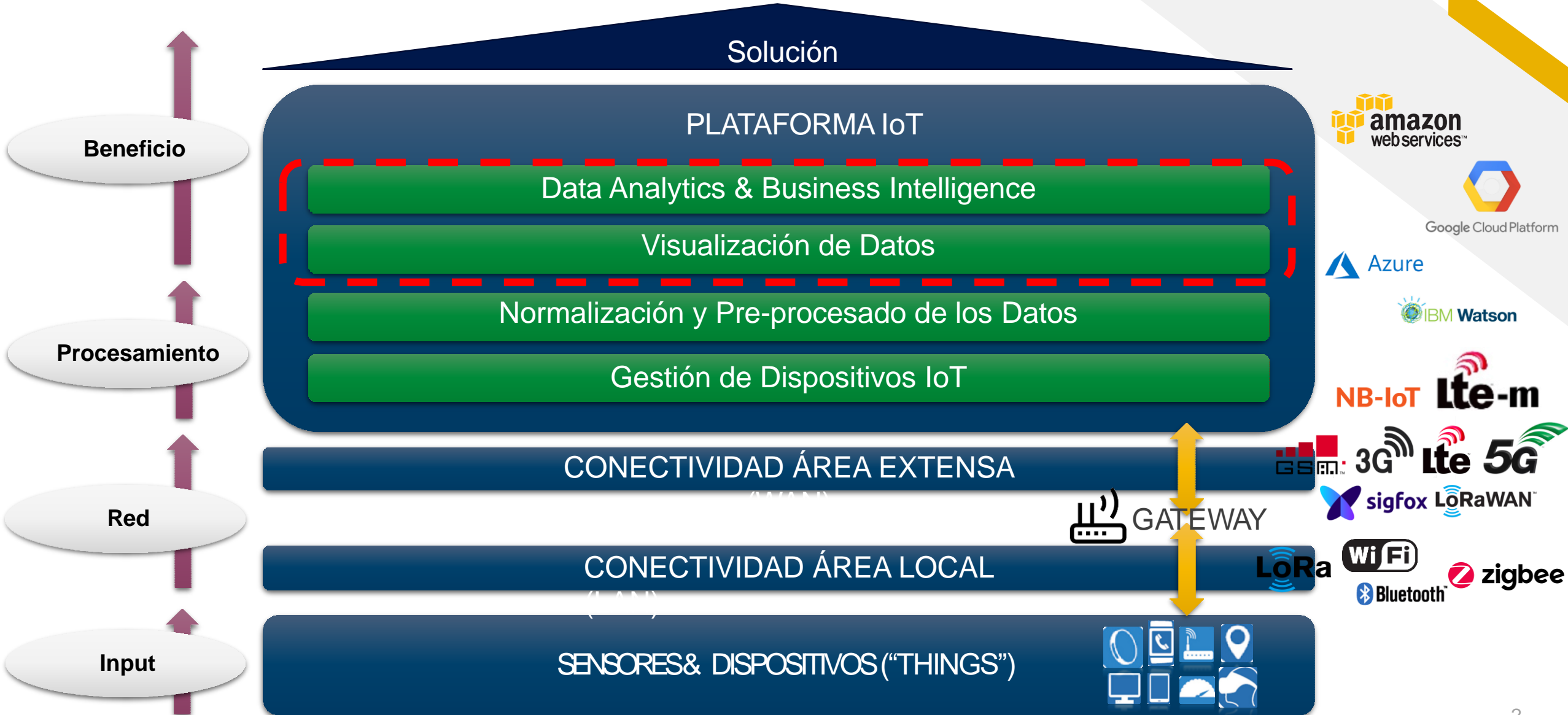


Eduard Lara

Eduard Lara

5.1 Nivel Data Science Plataformas IoT



5.1 ¿Qué es el Data Science y para qué lo necesitamos?

- Vamos a analizar las últimas dos capas de una plataforma IoT
 - La visualización de datos
 - La parte de analítica de datos (data science) y business intelligence.

