

Unidad 1. ¿Qué es la ciencia de datos?

Por Sarahí Aguilar

Universidad Panamericana

ITISI 2016

BDP COM145

Primavera 2022

Agenda



Historia



Estado del arte



Definición



Aplicaciones

Agenda

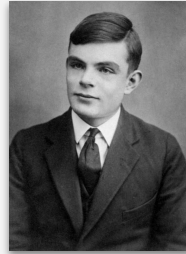


Historia

Estado del arte

Definición

Aplicaciones



1950

Prueba de Turing

Alan Turing propone un estándar para que una máquina se denomine *inteligente* en su paper Computing Machinery and Intelligence.



1943



1956



1960



1965

Threshold logic

Walter Pitts y Warren McCulloch crearon un modelo basado en redes neuronales del cerebro humano.

Conferencia de Dartmouth

John McCarthy decidió organizar un grupo para desarrollar ideas sobre las máquinas pensantes. Eligió el término *Inteligencia Artificial* para el nuevo campo.

Modelo de Back Propagation

Henry J. Kelley desarrolla las bases del modelo.

Esfuerzos de algoritmos de Deep Learning

Alekséi Ivájnénko usó modelos con funciones de activación polinomiales.



1974

Inicia primer invierno de la Inteligencia Artificial



1979

Redes convolucionales

Kunihiko Fukushima desarrolla la primer red neuronal con múltiples capas, Neocognitron., que *aprendía* patrones visuales.



1989

Demostración *práctica* del modelo de Back Propagation

En Bell Labs, Yann LeCun combina una red convolucional con un modelo de propagación para leer dígitos escritos a mano.



1990

Inicia segundo invierno de la Inteligencia Artificial



1997

Deeper Blue vence a Gary Kaspárov

Deep Blue, una supercomputadora desarrollada por el fabricante estadounidense IBM para jugar al ajedrez, derrota Gary Kaspárov, campeón del mundo.



1999

Nacimiento del GPU

La velocidad computacional incrementa 1,000 veces en menos de un año.



2000

Problema del desvanecimiento de gradiente

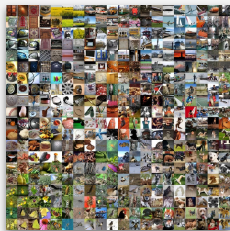
Se plantea que, en algunos casos, el gradiente se irá desvaneciendo a valores muy pequeños, impidiendo eficazmente el peso de cambiar su valor.



2001

Conceptualización del Big Data

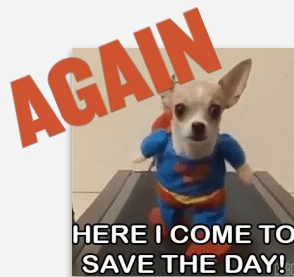
META Group (ahora llamado Gartner) describió los desafíos y oportunidades del crecimiento de datos como tridimensionales.



2009

ImageNet

Fei-Fei Li, crea y publica una base de datos de más de 14 millones de imágenes de Internet etiquetadas, ImageNet.



2011

Aceleración del GPU

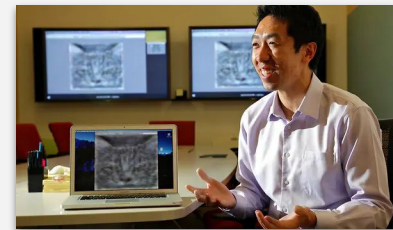
La velocidad de las GPU había aumentado significativamente, lo que hizo posible entrenar redes neuronales convolucionales sin el entrenamiento previo capa por capa.



2011

IBM Watson gana Jeopardy!

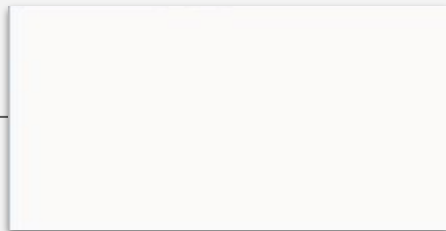
Watson, un sistema desarrollado por IBM y basado en inteligencia artificial capaz de responder a preguntas formuladas en lenguaje natural, derrota a dos oponentes humanos en Jeopardy!



