

PRIMER ENCUENTRO COMPLUTENSE "BIGDATA Y SOCIEDAD"

Patrocinado por el Banco Santander

1-2 de Diciembre de 2016, Facultad de Geografía e Historia
Universidad Complutense de Madrid

El poder de los datos



Francisco Herrera

Dpto. Ciencias de la Computación e Inteligencia Artificial

Universidad de Granada

herrera@decsai.ugr.es

El poder de los datos

Introducción a Big Data y Ciencia de Datos

El poder de los datos: Casos en diferentes ámbitos

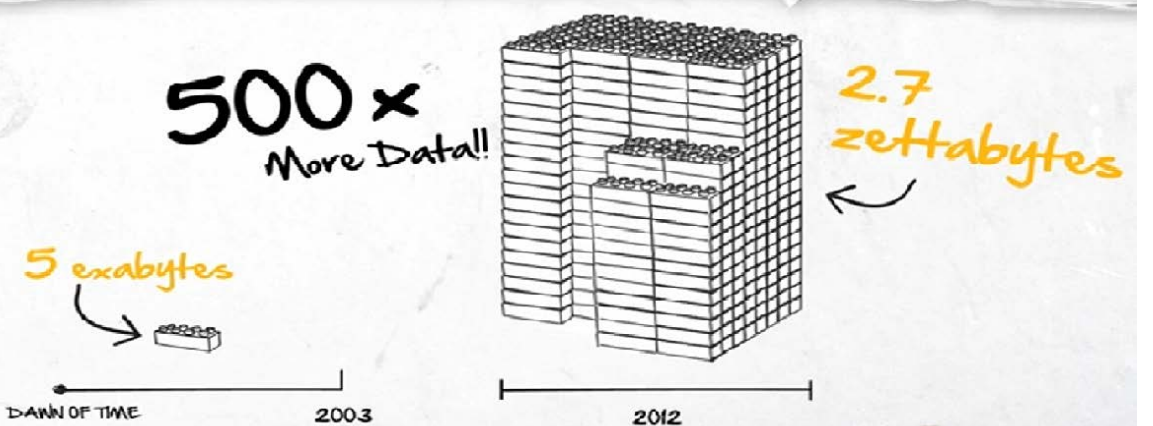
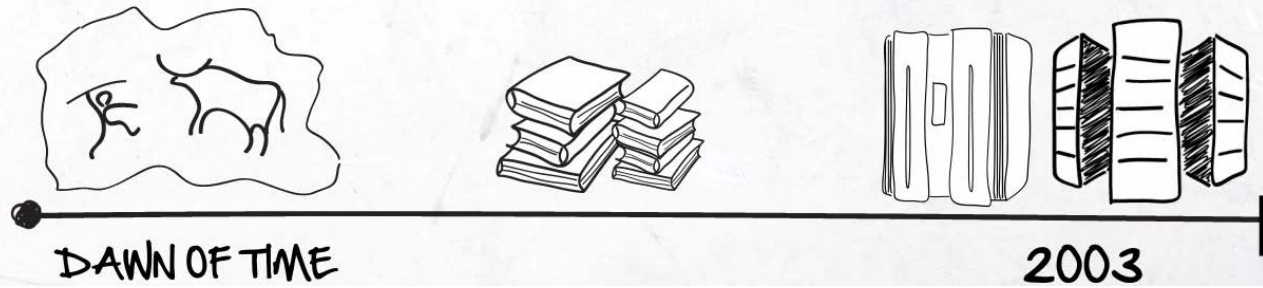
Analítica y tendencias tecnológicas

La realidad, lo mediático y el futuro

La explosión de los datos

5 exabytes

Exabytes: 10^{18} bytes.



El mundo comienza a girar alrededor de los datos



Data are becoming the new raw material of business.

— Craig Mundie —

Craig Mundie
Head of Research and Strategy at Microsoft



Data is a precious thing and will last longer than the systems themselves.

— Tim Berners-Lee —

Tim Berners-Lee
Inventor of the World Wide Web

El mundo comienza a girar alrededor de los datos

ELMUNDO.es
Líder mundial en español | Miércoles 04/09/2013. Actualizado 16:27h.

INTERNET | Campus Party Europa 2013

'Es la década de los datos y de ahí vendrá la revolución'



<http://www.elmundo.es/elmundo/2013/09/03/navegante/1378243782.html>



Alex Pentland
MIT Media Lab



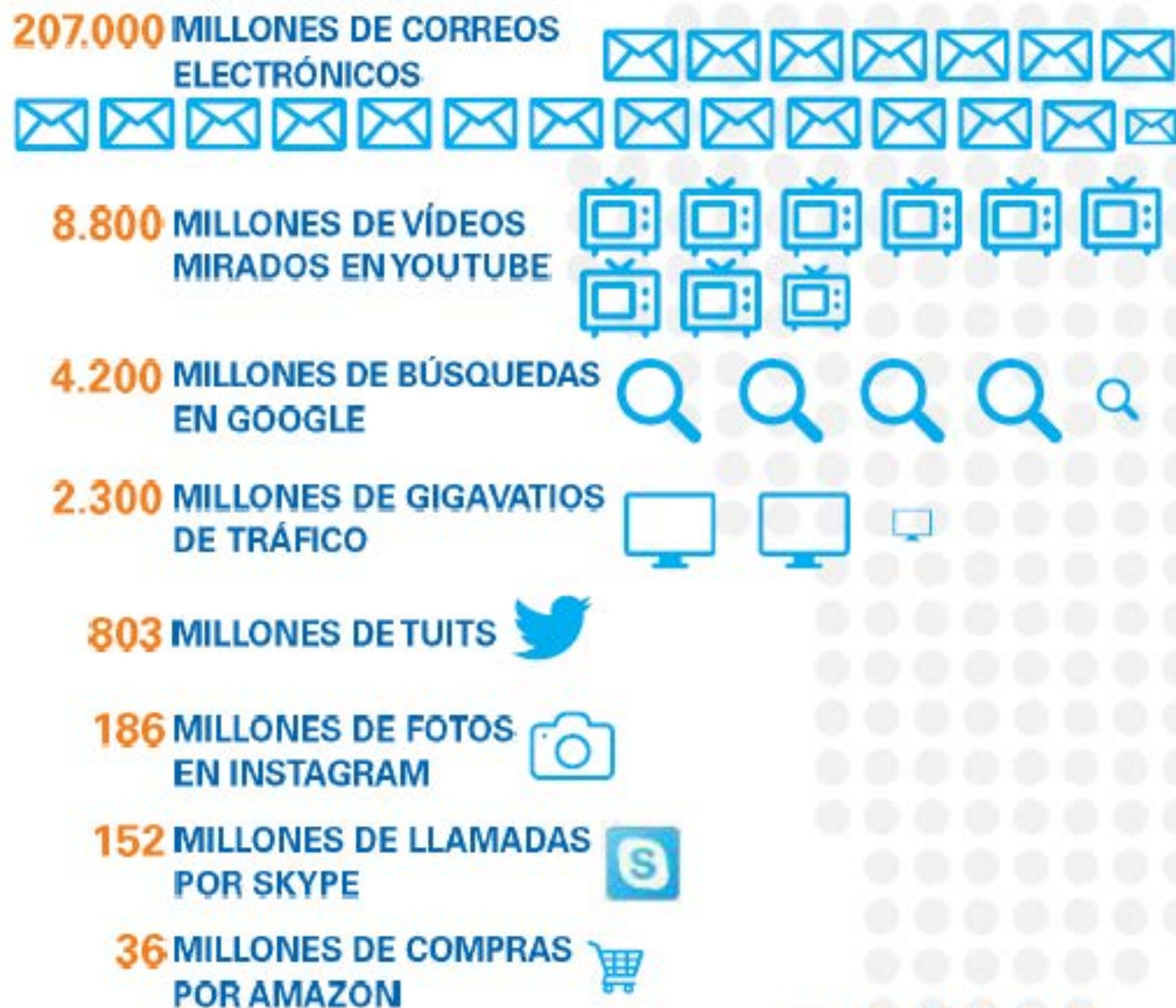
DIGITAL SINGLE MARKET
Digital Economy & Society

Towards a thriving data-driven economy

<https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>

<https://ec.europa.eu/digital-single-market/en/news/communication-data-driven-economy>

UN DÍA TÍPICO EN LA VIDA DE INTERNET*



*4 de abril de 2015.

