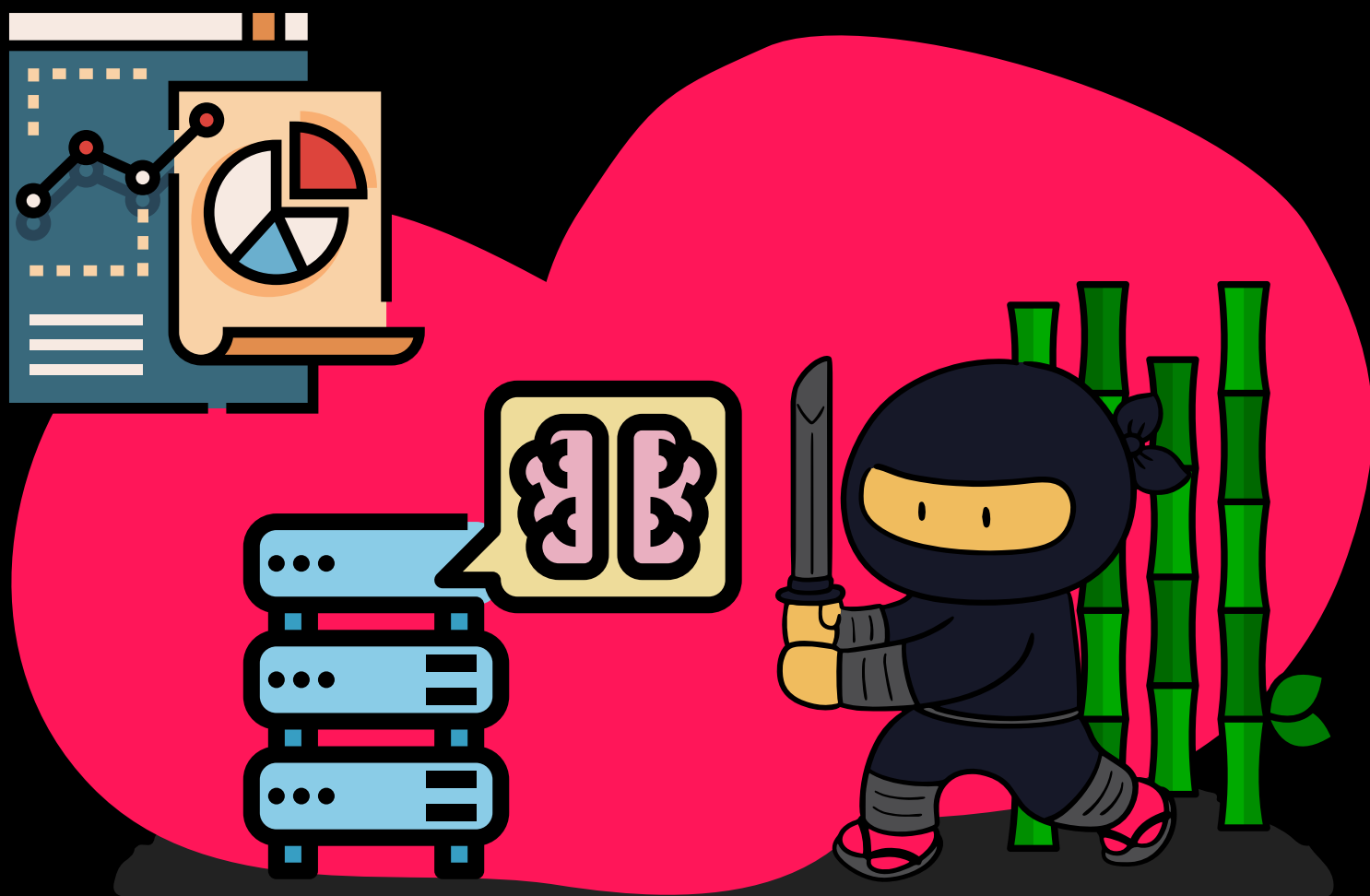


GUÍA NINJA DEL BIG DATA Y EL MACHINE LEARNING



WWW.DATOS.NINJA

WWW.DATOS.NINJA

CONTENIDO

BIENVENIDO AL MUNDO DEL BIG DATA Y LA INTELIGENCIA ARTIFICIAL	4
¿Qué es Big Data?	8
¿Qué es Data Science?	10
¿Qué es Machine Learning?	12
PERO... ¿POR DÓNDE EMPIEZO?	13
Estrategia basada en teoría	15
Estrategia basada en la Práctica	18
Títulos y certificados	19
Etapas de aprendizaje	20
5 mandamientos y 10 errores a evitar	21
Libro recomendado	28
HABILIDADES Y HERRAMIENTAS	30
Python, Análisis de Datos y Machine Learning	31
Bases de datos, lenguajes SQL y NoSQL	32
Arquitectura de computación distribuida	33
HARDWARE NECESARIO Y PREPARACIÓN DEL ENTORNO	35
CONCLUSIÓN	39



BIENVENIDO AL MUNDO DEL BIG DATA Y LA INTELIGENCIA ARTIFICIAL

Mucha gente se interesa por la inteligencia artificial, el machine learning y el Big Data hoy en día.

De toda esta gente, la mayoría no profundiza y lo acaba olvidando.

¿Por qué?

¿Por qué esas personas dejan de lado algo que en un principio parecía tan apasionante?

Principalmente hay dos razones que podrían explicar este comportamiento:

Hacen sus primeras investigaciones sobre la Ciencia de Datos y descubren una cantidad abrumadora de cosas por aprender para ser un ninja de los datos.

Se encuentran con gente que les dice que si no son unos expertos en matemáticas, en estadística aplicada y además saben Python nivel gurú no tienen nada que hacer.

Son suficientes cosas como para asustar a un científico de datos

experimentado, así que imagínate si es alguien que está empezando.

No sé tú...

Cuando yo comencé a interesarme por el Big Data, la Ciencia de Datos y la Inteligencia Artificial creía que me adentraba en un campo complejo, que aprendería técnicas y descubriría herramientas que me permitirían resolver problemas chulos. Los problemas chulos de los que hablaban en las notas de prensa de todas las empresas que decían implementar soluciones Big Data.

En realidad no tenía ni idea de donde me estaba metiendo.

Y tampoco me quedaba muy claro al preguntarle a Google. Todo el mundo parecía estar usando Big Data y no acababa de entender qué hacían exactamente y mucho menos cómo podía aprender a hacerlo yo.

Cumplía los requisitos básicos para dejar el Big Data olvidado en un cajón, menos mal que soy muy cabezota y sigo persistiendo cada día para convertirme en ninja del Big Data 🥷

Esta guía pretende ser un pequeño mapa para situarte en el universo Big Data.

Para contarte cómo puedes empezar en esto y espero que también puedas ahorrarte caer en algunos de los errores más comunes que la gente comete al adentrarse en este mundillo del Big Data y la Ciencia de Datos.

Personalmente, creo haberlos cometido todos 😓

Pero comencemos antes por los aciertos... Lo primero de todo:

ENHORABUENA POR HABER COMENZADO

Seguramente, el mejor momento para entrar en el mundo Big Data esté en el pasado. Sin embargo, el segundo mejor momento para empezar con ello es **ahora**.

Seguro.

Tienes internet. Puedes aprender cualquier cosa.

Si te apasiona aprender cosas nuevas y resolver misterios, has llegado al iceberg indicado. El campo del Data Science y del Machine Learning, es infinito. Y se expande cada día.

Pero cuidado... Tienes que tener tres cosas claras:

1. La Ciencia de Datos es una carrera de fondo. Imposible de dominar en 1, 3 ó 6 meses. Si te lo tomas en serio podrás adquirir las habilidades básicas de una manera relativamente rápida pero para ser un ninja de los datos tendrás que dar cera y pulir mucha cera de manera **consistente**.
2. No te dejes abrumar por la cantidad de recursos que hay en Internet. Si intentas encontrar el método perfecto para aprender Machine Learning o Data Science corres el riesgo de infectarte de **parálisis por análisis**.
3. La **persistencia** es clave. Te encontrarás (si no lo has hecho ya...) con dificultades y frustraciones a lo largo del camino. Reserva un poco de tiempo cada día para seguir aprendiendo.

Es normal, e incluso sano, que quieras dominar el poder de los datos en el menor tiempo posible. Al fin y al cabo eres consciente de que tu recurso máspreciado es el **tiempo**.