2012

El experimento del gato

Andrew Ng de Google Brain, desarrolla una red neuronal capaz de reconocer la imagen de un gato.



Agenda

A vertical line runs down the left side of the slide. It has four horizontal bars of different shades of gray intersecting it at regular intervals.

Historia

Estado del arte

Definición

Aplicaciones



Estado del arte

<https://paperswithcode.com/sota>

Retos a resolver en el 2022 en México

Operacionalización de la IA

Desplegar modelos de tal manera que comiencen a crear valor para la empresa es una tarea difícil técnica y políticamente.

Más plataformas de datos

Tener acceso a los datos *correctos* significa poder ver al cliente desde múltiples ángulos y adaptarse mejor al mercado.

Búsqueda de talento más profesional

La demanda de científicos de datos incrementará, pero la exigencia en la oferta también.

IA en las universidades

Los académicos que estén familiarizados con estos temas son también muy buscados por las empresas, lo cual pone a las universidades en la difícil posición de tener que atraer al talento que la industria también está buscando.

What about Ethics?

Common carbon footprint benchmarks

in lbs of CO2 equivalent

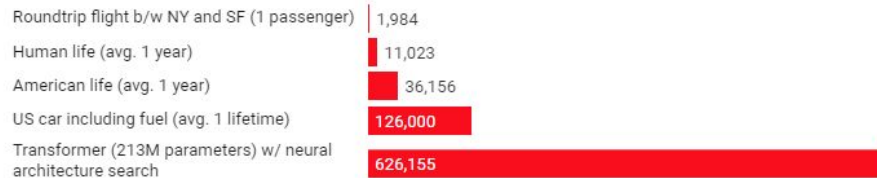


Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

Agenda

A vertical line runs down the left side of the slide. It has four horizontal bars of different shades of gray intersecting it at regular intervals.

Historia

Estado del arte

Definición

Aplicaciones

Definición

Ante la ausencia de datos, las creencias están desinformadas y las decisiones, en el mejor de los casos, se basan en las mejores prácticas o en la intuición.

Definición

*"Data Science...
It's not a real science.
It's an unnecessary label.
The name doesn't even make sense
There's no definition."*

Pete Warden, Technical Lead of the TensorFlow Mobile team, Google

¿Qué es la ciencia de datos?

"By 'Data Science' we mean almost everything that has something to do with data."

Journal of Data Science, 2003

"Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data."

Wikipedia

"Data science is the discipline of making data useful."

Cassie Kozyrkov, Head of Decision Intelligence, Google

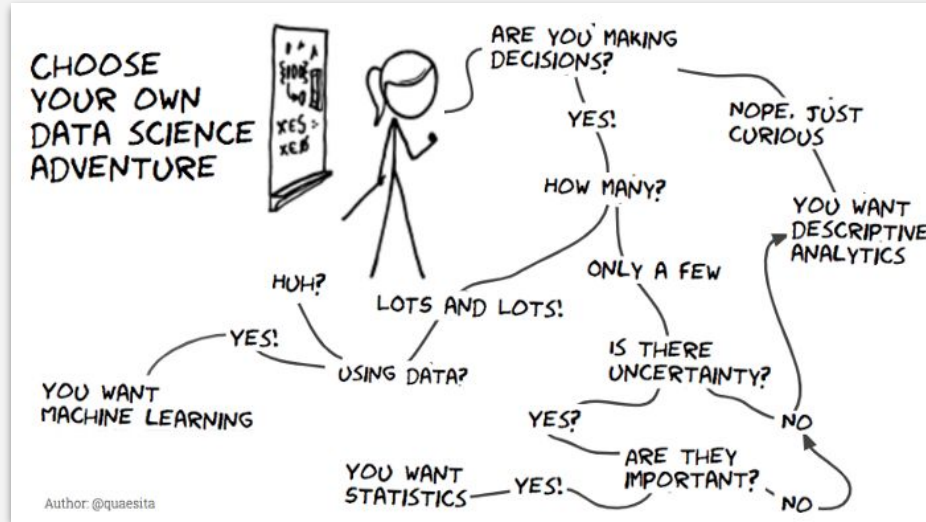
¿Qué es la ciencia de datos?

