

# CIENCIA DE DATOS

*“Tecnología aplicada al descubrimiento de conocimiento en grandes volúmenes de datos para una toma de decisiones más acertada”*

# CONTENIDO

1. Antecedentes
2. La Ciencia de Datos
3. Tipos de Técnicas
4. El proceso
5. Técnicas Supervisadas y no Supervisadas
6. Aplicaciones
7. Herramientas
8. Demanda
8. Algunas Conclusiones

# En el Mercado se observa ...

- *Crecimiento y complejidad de las organizaciones*

“Las compañías que ven los datos como un **recurso estratégico** son las que sobrevivirán y progresarán”

Ref. Marr, Bernard: Data Strategy. 2017

“Hay tres áreas fundamentales en la que los datos son realmente importantes para los **negocios**:

- Mejora en la toma de decisiones
- Mejora en la actividad
- Conversión de datos en dinero”

Ref. Marr, Bernard: Data Strategy. 2017

- *Mercado cada vez más competitivo*

“Para el 2020, la mitad del software de análisis de datos de Empresas incluirá **facultades predictivas** para analizar los datos”

Data Corporation, 2017

“Los datos se están convirtiendo en la nueva materia prima de los **negocios**”

Ref. “The Economist”

- *Incremento vertiginoso en la cantidad de datos que se generan diariamente*

“Cada dos días se crea **tantos datos** como lo habíamos hecho desde el principio del tiempo hasta el 2003”

Ref. Marr, Bernard: Data Strategy. 2017

“El Big Data y sus implicaciones **afectarán a todos** y cada uno de los negocios en su funcionamiento y oportunidad”

Ref. Marr, Bernard: Data Strategy. 2017

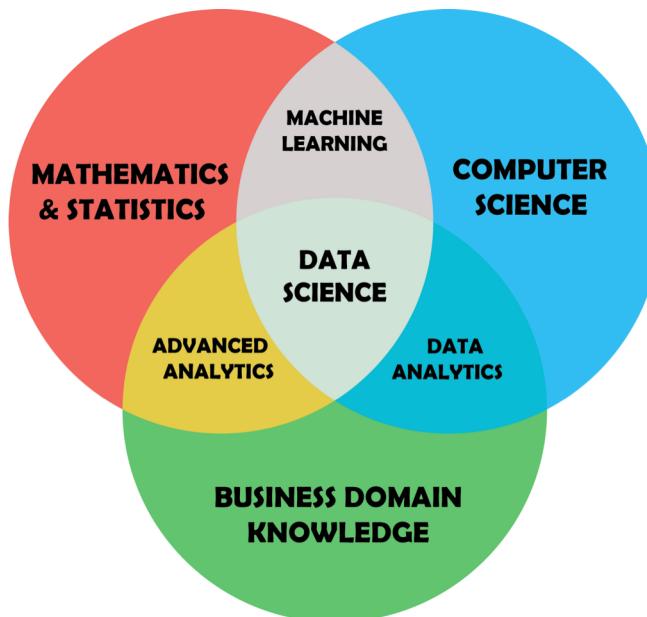
- ***Acceso a tecnologías de procesamiento de datos masivos al alcance de las organizaciones***

El acceso a potencia computacional en velocidad y volumen de procesamiento de datos es cada vez más posible en términos de costos.

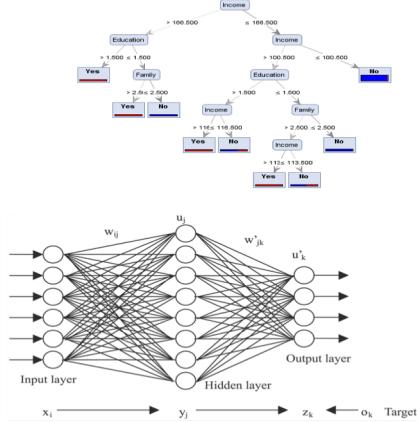
“El éxito de una organización está en función directa a su pericia en la utilización de técnicas de Ciencia de Datos en el **descubrimiento de conocimiento** para una toma de decisiones más acertada en su gestión estratégica con una base tecnológica esencial.”

**La Ciencia de Datos** es un campo multidisciplinario que utiliza métodos científicos, procesos, algoritmos y sistemas para extraer conocimiento de datos estructurados y no estructurados.

