



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Diplomado en Big Data y Ciencia de Datos

## *Curso: Ciencia de Datos y sus Aplicaciones*

### Clase 01: Introducción y Motivación

Jaime Caiceo, Roberto González  
y Roberto Muñoz

✉ [jcaiceo@uc.cl](mailto:jcaiceo@uc.cl)

✉ [regonzar@uc.cl](mailto:regonzar@uc.cl)

✉ [rmunoz@uc.cl](mailto:rmunoz@uc.cl)

El uso de apuntes de clases estará reservado para finalidades académicas. La reproducción total o parcial de los mismos por cualquier medio, así como su difusión y distribución a terceras personas no está permitida, salvo con autorización del autor.



## Objetivos

- Descubrir y evaluar oportunidades de negocio accionables a partir del análisis de datos.
- Entender y aplicar los conceptos y métodos fundamentales de Data Science a problemas reales de negocio, interpretando adecuadamente los resultados y generando acciones de valor agregado.
- Desarrollar modelos descriptivos, predictivos y prescriptivos sobre datos de diversas industrias.



## Evaluación

- Evaluaciones:
  - Se realizarán **2 trabajos en grupo**
    - Modelo de Predicción de Fuga de Clientes
    - Modelo de Segmentación
  - Se evaluará también la **participación en clases**
- Nota Final del Curso:
  - 80% Promedio Trabajos + 20% Nota Participación





# Planificación

Clase	Temas
1	Clases y evaluaciones Introducción y Motivación
2	Discusión metodologías: CRISP y TDSP Proyecto individual de detección de anomalías
3	Discusión proyecto detección anomalías
4	Industria financiera. Modelos de fuga de clientes. Comenzar proyecto grupal predicción de fuga
5	Desarrollo de proyectos grupales
6	Presentación alumnos de proyecto predicción de fuga Cierre análisis del proyecto
7	Segmentación de clientes Comenzar proyecto grupal de clustering
8	Discusión proyecto clustering





