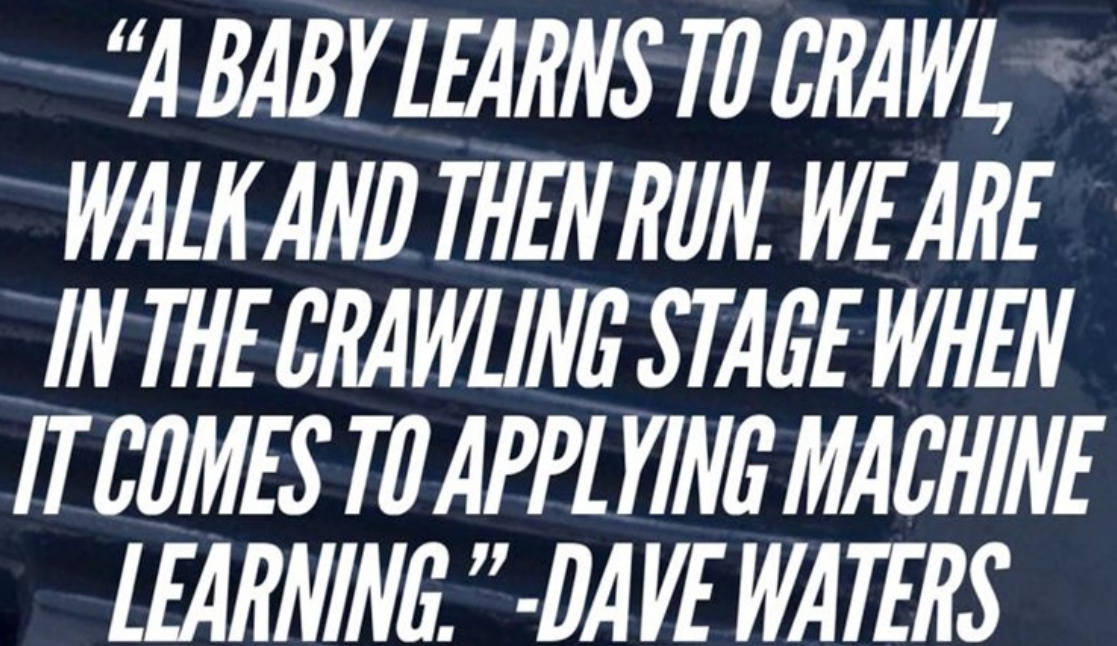


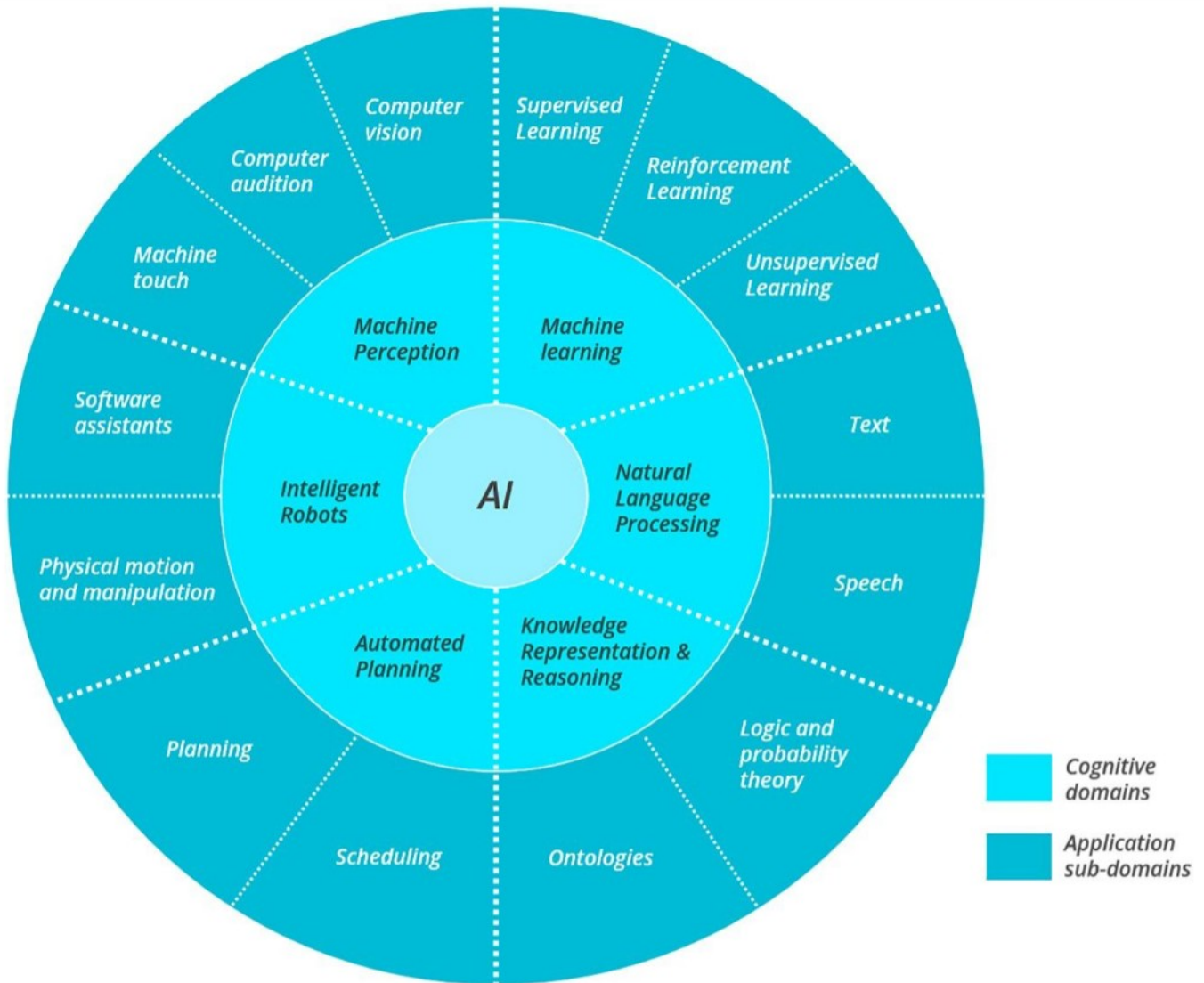
¿Qué es Machine Learning?



***“A BABY LEARNS TO CRAWL,
WALK AND THEN RUN. WE ARE
IN THE CRAWLING STAGE WHEN
IT COMES TO APPLYING MACHINE
LEARNING.” -DAVE WATERS***

¿Qué es Machine Learning?

Actualmente, la inteligencia artificial es una de las tecnologías que están revolucionando la industria en general. Consecuentemente, inteligencia artificial es un área de conocimiento con crecimiento constante, con muchas aplicaciones prácticas y temas de investigación activos. Sin embargo, la inteligencia artificial aún tiene el desafío de resolver problemas que los humanos resuelven intuitivamente a través de su experiencia.



Clasificación de los campos de conocimiento de inteligencia artificial

¿Qué es Machine Learning?

Estudiar todos los campos de conocimiento de la inteligencia artificial puede ser una tarea interminable, debido a que existe una gran variedad y combinación entre estas. Los 6 principales campos de conocimiento de inteligencia artificial son:

- **Machine Learning**

Es una forma de inteligencia artificial que hace predicciones con datos

- **Natural Language Processing NLP**

Es un subcampo de la lingüística, la informática y la inteligencia artificial que se ocupa de las interacciones entre las computadoras y el lenguaje humano.

- **Knowledge Representation and Reasoning**

La representación y el razonamiento del conocimiento (KRR, KR&R, KR²)

Es el campo de la inteligencia artificial (IA) dedicado a representar información sobre el mundo en una forma que un sistema informático puede usar para resolver tareas complejas como diagnosticar una condición médica o tener un diálogo.

- **Automated Planning**

La planificación y programación automatizadas, a veces denotadas simplemente como planificación de IA, es una rama de la inteligencia artificial que se refiere a la realización de estrategias o secuencias de acción, normalmente para que las ejecuten agentes inteligentes, robots autónomos y vehículos no tripulados.

- **Intelligent Robots**

Esta denominación también es utilizada por la norma técnica EN ISO 8373, que define el robot inteligente como “un robot capaz de realizar tareas sondeando su entorno y/o interactuando con fuentes externas y adaptando su comportamiento”.

- **Machine Perception**

La percepción de la máquina es la capacidad de un sistema informático para interpretar datos de una manera similar a la forma en que los humanos usan sus sentidos para relacionarse con el mundo que los rodea

¿Qué es Machine Learning?

¿Qué es Machine Learning?

Machine Learning es uno de los subcampos de la Inteligencia Artificial y puede ser definido como:

“Machine Learning es la ciencia que permite que las computadoras aprendan y actúen como lo hacen los humanos, mejorando su aprendizaje a lo largo del tiempo de una forma autónoma, alimentándolas con datos e información en forma de observaciones e interacciones con el mundo real.” — Dan Fagella

Busca construir algoritmos basados en un conjunto de datos de un determinado fenómeno. Estos datos pueden tener origen en la naturaleza, de actividad humana o generada por otro algoritmo. Es así que los algoritmos de Machine Learning, tienen la capacidad de adquirir conocimientos propios mediante la extracción de características y/o patrones de los datos.

El rendimiento de los algoritmos de Machine Learning depende de la representación de los datos que se proporcionan. Cada pieza de información que se incluye en la representación de nuestro problema se conoce como características. Estas características pueden ser números, vectores, matrices o tensores.

Dado que estas características en su representación más básica terminan siendo números, se puede ir deduciendo que es necesario tener un conocimiento de matemáticas. Es así, que los tópicos básicos que se recomiendan conocer antes de entender algoritmos de Machine Learning son los siguientes:

Álgebra lineal, básicamente hay que entender lo relacionado a vectores, matrices y tensores. Pues es de este modo en cómo se llega a representar los datos.

Estadística, es necesario conocer sobre medidas de tendencia central, distribuciones de probabilidad, análisis de correlación, pruebas de hipótesis, entre otros temas.

Optimización, hay que entender cómo encontrar mínimos y/o máximos de una función objetivo.

¿Qué es Machine Learning?

Programación, pues hay que conocer cómo implementar estos algoritmos. Actualmente, entre los lenguajes de programación que están siendo empleados con mayor frecuencia son Python y R.

Debido al aumento de la capacidad y al abaratamiento de las tecnologías de la información y de los sensores, podemos producir, almacenar y enviar más datos que nunca antes en la historia. De hecho, se calcula que el 90% de los datos disponibles actualmente en el planeta se ha creado en los últimos dos años, produciéndose actualmente en torno a 2,5 quintillones (2.500.000.000.000.000.000) de bytes por día, siguiendo una tendencia fuertemente creciente. Estos datos alimentan los modelos de Machine Learning y son el impulso principal del auge que esta ciencia ha experimentado en los últimos años.

Machine learning ofrece una manera eficiente de capturar el conocimiento mediante la información contenida en los datos, para mejorar de forma gradual el rendimiento de modelos predictivos y tomar decisiones basadas en dichos datos. Se ha convertido en una tecnología con una amplia presencia, y actualmente está presente en: filtros anti-spam para correo electrónico, conducción automática de vehículos o reconocimiento de voz e imágenes.