NÚMERO DE USUARIOS DE INTERNET



PROPIETARIOS DE TELÉFONOS MÓVILES





La tecnología inteligente alza el vuelo

60% de la población mundial no está conectada



ACCESO A INTERNET



31%
población de las economías en desarrollo



80%
población de las economías avanzadas

SEIS TECNOLOGÍAS DIGITALES A SEGUIR

- TELÉFONOS MÓVILES 5G
- INTELIGENCIA ARTIFICIAL
- ROBÓTICA
- AUTOMÓVILES SIN CONDUCTOR
- OBJETOS FÍSICOS INTERCONECTADOS
- IMPRESIÓN 3D



Preparado por Natalie Ramírez-Djuanena. Texto y gráfico basados en Informe sobre el desarrollo mundial 2016: Dividendos digitales, publicado por el Banco Mundial en enero de 2016. El informe está disponible en www.worldbank.org/en/publication/wdr2016



La tecnología inteligente alza el vuelo



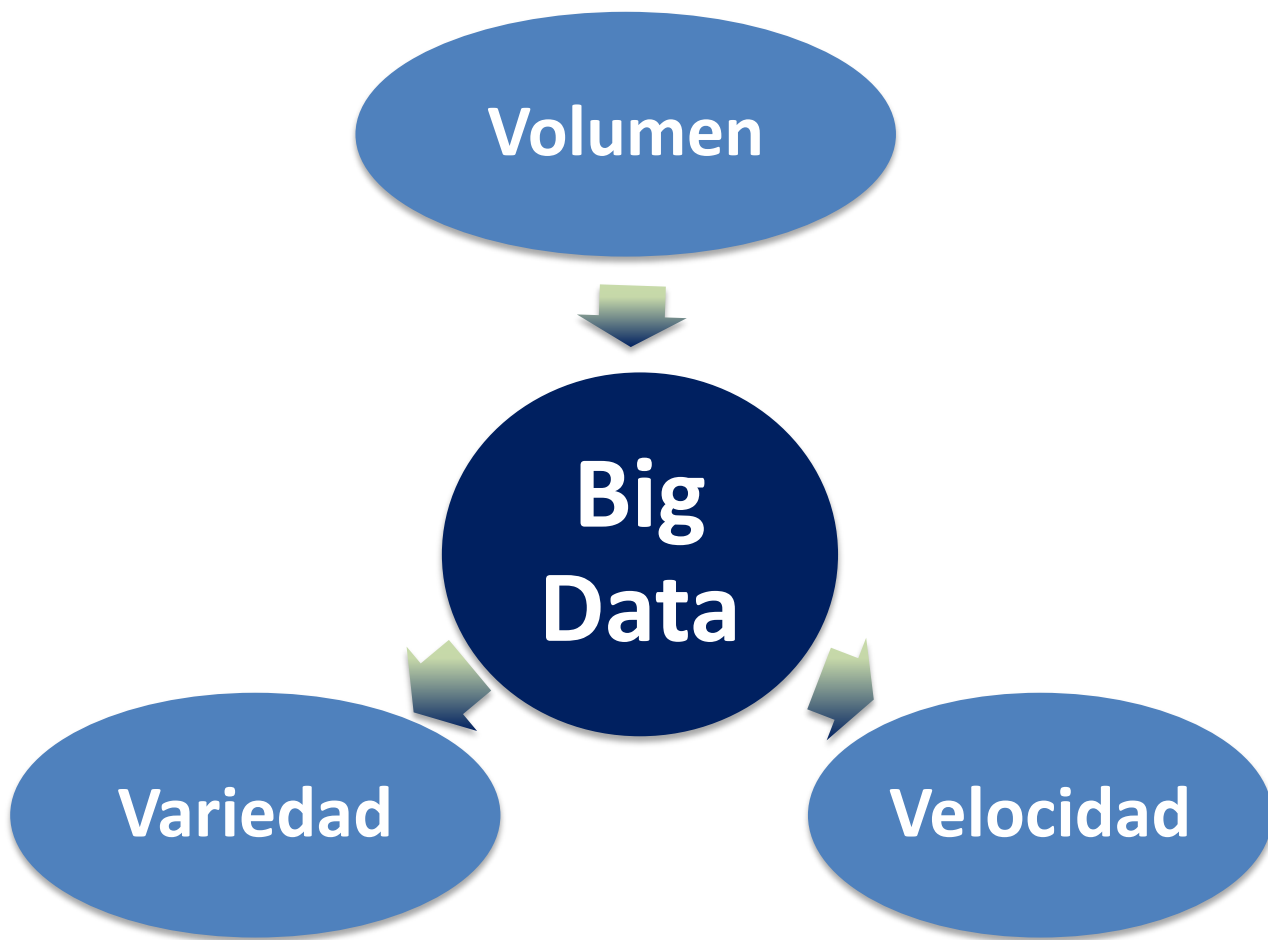
Cuando el planeta entero esté conectado, podemos esperar un aumento espectacular de la prosperidad humana.




Hal Varian
Chief Economist, Google

Hal Varian
Economista Jefe de Google

Big Data en 3 V's



 Lo difícil es determinar cómo traducir esos datos brutos en información que pueda servir para mejorar el rendimiento.

Hal Varian
Economista Jefe de Google




Doug Laney, Gartner Feb. 6, 2001

3-D Data Management: Controlling Data Volume, Velocity and Variety.

“Without Analytics, Big Data is just Noise”

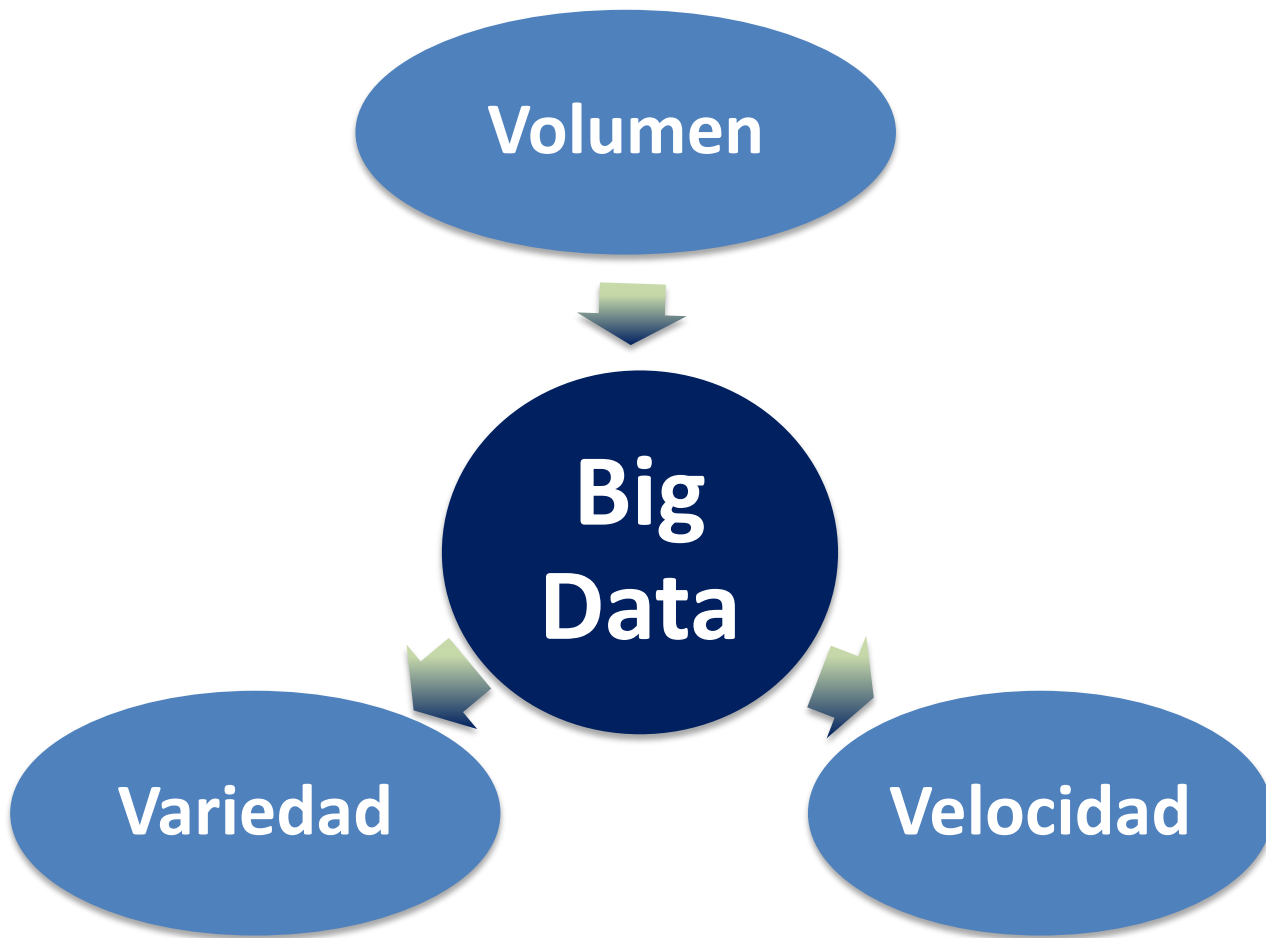
Guest post by Eric Schwartzman,
founder and CEO of [Comply Socially](#)



 Lo difícil es determinar cómo traducir esos datos brutos en información que pueda servir para mejorar el rendimiento.

Hal Varian
Economista Jefe de Google

Big Data en 3 V's



Ciencia de Datos (Analítica de datos): Ámbito de ciencia y tecnología que engloba las habilidades asociadas a la extracción de conocimiento de datos (incluyendo aprendizaje automático, visualización, tecnologías para big data, ...)

