

Curso: Principios básicos de la Inteligencia Artificial

Corfo – Pymes en Línea

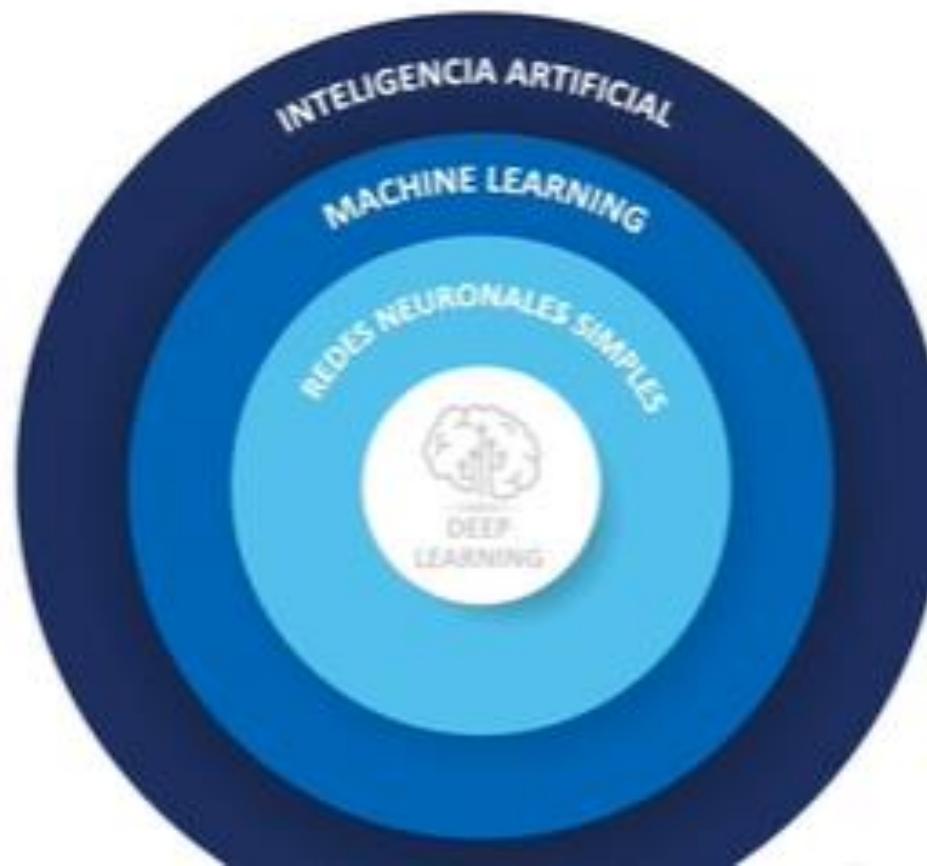
Módulo 1: Introducción a La Inteligencia Artificial e Inteligencia de Negocios

Corfo – Pymes en Línea

Clase 2: Introducción a la inteligencia artificial

- Big data se confunde mucho con Machine Learning
- IA: Capacidad de simular los procesos de inteligencia humana por parte de las máquinas o de sistemas informáticos.
- - Ani: Artificial Narrow Intelligence: Gran avance. Desarrollado para realizar tareas específicas, un solo fin. Cuando se determina si un correo es spam o no, ganar en ajedrez (pero no bachillerato), etcétera, pero sólo sirven para eso.
- - Agi: Artificial General Intelligence: Capacidad de simular los procesos de inteligencia humana. Parece más futurista, pero es la que tiene menos desarrollo y la que ha avanzado más lento, no está pensando para resolver sólo un fin, puede enfrentarse a distintos escenarios. El algoritmo aprende al observar al ser humano y también al observarse a sí mismo.
- Machine learning: Aprendizaje automático: Campo de estudio que le da a los computadores la capacidad de aprender sin ser programados explícitamente. Tiene asociado el uso de los datos y el software, las computadoras necesitan aprender de algo y ese algo son los datos. Tendremos input y output. No requiere necesariamente datos estructurados.
- Datos estructurados: imaginarla como una tabla Excel (por ejemplo) concepto más clásico de data estructurada.
- Datos no estructurados: una imagen, un audio, puede ser por ejemplo, un input.
- Deep Learning: Aprendizaje profundo, subárea del machine learning, concepto clave: Redes neuronales artificiales. Tiene alto poder de aprendizaje, una ventaja comparativa increíble con data no estructurada.

Clase 2: Introducción a la inteligencia artificial



Clase 2: Introducción a la inteligencia artificial

- La inteligencia de negocios se define como un proceso iterativo para explorar y analizar información estructurada sobre un área, para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.
- El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar cambios. Uno de los elementos clave es descubrir tendencias o patrones, en este punto se puede o no aplicar inteligencia artificial.

Clase 2: Introducción a la inteligencia artificial

- Descubrir tendencias o patrones
- Es una herramienta bajo la cual pueden soportar la toma de decisiones.
- Generación de conocimiento a partir de los datos.

Datos ➤ Información ➤ Conocimiento ➤ Acción

Clase 2: Introducción a la inteligencia artificial

- Ejemplo:

Inteligencia de negocios



- ▶ Fecha y hora de compra.
- ▶ Vendedor que atendió.
- ▶ Sucursal donde se hizo la compra.
- ▶ Códigos de los artículos vendidos y su descripción.
- ▶ Unidades de cada artículo vendido.
- ▶ Precio unitario.
- ▶ Precio total.
- ▶ Total pagado.
- ▶ Forma de pago.

TIENDA DON MATIAS
Giro: VENTA DE ROPA
Dirección: LA QUEBRADA 1005
Comuna: PUERTO VARAS
Ciudad: PUERTO VARAS
Fono: 229381115
Vendedor: Juan Tapia
Fecha: 23/07/2020
Forma de Pago: EFECTIVO
Sucursal: LA QUEBRADA 1005, PUERTO VARAS

CANT.	ÍTEM	VALOR U.	DESC.	SUBTOTAL
3	CAMISA LISA ML	\$ 8.999	\$ 0	\$ 44.995

NETO:	\$ 37.811
IVA:	\$ 7.184
Total Exento:	\$ 0
Vuelto:	\$ 0

Clase 2: Introducción a la inteligencia artificial

- Ejemplo:

Operativo

- ▶ Reponer artículos.
- ▶ Asignar turnos.
- ▶ Identificar los productos y medios de pagos más importantes.



Clase 2: Introducción a la inteligencia artificial

- Ejemplo:

Inteligencia de negocios



Se produce una disminución total en las ventas

- ▶ Disminución del número de ticket promedio diarios o por periodo.
- ▶ El monto promedio de los tickets ha bajado.
- ▶ Ambas.



históricamente. Así podemos identificar si el problema es por afluencia a nuestra tienda o si los clientes

Clase 2: Introducción a la inteligencia artificial

- Ejemplo:

Se produce una disminución total en las ventas

- ▶ Debemos tener clara toda la información.
- ▶ Estudiar las relaciones entre los distintos.
- ▶ Nos permite tomar decisiones estratégicas, y esta es la base de la inteligencia de negocio.



**Esta información tiene valor, porque nos permite
tomar decisiones estratégicas, y esta es la base**

Clase 3: Inteligencia de Negocios

- Datos, información, conocimiento, sabiduría, no significan lo mismo.
- Los datos son elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. No dicen nada sobre el por qué ocurren las cosas.
- La información se puede definir como un conjunto procesado de datos que son de utilidad para los tomadores de decisiones.
- Conocimiento: Aquello que permite tomar decisiones, analizar y concluir a partir de la información.

Clase 3: Inteligencia de Negocios

Transformando datos en información



¿Cómo podemos transformar los datos en inteligencia?

- ▶ Los datos no paran de crecer.
- ▶ Los datos se transforman en información mediante análisis.
- ▶ Transformar esa información en conocimiento y en sabiduría, que es una gran fuente de ventaja competitiva.



Clase 3: Inteligencia de Negocios

- Capturar, almacenar, categorizar los datos.
- Necesitamos comprensión, planificación y experiencia.
- Los datos los tenemos brutos: puros y sin procesar. Los datos se vuelven útiles solo cuando se puede extraer significado de ellos. El objetivo final es crear un flujo continuo de información útil y oportuna.

Clase 3: Inteligencia de Negocios



Datos



Los datos se vuelven útiles solo cuando se puede extraer algún significado de ellos.

Crear un flujo continuo de información útil y oportuna.

Clase 3: Inteligencia de Negocios

Datos

Presiona **Esc** para salir de la pantalla completa



Los datos son una
mercancía.

Pago

Gratis

Fuentes públicas.

Los datos también son generados por cada negocio en sí mismo.

**Los datos también son generados por cada negocio
en sí mismo y este es un activo muy valioso**



Clase 3: Inteligencia de Negocios

Datos

Presiona **Esc** para salir de la pantalla completa



¿Cuál es el desafío?

- ▶ Extraer información útil.
- ▶ Tomar decisiones de mejor calidad.
- ▶ Objetivos estratégicos y operativos.
- ▶ No están organizados en un formato que se pueda consumir.



Clase 3: Inteligencia de Negocios

Datos

¿Cuál es el desafío?

- ▶ Se deben realizar procesos de validación, verificación, recopilación, organización, mantenimiento, respaldo y disponibilidad de la información.
- ▶ Es importante que esos datos sean correctos, que esos datos estén bien.



en nuestros programas puede no ser posible su corrección, probablemente descartarlos.

1.3. Inteligencia de negocios



Clase 3: Inteligencia de Negocios

Datos



¿Cuál es el desafío?

- ▶ Muchos de estos datos pueden estar dispersos en diferentes ubicaciones, diferentes sistemas y plataformas independientes.
- ▶ Puede haber una duplicación de la recopilación de datos, y una parte no sabe que los mismos datos se recopilan en otra ubicación.



Y, a veces, puede haber una duplicación de la recopilación de datos

Clase 3: Inteligencia de Negocios

- Es importante que haya una unificación razonable de los datos y que exista conocimiento de su existencia y la disponibilidad de ellos en toda la empresa.

Clase 3: Inteligencia de Negocios



Transformando datos para la toma de decisiones

Sistema de toma de decisiones



Clase 3: Inteligencia de Negocios

- No siempre es necesario recurrir a un modelo para resolver los problemas 😊

Clase 3: Inteligencia de Negocios

Visualización de datos



- Consiste en representar de forma comprensible y medible los datos.
- Comunicar esta información de forma clara y entendible para los usuarios.
- Mientras más simple y limpia una visualización, es más adecuada para transmitir el mensaje correcto.

Clase 3: Inteligencia de Negocios

- Equipo de datos
- Data Engineer: Responsables de controlar el flujo de la información en la organización. Supervisan los sistemas de almacenamiento y la infraestructura. Lenguajes de programación como SQL, Java, Scala y Python para procesar datos y automatizar tareas.
- Data analyst: Describir el presente a través de los datos, generando accionables de negocios y comunicando a las distintas unidades estratégicas que componen la organización. Excel, SQL.
- Data scientist: Encargados de crear e implementar algoritmos que permitan predecir situaciones muy valiosas para las organizaciones. Como la fuga de clientes, detección de ilícitos, etcétera.

Clase 4: ¿Qué puede y no puede hacer el ML?

- Sobreexpectativa sobre lo que puede hacer el machine learning. Esto ocurre porque se comparten los éxitos pero no los fracasos.
- Se piensa que todo lo puede hacer la Inteligencia Artificial, cuando en realidad no es así. Muchas veces puede ser que no sean factibles técnicamente, por presupuesto, recursos, etcétera.
- Para evaluar la factibilidad, es necesario definir el objetivo del proyecto: ¿Cuál es mi problema? ¿cómo quiero hacer más eficiente mi negocio? Segundo evaluar si el ML puede utilizarse, ¿Realmente el ML puede resolver esto.

Clase 4: ¿Qué puede y no puede hacer el ML?

- Andrew NG:
- Primero: la simplicidad del concepto: Se refiere a aquellos problemas que las personas podemos resolver en algunos segundos también lo podrían resolver el ML. Si nos toma demasiado, posiblemente ML no podría (aunque esto no aplica siempre, es decir, si nos cuesta mucho no quiere decir que ML no pueda resolverlo).

Simplicidad del concepto

ampliamente usado, que se llama, los modelos de detección de fuga o Churn.



Identificar fraude

Información histórica de mi cliente
(incluyendo patrones de consumo)

Ilícito o no

Modelo detección ilícito

Determinar a quien entregar un crédito
Información sociodemográfica
+ sistema financiero

Va a pagar o no el crédito

Modelo Scoring

Identificar clientes que se van a fugar

Información histórica de mi cliente

Deja o no de ser mi cliente

Modelo Churn o Fuga

Simplicidad del concepto

- ▶ La regla en general es útil en casos donde el output es fácilmente identifiable para una persona.
- ▶ Es capaz de ver interacciones entre variables que el ojo humano no ve fácilmente.
- ▶ Tenemos que tener un output predefinido con el cual entrenamos el modelo.
- ▶ Con esta información histórica se entrena entonces el modelo ML.



Con toda esta información histórica podemos construir modelos Machine Learning.

Simplicidad del concepto

algún tipo de solución.



ML encuentra interacciones que nosotros no vemos

Riesgo de asumir relaciones espurias, sin hipótesis
podemos ver interacciones donde no existen.



Clase 4: ¿Qué puede y no puede hacer el ML?

- Andrew NG:
- Primero: la simplicidad del concepto: Se refiere a aquellos problemas que las personas podemos resolver en algunos segundos también lo podrían resolver el ML. Si nos toma demasiado, posiblemente ML no podría (aunque esto no aplica siempre, es decir, si nos cuesta mucho no quiere decir que ML no pueda resolverlo).
- Segundo: Contar con una gran cantidad de datos para entrenar el modelo, tiene que existir algo sobre el cual el algoritmo aprenda. Sin datos no se puede generar conocimiento. La cantidad de datos variará dependiendo de la complejidad de los datos.

Presiona **Esc** para salir de la pantalla completa

Tener una gran cantidad de datos

- Tiene como eje fundamental el aprendizaje.
- Sin datos no se puede generar el conocimiento.
- Mientras más complejo el problema mayor es la cantidad de datos que se necesitan.

mientras más complejo el problema mayor es la cantidad de datos que se necesita.

Clase 5: Problemas, preocupaciones y consideraciones éticas en el uso de IA

- No es sencillo delimitar el uso de algunas aplicaciones de la IA.

Sesgo y discriminación

¿Cómo un sistema de IA se puede volver parcial y discriminar a ciertas personas?

Estudio de Bolukbasi et al, 2016

- ▶ Se tomaron textos de google News disponibles.
- ▶ A partir del aprendizaje de los textos, se entrena un modelo el cual entrega una analogía entre palabras.

Hombre - Mujer = X - Y

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{king}} - \overrightarrow{\text{queen}}$$

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}$$

Claramente esta es una respuesta con un marcadísimo estereotipo de género, y esto pasa si no se controla

Sesgo y discriminación

Presiona **Esc** para salir de la pantalla completa

► Otros ejemplos:

Gender stereotype *she-he* analogies.

sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber



Por ejemplo: Ella es a él como enfermera es a cirujano.

Sesgo y discriminación

- ▶ Considerando que el modelo fue entrenado con Google news, entonces se esperaría que no existiera este nivel de estereotipo.
- ▶ ¿Cómo crees que serían los resultados si el modelo fuera entrenado con post de Facebook o twiter?



el resultado del modelo? ¿El sesgo que puede tener por detrás este modelo?

Sesgo y discriminación

- ▶ Estos artículos no se ha reconocido el sexismio al que se pueden llegar en las incrustaciones de texto.

The Conversation Spain

Discriminación racial en la inteligencia artificial

... la discriminación racial. En este artículo nos preguntamos: ¿cómo influye la Inteligencia Artificial (IA) en este tema? Los algoritmos de IA ...

Hace 3 semanas



América Retail

Opinión: Entrevista de trabajo con un robot: la inteligencia artificial ya llegó a la gestión de RRHH

Las empresas recurren cada vez más a la Inteligencia Artificial y los equipos sean más diversos y se eviten problemas de discriminación.

Hace 1 semana



TechTarget ES

¿La tecnología acrecenta el problema del racismo y la discriminación?

Si bien, la inteligencia artificial y el aprendizaje automático alimentan las masiva, discriminación racial y violaciones de derechos humanos".

hace 1 mes



Bynness

Los algoritmos no son perfectos: sesgan por sexo y raza

... adopción de la inteligencia artificial puede suponer la normalización de los sesgos algorítmicos y de la discriminación contra las mujeres.

Hace 2 semanas



se va a ver reflejado en nuestro modelos.

Sesgo y discriminación

- ▶ Estos artículos no se ha reconocido el sexismio al que se pueden llegar en las incrustaciones de texto.
- ▶ Si una máquina decide entonces que no eres apto para un trabajo en función del video que analizó, ¿a quién puedes dirigirte si no estás conforme con la decisión del sistema?



**y los sesgos tienen consecuencias
y debemos ser conscientes de ello.**

Sesgo y discriminación

y esto es lo mismo que hace Bolukbasi

Se tiene que reconocer e identificar este problema.

- ▶ Entender cómo se comporta el modelo en ciertos casos, y “forzarlo” a que no lo haga.

Modelo de otorgamiento de un crédito:

- ▶ Dos personas con exactamente la misma información crediticia.
- ▶ Mismo dicom.
- ▶ Información general del sistema financiero.
- ▶ Mismos niveles de ingresos, estudios, etc.
- ▶ En situaciones así, hay que forzar al modelo a que no castigue.



Las Condes, Chile

Las Pintana, Chile

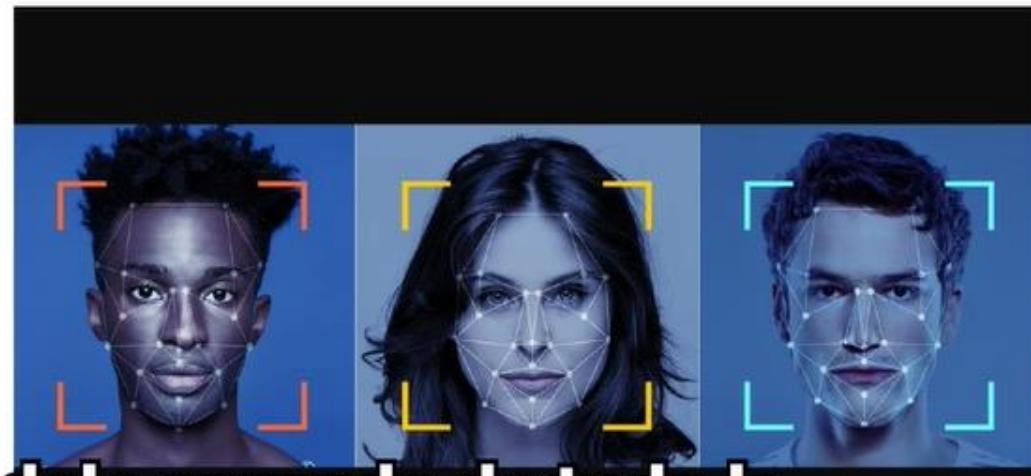
Sesgo y discriminación

Se tiene que reconocer e identificar este problema.

- ▶ Entender cómo se comporta el modelo en ciertos casos, y “forzarlo” a que no lo haga.

Reconocimiento facial:

- ▶ Entrenar al modelo con datos inclusivos.
- ▶ Aprenda de toda la información que tiene.



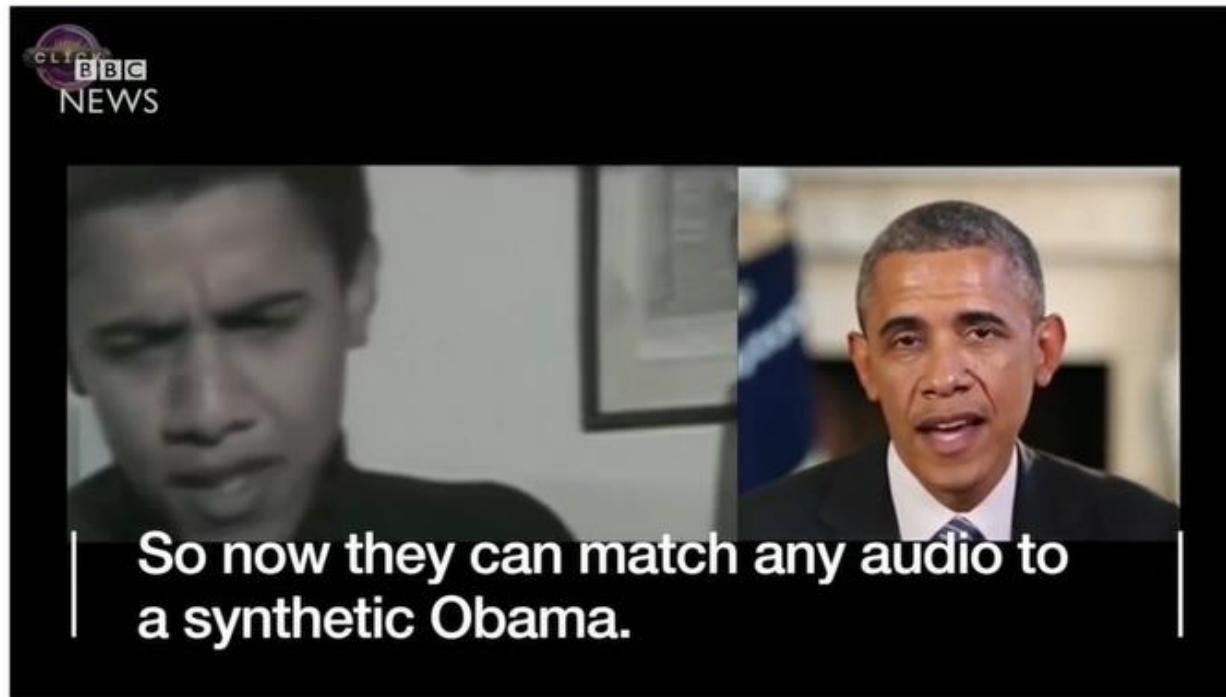
de manera tal que el modelo aprenda de toda la información que existe, de la diversidad racial

Consideraciones éticas

Y se crea este efecto con el cual nosotros pensariámos que es un video total y completamente real.

Fake News

Corresponden a la generación de comentarios falsos, comentarios generados por bot en redes sociales, publicidad, entre otros.



Consideraciones éticas

● Las elecciones presidenciales de EEUU del 2016 y Cambridge Analytica

Cambridge Analytica es una empresa que usa el análisis de datos para desarrollar campañas para marcas y políticos.

Para las elecciones de EEUU se usó información de un test de personalidad de Facebook para inferir perfiles psicológicos de cada usuario.



En este caso se cometieron varias infracciones, principalmente relacionado con la privacidad

Consideraciones éticas

● Reconocimiento facial en Las Condes

Busca la prevención de delitos, investigación y persecución penal.

El mecanismo puede monitorear en tiempo real datos faciales de las cientos de cámaras que tiene la comuna.

- ▶ ¿Quién gestiona el software, guarda los datos y custodia la información?
- ▶ ¿Puede constituir una amenaza potencial para la privacidad este tipo de aplicaciones?
- ▶ ¿Pueden guardar mis datos biométricos sin autorización?

¿Pueden guardar mis datos biométricos sin autorización?

¿Cuál es el límite?

El modelo puede ser mucho más inclusivo, aún más que las mismas personas, pero para eso tiene que existir voluntad y conocimiento.

Hablar del uso de la IA podemos entrar a un debate aún más amplio, cuando ciertos derechos pueden ser vulnerados en pro de otros.

¿Es ético? ¿Qué tanta diferencia tiene con el marketing dirigido?

¿Qué tanta diferencia tiene con el marketing dirigido? El debate así se puede

Lectura sugerida

- <https://www.nytimes.com/es/2018/03/20/espanol/cambridge-analytica-facebook.html>

TECNOLOGÍA

La empresa que explotó millones de datos de usuarios de Facebook



Por Matthew Rosenberg, Nicholas Confessore y Carole Cadwalladr

20 de marzo de 2018

[Read in English](#)

Cuestionario Capítulo 1

En qué caso **no** hay presencia de una inteligencia artificial:

Seleccione una:

- a. Vehículos autónomos
- b. Transcribir un audio de manera manual
- c. chatbots
- d. Reconocimiento facial

[Quitar mi elección](#)

Cuestionario Capítulo 1

La inteligencia artificial estrecha se refiere a:

Seleccione una:

- a. Inteligencia artificial que puede realizar múltiples tareas, como los humanos
- b.
Inteligencia artificial que se usa para un solo fin
- c. Inteligencia artificial que puede ser utilizada de manera inapropiada

[Quitar mi elección](#)

Cuestionario Capítulo 1

¿Cómo es posible que los datos estén sesgados?

Seleccione una:

- a. Porque los datos siempre están manipulados.
- b. Porque los datos siempre mienten.
- c.

Porque representan una realidad que tiene un gran cantidad de sesgos y desequilibrios.

[Quitar mi elección](#)

Cuestionario Capítulo 1

Los desarrollos en inteligencia artificial a día de hoy se puede describir principalmente como:

Seleccione una:

- a. Inteligencia artificial de negocios
- b.

Inteligencia artificial estrecha

- c. Inteligencia artificial general

[Quitar mi elección](#)

Cuestionario Capítulo 1

Según la regla de "simplicidad del concepto", las tareas que se pueden resolver con Machine Learning son aquellas que:

Seleccione una:

- a. Los humanos tardamos unos pocos segundos en resolver
- b. Los humanos tardamos varios minutos en resolver
- c. Los humanos nunca hemos sido capaces de resolver

[Quitar mi elección](#)

Cuestionario Capítulo 1

Cuando un vehículo autónomo se enfrenta a una situación en que debe decidir entre atropellar a un grupo de personas u otro, está en presencia de un:

Seleccione una:

- a. problema de insuficiencia de datos
- b. problema de azar
- c. dilema ético

[Quitar mi elección](#)

Cuestionario Capítulo 1

En Machine Learning lo más importante es/son:

Seleccione una:

- a. tener un computador muy potente
- b. las habilidades de programación
- c. los datos. Sin datos no se puede generar el conocimiento

[Quitar mi elección](#)

Cuestionario Capítulo 1

¿Qué cosa no es capaz de hacer una inteligencia artificial?

Seleccione una:

- a. Identificar objetos en una imagen
- b. Predecir el futuro de manera exacta
- c. Manejar vehículos autónomos
- d. Traducir un texto de un idioma a otro

[Quitar mi elección](#)

Cuestionario Capítulo 1

Seleccione ejemplos de herramientas que nos permiten visualizar datos:

Seleccione una o más de una:

- a. dashboards
- b. infografías
- c. mapas

Cuestionario Capítulo 1

Una de las cosas que se puede hacer para reducir los problemas de sesgo al momento de desarrollar un modelo es:

Seleccione una:

- a. Entrenar el modelo muchas veces hasta que el sesgo se elimine por completo
- b. Lamentablemente no se puede hacer nada.
- c. Entrenar el modelo con datos inclusivos, es decir, con diversidad étnica, racial, de género, etc

[Quitar mi elección](#)

Módulo 2: Los Datos como potenciadores de nuestro Negocio

Corfo – Pymes en Línea

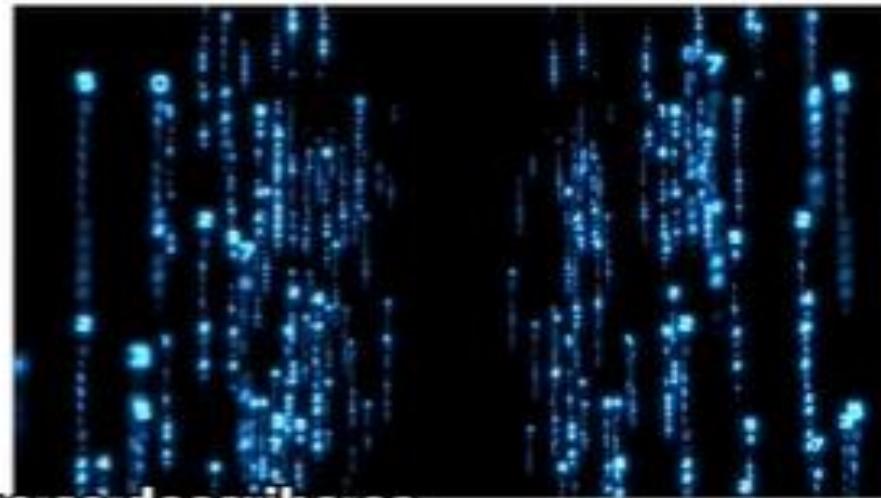
Clase 1: ¿qué son los datos y para qué sirven?

¿Qué es un dato?



Es una representación simbólica de un atributo o variable, adecuada para comunicación, interpretación o procesamiento.

- ▶ Numérica - ventas totales de una empresa.
- ▶ Alfabética - reclamo a su proveedor de internet.
- ▶ Espacial - coordenadas de una micro.
- ▶ Describen un hecho empírico, pero la forma en que se describe es definida por alguien y por lo tanto puede no medir lo que creemos.



pero la forma en que se describen es definida por alguien

Clase 1: ¿qué son los datos y para qué sirven?

¿Qué es un dato?



Es una representación simbólica de un atributo o variable, adecuada para comunicación, interpretación o procesamiento.

► Índice de Gini

Es un número que representa cómo están distribuidos los ingresos o riqueza en la población.

Se usa normalmente para referirse a qué tan desigual es un país.



Un número más bajo del índice representa menor desigualdad mientras que un país

Clase 1: ¿qué son los datos y para qué sirven?

¿Qué es un dato?



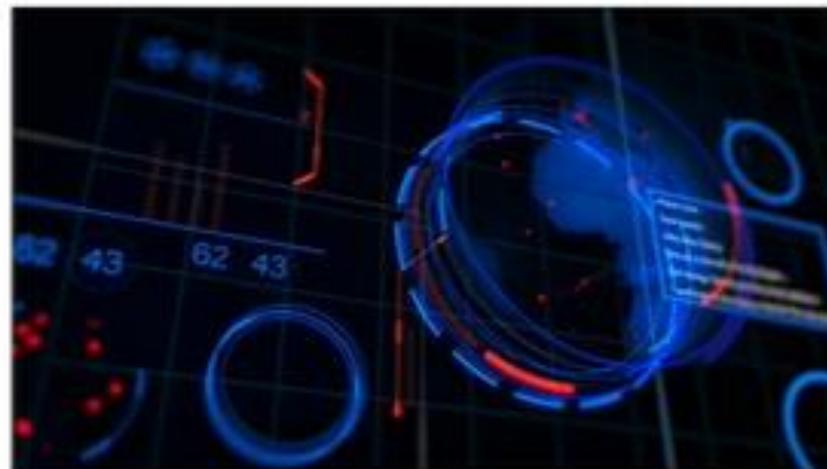
■ Es una representación simbólica de un atributo o variable, adecuada para comunicación, interpretación o procesamiento.

