantizar que los equipos de ciencia de datos operen en forma eficiente y segura. También pueden ser responsables de la creación y actualización de los entornos de TI para los equipos de ciencia de datos.
- **Directores de ciencia de datos:** Estos gerentes supervisan el equipo de ciencia de datos y su trabajo diario. Son creadores de equipos que pueden equilibrar el desarrollo del equipo con la planificación y el monitoreo del proyecto.
- Pero el participante más importante en este proceso es el científico de datos.

¿Qué es un científico de datos?

- Como especialidad, la ciencia de datos aún es nueva. Surgió de los campos del análisis estadístico y de la minería de datos. La revista *Data Science Journal* lanzada en el año 2002 y publicada por Consejo Internacional para la Ciencia: Comité de Información para Ciencia y Tecnología. En el año 2008, ya había surgido el título de científico de datos y el campo despegó rápidamente. Desde entonces, ha habido una escasez de científicos de datos, a pesar de que cada vez más escuelas y universidades han comenzado a ofrecer títulos en ciencia de datos.
- Las tareas de un científico de datos pueden incluir el desarrollo de estrategias para analizar datos; la preparación de datos para su análisis; explorar, analizar y visualizar datos; construir modelos con datos mediante lenguajes de programación como Python y R; e implementar modelos en aplicaciones.
- El científico de datos no trabaja solo. De hecho, la ciencia de datos más efectiva se ejecuta en equipos. Además de un científico de datos, este equipo puede incluir un analista empresarial que define el problema, un ingeniero de datos que prepara los datos y su método de acceso, un arquitecto de tecnología informática que supervisa los procesos subyacentes y la infraestructura, y un desarrollador de aplicaciones que implementa los modelos o las salidas del análisis en aplicaciones y productos.

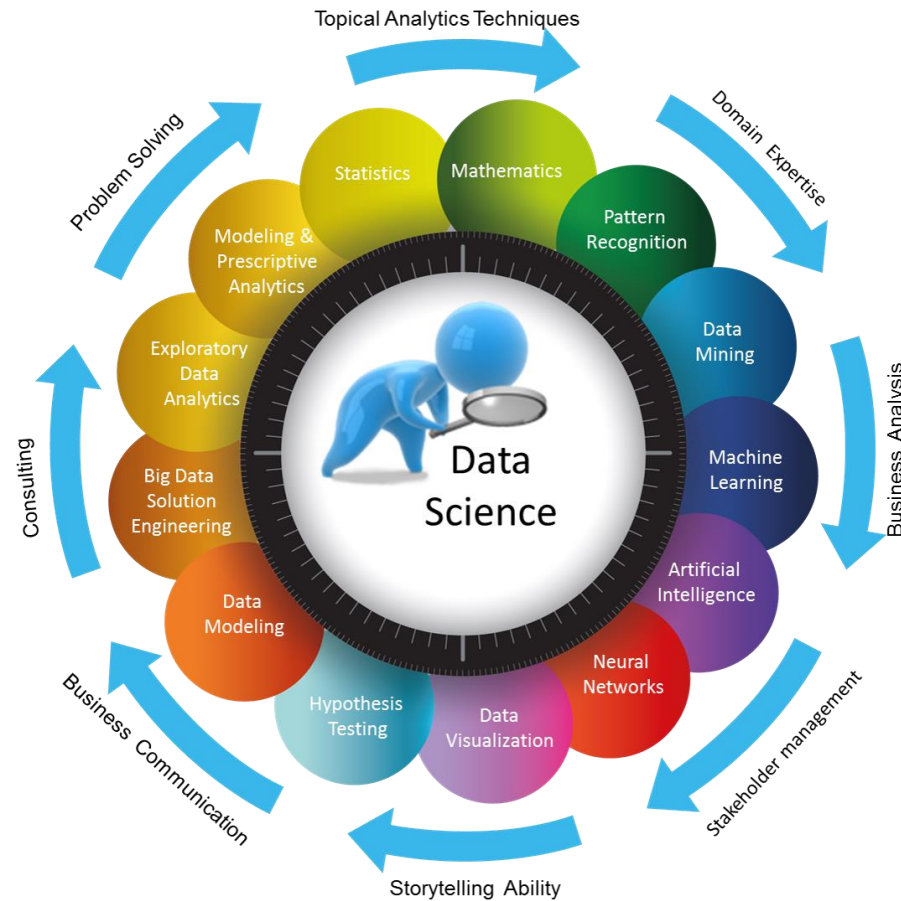
ALGUNAS PREGUNTAS

- A continuación, presentamos preguntas y situaciones que sirven como ejemplos de aplicaciones de Data Science:
- → Mostrar qué productos realmente interesan a los clientes de una empresa.
- → En los próximos dos meses, ¿qué clientes probablemente no pagarán?
- → En el período posterior a la crisis, ¿cuánto tiempo le tomará a una compañía en particular restablecer su negocio?
- → Para que esto suceda, ¿qué acciones deben tomar por adelantado?
- → ¿Es mejor invertir en acciones o comprar / alquilar equipos más modernos para la producción?
- → ¿Es mejor contratar a más personas el año que viene? Si es así, ¿en qué áreas será rentable el retorno?

Cómo un científico de datos entrega resultados

- Ante estos desafíos, tan vitales para la supervivencia de la empresa, este profesional confía en la tecnología y el desarrollo de habilidades digitales para analizar datos y ofrecer soluciones para el negocio.
- Con ellos, el científico de datos tiene el conocimiento necesario para comprender completamente el problema, identificar qué tipo de método usar y ayudar a resolver la situación, comenzando con el tratamiento de los datos.
- En este paso, el objetivo es proporcionar al método elegido solo información útil para resolver el problema. Cuando se completa, aplica la solución resultante, evaluándola con métricas específicas, las mismas que mostrarán las limitaciones contenidas en la solución. Luego, con toda certeza, es hora de llevar la solución a la junta, es decir, a los propietarios del problema.

TODO GIRA ALREDEDOR DE LOS DATOS



TODO ESTA EN COMO RAZONAMOS



NO OLVIDARSE DE QUE EN NUESTRA MENTE DEBE EXISTIR UN ORDEN

El proceso del análisis de datos

