

# Aprendizaje automatizado

PRESENTACIÓN DEL CURSO

---

Gibran Fuentes-Pineda

Febrero 2021

# Panorama actual del análisis de datos

- Muchos más datos y a mayor velocidad

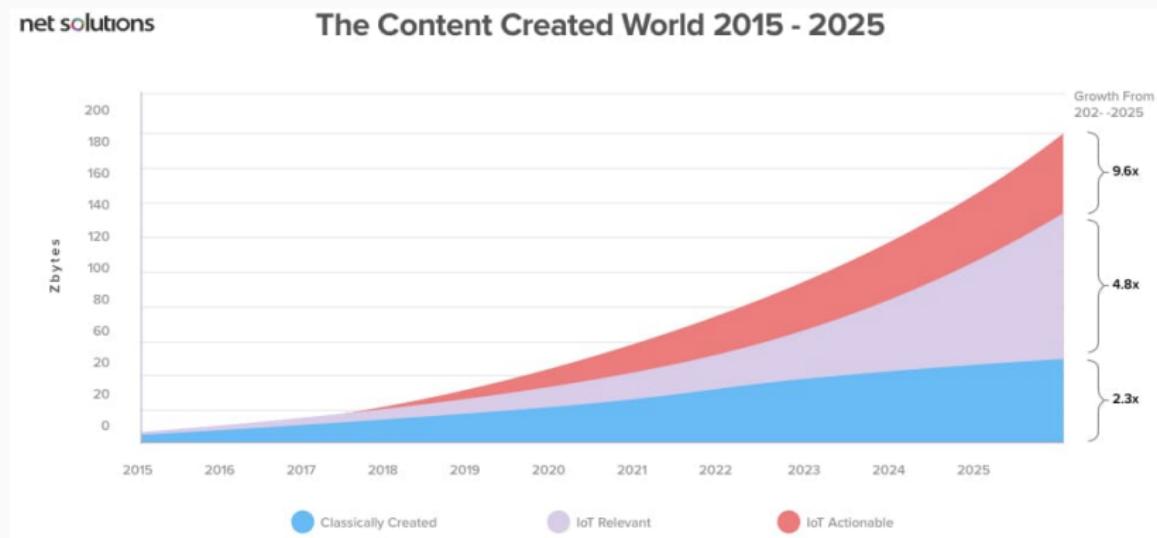


Imagen tomada de <http://www.tech-dynamics.com/wp-content/uploads/2014/02/BigDataChart.png>