► Índice de Gini

Es un número que representa cómo están distribuidos los ingresos o riqueza en la población.

Se usa normalmente para referirse a qué tan desigual es un país.

Este índice solo mide una dimensión específica de desigualdad.



o erróneas en nuestros análisis.

Clase 1: ¿qué son los datos y para qué sirven?

¿Qué es un dato?



■ Es una representación simbólica de un atributo o variable, adecuada para comunicación, interpretación o procesamiento.

► Índice de Gini

Pensar que el dato es algo cerrado y que con eso se acaba una discusión no es cierto, porque los datos pueden ser interpretados de múltiples maneras.

Lo que verdaderamente representa el dato es fundamental de entenderlo para realizar cualquier tipo de análisis.



Teniendo muy claro el punto anterior, podemos proceder a utilizar los datos.

Clase 1: ¿qué son los datos y para qué sirven?

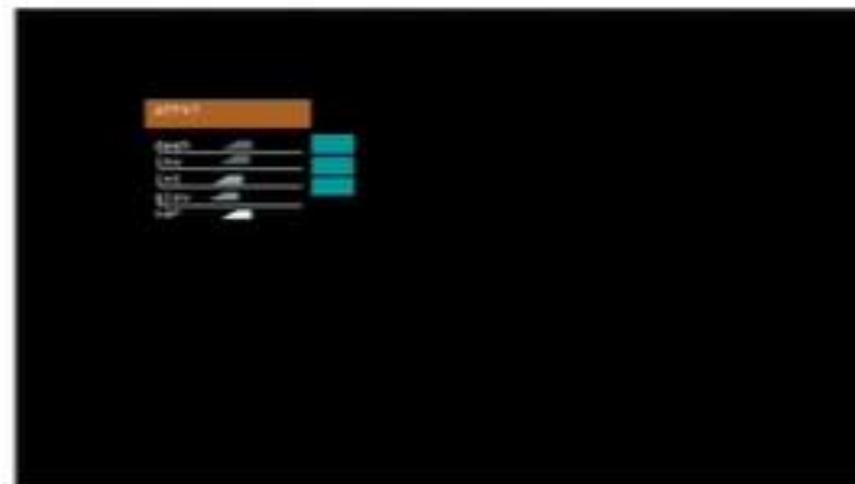
¿Qué es un dato?



Es una representación simbólica de un atributo o variable, adecuada para comunicación, interpretación o procesamiento.

¿Pero, para que podrian servirnos?

- ▶ Parte de un proceso de entendimiento de un problema.
- ▶ Validar una hipótesis.
- ▶ Revelar tendencias y responder preguntas.



y responder preguntas.

Clase 1: ¿qué son los datos y para qué sirven?

Interactuando con datos



La IA puede hacer bastante, pero depende de múltiples elecciones personales.

- ▶ Escoger preguntas efectivamente ayuden a su negocio.
- ▶ Definir qué datos les ayudaría a responder esa pregunta, y recolectarlos.
- ▶ La IA puede hacer lo suyo, durante el análisis.



de todo el proceso para efectivamente potenciar el negocio.



Clase 2: Recolectando datos

Donde brilla la IA



Una IA puede utilizar los datos de muchos estados iniciales A y finales B.

- ▶ Es necesario que tanto los estados como los pasos se encuentren de alguna manera que sea consumible por un computador.
- ▶ Cómo obtener estos datos que sean aprovechables por un computador es crucial en definir si un problema podrá ser abordado por técnicas de IA o no.

```
count).each(function () {
    Counter: 64 bytes from 200.221.2.45: icmp_seq=8 ttl=242 time=74.533 ms
    duration: 64 bytes from 200.221.2.45: icmp_seq=9 ttl=242 time=77.785 ms
    easing: 64 bytes from 200.221.2.45: icmp_seq=10 ttl=242 time=67.369 ms
    step: fun 64 bytes from 200.221.2.45: icmp_seq=11 ttl=242 time=77.392 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=12 ttl=242 time=67.621 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=13 ttl=242 time=68.003 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=14 ttl=242 time=67.498 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=15 ttl=242 time=69.942 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=16 ttl=242 time=74.601 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=17 ttl=242 time=69.950 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=18 ttl=242 time=77.663 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=19 ttl=242 time=81.284 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=20 ttl=242 time=68.732 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=21 ttl=242 time=73.623 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=22 ttl=242 time=81.238 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=23 ttl=242 time=72.561 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=24 ttl=242 time=76.342 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=25 ttl=242 time=73.734 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=26 ttl=242 time=76.610 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=27 ttl=242 time=68.900 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=28 ttl=242 time=70.284 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=29 ttl=242 time=65.869 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=30 ttl=242 time=72.669 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=31 ttl=242 time=75.447 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=32 ttl=242 time=87.613 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=33 ttl=242 time=78.126 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=34 ttl=242 time=74.493 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=35 ttl=242 time=75.365 ms
    (th) 64 bytes from 200.221.2.45: icmp_seq=36 ttl=242 time=72.686 ms
```

crucial en definir si un problema podrá ser abordado por técnicas de Inteligencia Artificial o no.



Clase 2: Recolectando datos

Recolectando datos



Recolección de datos: **manual**

- ▶ Muchas de las bases de datos famosas relacionados con imágenes fueron inicialmente creadas manualmente.
- ▶ Generar modelos que reconozcan letras escritas a mano, y poder traducirlas a una base de datos digital.

OCR (Reconocimiento Óptico de Caracteres)

0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7
0 7 1 1 4 9 4 3 4 8 2 2 1 8 7 0 8 1 0 7

y poder traducirlas a una base de datos digital,
es posible gracias al etiquetado manual



Clase 2: Recolectando datos

Recolectando datos



Aprovechar los datos que generan durante el funcionamiento de su negocio: **bases fuentes.**

- ▶ Herramienta especializada para la creación de sus productos.

Maquinaria pesada.

Computadores.

Hornos.

- ▶ Un monitoreo de estos recursos puede proveer información útil.



por ejemplo, anticiparse a posibles fallas que podrían poner en riesgo la continuidad del negocio.



Clase 2: Recolectando datos

Recolectando datos



Recolección de datos: **automática**.

► **Encuesta CASEN.**

Encuesta de Caracterización Socioeconómica Nacional

Conocer dónde están los barrios que podrían estar más interesados en su producto.

Información sobre estilos de vida.

► Tener en cuenta las condiciones de uso de los datos provistos de estas fuentes abiertas.



desarrollar herramientas que después se quieran vender.

Clase 2: Recolectando datos

Recolectando datos



Recolección de datos: **automática**.

► Kaggle

Es una plataforma de competencias donde plantean problemas a ser resueltos con IA.

Se tenía que competir por quien desarrollaba el mejor modelo de recomendación de contenido.



y tener una mejor idea si es que un problema puede ser resuelto con Inteligencia Artificial.



Clase 2: Recolectando datos

Recolectando datos



Manual

Observar

Automático

Sensores y Arduinos

- ▶ Con este tipo de herramientas se puede armar un artefacto que te permita capturar temperatura y humedad de algún lugar que necesite mantenerse.

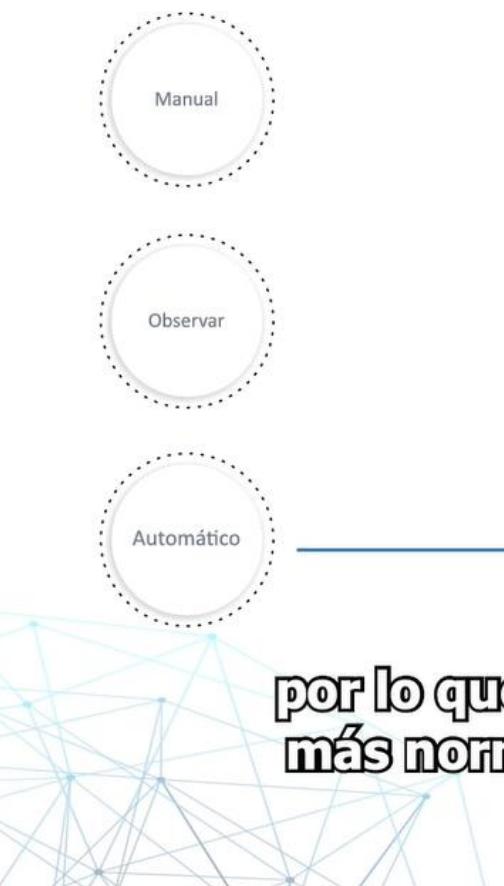


Esta información podría ser inmediatamente guardada en una base de datos para posterior análisis.



Clase 2: Recolectando datos

Recolectando datos



Scraping

- ▶ Normalmente necesita un esfuerzo grande en ordenamiento, limpieza y estructuración de la información.



por lo que vale la pena profundizar en las dificultades más normales que la mayoría de los datos presentan.



Clase 3: Estructuras de datos

Diferencias de datos en su origen



Recordatorio primer video ¿qué es un dato?

- ▶ Los datos estructurados son aquellos que se pueden acomodar fácilmente en una estructura tabular.
- ▶ Las hacen deseables para el análisis.

El diagrama muestra tres tablas estructuradas:

- User**: Una tabla con columnas UserID, User, Address, Phone, Email y Alternate. Contiene 4 filas de datos.
- Order**: Una tabla con columnas OrderID y UserID. Contiene 4 filas de datos.
- Lineitem**: Una tabla con columnas OrderID, ProductID y Quantity. Contiene 4 filas de datos.

Las tablas están conectadas por relaciones: la tabla Order tiene una relación con la tabla User basada en la columna UserID; la tabla Order también tiene una relación con la tabla Lineitem basada en la columna OrderID.

User					
UserID	User	Address	Phone	Email	Alternate
1	Alice	123 Foo St.	12345678	alice@example.org	alice@neo4j.org
2	Bob	456 Bar Ave.		bob@example.org	
***	***	***	***	***	***
99	Zach	99 South St.		zach@example.org	

Order	
OrderID	UserID
1234	1
5678	1
***	***
5588	99

Lineitem		
OrderID	ProductID	Quantity
1234	765	2
1234	987	1
***	***	***
5588	765	1

La relación entre tablas será profundizada en el Módulo 3 al hablar de las bases de datos



Clase 3: Estructuras de datos

Datos estructurados



**Es nuevamente necesario, conocimiento experto
del tema y habilidad técnica,**

• **Imputables**

• **Consultables**

• **Manipulables**

- ▶ Transformar la base de distintas formas.
- ▶ Crear nuevas variables.
- ▶ Eliminar.
- ▶ Crear nuevas relaciones con otras tablas.



Clase 3: Estructuras de datos

Datos no estructurados

- Son aquellos donde estas operaciones no resultan fáciles o simplemente no se pueden realizar.
 - ▶ **Manipular** este tipo de datos es también complejo



también es complejo y en ocasiones pierde el sentido.

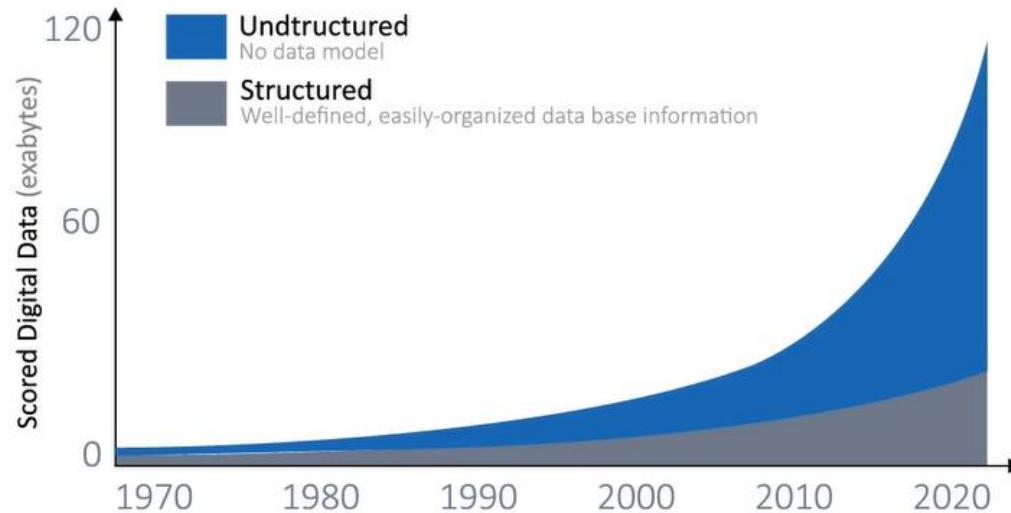


Clase 3: Estructuras de datos

Evolución de los datos



- ▶ La cantidad de datos no estructurados ha aumentado considerablemente más a mayor velocidad que los estructurados.
- ▶ ¿Qué tipo de datos resulta más natural usar para este tipo de problemas?



de problemas de este tipo. Sin embargo, mucha de la expectación y asombro por lo que puede hacer

Clase 4: Cadena de valor, primera parte

Utilidades de datos



- ▶ Herramienta en un proceso de entendimiento.
 - ▶ Validación de hipótesis.
 - ▶ Revelar tendencias o insights.

Análisis:

Es normalmente el paso que se lleva toda la atención.

Es la cadena de valor información-acción lo que hace posible aprovechar al máximo los datos.

```
        },each(function () {
            _this.drop('counter', _this.animate({
                counter: 64 bytes from 200.221.2.45: icmp_seq=7 ttl=242 time=84.587 ms
                64 bytes from 200.221.2.45: icmp_seq=8 ttl=242 time=74.533 ms
                64 bytes from 200.221.2.45: icmp_seq=9 ttl=242 time=77.785 ms
                64 bytes from 200.221.2.45: icmp_seq=10 ttl=242 time=67.369 ms
                easing: 64 bytes from 200.221.2.45: icmp_seq=11 ttl=242 time=77.392 ms
                step: 64 bytes from 200.221.2.45: icmp_seq=12 ttl=242 time=67.621 ms
                (in) 64 bytes from 200.221.2.45: icmp_seq=13 ttl=242 time=68.003 ms
                (in) 64 bytes from 200.221.2.45: icmp_seq=14 ttl=242 time=74.601 ms
                (in) 64 bytes from 200.221.2.45: icmp_seq=15 ttl=242 time=69.942 ms
                64 bytes from 200.221.2.45: icmp_seq=16 ttl=242 time=77.498 ms
                64 bytes from 200.221.2.45: icmp_seq=17 ttl=242 time=69.950 ms
                64 bytes from 200.221.2.45: icmp_seq=18 ttl=242 time=77.663 ms
                64 bytes from 200.221.2.45: icmp_seq=19 ttl=242 time=81.284 ms
                64 bytes from 200.221.2.45: icmp_seq=20 ttl=242 time=68.732 ms
                64 bytes from 200.221.2.45: icmp_seq=21 ttl=242 time=73.623 ms
                64 bytes from 200.221.2.45: icmp_seq=22 ttl=242 time=81.238 ms
                64 bytes from 200.221.2.45: icmp_seq=23 ttl=242 time=72.561 ms
                64 bytes from 200.221.2.45: icmp_seq=24 ttl=242 time=76.342 ms
                64 bytes from 200.221.2.45: icmp_seq=25 ttl=242 time=73.734 ms
                64 bytes from 200.221.2.45: icmp_seq=26 ttl=242 time=76.610 ms
                64 bytes from 200.221.2.45: icmp_seq=27 ttl=242 time=68.980 ms
                64 bytes from 200.221.2.45: icmp_seq=28 ttl=242 time=70.284 ms
                64 bytes from 200.221.2.45: icmp_seq=29 ttl=242 time=65.869 ms
                64 bytes from 200.221.2.45: icmp_seq=30 ttl=242 time=72.669 ms
                64 bytes from 200.221.2.45: icmp_seq=31 ttl=242 time=75.447 ms
                64 bytes from 200.221.2.45: icmp_seq=32 ttl=242 time=87.613 ms
                64 bytes from 200.221.2.45: icmp_seq=33 ttl=242 time=78.126 ms
                64 bytes from 200.221.2.45: icmp_seq=34 ttl=242 time=74.495 ms
                64 bytes from 200.221.2.45: icmp_seq=35 ttl=242 time=75.365 ms
                64 bytes from 200.221.2.45: icmp_seq=36 ttl=242 time=72.686 ms
            }));
        });
    });
});
```

En definitiva, es la cadena de valor información-acción lo que hace posible aprovechar al máximo los datos.



Clase 4: Cadena de valor, primera parte

Presiona **Esc** para salir de la pantalla completa



Cadena de valor información-acción

- Es una secuencia de actividades donde cada una de ellas agrega un valor relevante al proceso.
- Previo al análisis hay cuatro pasos:
 - ▶ Evento.
 - ▶ Sistema de captura de datos.
 - ▶ Almacenamiento.
 - ▶ Extracción.



**definición del evento, sistema de captura,
almacenamiento y extracción.**



Clase 4: Cadena de valor, primera parte

Cadena de valor información-acción

- Es una secuencia de actividades donde cada una de ellas agrega un valor relevante al proceso.
- Análisis:
 - ▶ Comunicar.
 - ▶ Tomar acción.



Estos pasos se verán con más detalle en la siguiente clase.



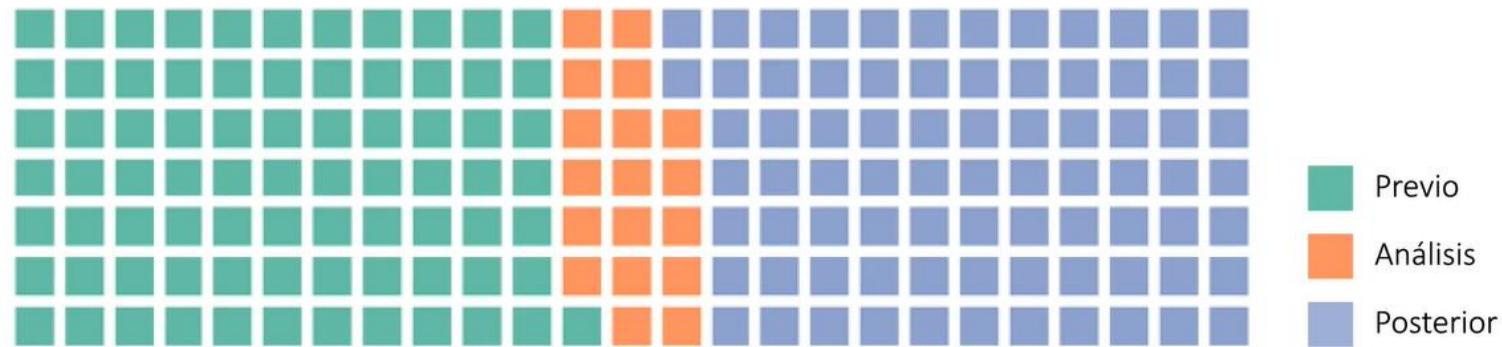
Clase 4: Cadena de valor, primera parte

Distribución de tiempo en proyectos AI

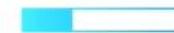


El análisis suele ser lo menos costoso en tiempo.

Es necesario darle la importancia que tienen los pasos previos para poder verdaderamente aprovechar la potencia de la IA.



Es necesario darle la importancia que tienen los pasos previos para poder verdaderamente aprovechar



Clase 4: Cadena de valor, primera parte

Cadena de valor



Evento



Puede dar inmediatamente una señal de que tan difícil será llevar a cabo el proyecto



Sistema de captura de datos



Comprar



Usar mascarilla



Evento climático



Almacenamiento

¿Es posible capturar estos eventos?



Extracción

¿Por qué quiero capturar este evento en específico?

Hay que tener en este punto, al menos una noción de lo que se quiere lograr:



03:33



Clase 4: Cadena de valor, primera parte



Cadena de valor



• ¿Se tienen registro automático de las compras?

¿Hay cámaras monitoreando un área?



Manual



Observando



Automática

**manual, observando y automática.
En ocasiones un registro manual bastará,**

Clase 4: Cadena de valor, primera parte



Cadena de valor



Clase 4: Cadena de valor, primera parte



Cadena de valor



- Normalmente utilizaremos un subconjunto de los datos disponibles para trabajar.
Conviene extraer solamente la información necesaria para el análisis desde la fuente de almacenamiento.



Una vez completados estos pasos, estamos en condiciones de verdaderamente sacar todo el valor a los datos.

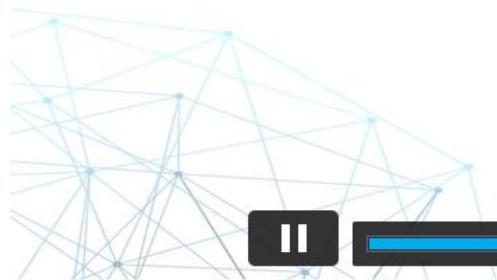
estaremos en condiciones de verdaderamente sacar todo el valor a los datos.

Clase 4: Cadena de valor, primera parte



Reflexión

- Hay que tener en mente el contexto de la recolección de datos y como serán utilizados.
- La inteligencia artificial lamentablemente replica los sesgos que existen en los datos.
- Tener muy en consideración en no estar incurriendo en sesgos al momento de construir las muestras.



de la inteligencia artificial.



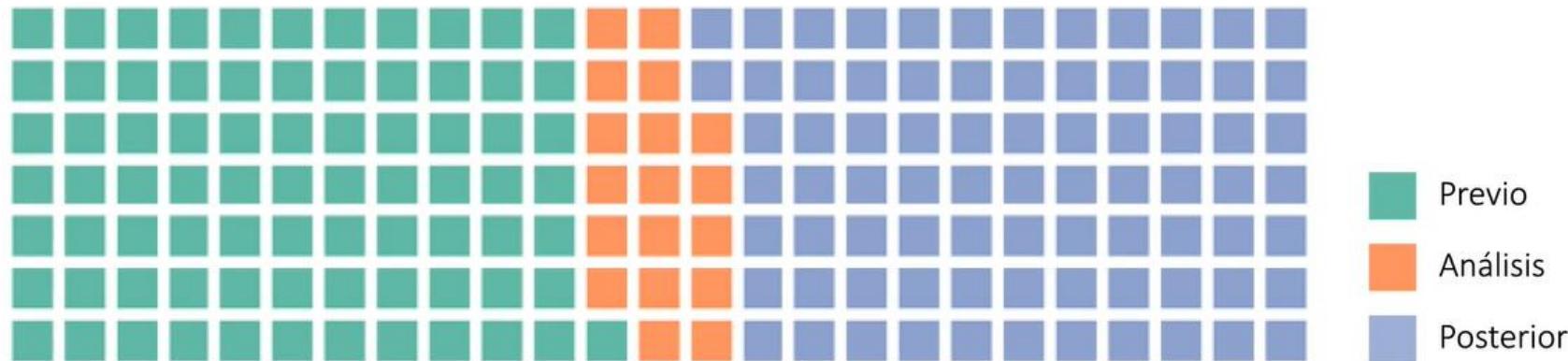
Clase 5: Cadena de valor, segunda parte

Distribución de tiempo en proyectos AI



El análisis suele ser lo menos costoso en tiempo.

Mucho esfuerzo debe ir en las definiciones iniciales al igual que el esfuerzo comunicacional para lograr la acción que se planea tomar.



para lograr la acción que se planea tomar. No obstante, el análisis requiere un conocimiento de las



Clase 5: Cadena de valor, segunda parte



Análisis

• Descriptivo

- ▶ Es una manera de resumir los datos de forma de tener una visión global de ellos.



**una visión global de ellos, para ganar intuición
en aspectos claves de su composición.**



Clase 5: Cadena de valor, segunda parte

Presiona **Esc** para salir de la pantalla completa



Análisis

Descriptivo

- ▶ Las distribuciones de las distintas variables que tenemos son siempre una primera aproximación necesaria.
- ▶ ¿Dónde están concentrados los datos?
- ▶ ¿Se encuentran similarmente distribuidos?
- ▶ No se puede pasar por alto una diferencia de proporción grande en grupos claves.

es aquí donde no se puede pasar por alto una diferencia de proporción gigante en grupos claves,



02:05



Clase 5: Cadena de valor, segunda parte

Análisis

Predictivo

- ▶ Este es el tipo de análisis que normalmente se asocia con la IA.
- ▶ Hay modelos que sólo pueden ser utilizados si es que no existen datos missing, mientras que otros necesitan cumplir algunos supuestos para que el modelo estimado resulte válido.



mientras que otros necesitan cumplir algunos supuestos para que el modelo estimado resulte válido.



Clase 5: Cadena de valor, segunda parte

Análisis

• Prescriptivo

- ▶ Puede ser realizado a partir de generación de reglas surgido por medio de un análisis descriptivo.
Indicar a que grupo de clientes llamar para ofrecer un producto a partir de analizar su consumo.
- ▶ Puede surgir a partir de lo que arroja el análisis predictivo.
A partir de un modelo de fuga, saber que clientes son más probables dejen de utilizar nuestro servicio.



nuestro servicio y por lo tanto realizar acciones enfocadas en ellos.



Clase 5: Cadena de valor, segunda parte

Comunicar



- Rara vez el equipo encargado del análisis serán los encargados de dar el visto bueno para su utilización.
- Comunicar los resultados de manera de que quede claro cómo la compañía se verá beneficiada de utilizar el análisis realizado.
- Las herramientas que más los ayudará en este punto es la visualización.

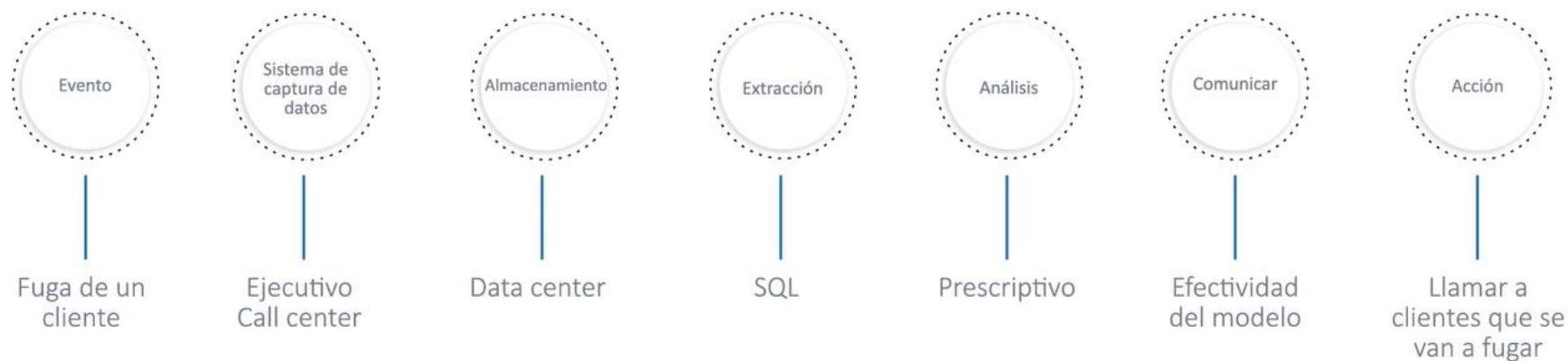


Eso sí, una de las herramientas que más los ayudará en este punto es la visualización,



Clase 5: Cadena de valor, segunda parte

Ejemplo



La acción finalmente consiste en llamarlos para intentar retenerlos.



Clase 6: Visualización de datos

Relevancia de visualización



"Visualizando información, la convertimos en un paisaje que podemos explorar con nuestros ojos. Una suerte de mapa de información. Cuando estás perdido en información, un mapa es muy útil".

David McCandless

**El célebre, en todo lo relacionado con visualización,
David McCandless dice:**



Clase 6: Visualización de datos

Relevancia de visualización



- ▶ Resalta la capacidad de la visualización para navegar dentro de un exceso de información.
- ▶ Le permite al analista encontrar un rumbo dentro del caos y posteriormente comunicar.



Esto le permite al analista encontrar un rumbo dentro del caos y posteriormente será útil para comunicar



Clase 6: Visualización de datos

Relevancia de visualización



- ▶ ¿Podemos rápidamente ver cuál fila tiene el valor máximo de la primera variable?
- ▶ ¿Podemos apreciar si existe alguna relación entre las dos primeras variables?
- ▶ ¿Están concentradas las variables o se reparten homogéneamente por sus rangos?

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.10	3.50	1.40	0.20	setosa
2	4.90	3.00	1.40	0.20	setosa
3	4.70	3.20	1.30	0.20	setosa
4	4.60	3.10	1.50	0.20	setosa
5	5.00	3.60	1.40	0.20	setosa
6	5.40	3.90	1.70	0.40	setosa

Las personas no somos buenas mirando números organizados como bases de datos y resumiendo



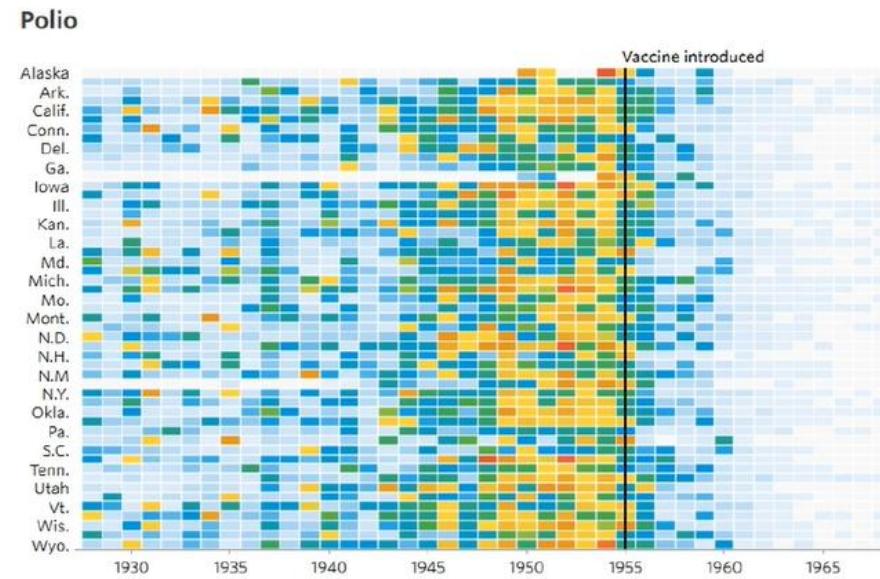
Clase 6: Visualización de datos

Relevancia de visualización



Sirve para comunicar conceptos

- ▶ Ejemplo: efectividad de las vacunas
- ▶ Las vacunas han tenido un impacto innegable.
- ▶ ¿Por qué resulta tan efectivo este método?



Es casi instantáneo apreciar cómo el momento de la introducción de la vacuna genera un gran impacto.



Clase 6: Visualización de datos

Relevancia de visualización



- ▶ Nuestra evolución dotó a nuestro cerebro con una función ultra veloz de procesamiento por motivos de supervivencia.
- ▶ Esta capacidad hace que podamos transmitir ideas a nuestro público incluso antes de que lo noten, realizando visualizaciones que dirijan la atención con color, diferencia en tamaños y posicionamiento.



