



SIGNAL PROCESSING &  
MACHINE LEARNING FOR BIG DATA

DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES



# Opening Session

Eduardo López Gonzalo  
[eduardo.lopez@upm.es](mailto:eduardo.lopez@upm.es)

Luis Hernández Gómez  
[luisalfonso.hernandez@upm.es](mailto:luisalfonso.hernandez@upm.es)

Departamento de Señales, Sistemas y  
Radiocomunicaciones  
E.T.S. Ingenieros de Telecomunicación  
Universidad Politécnica de Madrid

# Opening Session

- Program Overview
- Internships and Master Thesis
- Students' presentations
- Predictive & Descriptive Learning and Machine Learning Lab tools
- Level Test

# Opening Session

- Program Overview
- Internships and Master Thesis
- Students' presentations
- Predictive & Descriptive Learning and Machine Learning Lab tools
- Level Test

# Big Data?

May 5, 2016 | 931 views

## Big Data At Sprint: Turning Mobile Network Data Into Business Value

Thanks to its huge network of users, Sprint has access to vast amounts of user data. Three years ago it established subsidiary Pinsight Media to investigate ways of capitalizing on that data. Since then it has gone from serving zero to six billion ad impressions per months, based on "authenticated [...]



Apr 19, 2016 | 20,645 views

## How Big Data And Analytics Are Transforming The Construction Industry



Apr 22, 2016 | 14,793 views

## How Big Data And Analytics Are Transforming Supply Chain Management

Apr 28, 2016 | 9,109 views

## Big Data Overload: Why Most Companies Can't Deal With The Data Explosion

FEB 16, 2016 @ 02:54 AM 15,945 VIEWS

## How Big Data Is Transforming Medicine



Forbes / Tech

SEP 8, 2015 @ 02:26 AM 131,676 VIEWS

## 4 Ways Big Data Will Change Every Business



# The Data Revolution: Big Data

May 5, 2016 | 931 views

## Big Data At Sprint: Turning Mobile Network Data Into Business Value

Thanks to its huge network of users, Sprint has access to vast amounts of user data. Three years ago it established subsidiary Pinsight Media to investigate ways of capitalizing on that data. Since then it has gone from serving zero to six billion ad impressions per months, based on "authenticated [...]



Apr 19, 2016 | 20,645 views

## How Big Data And Analytics Are Transforming The Construction Industry



Apr 22, 2016 | 14,793 views

## How Big Data And Analytics Are Transforming Supply Chain Management

Apr 28, 2016 | 9,109 views

## Big Data Overload: Why Most Companies Can't Deal With The Data Explosion

FEB 16, 2016 @ 02:54 AM 15,945 VIEWS

## How Big Data Is Transforming Medicine



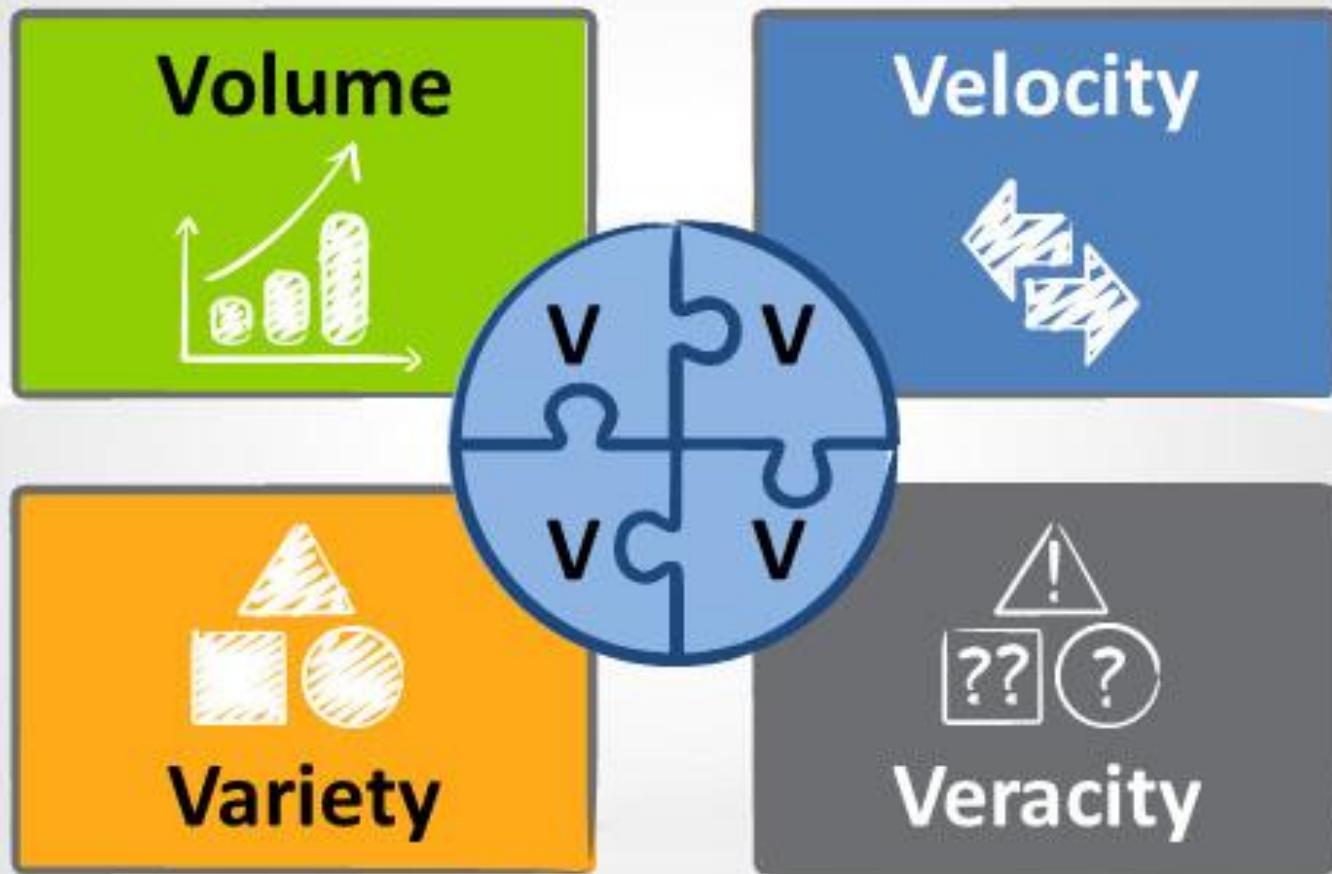
Forbes / Tech

SEP 8, 2015 @ 02:26 AM 131,676 VIEWS

## 4 Ways Big Data Will Change Every Business



## Big Data: 4V definition



# 1970s ETL (Extract, Transformation, Loading)

- **Data is extracted:** from the existing sources
- **Data is then transformed:** for its **specific** use
- **Data is loaded:** into the final target operational database or data warehouse.

