



UNIVERSIDAD
PANAMERICANA

Introducción a la Ciencia de Datos

Dr. Leon Felipe Palafox Novack
lpalafox@up.edu.mx

0

Noticias del día

Qué ha pasado en el mundo de Data
Science?

RUMBLE 2017

Inside the world's biggest and most important AI conference

By [Dave Gershgorn](#) • December 12, 2017





1

Anuncios parroquiales

Proyecto Final



Proyecto Final

El objetivo es que trabajen en el proyecto a lo largo del curso.

Conforme vayamos aprendiendo las herramientas, se recomienda las practiquen con sus propios datos.

Vayan formando sus equipos oportunamente.

- Los tutoriales de Python van a estar disponibles toda la clase
- El machote para el reporte final ya esta en la pagina web.
- La fecha de entrega es la última clase del curso, 18 de Septiembre

2

Anuncios parroquiales

Examen

- Va a salir del material de los slides.
 - ▶ No pregunto nada que no venga en los slides
- Es la segunda mitad de la siguiente clase.

3

Machine Learning

Haciendo que las computadoras
piensen

“ A breakthrough in machine learning would be worth ten Microsofts.

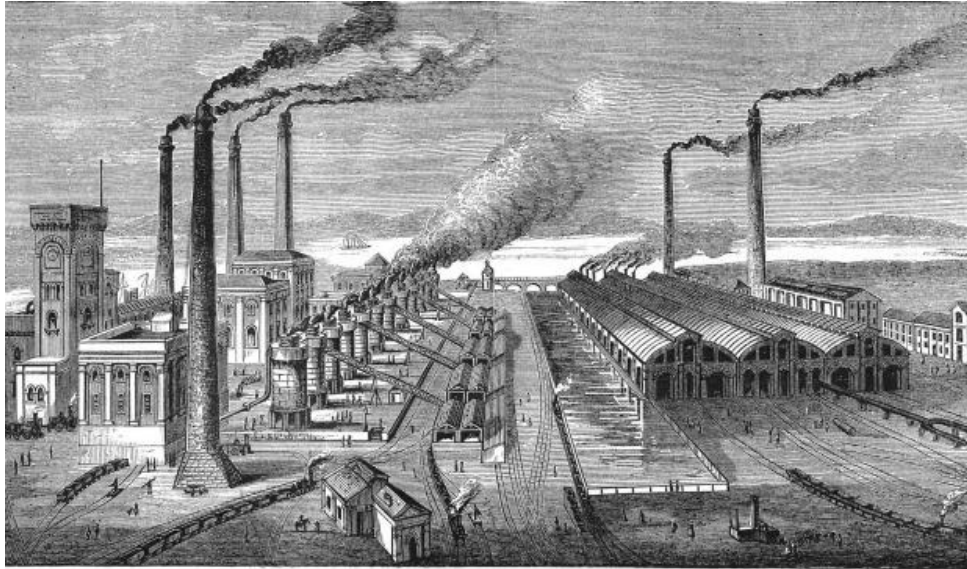
- Bill Gates

- Ciclo de vida del Dato

- Gobierno de Datos

- Primera Revolución Industrial
 - ▷ 1760:1820, Máquina de vapor, fábricas
- Segunda Revolución Industrial
 - ▷ 1900, Línea de producción, electricidad
- Tercera Revolución Industrial
 - ▷ 1980: 2000, Internet, Computación

■ Primera revolución industrial



Segunda revolución industrial



CONSUMERism



■ Tercera revolución industrial



■ ¿Y la cuarta revolución industrial?

Simplifying The Fourth Industrial Revolution: Data Science, IoE & AI



Neeraj Sanan CommunityVoice
Forbes Communications Council ⓘ

Mar 28, 2017, 09:00am • 2,512 views • #GettingBuzz

POST WRITTEN BY

Neeraj Sanan

Chief Marketing Officer at [Spire](#) - Contextual Search & Artificial Intelligence.