$$x_1 + x_2 + x_n \rightarrow Y$$

¿Qué es la ciencia de datos?

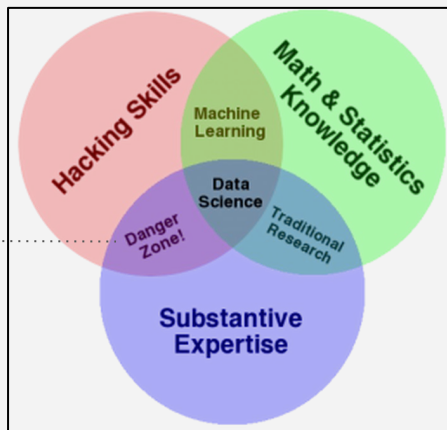
Datos + Herramientas -> Implicaciones

¿Cuáles son las **Herramientas** de la ciencia de datos?



¿Cuáles son las herramientas de la ciencia de datos?

"There are lies, damned lies and statistics."
Mark Twain

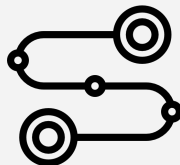


THE DATA SCIENCE VENN DIAGRAM, Drew Conway, CEO and founder of Alluvium

¿Cuáles son las herramientas de la ciencia de datos?



Analítica
Get inspired.



Estadística
Decide wisely.



Machine Learning
Make recipe.

Agenda

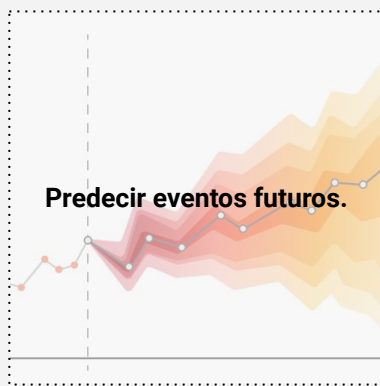


Historia

Estado del arte

Definición

Aplicaciones



¿Cómo aplican ciencia de datos las empresas hoy en día?

[Jamboard](#)

BBVA

**Recomendadores
internos**

**Análisis
predictivo
en tiempo
real**

**Segmentación
de clientes**

Banca

**Deteccion
de fraudes**

**Atención
al cliente**

**Gestión
de
Riesgos**

**Controles
internos**

**Relación
con el
cliente**

**Riesgo en
inversión**

NFL

Modelo semántico, llamado Video Game Ontology

Permite a los diseñadores y empresas de videojuegos intercambiar datos entre sus productos.

Detección de fraudes. Los niveles de seguridad en la industria de juegos deben ser de los más altos niveles, por lo tanto, los algoritmos de Machine Learning permiten la identificación más rápida de actividades

Predicciones sobre cuántas personas atenderán a un estreno

pronósticos, investigación de operaciones, y recomendación de contenido

Análisis de las emociones de los usuarios para medir la satisfacción con la página/producto

Entretenimiento

Analizar tendencias en los videojuegos multijugador para determinar si hay elementos que requieren ajustes de balance

En el deporte, páginas donde consultamos las estadísticas de un equipo o de un jugador sirven tanto para los aficionados como para los profesionales (tomar decisiones)

Netflix recolecta datos de los usuarios e implementa modelos de análisis de datos para descubrir su comportamiento del usuario y hábitos de compra, para posteriormente usar la información recogida y