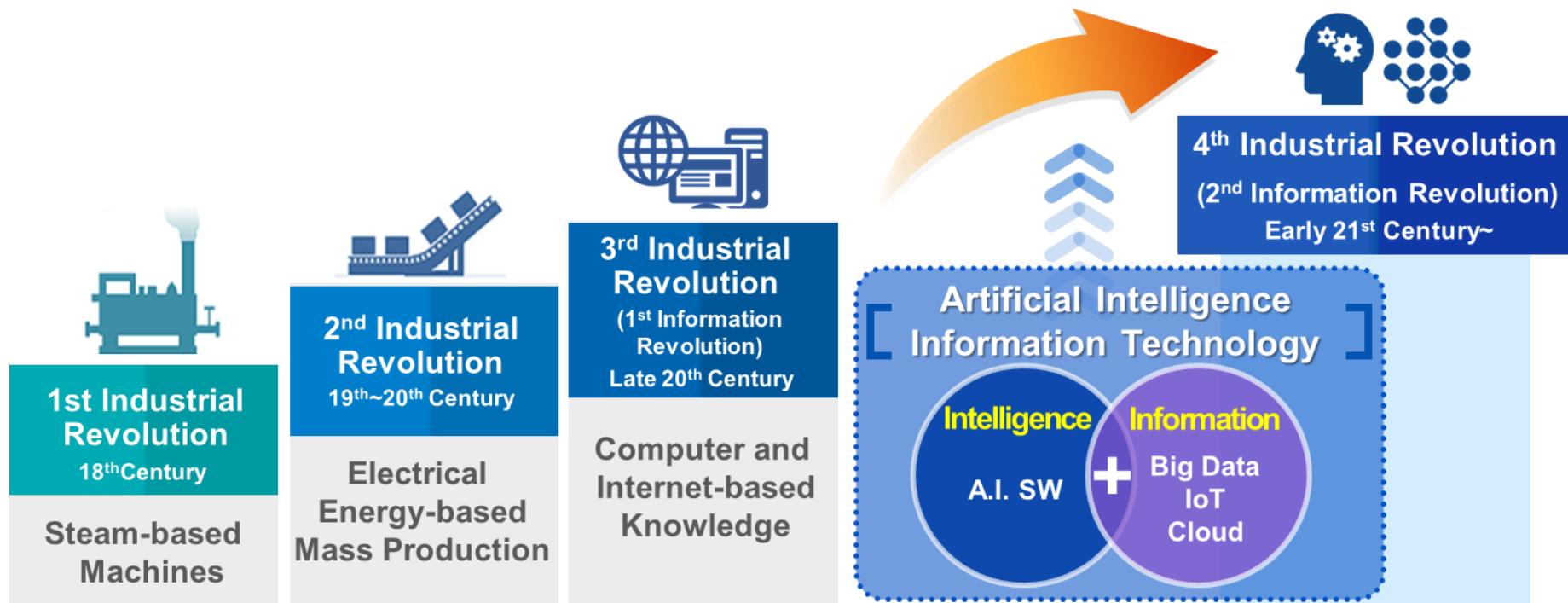


*Business Impact* —

**Attention to data management, particularly in a climate of e-commerce, and greater need for collaboration can enable enterprises to achieve greater returns on their information assets.**

*Bottom Line* —

In 2001/02, IT organizations must look beyond traditional direct brute-force physical approaches to data management. Through 2003/04, practices for resolving e-commerce accelerated data volume, velocity, and variety issues will become more formalized/diverse. Increasingly, these techniques involve tradeoffs and architectural solutions that involve/impact application portfolios and business strategy decisions.



---

Intelligent Machines

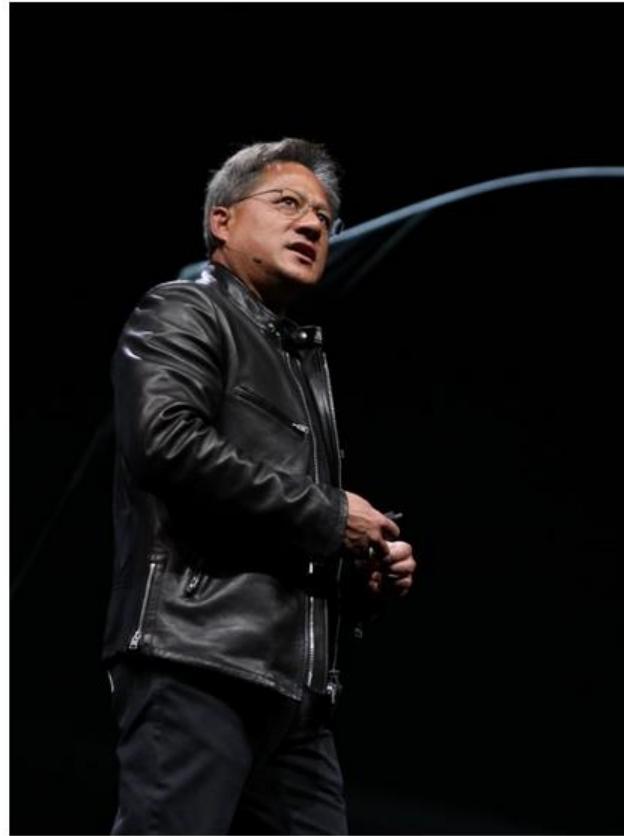
---

# Nvidia CEO: Software Is Eating the World, but AI Is Going to Eat Software

Jensen Huang predicts that health care and autos are going to be transformed by artificial intelligence.

by Tom Simonite May 12, 2017

Nvidia CEO Jensen Huang at the company's developer conference in San Jose, California.



# Amazon

Jeff Bezos' letter to Amazon shareholders - May, 2017

Newsletter signup

Cloud Tech

## Jeff Bezos explains Amazon's artificial intelligence and machine learning strategy

BY TODD BISHOP on May 6, 2017 at 9:46 am

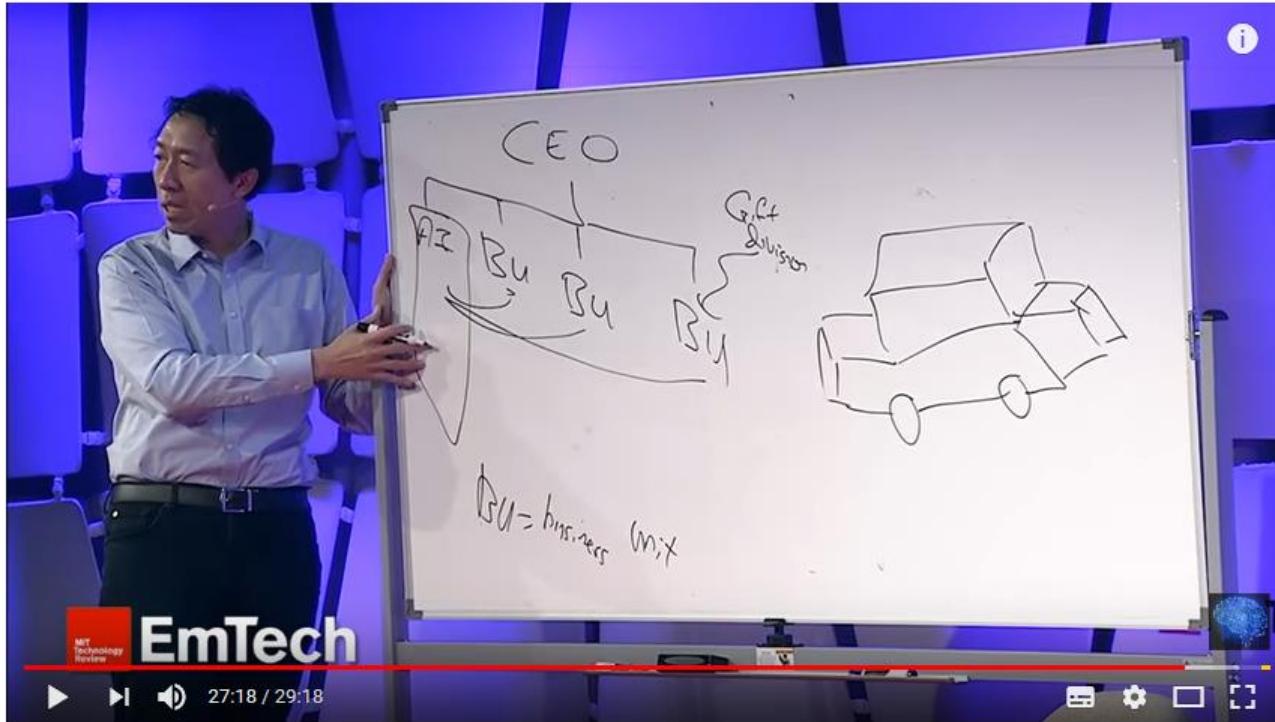
[Post a Comment](#) [f Share 567](#) [Tweet](#) [Share 1.6K](#) [Reddit](#) [Email](#)



