Platzi

Glosario para entender el mundo de

Data Science e Inteligencia Artificial

Algoritmo	Dentro de matemáticas y programación es una serie de instrucciones o reglas definidas con un orden y cantidad determinada de pasos para realizar algún cómputo, procesar datos y resolver problemas.
Algoritmo de machine learning	En machine learning existen diferentes algoritmos que sirven para generar modelos entrenados con los que se puede predecir información.
Almacenar datos	Se refiere a guardar y resguardar datos informáticos dentro de una base de datos para garantizar tener acceso a ellos.
API	API o interfaz de programación de aplicaciones es un conjunto de procesos escritos previamente dentro de librerías de programación.

Backend	Se refiere a la capa de desarrollo web donde se ejecutan procesos de acceso a datos de una aplicación web. Esto no es visible a usuarios de la app y se ejecuta en servidores.
Base de datos OLAP	Es el acrónimo en inglés de Procesamiento Analítico en Línea. Son bases de datos especializadas para generar consultas agilizadas de grandes cantidades de datos.
Base de datos OLTP	Sus siglas se refieren a Procesamiento de Transacciones En Línea (OnLine Transaction Processing). Se enfocan al registro y actualización de datos (transacciones). Son las bases de datos que se usan generalmente en todo tipo de aplicaciones donde se encuentran los datos que necesitan estas mismas.
Bases de datos	Es el conjunto de datos pertenecientes a un mismo contexto que son almacenados para su uso en algún sistema de software. Para manejar bases de datos usamos software conocido como motor de base de datos. Algunos comúnes son MySQL y PostgreSQL.

Cómputo en la nube

Son los servicios en la nube para almacenar, administrar y procesar datos, servidores, bases de datos, redes y software. Con “la nube” se refiere a que estos funcionan a través de internet. Servicios que rentamos de Amazon Web Service, Google Cloud Platorm o Microsoft Azure en lugar de tener físicamente un servidor en nuestra empresa.

Cómputo paralelo

Es la forma de cómputo en la que muchas instrucciones se ejecutan simultáneamente dividiendo los problemas grandes en problemas más pequeños.

Data pipeline

Un pipeline de datos es un proceso donde, con diferentes tecnologías, movemos y procesamos datos.
En data science se utilizan para mover datos de bases de datos OLTP a bases de datos OLAP.

Data science

Es un campo que involucra métodos científicos, procesos y sistemas para obtener información a través de análisis de datos y machine learning. Básicamente es el proceso de descubrir información valiosa de los datos.

Datos	En data science se refiere a esos registros de cifras, números o palabras que se almacenan en documentos informáticos y en bases de datos de sistemas informáticos. Son la materia prima con la que se trabaja.
ETL	ETL (Extract, Transform, Load). Es un tipo de pipeline de datos donde extraemos datos de diversas fuentes, los transformamos para poder almacenarlos y los cargamos en bases de datos especializadas para analítica.
Frontend	Se refiere a la capa de desarrollo web donde se crea la parte de la aplicación web a la que usuarios pueden acceder directamente generalmente desde un navegador. Todas las interfaces que ves de Platzi.com son la capa frontend de la aplicación web que te da acceso a todos los cursos.
Insights/Información de valor	Es información que se ha obtenido después de analizar datos. Generalmente sirve para tomar una decisión de negocio, generar estrategias u otra acción en una organización, por lo que se dice que tiene valor. Por ejemplo, puede ser información sobre qué clientes van a dejar de suscribirse a un servicio por ciertas situaciones.

Inteligencia artificial

En informática es la inteligencia expresada por máquinas. Es cuando una máquina imita funciones cognitivas de mentes humanas. Básicamente son algoritmos que hacen que las máquinas aprendan bajo repetición para emular la inteligencia humana natural. Para ello reconocen patrones en grandes cantidades de datos.

Lenguaje de programación

Es un lenguaje que sirve para dar instrucciones en forma de algoritmos a las computadoras.
Dentro del lenguaje vienen configuradas órdenes que escribimos para decirle a la computadora qué hacer.

Librería de programación

Una biblioteca o librería de programación es un conjunto de funciones que otras personas han escrito utilizando un lenguaje de programación. Básicamente es código que otras personas han escrito y que puedes utilizar de forma sencilla dentro de los programas que escribas.

En Python, por ejemplo, existe la biblioteca Pandas que tiene funciones útiles para manipular y analizar datos. Funciones que el lenguaje no tiene por defecto.

Limpieza de datos	En data science se refiere cuando se descubren datos erróneos o faltantes en una tabla o base de datos y se corrigen o eliminan para poder analizarlos.
Machine learning	Machine learning o aprendizaje automático es una rama de inteligencia artificial. Su objetivo es que las computadoras aprendan. En machine learning las computadoras observan grandes cantidades de datos y construyen un modelo capaz de generar predicciones para resolver problemas.
Modelo de machine	Es la salida de información que se genera cuando se entrena un algoritmo de machine learning con datos. Después del entrenamiento a los modelos predictivos podemos darle nuevos datos similares a con los que se entrenó y obtendremos una predicción de salida.
Servidores en la nube	Es un servidor virtual que se ubica en un servidor físico. Este servidor se hospeda con algún provedor de servicio como AWS, Microsoft Azure y Google Cloud Platform. Tiene funcionalidades similares a las de un servidor físico, pero puede ser más rentable ya que para usarlo se renta un servicio y no hay que comprarlo y tenerlo físicamente en una empresas.

Transformar datos

Transformar implica separar datos, limpiar datos nulos, agregar nuevas columnas con nueva información, e incluso cambiarles el formato.

Visualización de datos

Es el campo con el objetivo de representar datos en forma de gráficas. Presentar información de manera visual es más eficiente para comunicar cuando se tienen muchos datos. Con esto se pueden tener insights más rápido, más claros y con ello tomar acciones con mayor eficiencia.

Producción
Sistemas en producción

Poner software en “producción” significa que esa pieza de software está disponible para ser usada por sus usuarios.

El software no se desarrolla directamente donde los usuarios lo utilizan, esto se hace internamente y se prueba antes. Cuando se considera listo para que usuarios lo usan es cuando se pone en producción y es posible usarlo.

Cuando se habla de machine learning en producción se refiere a que el software que utiliza los modelos entrenados funciona dentro de las aplicaciones que usan usuarios.