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL



Motivación e Introducción

# CONTEXTO



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

## Datos



**3,596,745,387**

Internet Users in the world



**1,168,291,454**

Total number of Websites



**186,023,704,407**

Emails sent **today**



**4,129,452,089**

Google searches **today**



**3,857,666**

Blog posts written **today**



**524,933,660**

Tweets sent **today**



**4,768,172,638**

Videos viewed **today**  
on YouTube



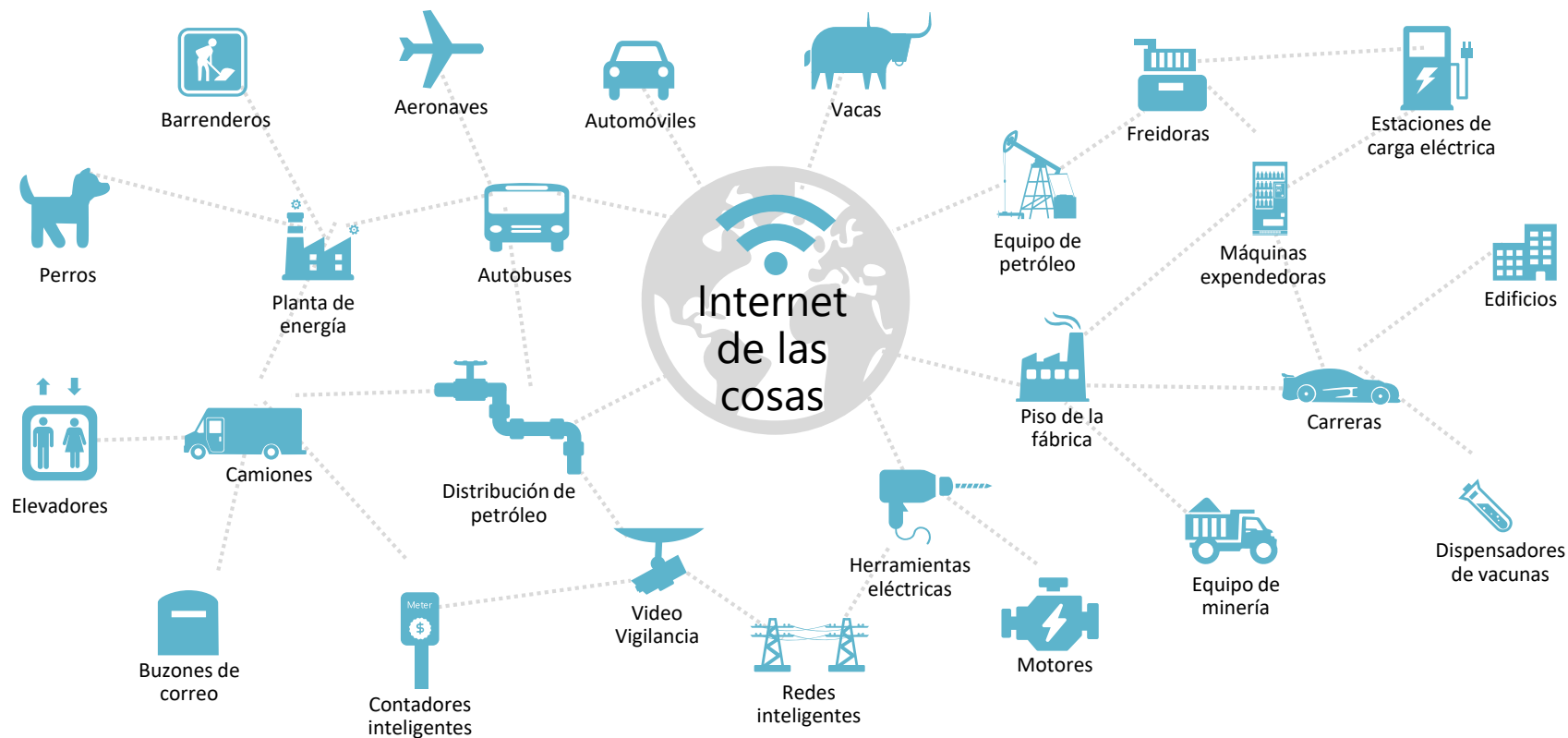
**53,688,557**

Photos uploaded **today**  
on Instagram



**84,978,753**

Tumblr posts **today**





# Big Data

“

**Big data es alto volumen, alta-velocidad y alta-variedad** de componentes de datos que demandan herramientas de procesamiento de información **costo-efectivas**, que habiliten **mayor inteligencia, toma de decisiones y automatización de procesos.** ”

– Gartner, Big Data Definition (\*)