Sin embargo, la cantidad de cursos, artículos e información que existen sobre este tema es inabordable para cualquier mortal y es probable que llegue a abrumarte.

Una cosa que puedes hacer es invertir algo de tiempo en descubrir la rama del Data Science que más te llama la atención y comenzar por ahí.

Hay muchos caminos para llegar a Roma y hay incluso más caminos para convertirse en un ninja de los datos.

Para no perderte en esta aventura solo necesitas seguir los 3 mandamientos ninja:

1. Empieza por algún sitio. No hay un orden correcto. Esta guía puede ser un buen comienzo para saber qué te resulta más interesante dentro del campo de Big Data.

2. No es necesario saberlo todo. Comienza a resolver problemas cuando aún no te sientas preparado. Empieza por entender lo mínimo necesario para continuar implementando el proyecto que hayas elegido y después profundiza.

3. Nunca dejes de aprender. Ser un ninja del Big Data es una maratón.

El Big Data, el Data Science y el campo del Machine Learning están en constante evolución. Nuevas técnicas, nuevos modelos y nuevas herramientas aparecen cada día para obtener información de valor de los datos que generamos.

Una vez hayas interiorizado los 3 mandamientos ninja estarás preparado para desencadenar el **potencial brutal de los datos**.

Recuerda que el Data Science, el Machine Learning y todas las disciplinas relacionadas con los datos se pueden aplicar a cualquier negocio que genere datos, es decir, al 99.9% de los negocios¹

Al ser un campo completamente transversal tiene la gran ventaja de que podrás elegir la industria que más te guste para aplicar tus conocimientos ninja. Salud, deporte, ventas, marketing... Cualquiera industria. Cualquiera.

Incluso tendrás la posibilidad de moverte de una a otra. Un sueño para cualquier mente inquieta.

¿No te parece lo suficientemente emocionante?

Y si te digo que la mayoría de las empresas aún no usan Machine Learning, ni Big Data, ni saben cómo sacar información de los datos ¿vislumbras ahora el potencial de convertirte en un ninja del Big Data?

¹ Porcentaje completamente inventado

Piensa que el proceso para poder implementar Big Data en un negocio concreto es complejo por lo que cualquier empresa que se meta en ello y lo haga bien, será capaz de diferenciarse de la competencia. Las compañías más potentes como Amazon, Google, Apple... , con músculo financiero para ponerse a ello, ya están metidas en Big Data desarrollando sus propias soluciones y, eventualmente, la tecnología llegará al resto. Cuanto antes encuentre el valor de los datos una empresa, más capacitada estará para diferenciarse de sus competidores.

Para eso necesitan ninjas de los datos.

Pero volvamos a esta guía, lo primero que voy a hacer es contarte las diferencias entre algunos conceptos que muchas veces aparecen mezclados.

Una explicación rápida de qué es el Big Data, en qué consiste la Ciencia de Datos y de qué va el Machine Learning. Dónde empieza y acaba cada uno.

Si tienes claras las diferencias puedes pasar directamente a la siguiente sección.

¿QUÉ ES BIG DATA?

Big Data son soluciones basadas en el manejo y procesamiento de **volúmenes muy grandes de datos**, tan grandes que tienen que repartirse entre varios ordenadores.

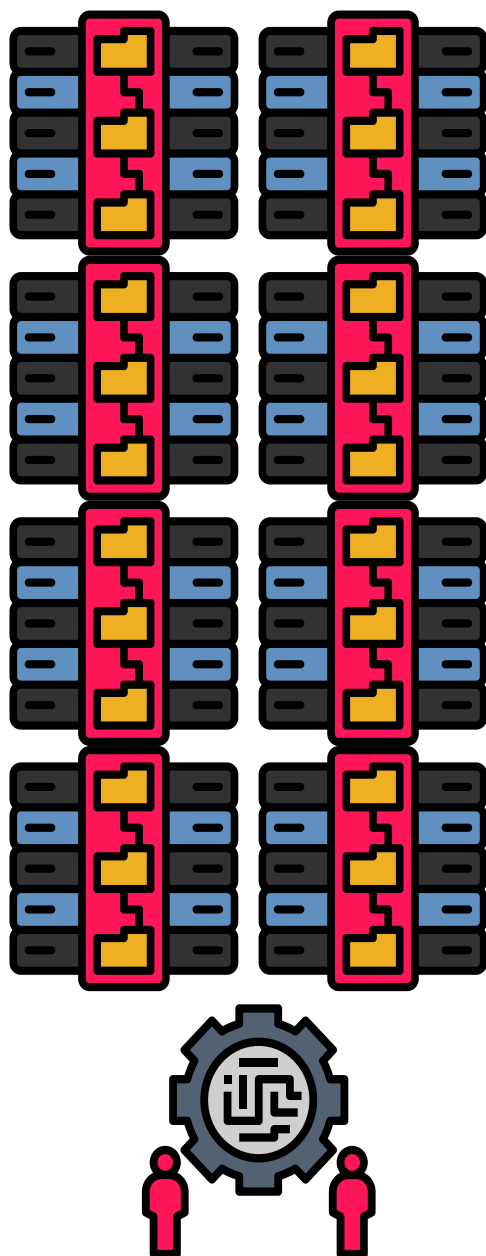
Estos datos se generan de forma continua y se tienen que poder tratar o almacenar a una **velocidad alta** y, por si fuera poco, son diversos. Algunos son registros con información estructurada, archivos de correo electrónico, imágenes, vídeos... son datos **variados**.

¿Y qué nos aporta manejar tantísima información? - Te preguntarás -...

Pues bien, cuando es posible acceder a la cantidad de datos necesaria como para encontrarnos en un entorno de Big Data el valor añadido por este ecosistema podría compararse a intentar buscar la solución a un problema en un libro o buscarla en una biblioteca entera.

¿Y por qué no se ha construido esa "biblioteca de soluciones" hasta ahora si ya éramos conscientes del valor de los datos?

Básicamente, el coste de adquirir los sistemas de almacenamiento para guardar grandes cantidades de datos ha sido demasiado elevado hasta hace unos años. Además, la tecnología disponible no era capaz de procesar volúmenes de información tan elevados a la velocidad adecuada y las soluciones para integrar los datos, mantenerlos actualizados y realizar copias de seguridad eran demasiado complejas.



Por suerte, esto ha cambiado últimamente 😊

Las herramientas Big Data y, por extensión, los ingenieros de Big Data, facilitan el trabajo de analistas de datos y consultores de Business Intelligence y aumentan la capacidad de gestionar cada vez cantidades más grandes de datos.

Por ejemplo, los bancos utilizan Big Data para dar solución a los problemas de detección de fraude o de blanqueo de capitales entre otros casos de uso.

Otro ejemplo que usas todos los días es el algoritmo de búsqueda de Google, que también está cimentado sobre el procesado de grandes volúmenes de información (lo que viene siendo todo Internet) para intentar devolver un resultado relevante a la consulta de un usuario.

¿Tienes ganas de saber más sobre Big Data? Puedes encontrar la versión completa del artículo [aquí](#).

¿QUÉ ES DATA SCIENCE?

El Data Science es la ciencia que tiene por objetivo extraer información de valor y el conocimiento de los datos.

Y con ciencia me refiero a métodos, a procedimientos y a las habilidades necesarias para trabajar con datos sin volverse loco en el intento.

Para empezar, voy a volver a subrayar para qué sirve el data science: **para extraer conocimiento e información de valor de un montón de datos.**

Para lo cual necesitamos:

1. Un problema concreto a resolver
2. Un conjunto de datos relacionados con nuestro problema

Con un científico de datos en el equipo (o mejor aún, con un grupo completo dedicado a los datos) y con la capacidad de obtener datos de valor, es posible incorporar rápidamente los beneficios del Data Science.