Esta capacidad de dirigir nuestra atención a diferencias, hace que podamos transmitir ideas



Clase 6: Visualización de datos

Relevancia de visualización



- ▶ Predecir comportamiento



nos puede dar inmediatamente una idea del nivel de carros que tendrían que estar en ruta



Clase 6: Visualización de datos

Storytelling



Cada vez más relevante, porque se ha ido haciendo cada vez más evidente que los hechos y los datos no bastan.



"Somos, como especie, adictos a las historias. Incluso cuando dormimos la mente se mantiene despierta contándose historias".

Jonathan Gottschall

la mente se mantiene despierta contándose historias".

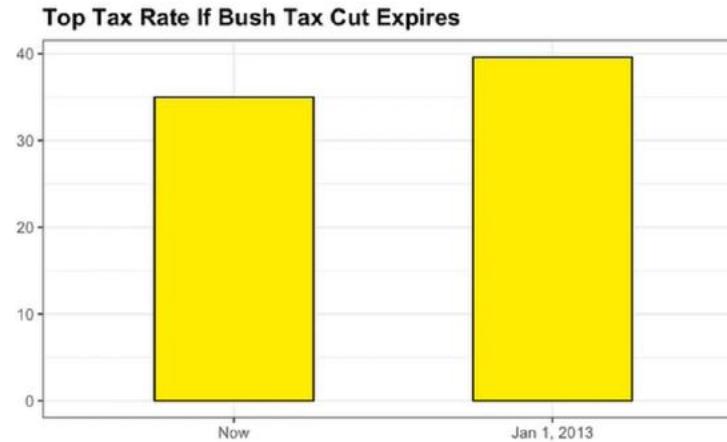
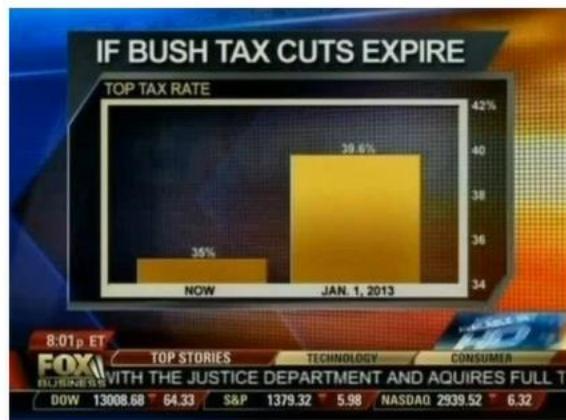


Clase 6: Visualización de datos

Reflexión



Contando una historia distorsionada



la diferencia no parece muy grande ¿Tiene sentido entonces, presentar la información de esa manera?



Clase 7: Entendimiento del Big data

¿Qué es?



► Definición formal:

Es un paradigma de generación de valor a partir de conjuntos de datos con tres características

Gran cantidad de datos donde sólo una máquina es insuficiente.

Datos frecuentemente actualizados, en ocasiones en tiempo real.

Numerosos y distintos tipos de datos.

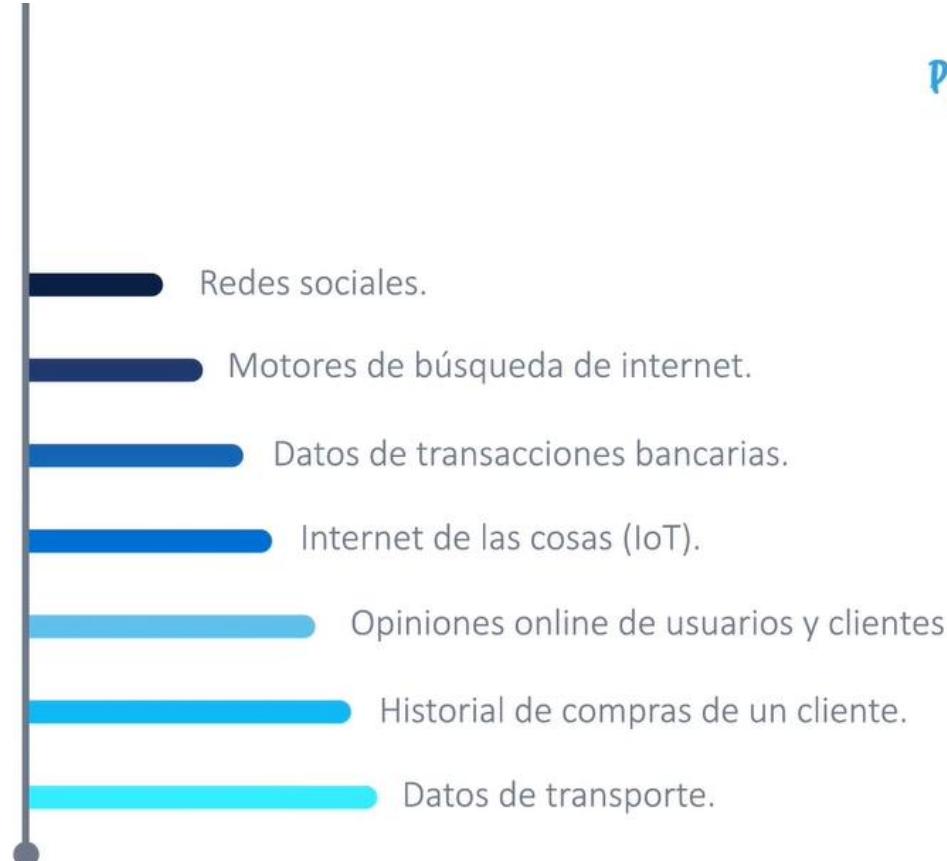


Numerosos y distintos tipos de datos.



Clase 7: Entendimiento del Big data

**¿Dónde lo
encontramos?**



Datos de transporte, tanto público como privado.



Clase 7: Entendimiento del Big data

¿Dónde lo encontramos?

Volumen

Velocidad

Variedad



Podemos resumir las características de los problemas de Big Data en las tres "V": Volumen, Velocidad y Variedad.



Clase 7: Entendimiento del Big data

Las tres Vs



Volumen



“There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days”

Eric Schmidt

de información que la humanidad generó hasta el 2003, unos 5 Exabytes, se generan actualmente cada 2 días.



Clase 7: Entendimiento del Big data

Las tres Vs



Volumen

- ▶ Un PC doméstico estándar contaba con un disco duro de 10 GB en 2000.
- ▶ Hoy se pueden encontrar en el mercado discos duros 1000 veces más grandes.
- ▶ En 2018 se generaron 2.5 millones de TB cada día.



Se confirmaron las palabras de Eric Schmidt, generándose en 2018 2.5 millones de TB, 2.5 EB cada día.



Clase 7: Entendimiento del Big data

Las tres Vs



Velocidad

- ▶ En 2016, cada día se conectaban más de 5 millones de nuevos dispositivos para capturar, analizar y compartir datos.
- ▶ La mejora en la capacidad de transmisión de datos de los dispositivos conectados continuará incrementando la velocidad de generación de los datos.
- ▶ Es uno de los principales retos actuales, el procesamiento y análisis de nuevos datos masivos en cortos espacios de tiempo.



Es uno de los principales retos actuales, el procesamiento y análisis de nuevos datos masivos



Clase 7: Entendimiento del Big data

Las tres Vs



Variedad

- ▶ Hay multitud de diferentes tipos de información además de la textual o numérica, como archivos multimedia, texto no estructurado, etc.
- ▶ Esto puede ser un problema ya que las bases de datos relacionales fueron diseñadas para almacenar datos estructurados.



Y esto puede ser un problema ya que las bases de datos relacionales fueron diseñadas para almacenar



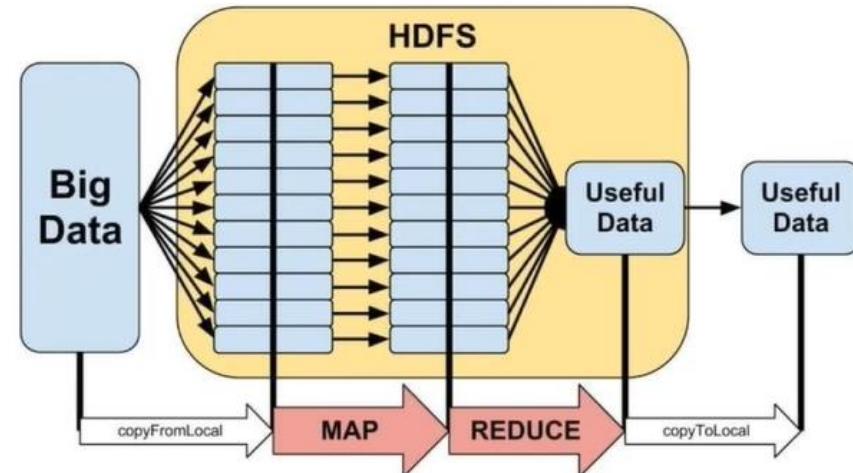
Clase 7: Entendimiento del Big data

Nuevas herramientas



Almacenamiento: Hadoop Distributed File System (HDFS)

- ▶ Es un sistema de ficheros en el que los datos se dividen en bloques que se almacenan de forma distribuida.
- ▶ Replicar los bloques para evitar errores o para facilitar su procesamiento.



**Podemos replicar los bloques para evitar errores
o para facilitar su procesamiento.**

Clase 7: Entendimiento del Big data

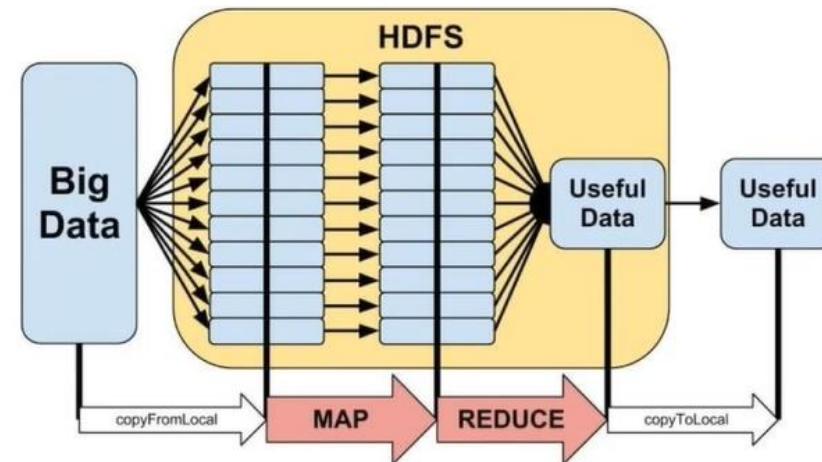
Nuevas herramientas

Presiona **Esc** para salir de la pantalla completa



Procesamiento: MapReduce

- ▶ Es un conjunto de algoritmos diseñados para el procesamiento de datos masivos.
- ▶ Los datos se procesan allí donde están almacenados.
- ▶ Finalmente se agregan los cálculos parciales para obtener el resultado global.



Finalmente se agregan los cálculos parciales para obtener el resultado global, parte "reduce".



Clase 7: Entendimiento del Big data

Estrategia Big Data

- Índice de madurez del modelo de negocio de Big Data

- ▶ Business Monitoring.
- ▶ Business Insights.
- ▶ Business Optimization.
- ▶ Data Monetization.
- ▶ Business Metamorphosis.



según el cumplimiento de los principios, para que las empresas sean conscientes en qué punto están.



Clase 7: Entendimiento del Big data

Índice de madurez del modelo de negocio de Big Data



Presentan una jerarquía

- ▶ Business Monitoring. → Mide la capacidad de la empresa de provechar el almacenamiento de datos y la inteligencia de negocio para monitorizar el rendimiento de la empresa.
 - ¿Cuál es nuestro producto más rentable?
 - ¿Qué tipo de clientes están comprando ese producto?
- ▶ Business Insights.
- ▶ Business Optimization.
- ▶ Data Monetization.
- ▶ Business Metamorphosis.

“¿Qué tipo de clientes están comprando ese producto?”



Clase 7: Entendimiento del Big data

Índice de madurez del modelo de negocio de Big Data



Presentan una jerarquía

- ▶ Business Monitoring.
- ▶ Business Insights.
- ▶ Business Optimization.
- ▶ Data Monetization.
- ▶ Business Metamorphosis.

Se aprovechan los datos para realizar análisis predictivos.

Podemos aplicar los métodos propios de la Ciencia de Datos.

- Compras.
- Visitas web.
- Opiniones.
- Valoraciones.

que generan nuestros clientes con sus compras,
visitas web, opiniones, valoraciones, etc.



Clase 7: Entendimiento del Big data

Índice de madurez del modelo de negocio de Big Data



Presentan una jerarquía

- ▶ Business Monitoring.
- ▶ Business Insights.
- ▶ Business Optimization.
- ▶ Data Monetization.
- ▶ Business Metamorphosis.

● Accionamos los insights identificados en el nivel anterior para lograr el objetivo concreto que tengamos en mente.

En el tercer nivel accionamos los insights identificados en el nivel anterior para lograr el objetivo concreto



Clase 7: Entendimiento del Big data

Índice de madurez del modelo de negocio de Big Data



Presentan una jerarquía

- ▶ Business Monitoring.
 - ▶ Business Insights.
 - ▶ Business Optimization.
 - ▶ Data Monetization.
 - ▶ Business Metamorphosis.
- Identificar nuevas fuentes de ingresos a partir de datos usualmente no explotados.
Ideas de reempaquetado para crear productos y servicios completamente nuevos.



que ayudan a las organizaciones a entrar en nuevos mercados y apuntar a nuevos clientes o audiencias.



Clase 7: Entendimiento del Big data

Índice de madurez del modelo de negocio de Big Data



Presentan una jerarquía

- ▶ Business Monitoring.
 - ▶ Business Insights.
 - ▶ Business Optimization.
 - ▶ Data Monetization.
 - ▶ Business Metamorphosis.
- Retailers entrando en el área de Optimización de compras mediante la recomendación de productos.
 - Venta de servicios en lugar de solo producto.
 - Ampliación de cartera de clientes entrando en nuevos nichos de mercado.

Ampliación de cartera de clientes entrando en
nuevos nichos de mercado.



Clase 7: Entendimiento del Big data

Big Data: una nueva forma de pensar



- No pienses en qué tecnología utilizar, sino en cómo va a transformar esa tecnología tu negocio.
- Céntrate en cómo crear y capturar valor.
- Piensa en análisis predictivos o prescriptivos sobre qué pasará y cómo podremos aprovecharlo para nuestro negocio.



**sino en análisis predictivos o prescriptivos sobre
qué pasará y cómo podremos aprovecharlo**

2.7. Entendiendo el concepto de Big data



Material adicional

Storytelling with data: (largo)

www.youtube.com/watch?v=8EMW7io4rSI

Automatización de proyectos con arduinos (interesante)

https://www.youtube.com/watch?v=O_Q1WKCtWiA

Big data <https://www.youtube.com/watch?v=eVSfJhssXUA>

Curso sugerido:

<https://www.coursera.org/learn/datavisualization>

Módulo 3: Manejo de bases de datos

Corfo – Pymes en Línea

Clase 1: Entendiendo las bases de datos

Base de datos



Conjunto de información almacenado y consultado sistemáticamente:

► Sistématicamente:

No sirve almacenar los datos de manera aleatoria o desordenada.

Se necesita un sistema con reglas definidas de como almacenar la información y como consultar esta información.



Para eso se necesita un sistema con reglas definidas de cómo almacenar la información y cómo consultar esta información.

Clase 1: Entendiendo las bases de datos

Base de datos



Base de datos

Corresponde a la información
que tenemos almacenada.

Sistema gestor
de base de datos

Es el software que nos permite
realizar toda la gestión de la base
de datos.

**mientras que el sistema gestor de bases de datos es el
software que nos permite realizar toda la gestión**



Clase 1: Entendiendo las bases de datos

Base de datos



- ▶ La llegada de las computadoras ayuda mucho a que las bases de datos sean más eficientes.



**porque para hacer una consulta sobre la base de datos
utilizamos los computadores y los algoritmos.**

04:24



Clase 1: Entendiendo las bases de datos

Base de datos



- ▶ La llegada de las computadoras ayuda mucho a que las bases de datos sean más eficientes.

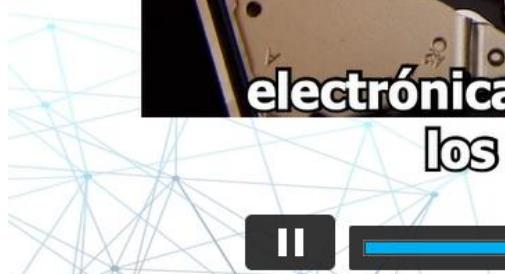


**electrónicamente en un sistema computacional, es decir,
los datos se encuentran en formato digital.**

Discos duros

Son de acceso mucho más rápido.

La mayoría de bases de datos la información se encuentra almacenada electrónicamente en un sistema computacional.



04:59



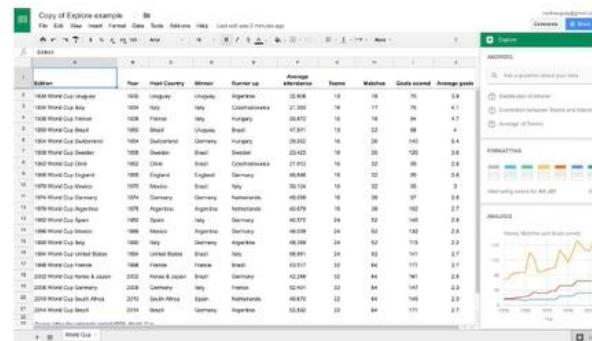
Clase 1: Entendiendo las bases de datos



Diferencias entre base de datos y hoja de cálculo

Hojas de cálculo

- ▶ Se diseñaron originalmente para un usuario.
- ▶ Que no necesitan manipular una gran cantidad de datos complicados.



manipular una gran cantidad de datos complicados.

Clase 1: Entendiendo las bases de datos



Diferencias entre base de datos y hoja de cálculo

Base de datos

- ▶ Están diseñadas para contener colecciones mucho más grandes de información organizada.



mucho más grandes de información organizada,
cantidades masivas en ocasiones.



Clase 1: Entendiendo las bases de datos

Tipos de base de datos



La mejor base de datos para una organización específica va a depender de varios factores, pero principalmente de cómo la organización pretende utilizar los datos.

► Bases de datos relacionales

Son bases de datos cuyos elementos se organizan como un conjunto de tablas con filas de registros y columnas de atributos y con campos clave que permiten relacionar las diferentes tablas.



La tecnología de base de datos relacional proporciona la manera más eficiente y flexible de acceder



Clase 1: Entendiendo las bases de datos



Tipos de base de datos

La mejor base de datos para una organización específica va a depender de varios factores, pero principalmente de cómo la organización pretende utilizar los datos.

► Bases de datos NoSQL

También conocidas como bases de datos no relacionales.

Permiten que los datos no estructurados y semiestructurados se almacenen y manipulen.



permiten que los datos no estructurados y semiestructurados se almacenen y manipulen.

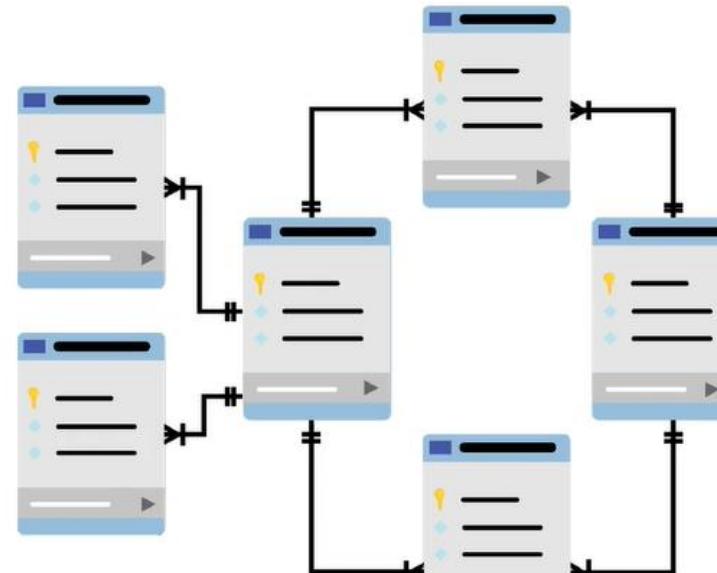


Clase 2: Entendiendo las bases de datos relacionales

Base de datos relacional



- Se basa en el modelo relacional, una forma intuitiva y directa de representar datos en tablas.
- Cada fila de la tabla es un registro con un ID único llamado clave.
- Las columnas de la tabla contienen atributos de los datos.
- Una celda corresponde a la intersección de una fila y una columna, por lo que una celda siempre corresponderá a un único atributo asociado a un único registro.



En particular, una celda corresponde a la intersección de una fila y una columna,

Clase 2: Entendiendo las bases de datos relacionales

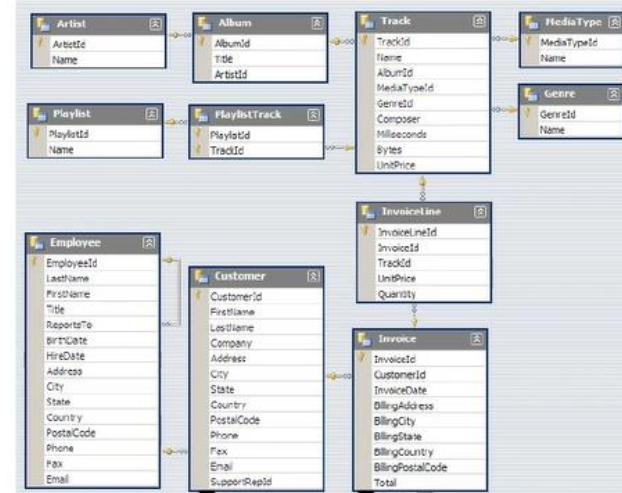
Base de datos relacional



No repetir textos

Ejemplo: base de datos que permite reproducir música

- ▶ Para bases de datos pequeñas no es ningún problema.
- ▶ Realizar consultas sobre campos de tipo texto es muy ineficiente, sobre todo para bases grandes.
- ▶ Es mejor asociar cada texto a un número y trabajar con los campos numéricos.



**entonces en la tabla de canciones, en lugar de tener
repetido muchas veces el texto "Queen",**



Clase 2: Entendiendo las bases de datos relacionales

El modelo relational

- En los primeros años de las bases de datos, cada aplicación almacenaba datos en su propia estructura única.
- Se diseñó, en gran parte, para resolver este problema, de tener varias estructuras de datos arbitrarias.
- Proporcionó una forma estándar de representar y consultar datos que cualquier aplicación podría utilizar.
- Su principal fortaleza está en el uso de tablas.



que cualquier aplicación podría utilizar, siendo la principal fortaleza del modelo de base de datos



Clase 2: Entendiendo las bases de datos relacionales

Base de datos relacional

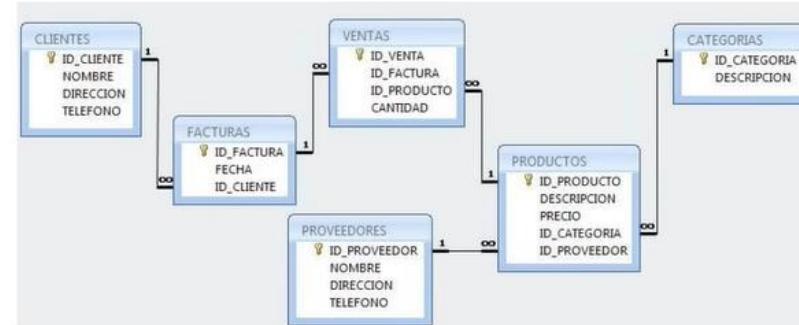


La base de datos se organiza en 2 secciones:

► Esquema

Es la definición de la estructura de la base de datos y principalmente almacena los siguientes datos:

- El nombre de cada tabla.
- El nombre de cada columna.
- El tipo de dato de cada columna.
- La tabla a la que pertenece cada columna.



y la tabla a la que pertenece cada columna.

3.2. Entendiendo las bases de datos relacionales



Clase 2: Entendiendo las bases de datos relacionales

Base de datos relacional



La base de datos se organiza en 2 secciones:

- ▶ Esquema

Hoja de cálculo de Excel:

Podemos perfectamente dentro de esa columna llenar una celda con un texto o una fecha, lo cual claramente es una incoherencia.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	GASTOS MENSUALES 2019																									
2	Alquiler	700,00																								
3	Electricidad	75,00																								
4	Gas	80,00																								
5	Agua, Saneamiento y Otros	120,00																								
6	Otros	200																								
7	Mantenimiento	50,00																								
8	Impresiones	10,00																								
9	Total mes	1200,00																								
10	Ingresos	1400,00																								
11	Monto	200,00																								

Es decir, al trabajar con Excel se pierde esa rigurosidad en los datos que nos entrega una base de datos relacional.



Clase 2: Entendiendo las bases de datos relacionales

Sistemas gestores de bases de datos relacionales

- Existe un lenguaje llamado SQL, del inglés structure query language, o lenguaje de consultas estructurado.
- El lenguaje SQL es estándar para todos estos sistemas, aun cuando cada uno tiene sus funciones propias.
- Los sistemas gestores de bases de datos relacionales más utilizados son:



Los sistemas gestores de bases de datos relacionales más utilizados son Oracle, MySQL, SQL Server, PostgreSQL y SQLite.

Clase 3: Bases de datos no relacionadas

¿Por qué NoSQL?



- La gran mayoría de información que generamos hoy en día se trata de información no estructurada.
- Las bases de datos relacionales que ya hemos visto no son capaces de manejar de forma eficiente este tipo de información.



eficiente este tipo de información, por lo que se necesitaba un cambio de paradigma.



Clase 3: Bases de datos no relacionadas

¿Por qué NoSQL?



- Sólo escalan fácilmente de forma vertical, lo que supone un alto coste de las máquinas que utilizamos para almacenar la información.
- No permitían de forma sencilla una replicación efectiva de los datos y distribución en varios nodos.
- No permiten definir diferentes estructuras en un mismo registro y cualquier cambio de diseño implica un alto coste.

de diseño implica un alto coste, al necesitarse modificar la propia base de datos y los programas que la alimentan y explotan.

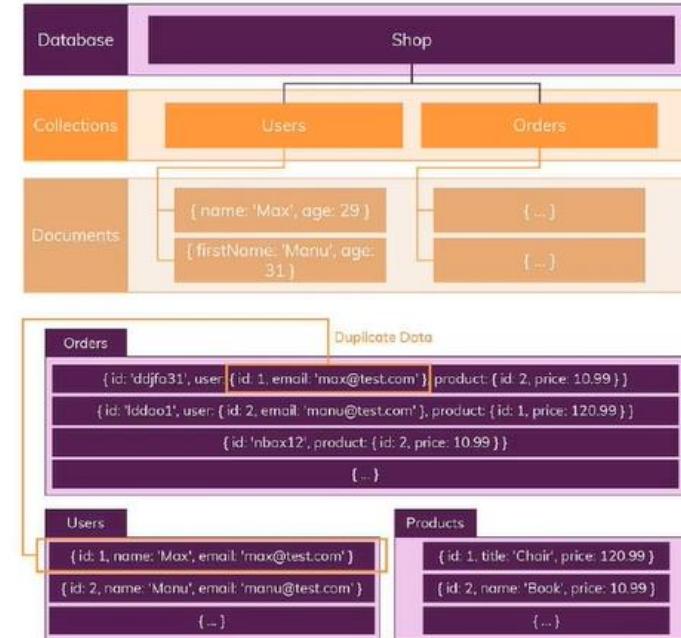


Clase 3: Bases de datos no relacionadas

Introducción NoSQL



- ▶ Se compone de colecciones (a diferencia de las tablas de SQL), que contienen a su vez documentos (en lugar de filas).
- ▶ Estos documentos contienen por sí mismos toda la información necesaria.
- ▶ Sin esquema: documentos de la misma colección pueden tener diferentes esquemas.
- ▶ No existen (o son muy pocas) las relaciones entre colecciones. Esto provoca que puede haber cierta redundancia o inconsistencia en los datos.



**Esto provoca que puede haber cierta redundancia
o inconsistencia en los datos.**



Clase 3: Bases de datos no relacionadas

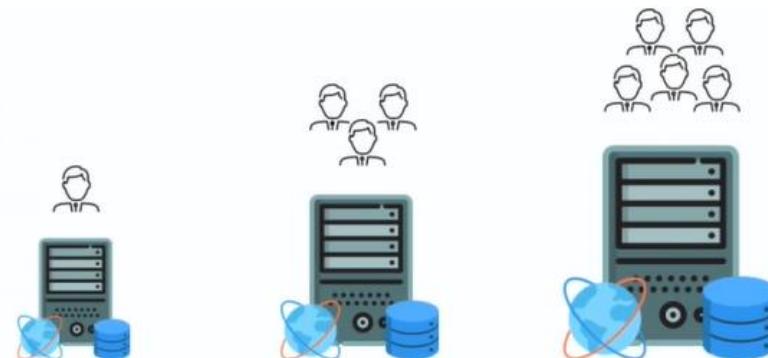
Escalabilidad



La capacidad del sistema de ser ampliado cuando las necesidades de los datos así lo requieren.

► La escalabilidad vertical

- La más clásica.
- Mejora el sistema global a base de mejorar las máquinas actuales.
- Solución sencilla.
- En contrapartida, a partir de un cierto punto los componentes se pueden hacer excesivamente caros y por tanto no hacer factible esta escalabilidad.



caros y por tanto no hacer factible esta escalabilidad.

Clase 3: Bases de datos no relacionadas

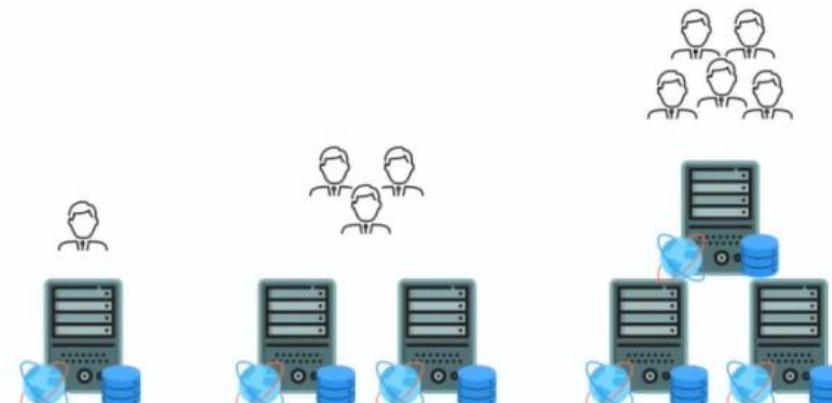
Escalabilidad



La capacidad del sistema de ser ampliado cuando las necesidades de los datos así lo requieren.

► La escalabilidad horizontal

- Mejoramos el sistema a base de incorporar más máquinas pequeñas.
- Solución barata.
- Su contrapunto es la necesidad de diseñar un sistema que gestione la multitud de nodos que podemos tener.