Terminología Básica y Notaciones

En Machine Learning generalmente se utilizan matrices y notaciones vectoriales para referirnos a los datos, de la siguiente forma:

Cada fila de la matriz es una muestra, observación o dato puntual.

Cada columna es una característica (o atributo), de la observación mencionada en el punto anterior (“feature” en la imagen inferior).

En el caso más general habrá una columna, que llamaremos objetivo, etiqueta o respuesta, y que será el valor que se pretende predecir. (“label” en la imagen inferior.)

← Features →					Label
Position	Experience	Skill	Country	City	Salary (\$)
Developer	0	1	USA	New York	103100
Developer	1	1	USA	New York	104900
Developer	2	1	USA	New York	106800
Developer	3	1	USA	New York	108700
Developer	4	1	USA	New York	110400
Developer	5	1	USA	New York	112300
Developer	6	1	USA	New York	114200
Developer	7	1	USA	New York	116100
Developer	8	1	USA	New York	117800
Developer	9	1	USA	New York	119700
Developer	10	1	USA	New York	121600

Existen algoritmos específicos cuyo propósito es “entrenar” los modelos de Machine Learning. Dichos algoritmos proporcionan datos de entrenamiento que permiten a los modelos aprender de ellos.

¿Qué es Machine Learning?

Con respecto a los algoritmos de Machine Learning, normalmente tienen determinados parámetros “internos”. Por ejemplo en los árboles de decisión, hay parámetros como profundidad máxima del árbol, número de nodos, número de hojas,...a estos parámetros se les llama “hiperparámetros”.

Llamamos “generalización” a la capacidad del modelo para hacer predicciones utilizando nuevos datos.

Definiendo Machine Learning

El Machine Learning -traducido al Español como “Aprendizaje Automático”- es un subcampo de la Inteligencia Artificial que busca resolver el **“cómo construir programas de computadora que mejoran automáticamente adquiriendo experiencia”**.

Esta definición indica que el programa que se crea con ML no necesita que el programador indique explícitamente las reglas que debe seguir para lograr su tarea si no que este mejora automáticamente.

Grandes volúmenes de datos están surgiendo de diversas fuentes en los últimos años y el Aprendizaje Automático relacionado al campo estadístico consiste en **extraer y reconocer patrones y tendencias para comprender qué nos dicen los datos**. Para ello, se vale de algoritmos que pueden procesar Gygas y/o Terabytes y obtener información útil.

Una Definición Técnica

A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

La experiencia E hace referencia a grandes volúmenes de datos recolectados (muchas veces el Big Data) para la toma de decisiones T y la forma de medir su desempeño P para comprobar que mejoran con la adquisición de más experiencia.

El **Machine Learning** es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo).

¿Qué es Machine Learning?

Diagrama de Venn

Drew Conway creó un simpático diagrama de Venn en el que interrelaciona diversos campos.



En esta aproximación al ML, podemos ver que es una intersección entre conocimientos de Matemáticas y Estadística con Habilidades de Hackeo del programador.

Aproximación para programadores

Los programadores sabemos que diversos **algoritmos de búsqueda** pueden tomar mucho tiempo en resolverse y que cuanto mayor sea el campo de búsqueda crecerán potencialmente las posibilidades de combinación de una respuesta óptima, haciendo que los tiempos de respuesta tiendan al infinito o que tomen más tiempo de lo que un ser humano tolerar (por quedarse sin vida o por impaciencia). Para poder resolver este tipo de problemas surgen soluciones de tipo heurísticas que intentan dar “intuición” al camino correcto a tomar para resolver un problema. Estos pueden obtener buenos resultados en tiempos menores de procesamiento, pero muchas veces su intuición es arbitraria y pueden llegar a fallar. **Los algoritmos de ML intentan utilizar menos recursos para “entrenar” grandes volúmenes de datos e ir aprendiendo por sí mismos.**

Entre los Algoritmos más utilizados en Inteligencia Artificial encontramos:

- Árboles de Decisión
- Regresión Lineal
- Regresión Logística
- k Nearest Neighbor
- PCA / Principal Component Analysis
- SVM
- Gaussian Naive Bayes
- K-Means
- Redes Neuronales Artificiales
- Aprendizaje Profundo ó Deep Learning
- Clasificación de imágenes

Podemos subdividir el ML en 2 grandes categorías:

- Aprendizaje Supervisado
- Aprendizaje No Supervisado.

¿Qué es Machine Learning?

Las Redes Neuronales Artificiales

Una mención distintiva merecen las RNAs ya que son algoritmos que utilizan un comportamiento similar a las neuronas humanas y su capacidad de sinopsis para la obtención de resultados, interrelacionándose diversas capas de neuronas para darle mayor poder. Aunque estos códigos existen desde hace más de 70 años, en la última década han evolucionado notoriamente –Breve Historia de las RNA– (en paralelo a la mayor capacidad tecnológica de procesamiento, memoria RAM y disco, la nube, etc) y están logrando impresionantes resultados para analizar textos y síntesis de voz, traducción de idiomas, procesamiento de lenguaje natural, visión artificial, análisis de riesgo, clasificación y predicción y la creación de motores de recomendación.

Aplicaciones del Machine Learning

Para comentar las diversas aplicaciones del ML primero vamos a diferenciar entre el Aprendizaje Supervisado y No Supervisado y el Aprendizaje por Refuerzo.

Aprendizaje Supervisado

En el Aprendizaje Supervisado los datos para el entrenamiento **incluyen la solución deseada, llamada “etiquetas” (labels)**. Un claro ejemplo es al clasificar correo entrante entre Spam o no. Entre las diversas características que queremos entrenar deberemos incluir si es correo basura o no con un 1 o un 0. Otro ejemplo son al predecir valores numéricos por ejemplo precio de vivienda a partir de sus características (metros cuadrados, nº de habitaciones, incluye calefacción, distancia del centro, etc.) y deberemos incluir el precio que averiguamos en nuestro set de datos.

Los algoritmos más utilizados en Aprendizaje Supervisado son:

- **Nearest Neighbors**

¿Dónde se aplica k-Nearest Neighbor? Aunque sencillo, se utiliza en la resolución de multitud de problemas, como en sistemas de recomendación, búsqueda semántica y detección de anomalías.

- **Linear Regression**

La regresión lineal múltiple permite crear modelos que emplean diferentes predictores, los cuales se usarán para dar una respuesta a Y. Podría utilizarse para predecir cuántos litros de gasolina consumirán varios coches, variable de respuesta Y, en función del peso y la potencia que tengan, variable predictiva X

- **Logistic Regression**

El método de regresión logística es un método estadístico que se usa para resolver problemas de clasificación binaria, donde el resultado solo puede ser de naturaleza dicotómica, o sea, solo puede tomar dos valores posibles. Por ejemplo, se puede utilizar para detectar la probabilidad que ocurra un evento

- **Support Vector Machines**

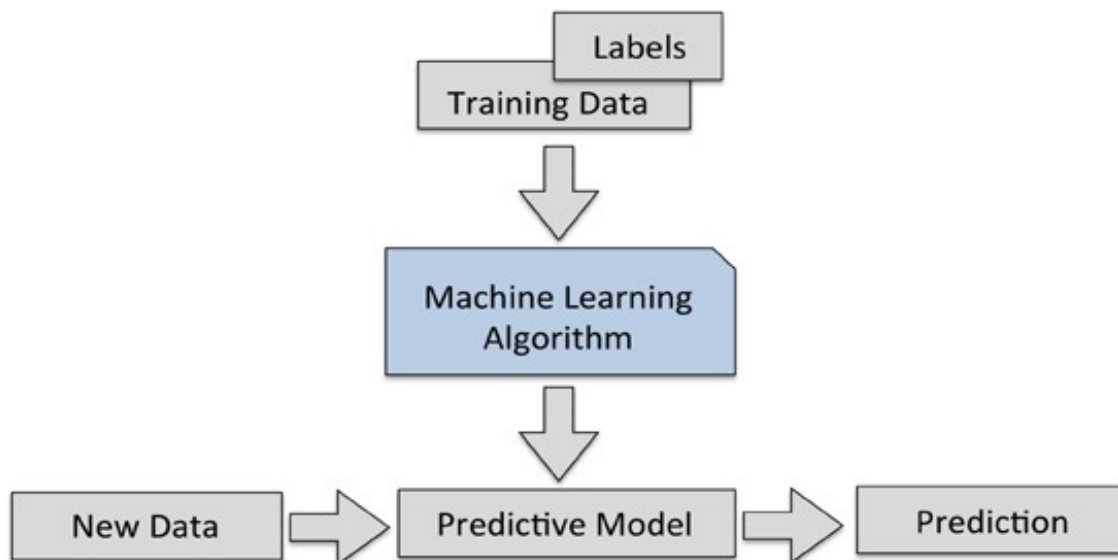
Support vector machine (SVM) es un algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión, incluidas aplicaciones médicas de procesamiento de señales, procesamiento del lenguaje natural y reconocimiento de imágenes y voz.

¿Qué es Machine Learning?

- **Bayesian Classifiers**
- **Decision Tress and Random Forest**
- **Neural Networks**
- **Deep Learning**

Se refiere a un tipo de modelos de Machine Learning que se entrenan con un conjunto de ejemplos en los que los resultados de salida son conocidos. Los modelos aprenden de esos resultados conocidos y realizan ajustes en sus parámetros interiores para adaptarse a los datos de entrada. Una vez el modelo es entrenado adecuadamente, y los parámetros internos son coherentes con los datos de entrada y los resultados de la batería de datos de entrenamiento, el modelo podrá realizar predicciones adecuadas ante nuevos datos no procesados previamente.

De forma gráfica:



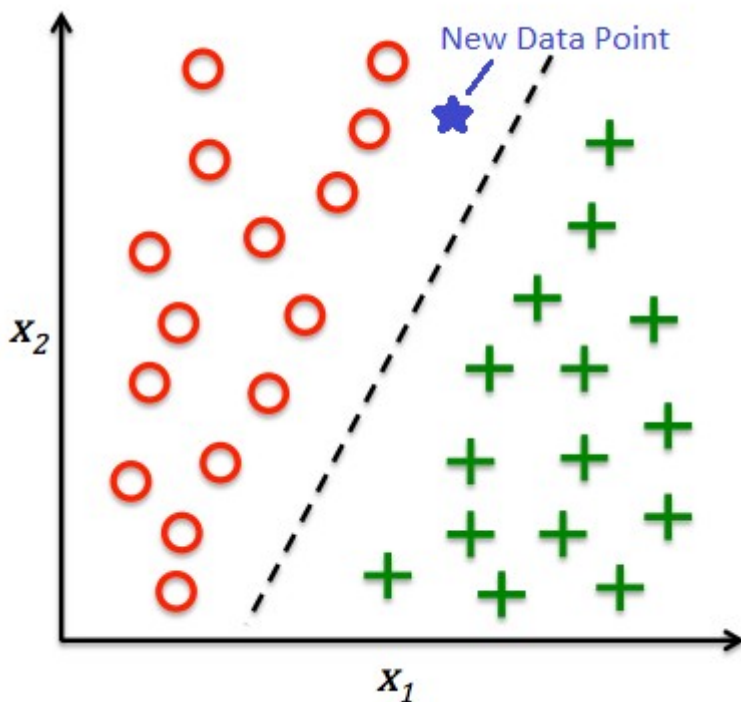
¿Qué es Machine Learning?