The Fourth Industrial Revolution Moves From Automated To Autonomous



Or Shani CommunityVoice
Forbes Technology Council ⓘ

Jun 4, 2018, 08:30am • 1,471 views • #NewTech

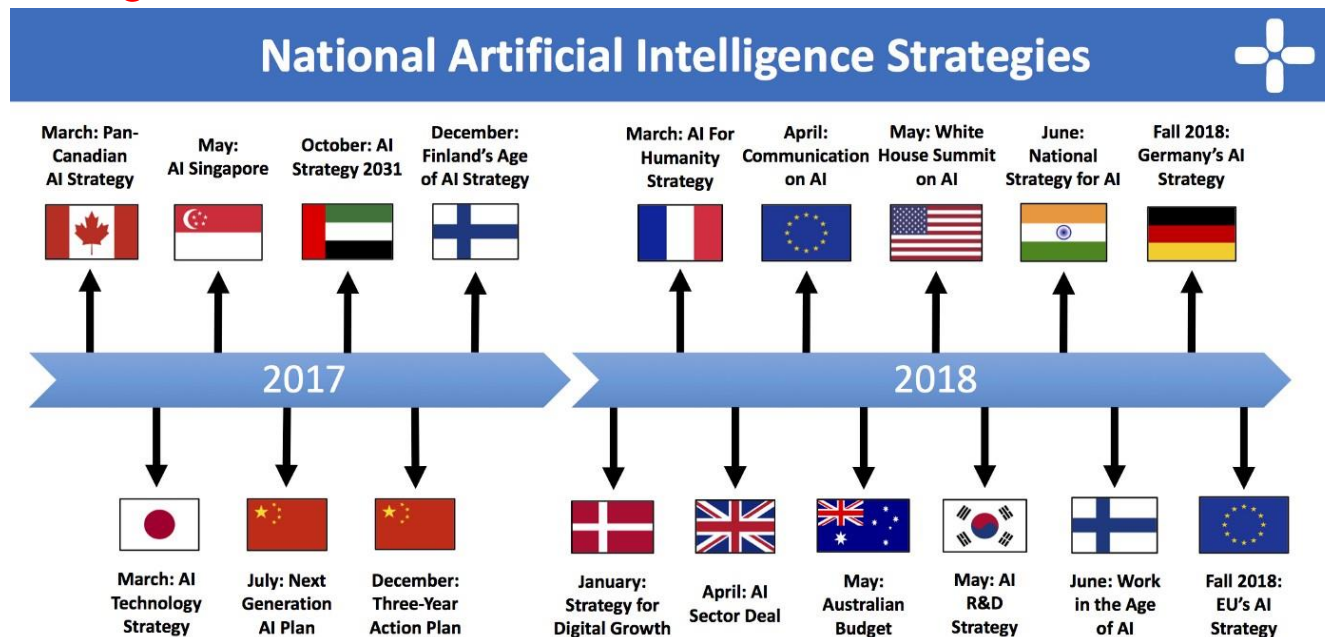
POST WRITTEN BY

Or Shani

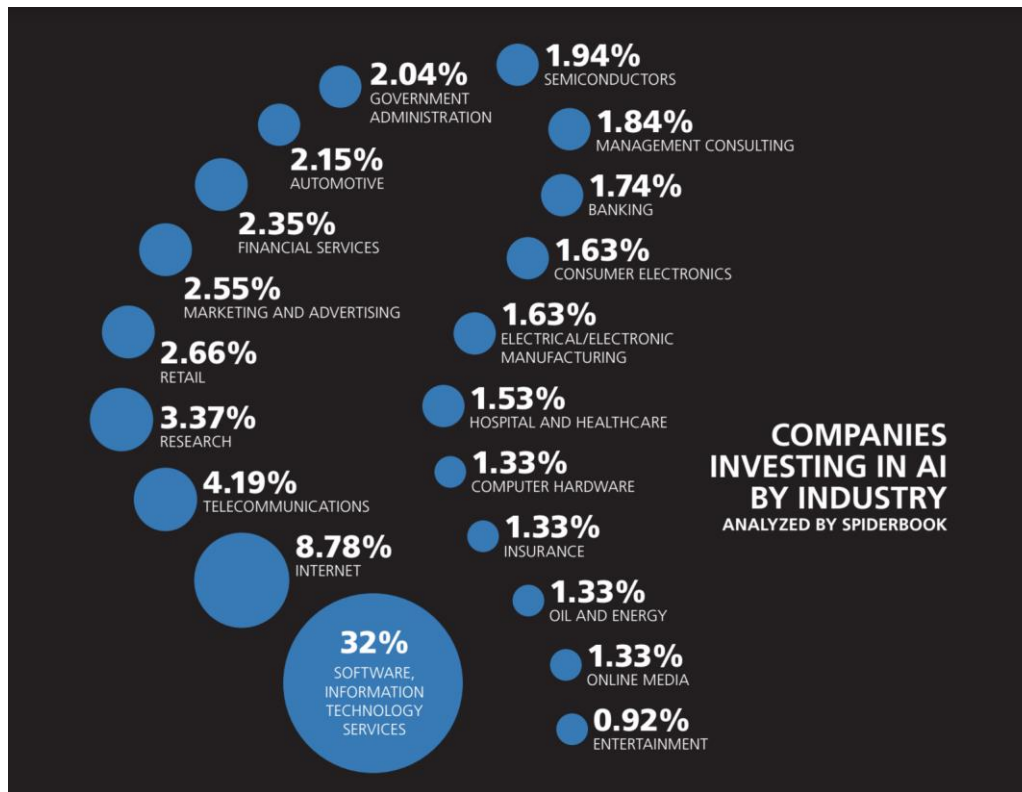
Or Shani is CEO and founder of [Albert Technologies](#), maker of Albert, the world's first and only autonomous digital marketer.