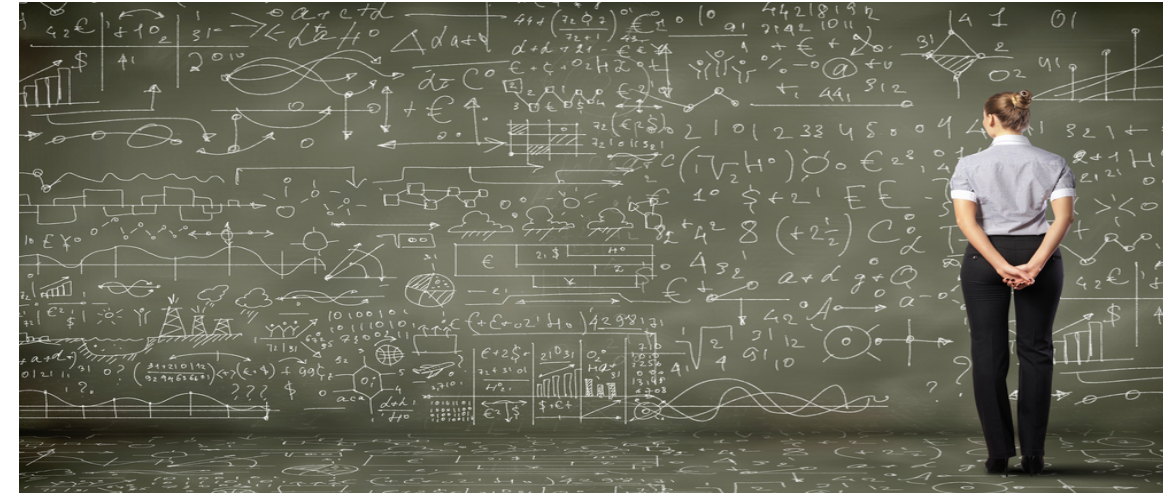
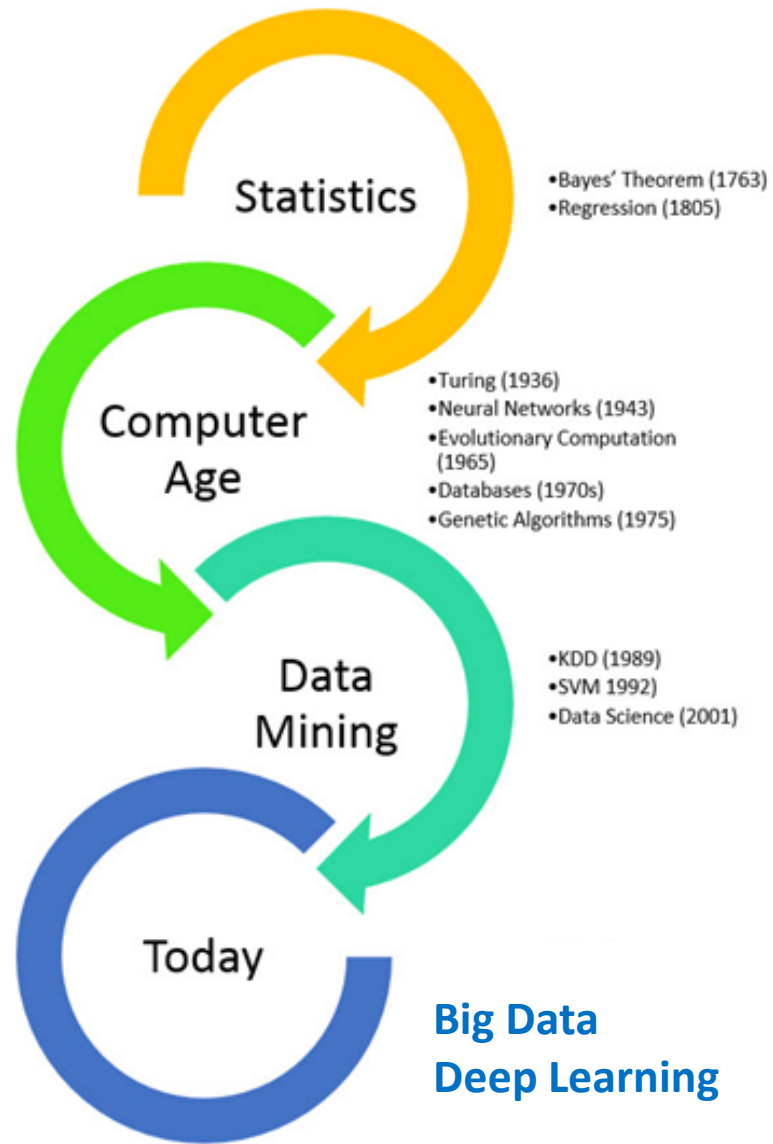


Doug Laney, Gartner Feb. 6, 2001

3-D Data Management: Controlling Data Volume, Velocity and Variety.

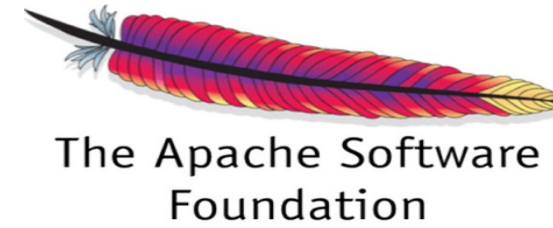


History of Data Science



2001-2010
2010-2016

BIG DATA Technology and Chronology



BIG DATA

2001
3V's Gartner
Doug Laney



2004
MapReduce
Google
Jeffrey Dean



2009-2013 Flink
TU Berlin
Flink Apache (Dec. 2014)
Volker Markl

Big Data



2008
Hadoop
Yahoo!
Doug Cutting

2010-2016:
Big Data
Analytics: Mahout, MLLib, ...
Hadoop Ecosystem
Applications
New
Technology



2010 Spark
U Berkeley
Apache Spark
Feb. 2014
Matei Zaharia



El poder de los datos

Introducción a Big Data y Ciencia de Datos

El poder de los datos: Casos en diferentes ámbitos

Analítica y tendencias tecnológicas

La realidad, lo mediático y el futuro

El poder de los datos: Análisis de asociaciones

Pañales y cerveza. Ficción y leyenda para ilustrar el análisis de transacciones



Si compro **cerveza**, entonces compro **pañales**

60%

Si compro **pañales**, entonces compro **cerveza**

100%



El poder de los datos: Análisis de asociaciones en Salud



El poder de los datos: Recomendaciones (Amazon, Netflix, ...)



Los datos incrementaron tremendamente las ventas
Ahora más de 1/3 de las ventas son gracias a las recomendaciones



Críticos y editores literarios
La voz de Amazon (1995)



Dilema: ¿Lo que los clics decían o lo que opinaban los críticos?



Greg Linde (1997) propuso un sistema de recomendaciones, filtrado colaborativo "artículo a artículo"



Netflix y Amazon son dos empresas cuyo plan de negocio está basada en big data y sistemas de recomendación

El poder de los datos: Predicción

Unos días después el director llamó al padre para disculparse.

Respuesta conciliadora del padre:

“He estado hablando con mi hija –dijo el padre– Resulta que en mi casa han tenido lugar ciertas actividades de las que yo no estaba del todo informado. Mi hija sale de cuentas en agosto. Soy yo el que les debe una disculpa”.



Análisis de transacciones
Fidelización de clientes



El poder de los datos: Información en Redes Sociales

Un sistema de análisis de datos de Twitter puede inferir aspectos de salud analizando 144M de tuits (2011-2013)



Se obtienen 13 grupos coherentes de mensajes correlacionados

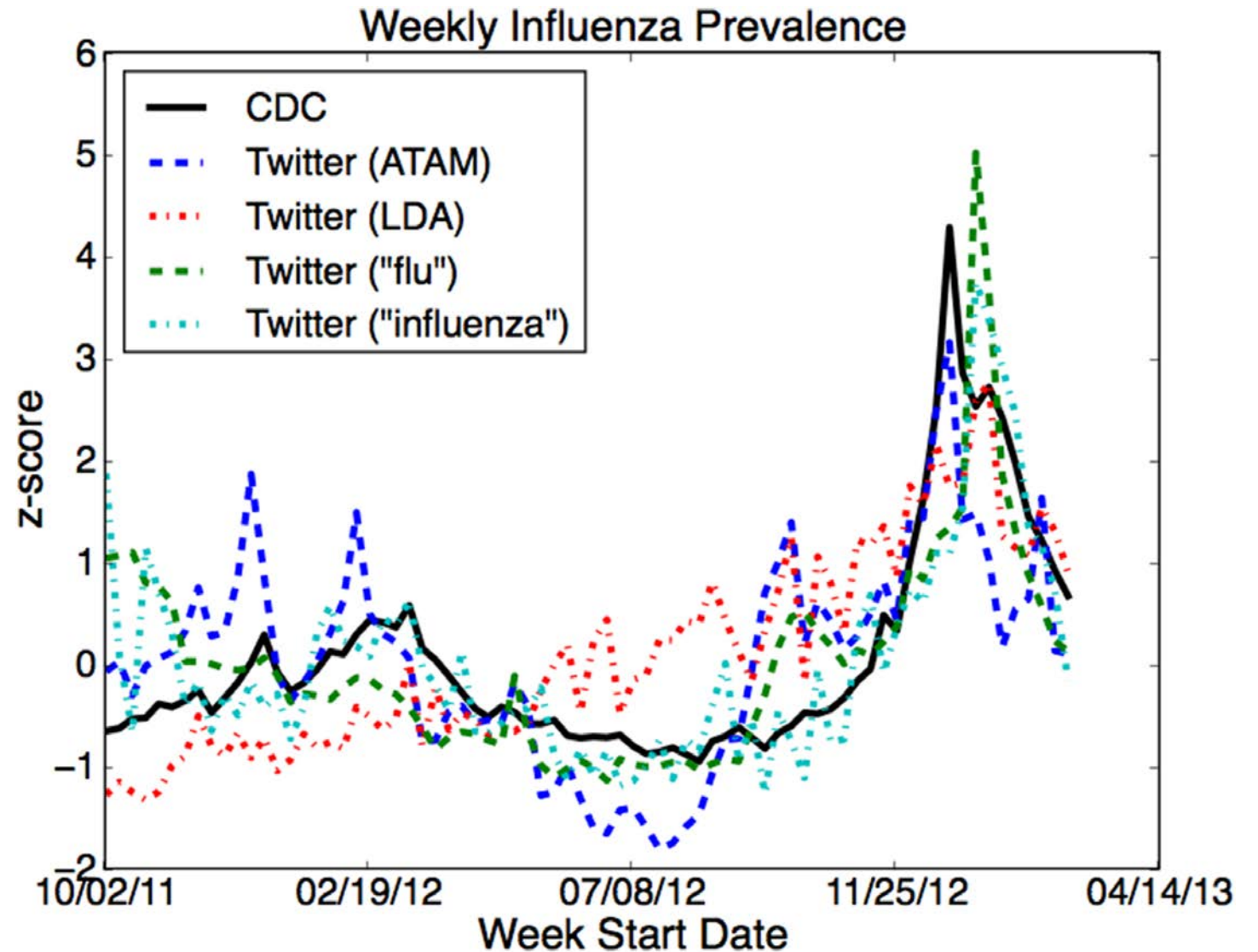
- Gripe estacional ($r= 0.689$) y alergias ($r = 0.810$)
- Ejercicio y obesidad relacionados con datos geográficos, ...