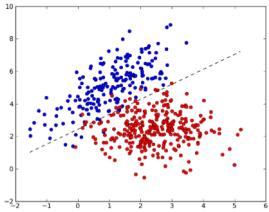
<sup>a b c</sup> Dhar, V. (2013). "Data science and prediction". *Communications of the ACM*. 56 (12): 64–73. doi:10.1145/2500499.



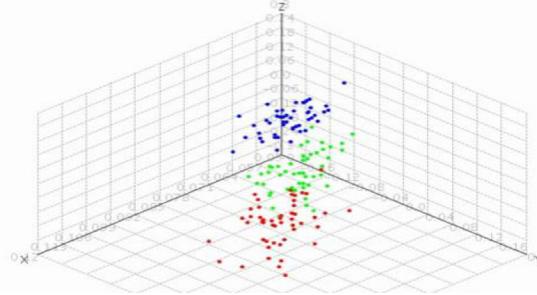
# Tipos de técnicas



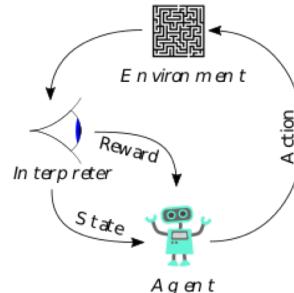
SUPERVISED  
LEARNING



UNSUPERVISED  
LEARNING



REINFORCEMENT  
LEARNING



[https://en.wikipedia.org/wiki/Reinforcement\\_learning](https://en.wikipedia.org/wiki/Reinforcement_learning)



DATASET				
sepal_length	sepal_width	petal_length	petal_width	class
5.9	3.1	5.4	2.1	Iris-virginica
6.5	3	5.8	2.2	Iris-virginica
4.9	3.1	1.5	0.1	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
5.2	2.7	5.1	1.9	Iris-virginica
5.4	3	4.5	1.5	Iris-versicolor
6.4	2.7	5.3	1.9	Iris-virginica
7.2	3	5.8	1.6	Iris-virginica
4.9	3	1.4	0.2	Iris-setosa
6.4	2.8	5.6	2.1	Iris-virginica
5.8	2.8	5.1	2.4	Iris-virginica
*	*	*	*	
6.4	3.1	5.5	1.8	Iris-virginica
6.2	3.4	5.4	2.3	Iris-virginica
4.6	3.1	1.5	0.2	Iris-setosa

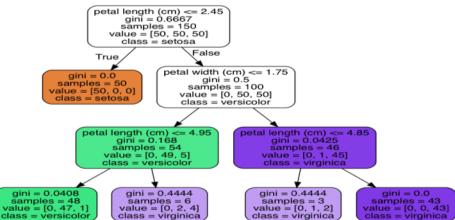
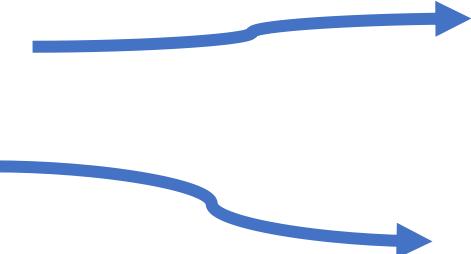
<https://archive.ics.uci.edu/ml/datasets/iris>

[https://www.eurocoms.co.uk/  
it-hardware/data-storage/](https://www.eurocoms.co.uk/it-hardware/data-storage/)