Amazon CEO Jeff Bezos appeared this week at the Internet Association's annual gala

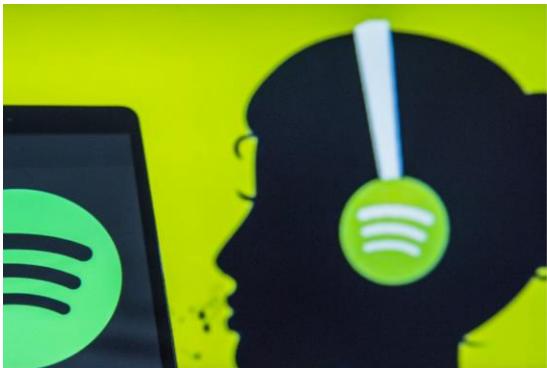
*"Machine learning and AI is a horizontal enabling layer. It will empower and improve every business, every government organization, every philanthropy — basically there's no institution in the world that cannot be improved with machine learning".*

*Jeff Bezos*



Andrew Ng - The State of Artificial Intelligence

## MANY DIFFERENT FIELDS & APPLICATION DOMAINS

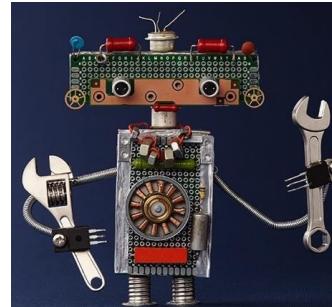


### Music Recommendation systems

- Is this song similar to that one?
- When can this song become a hit?
- Should I recommend this song?

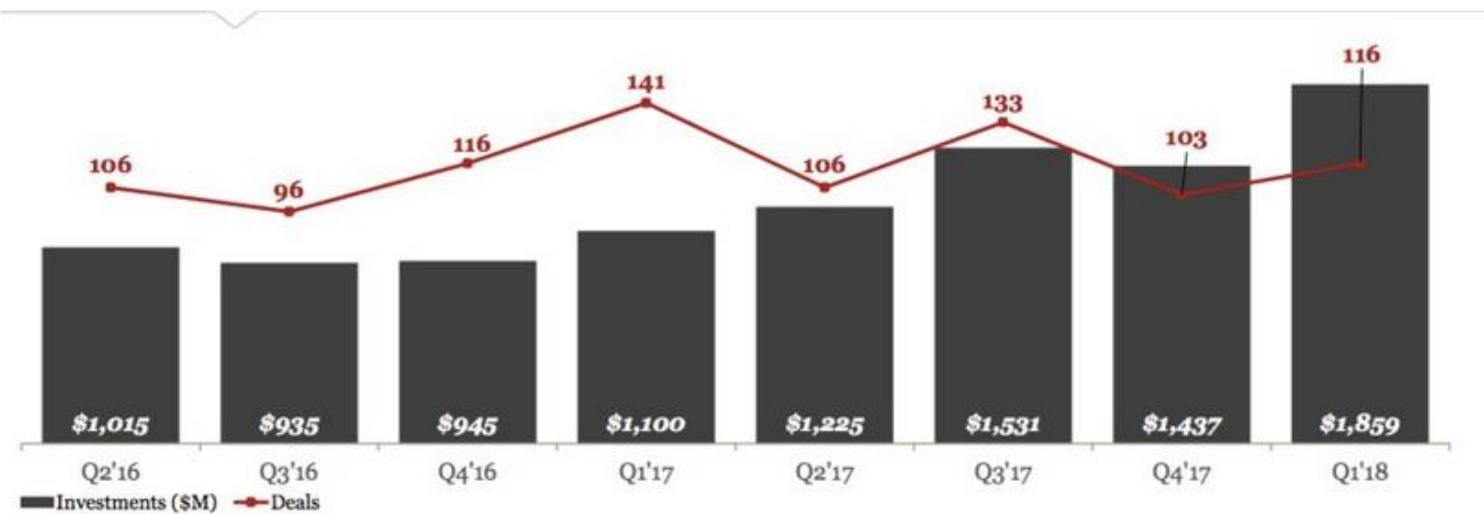
### Fixed Before It's Broken

The Art & Science of Predictive Maintenance



### Predictive Maintenance of machinery

- Is that behavior normal?
- When the machine is going to fail?
- Should I add this machine to the review list?