Machine Learning



2018-06-28 | Politics + AI | Tim Dutton



Que es Machine Learning

- ▷ Es el nombre comercial de una rama de Inteligencia Artificial.
- ▷ En 1984 aconteció el primer invierno de la Inteligencia Artificial.
 - ▷ Se prometieron muchas cosas que no se entregaron
 - ▷ Faltaban dos cosas:
 - ▷ Poder de procesamiento
 - ▷ Datos!
- ▷ Se perdió dinero y financiamiento por parte de las instituciones.



Que es Machine Learning

■ En los 2000, todo mundo contrataba desarrolladores de bases de datos.

▷ La aparición del developer Full Stack

▷ Una acumulación masiva de datos

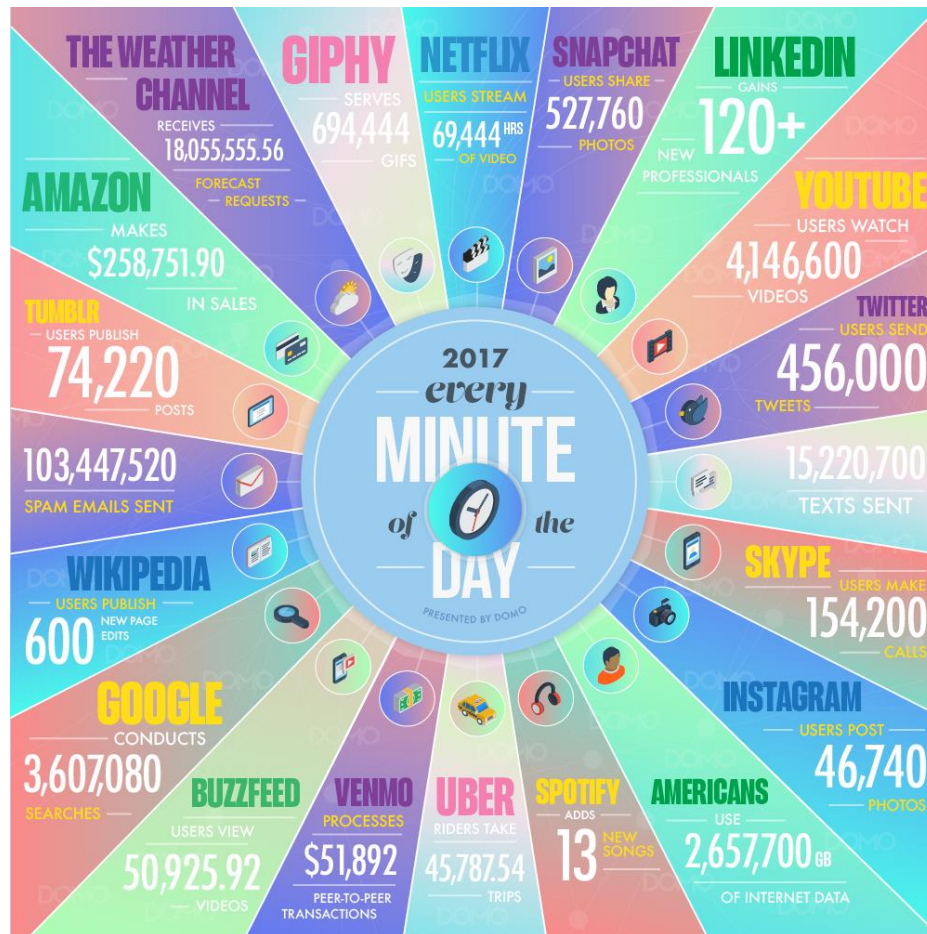
■ La aparición de las redes sociales masivas.