Hay dos aplicaciones principales de aprendizaje supervisado: clasificación y regresión:

Clasificación:

Clasificación es una sub-categoría de aprendizaje supervisado en la que el objetivo es predecir las clases categóricas (valores discretos, no ordenados, pertenencia a grupos). El ejemplo típico es la detección de correo spam, que es una clasificación binaria (un email es spam — valor “1”- o no lo es — valor “0” -). También hay clasificación multi-clase, como el reconocimiento de caracteres escritos a mano (donde las clases van de 0 a 9).

Un ejemplo de clasificación binaria: hay dos clases de objetos, círculos y cruces, y dos características de los objetos, X_1 y X_2 . El modelo puede encontrar las relaciones entre las características de cada punto de datos y su clase, y establecer la línea divisoria entre ellos. Así, al ser alimentado con nuevos datos, el modelo será capaz de determinar la clase a la que pertenecen, de acuerdo con sus características.



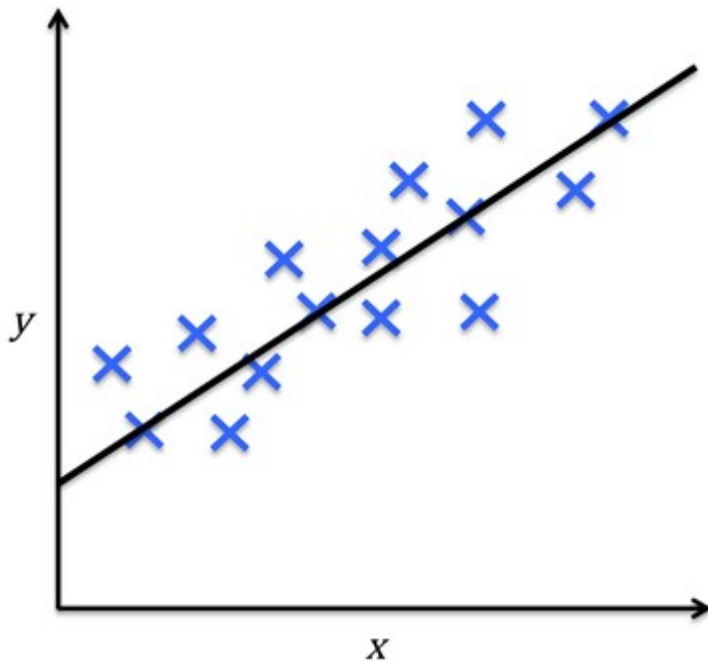
En este caso, el nuevo punto de datos entra en el área correspondiente al subespacio de círculos y por tanto, el modelo predecirá que la clase del objeto es círculo.

¿Qué es Machine Learning?

Regresión:

La regresión se utiliza para asignar categorías a datos sin etiquetar. En este tipo de aprendizaje tenemos un número de variables predictoras (explicativas) y una variable de respuesta continua (resultado), y se tratará de encontrar una relación entre dichas variables que nos proporcione un resultado continuo.

Un ejemplo de regresión lineal: dados X e Y , establecemos una línea recta que minimice la distancia (con el método de mínimos cuadrados) entre los puntos de muestra y la línea ajustada. Después, utilizaremos las desviaciones obtenidas en la formación de la línea para predecir nuevos datos de salida.



Aprendizaje No Supervisado

En el aprendizaje No Supervisado los datos de entrenamiento no incluyen Etiquetas y el algoritmo intentará clasificar o descifrar la información por sí solo. Un ejemplo en el que se usa es para agrupar la información recolectada sobre usuarios en una Web o en una app y que nuestra Inteligencia detecte diversas características que tienen en común.

Los algoritmos más importantes de Aprendizaje No supervisado son:

- Clustering K-Means (ejercicio paso a paso en español)
- Principal Component Analysis
- Anomaly Detection

En el aprendizaje no supervisado, trataremos con datos sin etiquetar cuya estructura es desconocida. El objetivo será la extracción de información significativa, sin la referencia de variables de salida conocidas, y mediante la exploración de la estructura de dichos datos sin etiquetar.

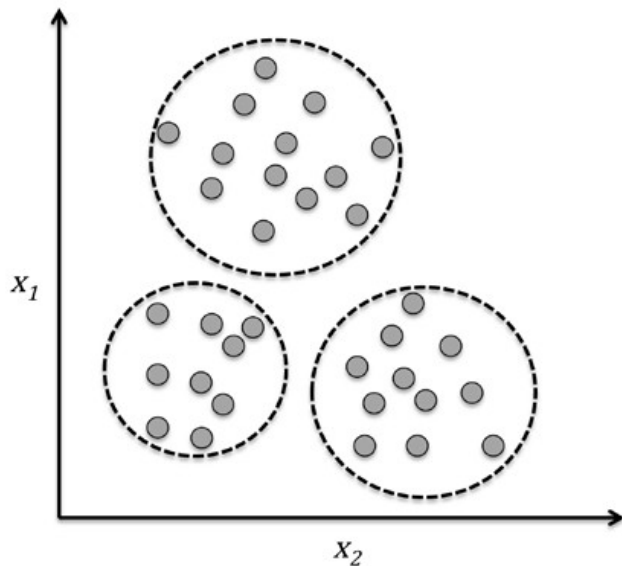
¿Qué es Machine Learning?

Hay dos categorías principales: agrupamiento y reducción dimensional.

Agrupamiento ó Clustering:

El agrupamiento es una técnica exploratoria de análisis de datos, que se usa para organizar información en grupos con significado sin tener conocimiento previo de su estructura. Cada grupo es un conjunto de objetos similares que se diferencia de los objetos de otros grupos. El objetivo es obtener un numero de grupos de características similares.

Un ejemplo de aplicación de este tipo de algoritmos puede ser para establecer tipos de consumidores en función de sus hábitos de compra, para poder realizar técnicas de marketing efectivas y “personalizadas”.

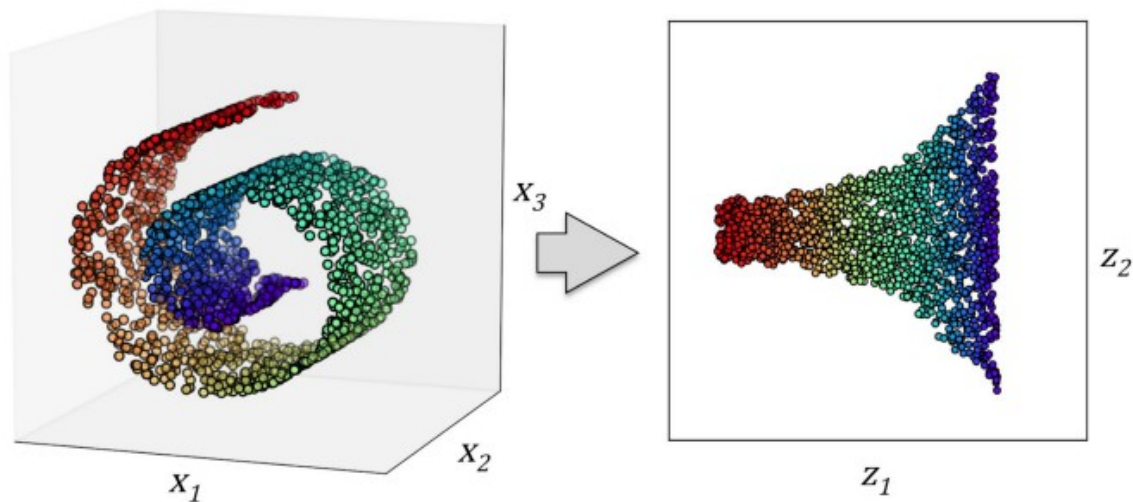


¿Qué es Machine Learning?

Reducción dimensional:

Es común trabajar con datos en los que cada observación se presenta con alto número de características, en otras palabras, que tienen alta dimensionalidad. Este hecho es un reto para la capacidad de procesamiento y el rendimiento computacional de los algoritmos de Machine Learning. La reducción dimensional es una de las técnicas usadas para mitigar este efecto.

La reducción dimensional funciona encontrando correlaciones entre las características, lo que implica que existe información redundante, ya que alguna característica puede explicarse parcialmente con otras (por ejemplo, puede existir dependencia lineal). Estas técnicas eliminan “ruido” de los datos (que puede también empeorar el comportamiento del modelo), y comprimen los datos en un sub-espacio más reducido, al tiempo que retienen la mayoría de la información relevante.



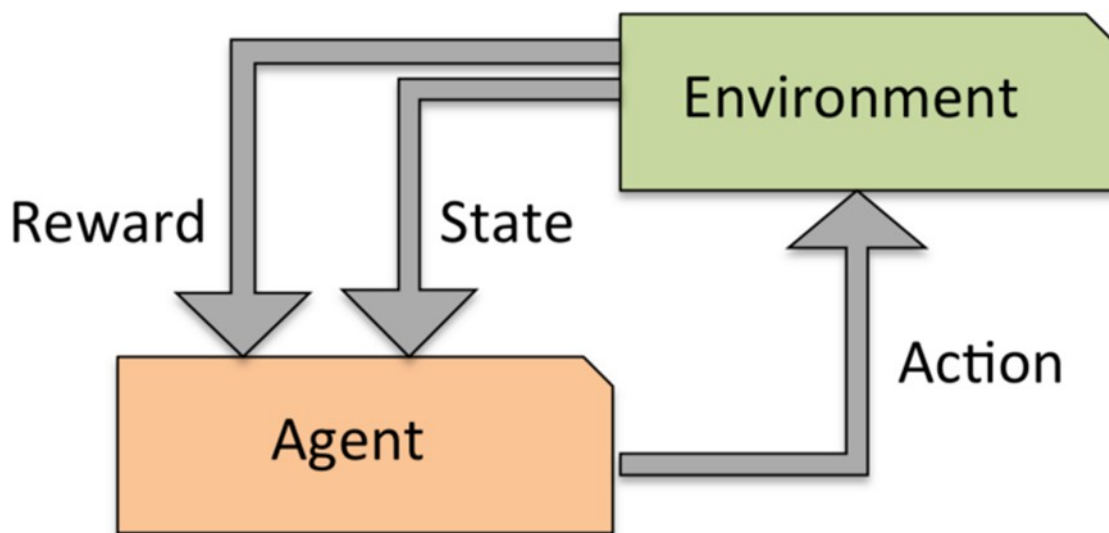
¿Qué es Machine Learning?