(\*) Big Data (Gartner, 2016), <http://www.gartner.com/it-glossary/big-data/>







## Motivación al Análisis

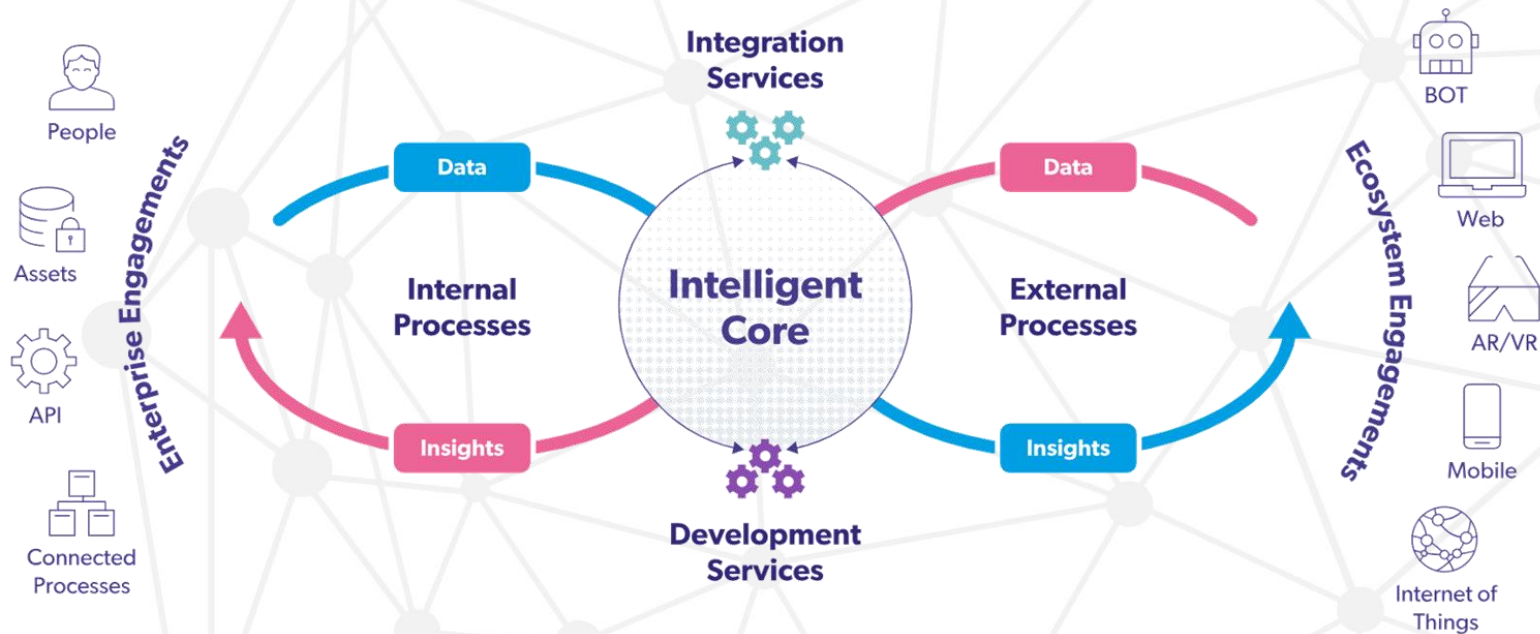
“Big data is not about the data!”

Gary King, Harvard University  
Noviembre 2016

La verdadera revolución no es sobre los datos,  
sino sobre el impresionante progreso en la  
estadística y otros métodos para extraer ideas  
(insights) desde los datos.

