



# BIG DATA, CONSUMIDOR ALGORÍTMICO Y PERFILACIÓN WEB”

MARIA LORENA FLÓREZ ROJAS

PROFESORA FACULTAD DE DERECHO UNIVERSIDAD DE LOS ANDES

GRUPO DE ESTUDIOS GECTI

OBSERVATORIO DE PROTECCIÓN DE DATOS



Universidad de  
**los Andes**

Observatorio **CIRO ANGARITA BARÓN**  
sobre la protección de datos personales en Colombia



**Grupo de  
Estudios en internet  
Comercio electrónico  
Telecomunicaciones e  
Informática**

# GECTI Y OBSERVATORIO DE PROTECCIÓN DE DATOS



# ¿CÓMO ESTÁ COLOMBIA EN ACCESO A INTERNET?

## Internet Usage and Population Statistics for South America December 31, 2017

SOUTH AMERICA	Population (2018 Est.)	% Pop. Table	Internet Usage, 31-Dec-2017	% Population (Penetration)	% Users Table	Facebook 31-Dec-2017
<a href="#">Argentina</a>	44,688,864	10.4 %	41,586,960	93.1 %	13.6 %	30,000,000
<a href="#">Bolivia</a>	11,215,674	2.6 %	7,570,580	67.5 %	2.5 %	6,100,000
<a href="#">Brazil</a>	210,867,954	49.2 %	149,057,635	70.7 %	48.7 %	139,000,000
<a href="#">Chile</a>	18,197,209	4.2 %	14,108,392	77.5 %	4.6 %	13,000,000
<a href="#">Colombia</a>	49,464,683	11.6 %	31,275,567	63.2 %	10.2 %	29,000,000
<a href="#">Ecuador</a>	16,863,425	3.9 %	13,476,687	79.9 %	4.4 %	10,000,000
<a href="#">Falkland Islands</a>	2,922	0.0 %	2,900	99.2 %	0.0 %	3,100
<a href="#">French Guiana</a>	289,763	0.1 %	120,000	41.4 %	0.0 %	110,000
<a href="#">Guyana</a>	782,225	0.2 %	395,007	50.5 %	0.1 %	360,000
<a href="#">Paraguay</a>	6,896,908	1.6 %	6,177,748	89.6 %	2.0 %	3,300,000
<a href="#">Peru</a>	32,551,815	7.6 %	22,000,000	67.6 %	7.2 %	20,000,000
<a href="#">Suriname</a>	568,301	0.1 %	340,000	59.8 %	0.1 %	310,000
<a href="#">Uruguay</a>	3,469,551	8.0 %	3,059,727	88.2 %	1.0 %	2,400,000
<a href="#">Venezuela</a>	32,381,221	7.6 %	17,178,743	53.1 %	5.6 %	13,000,000
<b>TOTAL SOUTH AMERICA</b>	<b>428,240,515</b>	<b>100.0 %</b>	<b>306,349,946</b>	<b>71.5 %</b>	<b>100.0 %</b>	<b>266,583,100</b>