Aprendizaje por Refuerzo

Nuestro sistema será un “agente autónomo” que deberá explorar “un espacio” desconocido y determinar las acciones a llevar a cabo mediante prueba y error. Aprenderá por sí mismo obteniendo premios -recompensas- y penalidades la forma óptima para recorrer un camino, resolver un puzzle o comportarse por ejemplo en el Pac Man o en el Flappy Bird. Creará la mejor estrategia posible (políticas) para obtener la mayor recompensa posible en tiempo y forma. Las políticas definirán qué acciones tomar ante cada situación a la que se enfrente.

Los modelos más utilizados son:

- Q-Learning basado en Procesos de Decisión de Markov (MDP: Markov Decision Process)



El aprendizaje reforzado es una de las ramas más importantes del aprendizaje profundo. El objetivo es construir un modelo con un agente que mejora su rendimiento, basándose en la recompensa obtenida del entorno con cada interacción que se realiza. La recompensa es una medida de lo correcta que ha sido una acción para obtener un objetivo determinado. El agente utiliza esta recompensa para ajustar su comportamiento futuro, con el objetivo de obtener la recompensa máxima.

Un ejemplo común es una máquina de ajedrez, donde el agente decide entre una serie de posibles acciones, dependiendo de la disposición del tablero (que es el estado del entorno) y la recompensa se recibe según el resultado de la partida.

Aplicaciones de ML y Ejemplos

Las aplicaciones más frecuentes del Machine Learning son:

Reconocimiento de Imágenes

Se utilizan diversos algoritmos para poder comprender imágenes sobre todo Deep Learning), encontrar algo en particular o agrupar zonas. Sabemos que las imágenes son conjuntos de píxeles continuos y cada uno contiene información del color que tiene que “iluminar” (por ejemplo RGB). Los ejemplos comunes son los de OCR Reconocimiento de Caracteres ópticos, es decir, encontrar letras, agruparlas, encontrar espacios y poder descifrar textos y el otro ejemplo es el de detectar personas en imágenes, presencia humana en cámaras de seguridad o más recientemente se utilizan redes neuronales convolucionales para detectar rostros: **reconocimiento facial**, seguramente más de una vez Facebook fue capaz de localizar a tus primos y hermanos en tus álbumes de fotos.

Reconocimiento de Voz

A partir de las ondas de sonido sintetizadas por el micrófono de tu ordenador, smartphone o de tu coche, los algoritmos de Machine Learning son capaces de limpiar ruido, intuir los silencios entre palabras y comprender tu idioma para interpretar tus ordenes, ya sea “Siri, Agregar un Recordatorio para el lunes que viene” o “Ok Google, Poner música de Coldplay” o hasta hacer reservas y pedir pizza. Parte del reconocimiento se hace mediante el Procesamiento del Lenguaje Natural (NLP en inglés).

Clasificación

Consiste en identificar a que Clase pertenece cada individuo de la población que estamos analizando. Se le asignará un valor discreto de tipo 1, 0 como en la clasificación de Spam o no. También podría ser clasificar si un tumor es benigno o no o al clasificar flores según las características obtenidas.

Predicción

Similar a la clasificación pero para valores continuos, nos permite intentar predecir qué valor obtendremos dado un conjunto de datos de entrada con resultado desconocido. Como comentábamos antes, se puede utilizar para predecir precios de inmuebles, alquileres, coches o la probabilidad de que ocurra algún evento utilizado con frecuencia en estadística con Regresión Lineal. Pero si lo que queremos hacer es Pronóstico de Series temporales, será mejor utilizar redes neuronales u otros modelos estadísticos como ARIMA.

Segmentación de Audiencia

Se suele utilizar el aprendizaje no supervisado para crear y descubrir patrones no conocidos en el comportamiento de los clientes de una web, app o comercio. Estos algoritmos pueden descubrir grupos que desconocíamos por completo o agrupar ciertas características que se correlacionan y que nosotros difícilmente lo habiéramos identificado.

Juegos

Podemos utilizar ML para dar inteligencia a los enemigos que tenga que enfrentar el usuario (protagonista) del juego. También se utiliza para hacer que nuestras tropas sepan como moverse y saltar obstáculos en un campo de batalla.

Coche Autónomo

En este campo se están ejecutando muchas de las investigaciones en Machine Learning para darle vida propia a los automóviles. Uno de los usos será para que el coche mueva el volante por si solo, analizando las imágenes, también para detectar otros vehículos y no chocar y hasta predecir cómo se están moviendo los demás para evitar accidentes. Detectar símbolos de velocidad, Stop, zona escolar en la carretera. El coche tiene muchísimas decisiones que tomar y en muy poco tiempo, siendo todos cruciales y de alta prioridad. Es un caso muy interesante de estudio para el Machine Learning y se está volviendo una realidad.

Salud

Dados los síntomas presentados por un paciente en una base de datos de pacientes anónimos nuestra máquina deberá ser capaz de predecir si es probable que esa persona pueda sufrir una enfermedad específica. Este tipo de Aprendizaje Automático es de suma delicadeza y puede servir como soporte para un equipo médico.

¿Qué es Machine Learning?

Economía y Finanzas

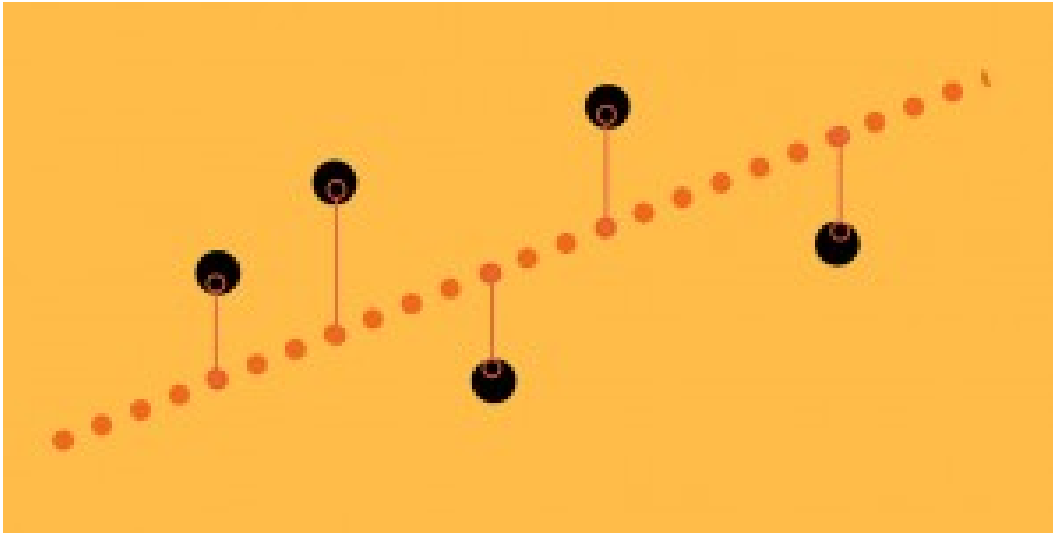
Se podrá dar soporte a analistas financieros intentando predecir determinadas cotizaciones de acciones en la Bolsa, ayudando a decidir si conviene comprar, mantener o vender.

Motores de Recomendación

Vemos una peli o una serie. Millones de personas ven esa misma serie. ¿No sería lógico que a mi me guste otra peli que vieron otros usuarios con un perfil similar al mío? Pues eso es lo que aprovechan Netflix y otros proveedores para poder recomendar contenidos a sus usuarios. Amazon es famoso por sus... “Si llevas ese producto, también te interesará este otro”. Esto también se consigue con este tipo de Aplicaciones del Aprendizaje Automático.

Principales Algoritmos usados en Machine Learning

Algoritmos de Regresión



Algoritmos de Regresión, Logístico o Lineal. Nos ayudan a clasificar o predecir valores. Se intentará compensar la mejor respuesta a partir del menor error.

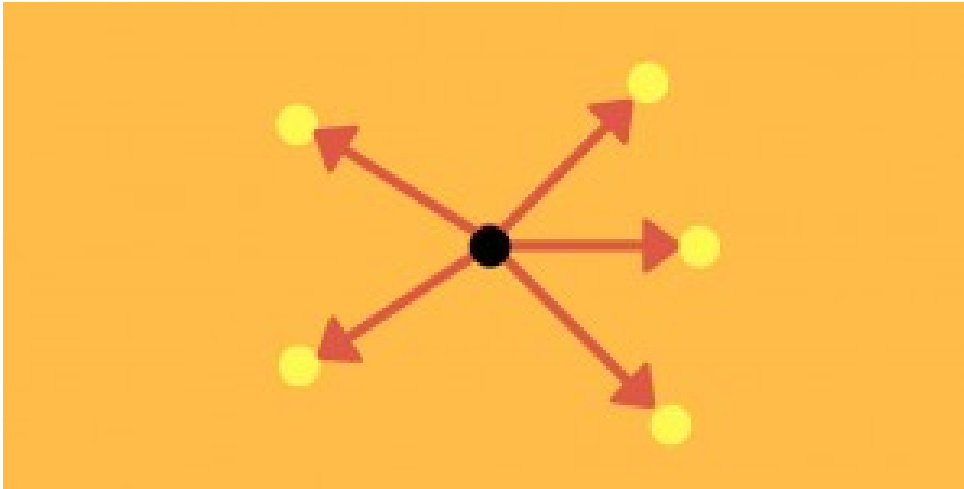
Los algoritmos de Regresión modelan la relación entre distintas variables (features) utilizando una medida de error que se intentará minimizar en un proceso iterativo para poder realizar predicciones “lo más acertadas posible”. Se utilizan mucho en el análisis estadístico. El ejemplo clásico es la predicción de precios de Inmuebles a partir de sus características: cantidad de ambientes del piso, barrio, distancia al centro, metros cuadrados del piso, etc.

Los Algoritmos más usados de Regresión son:

- Regresión Lineal
- Regresión Logística

¿Qué es Machine Learning?

Algoritmos basados en Instancia



Instance-Based Algorithms.

Son Modelos de Aprendizaje para problemas de decisión con instancias o ejemplos (muestras) de datos de entrenamiento que son importantes o requeridos por el modelo.

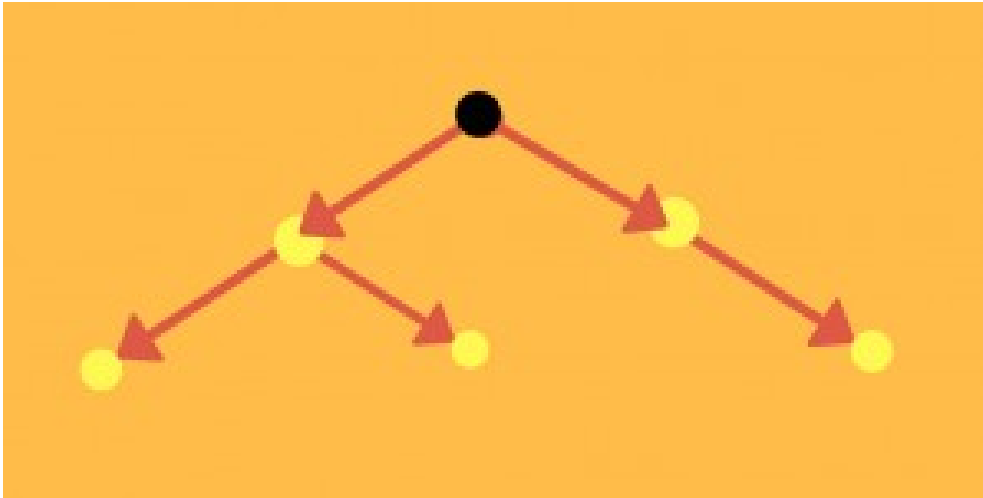
También son llamados Algoritmos “Ganador se lleva todo” y aprendizaje basado-en-memoria en el que se crea un modelo a partir de una base de datos y se agregan nuevos datos comparando su similitud con las muestras ya existentes para encontrar “la mejor pareja” y hacer la predicción.