**Descubrimiento expansión enfermedades
Salud y Redes sociales**

Discovering Health Topics in Social Media Using Topic Models
Michael J. Paul, Mark Dredze, Johns Hopkins University, Plos One 9(8), 2014

El poder de los datos

Predicción de la pandemia de la gripe

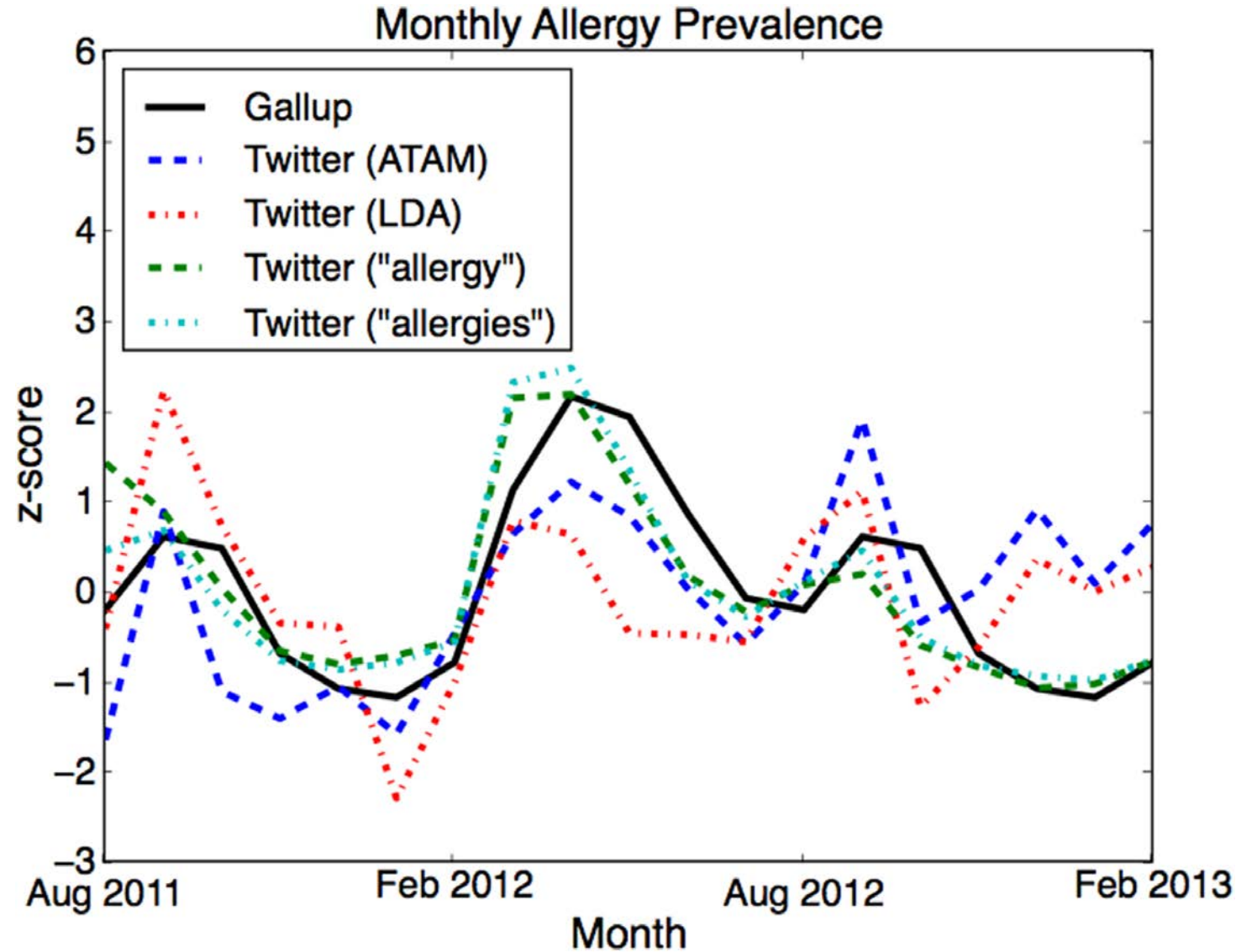


Descubrimiento expansión enfermedades Salud y Redes sociales

Discovering Health Topics in Social Media Using Topic Models
Michael J. Paul, Mark Dredze, Johns Hopkins University, Plos One 9(8), 2014

El poder de los datos

Incidencia de la alergia mensualmente



Descubrimiento expansión enfermedades Salud y Redes sociales

Discovering Health Topics in Social Media Using Topic Models
Michael J. Paul, Mark Dredze, Johns Hopkins University, Plos One 9(8), 2014

El poder de los datos: Wikipedia y la detección de la gripe

Detect pandemic risk in real time

El tráfico de consultas a la **Wikipedia** se puede utilizar para proporcionar en tiempo real la expansión de la **pandemia de gripe**.



¿Es mejor Wikipedia que Google para predecir la expansión de la gripe?

Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time

David J. McIver , John S. Brownstein, Harvard Medical School, Boston Children's Hospital, Plos Computational Biology, 2014.

El poder de los datos en Banca

Expansión.com

Domingo, 08.12.13. Actualizado a las 11:02

BBVA da un paso al frente en 'big data'

El banco ha actualizado todo su sistema de almacenamiento de datos corporativo para tener acceso desde un único punto a todas las 'islas de información' de la entidad en el mundo y consolidarse como un referente mundial en innovación tecnológica.

BBVA

Innova Challenge API

Available statistics services

Tendencias de negocio: Análisis de compras con tarjetas de crédito ...



El poder de los datos en Banca. Identificación de personas con las compras de tarjetas de crédito

PRIVACIDAD EN INTERNET »

Cuatro compras con la tarjeta bastan para identificar a cualquier persona

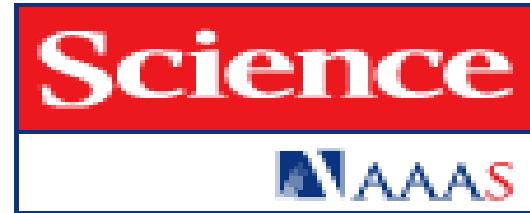
- Los patrones de uso de las tarjetas permiten descubrir la identidad del 90% de una muestra de 1,1 millones de personas anónimas, según demuestra un estudio del MIT



Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,^{1*} Laura Radaelli,² Vivek Kumar Singh,^{1,3} Alex “Sandy” Pentland¹

Large-scale data sets of human behavior have the potential to fundamentally transform the way we fight diseases, design cities, or perform research. Metadata, however, contain sensitive information. Understanding the privacy of these data sets is key to their broad use and, ultimately, their impact. We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals. We show that knowing the price of a transaction increases the risk of reidentification by 22%, on average. Finally, we show that even data sets that provide coarse information at any or all of the dimensions provide little anonymity and that women are more reidentifiable than men in credit card metadata.