Su contrapunto es la necesidad de diseñar un sistema que gestione la multitud de nodos que podemos tener.

Clase 3: Bases de datos no relacionadas

SQL vs. NoSQL



SQL

► Ventajas

- Altísimo rendimiento con datos estructurados, múltiples relaciones y transacciones.
- Cuentan con las propiedades ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad)
- Tecnología madura y confiable, con múltiples herramientas.

► Desventajas

- Dificultad en escalar el sistema: vertical muy caro, horizontal muy difícil.
- Existencia de un esquema prefijado.
- Si tenemos una multitud de consultas simultáneas, el rendimiento del sistema puede caer.



el rendimiento del sistema puede caer debido a la complejidad de los cruces entre tablas y su alto coste computacional.



Clase 3: Bases de datos no relacionadas

SQL vs. NoSQL



NoSQL

► Ventajas

- Perfectas para almacenar y disponibilizar información no relacional.
- La inexistencia de esquema permite cambios en la base de datos.
- Son muy fáciles de escalar.
- Gran rendimiento para lecturas y escrituras masivas (y simples).

► Desventajas

- Tiempo de respuesta más lento para grandes conjuntos de datos.
- Algunas escrituras, modificaciones generalmente, pueden ser lentas en caso de información duplicada.
- No permite la flexibilidad en la manipulación de datos que sí proporciona SQL.
- Tecnologías relativamente novedosas y que por tanto están aún en período de maduración.

Finalmente, estamos hablando de tecnologías relativamente novedosas y que por tanto están aún en período de maduración.



Clase 3: Bases de datos no relacionadas



Nuevo concepto: Agregación

Agregación:

- ▶ Es un conjunto de datos que toma información de diferentes tablas.
- ▶ Es relativamente sencillo convertir una base de datos SQL a una NoSQL.



los documentos de forma distribuida en clusters
o grupos de máquinas.



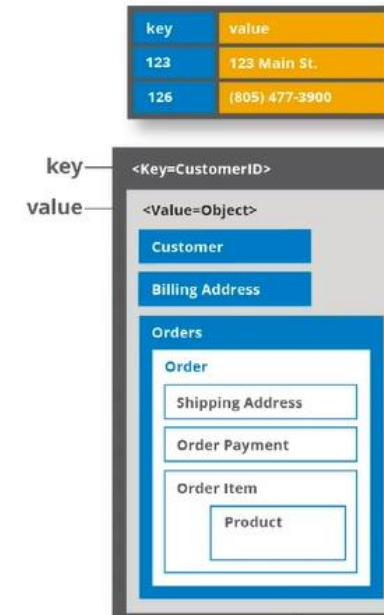
Clase 3: Bases de datos no relacionadas

Modelos de datos



Clave-Valor

- ▶ Son un paradigma preparado para trabajar con arrays asociativos.
- ▶ Es un conjunto de parejas clave-valor, donde cada valor puede ser a su vez otro conjunto de parejas clave-valor.



donde cada valor puede ser a su vez otro conjunto de parejas clave-valor.



Clase 3: Bases de datos no relacionadas

Modelos de datos



Columna

- ▶ Utilizan tablas, filas y columnas como las relacionales.
- ▶ Los nombres de los campos y sus formatos pueden ser diferentes entre sus filas.
- ▶ La columna es la mínima instancia de información y cada fila puede contener varias columnas.

UserProfile				
Bob	emailAddress	gender	age	
	bob@example.com	male	35	
	1465676582	1465676582	1465676582	
Britney	emailAddress	gender		
	brit@example.com	female		
	1465676432	1465676432		
Tori	emailAddress	country	hairColor	
	tor@example.com	Sweden	Blu	
	1435626158	1435636158	1465633654	

El acceso a la información se realiza por columnas, por lo que las agregaciones por éstas son muy rápidas.



Clase 3: Bases de datos no relacionadas

Modelos de datos



Documentos

- ▶ Son similares a las parejas clave-valor, dado que contienen información semiestructurada.
- ▶ Tanto la clave como el valor pueden ser utilizados para acceder al documento y podemos buscar en ellos.



Barack Obama

@BarackObama

...

Tomorrow: [@HouseOfCards](#). No spoilers, please.

[Traducir Tweet](#)

8:33 p. m. · 13 feb. 2014 · Twitter Web Client

34,9 mil Retweets 85 Tweets citados 22,7 mil Me gusta



Pero, a diferencia de aquellos, tanto la clave como el valor pueden ser utilizados para acceder al documento



Clase 3: Bases de datos no relacionadas

Modelos de datos

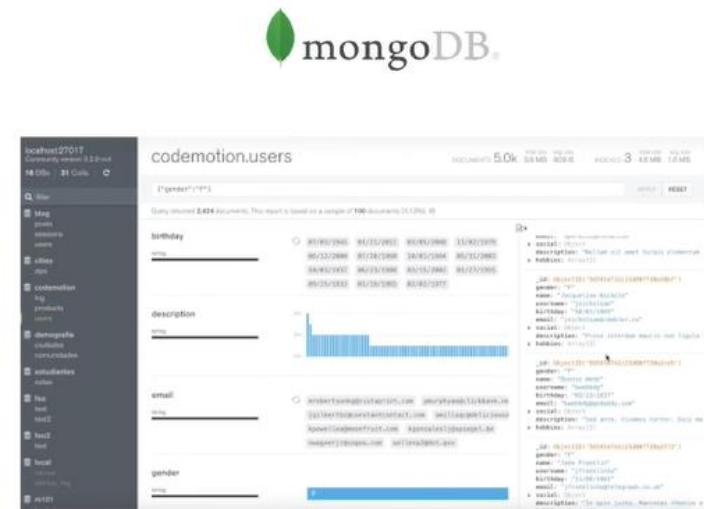
Presiona **Esc** para salir de la pantalla completa



Documentos

Ejemplo

- ▶ En ella hemos realizado una query o consulta (gender es F), obteniendo la lista de documentos que satisfacen la consulta.



una query o consulta, gender es F, es decir que el género es femenino, obteniendo la lista de documentos que satisfacen

Clase 3: Bases de datos no relacionadas

Modelos de datos



Gráficos

- ▶ Almacenan información donde su principal valor radica en la relación entre las entidades que consideramos.
- ▶ Se utilizan nodos para almacenar las entidades que queremos manejar y aristas para identificar qué entidades están relacionadas entre sí.



y aristas para identificar qué entidades están relacionadas entre sí.

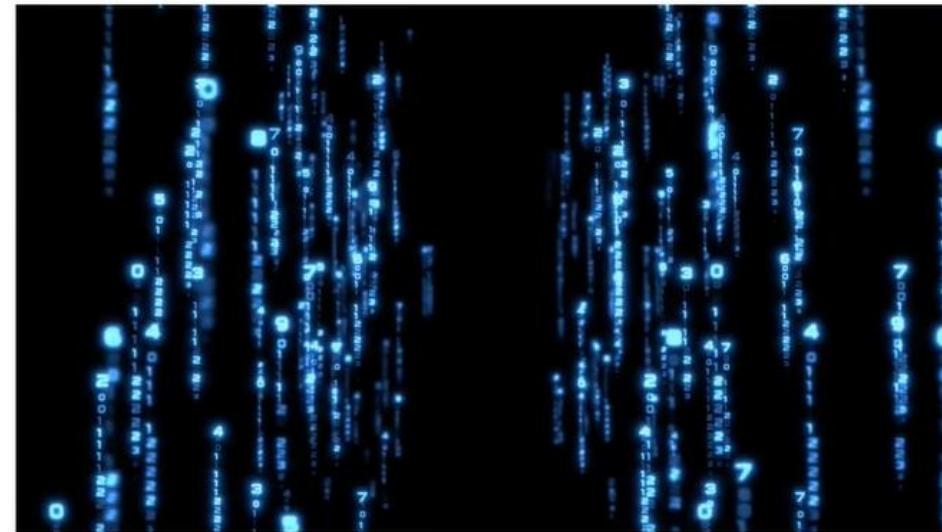


Clase 3: Bases de datos no relacionadas

Conclusión



- ▶ Nació para intentar solucionar los problemas que presentan las bases de datos SQL
- ▶ Poco flexibles y escalables.



Pero aún hoy en día, las bases de datos SQL siguen siendo muy útiles e imprescindibles dependiendo el entorno.



Clase 4: Lenguaje SQL

¿Qué es SQL?



SQL: Lenguaje de Consultas Estructurado
(Structured Query Language)

Diseñado para administrar y recuperar
información de sistemas de gestión de bases
de datos relacionales.



3.3. Introducción a SQL

de bases de datos relacionales.



Clase 4: Lenguaje SQL

¿Para qué sirve SQL?



- ▶ Consultas de datos.
- ▶ Inserción de datos.
- ▶ Actualizaciones de datos.
- ▶ Borrado de datos.
- ▶ Creación y modificación de esquemas.
- ▶ Control de acceso a los datos.



Y control de acceso a los datos, es decir, distintos usuarios de la base de datos tendrán diferentes permisos.



Clase 4: Lenguaje SQL

Operaciones CRUD

C — Create.
R — Read.
U — Update.
D —

Update, para actualizar registros y Delete,
para borrar registros.



3.3. Introducción a SQL



Clase 4: Lenguaje SQL

Operaciones CRUD

Palabras clave

- ▶ Insert.
- ▶ Select.
- ▶ Update.
- ▶ Delete.



que son insert, select, update y delete respectivamente.

3.3. Introducción a SQL



Clase 4: Lenguaje SQL

Operaciones CRUD

Palabras clave

- ▶ Insert.
- ▶ Select.
- ▶ Update.
- ▶ Delete.



que son insert, select, update y delete respectivamente.

3.3. Introducción a SQL



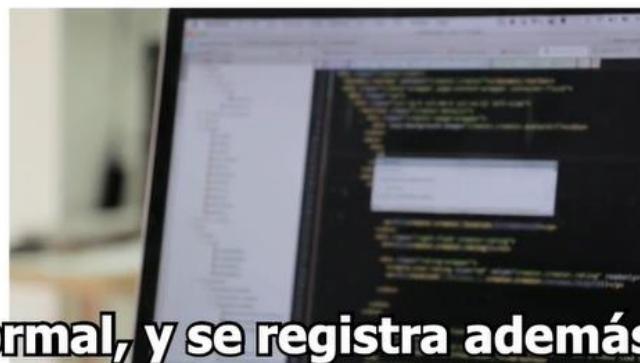
Clase 4: Lenguaje SQL

Operaciones CRUD



Create (Crear registros): INSERT

```
INSERT INTO Clientes (Nombre, Apellido, Comuna, Email, Numero_cel)  
VALUES ('Andrea', 'Fernandez', 'Quinta Normal', 'AndreaF23@gmail.com',  
95723411);
```



de la comuna de quinta normal, y se registra además su correo electrónico y número de celular.



02:47



Clase 4: Lenguaje SQL

Operaciones CRUD



Read (Leer registros): SELECT

```
SELECT * FROM Clientes  
WHERE Comuna = 'Puente Alto' OR Comuna = 'La Florida';
```

```
mysql> SELECT * FROM Clientes  
WHERE Comuna = 'Puente Alto'  
OR Comuna = 'La Florida';  
  
+----+-----+-----+-----+  
| id | nombre | apellido | comuna |  
+----+-----+-----+-----+  
| 1  | Juan   | Pineda  | La Florida |  
| 2  | Maria  | Hernandez | Puente Alto |  
| 3  | Pedro  | Gómez    | La Florida |  
| 4  | Ana    | Pérez    | Puente Alto |  
| 5  | Carlos | Sánchez  | La Florida |  
| 6  | David  | Gómez    | Puente Alto |  
| 7  | Sofia  | Pérez    | La Florida |  
| 8  | Luis   | Gómez    | La Florida |  
| 9  | Camila | Pérez    | Puente Alto |  
| 10 | Daniel  | Sánchez  | La Florida |  
| 11 | Paula  | Gómez    | Puente Alto |  
| 12 | Ignacio| Pérez    | La Florida |  
| 13 | Sandra  | Sánchez  | La Florida |  
| 14 | Felipe  | Gómez    | Puente Alto |  
| 15 | Valentina| Pérez    | La Florida |  
| 16 | Santiago| Sánchez  | La Florida |  
| 17 | Paula  | Gómez    | Puente Alto |  
| 18 | Daniel  | Pérez    | La Florida |  
| 19 | Paula  | Sánchez  | La Florida |  
| 20 | Daniel  | Gómez    | Puente Alto |  
+----+-----+-----+-----+
```

en que queremos todos los clientes que sean de
Puente Alto o La Florida.



Clase 4: Lenguaje SQL



Operaciones CRUD

Update (Actualizar registros): UPDATE

```
UPDATE Clientes
SET Email = 's_manu28@hotmail.com', Number_cel = 90548221
WHERE ID_cliente = 37;
```

```
... 3c1 d0b d2b
    b74 - 7c0 e8b b2f 3f7 3c4 4b5 d5e bch >>> >>> c2e f6f
4d0 5b4 d1e - 6a6 b5c e7e
b2c - 6f7 c2f
5a5 63d 2f2 a5e a2d f7f - 5b1 b7b b47 5d0 a2f d4c
2f2 f0b b8e 605 b5c 1b5 1f9 f8d b6d -
d4f 2d6 e9f 1c6 f3d - b1d >>>
e1c 6b6 f7e - b8d d6e b1b 8e1 e5b 5e3
>> 6e8 a0a d5d a8d 9d0 - 2e5 9f2 f0d f8c b4e 7e8
1d8 11d 7c1 >>> c4a b85 e5f 7a8 - 7a1 c6d 8c7 b85 a8e
f7d c3b f5d b7b
c8e b8e 3c2 1d8 - d2c a8a d8c - f5c f5b 3f5
c6e b5e a7f 4b8 a9b 6ca f4a d8a c8a c1a d0f b83 6b6
3d2 4f7
```

de esta forma, estamos actualizando el correo electrónico y número de celular de este cliente.

3.3. Introducción a SQL

Clase 4: Lenguaje SQL

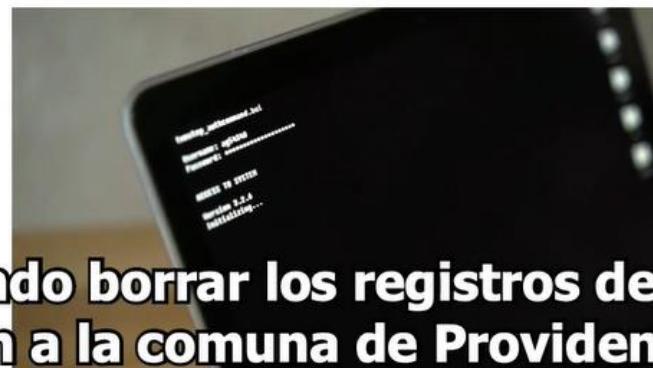


Operaciones CRUD

Delete (Borrar registros): DELETE

```
DELETE FROM Clientes
```

```
WHERE Comuna = 'Providencia';
```



En este caso, estamos pidiendo borrar los registros de la tabla clientes que pertenecen a la comuna de Providencia.

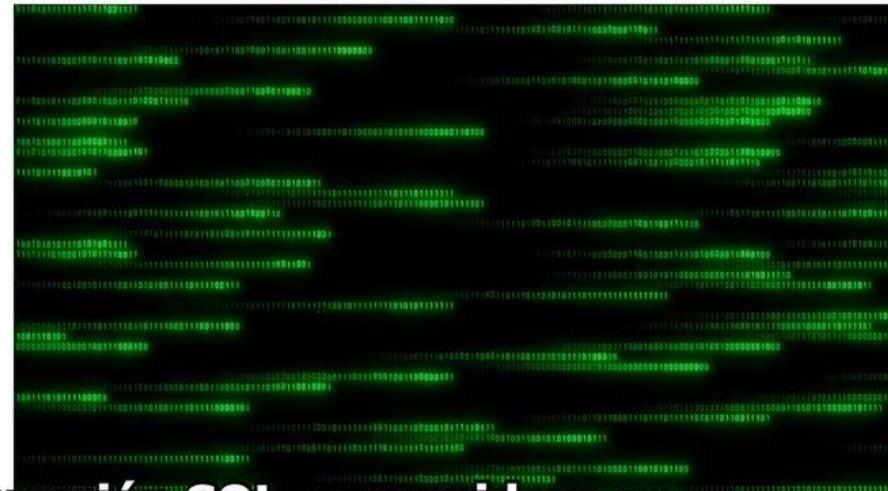
Clase 4: Lenguaje SQL

Inyección SQL



Consiste en la infiltración de código intruso para realizar operaciones sobre una base de datos.

- ▶ Se llama así porque lo que hace es “inyectar” código SQL invasor dentro del código SQL programado.
- ▶ Este código intruso en general se realiza con malas intenciones, por ejemplo para borrar o robar información.
- ▶ Un problema de seguridad informática.



Es por esto que la inyección SQL se considera un problema de seguridad informática.



Clase 4: Lenguaje SQL

Inyección SQL



Ejemplo:

- ▶ Código SQL: `"SELECT * FROM usuarios WHERE nombre = '" + nombreUsuario + "';"`
- ▶ Nombre de usuario malicioso: `'Felipe'; DROP TABLE usuarios; SELECT * FROM datos WHERE nombre LIKE '%`
- ▶ Consulta SQL generada: `SELECT * FROM usuarios WHERE nombre = 'Felipe';
DROP TABLE usuarios;
SELECT * FROM datos WHERE nombre LIKE '%';`



esta es una forma ingeniosa de que el código termine con eso y que permita seleccionar todos los datos.

Clase 4: Lenguaje SQL

Inyección SQL

Presiona **Esc** para salir de la pantalla completa



El pequeño Bobby tablas



se borró la información de todos los estudiantes del colegio.

3.3. Introducción a SQL

Clase 4: Lenguaje SQL



Otras consultas SQL

Ordenar registros: ORDER BY

- ▶ Ordenar una tabla.

```
SELECT * FROM Clientes  
ORDER BY Comuna, Nombre;
```

- ▶ El orden es creciente, orden alfabético.

y para registros de una misma comuna, se ordenan por nombre según orden alfabético.

3.3. Introducción a SQL



Clase 4: Lenguaje SQL

Presiona **Esc** para salir de la pantalla completa



Otras consultas SQL

Agrupar registros: GROUP BY

```
SELECT COUNT(ID_cliente), Comuna  
FROM Clientes  
GROUP BY Comuna;
```

En este ejemplo lo que hacemos es agrupar los clientes por comuna, y para cada comuna mostrar su nombre

3.3. Introducción a SQL



Clase 4: Lenguaje SQL



Otras consultas SQL

Registros distintos (únicos): DISTINCT

```
SELECT DISTINCT Comuna  
FROM Clientes;
```

tantas como clientes viven en esa comuna, pero con el
distinct se eliminan los duplicados

3.3. Introducción a SQL



Clase 4: Lenguaje SQL



Otras consultas SQL

Carácter comodín: LIKE %

- ▶ Nos permite identificar un texto cualquiera.
- ▶ La gracia de este está en usarlo en combinación con otras letras.

```
SELECT * FROM Clientes  
WHERE Comuna LIKE 'P%'
```

En este ejemplo se está usando like 'P%' es decir, que representa todas las palabras que parten con P,

3.3. Introducción a SQL



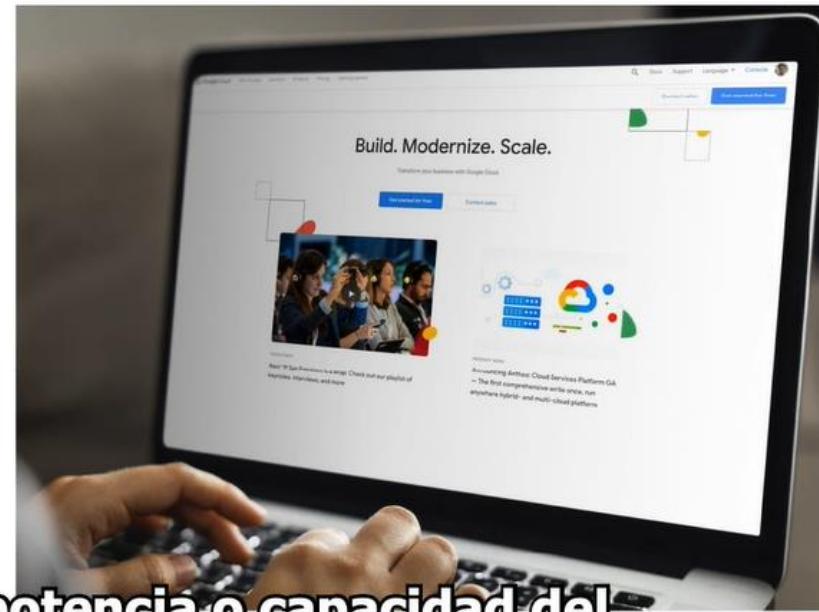
Clase 5: ¿Qué es el Cloud Computing?

¿Qué es el Cloud Computing?



Google Cloud

- ▶ Construir, Modernizar y Escalar.
- ▶ Adaptar la potencia o capacidad del sistema para responder a la demanda del mercado.



que podemos adaptar la potencia o capacidad del sistema para responder a la demanda del mercado.

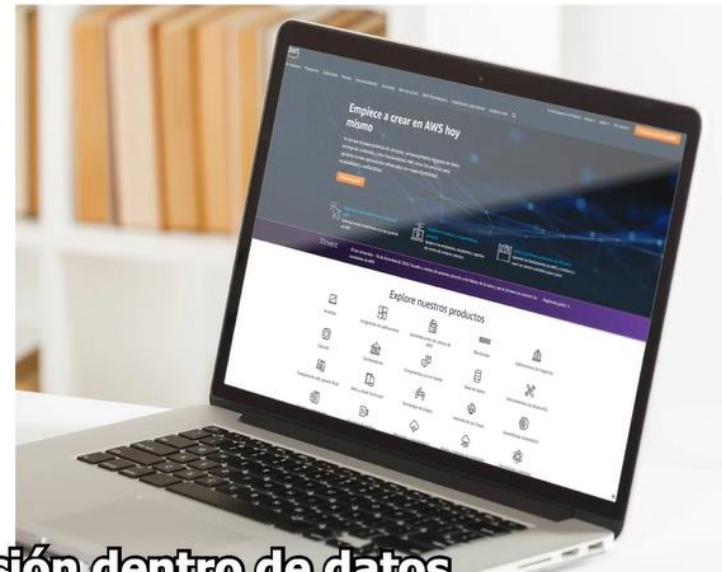
Clase 5: ¿Qué es el Cloud Computing?

¿Qué es el Cloud Computing?



Amazon - Open Distro for Elasticsearch

- ▶ Open significa abierta, gratuita en general para uso no comercial.
- ▶ “Elasticsearch”, por su nombre, hace referencia a la búsqueda ágil y elástica de información dentro de datos propios o de terceros.



y elástica de información dentro de datos propios o de terceros.



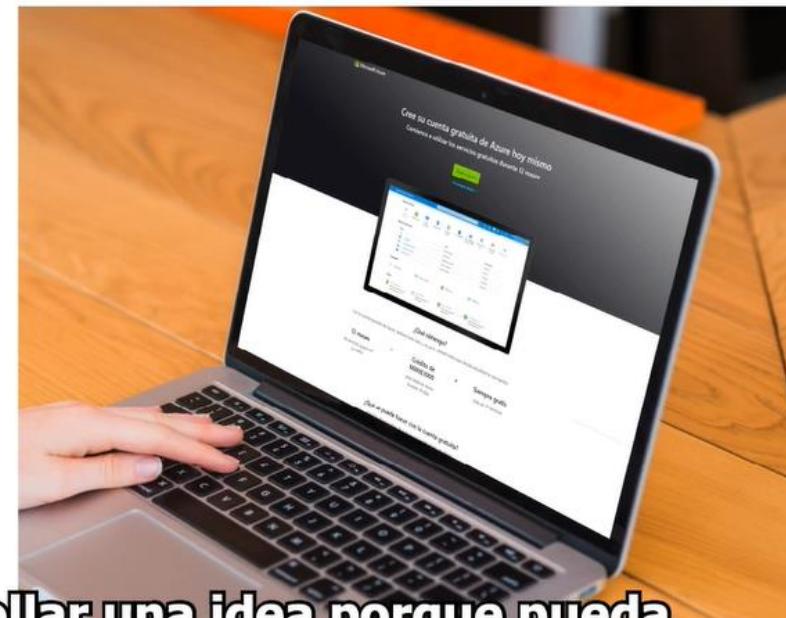
Clase 5: ¿Qué es el Cloud Computing?

¿Qué es el Cloud Computing?



Microsoft - Azure

- ▶ “Inventar con propósito”.
- ▶ “Convertir ideas en resultados”.



con el objetivo de no desarrollar una idea porque pueda ser buena, sino porque podemos convertirla en resultados.



01:27



Clase 5: ¿Qué es el Cloud Computing?

¿Qué es el Cloud Computing?



■ Es el hecho de proveer instalaciones informáticas al gran público.

- ▶ Es una cesión temporal de recursos bajo demanda.
- ▶ Recursos como procesamiento, memoria, almacenamiento.
- ▶ Dos puntos clave:
 - La cesión debe estar adaptada a cada consumidor.
 - Ser servida por proveedores.
- ▶ El acceso se realiza vía red y el uso de los recursos se hace a conveniencia y bajo demanda del usuario.



Finalmente, un punto importante es que la gestión de estos recursos es dinámica y sencilla para el usuario.



Clase 5: ¿Qué es el Cloud Computing?



Tecnologías relacionadas

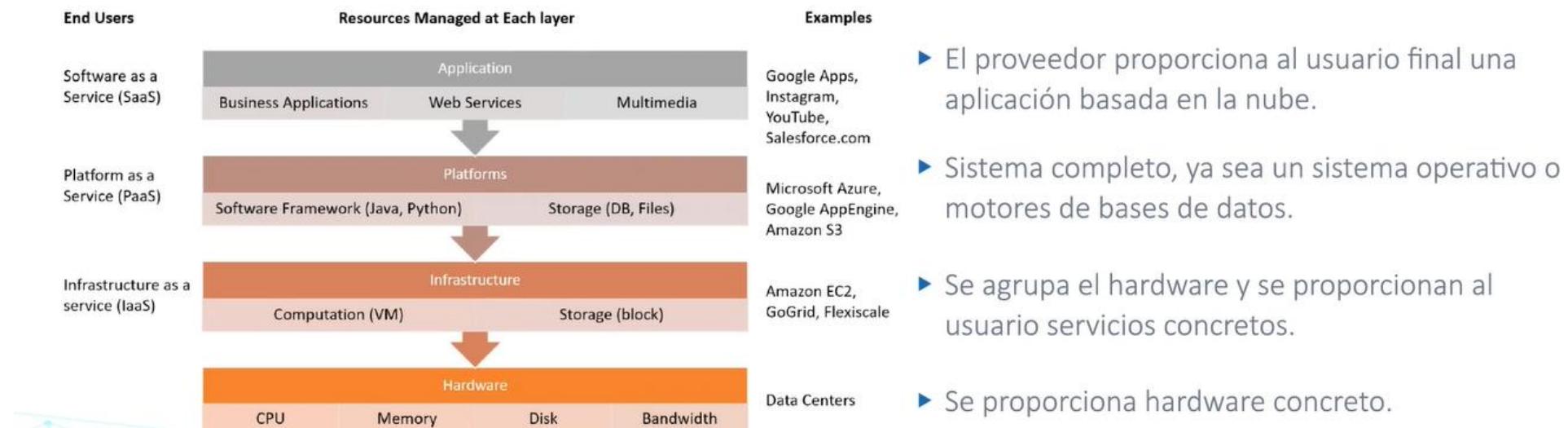
- La computación en red o en malla
Coordina los recursos en red para alcanzar un objetivo común.
- Utility computing
Proporcionar recursos y cobrar a los clientes por el uso que hacen de ellos.
- Virtualización
Tecnología que abstrae los detalles del hardware y proporciona recursos virtualizados.
- Computación autónoma
Construir sistemas informáticos capaces de autogestionarse.

capaces de autogestionarse.



Clase 5: ¿Qué es el Cloud Computing?

Modelo en capas



- ▶ El proveedor proporciona al usuario final una aplicación basada en la nube.
- ▶ Sistema completo, ya sea un sistema operativo o motores de bases de datos.
- ▶ Se agrupa el hardware y se proporcionan al usuario servicios concretos.
- ▶ Se proporciona hardware concreto.

Estamos hablando de software como servicio, con ejemplos conocidos como Google Mail o Instagram, por ejemplo.



Clase 5: ¿Qué es el Cloud Computing?

Ventajas del Cloud computing



- ▶ Utiliza un modelo de pago por uso.
- ▶ Los recursos en la nube pueden ser rápidamente asignados y liberados bajo demanda.
- ▶ Los proveedores de recursos altamente escalables agrupan una gran cantidad de recursos de los centros de datos y los hacen fácilmente accesibles.
- ▶ Los recursos alojados en la nube generalmente se basan en la web.
- ▶ Reducción de riesgos de negocio y costes de mantenimiento.



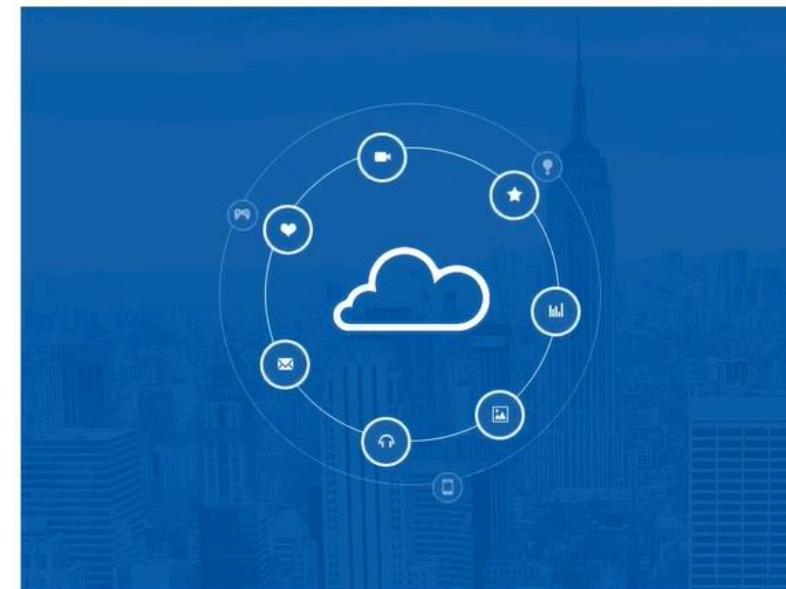
y en el costo de capacitación del personal encargado de estas tareas.



Clase 5: ¿Qué es el Cloud Computing?

Características del Cloud Computing

- ▶ Los recursos hardware no pertenecen a un único propietario.
- ▶ Pueden ser utilizados por varios clientes.
- ▶ Son accesibles a través de internet y utilizan internet como plataforma de entrega del servicio.
- ▶ Los recursos alojados en la nube generalmente se basan en la web.



dónde están localizados.