# Transformación basada en inteligencia

## ... Los datos no son la parte más importante



Fuente: IDC, Noviembre 2017



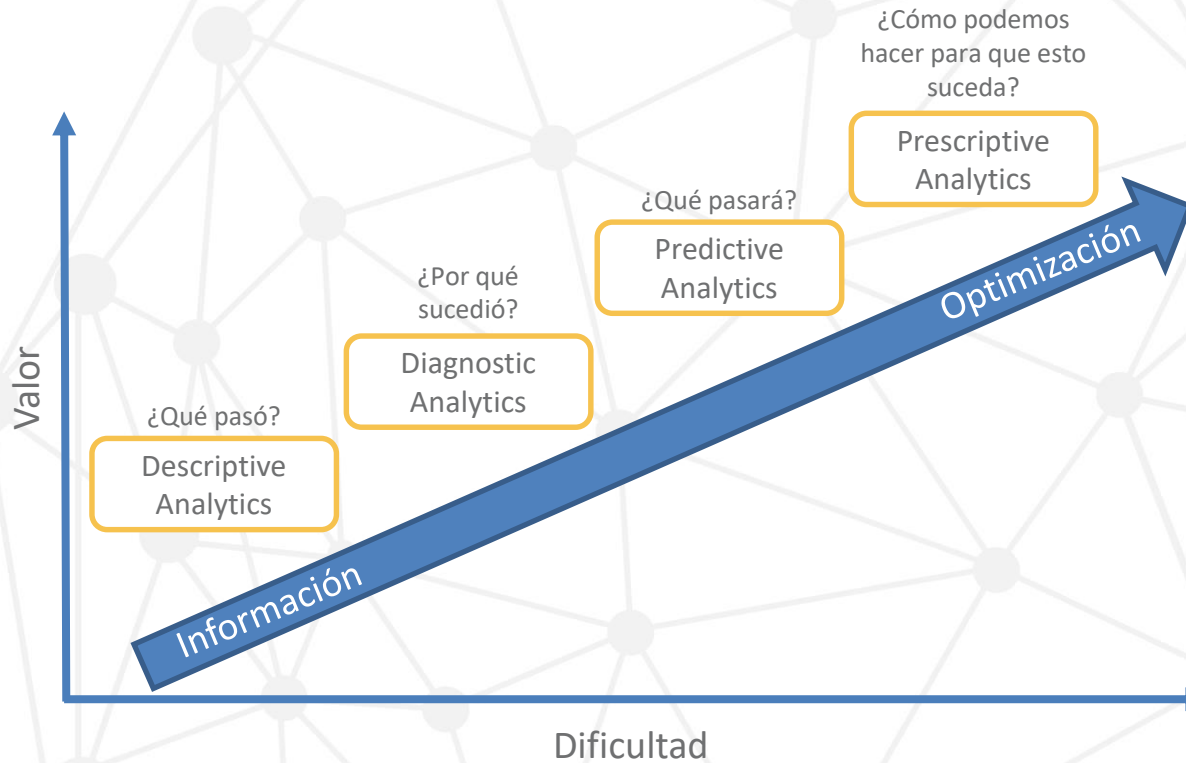
# Big Data Analytics

Desde los datos tomar decisiones y acciones...





# La evolución de los temas analíticos

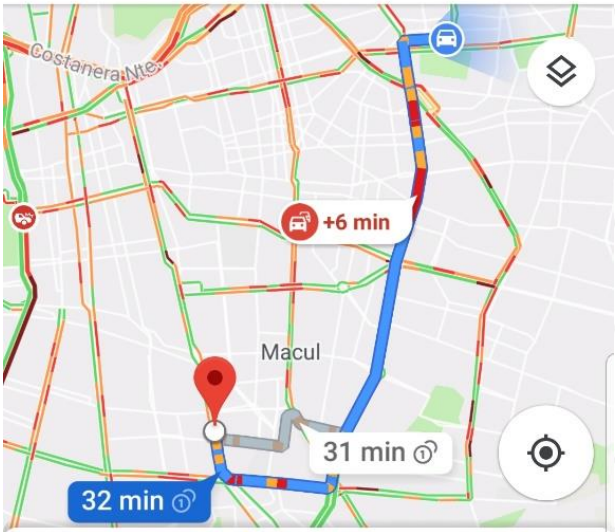
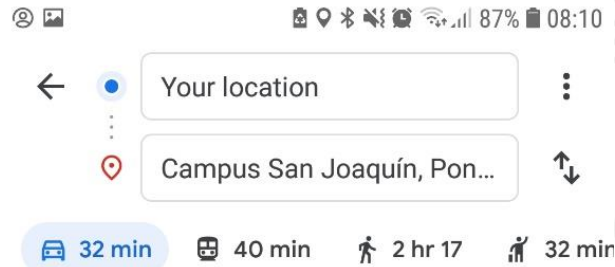