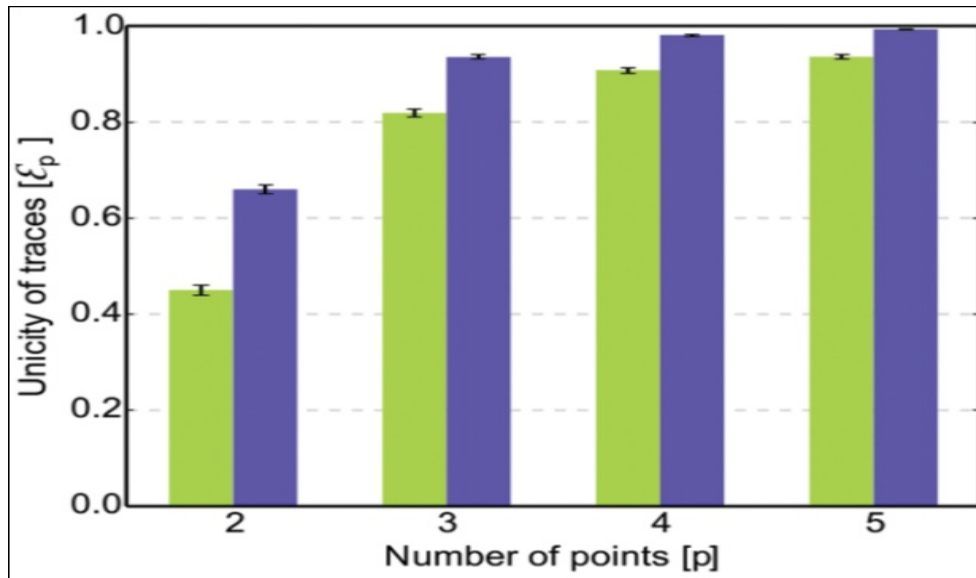


El poder de los datos en Banca. Identificación de personas con las compras de tarjetas de crédito

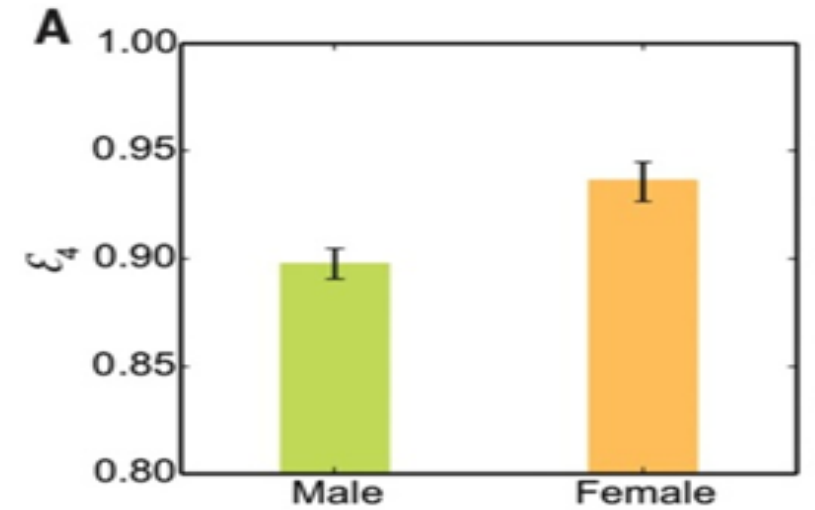
PRIVACIDAD EN INTERNET »

Cuatro compras con la tarjeta bastan para identificar a cualquier persona

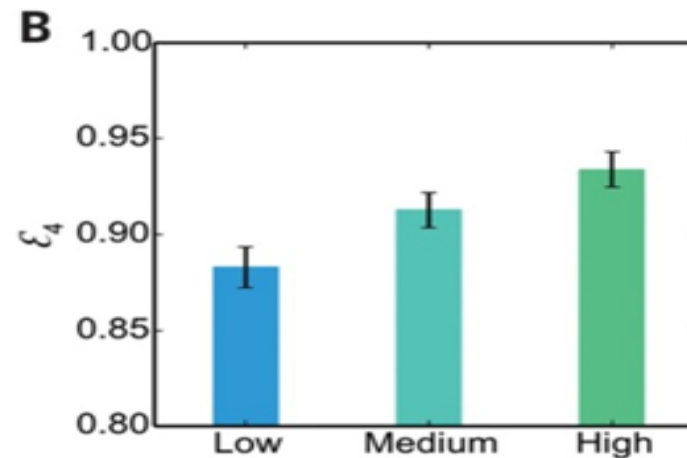
- Los patrones de uso de las tarjetas permiten descubrir la identidad del 90% de una muestra de 1,1 millones de personas anónimas, según demuestra un estudio del MIT



Identificación por el número de compras



Identificación por el género

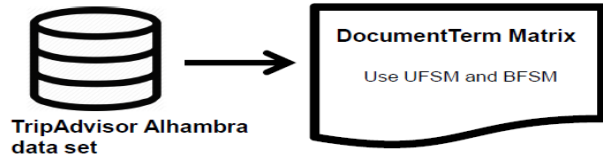


Identificación por el poder adquisitivo

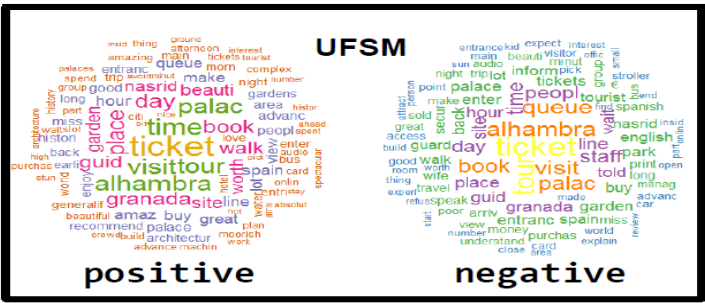
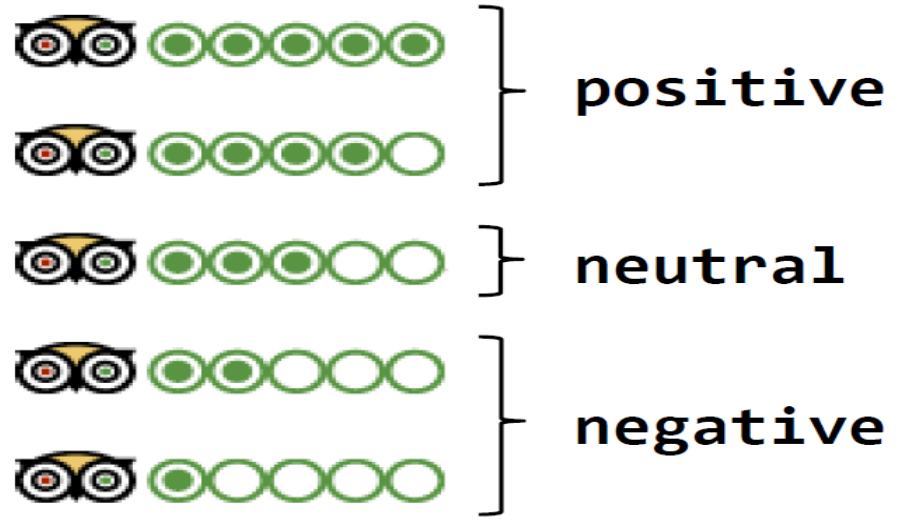
El poder de los datos en turismo

Sentiment Analysis For Touristic Attractions: A Case Study On The Alhambra

Ana Valdivia(Sept. 2016)



SentimentValue



SentimentCoreNLP

SentimentValue



El poder de los datos en turismo

Sentiment Analysis For Touristic Attractions: A Case Study On The Alhambra

Ana Valdivia(Sept. 2016)

SentimentValue	SentimentCoreNLP			Total
	positive	neutral	negative	
positive	4,049	1,071	2,508	7,628
neutral	51	32	260	343
negative	5	6	158	169
Total	4,105	1,109	2,926	8,140

Table 4.2: Correlation between SentimentValue and SentimentCoreNLP

53.08 % of coincidence



tripadvisor®

SentimentValue	SentimentCoreNLP			Total
	positive	neutral	negative	
positive	4,049	1,071	2,508	7,628
neutral	51	32	260	343
negative	5	6	158	169
Total	4,105	1,109	2,926	8,140