# Panorama actual del análisis de datos

- Más parámetros y más datos no estructurados

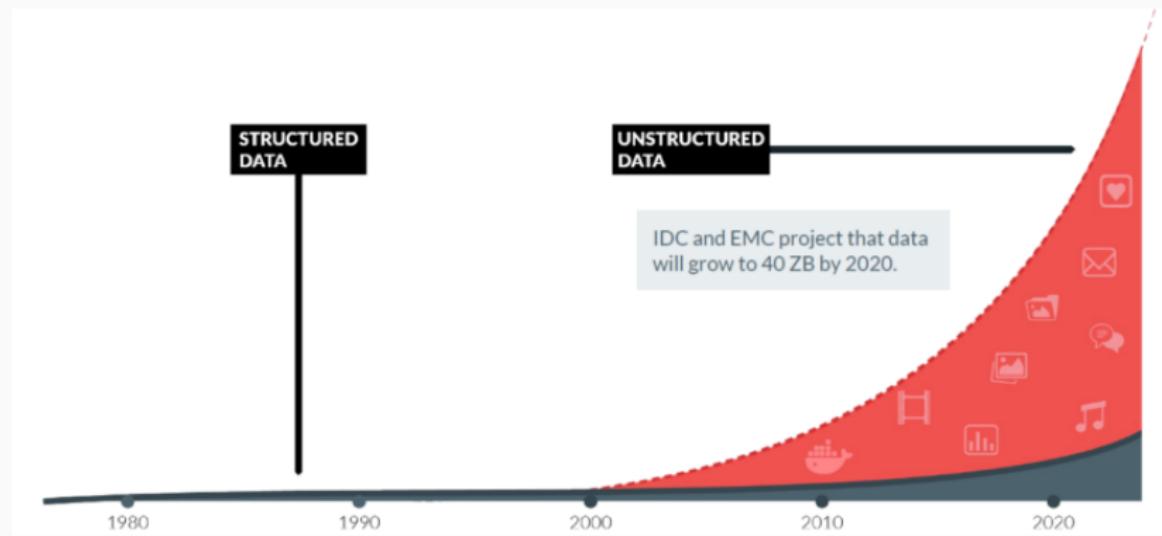


Imagen tomada de <https://www.datanami.com/2017/02/01/solving-storage-just-beginning-minio-ceo-periasamy/>

# Panorama actual del análisis de datos

- Más aplicaciones, nuevos retos



Imagen tomada de *The Washington Post*

# Panorama actual del análisis de datos

- Análisis menos descriptivo, más predictivo/prescriptivo

Involucramiento humano

Complejidad



Descriptivo   Diagnóstico   Predictivo   Prescriptivo

- Análisis menos descriptivo, más predictivo/prescriptivo

## Inferencia



Descriptiva   Diagnóstico   Predictiva   Prescriptiva

# Obteniendo reglas de los humanos

---

- Sistemas expertos
  - Reglas programadas directamente

# Obteniendo reglas de los humanos

---

- Sistemas expertos
  - Reglas programadas directamente
  - Ejemplos
    - Deducción de estructuras moleculares (DENDRAL)
    - Diagnóstico médico (MYCIN)

# Obteniendo reglas de los humanos

- Sistemas expertos
  - Reglas programadas directamente
  - Ejemplos
    - Deducción de estructuras moleculares (DENDRAL)
    - Diagnóstico médico (MYCIN)
- ¿Qué reglas necesito para reconocer dígitos escritos a mano?



# La idea: aprender de los datos

- El aprendizaje automatizado es el estudio de programas que aprenden a partir de ejemplos en lugar de ser manualmente programados



# La idea: aprender de los datos

---

*“Un programa aprende de la experiencia E con respecto a la tarea T y una medida de rendimiento P, si el rendimiento en T medido por P mejora con E”*

---

Tom M. Mitchell

# Ejemplos de aplicaciones

- Sugerir palabras conforme tecleas en el celular
- Predecir los términos de la búsqueda que deseas realizar



# Ejemplos de aplicaciones

- Selección de tweets relevantes

The image displays two separate Twitter feed snippets. The top snippet shows a tweet from IPN Cultura (@IPN\_Cultura) about the movie "Las sufragistas". The bottom snippet shows a tweet from Ryver (@RyverApp) comparing its service to Slack.

**Top Snippet (IPN Cultura):**

IPN retuiteó  
IPN Cultura @IPN\_Cultura 12m  
"Las sufragistas" (2015) es la primera película en la historia que se filma en el Parlamento del Reino Unido.

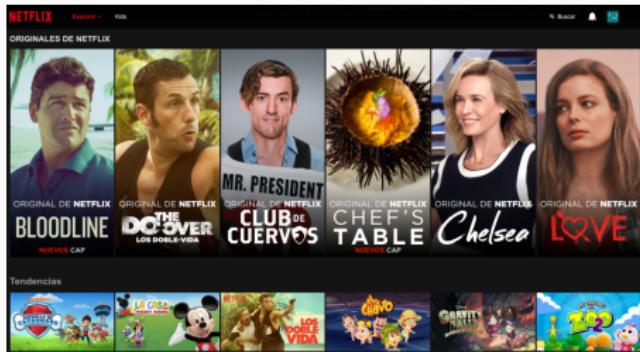
**Bottom Snippet (Ryver):**

Ryver @RyverApp Why pay for Slack when you can get a more powerful solution for free? #freeyourchats