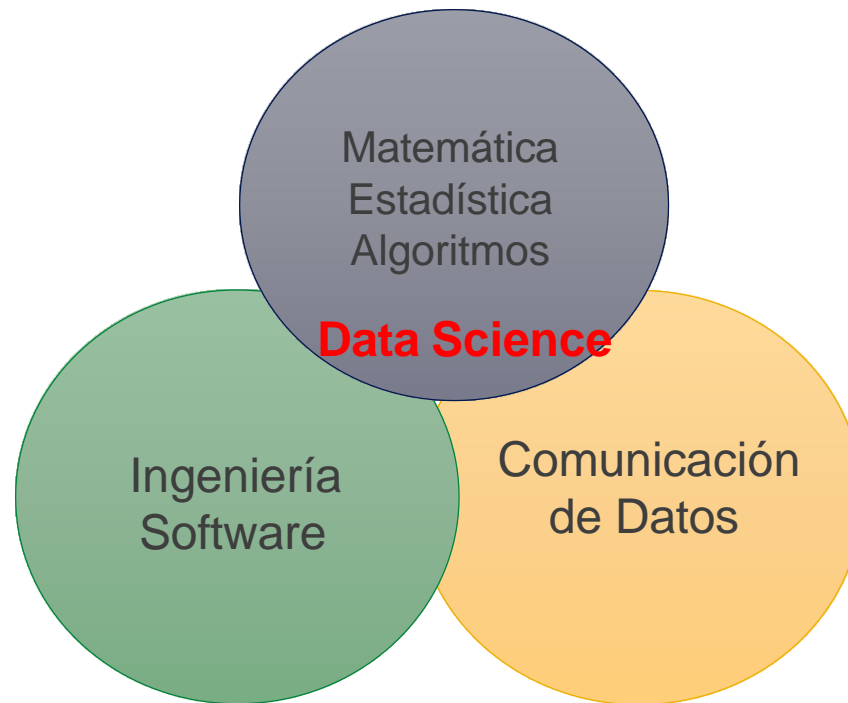
¿Que es la Data Science?

- La ciencia de datos (Data Science) es un campo interdisciplinario que incluye tanto métodos científicos como sistemas para extraer conocimiento y poder entender mejor esos datos en base a la estadística, aprendizaje automático y analítica predictiva.
- La Ciencia de datos (data science) permite explotar todo el valor que nos puede dar los datos tanto los recolectados de nuestras things como del resto de nuestro entorno

Es la ciencia que nos ayuda a responder preguntas a través del análisis de los datos.

5.1 ¿Qué es el Data Science y para qué lo necesitamos?

- Data Science es la mezcla de la matemática estadística con su algoritmo, la ingeniería de software y la comunicación de los datos. La conjunción de todas esas áreas sería el data science



IoT permite conseguir una comprensión sin precedentes tanto de nuestro negocio como de nuestros clientes, gracias a poder recabar toda esa información tan valiosa de todo el despliegue que tengamos de nuestros sensores.

5.1 ¿Qué es el Data Science y para qué lo necesitamos?

- Data Science permite anticiparnos a eventos inesperados, identificar tendencias y construir nuevos tipos de productos que nos den nuevas oportunidades de negocio a partir de los insights obtenidos de los datos
- Todos los nuevos casos de uso que se vayan generando a través de IoT que no sean dispositivos celulares van a tener un incremento del 22% anual
- Esto hace que tengamos un volumen de datos ingente que nos ayuden a la toma de decisiones y sobre todo a explotar todo nuestro negocio y poder conseguir un paso más respecto a nuestra competencia.

5.1 ¿Qué es el Data Science y para qué lo necesitamos?



Pregunta

¿Qué porcentaje de datos de los generados en el día a día en nuestro trabajo, estamos utilizando actualmente?

- Por ejemplo el encendido apagado de las luces de nuestra oficina (dato que probablemente no se esté utilizando).
- Numerosos estudios y encuestas indican que solamente estamos utilizando el 1% de los datos que generamos.
- Estamos perdiendo el otro 99% para poder obtener Insights y sacar conclusiones que puedan aportar un valor a nuestro negocio.