Table 4.2: Correlation between SentimentValue and SentimentCoreNLP

DISCOVERING INTERESTING PATTERNS IN
NEGATIVE OPINIONS

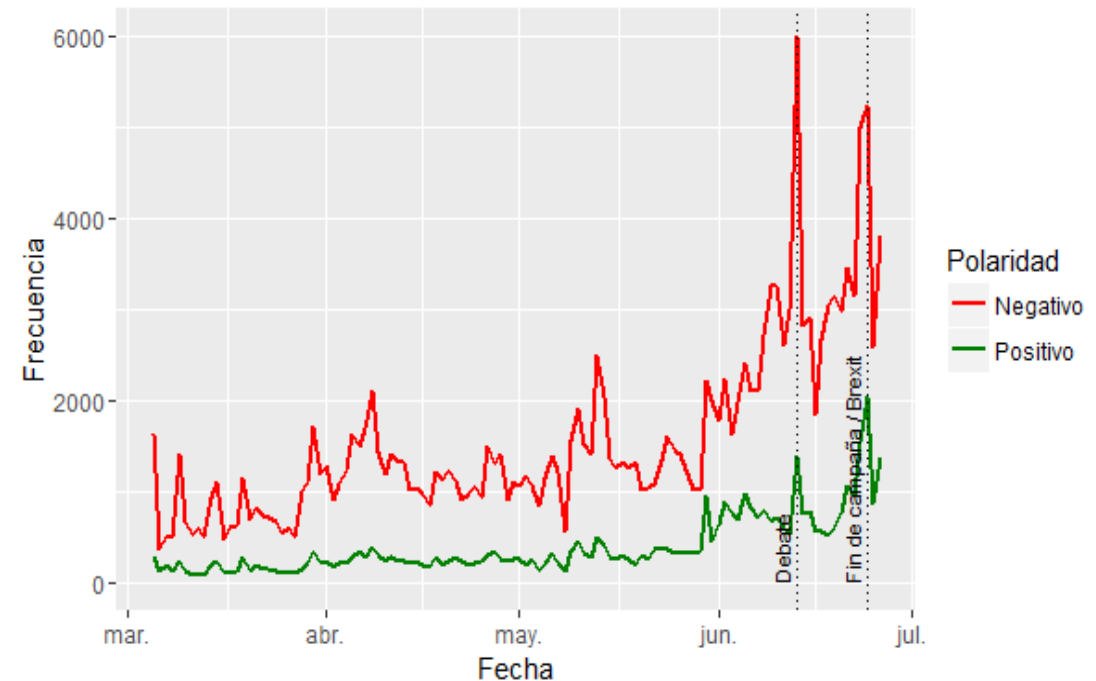
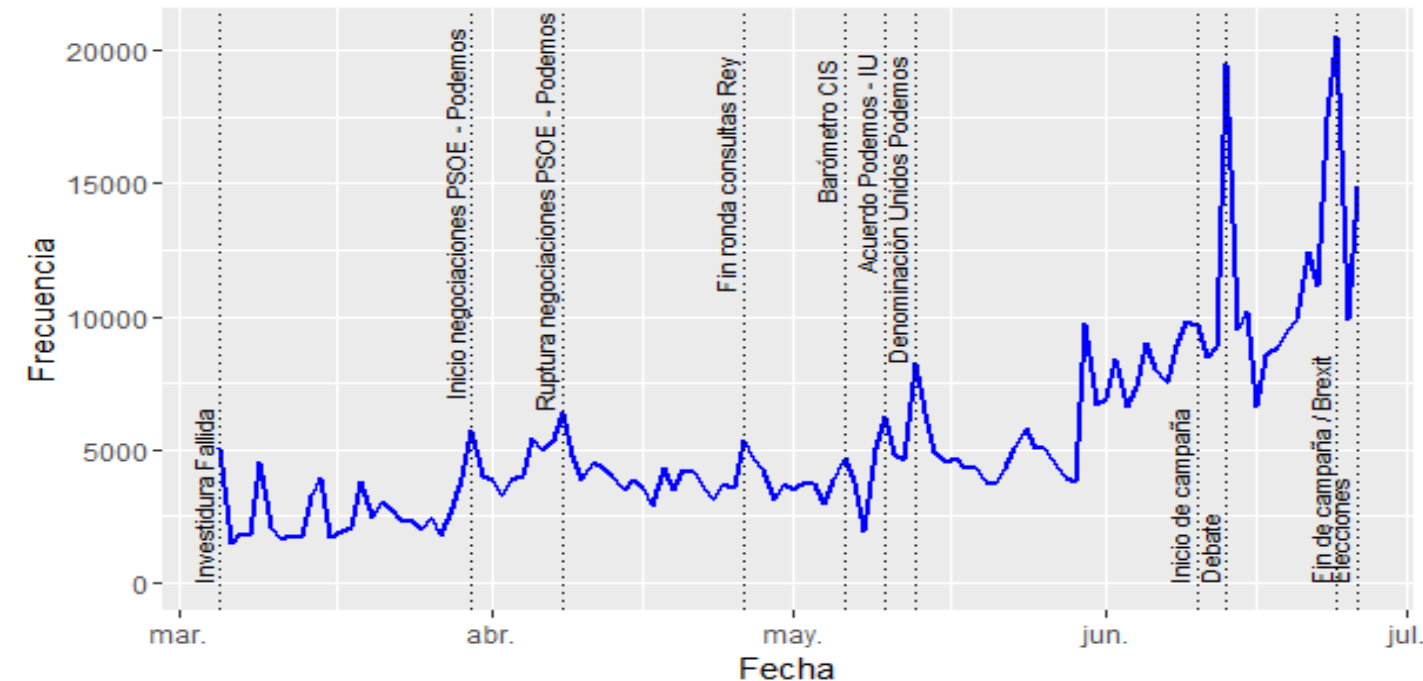
93.49 % of coincidence

El poder de los datos en política

Uso de técnicas de análisis de polaridad y fiabilidad de *tweets* en campaña electoral

Gabriel Quesada (Sept. 2016)

- Descarga de *tweets* a través de la *REST API* de *Twitter*.
- Los datos se obtienen a partir los *#hashtags* de los cuatro principales partidos y sus líderes.
- Período de recogida de datos: del 5 de marzo (tras la investidura fallida de Pedro Sánchez) hasta el 26 de junio (Elecciones Generales).
- Número total de *tweets* obtenidos: 1.593.624 (0,5 GB).



El poder de los datos

Introducción a Big Data y Ciencia de Datos

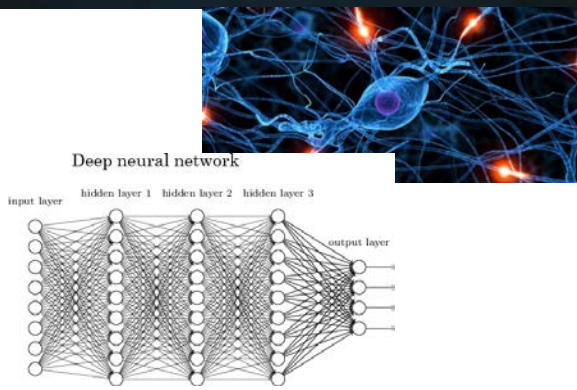
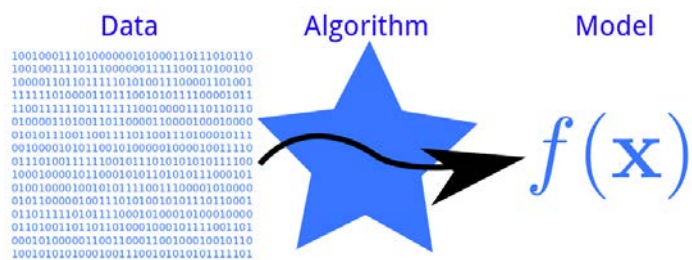
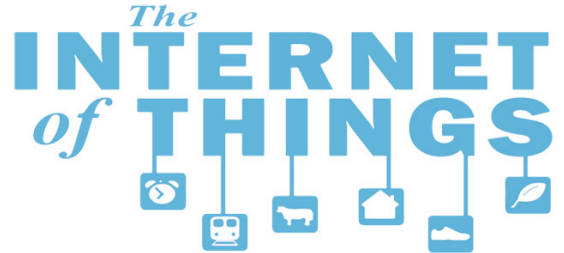
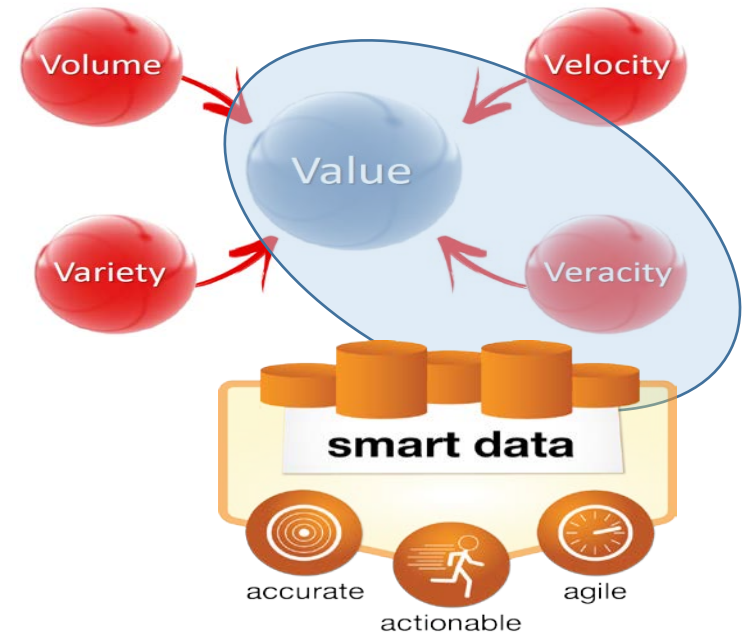
El poder de los datos: Casos en diferentes ámbitos

Analítica y tendencias tecnológicas

La realidad, lo mediático y el futuro



Analítica y Tendencias Tecnológicas



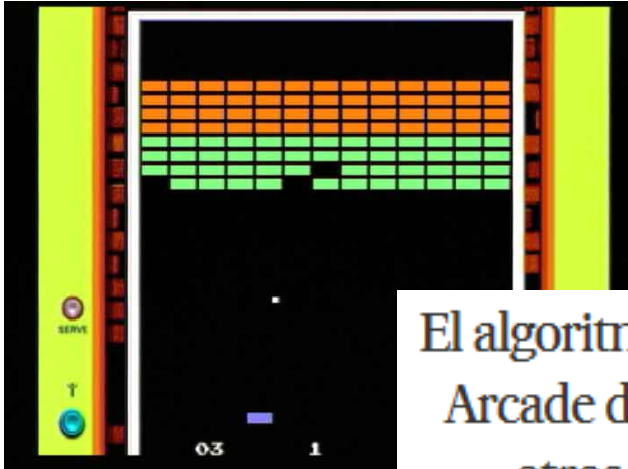
Aprendizaje automático
Ciencia de datos

Deep Learning

Inteligencia Cognitiva

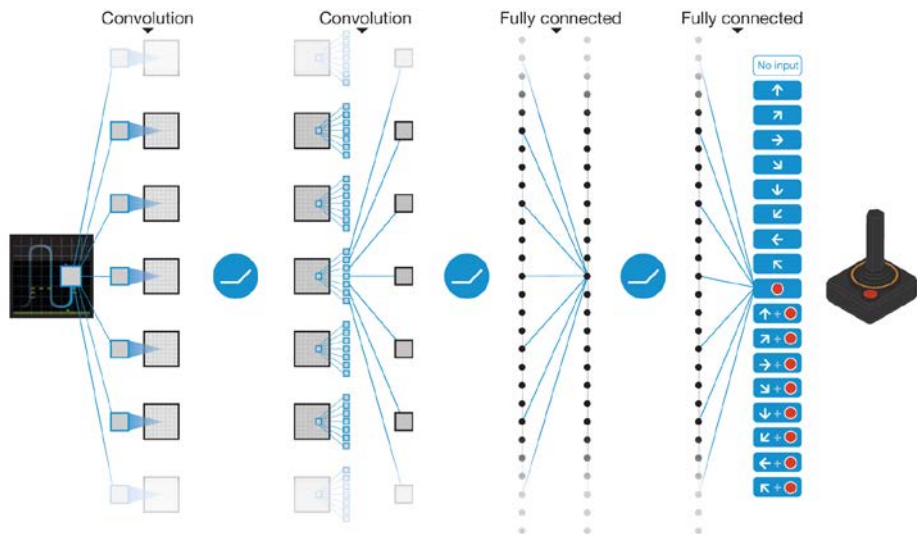
Social Big Data

Deep Learning: Google



Juegos Arcade
(Breakout)