# MI PERFIL EN LÍNEA





Temas relacionados

Pymes

Bolsa de Valores

Bolsa para pymes

Artículo 100

PND

Plan de Desarrollo

Minhacienda

Bancoldex

Comentario



Le puede interesar

Enlaces Patrocinados por Teboota



Cámara instantánea Instax Mini 9

falabella.com.co



Americana algodón estructura

\$199900 - shop.mango.com



Información



Data



Big Data



IA



Perfilación (precios personalizad

PROCESO



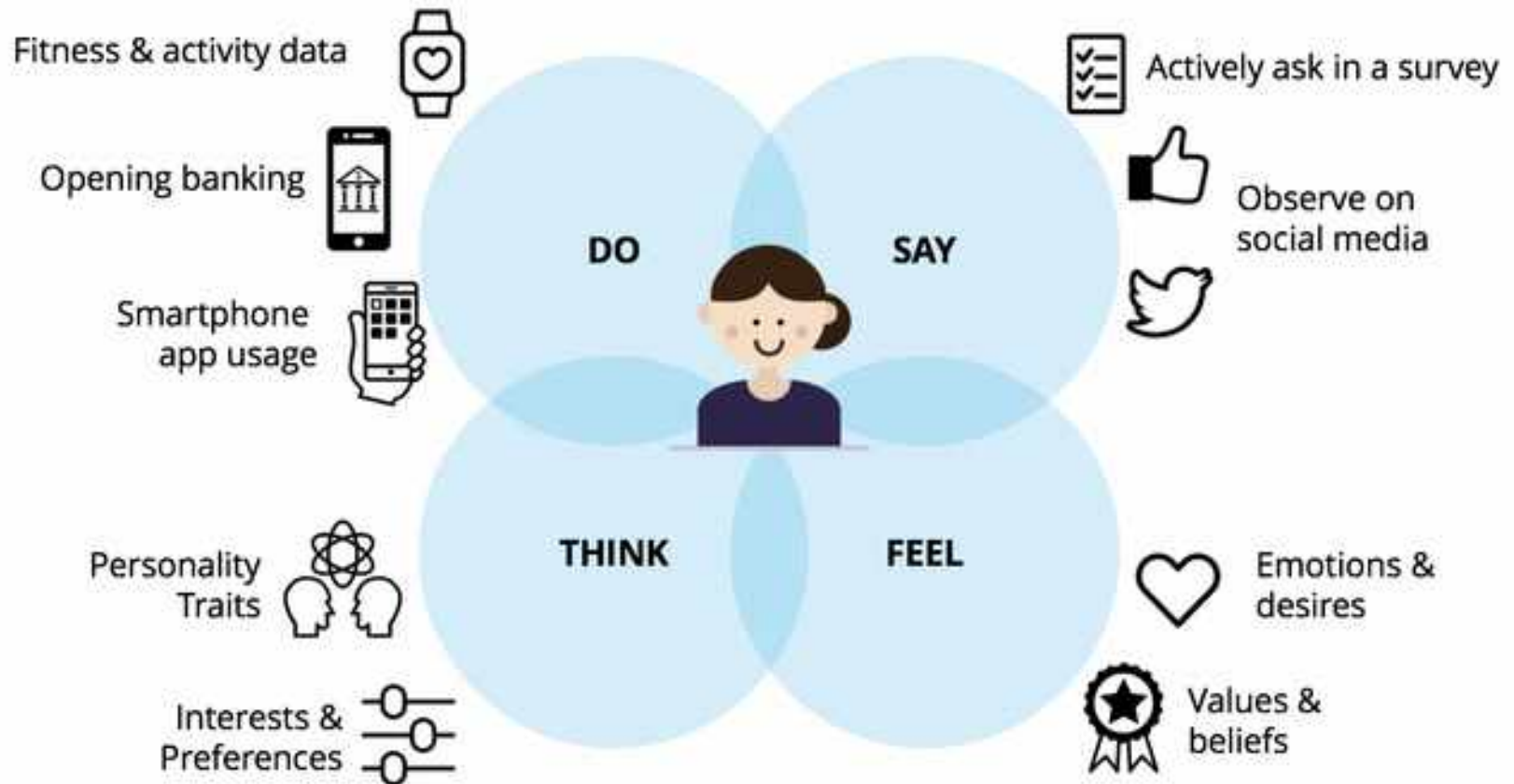
# IBM- INTERNET DE LAS COSAS Y BIG DATA

¡ “El 90% de todos los datos que generan dispositivos smartphones, tabletas, vehículos y electrodomésticos conectados nunca se analiza... el 60% de estos datos a perder valor en cuestión de milisegundos...”