5.1 ¿Qué es el Data Science y para qué lo necesitamos?

Flujo de la pirámide desde que tenemos los datos hasta que realmente tomamos plena conciencia de ellos y los podemos aprovechar para nuestro caso de uso

- **Datos** recolectados de nuestros Internet of Things. Nuestras things nos van a devolver una serie de datos que podemos transformar en información
- **Información** ya serían los datos depurados y tratados de tal manera que no sean una cadena de ceros y unos, sino que ya nos aporten información sobre nuestro entorno y eso lo que estamos detectando y sensorizando
- **Conocimiento:** Haríamos un análisis de toda esa información para poder tener conocimiento, saber interpretar tendencias y poder aprender de todo nuestro entorno gracias a los datos.
- **Inteligencia:** Con ese conocimiento ya podríamos aplicar la inteligencia de negocio necesaria para poder obtener mayor retorno de la inversión y captar el mayor valor posible de toda nuestra solución de IoT

5.2 ¿Qué es el Machine Learning y la Inteligencia Artificial?

El aprendizaje automático o Machine Learning habilita a los computadores aprender en base a la experiencia y el procesamiento de la información recolectada de nuestras things

En qué consiste el machine learning

En la programación tradicional, el funcionamiento habitual de cualquier computadora sería:

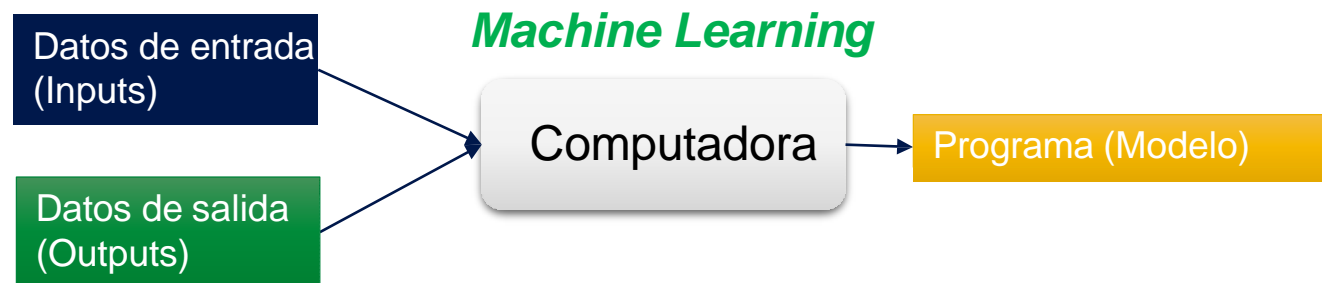
- Por un lado tenemos los datos de entrada que serían los inputs, una serie de valores que parametrizan un programa
- Por otro lado un programa realizado en base a unas funciones y un algoritmo concreto sería lo que entraría en la computadora
- Lo que un programador ha conseguido hacer en base a una lógica y a los datos de entrada, vamos a tener unos datos de salida



5.2 ¿Qué es el Machine Learning / Inteligencia Artificial?

El Machine learning cambia el enfoque totalmente.

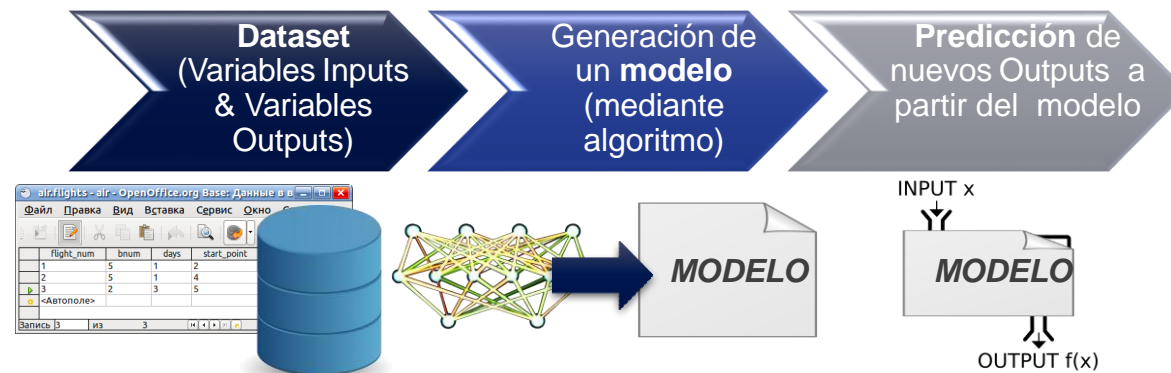
- De partida tenemos los datos de entrada y los datos de salida, el output . Toda esta información la insertaríamos en la computadora.
- La salida de la computadora va a ser el programa o un modelo que hace transformar los datos de entrada en los datos de salida.