A continuación te enumero solo algunos de estos beneficios:



- 1.** La toma de decisiones estratégicas es más potente gracias a las técnicas de análisis prescriptivo a partir de los datos.
- 2.** El diagnóstico de cualquier problema es más rápido a través de los patrones obtenidos de los datos históricos y actuales de la empresa / organización.
Si tienes datos sobre los resultados de varios departamentos en una empresa, es posible saber cuál tiene un mejor rendimiento e inferir el porqué.
- 3.** Es posible identificar oportunidades mediante el análisis predictivo de los datos. Como por ejemplo en el caso de datos de mercado y datos sobre el comportamiento de los consumidores.
- 4.** Aumentar las eficiencias de un negocio reduciendo el gasto en recursos innecesarios.
Imagínate que tienes una empresa que se anuncia en varios medios (anuncios de facebook, anuncios de google, anuncios en la tele o en periódicos) y de toda esa publicidad, la que mejor convierte es la que viene de google. Entonces será mejor gastar más en anuncios de google que en el resto de publicidad ¿verdad?
- 5.** De manera análoga a la identificación de oportunidades, también es posible estar mejor preparados ante posibles crisis.
- 6.** A partir de técnicas de análisis descriptivo es posible conocer mejor a nuestros usuarios / clientes permitiendo saber qué necesitan.
Conocer los patrones de consumo de nuestros clientes hacen posible que podamos ofrecer productos y servicios más adaptados a lo que buscan.

Si quieres saber más sobre la Ciencia de datos puedes echar un vistazo a [este artículo](#).

¿QUÉ ES MACHINE LEARNING?

El aprendizaje automático o machine learning es el conjunto de técnicas utilizadas para conseguir que un ordenador adquiriera una habilidad determinada sin necesidad de haberlo programado específicamente para realizar esa habilidad.

El Machine Learning forma parte del campo de la Inteligencia Artificial. Más concretamente, se trata de una **Inteligencia Artificial Débil**, que básicamente consiste en realizar una sola cosa y hacerla muy muy bien.

Generalmente son cosas que a una persona le llevaría pocos segundos de reflexión llevar a cabo, como reconocer un árbol en una imagen, identificar un mensaje de correo electrónico como SPAM o detectar si un material está defectuoso en una línea de producción.

De hecho, gracias al aprendizaje automático, los ordenadores son capaces de realizar estas tareas mejor que los humanos.

Pero eso no hace que sean inteligentes...

El algoritmo es capaz de ma-

pear datos de entrada con la salida esperada y utilizar ese aprendizaje sobre entradas que no ha visto anteriormente. Necesita ver muchos ejemplos antes de aprender a resolver el problema por él mismo.

Como imaginarás, la pega de este tipo de Machine Learning (supervisado), es que no siempre es fácil conseguir ejemplos con solución de un problema determinado.

Por ejemplo, si quisiéramos implementar un modelo de Machine Learning para distinguir perros y gatos habría que conseguir miles de imágenes de perros y gatos y etiquetarlas con el animal correspondiente. Esto generalmente tiene que hacerlo una persona y llega a ser muy costoso.

Puedes encontrar el artículo completo sobre qué es Machine Learning [aquí](#).



PERO... ¿POR DÓNDE EMPIEZO?

Espero que llegados a esta parte ya tengas un poco más claras las piezas importantes de este juego aunque probablemente aún no sepas por dónde empezar.

Para comenzar te diré por dónde NO empezar...

No te obsesiones con los algoritmos y modelos recién salidos... aparecen nuevos todo el rato.

Ignoralos.

No te distraigas.

En el campo del Machine Learning la mayor parte de los esfuerzos está enfocada en desarrollar nuevos modelos, cada vez más complejos, y la mayoría de las veces dan resultados ligeramente mejores.

En la vida real muchos de estos modelos son tan complejos y costosos de implementar que se quedan en el cajón de los modelos sin usar.

Te contaré una historia real.

NETFLIX

En 2006, Netflix ofreció una recompensa de 1 millón de dólares a aquella persona que fuera capaz de mejorar la precisión de su algoritmo de predicción de valoración de películas un 10%.

Un año más tarde, el equipo que más se había aproximado, había mejorado la precisión de Netflix un 8.43% después de combinar 107 algoritmos de Machine Learning y más de 2000 horas de trabajo.

En 2009, finalmente hubo un ganador que consiguió mejorar la precisión original de Netflix un 10% pero la solución era tan compleja que Netflix nunca llegó a implementarla.

El esfuerzo y coste eran muy altos para justificar la mejora de rendimiento y por aquel

entonces Netflix había comenzado a migrar su negocio desde el alquiler de DVDs hacia la reproducción de contenido en streaming por lo que los datos de los que empezaba a disponer eran diferentes.

Conclusión: La mayor parte de los modelos de Machine Learning que se utilizan en la vida real son mucho menos precisos que los de última generación pero son lo suficientemente precisos para abordar el problema y mucho menos costosos.

EMPIEZA POR ALGO MÁS SENCILLO.

En mi caso, comencé por el Machine Learning. El clásico, el de toda la vida. Sin redes neuronales y mucho menos sin Deep Learning.

Quería establecer unas bases muy sólidas antes de pasar a los todo-poderosos modelos de Deep Learning.

Sin embargo, antes de pasar al Deep Learning, me llamó la atención el entrenamiento y despliegue de modelos en la nube, en la de Google concretamente.

¿Por qué Google y no Amazon o Microsoft? No lo sé.

¿Por qué la nube en lugar de comenzar a entrenar modelos en local? Por mantener mi ordenador limpio, sin instalar librerías ni historias, lo cual puso muy contenta a la minimalista que hay en mí y mientras, mi modelo se entrenaba en las máquinas superpotentes de Google.

Tal vez tu camino sea diferente. Y estará bien... siempre y cuando sigas caminando.

Hay principalmente dos estrategias para aprender cualquier cosa:

- La estrategia en la que primero aprendes toda la teoría y luego empiezas a usarla. Spoiler alert: En Data Science nunca llegarás a saber toda la teoría.
- La estrategia en la que comienzas a implementar sin tener ni idea de lo que estás haciendo.

Cada una de estas estrategias tiene sus ventajas y sus inconvenientes.

Cada una de estas estrategias se adapta mejor a un tipo de persona diferente.

ESTRATEGIA BASADA EN TEORÍA

Tradicionalmente es la estrategia que se ha utilizado en la enseñanza reglada.

Primero te enseñan los conceptos y tú escuchas o tomas notas, después te ponen algún ejemplo sencillo y por último estudias para un examen. El 80% de las veces has olvidado todo lo estudiado al salir por la puerta del examen.

¿Sólo me pasaba a mí esto? 🤔

Los conceptos se van construyendo uno sobre otro, de más simples a más complejos.

Luego, un día llegas a la vida real y te das cuenta de los agujeros que tiene tu conocimiento. Toca reaprender algunas cosas (o muchas... 😊).

Es un proceso en el que se tarda mucho en ver resultados, se tarda mucho en detectar los errores de concepto que tenemos o las cosas que creíamos entender bien pero en realidad no habíamos pillado.

Cuando tenía 10 años mi padre me apuntó a clases de golf, yo quería ir a jugar al campo como hacía él y sin embargo, tenía que pasarme toda la mañana con un cubo de bolas para aprender a darle a la bola y aprender a hacer el movimiento de manera correcta.

Después de un tiempo en el campo de bolas le pregunté... ¿cuánto tiempo falta para poder salir

al campo como los mayores?

Me contestó que mucho, al menos 6 meses. Hasta que aprendiera a darle a la bola bien.

Me pareció tanto tiempo que nunca volví a coger un palo de golf.

En realidad tiene sentido que alguien que no sabe jugar no salga al campo antes de tener un nivel mínimo pero yo quería jugar al golf, no dar bolas toda la mañana.

No era divertido y me cansé.

El riesgo de tener que aprender toda la teoría antes de poder aplicar cualquier cosa es que te canses y lo dejes. Si hubiera habido un campo de golf "de pruebas" tal vez habría seguido con este hobby.

En Data Science y Big Data no tienes porqué saberlo todo antes de empezar a "jugar".

PUEDES SABER LO MÍNIMO, APLICARLO, IR PROFUNDIZANDO E ITERAR.

NIVEL DE MATEMÁTICAS ●●●

Una de las bases sobre las que se asienta la Ciencia de Datos son las matemáticas.

Eso es así.

Probabilidad, cálculo y álgebra matricial son los cimientos sobre los que se construyen todos los modelos de Machine Learning.

Esto desanima a muchos ninjas del Machine Learning en potencia.

Van a estudiar un grado en Ciencia e Ingeniería de Datos y el primer cuatrimestre se encuentran con:

- Álgebra Lineal
- Cálculo I
- Probabilidad

Y ya, si eso... Introducción a la ciencia de datos y algo de programación.

Uf... Complicado sobrevivir al primer cuatrimestre ¿no crees?