CHAT-ONLY CHAT + POSTS + UNLIMITED USERS

# Ejemplos de aplicaciones

- Recomendación de contenido



# Ejemplos de aplicaciones

- Predicción de anuncios con mayor probabilidad de ser explorados

A Google search results page for the query "samsung galaxy s7". The results include a header with the search term, a "All" filter, and a link to "Search tools". Below the header, it says "About 36,600,000 results (0.77 seconds)". A section titled "Shop for samsung galaxy s7 on Google" displays six sponsored ads for different Samsung Galaxy S7 models, each with a small image, the model name, price, and store information. Below this, there is a "Sponsored" section for the "Nuevo Samsung Galaxy S7" and a link to "Galaxy S7 and Galaxy S7 Edge | Samsung US".

Google search results for "samsung galaxy s7":

All Images News Videos More Search tools

About 36,600,000 results (0.77 seconds)

Shop for samsung galaxy s7 on Google Sponsored

Model	Price	Store
Samsung - Galaxy S7 ...	MX\$17,179.00	Best Buy Mex.
Samsung - Galaxy S7 Flat ...	MX\$14,969.00	Best Buy Mex.
Samsung - Galaxy A7 - ...	MX\$10,399.00	Best Buy Mex.
Samsung Galaxy S6 ...	MX\$12,499.00	Soriana.com
Samsung Galaxy Grand ...	MX\$2,939.00	FAMSA
Samsung Galaxy S7 ...	MX\$8,859.46	eBay

Nuevo Samsung Galaxy S7 - bestbuy.com.mx

[www.bestbuy.com.mx/samsung-galaxyS7](http://www.bestbuy.com.mx/samsung-galaxyS7) ▾

jDesubre lo que puede hacer! Aprovecha MSI y Envío gratis.  
18 MSI y 1 mes gratis - Hasta \$250x51000 en TVs  
Tipos: Pantallas, Videoguegos, Computadoras y Laptops, Tablets y Celulares, Línea Blanca, Cámaras...  
9 Ladera 16, Coyoacán - 01 800 237 8289 - Open today - 11:00 AM - 9:00 PM ▾

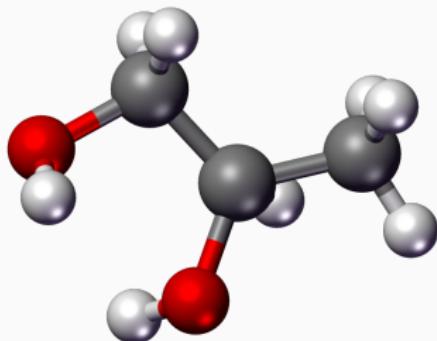
Galaxy S7 and Galaxy S7 Edge | Samsung US

[www.samsung.com/us/explore/galaxy-s7-features-and-specs/](http://www.samsung.com/us/explore/galaxy-s7-features-and-specs/) ▾

Be one of the first to own the newest Samsung phone! The Galaxy S7 & S7 Edge feature Table10-resistance, enhanced cameras and ability to add a microSD ...

# Ejemplos de aplicaciones

- Predicción de propiedades de moléculas



# Ejemplos de aplicaciones

- Detección de partículas exóticas

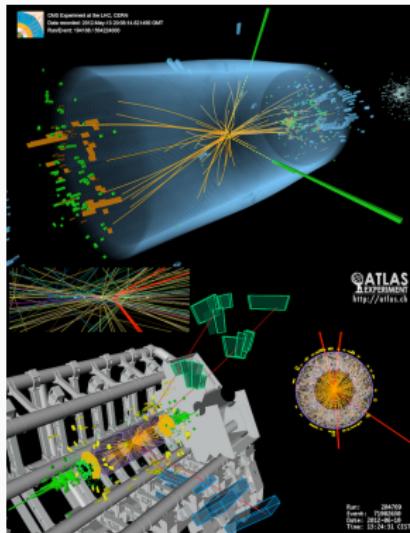
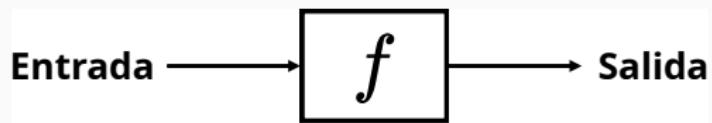


