



Desmitificando el Big Data

QUÉ ES Y PARA QUÉ SIRVE LA CIENCIA DE DATOS.

FELIPE ORTEGA. DSLAB, CETINIA, URJC.

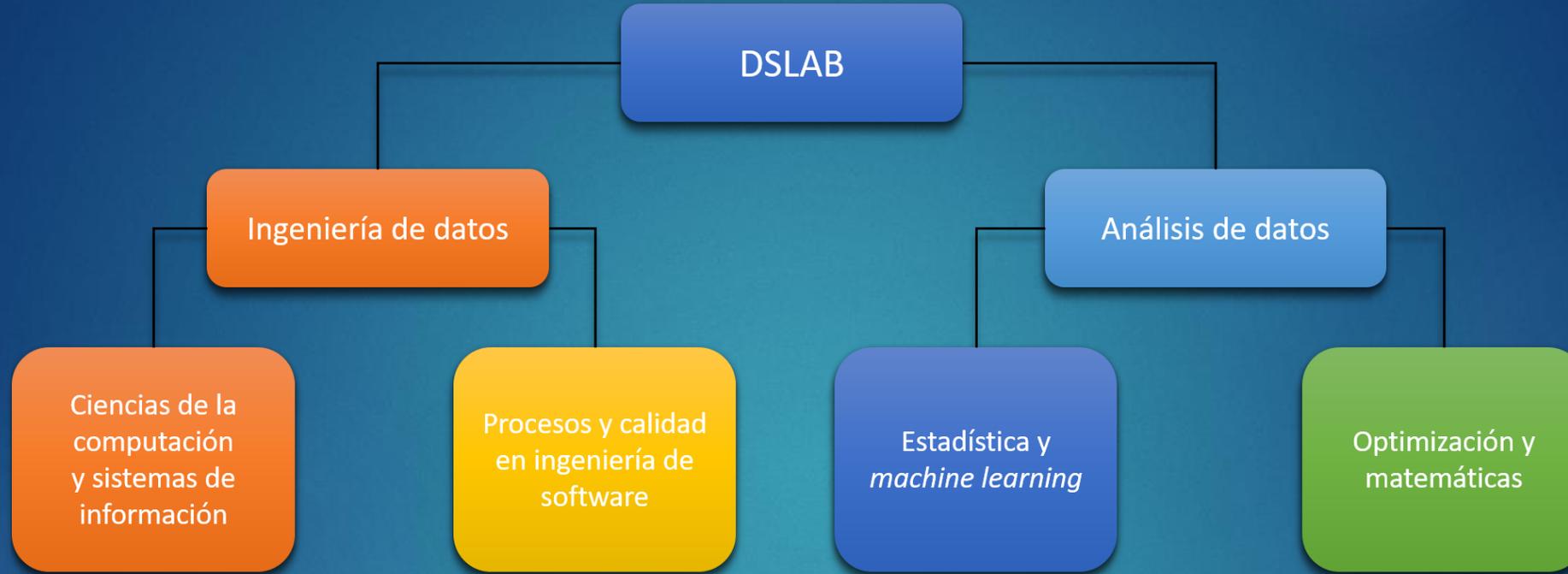
JORNADA "EXPERIENCIAS PRÁCTICAS – CALIDAD 4.0".

28-11-2019. AEC, MADRID.

Data Science Lab. CETINIA, URJC

- Grupo interdisciplinar.
- 15 profesores.
- 13 investigadores.
- Investigación en fundamentos y aplicaciones de la ciencia de datos.





Líneas de trabajo principales del DSLAB (URJC).



Isaac Martín de Diego
General Coordinator

Associate Prof. (Contratado Doctor)
Secretary, Master in Data Science
Dept. Computer Science and Statistics



Antonio Alonso Ayuso
Coordinator: Operations Research

Full Professor
Dept. Computer Science and Statistics



Javier M. Moguerza
Coordinator: Data Analytics

Full Professor
Dept. Computer Science and Statistics



Felipe Ortega
Coordinator: Data Engineering

Assistant Prof. (Ayudante Doctor)
Academic Dir., Master in Data Science
Dept. Signal & Comm.Theory & Telematic
Systems & Computing

Coordinadores DSLAB (URJC).