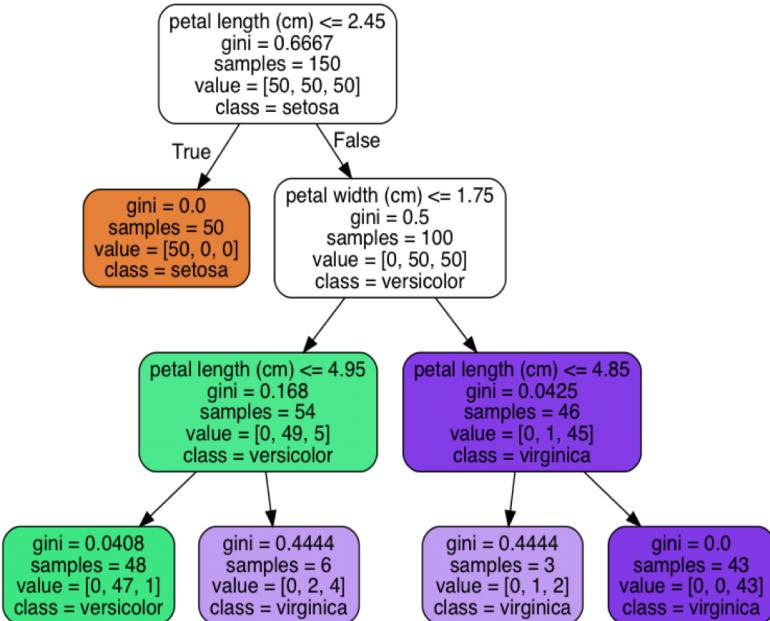
class
Iris-versicolor
Iris-virginica
Iris-virginica
Iris-setosa
Iris-versicolor
Iris-setosa
Iris-virginica

Ranking	Feature
1	sepal_width
2	petal_length
3	petal_width



# Resultados que apoyan la Gestión Estratégica

## Descriptivos



## Predicciones

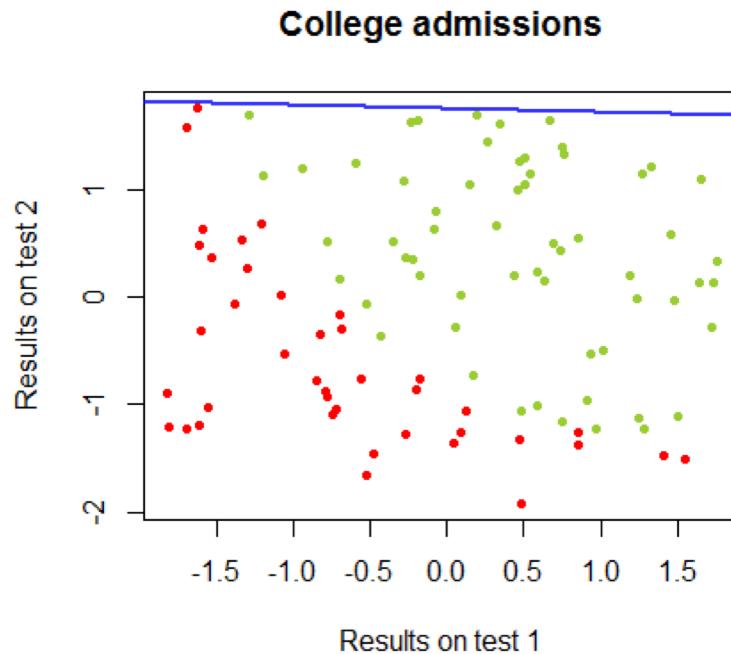
class
Iris-versicolor
Iris-virginica
Iris-virginica
Iris-setosa
Iris-versicolor
Iris-setosa
Iris-virginica

ej. GBM

## Ranking de Variables

Ranking	Feature
1	sepal_width
2	petal_length
3	petal_width

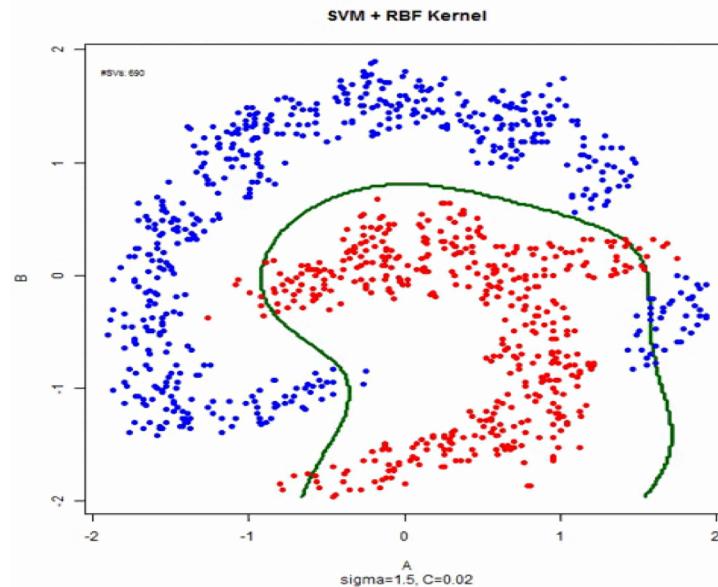
# Modelos Lineales – Regresión Logística