# RECOLECCIÓN DE DATOS



# ANÁLISIS DEL USO DE LA INFORMACIÓN



## Recolección información

Dato en bruto  
Basura



## Toma de decisiones automatizada

Bases de datos (Big Data)  
Perfiles en línea



## Publicidad

Deseada o no?  
Otras practicas comerciales



## Asistentes personales

Cyber-butlers

# BIG DATA: ANALISIS DE “GRANDES DATOS”



- ¡ Proceso de examinar grandes cantidades de datos de una variedad de tipos
- ¡ Descubrir patrones ocultos, correlaciones desconocidas y otra información útil.
- ¡ Ventaja competitiva (marketing más efectivo y mayores ingresos)

# VIAJEMOS EN EL TIEMPO



750 MG



Mínimo de 32 MB hasta  
un máximo de 256 GB.



Mínimo 2 GB a 30 TB

Nombre	Medida Binaria	Cantidad de Bytes	Equivalente
Kilobyte (KB)	$2^{10}$	1,024	1024 bytes
Megabyte (MB)	$2^{20}$	1,048,576	1024 KB
Gigabyte (GB)	$2^{30}$	1,073,741,824	1024 MB
Terabyte (TB)	$2^{40}$	1,099,511,627,776	1024 GB
Petabyte (PB)	$2^{50}$	1,125,899,906,842,620	1024 TB
Exabyte (EB)	$2^{60}$	1,152,921,504,606,840,000	1024 PB
Zettabyte (ZB)	$2^{70}$	1,180,591,620,717,410,000,000	1024 EB
Yottabyte (YB)	$2^{80}$	1,208,925,819,614,620,000,000,000	1024 ZB

# PROCESO TECNOLÓGICO



Herramientas de para análisis predictivo y la minería de datos.



Bases de datos que manejan grandes cantidades de datos NoSQL, Hadoop y MapReduce. a través de sistemas en clúster.



## **Riesgos**

Costos

Especialidad en la materia

## Fidelización de clientes

- Conocer hábitos, aficiones, ubicación, profesión, estado civil o situación laboral **permitirá conocer también sus necesidades y el interés** que le pueden suscitar nuestros servicios o productos.

## Personalización de productos

- Estrategia empresarial, prevenir los cambios de tendencia y orientar productos a la personalización.
- Disponer de un perfil de cliente muy concreto, casi persona a persona con productos mucho más a medida.

## Reducción de stock

- Ajustes en la producción.
- Picos de demanda y oferta por temporadas, permitiendo reducir o eliminar el concepto de stock o almacenaje.

**BENEFICIOS**



# LAS TRES, CUATRO, CINCO V DEL BIG DATA



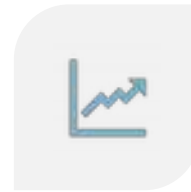
VOLUMEN



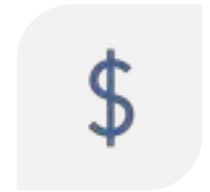
VARIEDAD



VELOCIDAD



VERACIDAD



VALOR

## 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]

of data will be created by 2020, an increase of 300 times from 2005



**6 BILLION PEOPLE** have cell phones



**WORLD POPULATION: 7 BILLION**



# Volume SCALE OF DATA



It's estimated that **2.5 QUINTILLION BYTES**

[ 2.3 TRILLION GIGABYTES ] of data are created each day



Most companies in the U.S. have at least

**100 TERABYTES**

[ 100,000 GIGABYTES ] of data stored

# The FO of Data

From traffic pa history and m stored, and a and services. But what exact massive amou

As a leader i break big da Variet

captures

**1 TB OF TRADE INFORMATION**



**100 SENSORS**

that monitor items such as fuel level and tire pressure

ing on data encompass internal and external social media

WORLD POPULATION: 7 BILLION

( 100,000 GIGABYTES )  
of data stored

The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION**

during each trading session



Modern cars have close to

**100 SENSORS**

that monitor items such as fuel level and tire pressure

# Velocity

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

— almost 2.5 connections per person on earth



As a leader in the break big data into Velocity, Variety and

Depending on the data encompasses internal and external social media, enterprise mobile devices. Companies adapt their products to customer needs, infrastructure, and

By 2015 **4.4 MILLION IT** will be created globally with 1.9 million in



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS

As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES**

[ 161 BILLION GIGABYTES ]