Los Algoritmos basados en instancia más usados son:

- k-Nearest Neighbor (kNN) –
- Self-Organizing Map

¿Qué es Machine Learning?

Algoritmos de Arbol de Decisión



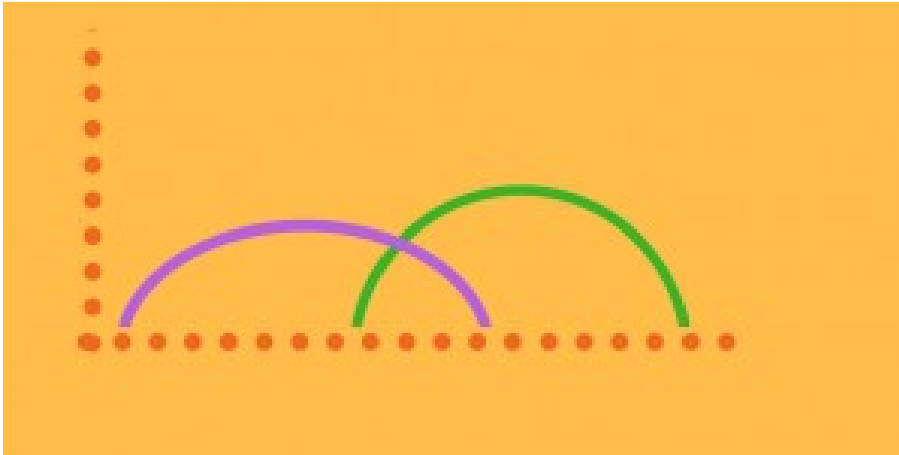
Arbol de Decisión. Buscará el mejor árbol, balanceando la posibilidad de ocurrencia y su importancia en cada rama y hojas para clasificar un resultado

Modelan la toma de Decisión basado en los valores actuales (reales) de los atributos que tienen nuestros datos. Se utilizan sobre todo para clasificación de información, bifurcando y modelando los posibles caminos tomados y su probabilidad de ocurrencia para mejorar su precisión. Una vez armados, los arboles de decisión ejecutan muy rápido para obtener resultados.

Los Algoritmos de árbol de decisión más usados son:

- Arboles de Clasificación y Regresión (CART)
- Decisión de Arbol condicional
- Random Forest

Algoritmos Bayesianos



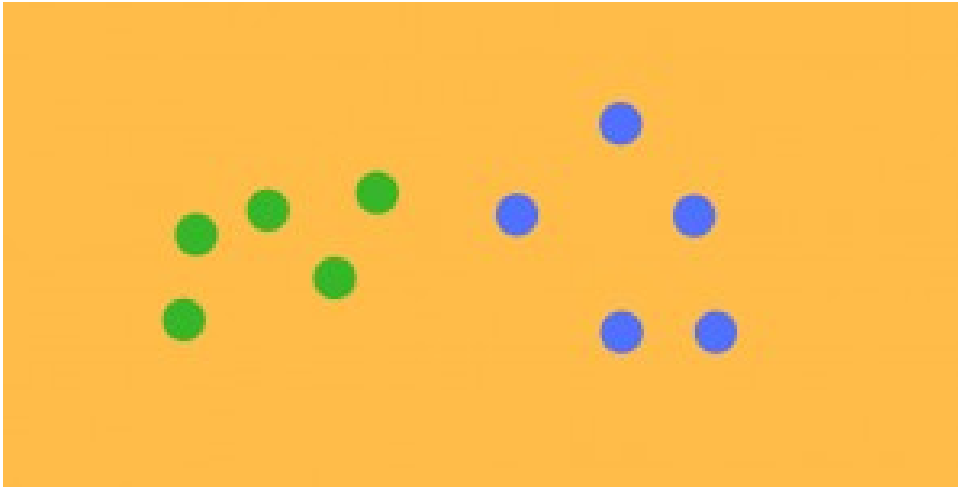
Algoritmos Bayesianos aprovechan los teoremas de probabilidad en Aprendizaje Automático

Son algoritmos que utilizan explícitamente el Teorema de Bayes de probabilidad para problemas de Clasificación y Regresión.
Los más utilizados son:

- Naive Bayes
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bayesian Network

¿Qué es Machine Learning?

Algoritmos de Clustering (agrupación)



Algoritmos de Agrupación son de Aprendizaje no supervisado y encontrarán relaciones entre los datos que seguramente no contemplamos a simple vista.

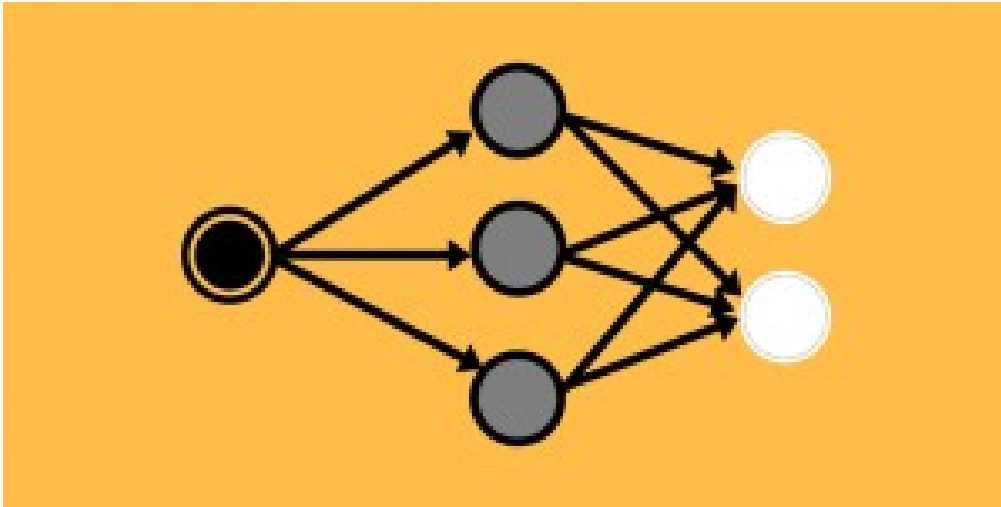
Se utilizan para agrupar datos existentes de los que desconocemos sus características en común o queremos descubrirlas.

Estos métodos intentan crear “puntos centrales” y jerarquías para diferenciar grupos y descubrir características comunes por cercanía.

Los más utilizados son:

- K-Means
- K-Medians
- Hierarchical Clustering

Algoritmos de Redes Neuronales



Las Redes Neuronales imitan el comportamiento de activación biológico y la interconexión entre neuronas para buscar soluciones no lineales a problemas complejos

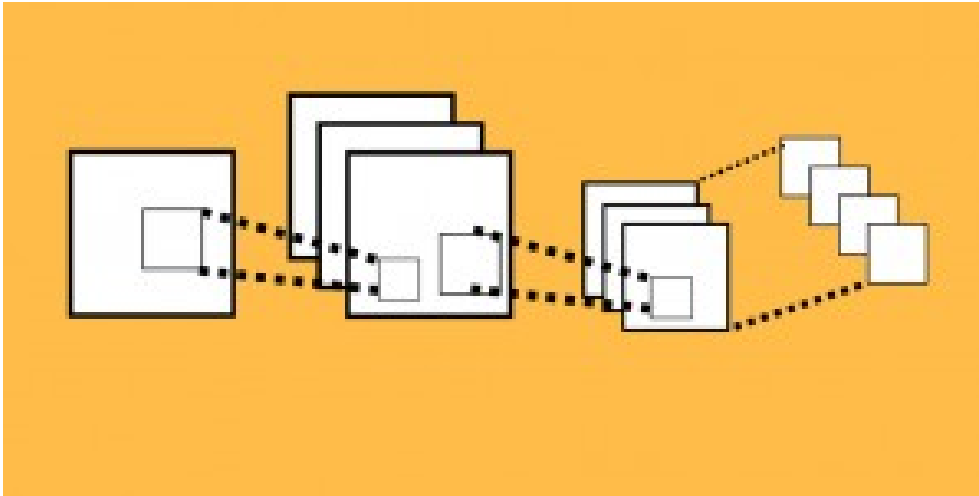
Son algoritmos y estructuras inspirados en las funciones biológicas de las redes neuronales.

Se suelen utilizar para problemas de Clasificación y Regresión pero realmente tienen un gran potencial para resolver multitud de problemáticas. Son muy buenas para detectar patrones. Las Redes Neuronales Artificiales requieren mucha capacidad de procesamiento y memoria y estuvieron muy limitadas por la tecnología del pasado hasta estos últimos años en los que resurgieron con mucha fuerza dando lugar al Aprendizaje Profundo (se detalla más adelante). Las redes neuronales básicas y clásicas son:

- Compuerta XOR
- Perceptron
- Back-Propagation
- Hopfield Network
- MLP: Multi Layered Perceptron

¿Qué es Machine Learning?

Algoritmos de Aprendizaje Profundo



Las redes convolucionales hacen que una red neuronal de aprendizaje profundo tenga la capacidad de reconocer animales, humanos y objetos dentro de imágenes

Son la evolución de las Redes Neuronales Artificiales que aprovechan el abaratamiento de la tecnología y la mayor capacidad de ejecución, memoria y disco para explotar gran cantidad de datos en enormes redes neuronales interconectarlas en diversas capas que pueden ejecutar en paralelo para realizar cálculos.

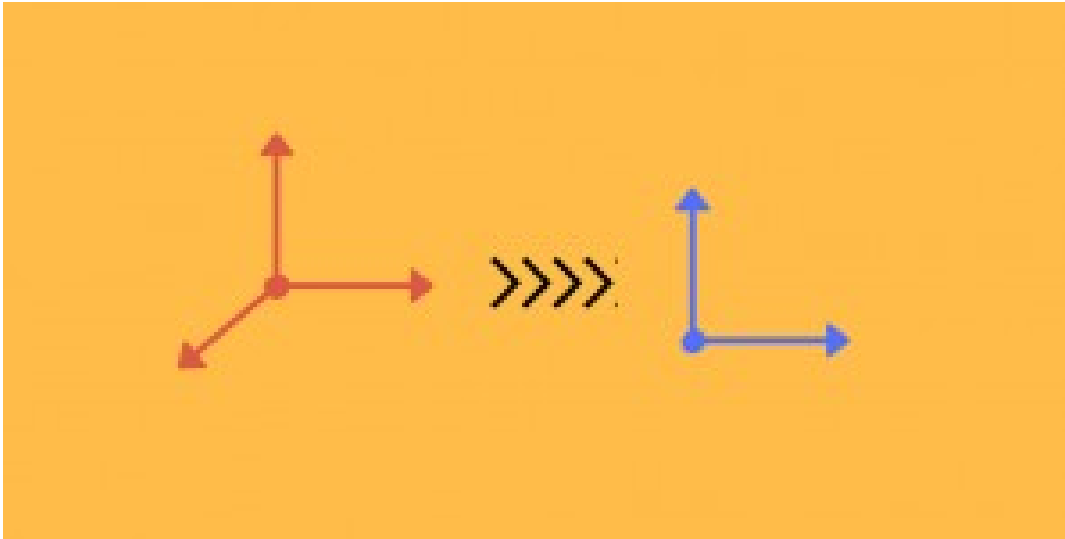
Comprende mejor Deep Learning con esta guía rápida que escribí.