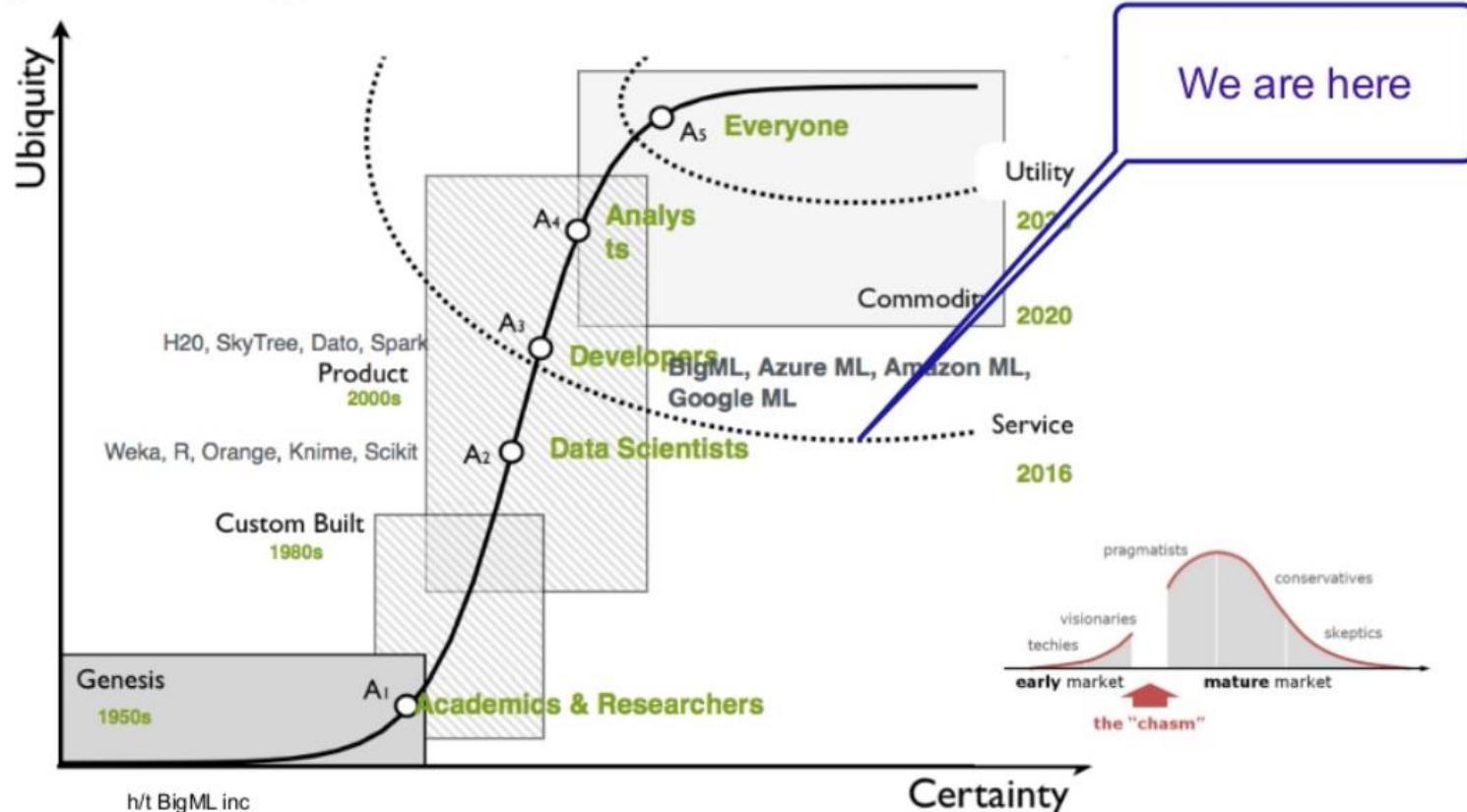


Funding for AI startups in the US hit an eight-quarter high in Q1 2018. 2018 Q1 MoneyTree Report

- US Artificial intelligent startups raked in \$1.9 billion in venture capital money in Q1 2018, up 29% from the quarter before.