El mito del *big data*

LA SOBREABUNDANCIA DE DATOS

Transformación digital

- ▶ Tres áreas de impacto: cliente, negocio y empresa.
- ▶ Mejora de procesos.
- ▶ Liderazgo.
- ▶ Desarrollo de una cultura científica.
- ▶ Organización para toma de decisiones.
- ▶ Combinar **evidencias** con **experiencia**.

Transformación digital en la empresa española

69%

Empresas españolas que han comenzado algún proceso de transformación digital (más táctico que estratégico).

65%

Empresas españolas que identifican el big data como herramienta para mejorar su competitividad.

50%

Empresas españolas y directores que no se ven preparados para abordar el proceso de transformación digital.



Jon Piehl
@EduWebGuy

Seguir

Let data drive decisions, not the Highest Paid Person's Opinion (HIPPO)

#HowGoogleWorks

Ver traducción



0:51 - 22 ago. 2014



Data-driven decision-making

EVITAR SITUACIONES POCO SENSATAS...



El mito del big data

¿Tener más *datos* garantiza
obtener más *conocimiento*?



“

80% of my time was spent cleaning data. Better data will always beat better models.

THOMSON NGUYEN
([HTTP://SUPPORTVECTORHUMANS.COM/](http://supportvectorhumans.com/))



La realidad

- ▶ No hay “balas de plata” en big data.
- ▶ Cuanto más complicados son los datos, más difíciles de preparar.
- ▶ Muestrear no es pecado.
- ▶ Muchas técnicas estadísticas no están preparadas para big data.
- ▶ Si introducimos muchos datos, todo es “significativo”.
- ▶ Hay que separar la información del ruido.
- ▶ Hay que entender los atípicos, no solo el caso frecuente.