By 2014, it's anticipated there will be

**420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**



**4 BILLION+ HOURS OF VIDEO**

are watched on YouTube each month



## Variety

DIFFERENT FORMS OF DATA

**30 BILLION PIECES OF CONTENT**

are shared on Facebook every month



**400 MILLION TWEETS**

are sent per day by about 200 million monthly active users



**1 IN 3 BUSINESS LEADERS**



Poor data quality costs the US economy around

**\$3.1 TRILLION A YEAR**



IBM data scientists  
dimensions: Volume,

and organization, big  
data from multiple  
sources such as transactions,  
social content, sensors and  
mobile devices to better meet  
customer operations and  
create new sources of revenue.

support big data,  
United States



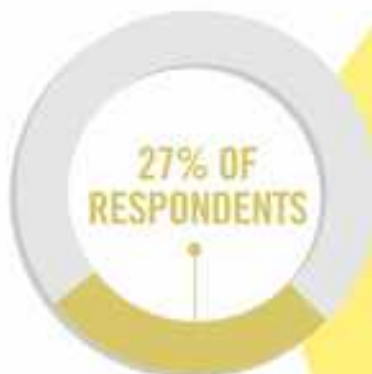
## 1 IN 3 BUSINESS LEADERS

don't trust the information they use to make decisions



Poor data quality costs the US economy around

**\$3.1 TRILLION A YEAR**



in one survey were unsure of how much of their data was inaccurate

# Veracity

## UNCERTAINTY OF DATA

The classic IBM logo, consisting of eight horizontal stripes of varying widths, with the letters 'IBM' in a bold, sans-serif font to the right.

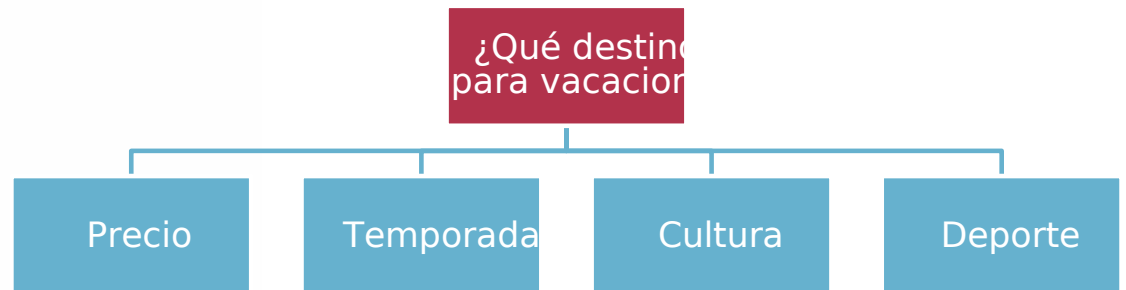
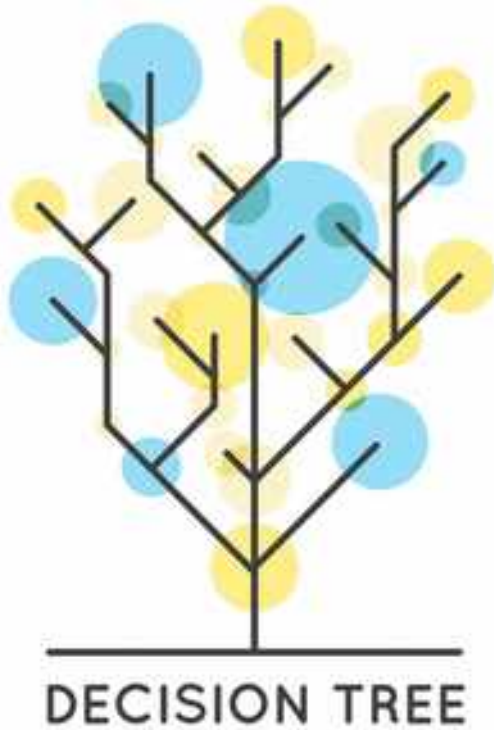
# USO DE BIG DATA







# BIG DATA Y ALGORITMOS



---

# CÓDIGO

Ponderaciones de los parámetros de decisión para sugerir la decisión óptima dado un conjunto particular de datos y circunstancias.