▷ Facebook tiene alrededor de 2 mil millones de usuarios

▷ Se generan datos iguales a alrededor de 500, 000, 000 millones de canciones cada día.

▷ Quien los va categorizar?

■ De pronto a hay datos de sobra, y el procesamiento también es sustancialmente mayor al que se tenía en 1984.



1er mensaje



- ▷ Mafia del poder
- ▷ Complot
- ▷ Voto por voto
- ▷ Fraude
- ▷ México
- ▷ Presidente

2ndo mensaje



- ▷ Tirititito
- ▷ Fútbol
- ▷ Estadio
- ▷ América
- ▷ México

Cómo supieron?



- ▶ Hay palabras asociadas a cada persona
- ▶ Su cerebro hace una correlación entre las palabras y las personas.
- ▶ Su cerebro calcula la probabilidad conjunta de las personas y las palabras.

Entrenamiento Supervisado



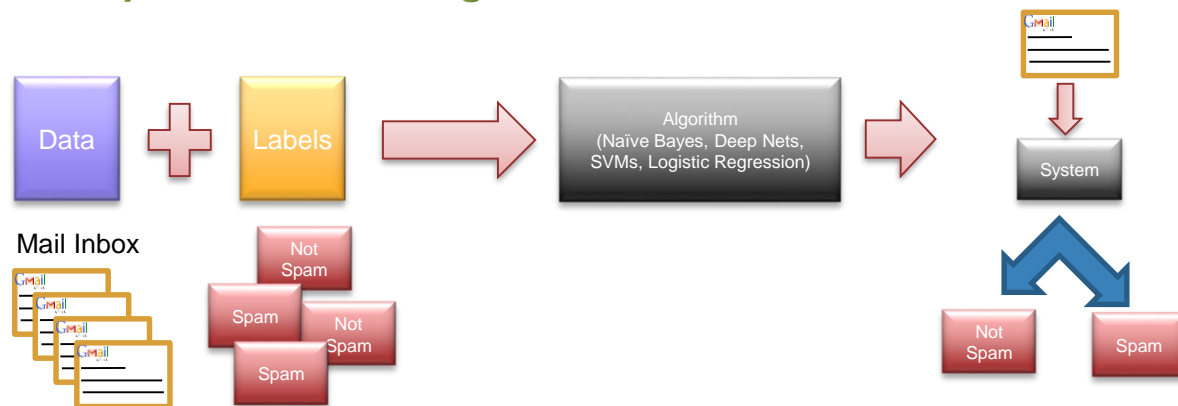
Set de datos **etiquetados**

- ▷ Set de mail etiquetados como spam/no spam
- ▷ Reviews de Amazon
- ▷ Status de Facebook
- ▷ Volumen de compra/venta de acciones

Algoritmo

- ▷ Regresión logística
- ▷ Regresión Linear
- ▷ Maquinas de soporte vectorial
- ▷ Deep Learning (Redes Neuronales)

Supervised Learning



Cada categoría tiene palabras claves que las caracterizan (feature)

Spam: Offer, Viagra, medicine, Free, Conference in China

Not Spam: Universidad, Artículo, Mama, Papa, Mia

■ Ajuste de los algoritmos

- ▷ *"The needs of the Many outweigh the needs of the few"*
 - ▷ *Spock*
- ▷ No quieres decirle a alguien que tiene Cancer si en realidad no tiene, pero **NO** quieres decirle a alguien que no tiene Cancer, cuando si tiene.