Spotify y Apple Music proveen a los artistas con información demográfica y geográfica para ayudar a entender a su audiencia

20th Century Fox usa IA para analizar trailers de películas para determinar a que tipo de audiencia le puede gustar

10 Most-Chosen FMCG Brands in the World
Consumer Reach Points measure every time a consumer chooses a brand
52-Week Period Between October 2013 and October 2014

Rank	Brand name	Manufacturer	Consumer Reach Points (m)	Penetration %	Frequency	Consumer Reach Points Growth %
1	Coca-Cola	The Coca-Cola Company	5,722	43.3%	13.4	-2%
2	Colgate	Colgate-Palmolive Company	3,992	64.6%	6.3	3%
3	Maggi	Nestlé	2,755	32.7%	8.6	5%
4	Lifebuoy	Unilever	2,338	26.5%	8.9	7%
5	Nescafé	Nestlé	2,137	22.8%	9.5	-6%
6	Pepsi	Pepsico	1,984	24.6%	8.2	1%
7	Lay's	Pepsico	1,747	27.2%	6.5	6%
8	Knorr	Unilever	1,647	29.4%	5.7	-4%
9	Dove	Unilever	1,457	34.7%	4.3	6%
10	Tide	P&G	1,438	29.8%	4.9	-1%

Coca cola realiza productos relacionados con datos recogidos por IA

Saber el mejor momento para enseñar anuncios y en dónde

Coca Cola está utilizando la ciencia de datos para encontrar nuevos patrones en el tráfico peatonal y la demanda para ubicar mejor sus máquinas para obtener la máxima utilidad.

Productos de consumo masivo

Desarrollo de nuevos productos o mejora de los existentes

Eficiencia para transporte del producto

En que regiones es aceptado el producto

Analizar los patrones del comportamiento del cliente

Mejorar la marca escuchando a los clientes, centrarse en el consumidor

Estimar ganancias en determinado tiempo

Ian Doshner Galland
Ricardo Ariel Flores Peregrina
Andrea Hinojosa Alborno
Manuel Kanahuati Ceballos
Andrés Robles Arce

Aplicaciones

Mejoras en el proceso de descubrimiento de fármacos que están impulsando la creación de nuevos conocimientos sobre cómo combatir enfermedades y dolencias

Diagnóstico temprano del cáncer (NCI)

Se utiliza ciencia de datos para diagnósticos virtuales, el usuario ingresa sus síntomas y el aplicativo da un diagnóstico conforme a estos.

Estudio de cortes bidimensionales del cerebro para la prevención de futuras Hemorragias Cerebrales en TC

Medicina

Detección temprana de enfermedades con datos de pacientes con algunas enfermedades

En el descubrimiento de las alteraciones que puede provocar una medicina en el cuerpo humano: simular la reacción de una medicina en proteínas del cuerpo, en células