Los algoritmos más populares de Deep Learning son:

- Convolutional Neural Networks
- Long Short Term Memory Neural Networks

¿Qué es Machine Learning?

Algoritmos de Reducción de Dimensión



Reducción de dimensión nos permite graficar o simplificar modelos muy complejos que en su condición inicial contenían demasiadas características.

Buscan explotar la estructura existente de manera no supervisada para simplificar los datos y reducirlos o comprimirlos.

Son útiles para visualizar datos o para simplificar el conjunto de variables que luego pueda usar un algoritmo supervisado.

Los más utilizados son:

- Principal Component Analysis (PCA)
- t-SNE (próximamente artículo)

¿Qué es Machine Learning?

Procesamiento del Lenguaje Natural (NLP)

El Natural Language Processing es una mezcla entre DataScience, Machine Learning y Lingüística. Tiene como objetivo comprender el lenguaje humano. Tanto en textos como en discurso/voz. Desde analizar sintáctica ó gramáticamente miles contenidos, clasificar automáticamente en temas, los chatbots y hasta generar poesía imitando a Shakespeare. También es común utilizarlo para el Análisis de Sentimientos en redes sociales, (por ejemplo con respecto a un político) y la traducción automática entre idiomas. Asistentes como Siri, Cortana y la posibilidad de preguntar y obtener

Otros Algoritmos

Otros algoritmos no desarrollados en este artículo:

- Algoritmos de Aprendizaje por Reglas de Asociación
- Algoritmos de Conjunto
- Computer Vision
- Sistemas de Recomendación
- Aprendizaje por Refuerzo
-

Recordemos que todos estos algoritmos son vulnerables a problemas de Underfitting y Overfitting que deberemos tener en cuenta y resolver.

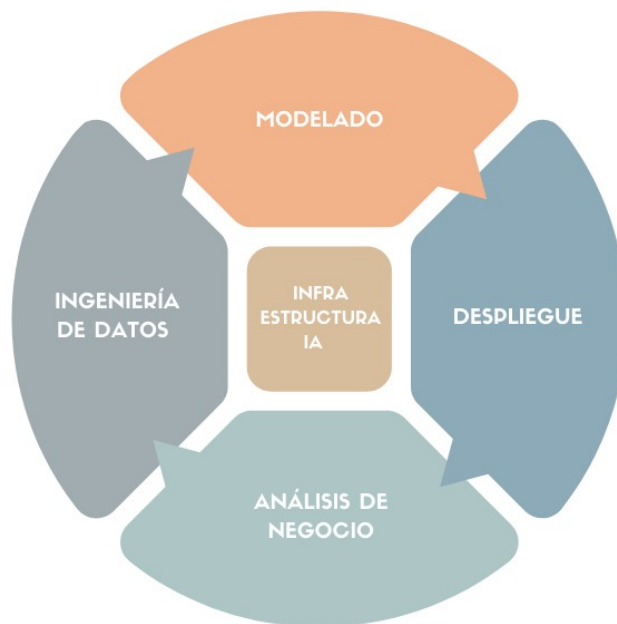
¿Qué es Machine Learning?

El proyecto de ML

Primero definamos en grandes rasgos las diversas etapas que conforman el desarrollo de un proyecto de Machine Learning.

1. Análisis de Negocio
2. Infraestructura de IA
3. Ingeniería de Datos
4. Modelado
5. Implementación / Despliegue

CICLO DE VIDA DE UN PROYECTO MACHINE LEARNING



¿Qué es Machine Learning?

1-Análisis de Negocio

En esta etapa se definen los objetivos que se desean alcanzar mediante el uso de la IA en el proyecto. Además se crean ó sugieren las métricas con las que se evaluará si los resultados que se obtengan darán valor -o no- al negocio.

Esta etapa es la primera e implica que al iterar el ciclo de vida del proyecto será la que decidirá si continuar o abandonar el proyecto. Entre otras cosas, se puede valer de visualizaciones y reportes (ó paneles) para lograr comprender el comportamiento de los modelos.

2-Infraestructura de IA

La infraestructura es una “etapa” transversal al resto de etapas, pues puede afectar al desarrollo de las mismas. Es conveniente contar con una mínima infraestructura sobre la que apoyarse desde el principio y que ese soporte sea flexible y escalable bajo demanda.

Necesitaremos ambientes para desarrollo, test y despliegue pero también herramientas, frameworks y entornos para negocio, seguridad, medios de almacenamiento, backups y auditoría. Debemos dar asistencia -o no- 24x7, tener algún mecanismo de feedback ó soporte técnico, repositorios para código, accesos restringidos a roles de usuario.

Una decisión importante también será si nosotros mismos montaremos y mantendremos la infraestructura (y sus costes) ó si nos apoyaremos en servicios existentes en la nube como AWS, Azure ó GCP.

3-Ingeniería de Datos

Esta etapa incluye la recolección de datos desde diversas fuentes y su tratamiento, preprocesado y futuro mantenimiento. Podemos tener uno ó muchos orígenes, bases de datos, archivos semi-estructurados ó datos sin estructura como ficheros de video o sonido. Acceso a información propia ó mediante APIs públicas ó de pago a terceros.

El tratamiento puede incluir múltiples transformaciones a datos existentes, cálculo de medias ó desvíos, agrupaciones, imputaciones ó reemplazo, descarte de columnas ó filas y hasta la generación de features sintéticas.

Todo esto sobre algunos megas ó a gigabits de información que debemos procesar en tiempo y forma. Aparecen cuestiones como el procesado masivo distribuido y en paralelo.

Y no olvidemos aquí, de ser necesario el asunto de las clases desbalanceadas y cómo equilibrarlas.

4-Modelado

Esta es la etapa donde ocurre la magia aquí es donde usaremos todos nuestros mayores skills en Data Science, Machine Learning, matemáticas y estadística, ingenio y creatividad.

Seleccionar modelos, algoritmos, selección de features, tuneo de hiperparámetros, entrenamiento, evaluación, split, métricas (técnicas y de negocio), optimización, interpretabilidad... predicción, clasificación, clustering, uso o no de redes neuronales, prueba y error, arquitecturas de redes, NLP, combinación, permutación, ensambles... BANG!

5-Implementación / Despliegue

Durante la fase de implementación y despliegue deberemos formalizar el código prototipo y pruebas realizadas anteriormente en un pipeline robusto y consistente.

Debemos combinar las etapas de obtención de datos con la de modelado para que puedan interrelacionarse y mantenerse en el tiempo.

Aquí es importante contar con un protocolo sobre cómo hacer el pasaje desde entornos de desarrollo a test y producción.

¿Qué es Machine Learning?

Además la fase de predicción puede requerir un despliegue como servicio web ó como un proceso batch ó puede tener otro tipo de endpoint como colas MQ u otro tipo de invocacion mediante sockets, definir su necesidad de tiempo real ó no.

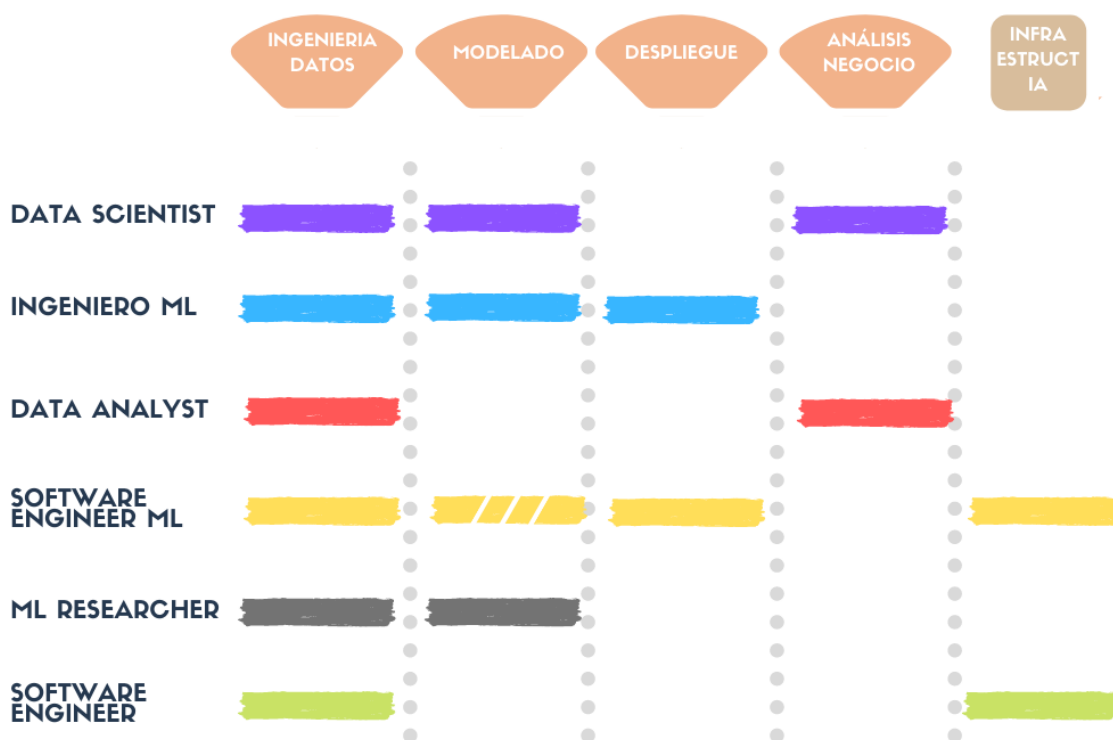
Se debe contemplar el reentrenamiento de los modelos, tiempos, triggers y condiciones para decidir si se reemplazará o no el modelo actual.

Y vuelta a empezar...

No olvidemos que estos 5 pasos son iterativos, con lo cual luego de la implementación volveríamos a comenzar en la etapa de Análisis de negocio, evaluando los resultados ó repercusiones obtenidas.

Los 6 Roles del equipo IA

A continuación comentaremos los 6 roles que cubren las diversas etapas del ciclo de vida de los proyectos de IA. Muchas veces estos roles se superponen unos a otros y muchas veces veremos cómo podemos identificarnos con más de uno de los roles. Estos están planteados como para lograr una especialidad en cada rol y obtener los beneficios -en una gran compañía sobretodo- que proporciona el supuesto de la separación de tareas.