# Machine Learning Adoption

Applying machine learning at scale





24  
HOURS

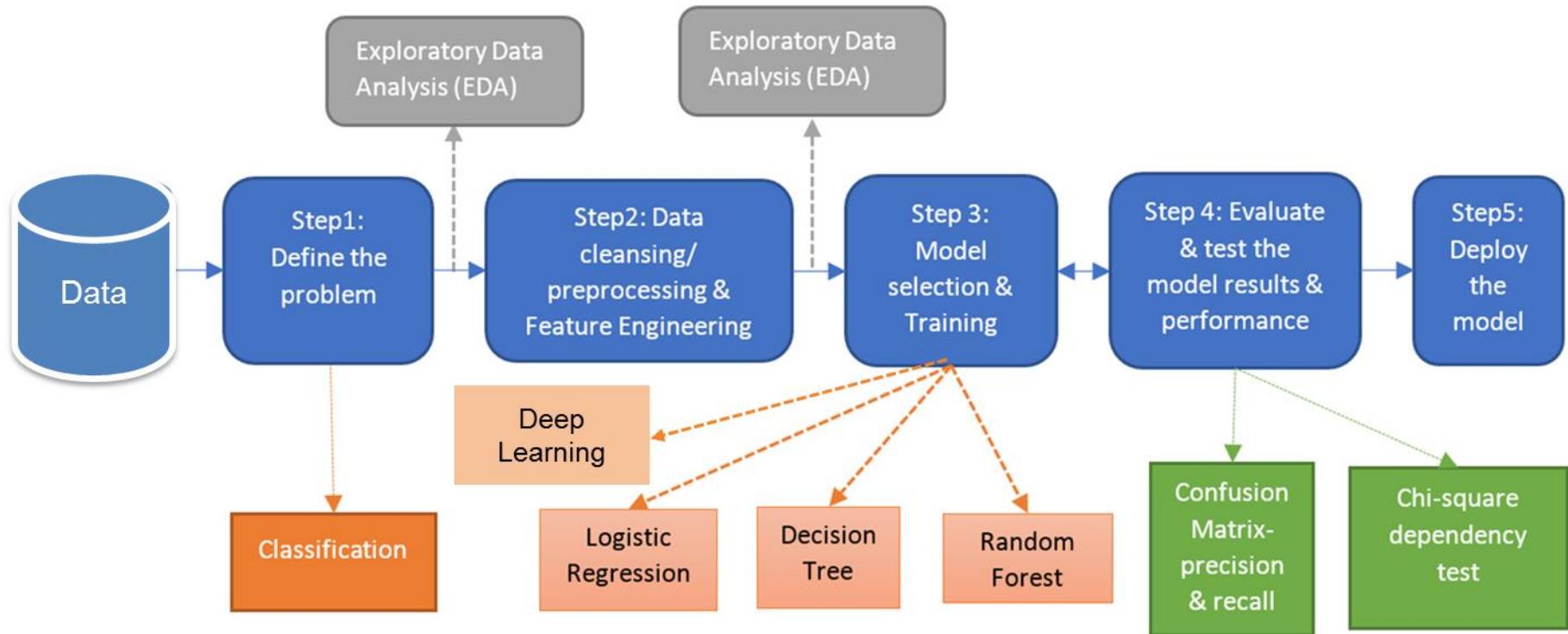
7  
DAYS A WEEK

365  
DAYS A YEAR



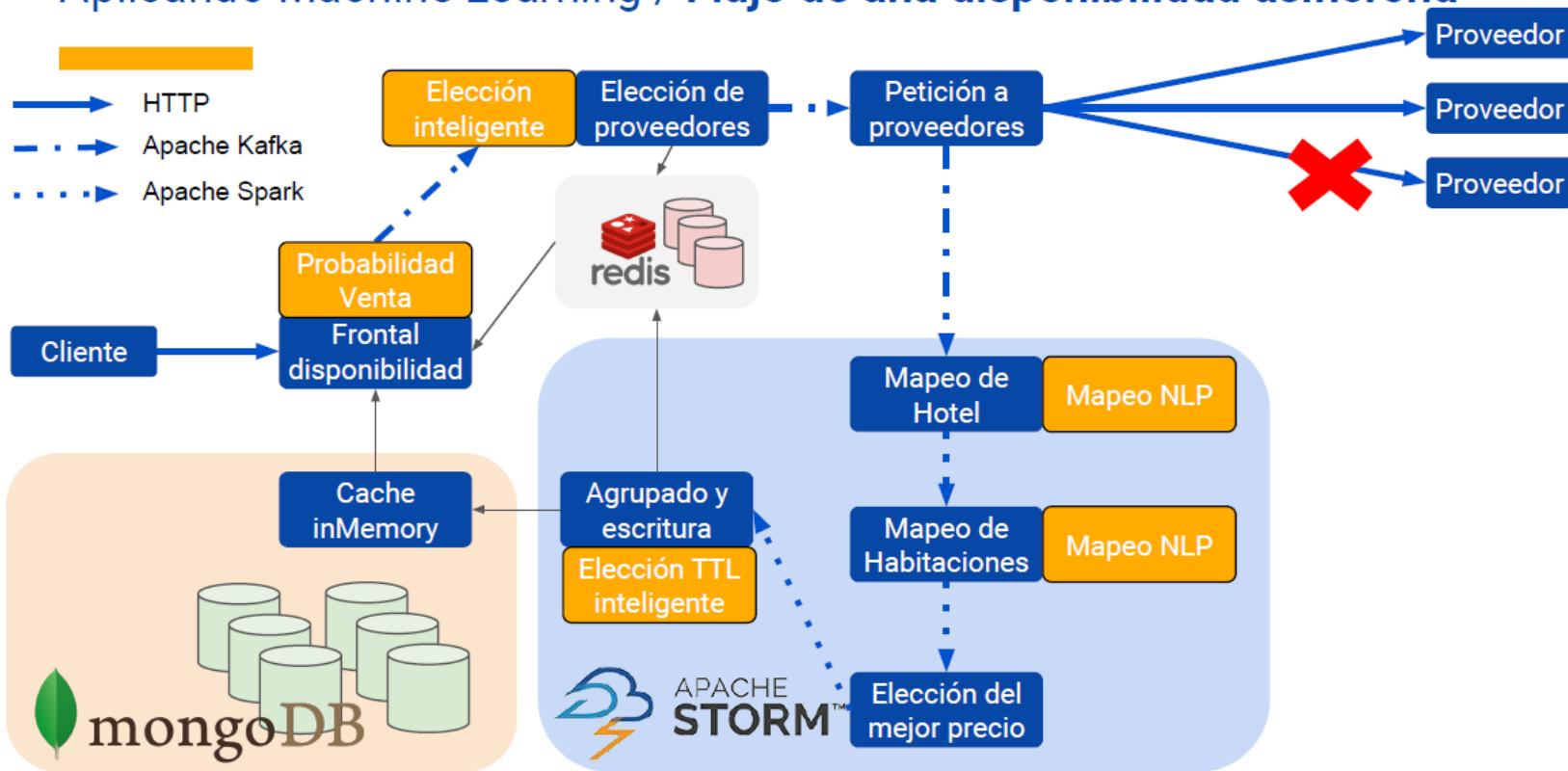
@sukiweb







## Aplicando Machine Learning / Flujo de una disponibilidad asincrónica



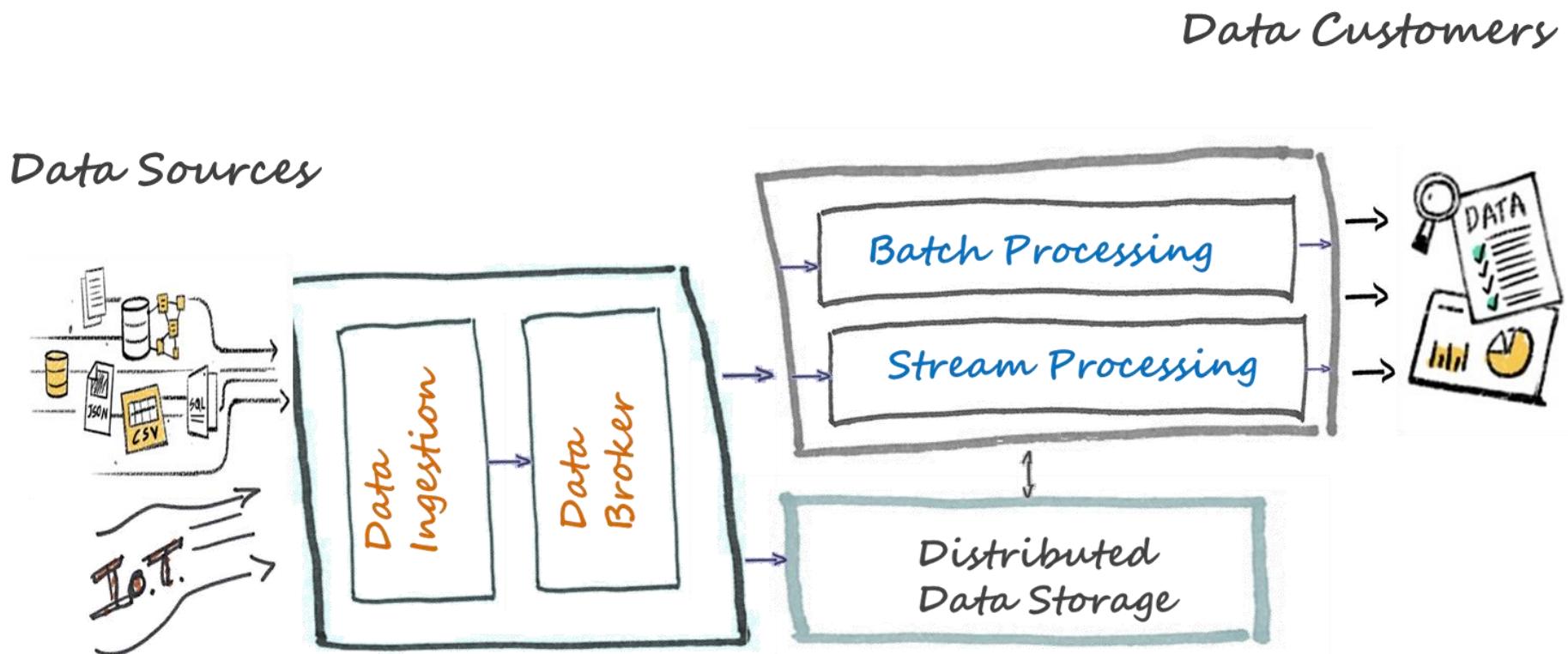
Logitravel Group



24 horas  
7 días  
365 días a la semana



# Big Data Architecture

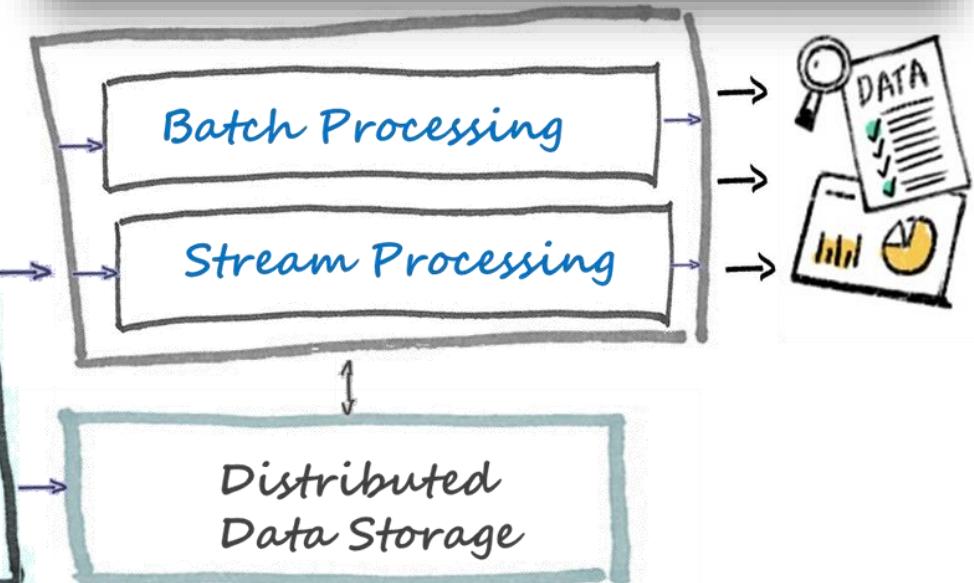
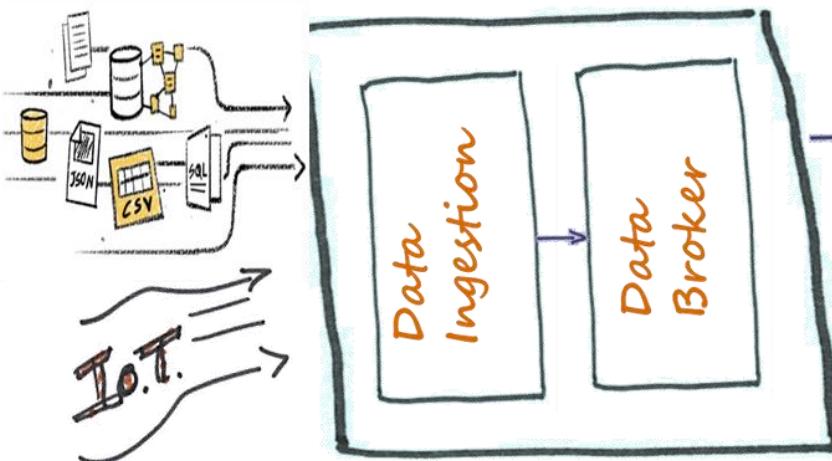