Imagen tomada de [https://en.wikipedia.org/wiki/Higgs\\_boson](https://en.wikipedia.org/wiki/Higgs_boson)

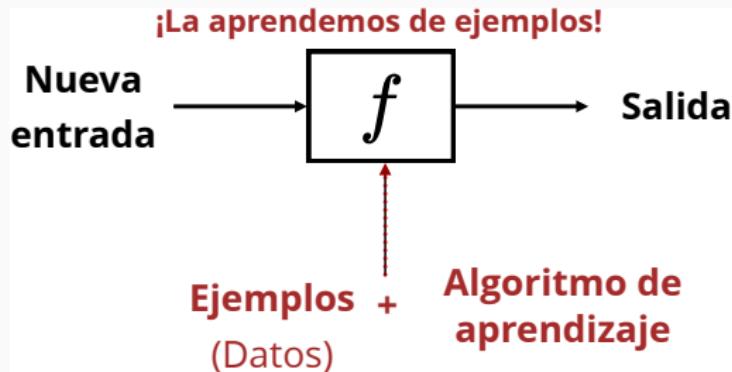
# ¿Cómo funciona?

- Podemos ver cualquier tarea como una función



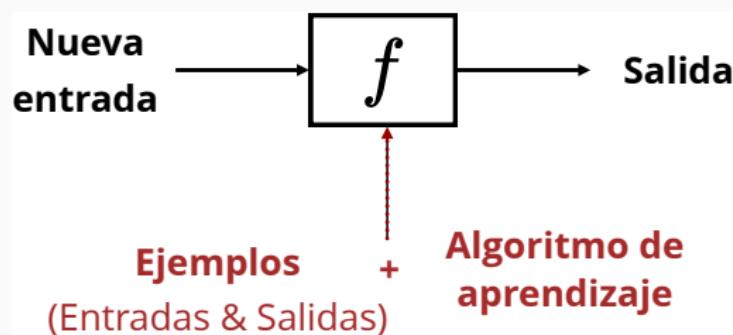
# ¿Cómo funciona?

- El objetivo es aprender la función  $f$  (es decir el modelo) a partir de un conjunto de ejemplos dados



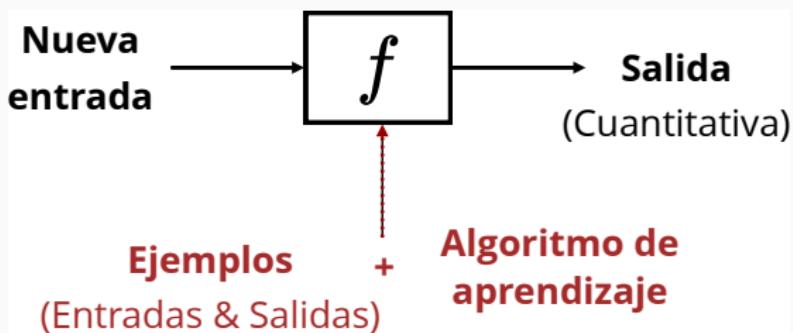
# Aprendizaje supervisado

- Los ejemplos incluyen tanto la entrada como la salida deseada



# Regresión

- Salida continua (cuantitativa)
- Ejemplos: predicción de temperatura de un cuarto, etc.