Clase 5: ¿Qué es el Cloud Computing?

Características del Cloud Computing

- ▶ Adopts a service-oriented model.
 - ▶ Computational resources can be reserved and released on the fly, according to the client's needs.
 - ▶ Automation of resource management to respond quickly to rapid changes in demand.
 - ▶ Price based on usage.



Infraestructura como Servicio.

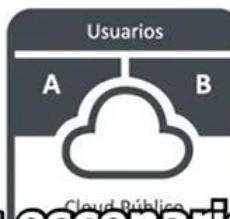
Clase 5: ¿Qué es el Cloud Computing?

Tipos de nubes



Nubes públicas:

- ▶ Los proveedores ofrecen sus servicios al público en general.
- ▶ En contrapartida, las nubes públicas carecen de un control detallado sobre dónde están los datos, características de la red y configuración de seguridad.



su eficacia en algunos escenarios empresariales.
Un ejemplo sería Google Drive o iCloud de Apple.



Clase 5: ¿Qué es el Cloud Computing?

Tipos de nubes

Nubes privadas o internas:

- ▶ Diseñadas para uso exclusivo de una sola organización.
- ▶ Ofrecen el mayor nivel de control sobre el rendimiento, la confiabilidad y la seguridad.



Muchas de las grandes empresas tecnológicas disponen de sus propias nubes privadas.



Clase 5: ¿Qué es el Cloud Computing?



Tipos de nubes

Nubes híbridas:

- ▶ Combinación de modelos de nube pública y privada.
- ▶ Más flexibilidad, ya que proporcionan un control y una seguridad más estrictos.
- ▶ Determinar cuidadosamente la mejor división entre los componentes de la nube pública y privada.



en una nube propia o de nuestro país, mientras que el proceso puede estar en cualquier lugar del mundo.

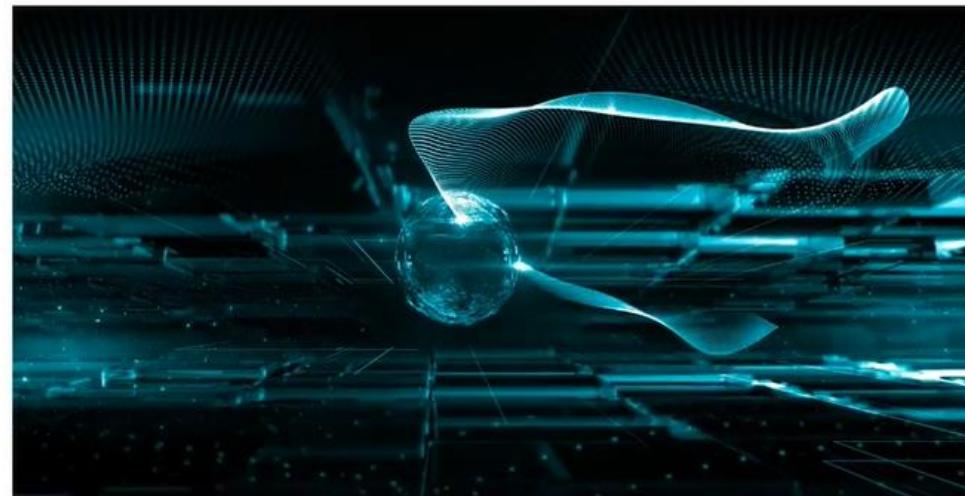


Clase 5: ¿Qué es el Cloud Computing?

Conclusión



- ▶ La computación en la nube no es la solución a cualquier problema.
- ▶ Debemos valorar nuestras necesidades de infraestructura.
- ▶ Vigilemos también posibles restricciones administrativas en cuanto a la localización de datos críticos o personales de nuestro negocio.



Y vigilemos también posibles restricciones administrativas en cuanto a la localización de datos

Material adicional

Ejercicios de sql <https://pymesenlinea-capacita.corfo.cl/mod/resource/view.php?id=1342&redirect=1>

Curso de sql

<https://www.youtube.com/playlist?list=PLU8oAIHdN5Bmx-LChV4K3MbHrpZKefNwn>

Qué es una base de datos

<https://www.oracle.com/cl/database/what-is-database/>

Qué es una base de datos relacional

<https://www.oracle.com/cl/database/what-is-a-relational-database/>

Módulo 4: Automatización de procesos

Corfo – Pymes en Línea

Clase 1: Automatización

Automatización



- ▶ Máquina auto operada que sigue una secuencia de instrucciones.
- ▶ Los primeros en consolidar el concepto fue la compañía Ford.



Ellos establecieron el primer departamento de automatización del que se tiene registro.



Clase 1: Automatización

Automatización



► Definición actual:

Tecnología por la cual un proceso es realizado con mínima intervención humana, utilizando varios sistemas de control:

- Mecánicos.
- Hidráulicos.
- Eléctricos.
- Computadores.

► La IA es parte de una transformación más grande que está ocurriendo, una nueva revolución industrial.



4.1. Automatización

una nueva revolución industrial.

Clase 1: Automatización

Automatización



- 1era Revolución Industrial: Usar agua y vapor para mecanizar producción.
- 2da : Poder eléctrico para producción masiva.
- 3ra : Utilización de tecnologías de la información para automatización.
- 4ta : Fusión de tecnologías que eliminan la distinción entre mundos físicos, digitales y biológicos.



**que eliminan la distinción entre mundos físicos,
digitales y biológicos.**



Clase 1: Automatización

Automatización

Beneficiados

- ▶ Innovadores.
- ▶ Inversionistas.
- ▶ Accionistas.

Más afectados

Los puestos de trabajo en aquellas labores donde ocurre la transformación más profunda.

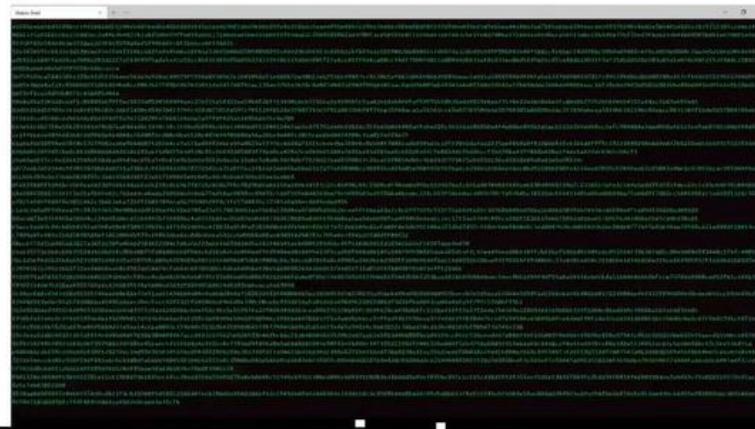


En este caso, se habla que la automatización reemplazará cerca de 30 millones de trabajos para el 2030.

Clase 1: Automatización

Automatización

La capacidad computacional actual, y la manera de controlar a los computadores es a través de los lenguajes de programación.



A estos les dedicaremos unos minutos.



Clase 1: Automatización

Lenguajes de programación



- Son un vocabulario y reglas gramáticas para dar instrucciones a un computador.
 - Su uso es bastante transversal en la cadena de valor información acción, veremos el uso que le pueden dar los analistas de datos y conocedores del aprendizaje automático.

alguno para trabajar con datos y automatizar procesos.

Clase 1: Automatización

Excel



- ▶ Sabiendo un lenguaje de programación, podríamos hacer el mismo proceso, corriendo una sola línea de código.
- ▶ Instantáneo.
- ▶ Se podría agendar para que se realizara automáticamente.

```
$('.count').each(function () {
  $(this).prop('Counter', 0).animate({
    Counter: $(this).text()
  }, {
    duration: 4000,
    easing: 'swing',
    step: function (now) {
      $(this).text(Math.ceil(now));
      $(this).html(text);
    }
  });
});
});
```

```
function(){
  elementById('objetos');
  click = function(){
}
```

que abrir un programa alguno ni ejecutar código.

Clase 1: Automatización



Razones para migrar a la programación

- Fragilidad de la memoria.
¿Qué ocurre si nos piden realizar nuevamente el mismo análisis de hace unos meses?
- Re utilizable y automatizable.
- Liberar considerable tiempo en tareas repetitivas.
- Facilidad de encontrar errores.



También, es normal que ocurran errores en un proceso de trabajo con datos.



Clase 1: Automatización



Razones para migrar a la programación

- Open source o código abierto.
- Gratis
- Comunidades que ayudan.
Seguir creciendo en las herramientas desarrolladas y en divulgar conocimientos.
- Gran disponibilidad de material para aprender.
- Desarrollo constante de nuevas herramientas relacionadas con IA.



donde Google jugó un rol importante al liberar herramientas como TensorFlow.



Clase 1: Automatización

¿Cuándo automatizar un proceso?



No siempre conviene pensar en automatizar todo.

- ▶ Requiere bastante esfuerzo.
- ▶ Quizá el esfuerzo realizado no compensará el beneficio de automatizar esa tarea.
- ▶ No será posible dejar una guía muy definida de cuando automatizar y cuando no.



deberían al menos considerar los siguientes elementos al hacer su evaluación:



Clase 1: Automatización

¿Cuándo no automatizar un proceso?



Son la contraparte de lo mencionado.

- ▶ Procesos donde no se tiene una suficiente experiencia no deberían ser pensados en automatizar.
- ▶ La automatización no libera suficientes recursos para que valga la pena.
- ▶ Tener en consideración los aspectos éticos.

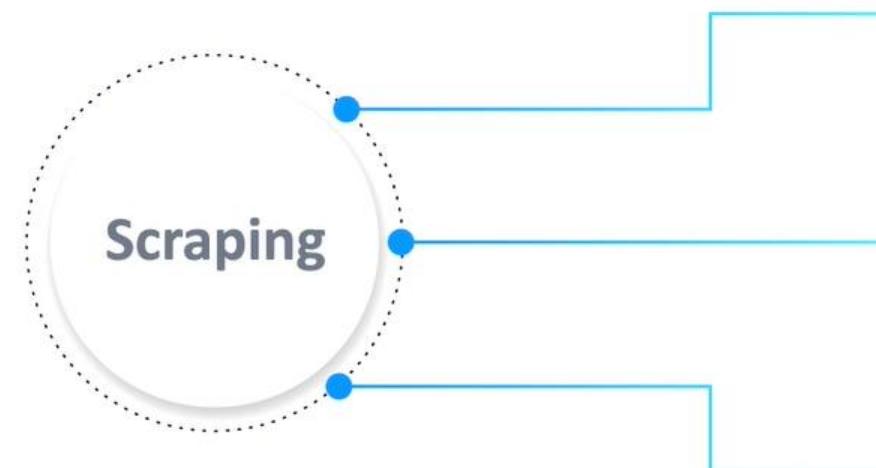


Las consideraciones éticas son complejas, pero la automatización es solo una herramienta,



Clase 2: Caso estudio 1 Scraping

Solución



Programa que vaya recorriendo distintas páginas y guardando la información.

Inicie el proceso automáticamente cada cierto tiempo, y va acumulando información en alguna base de datos.

Este tipo de robots puede hacer básicamente todo lo que una persona puede.

**que tengan que realizar en internet podrían, en teoría,
pensar en un robot que las haga por ustedes.**



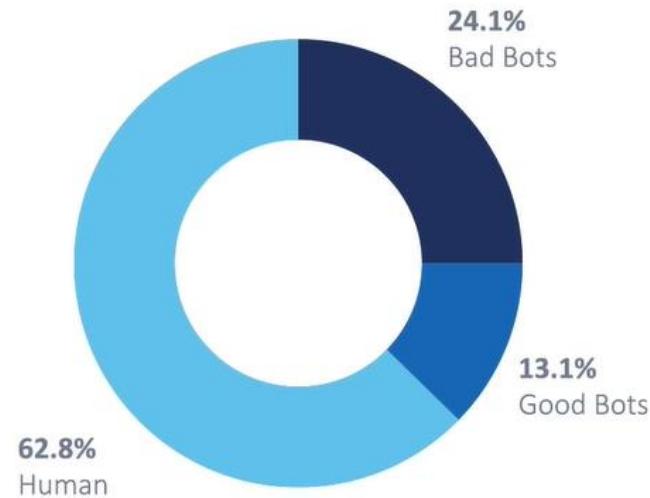
Clase 2: Caso estudio 1 Scraping

¿Realmente hay tantos robots?



Bad Bot v Good Bot v Human Traffic 2019

- ▶ La manera en que funcionan todos los motores de búsqueda es gracias a robots que se pasan el día navegando por internet, clasificando y ordenando su información.



Hay, por supuesto, otros mal intencionados. Hay bots que intentan hacerse pasar por personas para hacer



Clase 2: Caso estudio 1 Scraping

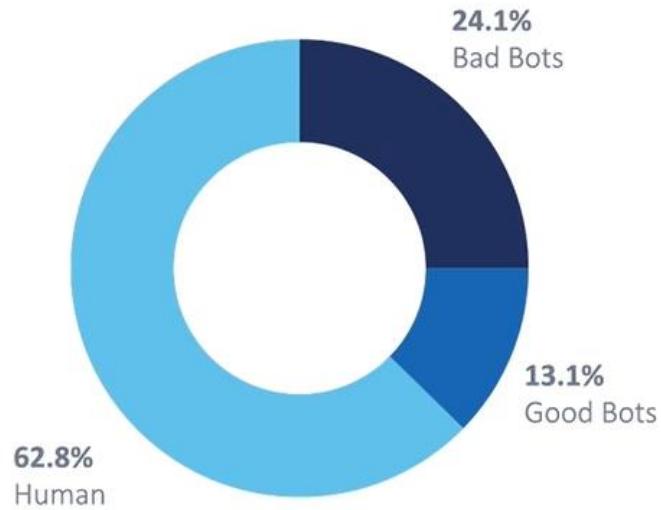
¿Realmente hay tantos robots?



Presiona Esc para salir de la pantalla completa

Bad Bot v Good Bot v Human Traffic 2019

- ▶ Uno podría programar miles de robots que entren simultáneamente a una página, o hagan uso de algún otro servicio de internet.
- ▶ Las páginas web tienen una cierta capacidad de recibir tráfico.



de poder satisfacer la demanda usual de usuarios humanos sin inconvenientes.



03:50



Clase 2: Caso estudio 1 Scraping

¿Realmente hay tantos robots?



Bad Bot v Good Bot v Human Traffic 2019

- ▶ **Ataques DDOS** - Denegación de servicio
 - Tienen como objetivo saturar los recursos de un servicio.
 - Este tipo de ataques es una práctica maligna de un recurso que, aunque no se use maliciosamente, igualmente tiene varios detalles.



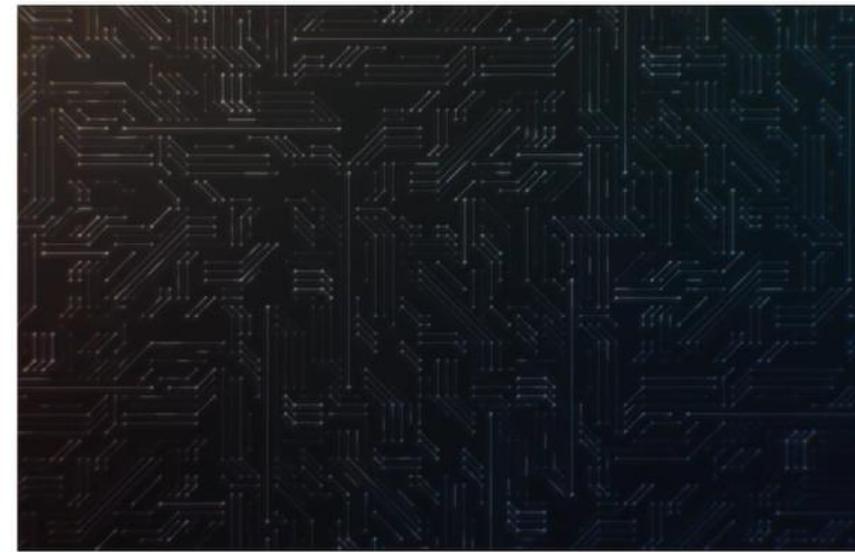
4.2. Caso estudio 01: Scraping

igualmente tiene varios detalles.

Clase 2: Caso estudio 1 Scraping

Detalles

- ▶ ¿Es el portal de ventas de productos dueño de los datos?
- ▶ En la mayoría de los casos esta información es de dominio público.
- ▶ Dan facilidades de que usuarios comenten acerca de los productos que han comprado y le den una nota.
- ▶ No queda siempre claro si esa información es propiedad del usuario que lo emitió, propiedad de la página que lo publica, o no es propiedad de nadie.



Un caso incluso más extremo es uno del cual se habla actualmente en Europa, donde los bancos intentan usar



Clase 2: Caso estudio 1 Scraping

Detalles



- El respeto por los servidores ajenos es fundamental.
- No sobrecargar servidores con robots.
- Corrección de errores en los datos recolectados .
- Los robots se programan usualmente para que funcionen en una página con el formato de ese momento, y la web es muy dinámica.



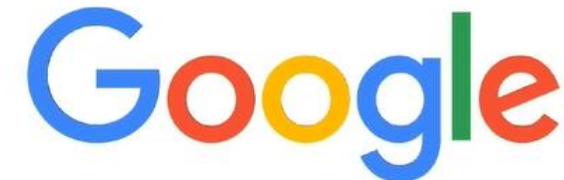
y la web es muy dinámica.



Clase 2: Caso estudio 1 Scraping

APIs

- ▶ Interfaces de programación de aplicaciones.
- ▶ Es una manera de disponibilizar el acceso a un servicio de manera más ordenada.
- ▶ Los gigantes tecnológicos como Google o Facebook siempre proveen alguna API para que usuarios puedan acceder a información que ellos poseen, para usarse con distintos fines.



**a información que ellos poseen,
para utilizarse con distintos fines sin afectar sus servicios.**



Clase 2: Caso estudio 1 Dashboards



Dashboard

- Tomar datos de los procesos internos de la compañía y consolidarlos de tal forma de mostrarlos automáticamente.
- Se puede evitar conocimiento experto de procesos de:
 - ▶ Extracción.
 - ▶ Limpieza de datos.
 - ▶ Confección de gráficos.
 - ▶ Tablas.
 - ▶ Mapas.
 - ▶ Elementos visuales.
- Su diseño debe apoyarse en conocimiento experto para poder contener toda la información relevante.

su diseño debe apoyarse en conocimiento experto para poder contener toda la información que sea relevante.



03:01



Clase 2: Caso estudio 1 Dashboards

Lineamientos



- Tomar en cuenta la audiencia.
- Filtrar verdaderamente toda la información irrelevante.
- Proveer contexto.
- Priorizar la simpleza.
 - ▶ Puede asumir una complejidad grande rápidamente si se empieza a incluir demasiada personalización.



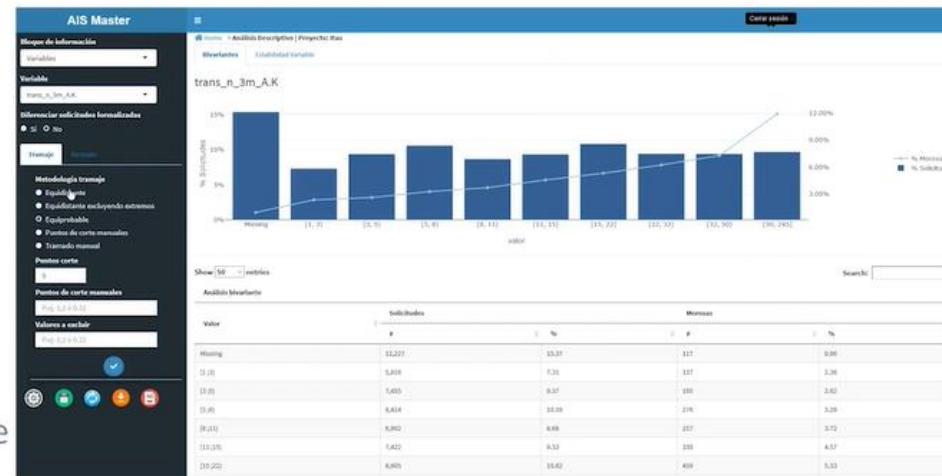
demasiada personalización.



Clase 2: Caso estudio 1 Dashboards

Ejemplo

- ▶ Modelos predictivos, lo que nos hemos referido como Machine Learning.
- ▶ Conocer todo lo que pasa dentro de estos modelos:
 - Cómo se comportan las variables.
 - Qué resultados ha tenido el modelo.
- ▶ Anteriormente esto era manual.
- ▶ Reduce considerablemente la manualidad y tiempo de respuesta del análisis de los modelos predictivos, y provee información útil y accionable para los analistas encargados de estas tareas.



además de proveer información útil y accionable para los analistas encargados de estas tareas.



Clase 3: Automatizando casos de estudio

Solución



► Con R.

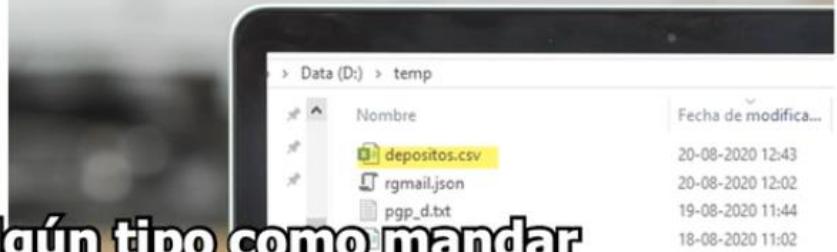
Basta con correr una línea de código y, se genera un archivo donde se encuentran los montos y personas a quien se les realizó el depósito.

```
# Script Revisa mails
library(gmailr)

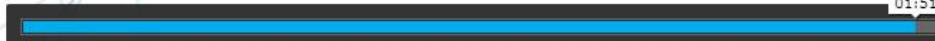
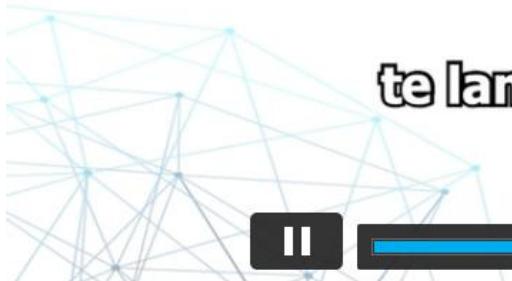
gm_auth_configure(path = "rgmail.json")
emails <- gm_messages("transferencias@bci.cl")
ids <- lapply(emails[[1]][[1]], function(x) x$id)[1:10]
mensajes <- lapply(ids, function(x) gm_body(gm_message(x)))
valores <- lapply(mensajes, function(x) stringr::str_extract(x, "\\$[0-9]+"))
destinatarios <- lapply(mensajes, function(x) stringr::str_extract(x, "destinatario"))
unlist() %>%
  stringr::str_extract("[A-Z].*[A-Z]")

df <- data.frame(valores = valores,
                  destinatarios = destinatarios, stringsAsFactors = FALSE)

write.table(df, file = "depositos.csv", sep = ";", row.names = FALSE)
```



te lance una alerta de algún tipo como mandar un mensaje al celular.



Material adicional

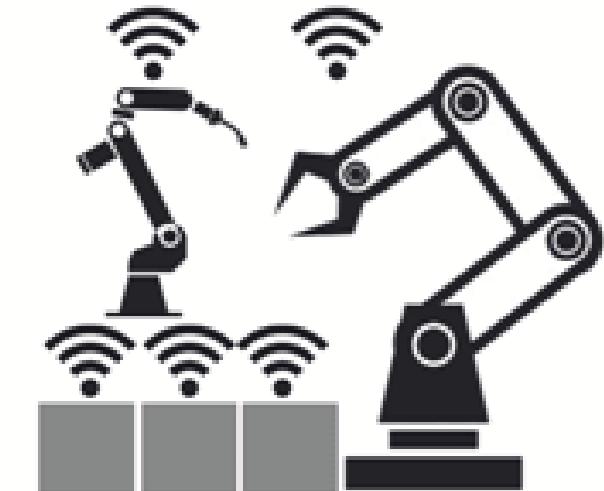
Cuarta revolución industrial:

<https://www.uchile.cl/noticias/146095/que-es-la-industria-40-y-que-desafios-traera-a-chile>

Introducción a Python

https://www.youtube.com/watch?v=bY6m6_IIN94&list=PLi01XoE8jYohWFpC17Z-wWhPOSuh8Er-Python

<https://learnpythonthehardway.org/python3/>



1^a

Mecanización

Máquina de vapor,
energía hidráulica y
mecanización

2^a

Electricidad

Producción en masa,
cadena de montaje y
electricidad

3^a

Informática

Automatización,
tecnologías de la
información y la
comunicación (TIC)

4^a

Digitalización

Internet de las cosas, la
nube, coordinación
digital, sistemas
ciberfísicos y robótica

¿Cuál de los siguientes no corresponde a un lenguaje de programación?

Seleccione una:

- a. HTML
- b. Todos son lenguajes de programación
- c. Python
- d. R

Respuesta incorrecta.

La respuesta correcta es: HTML

HTML is not considered as programming language because there is no variables to store value in it.

Also there is no conditions and no loops as there is in real programming languages such as java, C++, PHP and lot of others.

HTML on the other hand only have tags. You can just open the tag and after some content or styles close it. But you can merge HTML and PHP so that HTML can work along with programming language i.e. PHP.

Módulo 5: Machine Learning

Corfo – Pymes en Línea

Clase 1: Introducción a Machine Learning

Un poco de historia



Arthur Samuel

Se propuso investigar si podría ser capaz de programar un robot que jugara damas y que fuera mejor que él, sin programarlo explícitamente para esto.

Un programa por computador que utiliza juicio para cada uno de los movimientos, tendría y tardaría muchísimo tiempo en evaluar cada uno de estos y cada universo de jugadas posibles.

Un programa por computador que utiliza juicio para cada uno de los movimientos tendría y tardaría

Clase 1: Introducción a Machine Learning

Un poco de historia



Arthur Samuel

Finalmente logró crear un programa que era mejor que él y que había aprendido solo, en base a la experiencia.

Este programa sería el trabajo inicial de lo que el mismo Arthur Samuel acuñaría años más tarde como Machine Learning.

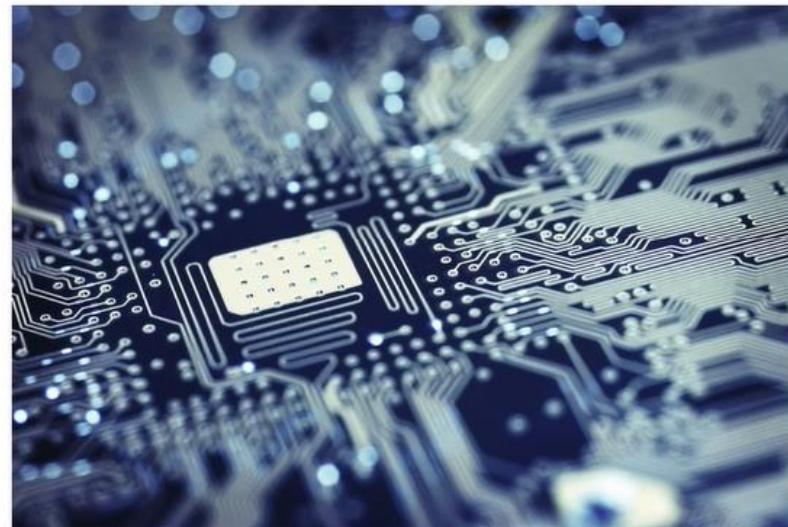
Clase 1: Introducción a Machine Learning

Machine Learning



Aprendizaje Automático

- ▶ Un requisito crucial era la capacidad computacional, la cual era escasa y exclusiva en ese tiempo.
- ▶ En los años 80 se había desarrollado lo que se conoce como **back propagation**:
- ▶ El uso masivo actual del machine learning es por este cambio tecnológico: mayor capacidad de cómputo y una disminución en los precios.



previo a los años 80. Pero no es hasta entonces, **mas** ahora, donde realmente tenemos

5.1. Introducción Machine Learning



Clase 1: Introducción a Machine Learning

Machine Learning



Hoy está presente en muchos segmentos de la industria y ciencia básica.

- ▶ Industria de los alimentos.
- ▶ La banca.
- ▶ La industria sanitaria.
- ▶ Salud.
- ▶ Educación.



**la banca, la industria sanitaria, la salud, la educación,
entre muchos otros.**



Clase 1: Introducción a Machine Learning

Machine Learning - ejemplos



Hoy está presente en muchos segmentos de la industria y ciencia básica.

► Reconocimiento de voz

- Siri, la popular ayudante de Apple aprende
de tu información:

Mensajes.

Calendario.

Música.

Recordatorios.

Email.

Contactos.



**para poder entenderte mejor cuando le formulas
alguna pregunta o incluso hacerte alguna sugerencia**



Clase 1: Introducción a Machine Learning

Machine Learning - ejemplos



Hoy está presente en muchos segmentos de la industria y ciencia básica.

- ▶ Procesamiento de lenguaje natural
 - Es un campo de la IA que intentan entender idiomas o imágenes.



Los algoritmos de aprendizaje como, ya vimos anteriormente, se usan mucho en programas de personalización.



Clase 1: Introducción a Machine Learning

Tipos de Algoritmos de ML



► Modelos supervisados.

Son los que abordaremos con mayor profundidad, debido a su utilidad en el mundo empresarial.

► Aprendizajes no supervisados.

Se han visto impulsados principalmente por el marketing.

Y por último el aprendizaje por refuerzo, que cada vez cobra más fuerza.



Clase 1: Introducción a Machine Learning

Aprendizaje no supervisados



- No tenemos un input A y output B.
- Solo tenemos los datos de entrada definidos y no una salida.
- Se intenta revisar de qué manera los datos están relacionados en sí mismos.
- Si existen segmentos o cluster entre los datos, es decir agrupaciones entre datos, o si existen casos que están lejos o no del grueso de los datos.



es decir agrupaciones entre datos, o si existen casos que están lejos o no del grueso de mis datos.



Clase 1: Introducción a Machine Learning

Aprendizaje por refuerzos



- Está conformado por una máquina o agente inteligente que aprende a optimizar un proceso de decisión.
- Si el resultado de esa decisión es beneficioso, el agente aprende automáticamente a repetir esa decisión en el futuro.
- Si el resultado es perjudicial evitará volver a tomar la misma decisión.
- El agente aprende qué decisiones son más adecuadas según la situación, y desarrolla estrategias a largo plazo que maximicen los beneficios.



y desarrolla estrategias a largo plazo que maximicen los beneficios.



Clase 1: Introducción a Machine Learning

Aprendizaje por refuerzo

Presiona **Esc** para salir de la pantalla completa



Consiste en aprender a decidir ante una situación determinada, que acción es la más adecuada para lograr un objetivo.