5.2 ¿Qué es el Machine Learning / Inteligencia Artificial?

Cambio de enfoque del machine learning

- El DataSet, en Machine learning, estaría formado de las variables de entrada (inputs) y las variables de salida (outputs).
- Este dataset podríamos insertarlo en la computadora para que a través de un algoritmo se genere un modelo. Este algoritmo transforma nuestros datos de entrada en esa salida esperada.
- Como ya tenemos un modelo, la idea es que ahora podamos predecir nuevos outputs a partir de este modelo. Es decir tenemos nuevos datos de entrada, como ya tenemos el modelo generado, cual sería la salida.



5.2 ¿Qué es el Machine Learning / Inteligencia Artificial?

Procedimiento de la inteligencia artificial AI

- En primer lugar **una fase de entrenamiento** es decir generar un DataSet y utilizarlo para entrenar una red neuronal y obtener así el modelo (lo visto anteriormente)
- El siguiente paso sería **una fase de producción** en la que ejecutaría ese modelo a gran escala para intentar llegar a los niveles de rendimiento esperado. Por ejemplo conseguir un 99 por ciento de acierto en la clasificación de un determinado utensilio o un determinado producto
- La siguiente fase ya sería de **aprendizaje activo**, donde no nos vamos a quedar únicamente con el modelo conseguido a través del algoritmo concreto sino que vamos a ir corrigiendo errores de predicción, mejorando el modelo paulatinamente.

5.2 ¿Qué es el Machine Learning / Inteligencia Artificial?

PROCEDIMIENTO INTELIGENCIA ARTIFICIAL

Aprendizaje Activo

Corregir errores de predicción mejorando el modelo paulatinamente

Entrenamiento

Generar el dataset y utilizarlo para entrenar la red neuronal y obtener el modelo

Producción

Ejecutar el modelo a gran escala e intentar llegar los niveles de rendimiento esperados (Ej: 99% acierto)

El flujo constante de la IA, está haciendo llegar a cotas impensables hace unos años de acierto en cuanto a los cálculos tanto de regresiones, clasificaciones y demás modelos de inteligencia artificial y machine learning

5.2 ¿Qué es el Machine Learning / Inteligencia Artificial?

¿Cuáles han sido los eventos clave de la inteligencia artificial durante los últimos años para poder llegar a un porcentaje de acierto elevado en esta materia de IA?

- Los algoritmos se han mejorado. En 2015 se produjo lo que se llama el “crossover point” en el cual el reconocimiento artificial por computador ya llega a mejorar lo que un humano era capaz de predecir gracias a las técnicas de Deep Learning
- Este reconocimiento se refiere tanto a reconocimiento de imágenes como reconocimiento de voz. En 2015 gracias a los algoritmos que se han ido mejorando, un computador viendo una foto o una imagen, era capaz de reconocer mejor que una persona, por ejemplo un determinado animal
- La capacidad de las GPUs, que mejora exponencialmente la velocidad de entrenamiento de los modelos, al trabajar con matrices de datos. Todos los modelos basados en Deep Learning en una red neuronal, necesitan una potencia de cálculo, que hace que al mejorarse esa capacidad podamos obtener mucho antes los modelos que transforman esas variables de entrada en variables de salida.

5.3 Tipos de Machine Learning

Tipos de Machine learning

1) Aprendizaje supervisado.