# Ejemplo de regresión

- Predecir el precio de una casa dado su tamaño

Tamaño ( $m^2$ )	Precio (USD)
433.71	61590.60
446.84	69741.66
582.16	56292.42
833.68	113220.61
920.39	145932.16
931.33	126830.14
1326.25	171278.58
⋮	⋮

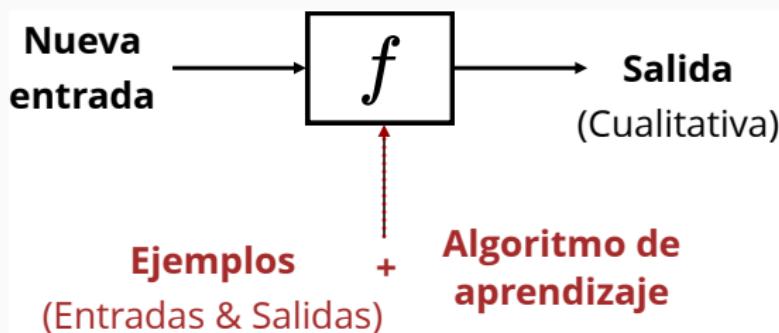
Característica o  
atributo      Respuesta



Imagen tomada de <https://en.wikipedia.org/wiki/House>

# Clasificación

- Salida discreta (cualitativa)
- Ejemplos: detección de spam, reconocimiento de rostros, etc.



# Ejemplo de clasificación

- Clasificar sub-especies de la flor Iris basado en el ancho y largo de su pétalo

Ancho	Largo	Especie
1.4	0.2	Setosa
1.7	0.4	Setosa
1.5	0.1	Setosa
:	:	:
4.7	1.4	Versicolor
4.5	1.5	Versicolor
3.3	1.0	Versicolor
:	:	:

Características o atributo      Respuesta

Setosa



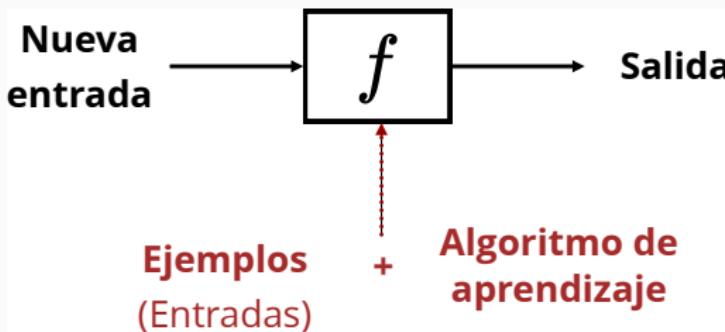
Versicolor



Imagen tomada de [https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set)

# Aprendizaje sin supervisión

- Los ejemplos sólo contienen las entradas sin las salidas deseadas
- Ejemplos de tareas: clustering y descubrimiento de patrones.



# Ejemplos de clustering

- Identificar las áreas de una ciudad que tienen mayor incidencia de un delito particular

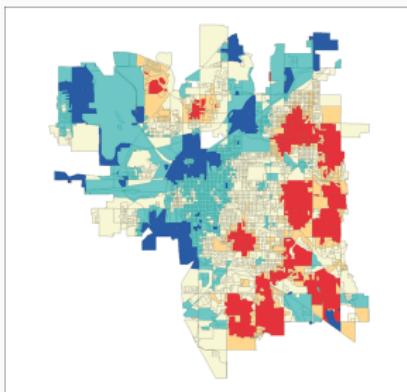


Imagen tomada de [http://www.esri.com/news/arcuser/0405/ss\\_crimestats1of2.html](http://www.esri.com/news/arcuser/0405/ss_crimestats1of2.html)

# Ejemplos de clustering

- Detección de comunidades en redes sociales basado en interacciones

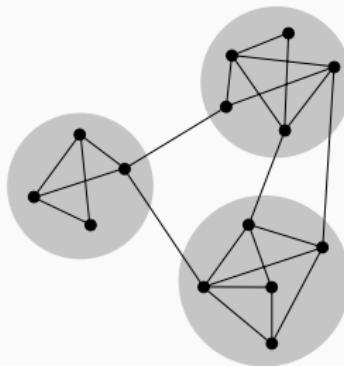


Imagen tomada de [https://en.wikipedia.org/wiki/Community\\_structure](https://en.wikipedia.org/wiki/Community_structure)

# Ejemplos de clustering

- Agrupar clientes con gustos similares



Imagen tomada de [https://en.wikipedia.org/wiki/Salsa\\_\(sauce\)](https://en.wikipedia.org/wiki/Salsa_(sauce))