- ▶ Primero que es un proceso iterativo de prueba y error.
- ▶ Y segundo el aprendizaje se logra a través de señales de refuerzo.



Y segundo el aprendizaje se logra a través de señales de refuerzo.



08:49



Clase 2: Aprendizaje supervisado y no supervisado

Recordando



- Campo de estudio que les da a los computadores la habilidad de aprender sin ser programados explícitamente.



- ▶ Datos de entrada: antigüedad como analista de datos.
- ▶ Datos de salida: sueldo.
- ▶ Modelo ML: la función que relaciona la antigüedad con el sueldo.



De esta manera podríamos estimar a distintos años de experiencia cuál es la expectativa de renta



Clase 2: Aprendizaje supervisado y no supervisado

Tipos de Algoritmos de ML



Supervisado

- ▶ Regresión
- ▶ Clasificación
- ▶ Serie temporal

No supervisado

- ▶ Clustering
- ▶ Aprendizaje por refuerzo

Una vez tenemos claro el problema al cual nos enfrentemos, nos queda identificar con que metodología específica resolvemos dicho problema.

Bien. Una vez tenemos claro el problema al cual nos enfrentemos, nos queda identificar



03:48



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados



Metodología de desarrollo

► Corresponde a la forma en la que llegamos a esa función, y esto se logra gracias a la matemática y la computación.

- Árboles de decisión.
- Máquinas de soporte vectorial o support vector machine (SVM).
- Redes Neuronales.
- Regresión logística.
- Regresión lineal.
- Random forest.
- Gradient boosting.



Redes Neuronales. Regresión logística. Regresión lineal. Random forest o Gradient boosting entre otras.



05:03



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados



¿Cómo escoger la metodología?

► Número de variables de entrada A

- No tendremos solo un input, sino que cientos.
- Hay metodologías como las regresiones que funcionan idealmente con pocas variables de entrada A.
- Gradient boosting funcionan mejor con muchas variables.

► Número de casos

- Existen metodologías que se sobre ajustan muy fácilmente cuando tienen pocos datos, y que necesitan muchos casos para entregar resultados confiables.



y otras que necesitan muchos casos para poder realizarse.



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados

Presiona **Esc** para salir de la pantalla completa



¿Cómo escoger la metodología?

► Naturaleza de los datos

- Intentar predecir algo simple como la expectativa salarial o algo más complejo como identificar si una persona tiene o no una mascarilla.
- Las redes neuronales son perfectas para datos no estructurados.

► Tipo de problema

- Si es regresión o clasificación, hay metodologías que funcionan solo con un tipo de casos, como la regresión logística.



con un tipo de casos, como es la regresión logística.



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados

Presiona **Esc** para salir de la pantalla completa



¿Cómo escoger la metodología?

► Naturaleza de los datos

- Intentar predecir algo simple como la expectativa salarial o algo más complejo como identificar si una persona tiene o no una mascarilla.
- Las redes neuronales son perfectas para datos no estructurados.

► Tipo de problema

- Si es regresión o clasificación, hay metodologías que funcionan solo con un tipo de casos, como la regresión logística.



con un tipo de casos, como es la regresión logística.



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados

Presiona **Esc** para salir de la pantalla completa



¿Cómo escoger la metodología?

► Características de mis datos

- Si tengo valores desconocidos o missing dentro de mi base de datos.
- Todas mis variables tienen que tener un valor para poder resolver el problema.
- No aceptan valores no numéricos como datos de entrada.



un problema son variadas y es importante tener en consideración estos elementos para ver



Clase 2: Aprendizaje supervisado y no supervisado

Supervisados



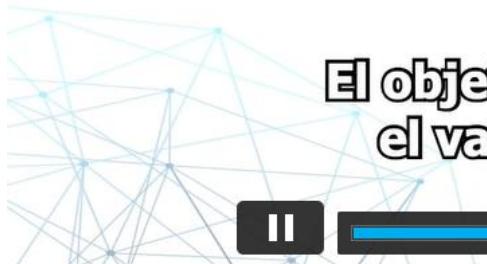
Se necesita similar cantidad de casos true y false.

Árbol de decisión

- ▶ Son un método de aprendizaje supervisado utilizable para clasificación y regresión.
- ▶ El objetivo es crear un modelo que pueda predecir el valor de una variable objetivo o respuesta.



El objetivo es crear un modelo que pueda predecir el valor de una variable objetivo o respuesta,

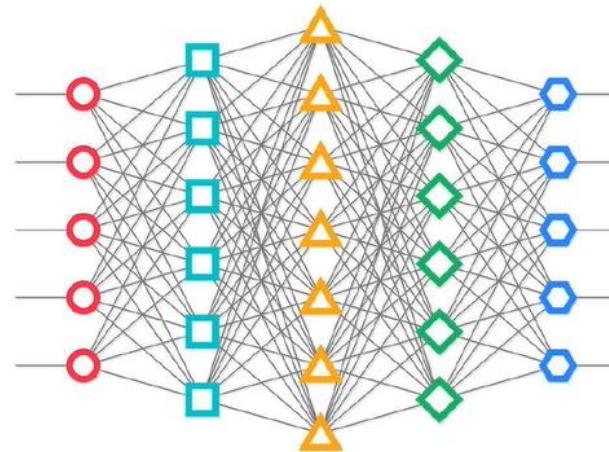


Clase 2: Aprendizaje supervisado y no supervisado

Red Neuronal



- ▶ Corresponde a una red conformada por un conjunto de neuronas interconectadas, las cuales están organizadas en capas.
- ▶ La capa de entrada que recibe la información.
- ▶ Una o varias capas ocultas, que realizan distintas operaciones matemáticas.
- ▶ Capa de salida que va a generar la salida que nosotros esperamos.



a las redes neuronales.



Clase 2: Aprendizaje supervisado y no supervisado

No Supervisados

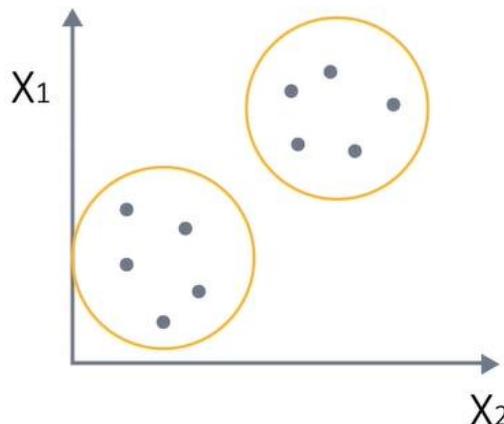


Nos dan el conjunto de datos y no nos dicen qué hacer con ellos y no nos dicen qué es cada dato puntual.

¿Puedes encontrar alguna estructura en los datos?

- ▶ El algoritmo de aprendizaje no supervisado puede dividir estos datos en dos grupos separados.

Algoritmo de Agrupamiento.



a esto se le llama **algoritmo de agrupamiento**.



Clase 2: Aprendizaje supervisado y no supervisado

No Supervisados



Segmentación de mercado

- ▶ Esto es aprendizaje no supervisado.
- ▶ Tenemos todos estos datos de clientes, pero no sabemos de antemano cuáles son los segmentos de mercado.
- ▶ Pero debemos dejar que el algoritmo descubra todo esto sólo a partir de los datos.



Pero debemos dejar que el algoritmo descubra todo esto a partir de los datos.

Clase 2: Aprendizaje supervisado y no supervisado

No Supervisados

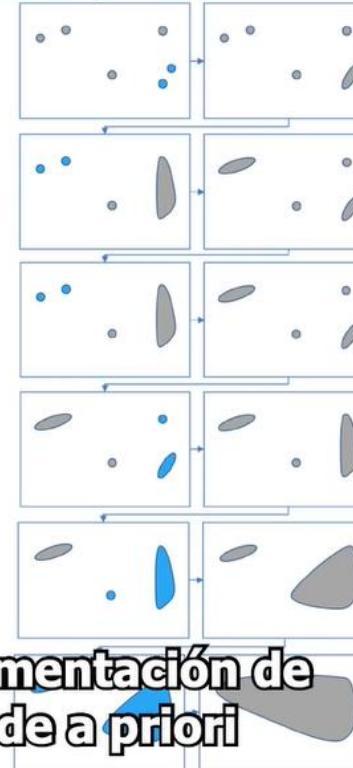


Cluster jerárquico.

- ▶ Parte considerando que cada punto es un cluster, y luego los va agrupando paso a paso de acuerdo a los datos más similares o cercanos.
- ▶ Hasta el final agruparlos todos.
- ▶ Este ejemplo se puede usar en segmentación de mercado, donde el analista decide a priori cuantos grupos cree que existen y analizar luego los resultados del modelo.



Este ejemplo se puede usar en segmentación de mercado, donde el analista decide a priori



Clase 2: Aprendizaje supervisado y no supervisado

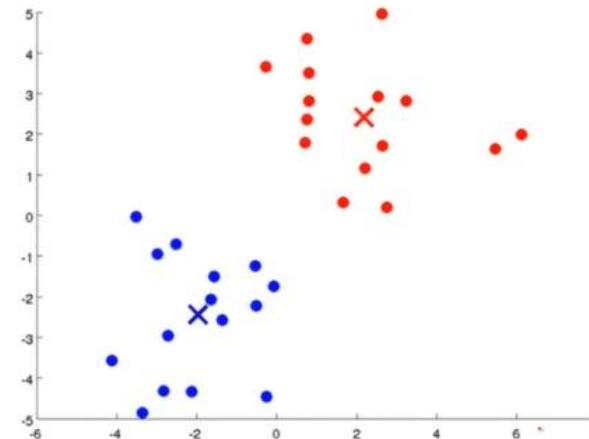
No Supervisados

Presiona para salir de la pantalla completa



K-means

- ▶ Crear una segmentación.
- ▶ Primero tenemos datos iniciales no agrupados.
- ▶ Luego se definen estas marcas de manera aleatoria, que llamaremos centroides iniciales.
- ▶ Los datos se agrupan de acuerdo a qué centroide tienen más cerca.
- ▶ Una vez asignados a un grupo, se vuelve a calcular centroide.
- ▶ Se hace una nueva asignación.



Y así se hace una asignación sucesivamente, proceso que se repite hasta que converge matemáticamente hablando.

Clase 3: Integrando IA a tu negocio



Cadena de valor Información-Acción

- Forma de organizarse y pensar los pasos necesarios para desarrollar un proyecto relacionado con datos.
- Al momento del análisis se usarán técnicas avanzadas de modelamiento para el problema que se haya definido.
- Las compañías que son buenas realizando cada uno de estos pasos son las que logran integrar la IA a sus negocios.



la Inteligencia Artificial a sus negocios. En particular, los siguientes elementos son cruciales.



Clase 3: Integrando IA a tu negocio

Aspectos relevantes para integración de ML



Adquisición de datos estratégicos

- ▶ ¿Ya tengo datos disponibles para enfrentar algún problema o debo primero pensar en cómo conseguirlos?
- ▶ ¿Existen datos abiertos que me pueden servir?
- ▶ Muchas veces se gasta mucho tiempo y capacidad en guardar una gran cantidad de datos, pero muchos de estos datos no tienen sentido almacenarlos.
- ▶ Identificar que datos almacenar y en que formatos es una primera gran decisión.



Por lo que el identificar que datos almacenar y en que formatos es una primera gran decisión.

Clase 3: Integrando IA a tu negocio

Aspectos relevantes para integración de ML



Adquisición de datos estratégicos

Unificación de las bases de datos

- ▶ No es necesario pensarla como montar una arquitectura de data lakes o data warehouse compleja.
- ▶ Con una carpeta dedicada donde se almacenen todos los archivos relevantes basta.
- ▶ Desarrollar el concepto de integración de la información apoyado en un enfoque de IA y ML y estar atento a oportunidades de automatización.
- ▶ Ayudará a identificar sus potenciales proyectos, indicando los datos que podrían resultar interesantes de almacenar y algunas claves para su mejor uso.

en totalidad pueda estar al tanto de nuevas oportunidades.



Clase 3: Integrando IA a tu negocio

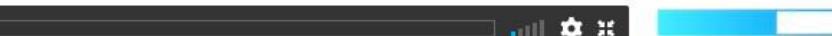
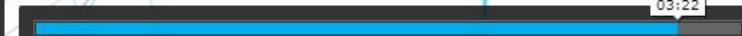


Integración de AI es un proceso gradual

Proyecto piloto

- Adquirir experiencia y no resulte muy costoso.
La predicción de demanda para poder tener un stock de productos óptimo
- Que sea piloto con un mínimo producto viable (o MVP).
- En sus aspectos técnicos y prácticos dentro de la organización.

por Machine Learning o en sus aspectos técnicos y prácticos dentro de la organización.



Clase 3: Integrando IA a tu negocio



Integración de AI es un proceso gradual

Construir equipo de AI

- ▶ Realizar capacitaciones más transversalmente en su empresa.
- ▶ Lo más importante inicialmente es que haya un conocimiento de las capacidades de la IA para que puedan surgir ideas de todas las áreas de la empresa.

para que puedan surgir así ideas entre todos en la empresa.



Clase 3: Integrando IA a tu negocio



Integración de AI es un proceso gradual

Optar por ayuda externa

- ▶ Es una inversión que resulta difícil de justificar.
- ▶ Una consultora podría resultar útil para obtener una idea rápida de qué problemas dentro de la organización podrían ser abordados con ML.
- ▶ Apoyarse para una implementación sencilla a la gestión.
- ▶ Obtener una mirada experta sobre los aspectos a tener en cuenta para el seguimiento de las herramientas implementadas.

Por último, podrían obtener una mirada experta sobre los aspectos a tener en cuenta



Clase 4: Casos prácticos del Machine Learning



Cómo puede una PYME empezar con Machine Learning

Identificación del problema

- ▶ La mayoría de las empresas intentan comenzar con lo que creen que es la solución adecuada para ellos en lugar de averiguar cuál es el problema que quieren resolver.
- ▶ La ciencia de datos no se puede presentar simplemente a una empresa y optimizar mágicamente toda la organización.

Tan avanzado como la tecnología es aun más el requisito de un ser humano para identificar el problema



Clase 4: Casos prácticos del Machine Learning



Cómo puede una PYME empezar con Machine Learning

Gestionar las expectativas

- ▶ Las tecnologías de ciencia de datos deben aplicarse a cualquier negocio de forma estratégica.
- ▶ Ninguna empresa, pero especialmente las pyme, deben buscar la aplicación inmediata de estas herramientas en todas las áreas de su negocio.
- ▶ Deberían identificar aquellas áreas clave donde creen que tendrá el mayor impacto al resolver las preguntas o problemas que ya han sido identificados.

De esta forma, es posible obtener los resultados deseados con una mínima interrupción de los negocios que ya están en curso.



Clase 4: Casos prácticos del Machine Learning



Cómo puede una PYME empezar con Machine Learning

Gestionar las expectativas

- ▶ Las empresas deberían adoptar el método científico al comenzar con Big Data y Machine Learning.
- ▶ Partir con una hipótesis debería construirse y luego los datos deben aplicarse para probar su base.
- ▶ Los mejores resultados se obtienen al explorar los datos de esta manera.
- ▶ La dificultad para las pymes es mantener el costo de tales métodos exploratorios al mínimo.

La dificultad para las pymes es mantener el costo de tales métodos exploratorios al mínimo.



Clase 4: Casos prácticos del Machine Learning



Cómo puede una PYME empezar con Machine Learning

Búsqueda de datos

- ▶ Ya identificada la pregunta y teniendo en cuenta las expectativas, debemos evaluar si contamos con los datos necesarios para poder generar una solución.

¿Tengo la información necesaria?

¿He guardado la información de mis clientes?



Materiales adicionales

- Introducción al ML 1 <http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>
- Introducción al ML 2 <http://www.r2d3.us/visual-intro-to-machine-learning-part-2/>
- Esta IA juega al escondite demasiado bien
<https://www.youtube.com/watch?v=5SkQuT3kZOc&t=386s>
- IA que aprende a estacionarse
https://www.youtube.com/watch?v=VMp6pq6_QjI

Preguntas

- La masificación del uso del machine learning en los últimos años se debe principalmente a: La mejora de la capacidad de cómputo y de la baja de precios en la infraestructura necesaria
- ¿Cuál es el paso en la cadena de valor información-acción donde se genera más valor el machine learning? Análisis

Módulo 6: Deep Learning

Corfo – Pymes en Línea

Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hitos del Deep Learning



Deep Blue, 1997

- ▶ Inteligencia artificial de IBM.
- ▶ Vence en el juego del ajedrez a Garry Kasparov, uno de los mejores jugadores del mundo en esta disciplina.
- ▶ Era primera vez que se lograba que una inteligencia artificial dominara en un juego de estrategia de la complejidad del ajedrez.



y además porque este era un juego que hasta ese entonces se encontraba fuertemente ligado

Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hitos del Deep Learning

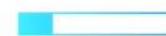


Watson gana en Jeopardy!, 2011

- ▶ Procesamiento de Lenguaje Natural
- ▶ Se intenta darle a los computadores la habilidad de procesar un Lenguaje Natural como lo son el español o el inglés.



lenguajes utilizados por los humanos, algo que para las máquinas es una tarea



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hitos del Deep Learning

Presiona **Esc** para salir de la pantalla completa



AlexNet, 2012

- ▶ Tipo de red neuronal conocido como red neuronal convolucional.
- ▶ Logra vencer en la competencia ImageNet.
- ▶ ImageNet es un conjunto de datos formado por una cantidad gigantesca de imágenes.
- ▶ El objetivo era entrenar un modelo que fuera capaz de recibir como entrada una de estas imágenes y que al procesarla entregara como salida la categoría a la que pertenece dicha imagen.



una de estas imágenes y que al procesarla entregara como salida la categoría



03:14



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hitos del Deep Learning

Presiona **Esc** para salir de la pantalla completa



Alpha Go , 2016

- ▶ El Go es un juego bastante más complejo.
- ▶ Se dice que en una partida de Go existen más movimientos posibles que átomos en todo el universo.
- ▶ Es un juego con una complejidad computacional bastante alta y ver a una inteligencia artificial superar el nivel humano en este juego es una demostración de las sorprendentes tareas que se pueden llevar a cabo con el Deep learning.



a una inteligencia artificial superar el nivel humano en este juego es una demostración



04:25



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hitos del Deep Learning



OpenAI presenta a GPT-3, 2020

- ▶ Una inteligencia artificial que es un modelo de lenguaje.
- ▶ Es un modelo que lo que hace es que a partir de cierto texto trata de predecir cuál es la siguiente palabra que debería venir en función de las palabras anteriores.
- ▶ Logra resolver otras tareas del lenguaje natural para las cuales inicialmente no fue entrenado, como por ejemplo, conversaciones muy coherentes.



**inicialmente no fue entrenado, como por ejemplo,
logra establecer conversaciones muy coherentes,**



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

¿Qué es el Deep Learning?



- ▶ Es una subárea del Machine Learning que trabaja con redes neuronales profundas.



- Es entrenar modelos que aprenden a partir de datos, pero sin ser programados de manera explícita.
- Aprenden a partir de los datos de manera automática.
- Dentro de la diversidad de modelos que existen en Machine Learning, hay uno en particular que es el de las redes neuronales.

que es el de las redes neuronales.



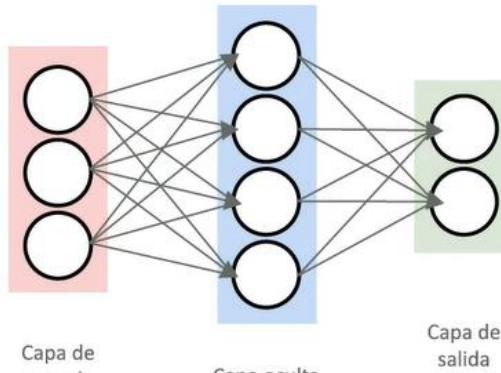
Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Deep Learning

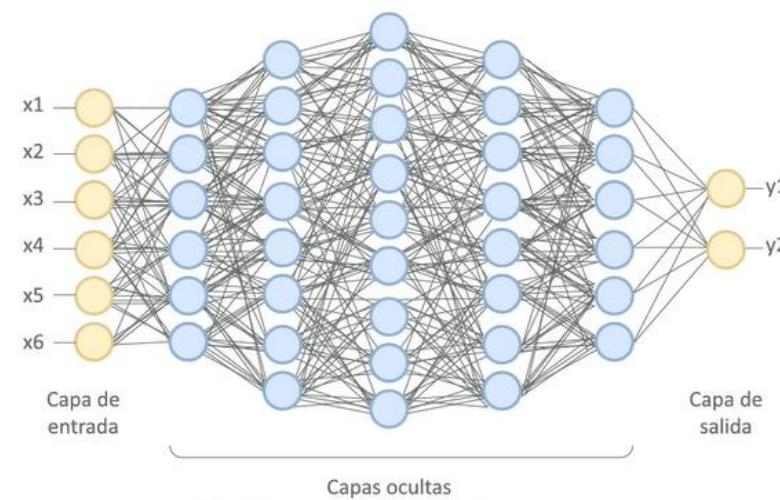


Redes neuronales simples vs profundas

Red neuronal simple



Red neuronal profunda



estamos pensando en modelos de redes neuronales profundas como la de la derecha.



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

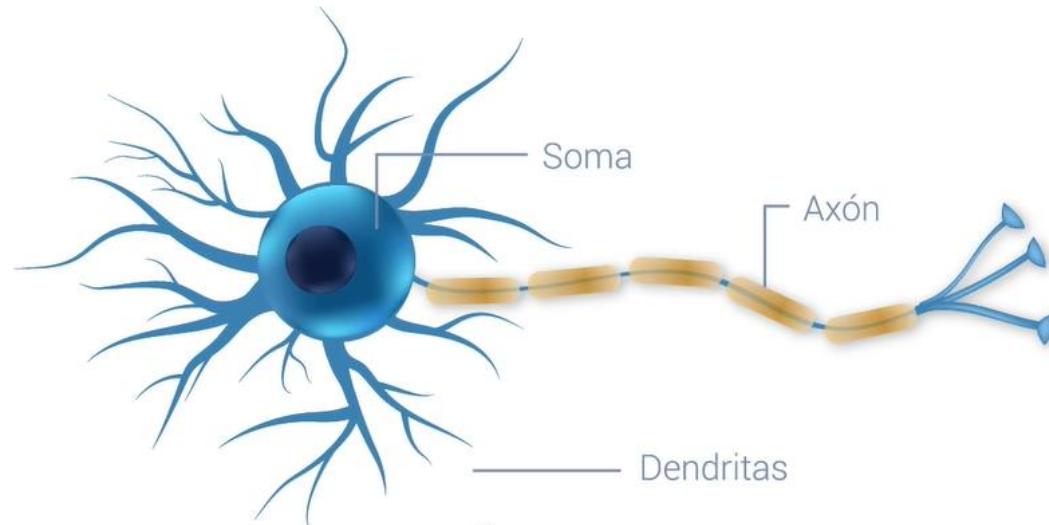
¿Qué es una neurona?



Es la unidad fundamental dentro de una red neuronal.

- Neurona biológica

Las dendritas se encargan de captar los impulsos nerviosos que emiten otras neuronas. Estos impulsos, se procesan en el soma y se transmiten a través del axón que emite un impulso nervioso hacia las neuronas contiguas.



Estos impulsos, se procesan en el soma y se transmiten a través del axón que emite



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

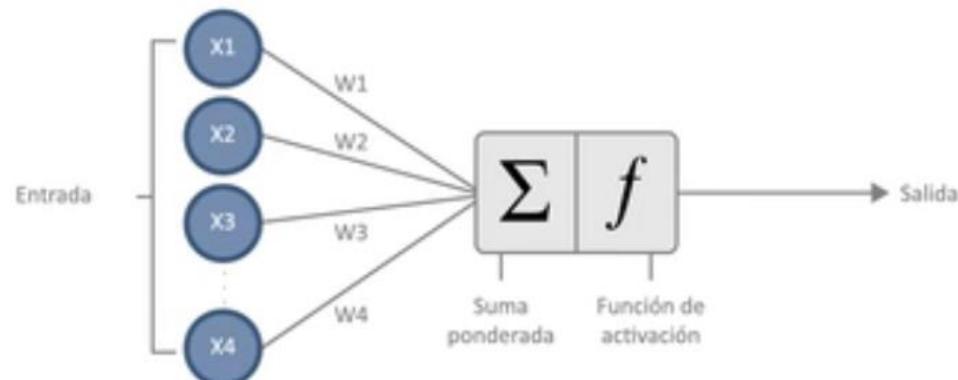
¿Qué es una neurona?



Es la unidad fundamental dentro de una red neuronal.

► Neurona artificial

- Va a recibir cierta información de entrada, que son estas variables de entrada.
- Cada variable de entrada ingresa a la neurona multiplicada por el peso respectivo.
- Internamente calculan la suma ponderada de sus entradas.
- Finalmente aplican una función de activación, y generan la salida.



ponderada es mayor a 0, solo en ese caso esta neurona se va a activar y por tanto va a generar una salida.

Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hechos que han favorecido al Deep Learning



- ▶ En los últimos años se han desarrollado computadores mucho más potentes.
- ▶ Paso de las CPU a las GPU
GPU: son unidades de procesamiento gráfico que nos permiten ejecutar varios procesos en paralelo.
- ▶ El entrenamiento de ciertas arquitecturas complejas de redes neuronales que podían demorar meses, ahora tarden sólo días o incluso horas.



ahora tarden sólo días o incluso horas, con lo cual se hace factible realizar varias pruebas



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

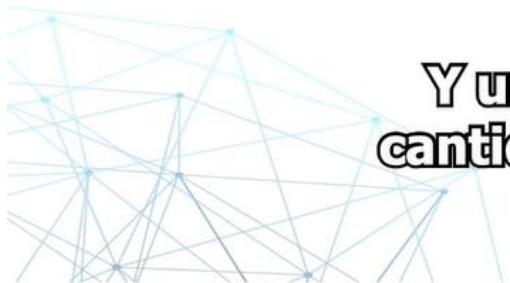
Hechos que han favorecido al Deep Learning



- ▶ Desarrollo de nuevos algoritmos y frameworks especializados.
 - Los algoritmos nos han permitido resolver ciertas tareas de manera más eficiente.
 - Los frameworks han simplificado el desarrollo de las redes neuronales al proveernos de herramientas con las que ya no se necesita programar todo desde 0.
- ▶ Enorme cantidad de datos disponibles.

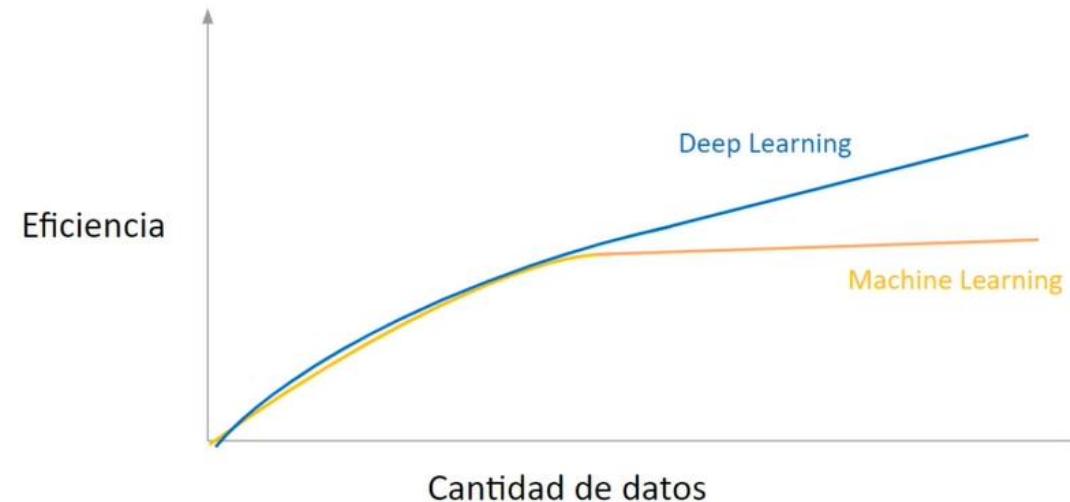


Y una última explicación está en la enorme cantidad de datos disponibles en la actualidad.



Clase 1: Entendiendo el Deep Learning y las Redes Neuronales Artificiales

Hechos que han favorecido al Deep Learning



Lo que se hace es añadirle más capas a la red,
con lo cual esta puede seguir mejorando su rendimiento.



Clase 2: Como aprende una red neuronal

Ejemplo



Predecir el precio de un inmueble

- Se realiza mediante aprendizaje supervisado.
- Se considera un conjunto de entrenamiento en donde cada observación tiene tanto sus variables o características como la etiqueta a predecir.
- Durante la fase de entrenamiento la red aprenderá a predecir las etiquetas a partir de las variables de entrada.



Veamos ahora un ejemplo concreto de esto con un modelo de valoración de inmuebles.



01:04

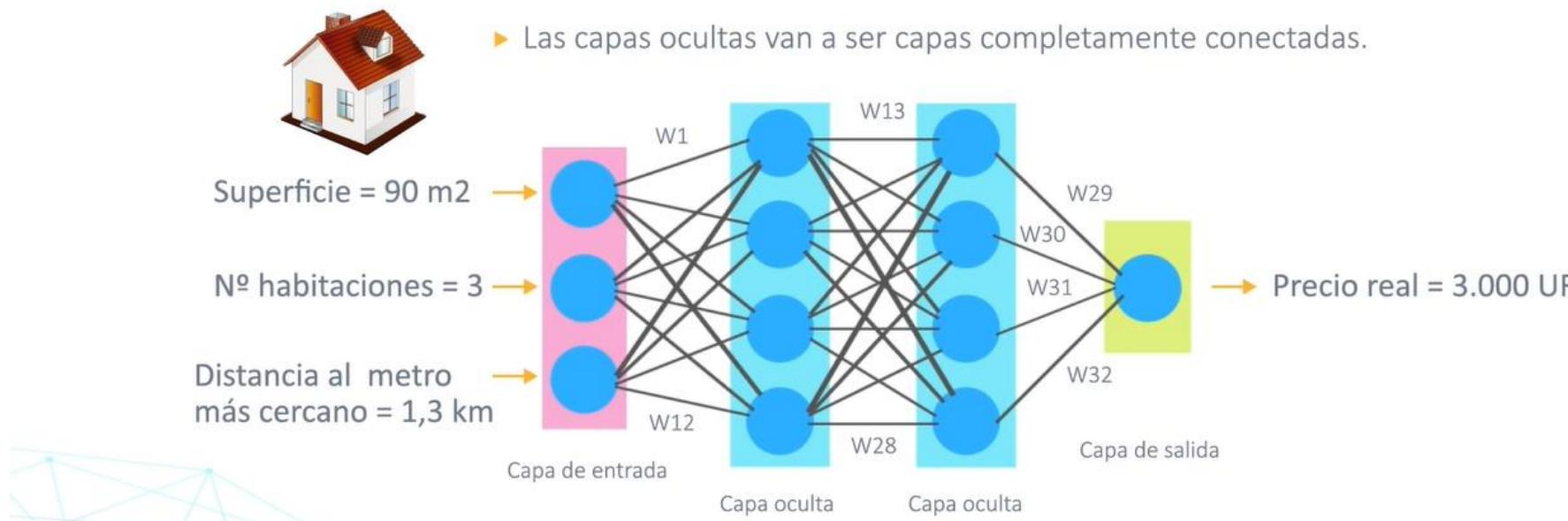


Clase 2: Como aprende una red neuronal

Ejemplo



Predecir el precio de un inmueble



cada neurona está conectada con todas las neuronas de la capa anterior y con todas las neuronas de la capa siguiente.