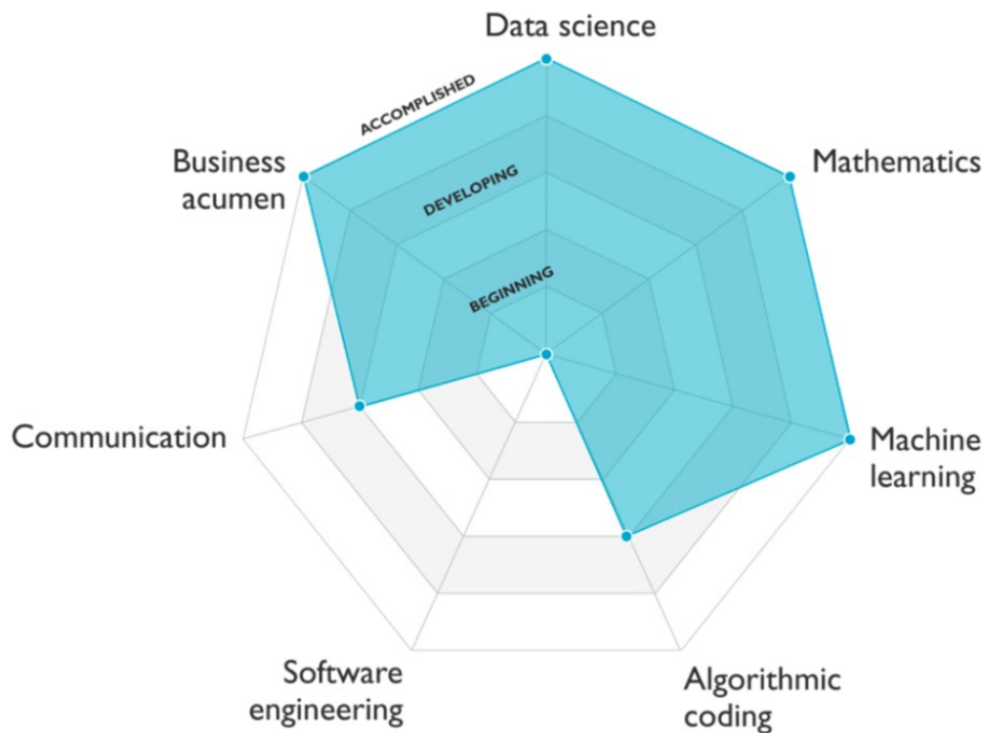


APRENDEMACHINELEARNING.COM

1. Data Scientist
2. Machine Learning Engineer
3. Data Analyst
4. Software Engineer ML
5. ML Researcher
6. Software Engineer

¿Qué es Machine Learning?

1-Data Scientist



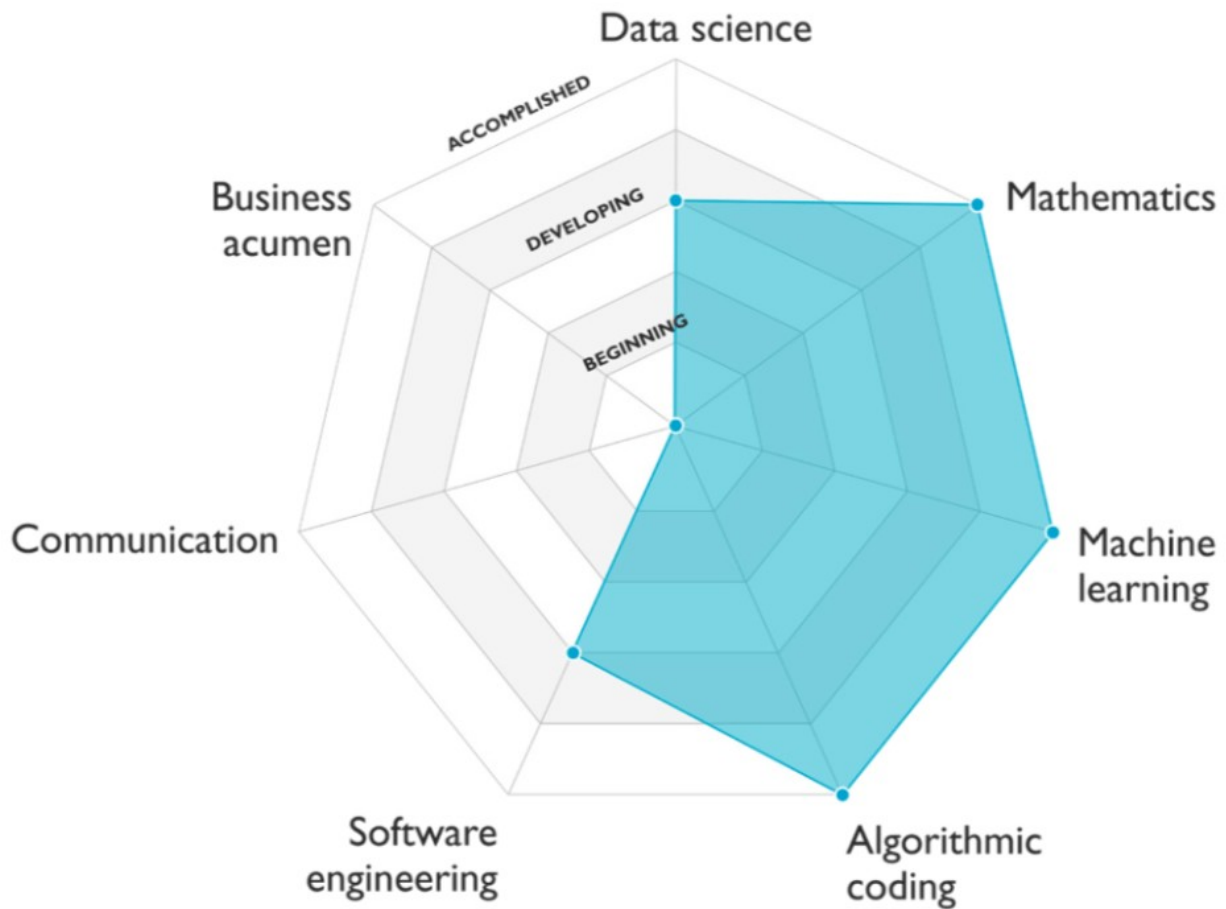
Skills de un Data Scientist – workera.ai

El científico de datos puede desarrollar las etapas de Ingeniería de Datos, modelado y Análisis de Negocio. Deberá tener fundamentos científicos sólidos así como habilidades en comunicación para poder transmitir los hallazgos a otros miembros del equipo ó a usuarios de negocio.

TOOLS: Python, numpy, scikit learn, Tensorflow, Pytorch, SQL, Tableau, Excel, PowerBI, Git, Jupyter Notebook.

¿Qué es Machine Learning?

2-Machine Learning Engineer



Skills de un Machine Learning Engineer – workera.ai

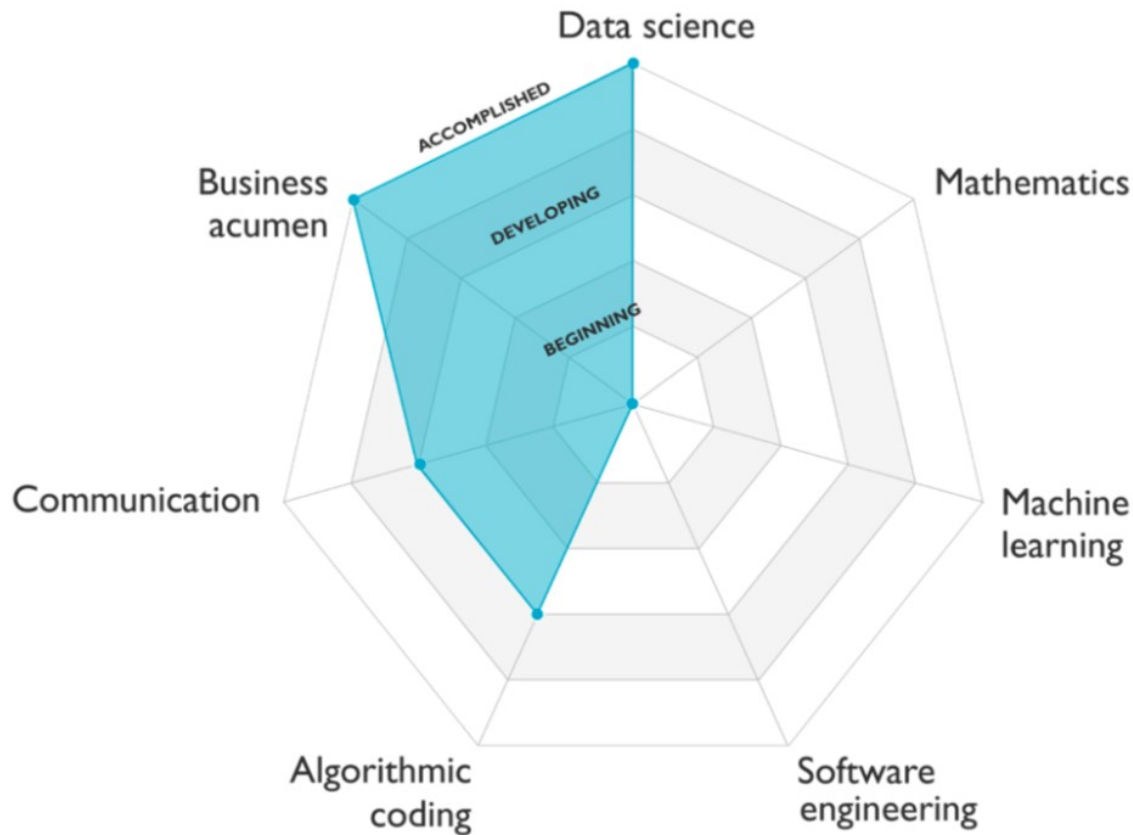
Pueden desarrollar las tareas de Ingeniería en datos, modelado y despliegue. En algunos casos también pueden colaborar con la analítica de negocio y en la infraestructura.

Deben tener habilidades en ingeniería pero también ciencias. Sus competencias en comunicación pueden depender de las necesidades del equipo y del resto de roles. Se prevé que este rol esté al tanto de las últimas tendencias en algoritmos y papers relacionados con Machine Learning y el uso y técnicas de Deep Learning.

TOOLS: Python, SQL, numpy, scikit learn, Tensorflow, Pytorch, Cloud: AWS, Azure, GCP, Git, Jupyter Notebook, JIRA.

¿Qué es Machine Learning?