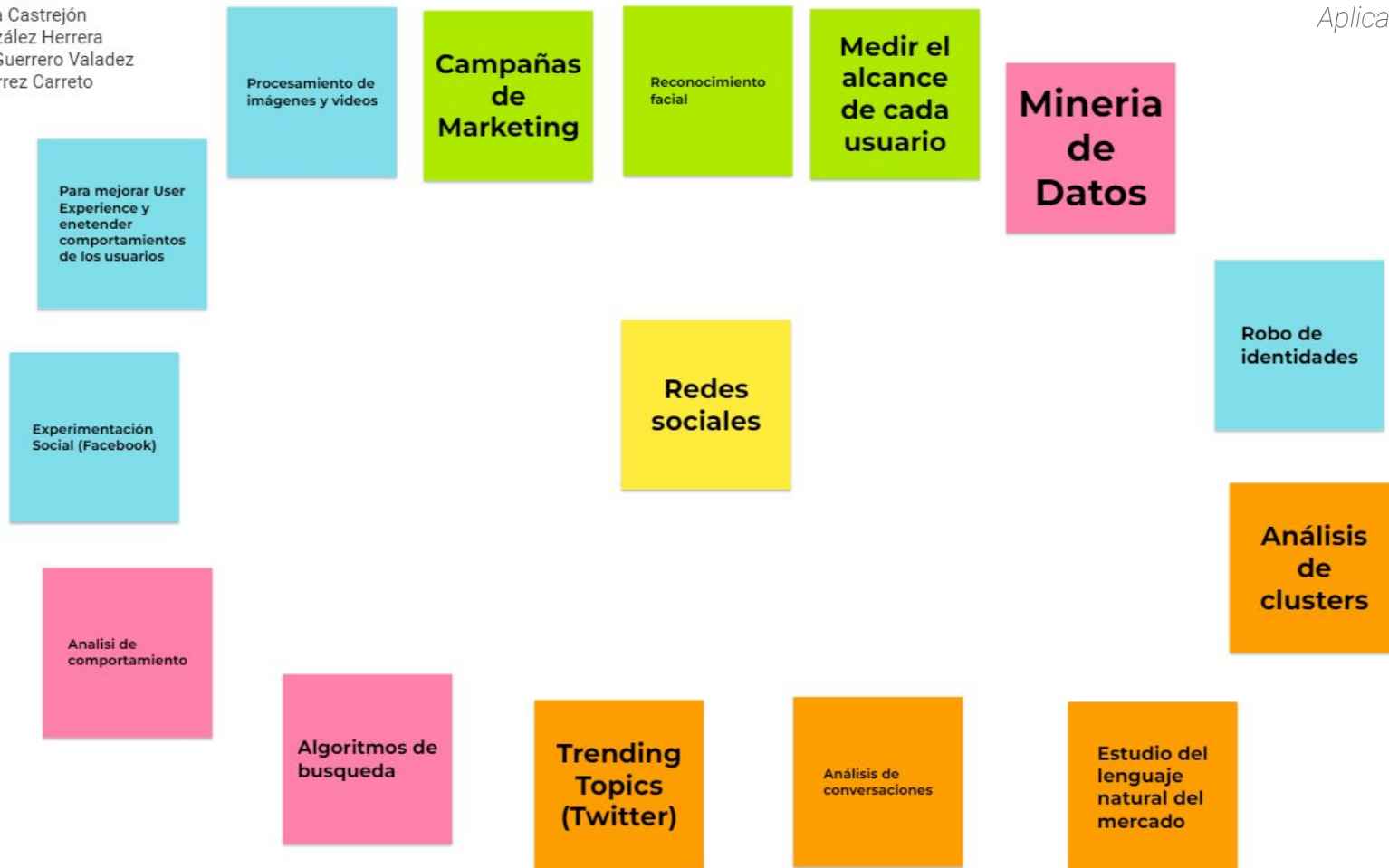
Interpretación por computadoras de imágenes médicas (estudios, radiografías, etc) para detectar enfermedades/anomalías en el cuerpo humano.

Universidad de Stanford hay investigadores que han desarrollado modelos para diagnosticar latidos irregulares de corazón y reconocimiento de lesiones benignas o malignas en la piel.

Elaborar estadísticas sobre los pacientes para elegir aquellos que sean más adecuados para ciertos estudios clínicos.

Carlos Arias Cortina
Carlos Cabrera Castrejón
Bernardo González Herrera
Juan Manuel Guerrero Valadez
Mariana Gutiérrez Carreto

Aplicaciones



¿Por qué la ciencia de datos llegó para quedarse?



¿Por qué la ciencia de datos llegó para quedarse?

La ciencia de datos brinda a los programadores una forma alternativa de decirle a las computadoras qué hacer.

¿Por qué la ciencia de datos llegó para quedarse?

La Inteligencia Artificial se trata de la autoexpresión humana.



¿Por qué la ciencia de datos llegó para quedarse?

**Programación
tradicional**

Instrucciones
explícitas

IA/ML

Ejemplos