- Sería permitir a las computadoras aprender a resolver un problema a partir de ejemplos.
- En base a un DataSet de partida que tenemos dividido tanto a la entrada como a la salida, el algoritmo nos va a dar un modelo que en base a esos ejemplos va a aprender a poder mapear o transformar nuevos inputs futuros, o clasificarlos o hacerles una regresión en base a todo este nuevo dataset
- El aprendizaje supervisado se diferencia entre clasificación y regresión



5.3 Tipos de Machine Learning

2) Aprendizaje no supervisado.

- El aprendizaje no supervisado nos ayuda a encontrar patrones ocultos o relaciones en los datos. No requiere de ejemplos concretos para entrenar el modelo sino que simplemente vamos a meter todo nuestro dataset y va a conseguir encontrar esos patrones que nos aporten valor en nuestras soluciones y casos de uso
- Se aplica sobre conjunto de datos no etiquetados y no requiere de un conjunto de entrenamiento. Esto es lo que se diferencia del aprendizaje supervisado
- Dentro del aprendizaje no supervisado estaría el clustering y estarían las reglas de asociación.

Aprendizaje No
Supervisado

Clustering

Reglas de
Asociación

El aprendizaje no supervisado nos ayuda a encontrar **patrones ocultos o relaciones ocultas en los datos.**

Se aplica sobre conjuntos de **datos no etiquetados y no requiere conjunto de entrenamiento.**

5.3 Tipos de Machine Learning

Supervisado - Clasificación

1) SUPERVISADO - CLASIFICACION

- Permite responder a la pregunta: ¿A que categoría pertenece un registro o un individuo?
- **Objetivo:** Entrenar un modelo para que automáticamente a partir de individuos ya etiquetados podamos utilizarlo para nuevos casos.

Ejemplo Supervisado - clasificación

- Tenemos este dataset de entrada que serían dos columnas para las variables input (la longitud y el peso) y la variable output-objetivo (el tipo de pez).
- Esta serie de datos nos dice que si la longitud de un individuo de longitud 50 y peso 12 se trata de un pez de tipo A, uno de longitud 29 y peso 18 es de este tipo B.
- En lugar de solamente 3 registros lo normal es que tengamos millones de registros para poder entrenar el modelo

Dataset entrada

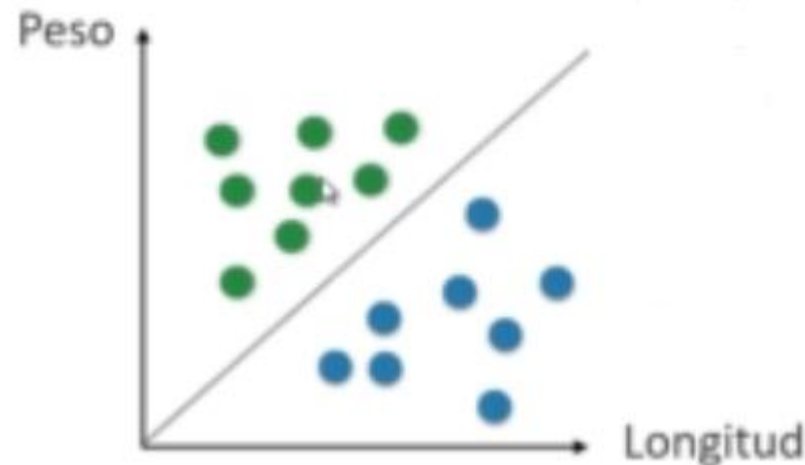
Longitud	Peso	Tipo de Pez
50	12	Pez A
22	18	Pez B
45	13	Pez A

5.3 Tipos de Machine Learning

Supervisado - Clasificación

Modelo clasificación

- En base a este dataset de entrada, dividido entre las variables de entrada y la variable objetivo entrenaríamos un modelo y este modelo en base a un algoritmo lo que nos haría es poder diferenciar o establecer un área o una división dentro del plano, de manera que los que estén de esa división hacia una parte serán de un tipo de pez y los que estén hacia la otra parte serán de otro tipo de pez.
- Estos solamente son un ejemplo con dos variables. Lo normal es que el dataset de entrada sean por ejemplo 30 variables lo que definen la variable objetivo de salida



Modelo Clasificación

5.3 Tipos de Machine Learning

Supervisado - Clasificación

Predicción de nuevos casos

- En base a que ahora ya tenemos un modelo de clasificación, vamos a ser capaces de predecir el tipo de pez cuando tengamos una serie de nuevos valores de longitud y peso.