# Big Data Architecture

Machine Learning

Artificial Intelligence

Data Sources



BIG DATA & AI LANDSCAPE 2018



Final 2018 version, updated 07/15/2018

© Matt Turck (@mattturck), Demi Obavomi (@demi\_obavomi), & FirstMark (@firstmarkcap)

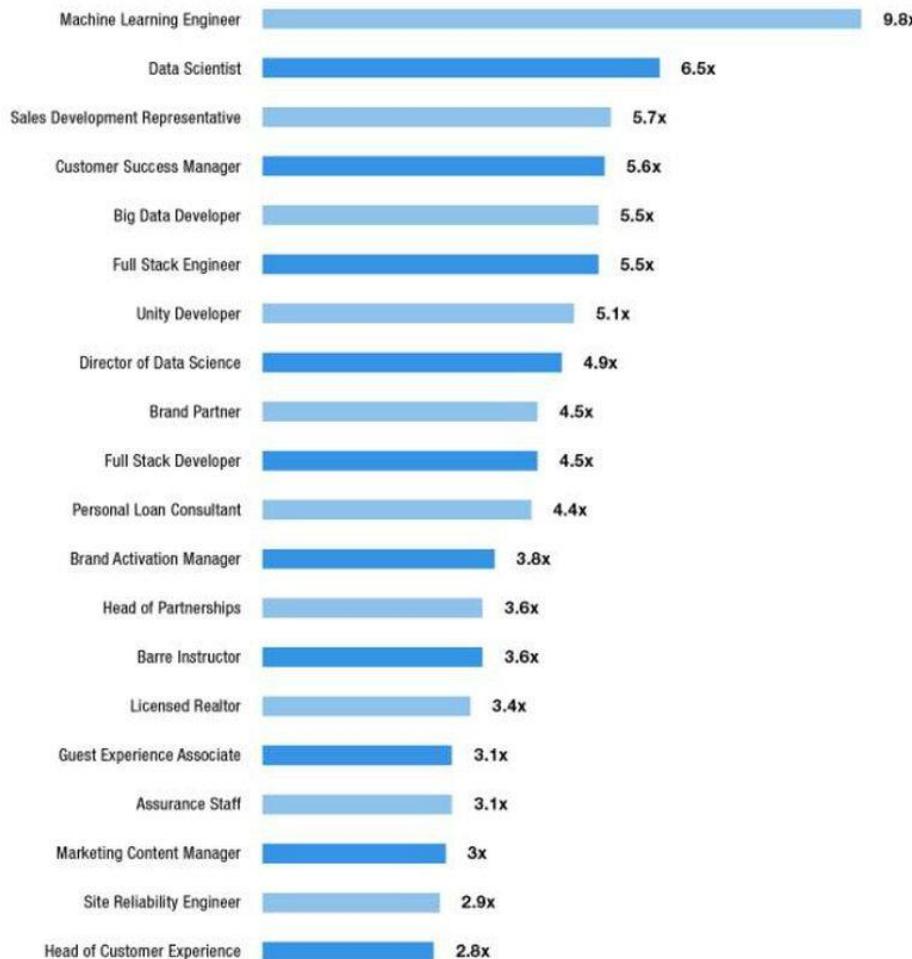
mattturck.com/bigdata2018

**FIRSTMARK**   
EARLY STAGE VENTURE CAPITAL

# LinkedIn's Fastest-Growing Jobs Today Are In Data Science And Machine Learning

Top 20 Emerging Jobs

LinkedIn Economic Graph



## Data Scientist: a search for unicorns

Real job post  
on Internet

The role requires deep knowledge in designing, planning, testing and deploying analytics solutions including the following:

- Proven experience implementing and deploying advanced analytics (Machine Learning) solutions for large enterprise or commercial customers in complex heterogeneous environments
- Expertise in Microsoft AzureML and/or Microsoft R Server
- Expertise in open source and/or competitive platforms
- Working knowledge of specialized tools and solution providers whether they be Microsoft partners or competitors
- Knowledge of full application life cycle design tools and methodologies
- Strong demonstrated consulting skills
- Experience with leading services delivery teams and delivery management
- Experience driving presales opportunities, leading customer workshops/briefings

The Data Scientist must have deep technical skills in the following:

- Data Science degree with enterprise experience in Machine Learning / Statistics
- At least one of the ML related technologies ( SAS, SPSS, RevR, Azure ML ).
- Data prep with industry standard ETL tools, R, Python, etc. and with Visualization tools (Tableau, QlikView, PowerBI, etc )

Additionally, experience in any of the following will be considered favorably:

- Big Data stack: MapReduce, Hadoop, Sqoop, Pig, Hive, Hbase, Flume
- ETL technologies: Talend, Pentaho, Informatica, etc
- DW: Relational (Oracle, Teradata, Netteza, SAP HANA), columnar (RedShift, Vertica), noSQL (MongoDB, Redis, Cassandra - key-value stores, graph databases, RDF triple stores)
- Languages: Java, Python, R, F#, C#
- Analytical tools, languages, or libraries: SAS, SPSS, R, Mahout)

# DATA Engineer

Develops, constructs, tests,  
and maintains architectures.  
Such as databases  
and large-scale  
processing systems.



DataCamp  
Learn Data Science By Doing

# DATA Scientist

Cleans, massages  
and organizes (big) data.  
Performs descriptive statistics  
and analysis to develop  
insights, build models and  
solve a business need.



# MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

## MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



## DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

## PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing package e.g. R
- ★ Databases SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS

## COMMUNICATION & VISUALIZATION

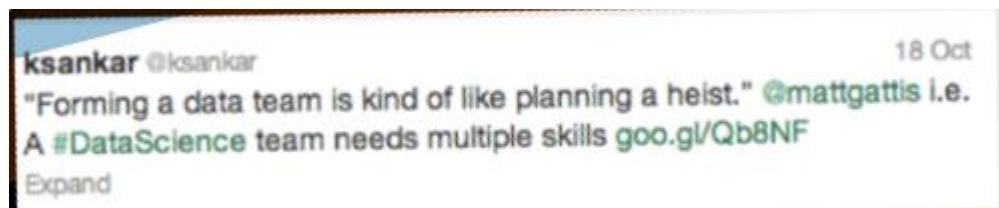
- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

# Data Science Workflow



# What is Data Science?

- **The curious case of the Data Scientist**



*Data Scientist (noun): Person who is worse at statistics than any statistician & worse at software engineering than any software engineer*

*– Will Cukierski (Kaggle)*

*Data Scientist (noun): Person who is better at statistics than any software engineer & better at software engineering than any statistician*

*– Josh Wills (Cloudera)*



**... but there are some  
“buts”!!!**



... there are “some” important issues to consider!

and not “only” those related to data privacy,



# WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY  
AND THREATENS DEMOCRACY

CATHY O'NEIL

# So focus your future big professional potential in improving our world!!!



HOME FIND DATA PREPARE DATA BUILD ANALYTICS RESOURCES STATUS LOGIN SIGN UP



We often hear that the world is drowning in data. Yet, when you want to learn and experiment, finding open, easily-accessible data sets remains a major challenge. Open data is data that can be freely used, intermixed with other datasets, and distributed to others without restrictions. We have compiled a small selection of publicly accessible data sources that should give an aspiring data scientist a quick start. We will be constantly adding new data sources to this page:



Artificial Intelligence

# Artificial Intelligence Is Our Future. But Will It Save Or Destroy Humanity?

by Patrick Caughill on September 29, 2017

37733



Why AI could destroy more jobs than it creates, and how to save them

By Nick Heath

Kumaran Akilan Software Engineer at  
Archie.ai  
Sep 2, 2017

# **Artificially Intelligent Engineers—How AI Will Kill All Engineering Jobs. *-And Why It Is a Good Thing.***

# Our Courses

<http://mstc.ssr.upm.es/>

Fundamentals

Signal Processing

Machine Learning

Applications and Practice

## SIGNAL PROCESSING AND MACHINE LEARNING FOR BIG DATA

SEMESTER 1			SEMESTER 2		
Statistical Modelling (3C)	Time Series Analysis (4.5C)	Optimization Fundamentals (3C)	Signal Processing for Big Data (4C)	Big Data for Image and Video Signals (4C)	Bio-inspired learning (3C)
Optimization techniques for big data analysis (3C)	Predictive and Descriptive Learning (6C)	Machine Learning Lab (4.5C)	Reinforcement learning (3C)	Application Projects (4C)	Large-scale Media Analytics (4C)
Data Science Foundations and Applications (2C)				Masters' Thesis (12C)	

**Review Schedule!!!**

# Our Courses: MUIT

<http://www.etsit.upm.es/da/estudios/master-universitario-en-ingenieria-de-telecomunicacion/plan-de-estudios/asignaturas.html>

ITINERARIO MACHINE LEARNING AND MULTIMEDIA DATA SCIENCE							
93000820	1	VISION ANALYSIS AND DEEP LEARNING(*)	VADL	6	guía		
93001071	1	DATA SCIENCE FOUNDATIONS AND APPLICATIONS (*)	DSFA	2	guía		
93001070	1	PREDICTIVE AND DESCRIPTIVE LEARNING (*)	PRDL	6	guía		
93001072	1	MACHINE LEARNING LAB (*)	MLLB	4,5	guía		
93000806	2	MEDIA DATA SYSTEM DESIGN (*)	MDSD	6	guía		
93000819	2	DEEP LEARNING FOR ACOUSTIC SIGNAL PROCESSING(*)	DLAS	6	guía		

# Opening Session

- Program Overview
- Internships and Master Thesis
- Students' presentations
- Predictive & Descriptive Learning and Machine Learning Lab tools
- Level Test

# Opening Session

## ● Internships and Master Thesis

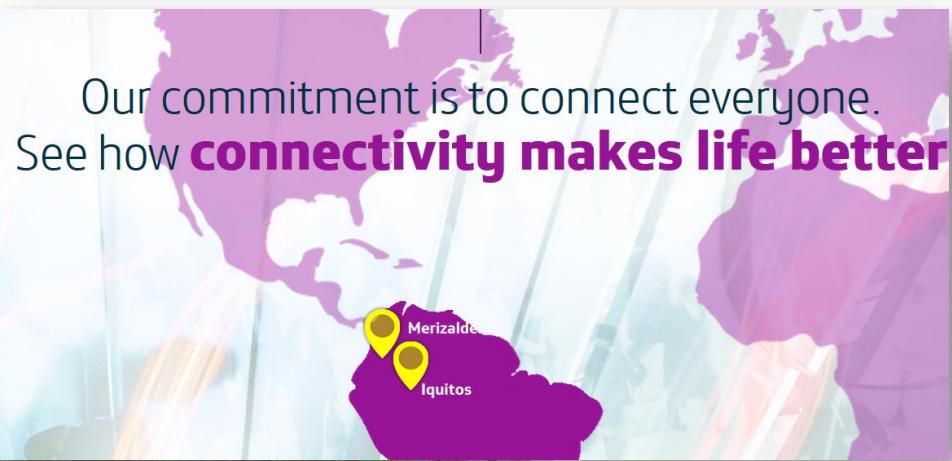
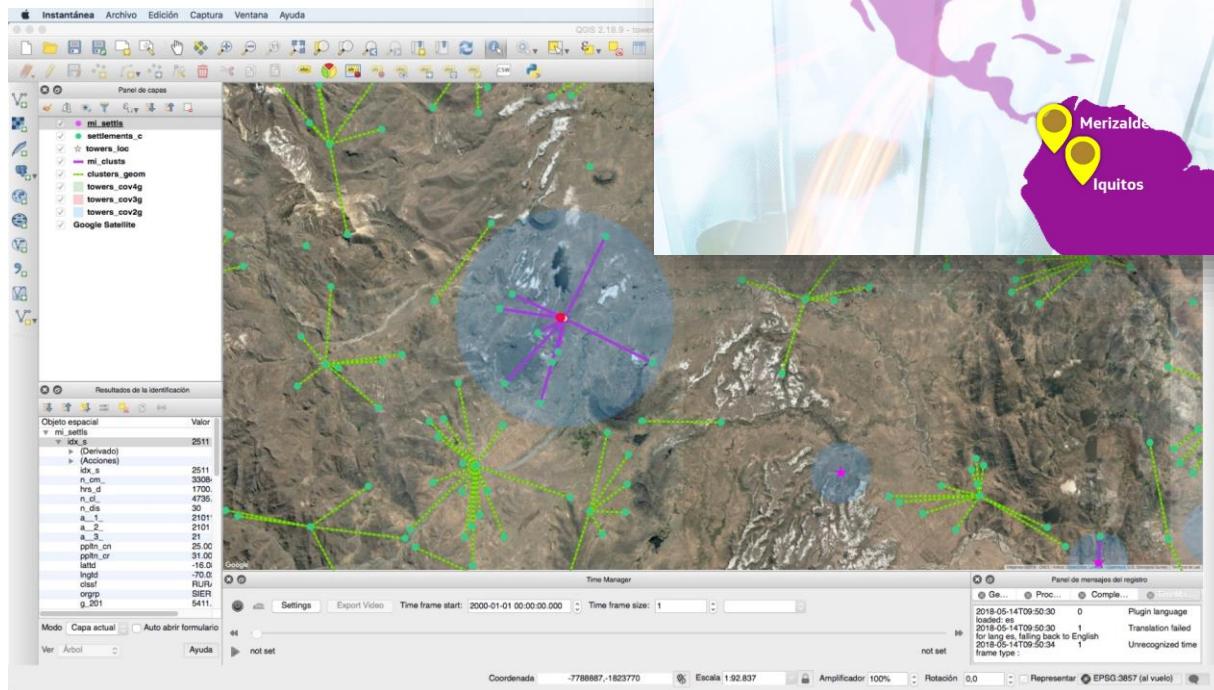
We recommend you:

- To be involved in internships with companies or research groups
- To start working on your Master Thesis

... as soon as possible!