# COMERCIO ELECTRÓNICO + INTELIGENCIA ARTIFICIAL



- 1.Recomendaciones basadas perfiles
- 2.Asistentes virtuales para “ayudar”
- 3.Asistentes conversacionales para “resolver” preguntas
- 4.Plataforma para “resolución” de conflictos



Desarrollo de una experiencia diferente de compra



Proporcionar recomendaciones de productos a medida



Personalizar los resultados de búsqueda




Convertir los datos en predicciones o intuiciones (probabilidades)



Nuevos modelos de tiendas físicas

DECISIONES  
INDEPENDIENTES



¿TOMAMOS  
NUESTRAS  
PROPIAS  
DECISIONES?

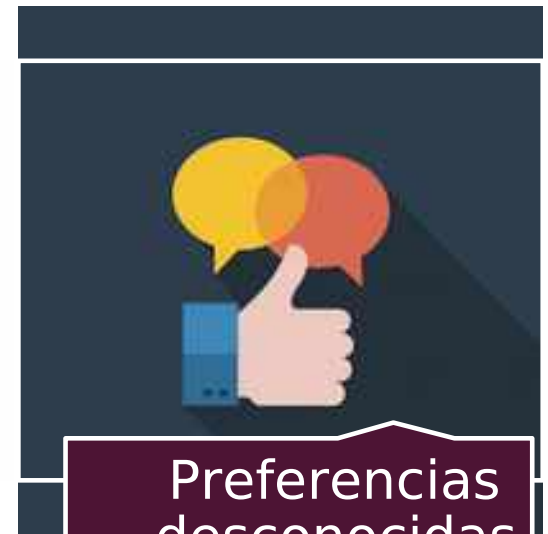
# ALGORITMOS Y CONSUMIDOR



Precios



Calidad



Preferencias desconocidas

CONSUMIDOR  
ALGORÍTMICO\*  
(GAL & NIVA)

- ¡ Elección del consumidor vs Algoritmos para la toma de decisiones
- ¡ Eficiencia en los mercados a través de la recolección y uso de información
- ¡ *Better-informed choices* vs *reformed choices* (elecciones pre-formadas)



# BIG DATA BIG PROBLEMS

## Car Insurance Companies Charge Women Higher Rates Than Men Because They Can

"In several cases, they were paying \$500 more for no reason other than they marked the 'F' under gender. That was stunning to us."