Fuente: Gartner Business Intelligence & Analytics Summit 2013



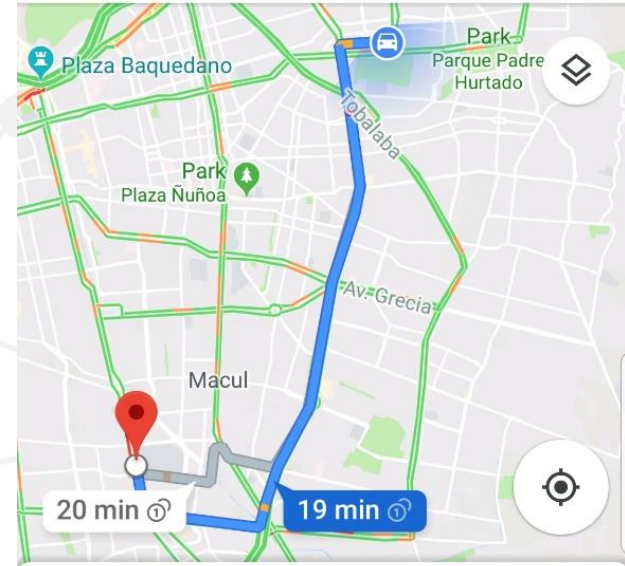
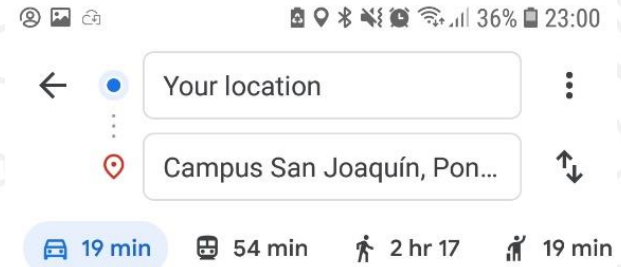


# Tipos de Análisis



**32 min** (13 km)

Best route, despite heavy traffic, but it's getting better



**19 min** (13 km)

Fastest route, the usual traffic



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL



Motivación e Introducción

**¿QUÉ HA PASADO EN LOS ÚLTIMOS 4 AÑOS?**



# Google DeepMind Challenge Match



vs

Lee Sedol\*

**AlphaGo Lee ganó 4-1**

- ▶ AlphaGo: una combinación de redes neuronales y algoritmos.
- ▶ Entre el 9 y 15 de Marzo del 2016.

(\*) 18 veces campeón mundial de Go.