# Internet for All

Telefonica



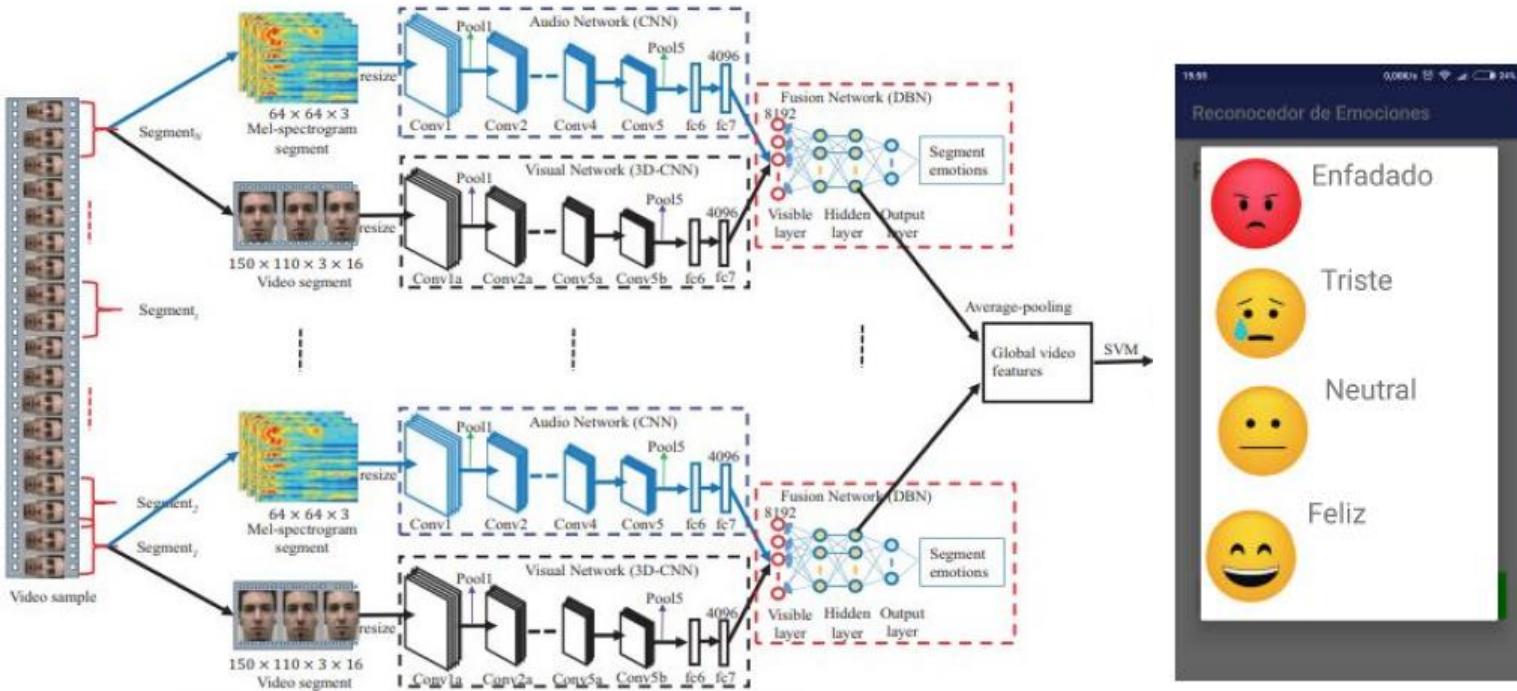
<https://www.telefonica.com/ext/public-policy/connectivity/>

# Road Accident Prediction and Characterization Convolutional Neural Networks



# Multimodal Emotion Recognition using Deep Learning

ETSIT – Research Group



# Opening Session

- Program Overview
- Internships and Master Thesis
- **Students' presentations**
- Predictive & Descriptive Learning  
and Machine Learning Lab tools
- Level Test

# Opening Session

.... your turn!

# Opening Session

- Program Overview
- Internships and Master Thesis
- Students' presentations
- Predictive & Descriptive Learning and Machine Learning Lab tools
- Level Test

# Opening Session

.... also for:

**Data Science Foundations and Applications**

(starting next Monday, September 10<sup>th</sup>)

# Opening Session

**Don't forget to bring your laptops to class**

- **Data Science Foundations and Applications**
- **Predictive & Descriptive Learning**
- **Machine Learning Lab**

# Opening Session



[https://twitter.com/mstc\\_ss/status/938720684192272384?ref\\_src=twsrc%5Etfw%7Ctwcamp%5Eembeddedtimeline%7Ctwterm%5Eprofile%3Amstc\\_ss%7Ctwcon%5Etimelinechrome&ref\\_url=http%3A%2F%2Fwww.mstc.ss.upm.es%2F](https://twitter.com/mstc_ss/status/938720684192272384?ref_src=twsrc%5Etfw%7Ctwcamp%5Eembeddedtimeline%7Ctwterm%5Eprofile%3Amstc_ss%7Ctwcon%5Etimelinechrome&ref_url=http%3A%2F%2Fwww.mstc.ss.upm.es%2F)

# Opening Session



**For Python we recommend ANACONDA distribution**

<https://www.anaconda.com/download/>



[What is Anaconda?](#) [Products](#) [Support](#) [R](#)

## Download Anaconda Distribution

Version 5.2 | Release Date: May 30, 2018

Download For:

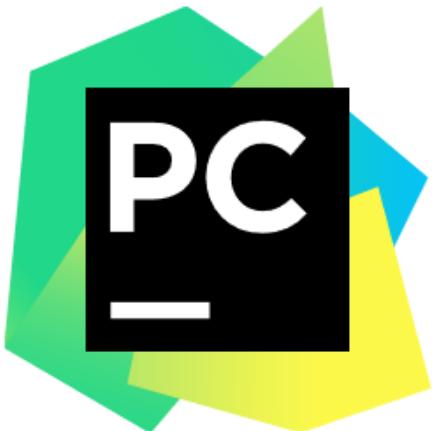
Python 3.6 version \*

Download

[64-Bit Graphical Installer \(631 MB\)](#)

[32-Bit Graphical Installer \(506 MB\)](#)

# Opening Session



Version: 2018.2.2

Build: 182.4129.34

Released: August 22, 2018

[System requirements](#)

[Installation Instructions](#)

[Previous versions](#)

## Download PyCharm

[Windows](#)

[macOS](#)

[Linux](#)

### Professional

Full-featured IDE  
for Python & Web  
development

[DOWNLOAD](#)

Free trial

### Community

Lightweight IDE  
for Python & Scientific  
development

[DOWNLOAD](#)

Free, open-source

# Opening Session



R Studio®

Products Resources Pricing About Us Blogs

## Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)



RStudio Server Pro +  
RStudio Connect  
Commercial License  
  
\$29,995 per  
year

RStudio Desktop Open Source License	RStudio Desktop Commercial License	RStudio Server Open Source License	RStudio Server Pro Commercial License
FREE	\$995 per year	FREE	\$9,995 per year
<a href="#">DOWNLOAD</a> <a href="#">Learn More</a>	<a href="#">BUY</a> <a href="#">Learn More</a>	<a href="#">DOWNLOAD</a> <a href="#">Learn More</a>	<a href="#">DOWNLOAD</a> <a href="#">Learn More</a>
			<a href="#">TALK</a> <a href="#">Try RStudio Server Pro</a>

# Opening Session

**Don't forget to bring you laptops with R Studio next Monday (September, 10<sup>th</sup>)**

**Data Science Foundations and Applications**

# Opening Session

- Program Overview
- Internships and Master Thesis
- Students' presentations
- Predictive & Descriptive Learning and Machine Learning Lab tools
- Level Test

# Opening Session

- Level Test

*Don't be afraid, it is just to know  
each other better!*

*(...on the academic side)*



SIGNAL PROCESSING &  
MACHINE LEARNING FOR BIG DATA

DEPARTAMENTO DE SEÑALES, SISTEMAS Y RADIOCOMUNICACIONES



## MSTC Overview

Eduardo López Gonzalo  
[eduardo.lopez@upm.es](mailto:eduardo.lopez@upm.es)

Luis Hernández Gómez  
[luisalfonso.hernandez@upm.es](mailto:luisalfonso.hernandez@upm.es)

Departamento de Señales, Sistemas y  
Radiocomunicaciones  
E.T.S. Ingenieros de Telecomunicación  
Universidad Politécnica de Madrid