Está claro que es necesario tener algunos conceptos matemáticos claros para entender el Big Data y la Ciencia de Datos pero...

¿Qué nivel de caminar necesitas antes de saber correr?

Habrás visto bebés tambaleándose mientras caminan y cogiendo una velocidad tal que sus padres tienen que esforzarse para alcanzarlos. Probablemente también has visto que muchos de esos sprints acaban en accidente.

Lo que seguramente no has visto es a un bebé sentado mientras sus padres le explican la teoría y la biomecánica que hay detrás de caminar.

**LA MANERA NATURAL DE
APRENDIZAJE ES LA PRÁCTICA.**



No te preocupes en exceso por las matemáticas. Podrás aprender lo necesario por el camino. Comenzar por aprender todos los conceptos estadísticos, todo el cálculo y todas las mates antes de dar pasitos en la Ciencia de Datos te dejará sin energía para el resto.

Además, puedes implementar modelos de Machine Learning sin tener claros los conceptos matemáticos que hay detrás pero no te servirán de nada las mates si no sabes programar.

Comienza aprendiendo algo de código e investiga las matemáticas después. Las necesitarás en algún punto del camino.

ESTRATEGIA BASADA EN LA PRÁCTICA

Con esta estrategia comienzas a resolver problemas sin tener ni idea de lo que estás haciendo. Aunque mediante esta estrategia te has quitado un cuatrimestre de matemáticas del camino, probablemente generarás problemas intermedios durante el proceso de aprendizaje.

Nadie dijo que fuera a ser fácil.

Está claro que esta estrategia no es válida para cualquier campo... Imagínate un cirujano aprendiendo medicina practicando directamente sobre pacientes 🤖

Por suerte, en Big Data y Data Science sí se puede llevar a cabo sin poner en peligro la vida de nadie...

La idea es adquirir los conocimientos mínimos y ponerlos en práctica sin esperar a ser un crack de las matemáticas o dominar la teoría detrás del deep learning.



Al empezar a implementar pequeños proyectos enseguida te chocarás contra el muro de todo lo que no sabes. Al principio será un muro altísimo pero no te abrumes. Busca la siguiente pieza que necesitas, el concepto que te falta para continuar, apréndelo y prosigue.

De la misma manera que no hace falta tener conocimientos de mecánica para

aprender a conducir, tampoco es necesario dominar toda la teoría antes de comenzar a trastear en el mundo de los datos.

Lo que sí vas a necesitar es conseguir unas bases de programación. El lenguaje que más te guste, aunque te recomiendo Python por ser el más extendido en Ciencia de Datos.

TÍTULOS Y CERTIFICADOS

Los ninjas de los datos no necesitan certificados. Buscan el conocimiento. Persiguen la maestría de habilidades.

No conviertas tu aprendizaje en una caza de títulos y certificados.

Recuerda que los cursos pueden ayudarte a adquirir los conocimientos fundamentales

pero trabajando en tus propios proyectos potenciarás habilidades específicas que no se pueden enseñar y fijarás los conceptos aprendidos en los cursos.

1. En vez de apilar certificados, muestra lo que sabes hacer a través de proyectos
2. En lugar de hacer más cursos, repite las lecciones que no te hayan quedado claras de alguno que ya hayas hecho.
3. En vez de buscar más recursos de aprendizaje en Internet vuelve a leer tu libro favorito de Big Data y Machine Learning.

Sin embargo, es cierto que en la mayoría de entrevistas de trabajo te pedirán títulos y certificados. Es esperanzador que en algunas empresas esto está cambiando pero aún son muchas las oportunidades que



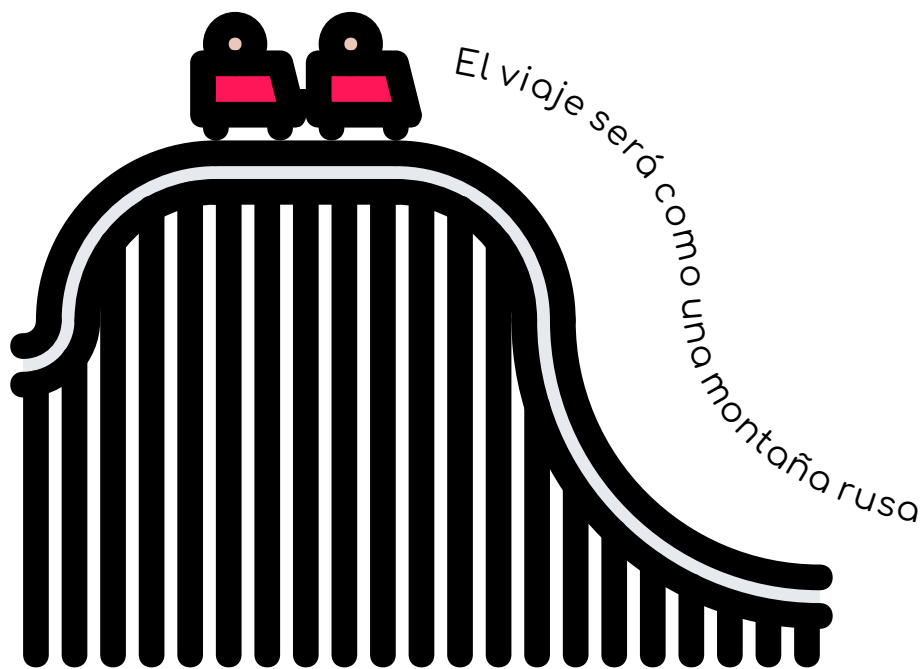
tienen como primer filtro un certificado .

La estrategia que te propongo es comenzar por los conocimientos, desarrolla tu propio plan de aprendizaje Big Data en sincronía con tus intereses y muestra los proyectos que has desarrollado a partir de tus habilidades.

Después busca una certificación si la necesitas. Te resultará más fácil obtenerla y la aprovecharás mejor.

Recuerda, sólo si la necesitas.

ETAPAS DE APRENDIZAJE



¿ESTÁS PREPARADO?

1. En la primera etapa del viaje aprenderás los fundamentos del Big Data y la Ciencia de Datos. Comprender la idea a alto nivel no es demasiado complicado así que empezarás a entender de qué va esto de sacar valor de los datos y por qué es tan interesante. Esta guía ninja ha sido creada para ayudarte en esta primera etapa. Cuando pasas de no saber nada a entender los fundamentos tendrás la sensación de ser imparable.

2. Pero mientras adquieres una visión global del universo Data Science empezarás a ser consciente de la **infinidad de este universo**. De lo poco que sabes y de lo mucho (MUCHO) que te queda por aprender.

🧐 Esta etapa es la más peligrosa, hay mucha gente que llega a abandonar. Te sentirás sobrepasado por todo lo que te falta por aprender. Aquí es donde hay que **perseverar y ser consistente**

- a. Recuerda que vale más **aprender un poco cada día** que querer hacer demasiado muy rápido, quemarse y abandonar.
 - b. Te será de utilidad intentar **dividir los problemas** en otros más pequeños que te resulten más manejables.
 - c. **PACIENCIA**
4. **Continúa aprendiendo. Continúa practicando.** De vez en cuando mira hacia atrás para darte cuenta de dónde vienes y todo lo que has avanzado. **Práctica, práctica y práctica.**
 5. En el campo del Big Data y la Ciencia de Datos tienes que ser consciente de que nunca dejarás de aprender. **Todos los maestros saben que son aprendices perpetuos.**

5 MANDAMIENTOS Y 10 ERRORES A EVITAR

Tenemos clara la estrategia a seguir. Sabemos que el viaje será largo y que habrá piedras en el camino.

Lo siguiente que necesitamos son unas pocas pautas para no perdernos:

Los 5 mandamientos para convertirnos en ninjas de los datos.



1.

Define bien el problema que quieres resolver.

La finalidad última del Big Data y la Ciencia de Datos es aportar valor en el negocio en el que se quiere aplicar.

Puedes pasar meses recolectando y preparando los datos, diseñar el modelo más potente y acabar concluyendo que a nadie le importa la información que has obtenido.

Es más importante hacer las preguntas adecuadas que buscar respuestas complejas a preguntas que nadie se ha hecho.

Pasa un tiempo aclarando las necesidades de tu cliente o de tu jefe, definiendo el problema que estás intentando resolver. Después ponte en acción.

2.

Pasa tiempo con tus datos.