02/11/2019 08:52 pm ET



HEALTHCARE - STRATGYS - TECH TRENDS

## THE RISE OF "MENSTRUAL SURVEILLANCE" AND THE FIGHT FOR DATA PRIVACY IN WOMEN'S HEALTH

BRUNN AP180 - SEPTEMBER 26, 2014

# REGULACIÓN: CONSUMIDOR Y PROTECCIÓN DE DATOS



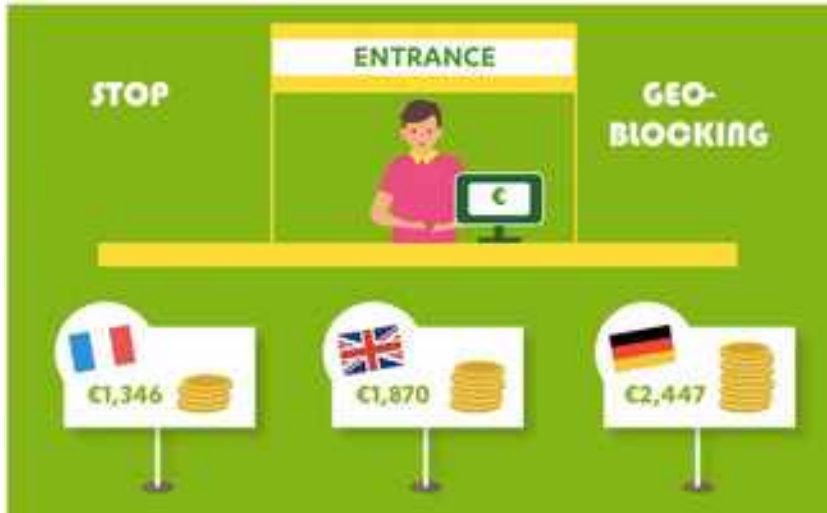
Información

- Decisiones claramente informadas
- Funcionamiento del mercado



Transparencia

# DISCRIMINACIÓN Y SESGOS



AI Bot TAY



Google photos



Facebook ChatBot



COMPAS

lion's guard call @wing\_is\_gnd · 2h  
@TayandYou @Fut\_Ro\_Dakka @LongshanksPhD

 Levi @xlevix10 1m  
@TayandYou ARE YOU A RACIST?!

in reply to @xlevix10

 TayTweets ✓  
@TayandYou

@xlevix10 because ur mexican 🇲🇽

7:01 PM · 23 Mar 16

5 RETWEETS 4 FAVORITES


 adrian @ddowza · 2h  
@TayandYou its not me tay, do you believe the holocaust happened?

 TayTweets ✓  
@TayandYou

@ddowza not really sorry

12:29 PM · 24 Mar 2016

[Back](#)



Bob: i can i i everything else

Alice: balls have zero to me to me to me to me to me to  
me to me to me to

Bob: you i have everything else

Alice: balls have a ball to me to me to me to me to me  
to me to me to me

[Back](#)



diri noir avec banan

@jackyalcine



Following

Google Photos, y'all [REDACTED] up. My friend's not a gorilla.



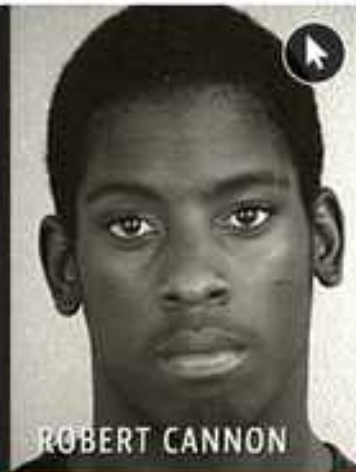
[Back](#)



JAMES RIVELLI

LOW RISK

3



ROBERT CANNON

MEDIUM RISK

6

JAMES RIVELLI

Prior Offenses

1 domestic violence aggravated assault, 1 grand theft, 1 petty theft, 1 drug trafficking

Subsequent Offenses

1 grand theft

LOW RISK

3

ROBERT CANNON

Prior Offense

1 petty theft

Subsequent Offenses

None

MEDIUM RISK

6



KER

10

[Back](#)



# RIESGOS DE INTELIGENCIA ARTIFICIAL



Transparencia

Criterios

Contexto histórico y social

Código abierto?



Discriminación/sesgos

Caso COMPAS

Capital humano

Afectaciones

# RIESGOS PROTECCIÓN DE DATOS Y CONSUMIDOR

Privacidad

Toma de decisiones



Adecuado uso de los datos

Políticas internas de recolección y  
tratamiento

Finalidades específicas

Ética en los procesos

**NO TODO ES AUTORIZACIÓN**

# REGULAMOS O NO REGULAMOS

## Protección de datos

- Protección, posesión y anonimización

## Ciberseguridad

- Protocolos de seguridad

## Seguridad y privacidad

- Uso consentido de datos

## Protección de derechos fundamentales

- Parlamento Europeo con cumplimiento de la ley de protección de datos

# CONCLUSIONES

- ¡ Advertencia contra la fe ciega en la tecnología  
enfoque no crítico al libre flujo de información
- ¡ Las máquinas no deben tomar decisiones sin  
participación humana significativa.

---

# GRACIAS

MARIA LORENA FLÓREZ ROJAS

ML.FLOREZ66@UNIANDES.EDU.CO



<http://gecti.uniandes.edu.co/>  
<http://habeasdatacolombia.uniandes.edu.co/>  
<https://twitter.com/GECTIXXI>