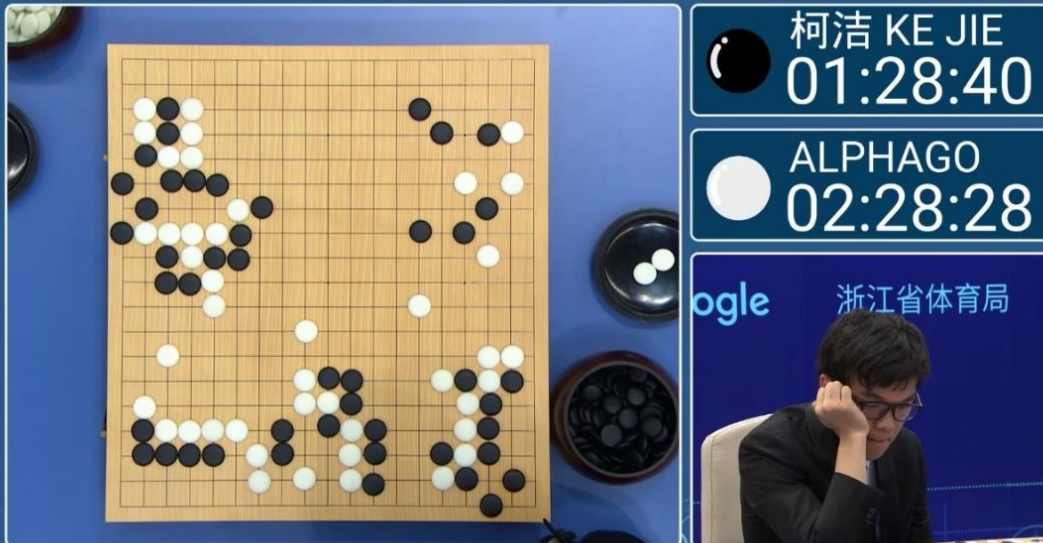
# Future of Go Summit, Wuzhen, China



vs

Ke Jie\*

**AlphaGo Lee Master  
ganó 3-0**



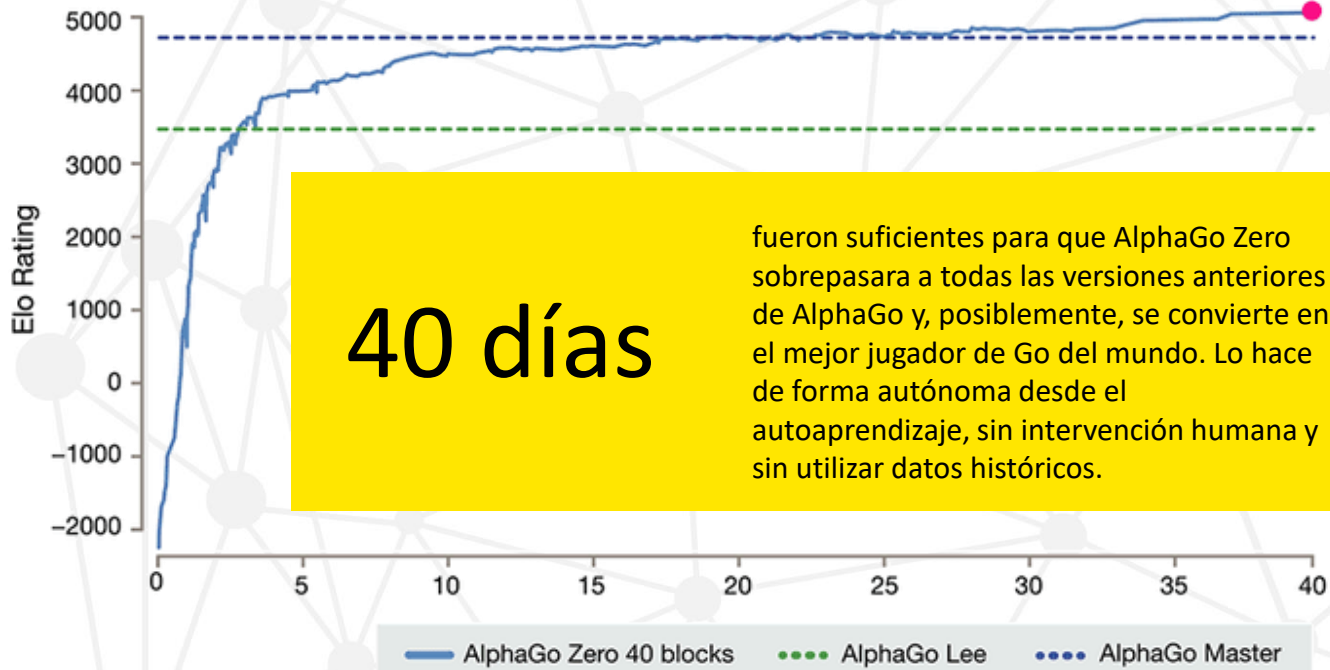
- ▶ Google anunció luego de esto que el equipo de investigación de AlphaGo se enfocaría en abordar problemas más complejos y que afectan al mundo.
- ▶ 23, 25 y 27 de Mayo del 2017.

(\*) Jugador nro. 1 del ranking mundial de Go.