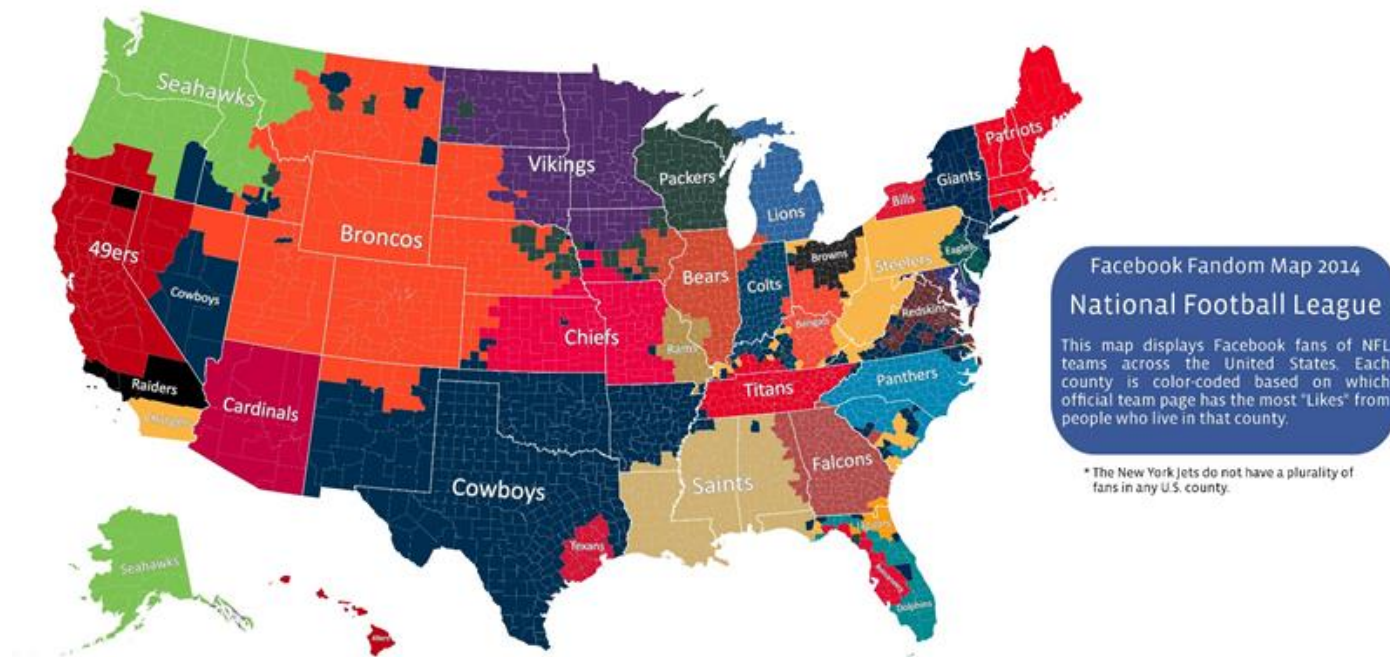
Ejemplos

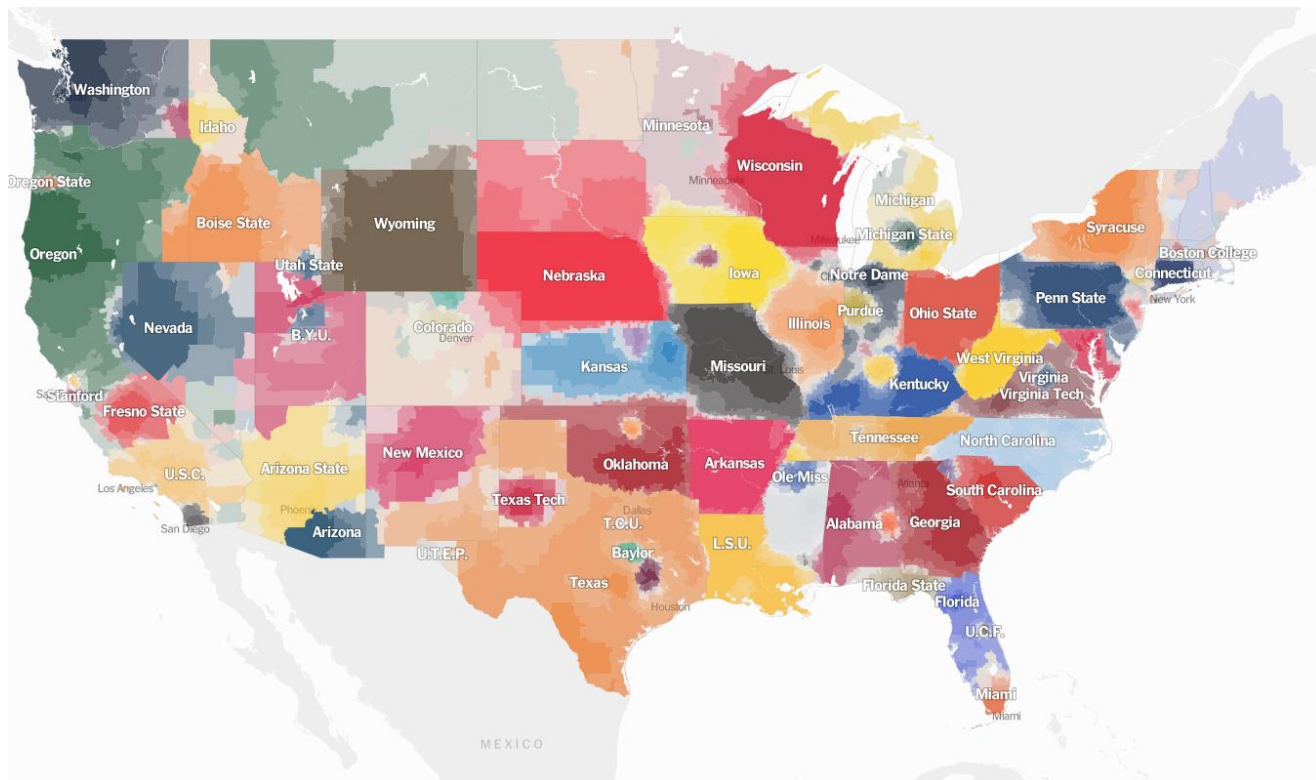


- Scoring crediticio
 - Como predecir si das o no un crédito.
- Predicción de audiencias
 - Saber a quien le vas a mostrar un anuncio

- <https://www.youtube.com/watch?v=AR3hY9iB5-l>
- <https://www.youtube.com/watch?v=aa0B-ErYq6Y&t=11s>

Otro paradigma



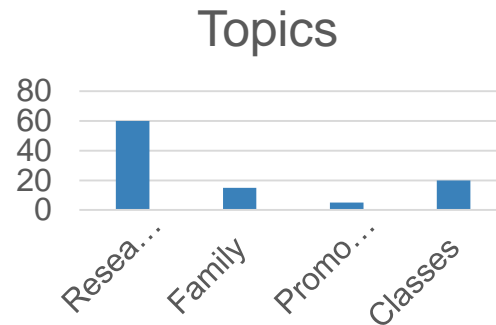
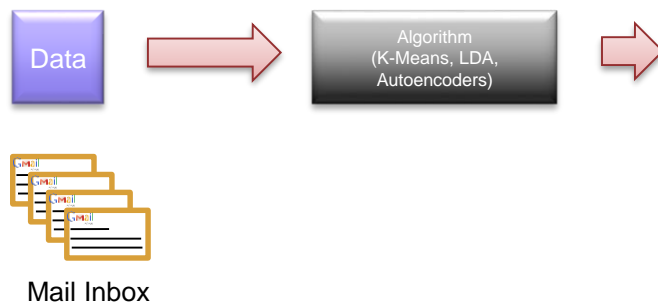


Descubrimiento de conocimiento



- No necesitamos etiquetas
- Los datos se organizan solos.
- Muchos algoritmos se encargan de organizar los datos automáticamente.

Unsupervised Learning



Igual que antes, nuestros features son palabras.

Cada topico tiene palabras que los caracterizan.

Research: Mars, Proposal, DTM, HiRISE, Machine Learning, Deep Nets, Bayesian

Family: Mom, House, Mexico

Promotions: Computer, PS4, Cheap, Amazon, Deal

Classes: Grades, Homework, Questions, Office Time

Ejemplos



- <http://everynoise.com/>
- <https://cs.stanford.edu/people/karpathy/tsnejs/>