3-Data Analyst



Skills de un Data Analyst - workera.ai

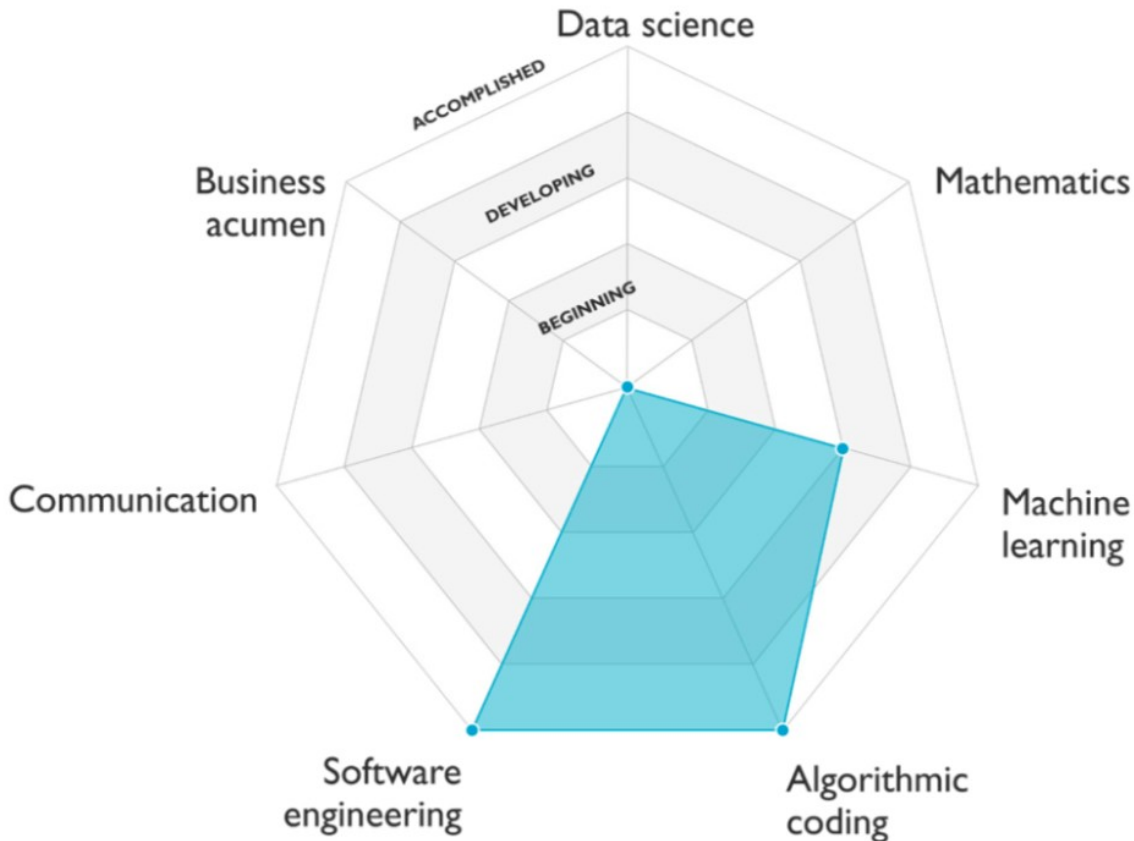
Preparado para las fases de Ingeniería de Datos y Análisis de Negocios. Por lo general tienen un gran conocimiento de SQL y manejo de bases de datos y analítica avanzada de negocios, visualización y reporting.

Por lo general se pretende de este rol muy buenas habilidades comunicativas, y se les exige menos en capacidad algorítmica ó de programación.

TOOLS: Python, SQL, Tableau, PowerBI, Excel, PowerPoint, A/B testing.

¿Qué es Machine Learning?

4-Software Engineer ML



Skills de un Software Engineer en Machine Learning – workera.ai

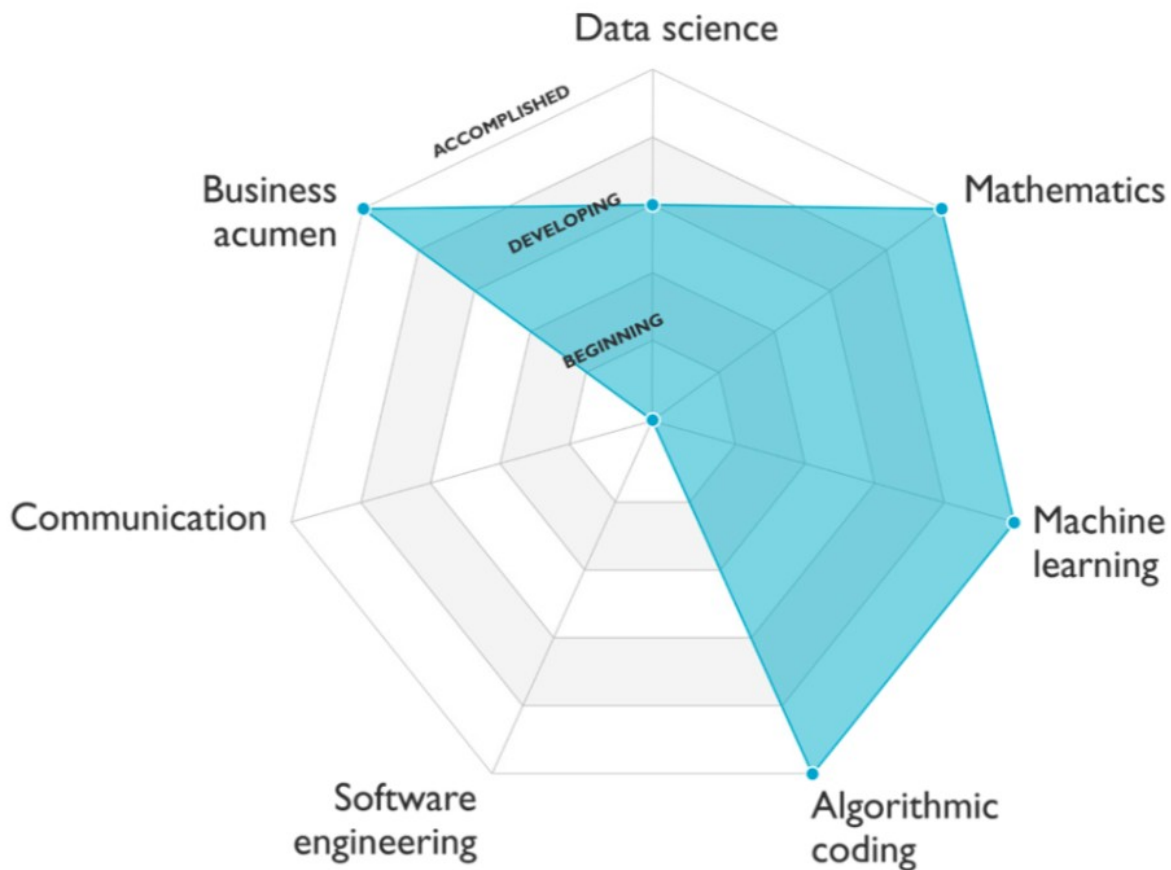
Personas con este título podrán desarrollar las etapas de Ingeniería de Datos, modelado, despliegue e infraestructura.

Deberían trabajar bien con otros miembros que trabajen con la parte de negocio. Este rol es conveniente para etapas tempranas de equipos o startups pues puede cubrir multitud de tareas, logrando prototipos y demostrando gran versatilidad.

TOOLS: Python, SQL, numpy, scikit learn, Tensorflow, Pytorch, AWS, GCP, Azure, Git, Jupyter Notebook, JIRA.

¿Qué es Machine Learning?

5-ML Researcher



Skills de un Machine Learning Researcher – workera.ai

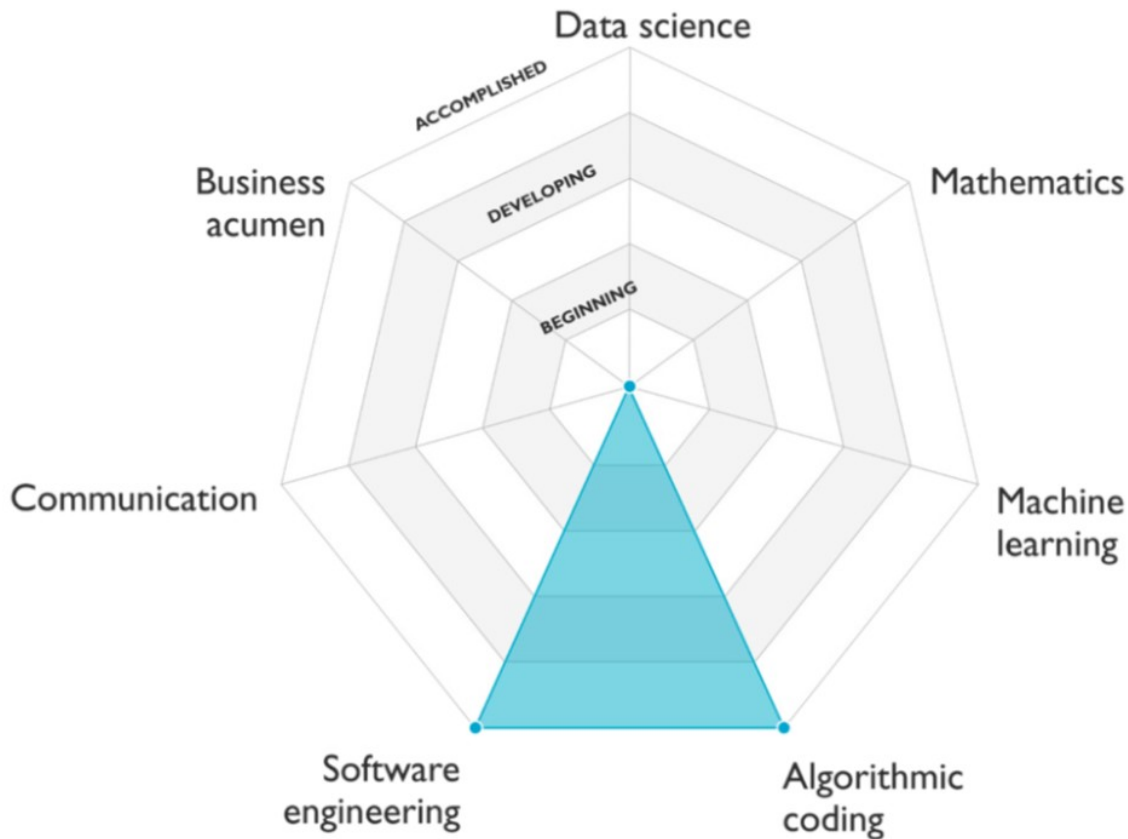
Este rol puede llevar las etapas de Ingeniería de datos y modelado. Desarrollan su mayor potencial en un ambiente de investigación, donde puedan dar rienda a sus conocimientos para buscar y descubrir patrones en los datos. Debe tener excelentes habilidades y conocimiento científico.

También puede especializarse en Deep Learning ó en algún área en particular como NLP, visión artificial ó motores de recomendación.

TOOLS: Python, SQL, numpy, scikit learn, Tensorflow, Pytorch, Git, Jupyter Notebook, Arxiv, NeurIPS, papers.

¿Qué es Machine Learning?

6-Software Engineer



Skills de un Software Engineer - workera.ai

Este rol -ahora muy ligado a “devops”- puede ocuparse de las etapas de Ingeniería de Datos e Infraestructura.

Demuestran gran habilidad en programación y manejo de software y diversas herramientas ó plataformas especializadas.

TOOLS: Python, SQL, AWS, GCP, Azure, Git, Jupyter Notebook, JIRA.

a Science tech Stack 2020