La mayoría de la gente pasa la mayor parte del tiempo desarrollando un modelo complejo, tuneando dicho modelo hasta la perfección. Sin embargo, el valor está en los datos. Aprende a conocerlos y te darán contexto sobre el problema que intentas resolver. Una de las mejores maneras es aplicar técnicas de visualización. También puede ser buena idea hablar con los expertos en el campo del que proceden los datos.

¿Son datos financieros? Habla con un experto en finanzas.

¿Tus datos son radiografías torácicas? Interroga a un médico experto en la materia.

Tal vez tú mismo seas un experto en alguna materia y según vayas aprendiendo sobre Big Data se te irán ocurriendo preguntas que poder responder en tu dominio.

3.

Aprende a comunicar la historia que te cuentan los datos.

Una vez que exprimas al máximo la información que tienen tus datos, tendrás claro el valor que puede aportar la solución basada en Big Data que has desarrollado para tu sector.

¿Lo tendrán claro los demás? ¿lo tendrá claro la persona

encargada de aprobar el presupuesto?

La comunicación será tal vez la habilidad más importante que tendrás que adquirir. Si eres tú la única persona consciente del valor de la solución, no tendrá ningún valor. Es importante que seas capaz de transmitir la importancia de los datos al resto de personas implicadas en el negocio.

4.

Comienza a resolver problemas cuando no estés preparado.

Si cuando empieces a desarrollar tus propios modelos de Machine Learning, a desplegar tus propias aplicaciones te sientes preparado es que has esperado demasiado tiempo para ponerte en marcha.

No esperes a sentirte preparado para enfrentarte a tu primer problema de Machine Learning.

5.

Busca el equilibrio entre teoría y práctica.

Imagínate que aprender Data Science fuera como respirar. Seguir tutoriales, leer libros y hacer cursos es inhalar. Crear tus propios proyectos es exhalar.

Para mantenerte saludable tendrás una respiración equilibrada. No aguantarías la respiración siguiendo muchos tutoriales y cursos sin crear nada, ni vivirías en apnea trabajando en proyectos sin afianzar los conceptos con un poco de teoría. Los ninjas de los datos buscan el equilibrio.

Una vez tengas claros los 5 mandamientos del ninja de los datos, abre bien los ojos para no caer en los errores más comunes en el aprendizaje y aplicación del Data Science. Muchos de ellos ya los hemos ido tratando a lo largo de esta guía.



1.

Para aplicar Machine Learning se necesita de una infraestructura alrededor del modelo.

Ya he comentado en esta guía que la mayoría de la gente se centra en desarrollar los modelos más potentes de Machine Learning para obtener la mayor precisión posible. Parten de un dataset de pruebas y se encierran en una habitación a entrenar el mejor modelo posible para ese problema. Pero... ¿cómo aplicamos ese modelo en el mundo real? No caigas en el error de dejar de lado toda la infraestructura necesaria alrededor del modelo para llegar a resolver un problema en el mundo real. No dejes que el modelo muera en un Jupyter Notebook (más sobre estos notebooks en un momento)

2.

Resolver un problema para el que aún no existen datos.

Un problema muy común en el mundo profesional: Te plantean un proyecto, haces las preguntas correctas y te haces una idea de cómo vas a conseguir las respuestas. Te vas a poner manos a la obra y descubres que aún no hay datos.

Hay que replantear el proyecto y comenzar por el principio: qué datos necesitamos y cómo conseguirlos.

3.

Asumir que los datos están preparados para ser utilizados.

Otro escenario muy común cuando te plantean un proyecto de Data Science es el siguiente:

[Cliente] Hola, mira, tenemos aquí unos datos y queríamos sacar valor de ellos...

[Data Scientist] Oh sí... sin problemas, me pongo con ello.

(Data Scientist se pone a desarrollar directamente un modelo a partir del dataset que le pasa su cliente)



Los datos no estaban preparados o tal vez no eran útiles para el problema que se quería resolver. Que una organización se haya dedicado a acumular datos a lo largo del tiempo sin una estrategia clara con la esperanza de poder aplicar técnicas de Big Data no garantiza que esos datos sean adecuados para el problema que quieren resolver. Mucho menos que estén listos para utilizarse. Un ninja de los datos pasa gran parte de su tiempo limpiando y preparando sus datos. Cuidando de ellos.

La recolección y preparación de datos siempre se infravalora y encontrarás mucha fricción en esta parte del proyecto

4.

Vas a lanzar un nuevo producto que implementa Machine Learning y tienes tanto hype que se te nubla el juicio y crees que el valor añadido está en el Machine Learning por sí mismo.



Has trabajado mucho en añadir Machine Learning a tu producto y el Machine Learning está muy de moda. El Machine Learning es lo que le da calidad a tu producto. Siento cortarte el rollo pero en realidad no.

Al usuario final no le importa si un producto implementa Machine Learning o no. Simplemente le interesa que ese producto les sea de utilidad. La mayor parte de las veces el usuario final no es consciente de qué soluciones implementan Machine Learning y cuáles no.

5.

Optimizar el modelo de Machine Learning con el foco equivocado.

He hablado antes de la importancia de definir bien el problema antes de ponerte manos a la obra. De tener claras las preguntas que tienen que responder los datos. Pues bien, otro punto muy importante es pensar muy bien los parámetros según los cuales vamos a

optimizar el modelo de Machine Learning. Imagínate que Google optimizara su modelo de búsqueda intentando maximizar el número de clicks en una determinada web para un término de búsqueda. A priori no parece una aproximación muy loca. Sin embargo, esto provocaría que si un usuario busca algo en Google y Google le muestra un sitio web con un contenido irrelevante en los resultados de búsqueda pero con un título llamativo el usuario entraría en ese resultado y enseguida saldría de la página web provocando un rebote, ya que el contenido de la web no era lo que buscaba. Como el modelo de Machine Learning del buscador está optimizado según el número de clicks sin tener en cuenta los rebotes, la web sin interés seguiría apareciendo en los resultados de búsqueda.

6.

Pensar que los modelos de Machine Learning se entrenan únicamente una vez.

Bueno, ya vas teniendo claro el proceso...

Planteas el problema, realizas las preguntas correctas para resolver dicho problema y buscas el parámetro correcto para optimizar el modelo de Machine Learning. Analizas los datos, los preparas y los limpias. Seleccionas el modelo de Machine Learning que mejor se adapta al problema. Lo entrenas y lo validas. Funciona perfectamente en el dataset que has usado.

Ha sido un arduo trabajo pero ya lo tienes 🤔

Siento ser yo quién te diga que en realidad solo llevas el 10% del trabajo hecho... **El modelo tendrá que ser re-entrenado muchas veces** y va a ser importante que ese re-entrenamiento se realice de la manera más fácil y eficiente posible.

7.

Crear que algoritmos más complejos llevan a mejores resultados.

Hay una manera muy sencilla de distinguir a un novato de un ninja de los datos. El novato siempre intentará aplicar el modelo más complejo a cualquier problema.

Recuerda que la solución al reto de Netflix de mejorar su rendimiento un 10% nunca llegó a ser implementada. Muchas veces es mejor utilizar un modelo pre-entrenado a construirlo desde cero. Hay que analizar en detalle el coste del modelo respecto a las ganancias que nos va a aportar 🤔

8.

Dejar de lado la visualización de datos.

Querer pasar a la acción demasiado rápido hace que la parte de visualización de los datos sea frecuentemente la primera en sufrir recortes. Un vistazo rápido a los datos y comenzar a trabajar en el modelado.

Muchas veces ganarás tiempo dedicando un poco de esfuerzo extra a la parte de visualización. No solo porque la exploración de los datos será mucho más fructífera y descubrirás patrones que tal vez se te escapen si no utilizas visualizaciones sino también porque al desarrollar habilidades en este campo te será mucho más sencillo, con práctica, transmitir a otras personas lo que has obtenido de los datos.

9.

No tener en cuenta el contexto del problema y no investigar a fondo el dataset.

No es lo mismo tratar con datos tabulares de los que no tienes ningún contexto que poder inferir relaciones entre tus datos porque conoces el contexto del que provienen dichos datos.

Ejemplo sencillo: Tienes dos columnas de datos numéricas. Si sabes que representan la edad y la altura de una persona puedes inferir que una persona de 5 años no puede medir 180 cm. Si no lo sabes, no puedes detectar registros de datos de mala calidad en tu dataset.