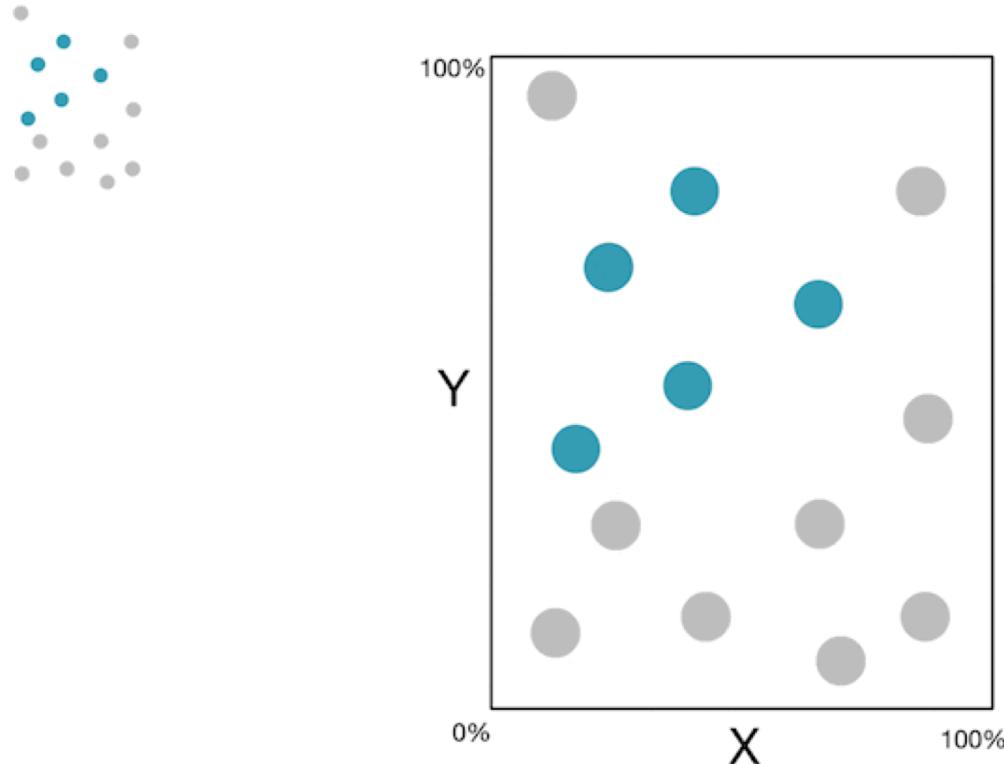
[https://stats.stackexchange.com/questions/162257/  
whats-the-difference-between-logistic-regression-and-perceptron](https://stats.stackexchange.com/questions/162257/whats-the-difference-between-logistic-regression-and-perceptron)

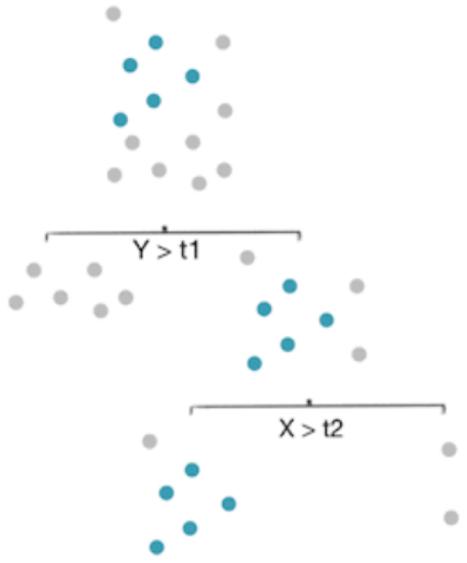
# Modelos No Lineales – Máquinas de Vectores de Soporte (Radial)