Longitud	Peso	Tipo de Pez
48	11	?
25	22	?
64	16	?

Predicción de nuevos casos

5.3 Tipos de Machine Learning

Supervisado - Clasificación

Aplicaciones Supervisado Clasificación

- Podemos tener una cámara en mitad del océano (nuestra thing) que esté reportando la longitud y el peso que está detectando de los peces que pasan en frente suya.
- Detección de enfermedades. Si tenemos una base de datos enorme con todos los parámetros y características de los individuos y en base a todo ello tenemos la enfermedad que tiene esa persona concreta, a futuro cuando venga una persona con un determinado patrón de determinadas características y propiedades vamos a poder predecir qué tipo de enfermedad tiene ese paciente.
- Clasificación de imágenes
- Reconocimiento de voz.
- Hay multitud de aplicaciones...

5.3 Tipos de Machine Learning

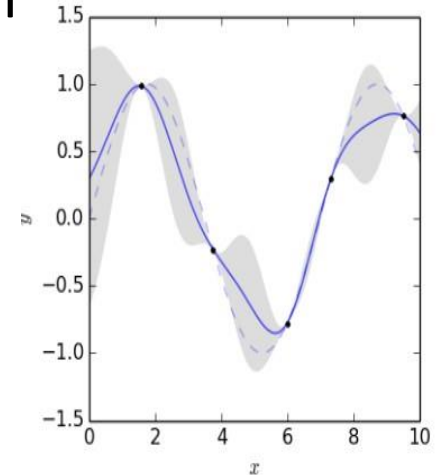
Supervisado - Regresión

2) SUPERVISADO REGRESIÓN

- Nos permite a la pregunta: ¿cuál es el siguiente valor?
- **Objetivo:** Construir un modelo automáticamente a partir de los valores que tenemos actualmente pueda predecir los siguientes valores.

Ejemplo Supervisado - regresión

- En este caso básico tenemos dos variables que en función de sus valores va a ir teniendo una tendencia, una estacionalidad, un comportamiento determinado y lo que queremos con la regresión sería averiguar cuál es el siguiente valor



5.3 Tipos de Machine Learning

Supervisado - Regresión

Aplicaciones Supervisado Regresión

- Previsiones de cotizaciones en bolsa
- previsiones de consumo eléctrico en base a una serie de variables de entrada
- previsión de la oferta y la demanda de determinados productos.

Pensar que podemos tener 20 o 30 variables de entrada y en base a la combinación de todas ellas vamos a tener un modelo entrenado que nos va a permitir a futuro con nuevas variables de entrada predecir cuál es el output cuál es el valor concreto

5.3 Tipos de Machine Learning

No Supervisado – Clustering

3) NO SUPERVISADO – CLUSTERING

- El objetivo es que nos va a permitir identificar estructuras ocultas y patrones en los datos de tal manera que detecte clusters, a partir de centroides que minimicen el error
- No requiere de un conjunto de entrenamiento y trabaja con cualquier tipo de datos. No vamos a tener que entrenar previamente el modelo con un dataset de entrada para poder entrenar y obtener ese modelo y después predecir el futuro sino que directamente le vamos a dar nuestro dataset y el clustering lo que va a hacer es intentar diferenciar dentro de nuestro DataSet conjuntos de datos.