Y esto es un ejemplo sencillo porque todos sabemos que un humano de 5 años no puede medir 180 cm pero ¿y si se tratara de relaciones entre datos en un campo más complicado con el que no estás familiarizado?

Pasar tiempo aprendiendo un poco sobre el campo del que provienen tus datos te ayudará. Mucho.

Y además siempre estarás aprendiendo cosas nuevas 😊

10.

Pensar que más certificados equivalen a más conocimientos

La caza de certificados y cursos no te llevará a tener más conocimientos.

En serio... He pasado semanas enteras siguiendo cursos y más cursos de los cuales apenas recuerdo el título.

No te voy a negar que es mucho más cómodo que te expliquen las cosas a tener que enfrentarte a ellas sin ayuda y muchas veces, un curso con los fundamentos bien explicados te ahorrará tiempo.

Sin embargo, así solo conseguirás una falsa sensación de conocimientos. Crees que sabes pero en realidad no. En realidad en cuanto rascas un poco la superficie, te das cuenta de todo lo que queda por aprender.

Busca aplicar los conocimientos en cuanto sea posible para así poder afianzarlos.

El resto siempre se olvida.

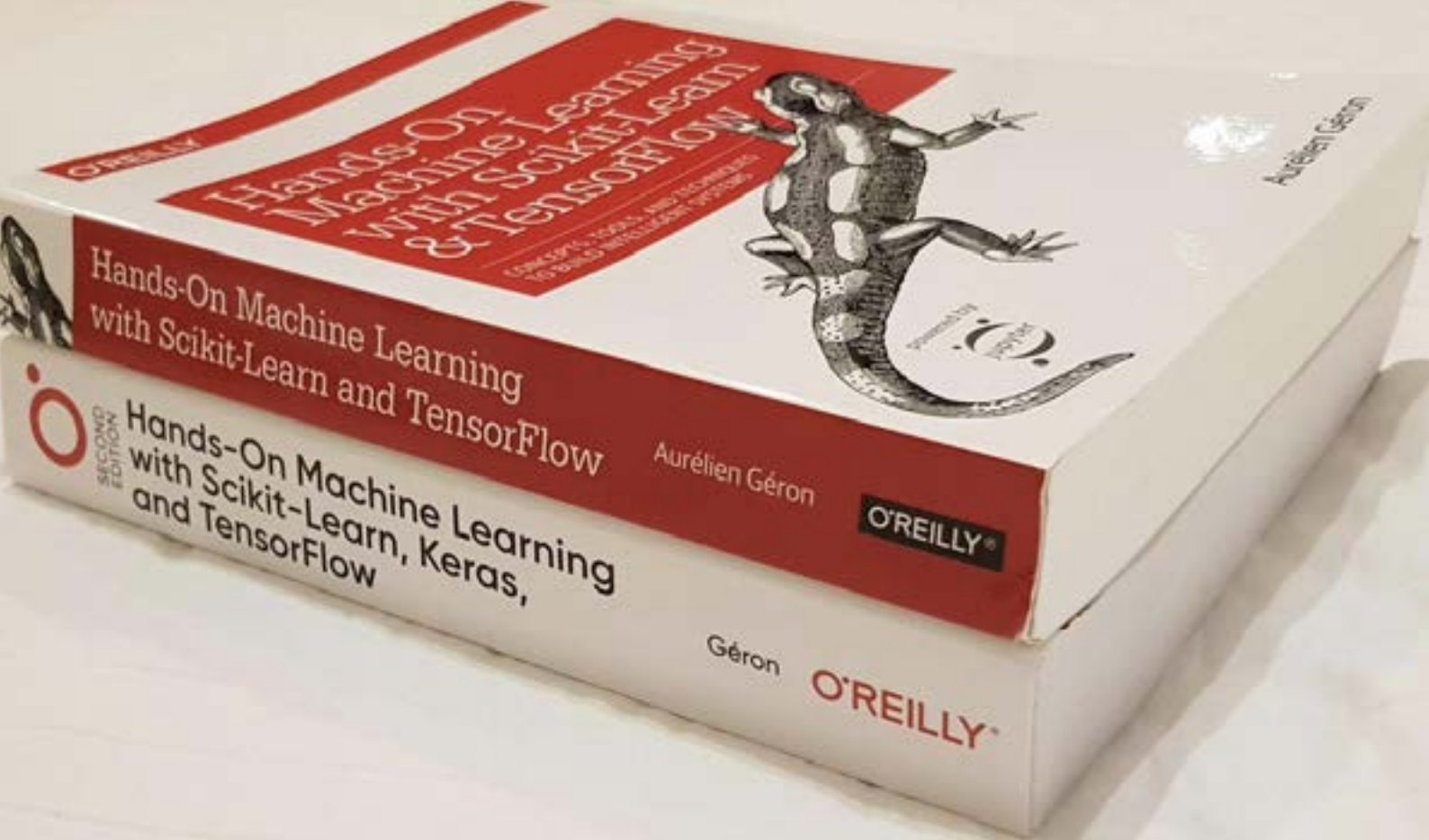
LIBRO RECOMENDADO

Hay muchos estilos de aprendizaje y cada uno se adapta mejor a un tipo de persona.

Tal vez asimiles los conceptos más fácilmente a través de vídeos o de diagramas. Tal vez lo tuyo es escuchar podcasts y absorber el conocimiento a través de los oídos. Hay gente que aprende mejor con la práctica (parte de la estrategia que defiendo en esta guía) y gente que estudia mejor leyendo o escribiendo.

Si te identificas con este último grupo, el libro que estás buscando es:

HANDS-ON MACHINE LEARNING WITH SCIKIT-LEARN, KERAS, AND TENSORFLOW



El libro introduce las técnicas de Machine Learning desde un enfoque práctico a través de pequeños proyectos y ejemplos. Además, el autor hace simples los conceptos complicados mediante el uso de analogías.

No sé tú pero yo siempre he aprendido mejor con ejemplos.

Se divide en dos partes:

- La primera trata los modelos de Machine Learning clásico. Introduce los fundamentos de clasificación y regresión, aprendizaje no supervisado y árboles de decisión. En esta parte manejarás librerías como Scikit Learn, Numpy y Pandas.
- La segunda parte se centra en redes neuronales, deep learning utilizando las librerías de Tensorflow y Keras.

Y recuerda que es un libro con un enfoque práctico. Ya en el capítulo 2 muestra un proyecto de aprendizaje automático completo.

Predecir el precio de la vivienda en California

... por si estás pensando en mudarte allí y comprar una casa 😎



HABILIDADES Y HERRAMIENTAS

Llegados a este punto, idealmente tendrás un poco más claro en qué consiste el Big Data, el Data Science y el Machine Learning y cómo comenzar tu camino en este campo. Un camino que será largo y que tendrá sus momentos menos buenos.

Da igual, tú tienes las cualidades perfectas para convertirte en un ninja de los datos: **Paciencia y tenacidad.**

Está claro que los datos es lo importante, que pasarás tiempo con ellos y que no caerás en la adicción a los certificados. Un puñado de certificados está bien y algún curso te ayudará a asentar las bases, pero mejor pulir las habilidades mediante proyectos propios.

¿Qué habilidades? Te preguntarás.

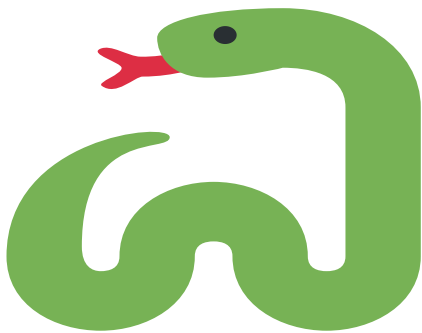
Pues dependiendo de hacia dónde quieras dirigir tu vida, profundizarás en alguna o varias de las habilidades que te cuento a continuación.

Eso sí, tendrás nociones de casi todas y cuanto más sepas de cada una más te acercarás a ser un maestro ninja de los datos.

PYTHON, ANÁLISIS DE DATOS Y MACHINE LEARNING

Lo primero que necesitarás para entrar en el mundo de la Ciencia de Datos es un lenguaje para comunicarte. Un lenguaje de programación. Da un poco igual cuál.

El más extendido en la actualidad es Python.