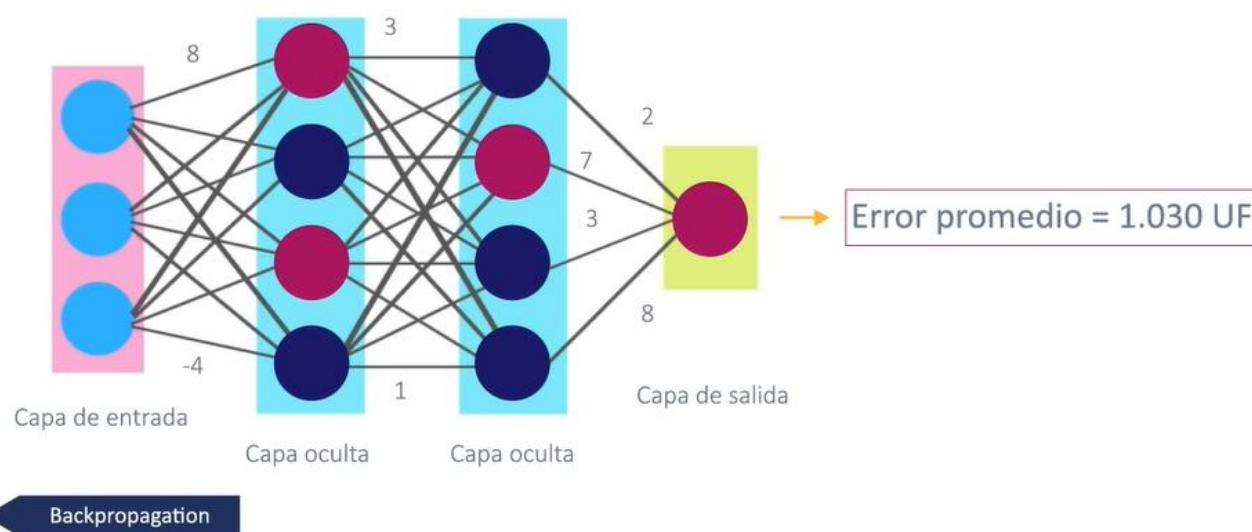
Clase 2: Como aprende una red neuronal

Propagación del error hacia atrás



Predecir el precio de un inmueble

- El error que comete la neurona en la capa de salida depende de lo que recibió como entrada.



es decir, de los errores de las neuronas de la capa anterior,
y estos a su vez de los errores de la capa anterior,

Clase 2: Como aprende una red neuronal

Entrenamiento de la red neuronal



- Es un proceso iterativo.
- Cada iteración contiene un paso de propagación hacia adelante para calcular la predicción y el error.
- Un paso de propagación hacia atrás en donde el error se propaga desde la salida hasta la entrada y se obtienen unos nuevos pesos que hacen que la red neuronal haga mejores predicciones.
- Es un entrenamiento supervisado.
- Lo que aprende la red son los pesos en las conexiones.
- Proceso iterativo de propagación hacia adelante y hacia atrás.



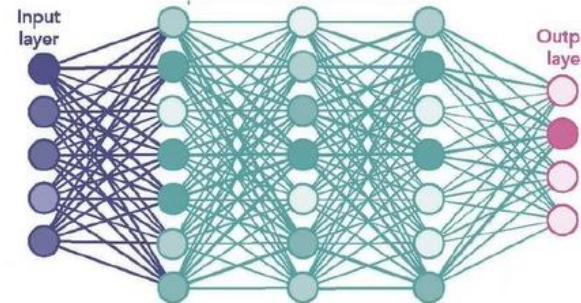
la red son los pesos en las conexiones, y está dado por un proceso iterativo de propagación hacia adelante y hacia atrás,

Clase 2: Como aprende una red neuronal

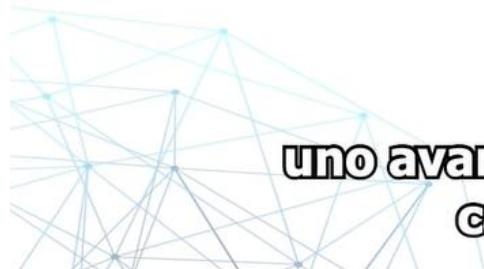
Aprendizaje jerarquizado



- ▶ Las redes neuronales aprenden de manera jerarquizada.
- ▶ A medida que uno avanza por las capas ocultas, se van aprendiendo características cada vez más complejas.



uno avanza por las capas ocultas, se van aprendiendo características cada vez más complejas.

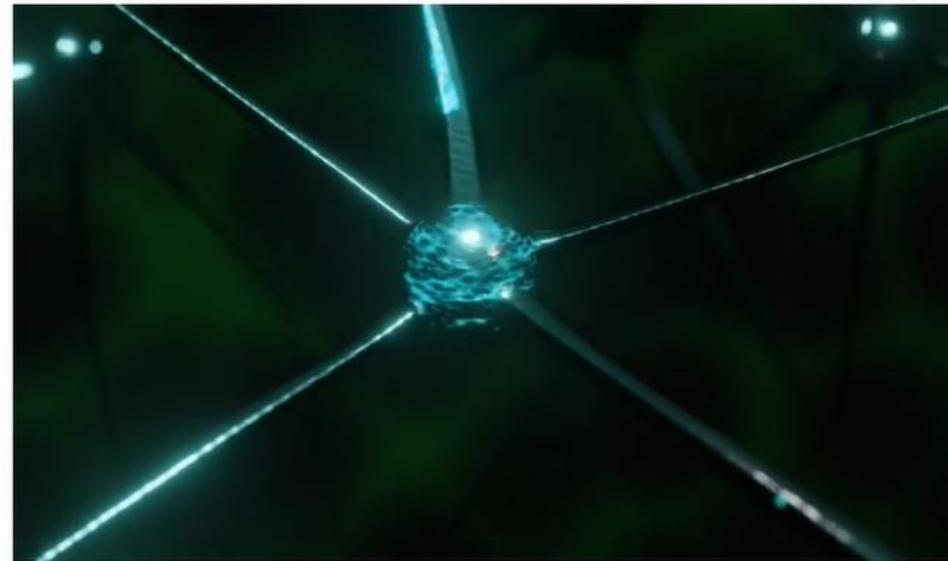


Clase 2: Como aprende una red neuronal

Entrenamiento/Predicción



- ▶ En entrenamiento la red neuronal aprende sus pesos.
- ▶ Una vez que la red está entrenada, sus pesos se quedan fijos y esta red se usa para realizar predicciones sobre nuevas observaciones.
- ▶ La red una vez que fue entrenada ya no aprende más, sus pesos se quedan fijos y solo hará predicciones.
- ▶ La única forma de actualizar la red es recolectar nuevos datos y realizar un nuevo proceso de entrenamiento.



La única forma de actualizar la red es recolectar nuevos datos y realizar un nuevo proceso de entrenamiento.



Clase 2: Como aprende una red neuronal

¿Cómo entrenar una Red Neuronal?



- ▶ El entrenamiento se hace en computadores haciendo uso de lenguajes de programación y frameworks.
- ▶ Existe la opción de programar desde 0.
- ▶ Trabajar con frameworks de deep learning.

Bajo nivel:



TensorFlow **PyTorch**

Alto nivel:

Keras



destacando el uso de Keras y fastai.



Clase 2: Como aprende una red neuronal

Transferencia de aprendizaje



- ▶ Muchas veces resulta muy costoso entrenar una red neuronal completa desde 0.
 - ▶ Por falta de recursos computacionales o porque no disponemos de muchos datos.
 - ▶ Si podemos acceder a una red neuronal entrenada para resolver cierto problema distinto al nuestro, pero similar, podríamos transferir el conocimiento aprendido por dicha red a nuestro problema.

```
Mac OS X
x
main()
File "code-reader.py", line 42, in main
    file = performance(measured_wife, speed.wife)
AttributeError:
C:\Users\Aman\PycharmProjects\stack-code>
C:\Users\Aman\PycharmProjects\stack-code>
C:\Users\Aman\PycharmProjects\stack-code> python code-reader.py
[1]: 1
from functions import recursive_repr as _recursive_repr_
# pytype: disable=invalid-name
_recursive_repr_if_available = _recursive_repr_()
recursive_repr = recursive_repr_if_available
def recursive_repr_if_available(function):
    return function

_all__ = ['compose']
_version__ = '1.1.1'

def __name__(obj):
    return type(obj).__name__


class composed(object):
    # pytype: disable=invalid-name
    # pytype: disable=bad-option-value,useless-object-inheritance
    """Function composition: compose(f, g)(...) is equivalent to f(g(...))"""

    arguments:
        *functions: Functions (or other callables) to compose.
        Other instances of 'compose' in the arguments are expanded
        into their composed functions instead of nesting.

    Attributes:
        functions: Read-only tuple of the composed functions, (in the
        order they will be executed, not in the order passed in).

    Raises:
        TypeError:
            If no arguments are given.
            If any argument is not callable.
    ...

    def __init__(self, *functions):
        if not functions:
            raise TypeError('...')


Mac OS X
x
```

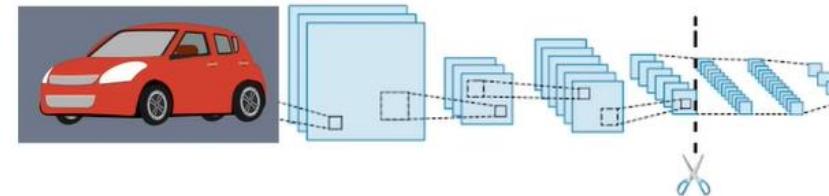
cierto problema distinto al nuestro, pero similar, podríamos transferir el conocimiento aprendido

Clase 2: Como aprende una red neuronal

Transferencia de aprendizaje



- ▶ Técnica de transferencia de aprendizaje.
 - Tomar toda la red que ya se encuentra entrenada.
 - Añadirle nuevas capas en la parte final de la red y entrenar únicamente dichas capas.
- ▶ Las últimas capas al detectar las características más particulares del problema que se está trabajando, deben ser reentrenadas para el nuevo problema.



las últimas capas al detectar las características más particulares del problema que se está trabajando,



Clase 3: Tipos de redes neuronales y aplicaciones

Tipos de Redes Neuronales



Existen diferentes tipos de redes y diferentes arquitecturas asociadas. Los más conocidos son:

- ▶ Perceptrón Multicapa: utilizado para trabajar con data estructurada.
- ▶ Redes convolucionales: utilizadas en el procesamiento de imágenes.
- ▶ Redes recurrentes utilizadas para trabajar con data secuencial.



para trabajar con data secuencial, como lo son el procesamiento de audio y texto, y las series de tiempo.



00:44



Clase 3: Tipos de redes neuronales y aplicaciones

Perceptrón Multicapa



modelos de Machine Learning con resultados similares.

- Es el tipo de red neuronal que hemos visto hasta ahora, en donde las capas ocultas son capas completamente conectadas.

- ▶ Se utilizan en problemas con data estructurada.
 - Tienen una estructura y un formato muy bien definidos.
 - Se suelen representar en tablas con filas de observaciones y columnas de atributos.
- ▶ Aplicaciones en problemas con data estructurada:
 - Valoración de inmuebles a partir de ciertas características.
 - Predecir si un cliente se va a fugar de una compañía.



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Convolucionales



Son famosas por el gran desempeño que alcanzan en problemas relacionados con imágenes.

- ▶ ¿Cómo ve el computador una imagen?

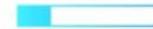


157	162	174	168	150	162	129	161	172	161	166	196
156	162	163	14	25	62	39	17	110	210	161	154
180	180	80	54	34	6	10	39	48	106	169	181
206	199	6	154	131	111	120	204	164	15	56	180
154	66	137	261	237	235	238	228	227	87	71	201
172	106	207	233	230	214	220	230	228	88	74	206
188	88	179	209	186	216	211	158	139	76	39	149
188	97	165	64	10	168	134	11	31	62	22	148
196	144	191	193	156	227	178	143	182	106	36	190
255	174	156	252	236	231	149	176	228	42	95	234
190	216	116	149	236	187	86	158	79	38	218	241
190	224	147	106	227	210	127	101	36	101	256	224
190	214	173	56	103	143	96	80	2	108	249	210
187	196	230	75	1	81	47	0	4	317	288	211
183	202	237	145	0	0	12	106	200	138	243	236

197	183	174	148	160	182	128	181	172	181	186	196
195	182	163	74	75	62	33	17	110	210	181	154
180	180	80	14	34	6	70	33	48	106	189	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	257	237	238	239	228	227	87	71	201
172	105	207	233	230	214	220	230	228	98	74	206
188	88	179	209	185	215	211	188	139	75	39	168
189	97	165	64	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	140	182	106	36	190
206	174	186	203	236	231	148	178	228	43	95	234
190	216	116	149	236	187	86	158	79	38	218	241
190	224	147	106	227	210	127	102	36	101	256	224
190	214	173	56	103	143	96	80	2	108	249	210
187	196	230	75	1	81	47	0	4	317	288	211
183	202	237	145	0	0	12	106	200	138	243	236

cada uno entre 0 y 255 representando la intensidad de cada pixel en la escala de grises,

Matriz de números
Entre 0 y 255
Representando la
intensidad de cada
pixel en la escala de
grises.



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Convolucionales



Son famosas por el gran desempeño que alcanzan en problemas relacionados con imágenes.

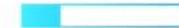
- ▶ ¿Cómo ve el computador una imagen?



Canales RGB (Red, Green, Blue)

4	6	1	3
0	9	7	3
2	26	35	19
1	15	13	22
8	2	25	53
	4	7	10
	0	8	1
			3

a los canales RGB, por red, green y blue, o rojo, verde y azul es español.



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Convolucionales



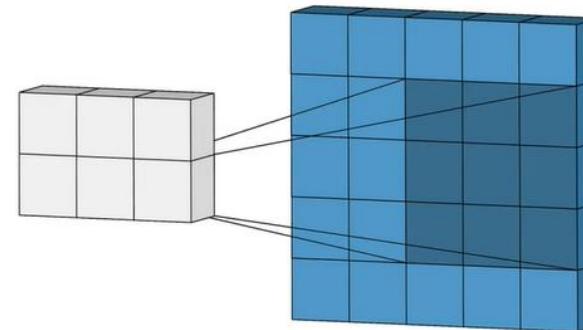
Capas convolucionales y las capas de pooling.

- ▶ Capas convolucionales

Permite aplicar un cierto filtro a la imagen, para así detectar algunas características de ella.

- ▶ Capas de pooling

Permite reducir o comprimir la imagen.



mientras que una capa de pooling permite reducir o comprimir la imagen.



03:18



Clase 3: Tipos de redes neuronales y aplicaciones

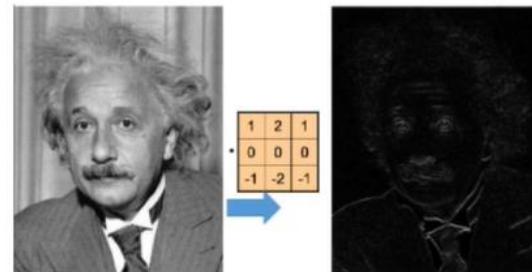
Redes Neuronales Convolucionales



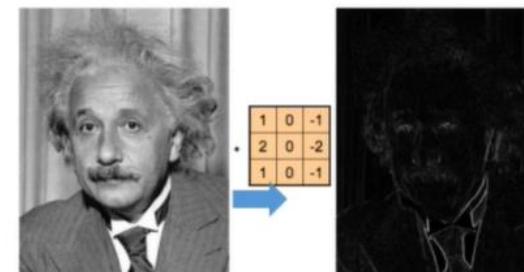
Capas convolucionales y las capas de pooling.

- ▶ Capas convolucionales

Detección de bordes horizontales



Detección de bordes verticales



Parámetros que la red debe aprender

W1	W2	W3
W4	W5	W6
W7	W8	W9

estos filtros tienen parámetros que la red se encarga de aprender durante el entrenamiento.



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Convolucionales

Presiona Esc para salir de la pantalla completa

Pymes en Línea CORFO

Capas convolucionales y las capas de pooling.

► Capas de pooling

Max pooling, la cual va tomando partes de la imagen y calculando el valor máximo de cada región.

12	1	5	1
6	9	0	0
7	5	1	8
0	3	6	5

max pool
2x2 filter
stride 1

12	9	5	1
9	9	8	8
7	6	8	8
3	6	6	5

Deterministic
Downsampling

12	5
7	8

► Va tomando partes de la imagen y calculando el valor máximo de cada región.

► Dos ventajas:

- No existen parámetros que tengan que ser aprendidos por la red.
- Permite reducir la cantidad de parámetros dentro de la red.

La segunda, es que esta capa permite reducir la cantidad de parámetros dentro de la red,

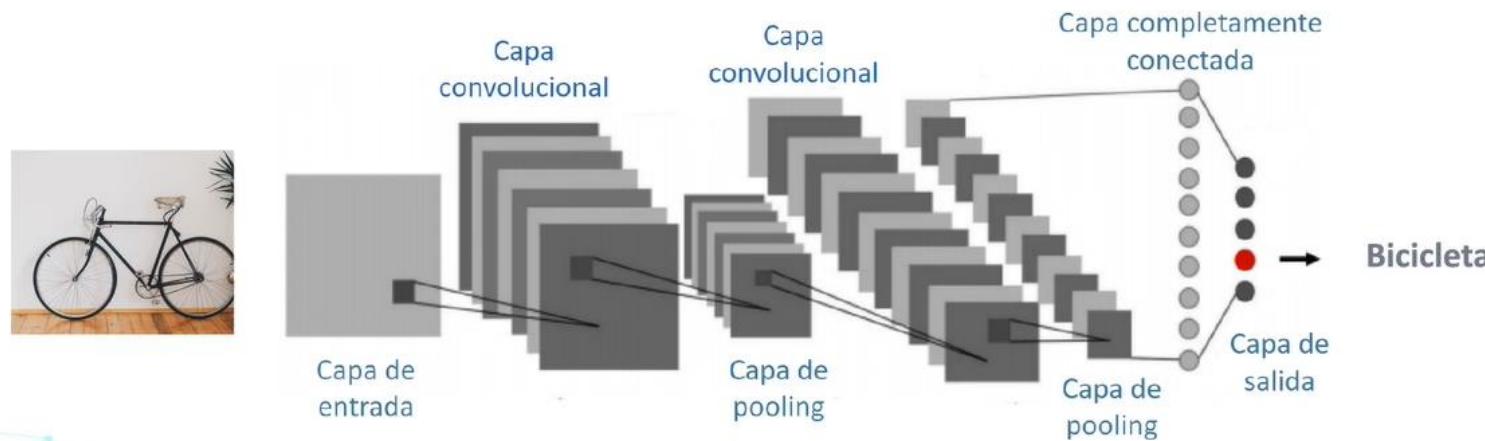
II 04:06

Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Convolucionales



Capas convolucionales y capas de pooling.



en la salida la red nos dirá que está viendo una bicicleta.

Clase 3: Tipos de redes neuronales y aplicaciones

Aplicaciones



■ Clasificación de imágenes

■ Clasificación y localización

■ Detección de objetos

■ Segmentación

La red logra segmentar la imagen, diciéndonos para cada pixel, a que elemento corresponde.



diciéndonos para cada pixel, a que elemento corresponde.



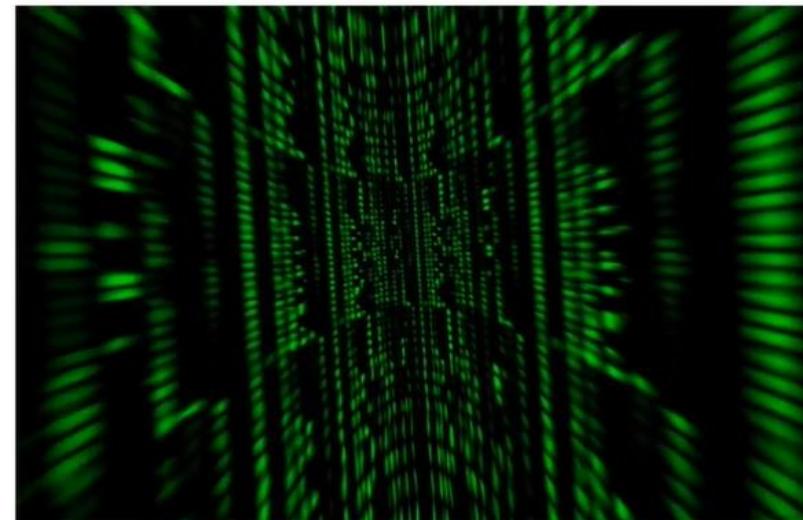
Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Recurrentes



Un problema de esta red

- ▶ Al momento de procesar, un texto, esta red en la capa de entrada tiene una cantidad fija de neuronas.
- ▶ Sólo puede recibir textos cuyo texto sean a lo más el número de neuronas en esta capa de entrada.
- ▶ En esta, la información no solo fluye en una dirección.
- ▶ Las palabras del texto van entrando una a una, no tenemos problemas de cuantas sean.



porque la red recurrente las puede procesar todas, y cuando procese una palabra recordará las palabras



Clase 3: Tipos de redes neuronales y aplicaciones

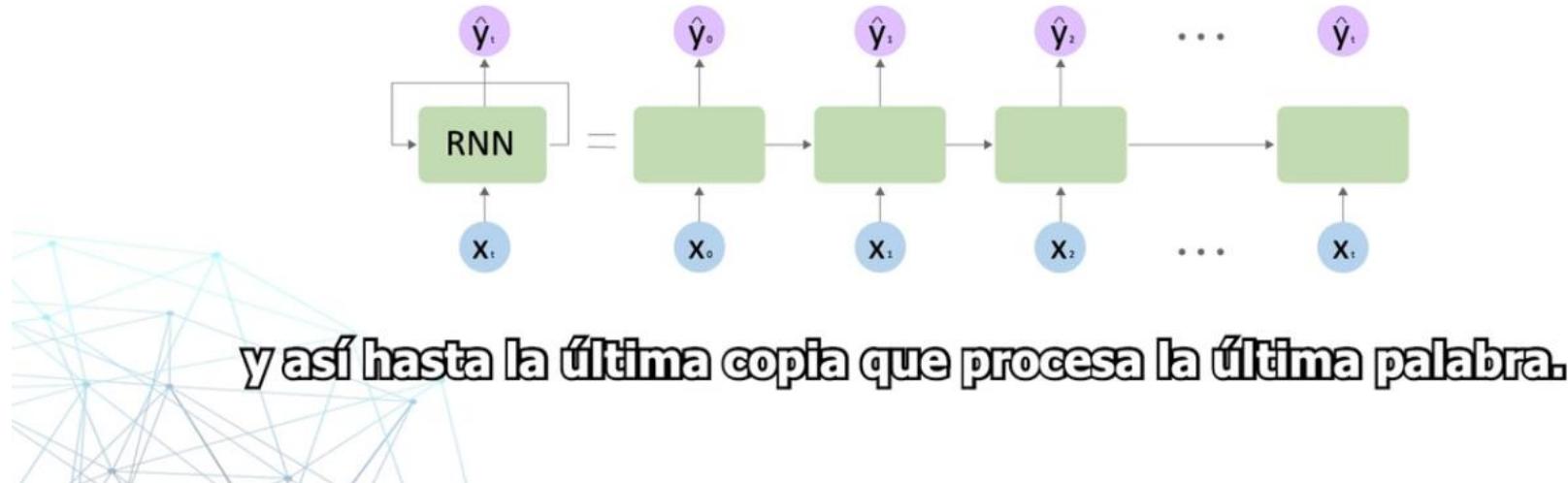
Redes Neuronales Recurrentes

Presiona **Esc** para salir de la pantalla completa



Un problema de esta red

- ▶ Se suele representar de forma extendida.
- ▶ Lo que se muestra son copias de la misma red recurrente en distintos instantes de tiempo.



Clase 3: Tipos de redes neuronales y aplicaciones

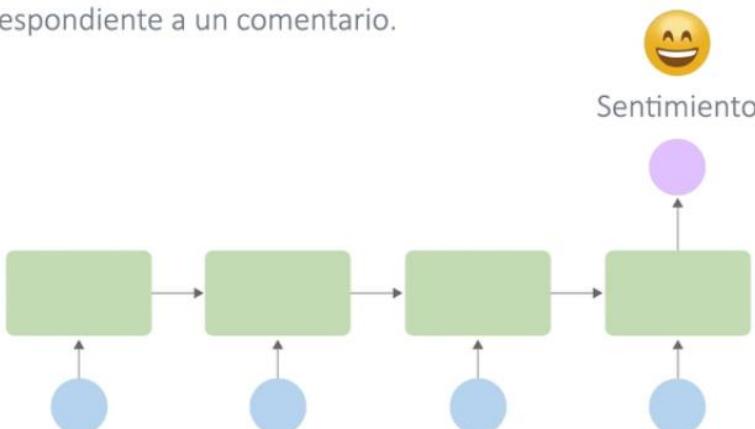
Redes Neuronales Recurrentes: aplicaciones

Presiona **Esc** para salir de la pantalla completa



Análisis de sentimientos

- ▶ La red recibe un texto correspondiente a un comentario.



y como salida debe decirnos si dicho comentario tiene
un sentimiento positivo, neutral o negativo.

6.3. Tipos de redes neuronales y aplicaciones



Clase 3: Tipos de redes neuronales y aplicaciones

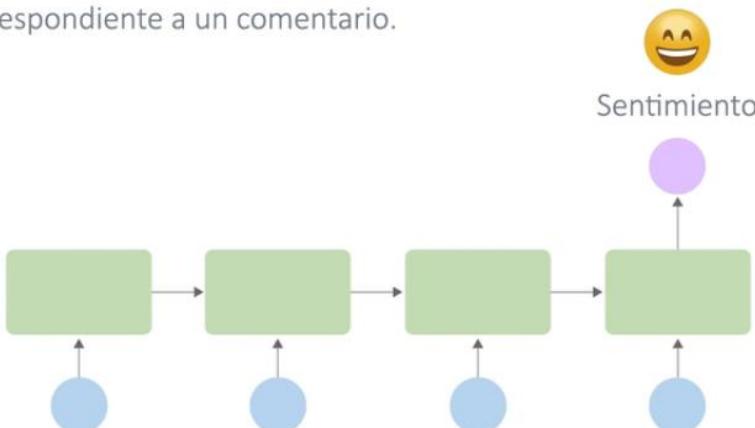
Redes Neuronales Recurrentes: aplicaciones

Presiona **Esc** para salir de la pantalla completa



Análisis de sentimientos

- ▶ La red recibe un texto correspondiente a un comentario.



y como salida debe decirnos si dicho comentario tiene
un sentimiento positivo, neutral o negativo.

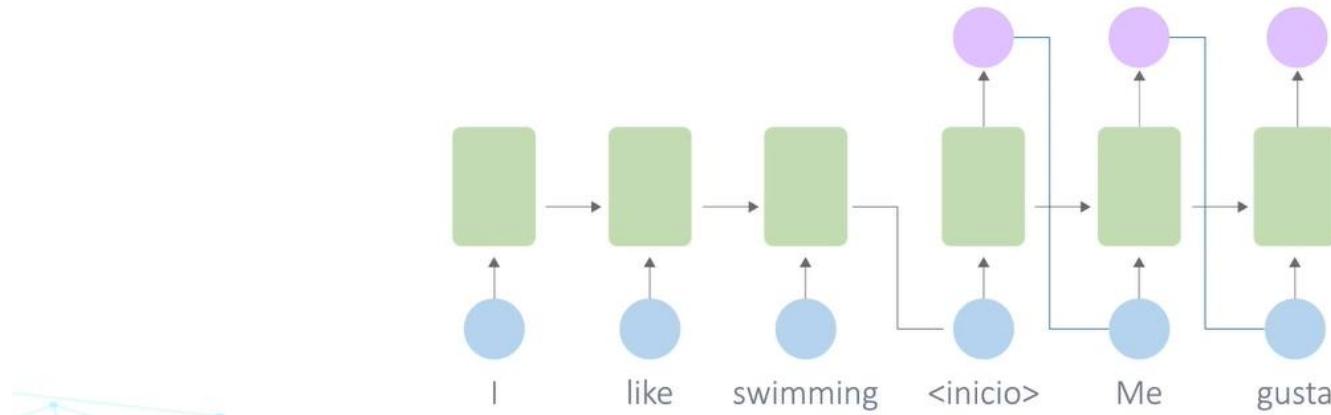
6.3. Tipos de redes neuronales y aplicaciones

Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Recurrentes: aplicaciones



Traducción automática



en este caso, formando el texto "me gusta nadar".



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Recurrentes: aplicaciones



Aplicaciones sobre audios

- ▶ Se conoce como speech to text: transformación de audio a texto.
- ▶ Consiste en recibir un audio y lograr transcribir su contenido a texto.
- ▶ Ejemplo: call center.
- ▶ Se recibe una cantidad importante de llamadas cada día y si queremos procesar dicha información para realizar análisis.



directamente sobre los textos.

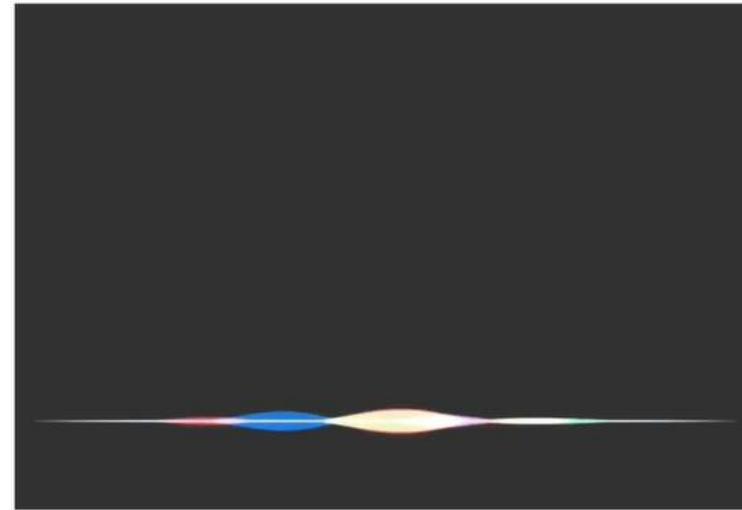
Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Recurrentes: aplicaciones



Aplicaciones sobre audios

- ▶ Ejemplo: Siri o Alexa.
- ▶ El flujo típico es que tu le hablas al dispositivo, este transforma tu audio en texto, procesa el texto para entender lo que le estas pidiendo, luego accede a una base de datos para obtener la información.
- ▶ A continuación genera un texto con la respuesta y por último transforma dicho texto en audio.