SVM Radial

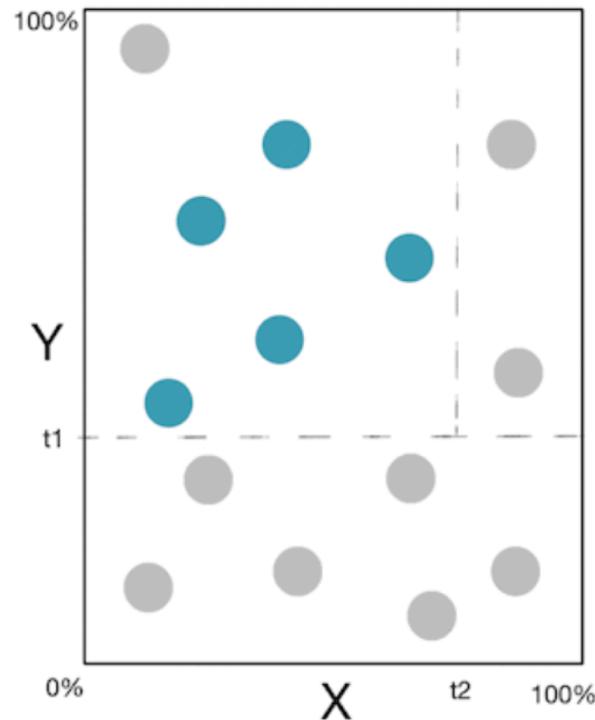


<https://gfycat.com/gifs/detail/RaggedShorttermHalcyon>

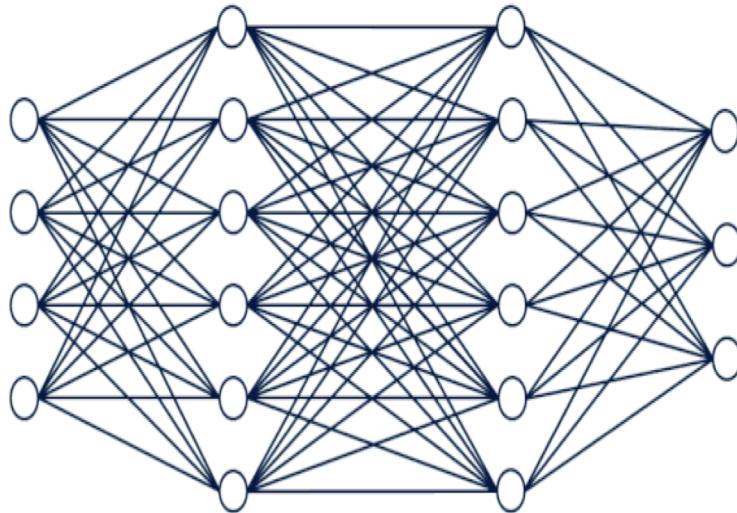




DONE

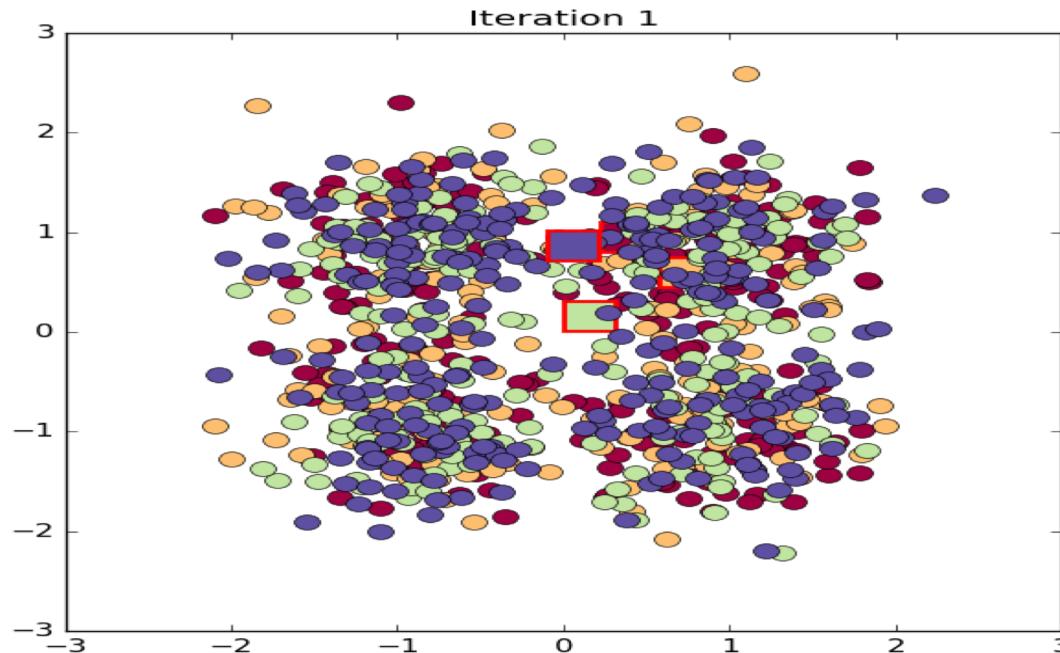


# Modelos Lineales y No Lineales – Redes Neuronales Artificiales



<https://community.alteryx.com/t5/Data-Science-Blog/It-s-a-No-Brainer-An-Introduction-to-Neural-Networks/ba-p/300479>

# Modelos de Segmentación - Clustering



# Aplicaciones ...

## **CASO:** Riesgo crediticio

**DATOS:** Datos demográficos, antecedentes crediticios, solvencia, tiempo de crédito, Plan de pagos, interés, cuotas, etc.

**RESULTADO:** Grado de riesgo de acuerdo al cliente

**CONOCIMIENTO:** Identificación de servicios personalizados a clientes de riesgo.

## **CASO:** Servicios en telefonía

**DATOS:** Tipos de servicios, Consumo histórico de servicios ofertados, Datos demográficos, Características de consumo de servicios.

**RESULTADO :** Perfiles de consumo de clientes

**CONOCIMIENTO:** Elaboración de servicios con grados de respuesta mayores

## **CASO:** Compras en supermercados

**DATOS:** Registro de compras según clasificación de productos, estación del año, hora, día de compra, cantidad, monto, etc.

**RESULTADO :** Patrones de compra de clientes

**CONOCIMIENTO:** Proyecciones más exactas de inventario, oferta óptima de productos

# Aplicaciones ...

**CASO:** Análisis del sentimiento de productos y servicios en Redes Sociales

**DATOS:** Estudio de los textos que publican los usuarios, características de productos o servicios, histórico de resultados, etc..

**RESULTADO :** Grado de percepción

**CONOCIMIENTO:** Campañas más acertadas de ofertas de productos y servicios.