# AlphaGo Zero





## ¿Qué ha sucedido últimamente?

96%

RESNET vision  
test 152 layers

**2016**

**Object recognition  
Human parity**

94.9%

Switchboard speech  
recognition test

**2017**

**Speech  
recognition  
Human parity**

88.5%

SQuAD reading  
comprehension  
test

**Enero 2018**

**Machine reading  
comprehension  
Human parity**

69.9%

MT research  
system

**Marzo 2018**

**Machine  
translation  
Human parity**

98.6%

Neutral TTS MOS  
similarity to human  
recording

**Agosto 2018**

**Speech synthesis  
near-human parity**



# The Guardian

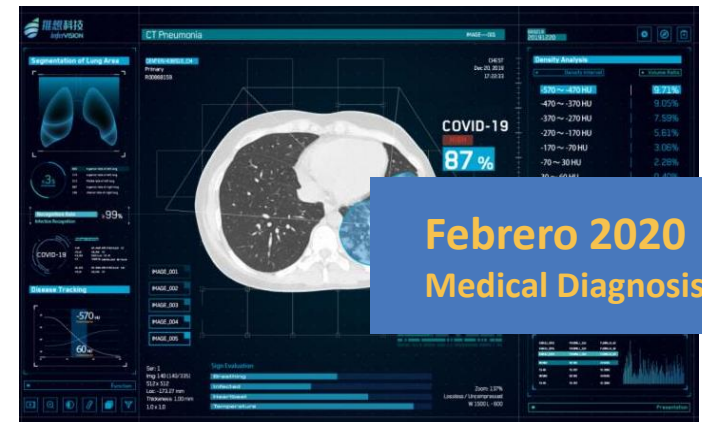
## Culture

## Research suggests AI able to interpret medical images using deep learning algorithm

**WIRED**

BUSINESS 02.26.2020 07:00 AM

Software that reads CT lung scans had been used primarily to detect cancer. Now it's retooled to look for signs of pneumonia caused by coronavirus.



## Febbrero 2020

### Medical Diagnosis





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

## Algunos desafíos

PROPUBLICA



### Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica  
May 23, 2016

Sesgo

The New York Times

### *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*



Ética





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL



Motivación e Introducción

# VÍNCULO CON LA TRANSFORMACIÓN DIGITAL

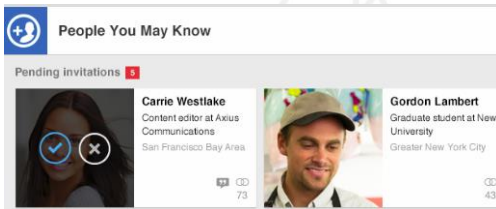


ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

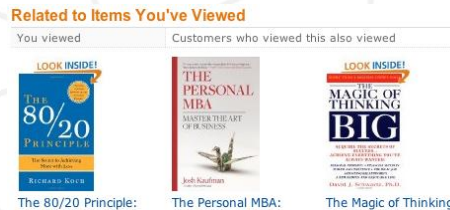
EDUCACIÓN  
PROFESIONAL

# Aparición de productos analíticos

LinkedIn



amazon.com



UBER





CATERPILLAR




# Caterpillar

IN THE AGE OF SMART IRON

**SMARTER FLEETS.  
SAFER CREWS.**



“Nuestra estrategia digital es una emocionante inversión para el largo plazo. Trabajamos duro, dentro de Caterpillar y con nuestros socios digitales, desarrollando la arquitectura de datos y las aplicaciones que permitirán que nuestros productos sean más inteligentes, ayudando a nuestros clientes a mejorar su productividad y seguridad”.



“Hoy en día, tenemos 400.000 activos conectados y creciendo. Para este verano, cada una de nuestras máquinas saldrá de la línea pudiendo conectarse, proporcionando algún tipo de retroalimentación en productividad operativa al propietario, al distribuidor y a nosotros, queremos llegar a mostrar al cliente en su iPhone todo lo que pasa con su máquina, su flota, su salud, su productividad y así sucesivamente”.

Dough Oberhelman, Caterpillar CEO. 25 de Abril 2016





ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

EDUCACIÓN  
PROFESIONAL

# Diplomado en Big Data y Ciencia de Datos

## *Curso: Ciencia de Datos y sus Aplicaciones*

### Clase 01: Introducción y Motivación

Jaime Caiceo, Roberto González  
y Roberto Muñoz

✉ [jcaiceo@uc.cl](mailto:jcaiceo@uc.cl)

✉ [regonzar@uc.cl](mailto:regonzar@uc.cl)

✉ [rmunoz@uc.cl](mailto:rmunoz@uc.cl)

El uso de apuntes de clases estará reservado para finalidades académicas. La reproducción total o parcial de los mismos por cualquier medio, así como su difusión y distribución a terceras personas no está permitida, salvo con autorización del autor.