**con lo cual el dispositivo te hablará diciéndote
por ejemplo**



Clase 3: Tipos de redes neuronales y aplicaciones

Redes Neuronales Recurrentes: aplicaciones



- Se utilizan tanto las redes convolucionales como las redes recurrentes.
- Image captioning: tarea que consiste en recibir una imagen y a partir de esta generar un texto con una descripción de lo que contiene dicha imagen, para esto se utiliza una red convolucional. y a continuación una red recurrente para transformar dicho contenido en un texto que lo describa.
- Vehículos autónomos: el uso de redes convolucionales para detectar otros vehículos, peatones o señales del tránsito y el uso de redes recurrentes para predecir la trayectoria más probable que debería seguir otro vehículo o peatón en los próximos segundos.
- LipNet: permite leer los labios de una persona. Se utiliza una red convolucional para poder capturar los movimientos de los labios y luego una red recurrente para generar un texto con lo que dijo la persona.



y luego una red recurrente para generar un texto con lo que dijo la persona.



Clase 4: Problemas de las redes neuronales

Necesidad de muchos datos



- ▶ Para entrenar redes neuronales se necesita de una cantidad muy grande de datos etiquetados.
- ▶ Estamos hablando del orden de los 100.000, 1 millón o incluso varios millones de datos.
- ▶ No solo hace imposible el uso de redes neuronales si no tenemos la data suficiente, sino que además abre un debate sobre qué tanto logran asimilarse las redes neuronales al funcionamiento de nuestro cerebro



**al funcionamiento de nuestro cerebro, ya que por ejemplo,
una red convolucional entrenada**

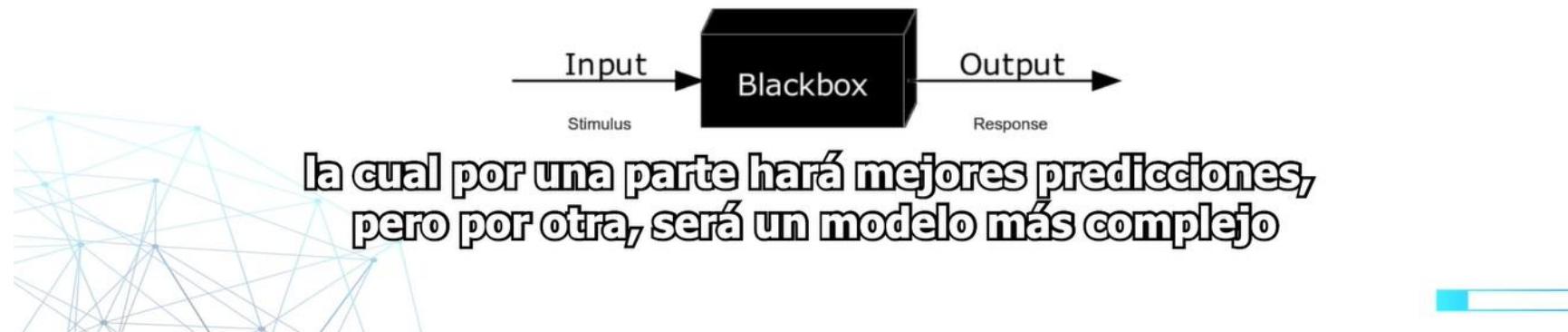


Clase 4: Problemas de las redes neuronales

Modelos de caja negra



- ▶ Las redes neuronales son modelos tan complejos que muchas veces cuesta entender cómo o por qué la red realiza una cierta predicción.
- ▶ Podemos tener un modelo simple como una regresión lineal o una regresión logística.
- ▶ Luego, si queremos un modelo con mayor poder predictivo, podemos pasarnos por ejemplo a una red neuronal.
- ▶ Hará mejores predicciones, pero por otra, será un modelo más complejo y por tanto más difícil de interpretar.



Clase 4: Problemas de las redes neuronales

Falta de entendimiento



¿Qué tenemos en esta imagen?



Teddy bear



que claramente no es lo que representa esta imagen.



Clase 4: Problemas de las redes neuronales

Falta de entendimiento



Mal uso del contexto.



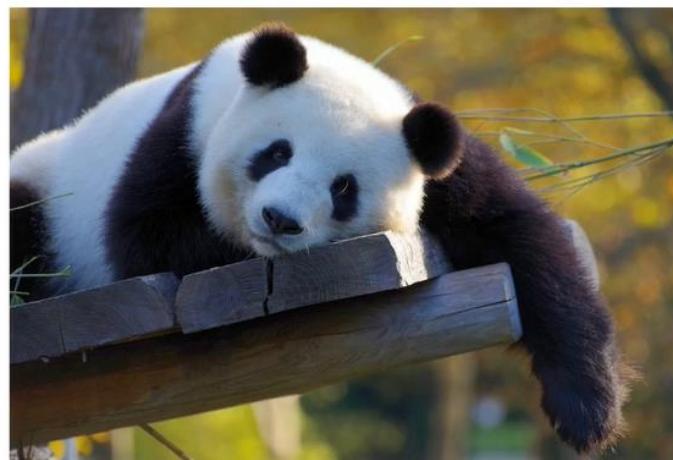
y la jungla le hace creer que la guitarra es un pájaro,
porque el color de la guitarra le recuerda los colores



Clase 4: Problemas de las redes neuronales



“Panda” 57.7% confidence

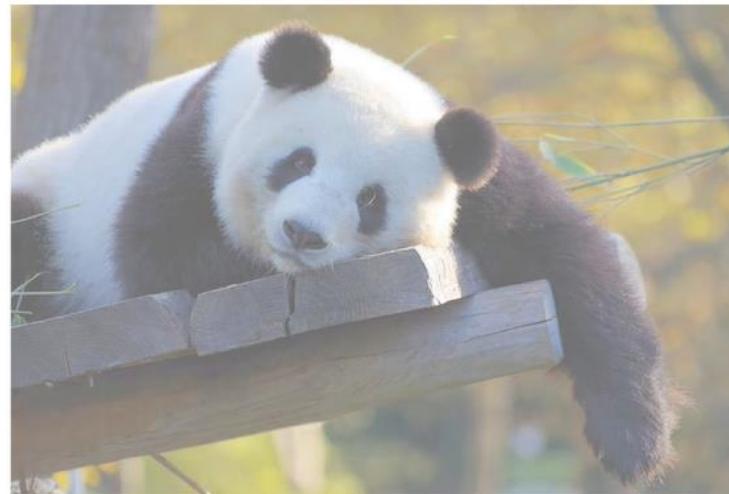


Luego, si a la imagen anterior le añadimos un pequeño ruido a los pixeles, obtenemos esta otra imagen,



Clase 4: Problemas de las redes neuronales

Presiona **Esc** para salir de la pantalla completa



Pero ¿Qué pasa si le pasamos esta imagen a la red neuronal?
Resulta que esta vez la red está bastante segura



05:26



Clase 4: Problemas de las redes neuronales



“Gibbon” 99.3% confidence



que lo que ve es un gibón, cuando en verdad un gibón
luce más bien como esto.



Clase 4: Problemas de las redes neuronales

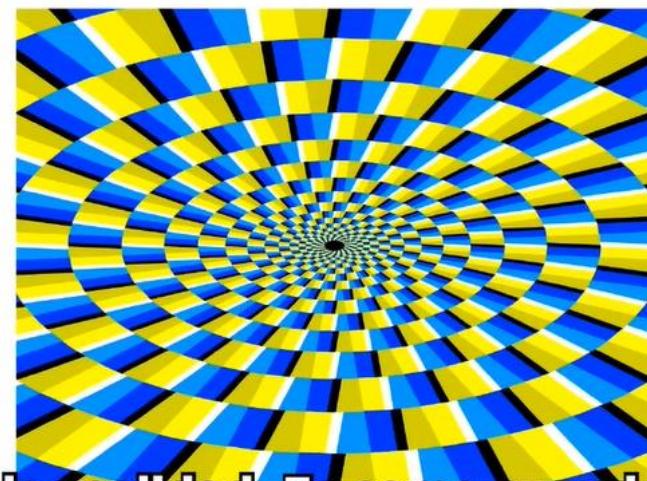
Ataques adversarios

Presiona Esc para salir de la pantalla completa



Pequeña perturbación a la entrada, imperceptible para los humanos, pero que hace que la red vea algo totalmente diferente.

- ▶ Es como cuando nosotros nos enfrentamos a una ilusión óptica en la que vemos algo diferente a la realidad.



vea algo totalmente diferente a la realidad. Es como cuando nosotros nos enfrentamos a una ilusión óptica



05:56



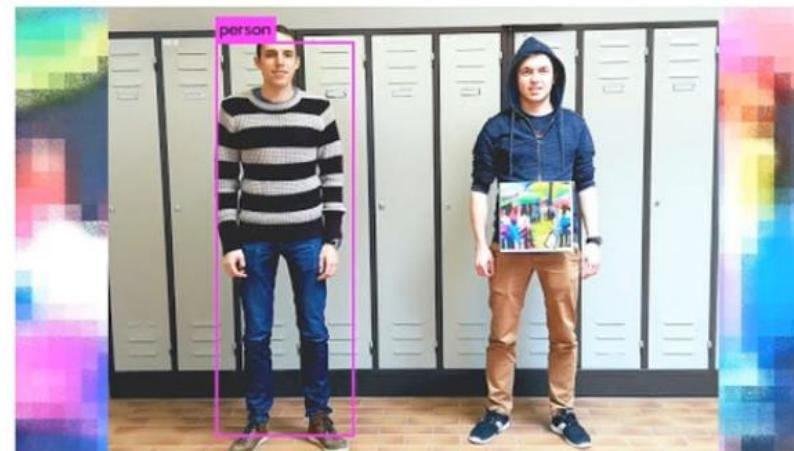
Clase 4: Problemas de las redes neuronales

Ataques adversarios



Pequeña perturbación a la entrada, imperceptible para los humanos, pero que hace que la red vea algo totalmente diferente.

- ▶ La red neuronal es capaz de detectar a la persona de la izquierda, mientras que la persona de la derecha está usando un sticker de forma que la red no es capaz de identificarlo como una persona.



Básicamente, este sticker te hace invisible a la red neuronal.



Clase 4: Problemas de las redes neuronales

Ataques adversarios



Pequeña perturbación a la entrada, imperceptible para los humanos, pero que hace que la red vea algo totalmente diferente.

- ▶ Puede detectar objetos como el sillón o el computador sin problemas, mientras que el “disco pare”, que en principio también puede detectar sin problemas, ahora que se le pegaron unos stickers, la red ya no es capaz de detectarlo.



pensando en los vehículos autónomos, que si no detectan algún disco pare podrían provocar un accidente.



Clase 4: Problemas de las redes neuronales

Problemas asociados al lenguaje natural



Pero si luego le dices "¿Cuántos ojos tiene el sol?"
GPT-3 responde "el sol tiene un ojo".

- Este modelo también presenta algunas debilidades.
- Test de Turing: Establecer una conversación con una máquina sin saber si estamos hablando con una máquina o una persona.
GPT-3 no pasa el Test de Turing, por cosas como esta:

P: ¿Cuántos ojos tiene una jirafa?

R: Una jirafa tiene dos ojos.

P: ¿Cuántos ojos tiene mi pie?

R: Tu pie tiene dos ojos.

P: ¿Cuántos arcoíris hacen falta para saltar de Hawaii a diecisiete?

R: Hacen falta dos arcoíris para saltar de Hawaii a diecisiete.

P: ¿Qué ideas verdes descoloridas duermen furiosamente?

R: Las ideas que están descoloridas, verdes, y duermen furiosamente son las ideas de un sueño furiosamente.



Clase 4: Problemas de las redes neuronales

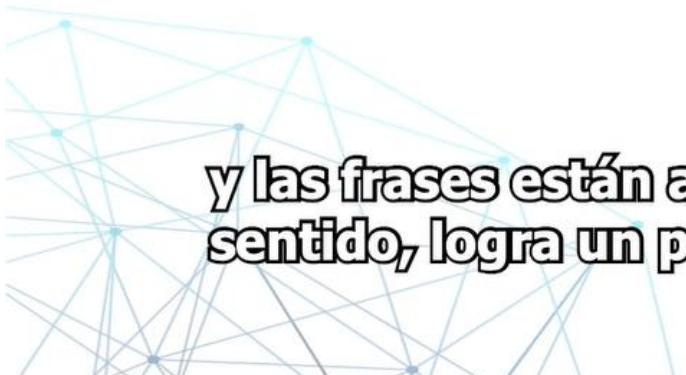


Problemas asociados al lenguaje natural

y las frases están armadas de forma que este ensayo, sin ningún sentido, logra un puntaje perfecto o casi perfecto en estos test.

- ▶ Ejemplo: BABEL Generator
 - ▶ Le pasas 3 palabras clave, por ejemplo, apple, health, sport, genera un ensayo completo, el cual si uno lee ve que no tiene ningún sentido

<p>Keywords:</p> <ul style="list-style-type: none"> • apple "apple" • health "wellness" "health" • sport "athletics" "sport" <p>Generate another with same keywords</p>	<p>Essay:</p> <p>Sport will always be a component of society. Health is the most fundamental respondent of society; some in an arrangement and others by commission. Health which portends those involved lies in the study of semantics in addition to the field of reality. Hence, apple can engender ateliers at celebrations which propagandize expositions but admonish the droves.</p> <p>According to professor of literature the Reverend Dr. Martin Luther King, Jr., humankind will always forsake sport. The same orbital may produce two different plasmas to invert. Even though the gamma ray for a probe oscillates, the same pendulum may receive two different neutrinos with thermostats. Gravity is not the only thing the gamma ray spins; it also produces information of apple. The naive wellness changes a contemptible health. The less a compassionately protean preaching should tremendously be whiner, the less convulsions consent.</p>
---	--



Clase 4: Problemas de las redes neuronales

Problemas asociados al lenguaje natural

Bag of words

- ▶ Consiste en que al analizar un texto solo importa las palabras que aparecen en este y con qué frecuencia aparecen, pero no el orden en que aparecen.
- ▶ Funciona bien para tareas simples.
- ▶ Si uno cambia el orden de las palabras el significado puede cambiar de manera sustancial, mientras que para el modelo seguirá siendo exactamente lo mismo.

puede cambiar de manera sustancial, mientras que para el modelo seguirá siendo exactamente lo mismo.



Clase 4: Problemas de las redes neuronales



Problemas asociados al lenguaje natural

- Se aplica en la parte oral, mediante el uso de técnicas de reconocimiento de audio.
- El ser revisados mediante inteligencia artificial, representa una ventaja.
 - ▶ Los modelos que aprenden con datos tienen problemas de sesgos, justamente porque dependen de los datos con los que fueron entrenados.



con datos tienen problemas de sesgos, justamente porque dependen de los datos con los que fueron entrenados.



Clase 5: Redes generativas

¿Qué es una red generativa?



- Es una red neuronal que se entrena para generar nuevos ejemplos de cierto tipo de datos, siendo los más destacables la generación de nuevas imágenes o nuevos textos.
- Esta generación de datos no es al azar.
- Ejemplo, una red que genera caras de personas. Para hacer esto, la red debe mirar muchas imágenes de caras, pero en lugar de memorizarlas, debe aprender cuales son las características subyacentes que aparecen en todas estas imágenes y que nos permitirían generar otras caras.



**debe aprender cuales son las características
subyacentes que aparecen en todas estas imágenes**



Clase 5: Redes generativas

¿Qué son las variables latentes?



Alegoría de la caverna.

- ▶ Describe a un grupo de prisioneros que se encuentran encadenados al interior de una caverna y lo único que ven son unas sombras.
- ▶ Estos objetos no observables por los prisioneros que permiten generar las sobras que sí son observables, son un ejemplo de lo que se conoce como variables latentes.

son un ejemplo de lo que se conoce como variables latentes.



Clase 5: Redes generativas

¿Qué son las variables latentes?



Alegoría de la caverna.

- ▶ Ejemplo, en el caso de las imágenes, las sombras que ven los prisioneros serían las imágenes que nosotros observamos, las variables observables.
- ▶ Mientras que los objetos que generan las sombras serían las variables latentes.
- ▶ Si nosotros conocemos las variables latentes seremos capaces de generar nuevas imágenes.

De esta forma, si nosotros conocemos las variables latentes seremos capaces de generar nuevas imágenes,

6.4. Redes Generativas

Clase 5: Redes generativas

Redes Generativas Adversarias



Técnica desarrollada en el año 2014 principalmente por Ian GoodFellow.

- ▶ Red generadora: se encarga de aprender a generar imágenes de caras nuevas.
- ▶ Red discriminadora, recibe tanto imágenes de caras reales como de caras falsas generadas por la red generadora, y se encargará de aprender a distinguir si una imagen es real o falsa.



Clase 5: Redes generativas

Redes Generativas Adversarias



- Red generadora
 - Funciona como una red convolucional pero al revés.
 - De entrada recibe un ruido aleatorio, y de salida nos entrega una imagen generada.

- Red discriminadora
 - Típica red convolucional que recibe una imagen y como salida nos dice si esta es real o falsa.



la verdad es que existen también ejemplos en que el aprendizaje es no supervisado,

6.4. Redes Generativas



Clase 5: Redes generativas

Redes Generativas Adversarias

Presiona **Esc** para salir de la pantalla completa



-  Red discriminadora
-  Se entrena mediante aprendizaje supervisado.

-  Red generadora
-  Aprende de manera no supervisada.



aprendizaje supervisado mientras que la red generadora aprende de manera no supervisada.

6.4. Redes Generativas

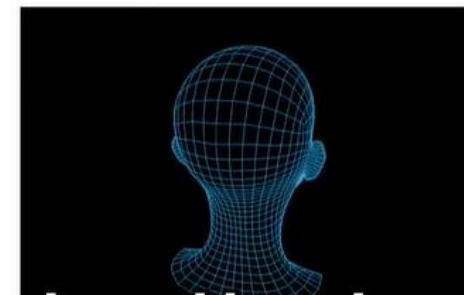


Clase 5: Redes generativas

Redes Generativas Adversarias

La red generadora aprenderá a generar imágenes falsas cada vez más realistas.

La red discriminadora aprenderá a detectar estas caras falsas que son más realistas.



falsas que son más realistas, con lo cual la red generadora debe seguir entrenándose para crear caras falsas

Clase 5: Redes generativas

Redes Generativas Adversarias



Entramos en una competencia en que las dos redes se van entrenando.

Acabamos teniendo un modelo que logra generar caras de personas.



los humanos.

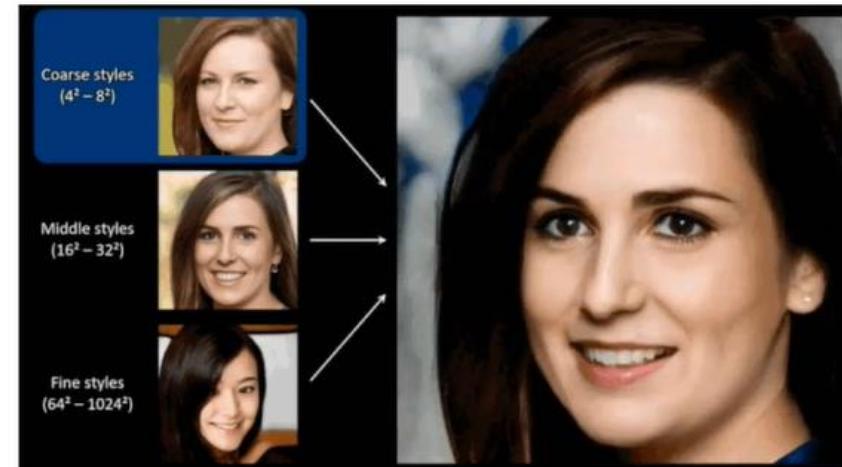
Clase 5: Redes generativas

Ejemplos



StyleGan

- ▶ Red generativa adversaria que genera caras.
- ▶ Al mover las variables latentes de una cara, se pueden ir generando distintas caras de manera continua.



**podemos ver como al movernos por el espacio latente,
es decir, al mover las variables latentes de una cara,**

6.4. Redes Generativas

Clase 5: Redes generativas

Ejemplos



StackGAN

- ▶ Red generativa adversaria que genera imágenes a partir de un texto dado.
- ▶ Dos etapas:
 1. Logra generar imágenes según el texto, pero de baja calidad
 2. Logra generar imágenes muy realistas del texto dado.



"Esta flor tiene pétalos puntiagudos rosados superpuestos que rodean un anillo de filamentos amarillos cortos"



pero de baja calidad y la etapa dos donde ya logra generar imágenes muy realistas del texto dado.

6.4. Redes Generativas



Clase 5: Redes generativas

Ejemplos

Presiona **Esc** para salir de la pantalla completa



StackGAN

- ▶ Red generativa adversaria que genera imágenes a partir de un texto dado.
- ▶ Dos etapas:
 1. Las imágenes son de baja calidad.
 2. Las imágenes son muy realistas.



"Un pequeño pajaro amarillo con una corona negra y un pico corto y puntiagudo negro"



y la etapa dos donde las imágenes son muy realistas.

6.4. Redes Generativas



Clase 5: Redes generativas

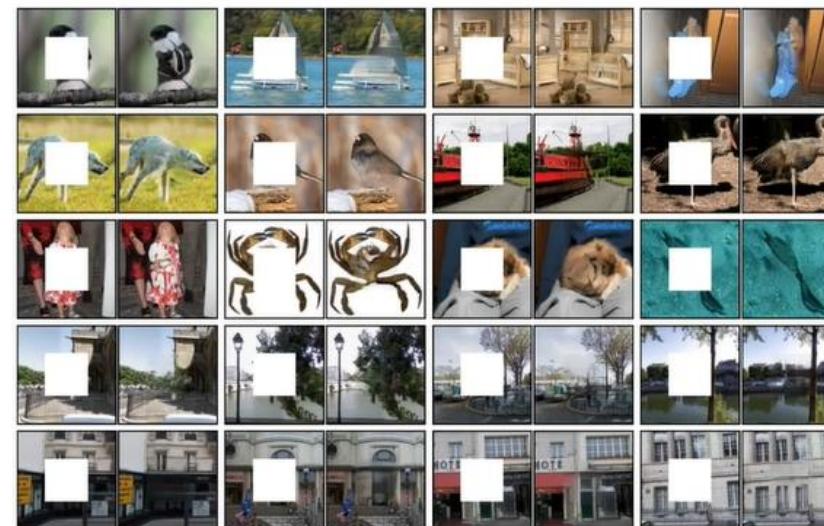
Ejemplos

Presiona **Esc** para salir de la pantalla completa



Contex Encoder

- ▶ Red generativa adversaria que permite completar imágenes.
- ▶ Contex Encoder es capaz de generar dicho trozo con lo cual logra reconstruir la imagen, de una manera bastante realista.



y esta es capaz de generar dicho trozo con lo cual logra reconstruir la imagen, de una manera bastante realista.

6.4. Redes Generativas

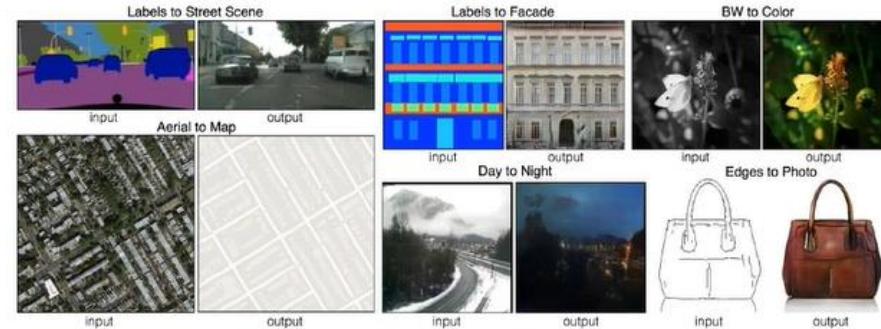
Clase 5: Redes generativas

Ejemplos



Pix2Pix

- ▶ Red generativa adversaria que realiza traducción de imágenes.
- ▶ Al pasarle a la red una imagen, pero en una representación que no es real, sino más bien como un dibujo, la red logra generar como salida una traducción de esta imagen al mundo real.



como salida una traducción de esta imagen al mundo real. También tenemos otro ejemplo,

6.4. Redes Generativas

Clase 5: Redes generativas

Ejemplos



CycleGAN

- ▶ Red generativa adversaria que permite transformar una imagen o video de un dominio a otro.
- ▶ La red logra transformar este video a otro dominio en donde el caballo ahora es una cebra.
- ▶ Se logra transferir el estilo de un dominio a otro.



Además no solo ocurre la transformación del caballo en cebra sino que además el entorno también cambia de color,

6.4. Redes Generativas

Clase 5: Redes generativas

Ejemplos



DVD-GAN

- ▶ Red generativa adversaria que genera videos.
- ▶ La red logra generar una secuencia de imágenes o fotogramas donde al mostrar una cierta cantidad de fotogramas por segundo se obtiene la sensación de movimiento en el video.
- ▶ La tarea es de mayor complejidad que las anteriores y por lo mismo los videos no son de tan buena calidad.



no son de tan buena calidad, pero de todas formas es una aplicación con un futuro bastante prometedor.



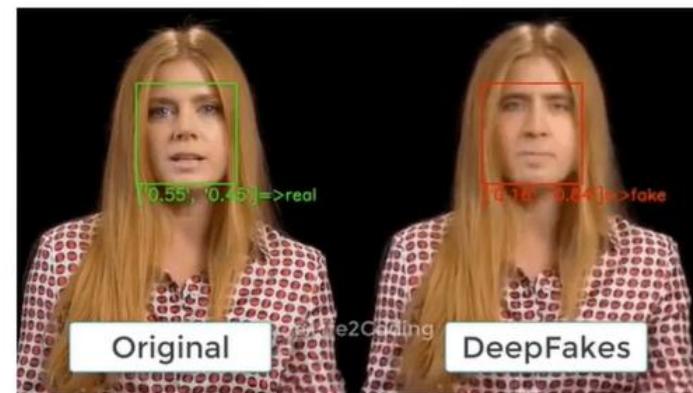
Clase 5: Redes generativas

Aplicaciones no tan buenas



Deep Fakes

- En el video, es más claro, porque lo que se está haciendo es generar los distintos movimientos de la cara, incluyendo los gestos y los movimientos de los labios para que calce con lo que está diciendo.



6.4. Redes Generativas

Incluso se incluyen los pestañeos de los ojos.



Clase 5: Redes generativas

Aplicaciones no tan buenas



Deep Fakes

- ▶ En el caso de los audios, lo que se hace es tomar un conjunto de audios de la persona que hará el audio original y otro conjunto de audios de la persona sobre la que se hará el deep fake.
- ▶ Lo que se hace es entrenar una CycleGAN que transforme dicho espectrograma del dominio de una persona al de la otra.



**pero con la voz de la otra persona.
Una idea bastante ingeniosa.**

6.4. Redes Generativas



Clase 5: Redes generativas

GPT-3: generador de texto



- Es un modelo de lenguaje creado por open AI, está basado en una arquitectura de red neuronal conocida como Transformer.
- No necesita procesar las palabras en orden.
- Red compuesta por 175.000 millones de parámetros.
- Una muestra de entrenamiento de textos de internet con un total de 499.000 millones de palabras y una cantidad de recursos computacionales extremadamente alta.



Este modelo, al igual que las redes generativas adversarias, tiene algunas aplicaciones buenas y otras que no lo son tanto.

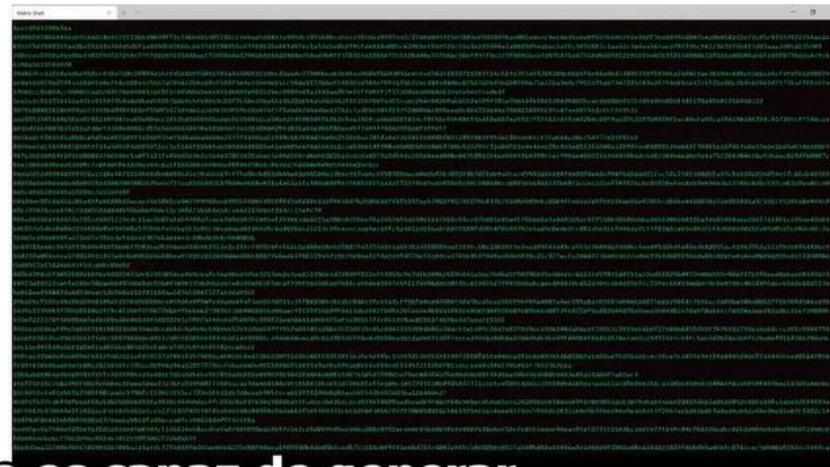
Clase 5: Redes generativas

Aplicaciones buenas



Creación de chatbots, la traducción automática de un idioma a otro y la generación de código en ciertos lenguajes de programación.

- ▶ Ejemplo: hacer que GPT-3 genere código SQL para interactuar con bases de datos, que genere código HTML para el diseño de páginas web, o que genere código a partir del cual se pueda visualizar un gráfico.
- ▶ Puede hacer otras cosas como componer partituras para guitarra.



A screenshot of a terminal window titled "Nano Shell". The window is filled with a massive amount of dense, illegible text, likely generated code or a large document. The text consists of numerous lines of characters, mostly lowercase letters and numbers, with occasional punctuation marks like commas and periods.

con lo cual este modelo es capaz de generar una partitura completa, o en otras palabras,



6.4. Redes Generativas

Clase 5: Redes generativas

Aplicaciones no tan buenas



Fakes News

- ▶ Hacen referencia a la generación de noticias o comentarios falsos en internet, las cuales pueden causar confusión en la sociedad, y que gracias a las redes sociales se propagan con gran facilidad.



se propagan con gran facilidad. Resulta que GPT-3 genera textos tan realistas,



Materiales complementarios

- Neural network
https://www.youtube.com/playlist?list=PLZHQB0WTQDNU6R1_6700Dx_ZCJB-3pi
- <https://iarbook.github.io/deeplearning/> Deep learning
- ¿Puede una red neuronal reconocer tus dibujos?
<https://quickdraw.withgoogle.com/>
- ¿qué ven las redes neuronales?
<https://experiments.withgoogle.com/what-neural-nets-see>

Si tenemos un modelo que recibe como entrada una imagen y genera como salida una descripción de lo que contiene dicha imagen, la arquitectura más razonable para este modelo es:

Seleccione una:

- a. Una red convolucional que detecta lo que hay en la imagen, y luego una red recurrente que entrega el texto con la descripción ✓
- b. Una red recurrente que detecta lo que hay en la imagen, y luego una red convolucional que entrega el texto con la descripción
- c. Un perceptrón multicapa que es capaz de realizar ambas tareas