**CASO: Marketing empresarial**

**DATOS:** Campañas emitidas, duración, Época del año, datos demográficos, impacto, Número de contactos, respuestas recibidas, etc.

**RESULTADO:** Grado de respuesta

**CONOCIMIENTO:** Marketing focalizado en clientes potenciales.

**CASO: Deserción de estudiantes universitarios**

**DATOS:** Información demográfica, rendimiento académico, información financiera, etc.

**RESULTADO :** Perfil de estudiantes desertores.

**CONOCIMIENTO:** Mecanismos de reducción de la deserción.

# Aplicaciones ...

**CASO:** Cultivo de productos

**DATOS:** Condiciones de la tierra, altura, humedad, temperatura, riego, precipitación, etc.

**RESULTADO:** Perfil de productos

**CONOCIMIENTO:** Condiciones de cosecha más óptimas.

**CASO:** Detección de personas en imágenes

**DATOS:** Imágenes en áreas de ingreso, etc.

**RESULTADO:** Detección de personas

**CONOCIMIENTO:** Condiciones de ingreso con niveles de seguridad más óptimos.



# RETOS EN MACHINE LEARNING

(Géron, Aurélien: "Hands-On Machine Learning with Scikit-Learn & TensorFlow")

Cantidad insuficiente de  
datos de entrenamiento

Datos de entrenamiento  
no representativos

Datos de baja calidad

Variables irrelevantes

# Herramientas ...

## Lenguajes

- Lenguaje Python – Anaconda®
- Lenguaje R – Rstudio®
- TensorFlow®
- SQL
- entre otros ...

## Entornos gráficos

- Rapidminer®
- Knime®
- entre otros ...

## Procesamiento en Big Data

- Bases de Datos NoSQL
- Hadoop, Spark

# Algunas conclusiones . . .

- Tecnología con potencial para la Gestión Estratégica de las organizaciones
  - Factibilidad de acceso a tecnología computacional en el medio
- Un campo muy amplio de aplicaciones de diversos ámbitos
- Un crecimiento acelerado en la demanda de profesionales en la Ciencia de Datos