Aunque cada vez hay más herramientas no-code en el mundo de la ciencia de datos, aún hay que saber programar para realizar las tareas necesarias en el día a día de un data scientist... Explorar un dataset, crear visualizaciones, entrenar un modelo, etc...

No es necesario dominar todo Python antes de entrenar tu primer modelo (en el caso de que te decidas por Python). Comienza por algo sencillo pero fundamental:

- Tipos de datos y sus operaciones más importantes
- Implementar condicionales y bucles para controlar el flujo de tu programa
- Funciones

Después de esto, la librería **Pandas** te ayudará a trabajar con tus datos en formato de tabla o dataframes (parecido a como lo harías con un Excel). La librería de **Numpy** te proporcionará herramientas para realizar operaciones numéricas con tus datos y podrás empezar a visualizarlos mediante las librerías **Matplotlib** y **Seaborn**.

En seguida estarás listo para tus primeros modelos de machine learning con **scikit-learn**.

Para afianzar esta etapa también puedes leer la primera parte del libro *Hands-On Machine Learning* en la que se explica el machine learning clásico.

BASES DE DATOS, LENGUAJES SQL Y NOSQL

Antes de comenzar con el deep learning y las redes neuronales...

¿Sabes que en las empresas la mayor parte de los datos se almacenan en bases de datos? Generalmente, cuando aprendes data science, los datos son sencillos y manejables y ya vienen casi listos en CSV para no darte muchos quebraderos de cabeza... En el mundo real lo primero que vas a necesitar es obtener tu dataset y, como digo, muchas veces será desde una base de datos.

Pueden presentarse dos opciones:

1. Que los datos estén almacenados en bases de datos relacionales: Los datos se guardan en tablas, organizados en filas y columnas. Cada fila representa una instancia de la tabla (por ejemplo los datos de un usuario concreto) y cada columna una característica (todos los valores de todos los usuarios correspondientes aun campo determinado, por ejemplo, el teléfono).

2. Que estén organizados de otra manera...

En el primer caso necesitarás aprender SQL y escribir las consultas (queries) correctas para llegar a conocer tus datos.

Mediante SQL podrás:

Seleccionar datos de la tabla en una base de datos relacional

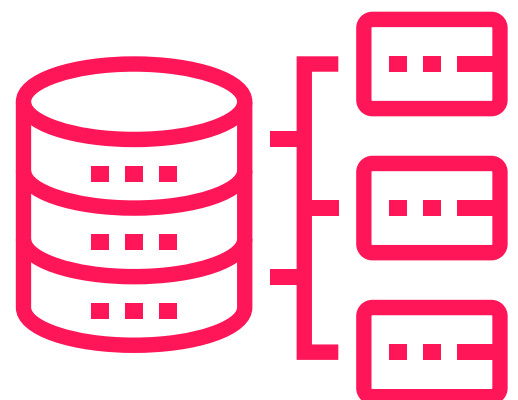
Unir información almacenada en distintas tablas

Agrupar los datos según el valor de cierta columna

Filtrar los datos y obtener los de tu interés

Para el segundo caso hay más variantes. Bases de datos en las que la información está estructurada en documentos, en columnas en lugar de en filas o bases de datos que mapean claves con valores.

Comienza por aprender algo de SQL 😊



ARQUITECTURA DE COMPUTACIÓN DISTRIBUIDA

Uno de los valores fundamentales del Big Data es precisamente la parte de tratar con un volumen de datos muy elevado. Esto hace que se pueda obtener más valor para el negocio que lo implementa.

El problema de tener tantos datos es dónde guardarlos y cómo procesarlos.

Para ello, y especialmente si te quieres enfocar hacia el perfil de **Ingeniero de datos**, es importante tener conocimientos de las herramientas que permiten construir la infraestructura para almacenar y procesar el Big Data.

ECOSISTEMA HADOOP ●●●

Hadoop nace como alternativa a necesitar un superordenador inmenso, ultrapotente y muy caro (en el caso de que existiera...😓) para poder almacenar y procesar Big Data.

Como en un entorno Big Data no es suficiente con un único servidor para almacenar y procesar todos los datos, la solución es utilizar más servidores.

De esta manera, se puede dividir la totalidad de los datos en conjuntos más pequeños y repartirlos entre todos los ordenadores de los que se dispone.

Además, se paralelizan las tareas y el procesamiento de los datos entre todos los ordenadores disponibles en un cluster (que es como se llama al conjunto de ordenadores que tenemos disponibles).

Básicamente se utiliza la estrategia de divide y vencerás.

Con esta idea en mente nace Hadoop.

Hadoop está formado por varios componentes:

- HDFS (Hadoop Distributed File System), encargado de dividir los datos en trozos más digeribles y distribuirlos entre los nodos del clúster.

- Yarn, dedicado a la administración de recursos del clúster.
- MapReduce para el procesamiento en paralelo de los datos.

Estas herramientas se utilizan cuando la gestión de los datos se realiza en las propias instalaciones de la empresa u organización que quiere implementar Big Data. Es decir, es la propia empresa la encargada de adquirir, mantener y gestionar los servidores que conformarán los nodos de su clúster.

LA NUBE ●●●

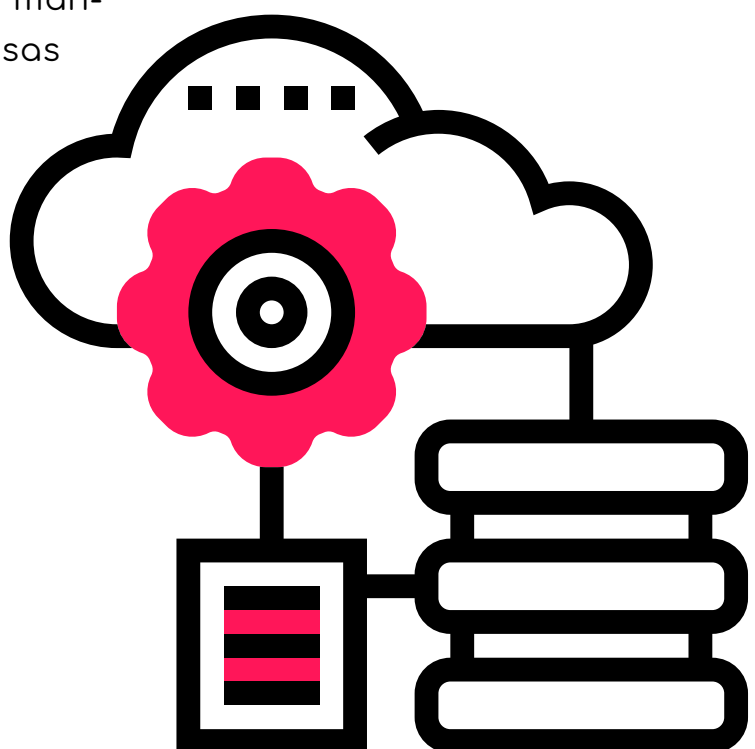
En los últimos años la tendencia está siendo mover estos recursos a la nube. Lo cual no es más que **utilizar los ordenadores de otro**.

Las tecnologías cloud se basan en subcontratar la gestión de los clusters a una empresa externa, que es la que se encarga de adquirir, mantener y gestionar el clúster para montar la infraestructura de Big Data.

Las soluciones más populares en la nube son las de Amazon (AWS), Google (GCP) y Microsoft (Azure). Tal vez te suenen...

Estas empresas han desarrollado sus propias herramientas para el almacenamiento y procesamiento de Big Data.

En realidad, son distintas maneras para implementar cosas similares por lo que solo tendrás que elegir si quieres aprender cómo funciona la nube de Google, la de Amazon, la de Microsoft o directamente la herramienta del ecosistema Hadoop y después será “relativamente rápido” pasar de una a otra.





HARDWARE NECESARIO Y PREPARACIÓN DEL ENTORNO

Hace tiempo hice un curso de Big Data presencial. Había comenzado con algún curso online y con la programación en Python pero me faltaba mucho contexto.

Pasaba muchísimo tiempo navegando por las ininidad de recursos e información que hay en Internet, intentando averiguar qué camino seguir para continuar aprendiendo.

Era un caso agudo de parálisis por análisis.

Así que decidí apuntarme a una formación después del trabajo que me sirviera de guía para conocer las posibilidades de la ciencia de datos.

Para seguir el curso era necesario contar con un ordenador con unos requisitos mínimos de almacenamiento, RAM y capacidad de procesamiento. Alguno de mis compañeros compró un ordenador específicamente para la ocasión y yo aproveché el portátil del trabajo.