El algoritmo se enfrentó a 49 juegos
Arcade de los años 80 superando a
otros algoritmos y a un jugador
humano profesional



La inteligencia artificial hace historia: AlphaGo, de Google, derrotó al campeón del juego chino go

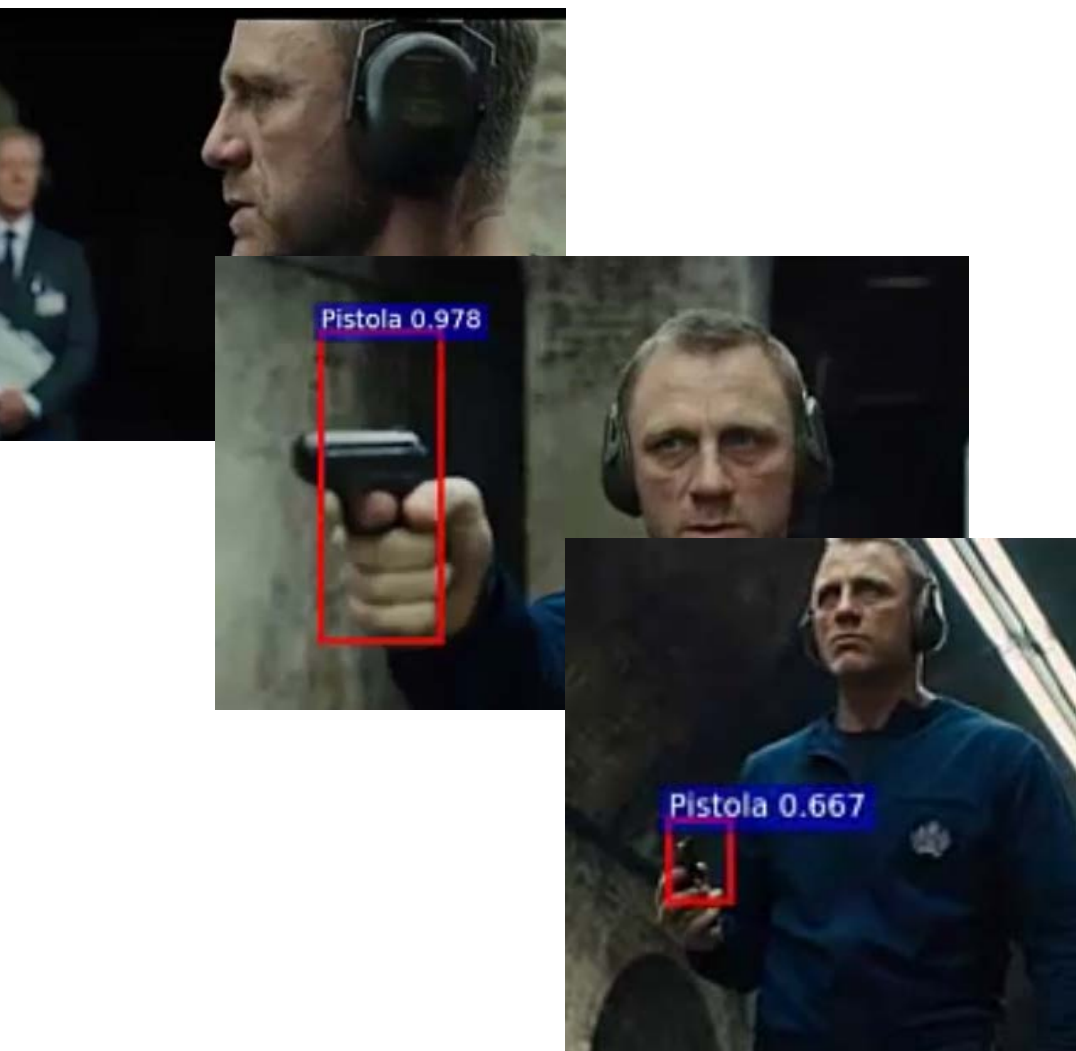
Publicado: 9 mar 2016 14:35 GMT

En un hito de la inteligencia artificial, la unidad DeepMind consiguió una victoria en la primera partida ante el surcoreano Lee Se-dol. "Estoy muy sorprendido, no esperaba perder", reconoció el jugador.



AFP PHOTO / GOOLGLE DEEPMIND

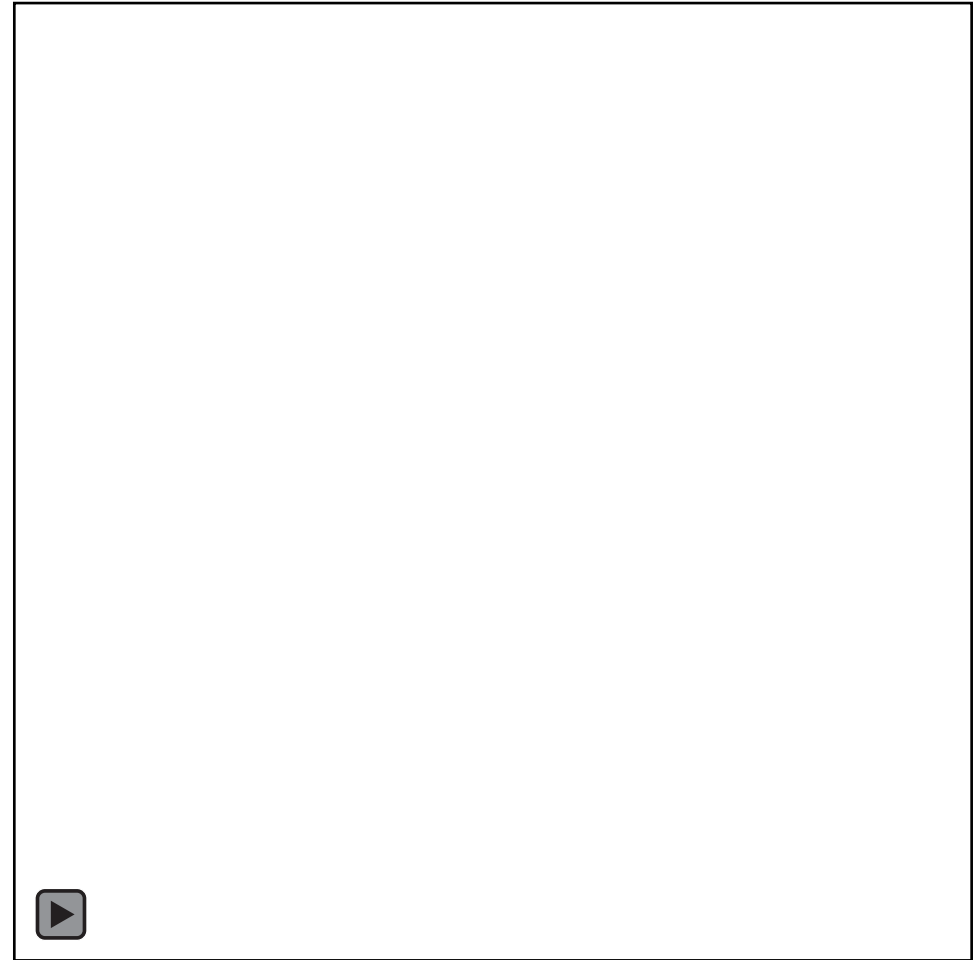
Deep Learning and Smart Data: Detección de Armas en Video



R. Olmos, S. Tabik, F. Herrera (UGR)

Video: Skyfall, 2012

Deep Learning and Smart Data: Detección de Armas en Video



El poder de los datos

Introducción a Big Data y Ciencia de Datos

El poder de los datos: Casos en diferentes ámbitos

Analítica y tendencias tecnológicas

La realidad, lo mediático y el futuro

Gran impacto en la sociedad y presencia en los medios de comunicación

EL MUNDO
Líder mundial en español | Miércoles 04/09/2013. Actu
INTERNET | Campus Party Europa 2013
'Es la década de los datos'
ahí vendrá la revolución



EL PAÍS
ECONOMÍA
ECONOMÍA EMPRESAS MERCADOS BO

El maná de los datos

- La conversión de datos en información útil por millones de dólares en 2015. La herramienta

SUSANA BLÁZQUEZ | Madrid | 29 SEP 2013 - 01

Archivado en: Citigroup Cap Gemini Sogeti IBM Telefónica Aplicaciones informáticas Tecn



EL PAÍS

ECONOMÍA

ECONOMÍA EMPRESAS MERCADOS BOLSA MIS AHORROS VIVIENDA TECNOLOGÍA OF

EMPRENEDORES »

El Big Data echa una mano al campo

- Una empresa española recoge miles de datos para predecir las cosechas

España necesitará 60.000 profesionales de Big Data hasta 2015

22 octubre, 2013 Eventos 18



España necesitará 60.000 profesionales de Big Data hasta 2015

"España va a necesitar alrededor de sesenta mil profesionales del Big Data de aquí a 2015", así lo ha asegurado Francisco Javier Antón, Subdirector General de Tecnologías del Ministerio de Educación, Cultura y Deportes en una mesa redonda sobre beneficio y aplicación de Big Data en pymes, moderada por Daniel Tapias de **Sigma Technologies**, celebrada durante el 4º Congreso Nacional de CENTAC de

Toledo.