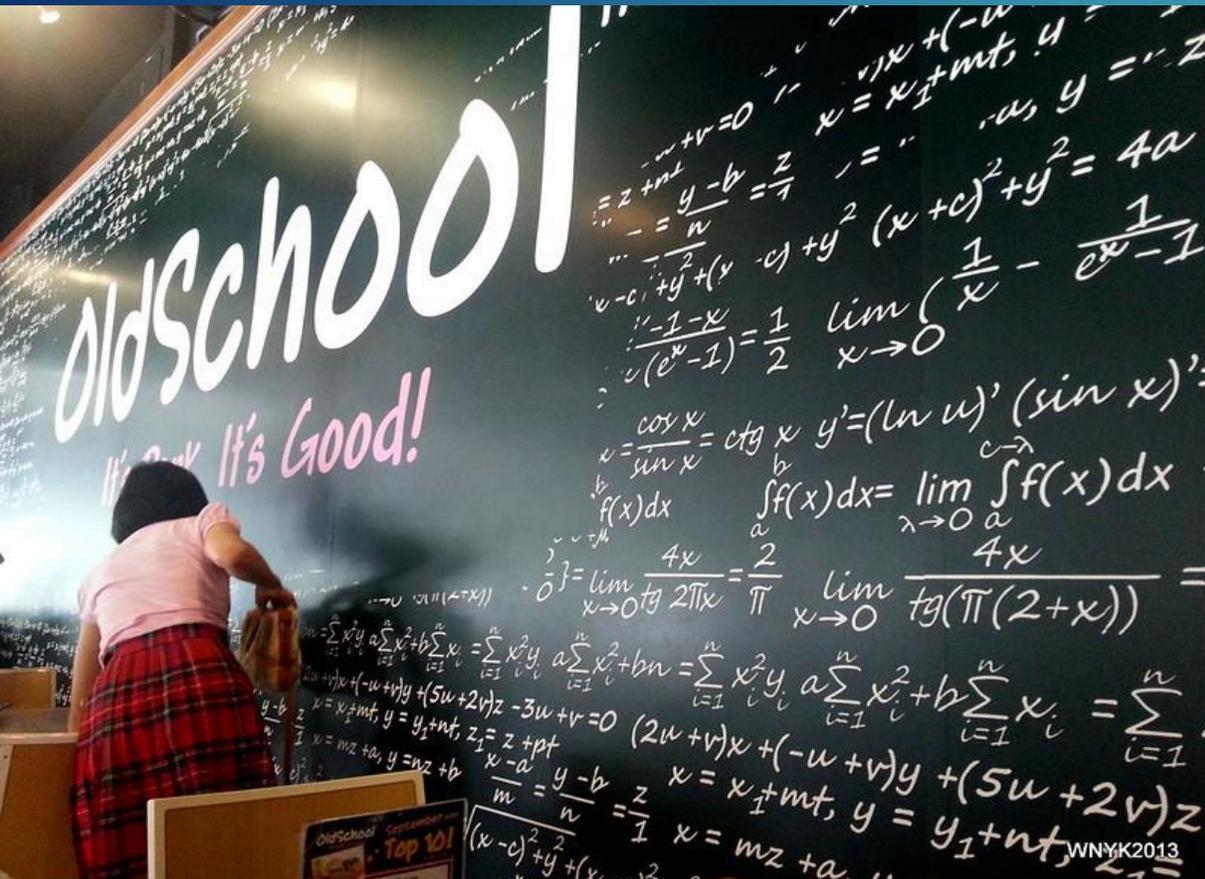
La ciencia de datos

RETOS Y OPORTUNIDADES PARA LA EMPRESA

Objetivos de la ciencia de datos

► INFERENCIA

► PREDICCIÓN/CLASIFICACIÓN



Pilares de la ciencia de datos



MULTIDISCIPLINAR

Combinación de conocimientos



COMPUTACIÓN

Software y sistemas de información



MATEMÁTICAS Y ESTADÍSTICA

Modelos, algoritmos, optimización



INVESTIGACIÓN

Aplicación del método científico



NEGOCIO

Conocimiento del área de aplicación



CALIDAD

Mejora de procesos empresariales



Ciclo de proyectos de ciencia de datos

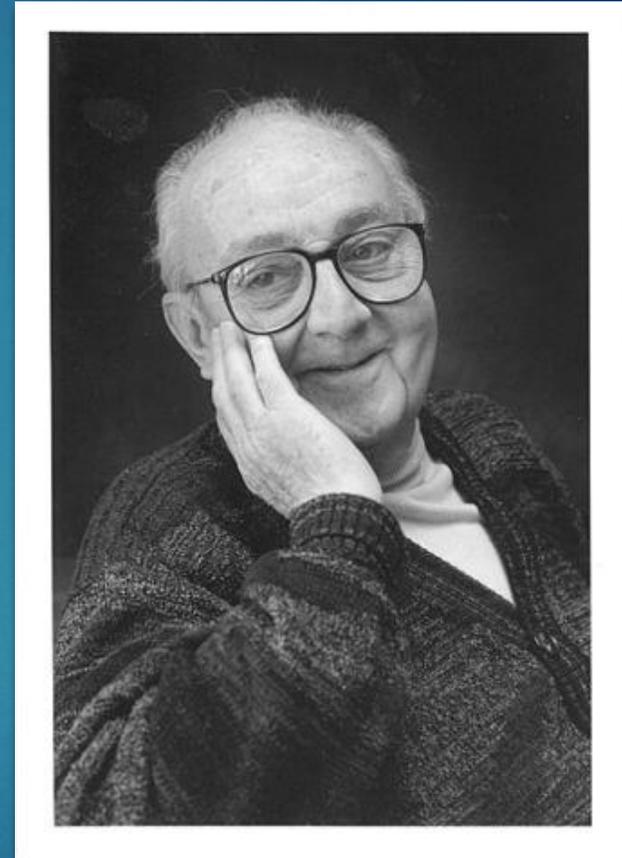
Calidad y gestión de los datos

- ▶ Partir siempre de una buena pregunta.
- ▶ Importa tener datos de calidad.
- ▶ Importan más las personas que las herramientas.
- ▶ No desperdiciar conocimiento de negocio acumulado.
- ▶ Automatizar procesos siempre que sea posible.
- ▶ Reutilizar datos, modelos y código.

“

All models are
wrong.
Some models are
useful.

GEORGE E. P. BOX



Gracias

FELIPE ORTEGA. DSLAB, CETINIA, URJC.

FELIPE.ORTEGA@URJC.ES