¿Qué equipo necesitas entonces para comenzar a aprender Data Science o Machine Learning?

PUES EN REALIDAD, PARA EMPEZAR TE SIRVE CUALQUIERA.

No necesitas adquirir un ordenador con una capacidad de procesamiento brutal y la tarjeta gráfica más potente del mercado.

Con un ordenador normal y conexión a internet tienes suficiente. Google pone el resto.

Google pone a disposición de los ninjas de los datos [Google Colab](#).

Una cosa importante: Google Colab es gratis.

Google Colab es un servicio que Google pone a tu disposición para poder ejecutar cuadernos de Jupyter (o Jupyter Notebooks) a través del navegador web.

No hace falta instalar nada.

Los cuadernos de Jupyter han convertido la programación en una experiencia interactiva ya que es posible ejecutar celdas de código y obtener resultados intermedios como gráficos, imágenes, valores de variables, etc... con tan solo darle al play.

Además, son muy fáciles de compartir, no dejan de ser documentos HTML, y potencian que el usuario añada texto explicativo sobre lo que hace cada pequeña celda de código. Esto hace que estos cuadernos estén muy extendidos en el mundo de la ciencia de datos y que sean la manera más sencilla de iniciarse en este mundo (y te aseguro que seguirás utilizándolos durante muuucho tiempo).

Volviendo a Google Colab...

Tenemos a nuestra disposición un cuaderno de Jupyter listo para usar, que además se va a ejecutar en los servidores de Google de manera completamente transparente para nosotros y de manera gratuita.

Y no solo eso...

Google también pone a nuestra disposición las codiciadas GPUs para que podamos entrenar modelos de deep learning más avanzados de una manera más eficiente que en una CPU convencional.

Cuando comiences a usar los cuadernos de Colab podrás almacenarlos y compartirlos como compartes cualquier otro documento de Google Drive. Ten en cuenta que cualquier cosa que hayas instalado en la máquina virtual de Google donde se ejecuta tu cuaderno no se salvará.

ES UNA BUENA PRÁCTICA INCLUIR TODOS LOS PASOS DE INSTALACIÓN DE LOS PAQUETES QUE NECESITES PARA EJECUTAR EL CUADERNO.

Instalación de librerías en Google Colab.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Intentamos importar una librería que no está instalada en la máquina desde donde se ejecuta nuestro cuaderno de Colab.

```
[11] import timezonefinder
```

```
-----
ModuleNotFoundError                                Traceback (most recent call last)
<ipython-input-11-34c219114d8a> in <module>()
----> 1 import timezonefinder

ModuleNotFoundError: No module named 'timezonefinder'

-----
NOTE: If your import is failing due to a missing package, you can
manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the
"Open Examples" button below.
-----
```

OPEN EXAMPLES SEARCH STACK OVERFLOW

Como no está instalada, python no la encuentra y obtenemos un error. Tenemos que proceder a instalarla nosotros.

```
[12] !pip install timezonefinder
```

```
Collecting timezonefinder
  Downloading timezonefinder-5.2.0-py36.py37.py38-none-any.whl (43.0 MB)
    |████████████████████████████████████████| 43.0 MB 2.1 MB/s
Requirement already satisfied: numpy>=1.16 in /usr/local/lib/python3.7/dist-packages (from timezonefinder) (1.19.5)
Installing collected packages: timezonefinder
Successfully installed timezonefinder-5.2.0
```

```
[13] import timezonefinder
```

Esta vez podemos importar la librería sin problemas 😊

Es una buena práctica incluir la instalación de todas las librerías que sean necesarias para la ejecución de un cuaderno

Pero el mundo no acaba en Google Colab en lo que respecta a entornos de desarrollo para Data Science y Machine Learning libres de instalar cualquier cosa en tu ordenador.

Deepnote es un Google Colab vitaminado en cuanto a la poder trabajar en equipo.

Si te has embarcado en esta aventura de la ciencia de datos con algún amigo estáis de suerte porque con Deepnote **podéis colaborar en tiempo real sobre los mismos cuadernos**.

En cierta manera, también se pueden compartir los cuadernos en Google Colab y hacer comentarios sobre ellos pero para que le aparezcan al resto del equipo es necesario que refresquen el navegador.

No es lo mismo.

Deepnote tiene una versión

gratuita para equipos de menos de 3 personas y otro punto a su favor es que permite la ejecución de tus proyectos en sus máquinas durante 750 horas al mes... o lo que es lo mismo, un único proyecto ejecutándose de manera continuada durante todo el mes.

Como ves, de momento puedes dejar de romperte la cabeza pensando en qué equipo tienes que tener o qué tienes que instalar en tu ordenador para poder iniciarte en el mundo de los datos.

Puedes empezar con un entorno web sencillo y gratuito que no requiere ninguna instalación de nada.

Simplemente tener una cuenta de Google o darte de alta en Deepnote.com.

Fácil, ¿no?



CONCLUSIÓN

Gracias por llegar hasta aquí.

Espero que esta guía te haya sido de utilidad y que te haya ayudado a trazar un pequeño mapa para que puedas seguir caminando en este apasionante mundo de los datos.

Quizás ya sabes que el procesamiento de datos masivo se considera la cuarta revolución industrial y ya está entre nosotros. El Big Data está cambiando el mundo tal y como lo conocemos: la manera en la que nos relacionamos, en la que consumimos y los puestos en los que trabajamos.

El impacto es claro.

Así que entender las tecnologías que conforman el ecosistema Big Data es y será clave para acceder a los puestos de trabajo más demandados.

Es una buena razón para interesarse por la Ciencia de Datos o por el Machine Learning.

Tal vez sea la tuya.

Pero creo que la clave para convertirte en un ninja de los datos y de la vida es la pasión a la hora de aprender Data Science y poder aplicarlo a cualquier campo en el que quieras trabajar.

Tienes entre manos un reto enorme por lo que el objetivo debe ser el progreso y no la perfección.

Convertirte en un ninja de los datos no va a suceder de la noche a la mañana sino que necesitarás tiempo, esfuerzo, energía, paciencia y compromiso.

Para ahorrarte alguna que otra frustración, que las habrá, cuando vayas a aprender un concepto concreto puedes probar a buscar entender la idea general rápido mediante pequeños cursos o tutoriales muy enfocados en esa idea concreta y construir el resto a partir de ahí (preferiblemente con proyectos personales).

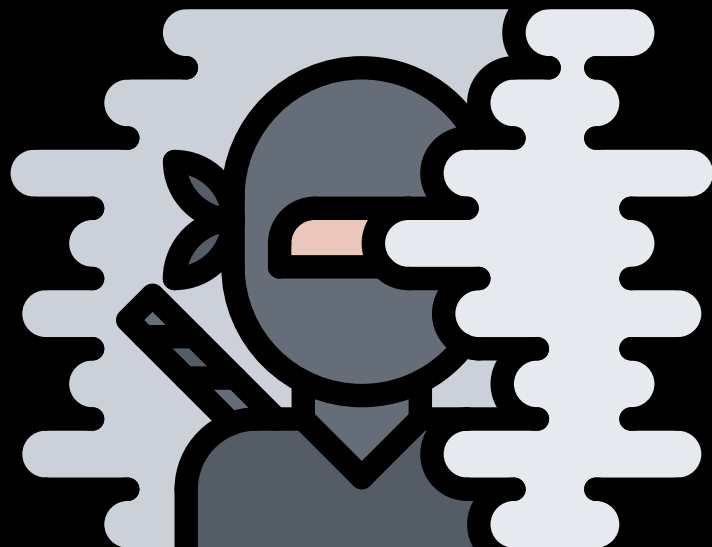
Y sobre todo, no te preocupes por no haber hecho un grado de Ciencia o Ingeniería de Datos en la Universidad ya que, como dijo Albert Einstein:

“LA SABIDURÍA NO ES UN PRODUCTO DE LA EDUCACIÓN, SINO DEL INTENTO DE ADQUIRIRLA DURANTE TODA LA VIDA.”

— ALBERT EINSTEIN

**IMPRIME ESTA GUÍA,
MANDASELA A ALGUIEN,
COMPÁRTELA EN DONDE
QUIERAS.**

**LLENEMOS EL MUNDO DE
NINJAS**



WWW.DATOS.NINJA