# Ejemplos de descubrimiento patrones

- Descubrir productos que los clientes compran frecuentemente juntos (por ej. leche y pan)



Imagen tomada de <http://www.iworkwithdata.com/2016/01/>

# Ejemplos de descubrimiento de patrones

- Extraer los tópicos de los que habla una colección de documentos

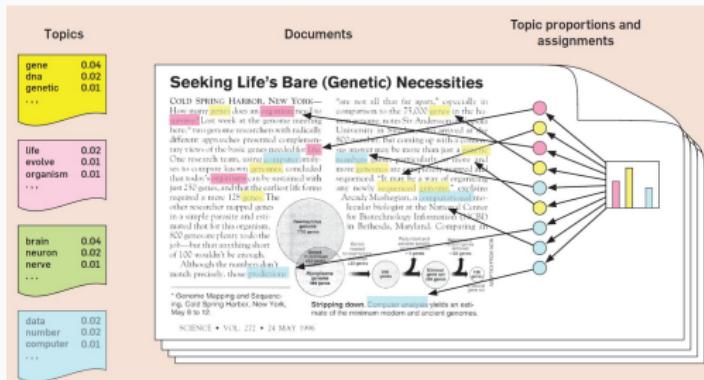


Imagen tomada de Blei. Probabilistic Topic Models, Communications of the ACM, 2012

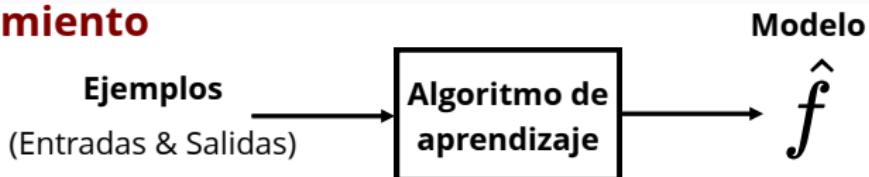
# Proceso de diseño

---

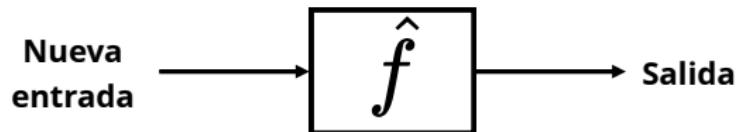
1. Recolectar los ejemplos (datos)
2. Pre-procesar ejemplos y seleccionar/extraer características relevantes (variables de entrada)
3. Elegir el algoritmo de aprendizaje apropiado para el problema
4. Producir diferentes modelos usando diferentes parámetros del algoritmo de aprendizaje y decidir el mejor

# Etapas del aprendizaje automatizado

## Entrenamiento



## Prueba



1. Clasificador bayesiano ingenuo
2. Métodos lineales de regresión y clasificación
3. Selección de modelos
4. Modelos gráficos probabilísticos
5. Modelos de variables latentes
6. Máquinas de vectores de soporte y kernels
7. Ensamblles

# Información del curso

- Horario
  - Martes y jueves de 13:00 a 14:30 hrs.
- Zoom:  
<https://cuaed-unam.zoom.us/j/87323873226>
- Google Classroom: 5urc3yg
- Contacto
  - Gibran Fuentes Pineda: gibranfp@unam.mx
  - Berenice Montalvo Lezama: bere.mcic@gmail.com
  - Ricardo Montalvo Lezama: richardt.pcic@gmail.com
- Material del curso
  - Página: [http://turing.iimas.unam.mx/~gibranfp/cursos/aprendizaje\\_automatizado](http://turing.iimas.unam.mx/~gibranfp/cursos/aprendizaje_automatizado)
  - Repositorio: <https://github.com/gibranfp/CursoAprendizajeAutomatizado>

# Criterio de evaluación

- Proyecto: 60 %
  - Individual o en equipo de máximo 3
  - 2 entregas: propuesta y final
- Tareas: 40 %
  - Ejercicios de programación