Aplicación

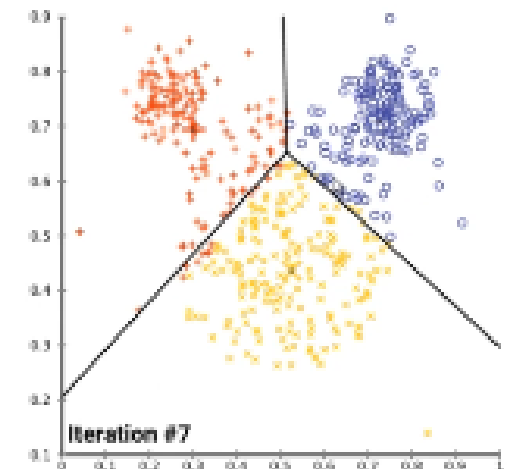
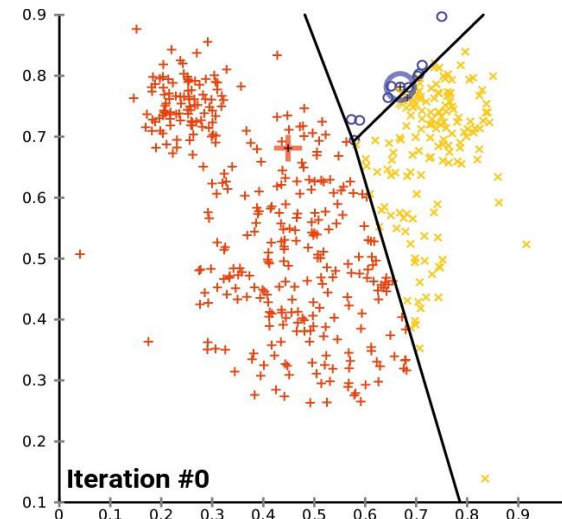
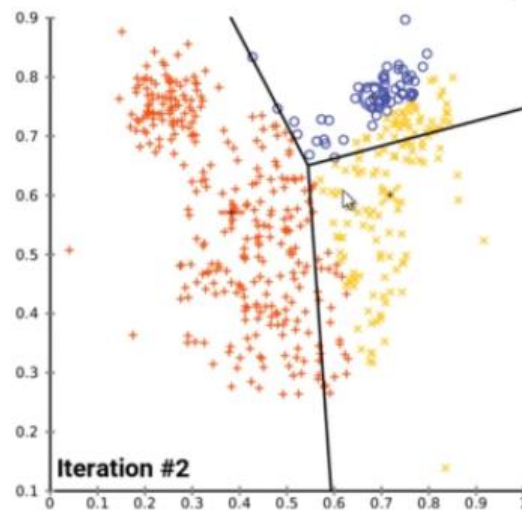
- Segmentación de clientes
- identificación de influencers en redes sociales
- agrupaciones de documentos en base a las etiquetas y palabras que se utilizan
- Tienen muchísimas aplicaciones en la parte clustering.

5.3 Tipos de Machine Learning

No Supervisado – Clustering

Ejemplo

- Tenemos la combinación de dos variables y lo que va haciendo es obtener determinados centroides para conseguir clusterizar todos nuestros datos.
- Por ejemplo si imaginamos que es la combinación de la anchura y la longitud de un pétalo de determinados tipos de flores o plantas. Intenta hacer clústeres de tres diferenciaciones hasta que podamos ver ese patrón gracias a la clusterización



5.3 Tipos de Machine Learning

No Supervisado - Reglas de Asociación

4) NO SUPERVISADO – REGLAS DE ASOCIACION

- **Objetivo:** nos permite identificar individuos o productos que tienen asociaciones entre ellos para poder tomar acciones en base justo a esa asociación y poder obtener un retorno.

Aplicaciones

- Detectar qué productos se compran frecuentemente juntos: Por ejemplo si compro el producto A, ¿Qué probabilidad tengo de comprar también el producto B?
- Recomendaciones de amistad en redes sociales: Si la persona A acepta recomendaciones de amistad de personas amigas de B, ¿Qué probabilidad hay de que vuelva a aceptar nuevas personas amigas de B?
- Hay multitud de aplicaciones en las reglas de la asociación.

5.3 Tipos de Machine Learning

No Supervisado - Reglas de Asociación

4) NO SUPERVISADO – REGLAS DE ASOCIACION

- **Objetivo:** nos permite identificar individuos o productos que tienen asociaciones entre ellos para poder tomar acciones en base justo a esa asociación y poder obtener un retorno.