Gran impacto en la sociedad y presencia en los

Expansión

economía digital

COMPañIAS ♦ PROTAGONISTAS ♦ INNOVACIÓN ♦ PREMIOS

ECONOMIA DIGITAL

Los directivos dudan de la fiabilidad de 'big data'

► El 'big data' revoluciona tu forma de buscar empleo

► ¿Podemos engañar al Big Data?



Sainsbury usaba los datos de compra de productos dietéticos como indicador de si sus clientes planeaban irse de vacaciones. | BLOOMBERG NEWS | EXPANSION

0 compartido 2 comentarios

Suscríbete

CincoDías

MARTES, 29 DE NOVIEMBRE DE 2016

Inicio

Mercados

Empresas

Economía

Tecnología

ESTÁ PASANDO:

IBEX 35

IPC

Banco Popular

Buscador Lotería Navidad

GESTIÓN

'Big data', de la teoría a la práctica

Empresas de sectores tradicionales como el inmobiliario ya apuestan por estas técnicas

Y después del 'big data', llegó el 'small data'

Sunil Gupta: "El poder del 'big data' se ha exagerado mucho"

PABLO SEMPERE DE LA PLATA - MADRID 26-11-2016 12:08

El poder de los datos: La realidad y las noticias mediáticas

MARCA

LaLiga Santander

Liga Santander El Real Madrid sería campeón y bajarían Depor, Granada y Osasuna

Así quedará LaLiga según el buscador de Microsoft



Actualizado 14/10/2016 11:11 CET (Central European Time)

El motor de búsqueda de Microsoft, Bing, se ha atrevido a predecir cómo va a quedar esta Liga.

Actualizado 14/10/2016

El Madrid campeón con 96 puntos

Posición	Equipo	Puntos
1	Real Madrid Club de Fútbol	96
2	Fútbol Club Barcelona	95
3	Club Atlético de Madrid	88
4	Sevilla Fútbol Club	70
5	Athletic Club	68
6	Real Club Celta de Vigo	64
7	Villarreal Club de Fútbol	64
8	Unión Deportiva Las Palmas	63
9	Málaga Club de Fútbol	62
10	Valencia Club de Fútbol	60
11	Real Betis	58
12	Sociedad Deportiva Eibar	56
13	Real Sociedad de Fútbol	54
14	Club Deportivo Leganés	40
15	Deportivo Alavés	38
16	Real Club Deportivo Español	33
17	Real Sporting de Gijón	31
18	Real Club Deportivo de La Coruña	29
19	Granada Club de Fútbol	21
20	Club Atlético Osasuna	17



“Qué puede hacer el hombre es tal vez una pregunta que puede dirigirse a la tecnología.

Pero qué hará el hombre debe preguntarse a la política, a la sociología o a la psicología”

Zygmunt Bauman

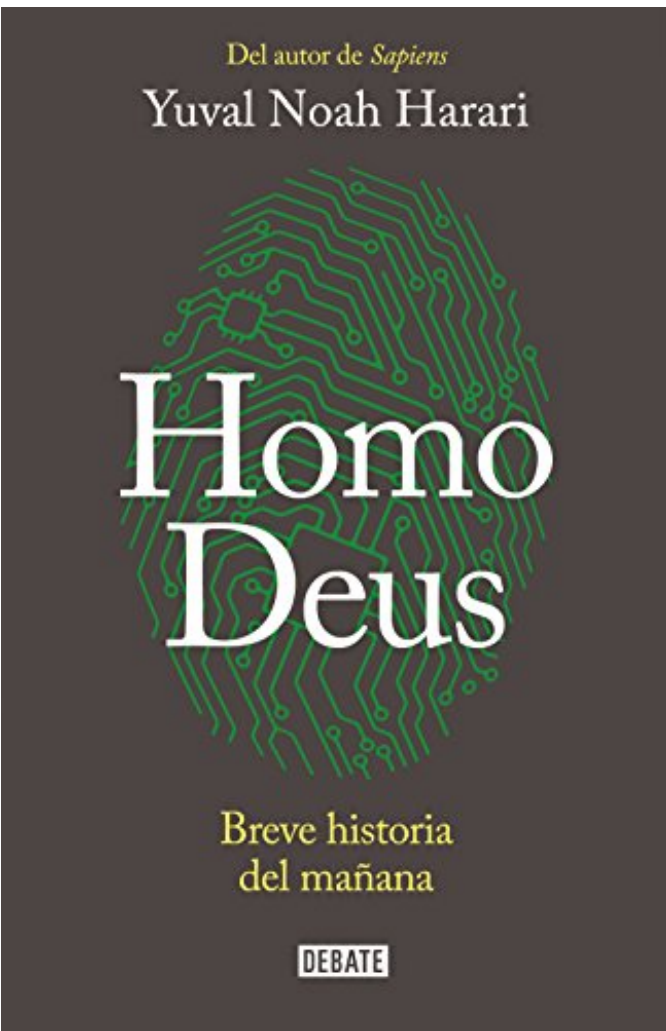
“Vivimos en dos mundos paralelos y diferentes: el online y el offline”

Copyright Clarín, 2014.

Zygmunt Bauman

Premio Príncipe de Asturias de Comunicación y Humanidades 2010



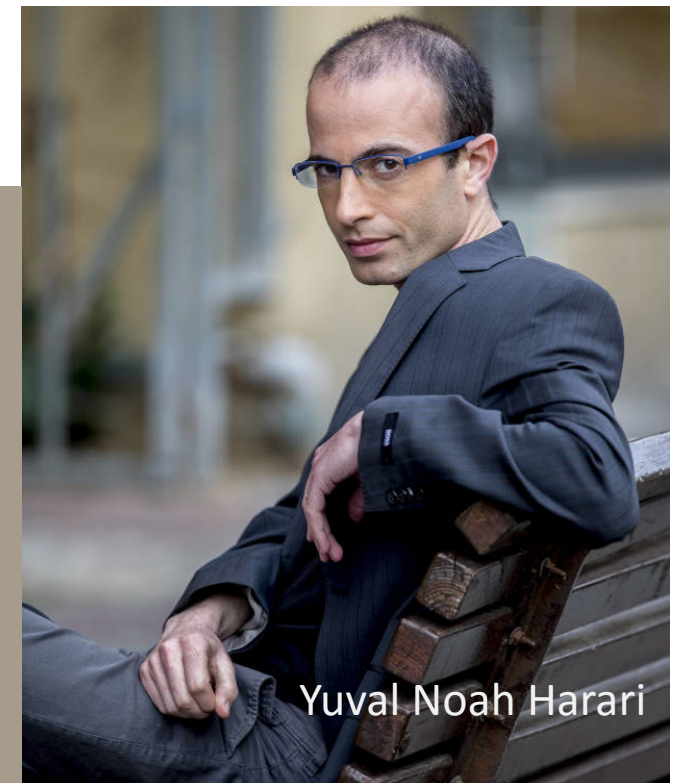
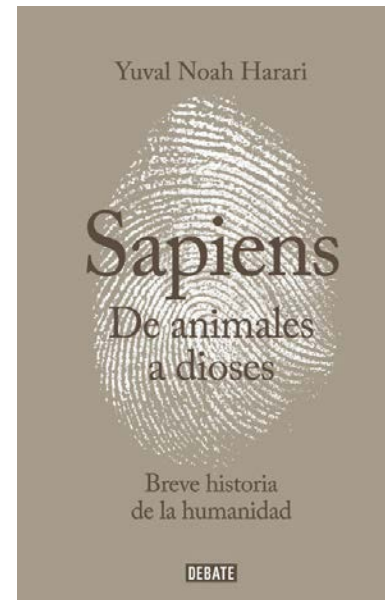


UNA LECTURA PARA EL DEBATE

“El mundo va a cambiar radicalmente gracias a los algoritmos, el big data y la inteligencia artificial”

Yuval Noah Harari:
"El poder está en manos de quien controla los algoritmos"
(yo diría los datos)

Lectura recomendada por Mark Zuckerberg



Yuval Noah Harari

Yuval Noah Harari, escritor e historiador israelí
Universidad Hebrea de Jerusalen

http://www.elconfidencial.com/alma-corazon-vida/2016-10-14/harari-poder-control-algoritmo-elite-salud_1274660



*Nube de datos
(Sergio García y Lola Moral,
Noviembre 2014)*

**¡Datos de calidad, analítica de datos,
conocimiento, decisiones de calidad!**

¡GRACIAS!