Aplicaciones

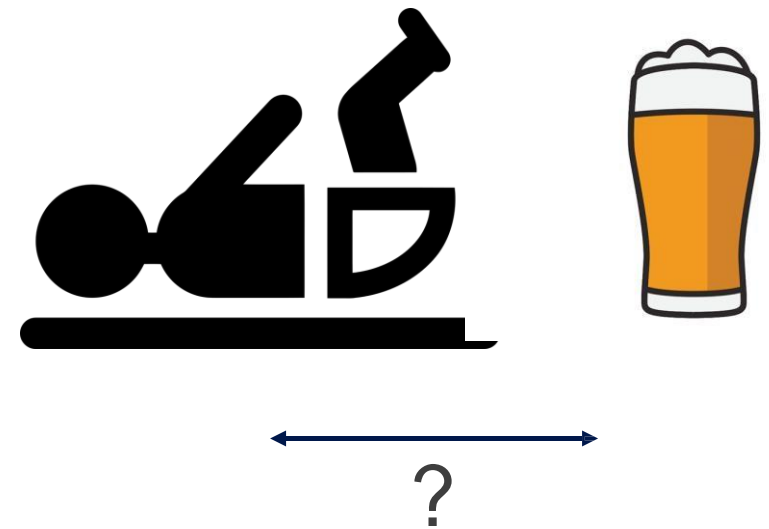
- Detectar qué productos se compran frecuentemente juntos. Por ejemplo si compro el producto A, ¿Qué probabilidad tengo de comprar el producto B?
- Recomendaciones de amistad en redes sociales: Si la persona A acepta recomendaciones de amistad de personas amigas de B, ¿Qué probabilidad hay de que vuelva a aceptar nuevas personas amigas de B?
- Hay multitud de aplicaciones en las reglas de la asociación.

5.3 Tipos de Machine Learning

No Supervisado - Reglas de Asociación

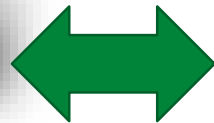
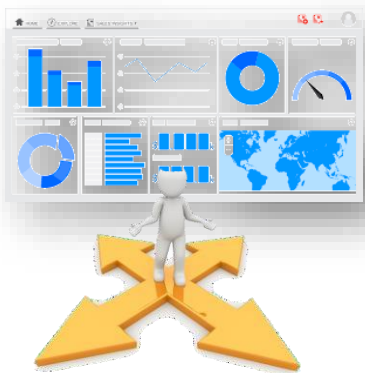
Ejemplo Reglas de asociación

- Pensar primero en estas dos imágenes: que se os puede venir a la mente en cuanto a la asociación? Hay numerosos estudios y algoritmos que han detectado que en base a reglas de asociación cuando un padre o una madre va a comprar pañales para su bebé también se compran cervezas en la tienda en concreto.
- Probablemente ocurre porque esa persona va a estar en casa y no va a poder salir demasiado pues aprovecha y compra cervezas para tomárselos en casa.
- Si fuéramos el dueño de esa tienda, intentaríamos poner juntos aproximadamente los pañales y la cerveza para que así cuando vaya un padre a comprar un producto, intente comprar también el otro a mano para que lo compre.
- Esta es la gracia de la regla de asociación.



5.4 Plataformas de Business Intelligence (POWER BI, QLIK, TABLEAU)

- El término Business Intelligence hace referencia al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa
- Plataformas de Business Intelligence <--> Complemento de módulos Analytics de las Plataformas IoT



5.4 Plataformas de Business Intelligence (POWER BI, QLIK, TABLEAU)



- Integración muy sencilla con plataformas IoT (especialmente Azure)
- Potente integración con múltiples tipos de fuentes BigData (Hadoop, Spark,...)
- Fácil manejo para usuarios con conocimientos de Office
- Actualización mensual
- Precio competitivo



- Herramientas de Analytics muy potentes incluyendo Redes Sociales
- Almacenamiento de datos en la nube elevado y tratamiento ETL



- Interfaz muy intuitiva
- Visualizaciones potentes